NOKIA

7450 Ethernet Service Switch
7750 Service Router
7950 Extensible Routing System

Advanced Configuration Guide - Part I
Releases Up To 14.0.R5

3HE 11598 AAAA TQZZA 01

Issue: 01

November 2016

Nokia — Proprietary and confidential.
Use pursuant to applicable agreements.

Nokia is a registered trademark of Nokia Corporation. Other products and company names mentioned herein may be trademarks or tradenames of their respective owners.

The information presented is subject to change without notice. No responsibility is assumed for inaccuracies contained herein.

© 2016 Nokia.

Contains proprietary/trade secret information which is the property of Nokia and must not be made available to, or copied or used by anyone outside Nokia without its written authorization. Not to be used or disclosed except in accordance with applicable agreements.

Table of Contents

Preface **29**
About This Guide 29

Basic System **33**

IEEE 1588 for Frequency, Phase, and Time Distribution **35**
Applicability 35
Overview 35
Configuration 48
Conclusion 68

Synchronous Ethernet **69**
Applicability 69
Summary 69
Overview 70
Configuration 76
Conclusion 82

System Management **83**

Distributed CPU Protection **85**
Applicability 85
Overview 85
Configuration 86
Conclusion 107

Event Handling System **109**
Applicability 109
Overview 109
Configuration 110
Conclusion 122

Interface Configuration **123**

Multi-Chassis APS and Pseudowire Redundancy Interworking **125**
Applicability 125
Overview 125
Configuration 128
Conclusion 143

Multi-Chassis LAG and Pseudowire Redundancy Interworking **145**
Applicability 145
Overview 145

Configuration 147
Conclusion 165

Router Configuration 167

Aggregate Route Indirect Next-Hop Option 169
Applicability 169
Overview 169
Configuration 170
Conclusion 177

Bi-Directional Forwarding Detection 179
Applicability 179
Overview 179
Configuration 181
Conclusion 220

LFA Policies Using OSPF as IGP 223
Applicability 223
Overview 223
Configuration 225
Conclusion 242

Unicast Routing Protocols 243

Associating Communities with Static and Aggregate Routes 245
Applicability 245
Overview 246
Configuration 247
Conclusion 272

BGP FlowSpec for IPv4 and IPv6 273
Applicability 273
Overview 273
Configuration 277
Conclusion 292

BGP Multipath 293
Applicability 293
Overview 293
Configuration 295
Conclusion 316

BGP Route Leaking 317
Applicability 317
Overview 317
Configuration 319

Conclusion 351

IS-IS Link Bundling 353
Applicability 353
Overview 353
Configuration 357
Conclusion 369

Policy Chaining and Logical Expressions 371
Applicability 371
Overview 371
Configuration 375
Conclusion 392

Separate BGP RIBs for Labeled Routes 393
Applicability 393
Overview 393
Configuration 399
Conclusion 419

MPLS 421

Automatic Bandwidth Adjustment in P2P LSPs 423
Applicability 423
Overview 423
Configuration 434
Conclusion 458

Automatic Creation of RSVP-TE LSPs 459
Applicability 459
Overview 459
Configuration 460
Conclusion 477

Class-Based Forwarding 479
Applicability 479
Overview 479
Configuration 480
Conclusion 514

DiffServ Traffic Engineering 515
Applicability 515
Overview 515
Configuration 527
Conclusion 551

Entropy Label 553
Applicability 553

Overview 553
Configuration 556
Conclusion 567

IGP Shortcuts 569
Applicability 569
Overview 569
Configuration 571
Conclusion 620

Inter-Area TE Point-to-Point LSPs 621
Applicability 621
Summary 621
Overview 623
Configuration 624
Conclusion 642

LDP FEC to BGP Label Route Stitching 643
Applicability 643
Overview 643
Configuration 646
Conclusion 671

LDP over RSVP Using OSPF as IGP 673
Applicability 673
Overview 673
Configuration 675
Conclusion 708

LDP Point-to-Point LSPs 709
Applicability 709
Overview 709
Configuration 713
Conclusion 725

LDP-SR Stitching for IPv4 Prefixes (IS-IS) 727
Applicability 727
Overview 727
Configuration 728
Conclusion 742

MPLS LDP FRR using ISIS as IGP 743
Applicability 743
Overview 743
Configuration 744
Conclusion 767

MPLS Transport Profile 769
Applicability 769
Summary 769
Overview 770
Configuration 771
Conclusion 794

Multicast Label Distribution Protocol 795
Applicability 795
Overview 795
Configuration 797
Conclusion 826

RSVP Point-to-Point LSPs 827
Applicability 827
Overview 827
Configuration 831
Conclusion 867

RSVP Signaled Point-to-Multipoint LSPs 869
Applicability 869
Overview 869
Configuration 870
Conclusion 911

Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP 913
Applicability 913
Overview 913
Configuration 915
Conclusion 933

Segment Routing with IS-IS Control Plane 935
Applicability 935
Overview 935
Configuration 937
Conclusion 957

Shared Risk Link Groups for RSVP-Based LSP 959
Applicability 959
Overview 959
Configuration 961
Conclusion 978

Static Point-to-Point LSPs 979
Applicability 979
Overview 979
Configuration 983
Conclusion 989

Tunneling of ICMP Reply Packets over MPLS LSPs 991
Applicability 991
Overview 991
Configuration 1002
Conclusion 1014

Unnumbered Interfaces in RSVP-TE and LDP 1015
Applicability 1015
Overview 1015
Configuration 1024
Conclusion 1060

Services Overview 1061

G.8032 Ethernet Ring Protection Multiple Ring Topology 1063
Applicability 1063
Overview 1063
Configuration 1071
Conclusion 1106

G.8032 Ethernet Ring Protection Single Ring Topology 1107
Applicability 1107
Overview 1107
Configuration 1110
Conclusion 1127

Layer 2 Services and EVPN 1129

BGP Multi-Homing for VPLS Networks 1131
Applicability 1131
Summary 1131
Overview 1133
Configuration 1135
Conclusion 1164

BGP Virtual Private Wire Services 1165
Applicability 1165
Overview 1166
Configuration 1169
Conclusion 1194

BGP VPLS 1195
Applicability 1195
Summary 1195
Overview 1196
Configuration 1197
Conclusion 1228

EVPN for MPLS Tunnels 1229
Applicability 1229
Overview 1229
Configuration 1231
Conclusion 1281

EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS) 1283
Applicability 1283
Overview 1283
Configuration 1285
Conclusion 1311

EVPN for PBB over MPLS (PBB-EVPN) 1313
Applicability 1313
Overview 1313
Configuration 1316
Conclusion 1350

EVPN for VXLAN Tunnels (Layer 2) 1351
Applicability 1351
Overview 1351
Configuration 1353
Conclusion 1377

EVPN for VXLAN Tunnels (Layer 3) 1379
Applicability 1379
Overview 1379
Configuration 1380
Conclusion 1418

EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services 1419
Applicability 1419
Overview 1419
Configuration 1421
Conclusion 1444

Fully Dynamic VSD Integration Model 1445
Applicability 1445
Overview 1446
Configuration 1453
Conclusion 1496

Inter-AS Model C for VLL 1497
Applicability 1497
Overview 1497
Configuration 1499
Conclusion 1515

LDP VPLS using BGP-Auto Discovery 1517
Applicability 1517
Summary 1517
Overview 1518
Configuration 1520
Conclusion 1542

Multi-Chassis Endpoint for VPLS Active/Standby Pseudowire 1543
Applicability 1543
Overview 1543
Configuration 1547
Conclusion 1567

Multi-Segment Pseudowire Routing 1569
Applicability 1569
Summary 1569
Overview 1570
Configuration 1573
Conclusion 1617

P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services 1619
Applicability 1619
Overview 1619
Configuration 1624
Conclusion 1643

PBB-Epipe 1645
Applicability 1645
Overview 1645
Configuration 1647
Conclusion 1664

PBB-VPLS 1665
Applicability 1665
Summary 1665
Overview 1666
Configuration 1667
Conclusion 1700

Shortest Path Bridging for MAC 1701
Applicability 1701
Overview 1701
Configuration 1703
Conclusion 1730

Layer 3 Services 1731

Carrier Supporting Carrier IP VPNs 1733
Applicability 1733
Overview 1733
Configuration 1736
Conclusion 1756

Layer 3 VPN: VPRN Type Spoke 1757
Applicability 1757
Summary 1757
Overview 1757
Configuration 1760
Conclusion 1770

NG-MVPN Configuration with PIM 1771
Applicability 1771
Summary 1771
Overview 1774
Configuration 1776
Conclusion 1817

NG-MVPN Configuration with MPLS 1819
Applicability 1819
Summary 1819
Overview 1821
Configuration 1823
Conclusion 1873

Rosen MVPN Core Diversity 1875
Applicability 1875
Overview 1875
Configuration 1881
Conclusion 1900

Rosen MVPN Inter-AS Option B 1901
Applicability 1901
Overview 1901
Configuration 1909
Conclusion 1926

NG-MVPN Sender-Only, Receiver-Only 1927
Applicability 1927
Overview 1927
Configuration 1929
Conclusion 1974

NG-MVPN Wildcard S-PMSI 1975
Applicability 1975
Overview 1975
Configuration 1979
Conclusion 2003

NG-MVPN Source Redundancy 2005
Applicability 2005
Summary 2005
Overview 2006
Configuration 2008
Conclusion 2037

Spoke Termination for IPv6-6VPE 2039
Applicability 2039
Overview 2039
Configuration 2041
Conclusion 2061

Traffic Leaking from VPRN to GRT 2063
Applicability 2063
Overview 2063
Configuration 2064
Conclusion 2084

VPRN Inter-AS VPN Model C 2085
Applicability 2085
Overview 2085
Configuration 2089
Conclusion 2099

Quality of Service 2101

Class Fair Hierarchical Policing for SAPs 2103
Applicability 2103
Summary 2103
Overview 2104
Configuration 2113
Conclusion 2140

FP and Port Queue Groups 2141
Applicability 2141
Overview 2141
Configuration 2143
Conclusion 2181

Pseudowire QoS 2183
Applicability 2183

Overview 2183
Configuration 2188
Conclusion 2204

QoS Architecture and Basic Operation 2205
Applicability 2205
Overview 2205
Configuration 2206
Conclusion 2245

List of tables

Synchronous Ethernet 69
Table 1 Revertive, Non-Revertive Timing Reference Switching Operation 74

BGP FlowSpec for IPv4 and IPv6 273
Table 2 FlowSpec Component Types 274
Table 3 FlowSpec Extended Community Attributes 275

IS-IS Link Bundling 353
Table 4 Status of the links A, B, C, and D 356

Policy Chaining and Logical Expressions 371
Table 5 Policy Chaining Versus Policy Logical Expressions 372
Table 6 Boolean Values for the Policy Actions 374
Table 7 Actions for the Logical Operators 374
Table 8 Mapping Final Result of an Expression to a Policy Action 375
Table 9 Assigned LP and Communities for the Import Logical Expressions 380
Table 10 Assigned LP and Communities for the Import Logical Expressions 383
Table 11 Assigned LP for the Import Logical Expressions 389

Separate BGP RIBs for Labeled Routes 393
Table 12 Comparison of the BGP RIB Architectures 393

DiffServ Traffic Engineering 515
Table 13 Comparison Bandwidth Constraint Models 522

IGP Shortcuts 569
Table 14 RSVP LSP Role As Outcome of LSP Level and IGP Level Configuration Options 591

Multicast Label Distribution Protocol 795
Table 15 Terminology 796

Segment Routing with IS-IS Control Plane 935
Table 16 Mode Comparison 939

G.8032 Ethernet Ring Protection Multiple Ring Topology 1063
Table 17 Terminology Comparison 1066

BGP VPLS 1195
Table 18 VE-IDs and Labels 1204
Table 19 VE-IDs and Number of Labels 1205

EVPN for MPLS Tunnels 1229
Table 20 Comparing EVPN Multi-homing and BGP Multi-homing 1274

EVPN for PBB over MPLS (PBB-EVPN) 1313
Table 21 EVPN and PBB-EVPN SR OS Feature Comparison 1313
Table 22 PBB-EVPN Multi-Homing Supported Combinations in SR OS 1334

P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services 1619
Table 23 IMET routes and Tunnel Types advertised based on the configuration 1627

NG-MVPN Configuration with PIM 1771
Table 24 Next Generation MVPN Components 1772

NG-MVPN Wildcard S-PMSI 1975
Table 25 S-PMSI Auto-Discovery BGP NLRI 1978

Class Fair Hierarchical Policing for SAPs 2103
Table 26 Burst Levels 2105
Table 27 Policer stat-mode 2114

FP and Port Queue Groups 2141
Table 28 Default QoS and Queue Group Comparison 2142
Table 29 Queue Group Templates - Ingress 2145
Table 30 Queue Group Templates - Egress 2145
Table 31 Network Ingress FP Queue Group Policer Usage 2152

QoS Architecture and Basic Operation 2205
Table 32 SAP Ingress Classification Match Criteria 2209
Table 33 QinQ Dot1p Bit Classification 2211
Table 34 Forwarding Classes 2211
Table 35 Queue Priority vs. Profile Mode 2216
Table 36 Network QoS Policy DSCP Remarking 2222

List of figures

IEEE 1588 for Frequency, Phase, and Time Distribution 35
Figure 1 PTP Messages and Timestamp Exchange 37
Figure 2 1588 Topology for Frequency Distribution 39
Figure 3 1588 Topology for Time Distribution 39
Figure 4 Frequency Distribution with 1588 as Last Mile 40
Figure 5 Unicast Message Negotiation 42
Figure 6 Floor Packet Counting for FPP (n, W, δ) 45
Figure 7 G.8271.1 Time Error Budget 47
Figure 8 Master and Slave Clocks for Frequency 49
Figure 9 Boundary Clock 58
Figure 10 Boundary Clocks with Edge VPRN Access 63

Synchronous Ethernet 69
Figure 11 SyncE Hypothetical Reference Network Architecture 71
Figure 12 Packet Based Network Timing Infrastructure 72
Figure 13 Current 7x50 Timing Sub-System Architecture 73
Figure 14 Network Considerations for Ethernet Timing Distribution 76

Distributed CPU Protection 85
Figure 15 Test Topology 86
Figure 16 Count Traffic with DCP Policy Count 92
Figure 17 Limit Traffic with dcp-static-policy-1 95
Figure 18 Dynamic Policing – Local Monitor 102
Figure 19 Dynamic Policers Instantiated 103

Event Handling System 109
Figure 20 Example Topology 110

Multi-Chassis APS and Pseudowire Redundancy Interworking 125
Figure 21 MC-APS Network Topology 127
Figure 22 Access Node and Network Resilience (Part 1) 127
Figure 23 Access Node and Network Resilience (Part 2) 128
Figure 24 Association of SAPs/SDPs and Endpoints 134
Figure 25 ICB Spoke SDPs and Association with the Endpoints 137
Figure 26 Additional Setup Example 1 (Part 1) 140
Figure 27 Additional Setup Example 1 (Part 2) 140
Figure 28 Additional Setup Example 2 (Part 1a) 141
Figure 29 Additional Setup Example 2 (Part 1b) 141
Figure 30 Additional Setup Example 2 (Part 2) 142

Multi-Chassis LAG and Pseudowire Redundancy Interworking 145
Figure 31 MC-LAG Example Topology 146
Figure 32 Network Resiliency 147
Figure 33 Association of SAPs/SDPs and Endpoints 155

Figure 34 ICB Spoke SDPs and Their Association with the Endpoints 159
Figure 35 Additional Setup Example 1 162
Figure 36 Additional Setup Example 2 163

Aggregate Route Indirect Next-Hop Option 169
Figure 37 Aggregate Routes 170
Figure 38 Example topology 171

Bi-Directional Forwarding Detection 179
Figure 39 BFD Multi-Scenarios 181
Figure 40 BFD Centralized Sessions 182
Figure 41 BFD Interface Configuration 183
Figure 42 BFD for ISIS 188
Figure 43 BFD for OSPF 190
Figure 44 BFD for OSPF and PIM 192
Figure 45 BFD for Static Routes 193
Figure 46 BFD for IES over Spoke SDP 196
Figure 47 BFD for RSVP 201
Figure 48 BFD for T-LDP 204
Figure 49 BFD for OSPF PE-CE I/F 207
Figure 50 BFD Sessions within IPSec Tunnels 210
Figure 51 Logic for Shared BFD Sessions 213
Figure 52 BFD for VRRP 214

LFA Policies Using OSPF as IGP 223
Figure 53 Example Topology 225

Associating Communities with Static and Aggregate Routes 245
Figure 54 Example Topology 246
Figure 55 CE Connections for Next-Hops 249
Figure 56 CE-7 Connectivity 262
Figure 57 CE-6 Connectivity 267

BGP FlowSpec for IPv4 and IPv6 273
Figure 58 Example Topology 276

BGP Multipath 293
Figure 59 Example Topology 295
Figure 60 BGP Multipath with eBGP Limit 2 298
Figure 61 eBGP Multipath with Limit 2 and ECMP Disabled 298
Figure 62 BGP Multipath with iBGP Limit 3 and ECMP Limit 8 300
Figure 63 BGP Multipath with Limit 6 and eBGP Preferred 302
Figure 64 BGP Multipath with Limit 6, eBGP Equal to iBGP, and Other Path Options Identical 304
Figure 65 BGP Multipath Configured with Restriction to the Same Neighbor AS 305
Figure 66 BGP Multipath Restricted to the Same Neighbor AS: AS Paths with Same Length 307

Figure 67 BGP Multipath Restricted to the Same Neighbor AS: AS Paths of Different Lengths 308
Figure 68 BGP Multipath Restricted to the Same Neighbor AS: AS Paths of Different Lengths, AS Path Ignored 309
Figure 69 BGP Multipath Restricted to Exact Same AS. All AS Paths are Different. 311
Figure 70 BGP Multipath Restricted to Exact Same AS. All AS Paths are Identical 312
Figure 71 EBGP Equal to IBGP: No EIBGP Load-Balancing 315
Figure 72 EBGP Equal to IBGP: EIBGP Load-Balancing Enabled 316

BGP Route Leaking 317
Figure 73 BGP Route Leaking Process 318
Figure 74 Example Topology 319
Figure 75 BGP IPv4 Route Leaking Between VPRNs 320
Figure 76 BGP IPv4 Route Leaking from VPRN to GRT 331
Figure 77 BGP IPv4 Route Leaking from GRT to VPRN 336
Figure 78 BGP IPv6 Route Leaking between VPRNs 340
Figure 79 BGP IPv6 Route Leaking from GRT and VPRN to VPRN 344

IS-IS Link Bundling 353
Figure 80 Link Bundle Schematic 354
Figure 81 Effect of Single Link Failure on Bundle Group 355
Figure 82 Double Link Failure 356
Figure 83 Example Topology 357
Figure 84 Link Failure 364
Figure 85 Second Link Failure 367

Policy Chaining and Logical Expressions 371
Figure 86 Example Topology 375

Separate BGP RIBs for Labeled Routes 393
Figure 87 RR-1 with Common IPv4 RIB 395
Figure 88 RR-1 with Separate Labeled-IPv4 RIB Implementation 396
Figure 89 Seamless MPLS - Separate Labeled-IPv4 Implementation 397
Figure 90 System Architecture with Separate RIBs for Labeled-Unicast and Unlabeled Routes 398
Figure 91 Example IPv4 Topology 399
Figure 92 BGP Sessions 400
Figure 93 PE-1 Applies Next-Hop-Self toward Neighbor PE-2 406
Figure 94 Applying Next-Hop-Self to Unlabeled IP-4 Routes to Neighbor PE-2 407
Figure 95 PE-1 Advertises Prefixes 1.1.1.1/32 and 11.11.11.11/32 410
Figure 96 RR with Labeled and Unlabeled BGP Sessions 413
Figure 97 Updates from Unlabeled Sessions Not Propagated to Labeled Sessions (Default) 415
Figure 98 RIB Leaking from IPv4 BGP RIB to Labeled-IPv4 BGP RIB 417

Automatic Bandwidth Adjustment in P2P LSPs 423
Figure 99 Auto-Bandwidth Adjustment Implementation 425
Figure 100 Underflow-Triggered Auto-Bandwidth Implementation 430
Figure 101 Example Setup for Auto-Bandwidth Point-to-Point LSPs 435

Automatic Creation of RSVP-TE LSPs 459
Figure 102 Example Topology 461
Figure 103 IGP Shortcuts with RSVP-TE Auto-Mesh 465
Figure 104 Test Topology for Single-Hop LDP-over-RSVP with ECMP 473

Class-Based Forwarding 479
Figure 105 Test Topology for CBF on RSVP LSPs 481
Figure 106 LSPs with Direct and Indirect Path toward PE-2 483
Figure 107 CBF on RSVP LSPs - Services 485
Figure 108 Test Topology for CBF of LDP Prefix Packets over IGP Shortcuts 493
Figure 109 Different Paths for Different FCs 498
Figure 110 Traffic on Default LSP 504
Figure 111 Traffic with FC EF on Direct Path 505
Figure 112 Traffic on Default LSP with Lowest ID on P-3 507
Figure 113 Traffic on System-Selected Default LSPs with Lowest Tunnel ID 509
Figure 114 Inconsistent CBF Configuration. Revert to ECMP Forwarding 512

DiffServ Traffic Engineering 515
Figure 115 Example Topology 516
Figure 116 Mapping of TE Classes 517
Figure 117 Bandwidth Reservation for the CTs 518
Figure 118 Bandwidth Reservation in Maximum Allocation Model for Three CTs 522
Figure 119 Bandwidth Reservation in Russian Doll Model for Three CTs 523
Figure 120 Paths from PE-1 to PE-3 528
Figure 121 MAM Bandwidth Allocation 531
Figure 122 Reserved and Unreserved Bandwidth 533
Figure 123 Reserved and Unreserved Bandwidth on PE-1 535
Figure 124 Bandwidth Reservation 537
Figure 125 Russian Doll Model for Three Class Types 538
Figure 126 Reserved Bandwidth for LSP with CT2 (One Session) 542
Figure 127 Bandwidth Reservation for LSP with CT2 and LSP with CT1 (2 Sessions) 543
Figure 128 Reserved Bandwidth on Both Interfaces of PE-1 (3 Sessions) 544
Figure 129 Reserved Bandwidth on Both Interfaces on PE-1 (4 Sessions) 545
Figure 130 Reserved Bandwidth on Both Interfaces of PE-1 (5 Sessions) 547
Figure 131 Reserved Bandwidth on Both Interfaces on PE-1 (6 Sessions) 549

Entropy Label 553
Figure 132 Load-Balancing of Flows Based on Hash Label or Entropy Label 554
Figure 133 Label Stack with Hash Label versus Label Stack with EL and ELI 555
Figure 134 Downstream LERs Signal EL Capability to ILER 555
Figure 135 Example Topology 557

Figure 136 RSVP LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via P-3 and P-4 558

IGP Shortcuts 569
Figure 137 Normal SPF Tree Sourced by PE-1 570
Figure 138 SPF Tree Sourced by PE-1 Using LSP Shortcuts 570
Figure 139 Example Topology 571
Figure 140 LSPs Between PE-1 and PE-6 585
Figure 141 RSVP Shortcuts LFA Use Case Example 596
Figure 142 Network Topology to Verify Installation of Shortcuts into RTM 600
Figure 143 Shortcuts Within a VRF Topology Network 615

Inter-Area TE Point-to-Point LSPs 621
Figure 144 Inter-Area TE LSP Setup 623
Figure 145 Inter-Area TE LSP Path 623
Figure 146 ABR Protection 632
Figure 147 Protection of All Nodes/Links Along the LSP Path 632
Figure 148 Admin Group Example 635
Figure 149 Share Risk Link Groups 638

LDP FEC to BGP Label Route Stitching 643
Figure 150 LDP FEC to BGP Label Route Stitching 644
Figure 151 Example Topology 646
Figure 152 BGP Enabled with P-4 as RR 649
Figure 153 End-to-End Epipe Service 655
Figure 154 Label Stacks for Traffic from AN-1 to AN-8 661
Figure 155 Block BGP Label Bindings to LDP DU Peer PE-9 667

LDP over RSVP Using OSPF as IGP 673
Figure 156 Initial Topology 675
Figure 157 VPRN 1 with LDPoRSVP and No Intra-Area PE Connectivity 688
Figure 158 VPRN 1 with LDPoRSVP and Intra-Area PE Connectivity 705

LDP Point-to-Point LSPs 709
Figure 159 Generic MPLS Network, MPLS Label Operations 711
Figure 160 MPLS Example Topology 712

LDP-SR Stitching for IPv4 Prefixes (IS-IS) 727
Figure 161 Example Topology 728

MPLS LDP FRR using ISIS as IGP 743
Figure 162 Initial Topology 745
Figure 163 Data Verification, Direction PE-1 => PE-5 Using VLL Service 755
Figure 164 LFA Computation, Inequality 1 for Prefix PE-5 (D) on PE-1 (S) 762
Figure 165 LFA Computation, Inequality 3 for Prefix PE-5 (D) on PE-1 (S) 762
Figure 166 IS-IS Overload on PE-2, Inequality 1 for 192.168.24.0/30 (D) on PE-1 (S) 767

MPLS Transport Profile 769
Figure 167 MPLS-TP Example Network Showing LSPs 772
Figure 168 MPLS-TP Example Network Showing Services Detail 773
Figure 169 MPLS-TP Configuration Steps 773
Figure 170 LSP Path Label Value Configurations 782

Multicast Label Distribution Protocol 795
Figure 171 Setup of mLDP P2MP LSP 797
Figure 172 Test Topology 798
Figure 173 LDP P2MP LSP 805
Figure 174 New LDP P2MP LSP after Metric Change 822

RSVP Point-to-Point LSPs 827
Figure 175 Generic MPLS Network, MPLS Label Operations 829
Figure 176 MPLS Example Topology 831
Figure 177 LSP with Dynamic Path Takes IGP Best Route 839
Figure 178 RSVP-TE LSP with Dynamic Path Using TE Metric 842
Figure 179 Fast Reroute One-to-One Detour Tunnels 844
Figure 180 Fast-Reroute Facility Bypass Tunnels 848
Figure 181 FRR Facility without Node Protection 851
Figure 182 Admin Groups 'Blue' and 'Red' 853
Figure 183 LSP and Bypass within Admin Group 'Blue' 855
Figure 184 LSP and FRR Bypass Tunnel Excluding Admin Group 'Red' 857

RSVP Signaled Point-to-Multipoint LSPs 869
Figure 185 P2MP Network Topology 870
Figure 186 P2MP LSP LSP-p2mp-1 875
Figure 187 P2MP LSP p-to-mp-1 with Metric Change 893
Figure 188 P2MP LSP LSP-p2mp-1 with Strict S2L Path toward PE-7 896
Figure 189 Intelligent Remerge, Case 1 898
Figure 190 Intelligent Re-merge, Case 2 902
Figure 191 Intelligent Re-merge, Case 3 906

Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP 913
Figure 192 Seamless MPLS - Network Topology, control and data plane 914
Figure 193 Seamless MPLS - IGP/LDP domains 915
Figure 194 Seamless MPLS - BGP 917
Figure 195 End-to-End Epipe Service 922
Figure 196 L3 VPN service 924
Figure 197 Label Stacks for Traffic from AN-1 to AN-5 927

Segment Routing with IS-IS Control Plane 935
Figure 198 Network Topology 937
Figure 199 RLFA Traffic Path During Protection 949

Shared Risk Link Groups for RSVP-Based LSP 959
Figure 200 Example Topology 960
Figure 201 SRLG Topology 962

Figure 202 Path Primary RSVP_TE LSP....968
Figure 203 FRR Bypass Tunnels Originating in PE-1 With and Without SRLG....971
Figure 204 SRLG for Secondary Path....974
Figure 205 SRLG Database Example....977

Static Point-to-Point LSPs....979
Figure 206 Generic MPLS Network, MPLS Label Operations....981
Figure 207 MPLS Example Topology....982
Figure 208 Static LSP Running over PE-1, PE-2, PE-5, PE-6....983

Tunneling of ICMP Reply Packets over MPLS LSPs....991
Figure 209 Use of TTL: Uniform versus Pipe....992
Figure 210 Use of TTL in an L2 VPN Service in Pipe Mode....993
Figure 211 Use of TTL in an L3 VPN Service in Pipe Mode....994
Figure 212 Tunneling of ICMP Reply Packets over an MPLS LSP....994
Figure 213 MPLS Label Stack Object....1000
Figure 214 ICMP Extension Header....1000
Figure 215 ICMP Extension Object: Object Header and Payload....1001
Figure 216 Example Configuration....1002
Figure 217 Tunnel from iLER PE-3 to eLER PE-6 via LSR PE-2....1005
Figure 218 UDP Traceroute in VPRN with iLER in Uniform Mode....1007
Figure 219 UDP Traceroute in VPRN without TTL Propagation to LDP....1009
Figure 220 Tunnel from iLER PE-3 to eLER PE-6 with Multiple LSRs....1010
Figure 221 UDP Traceroute with iLER in Uniform Mode....1011

Unnumbered Interfaces in RSVP-TE and LDP....1015
Figure 222 Example Topology for Unnumbered Interfaces in RSVP and LDP....1016
Figure 223 Configuration Example for Unnumbered Interfaces in RSVP and LDP....1024
Figure 224 LSP-PE-4-PE-2 on Unnumbered Interfaces....1028
Figure 225 LSP and FRR Facility Bypass Tunnels....1035
Figure 226 FRR One-to-One Only Supported on Numbered Interfaces....1038
Figure 227 FRR on iLER: No Bypass on Unnumbered Interfaces....1040
Figure 228 FRR Facility and Admin Groups....1042
Figure 229 SRLG-FRR Strict: No Bypass on PE-4....1044
Figure 230 LDP FRR LFA Link Protection on PE-4....1059

G.8032 Ethernet Ring Protection Multiple Ring Topology....1063
Figure 231 G.8032 Major Ring and Sub-Ring....1067
Figure 232 G.8032 Ring Components....1069
Figure 233 G.8032 Sub-Ring Interconnection Components....1070
Figure 234 Ethernet Test Topology....1075
Figure 235 ETH-CFM MEP Associations....1077
Figure 236 Sub-Ring to VPLS Topology....1099

G.8032 Ethernet Ring Protection Single Ring Topology....1107
Figure 237 G.8032 Operation and Topologies....1110
Figure 238 Example Topology....1111

Figure 239 Ethernet CFM Configuration 1115

BGP Multi-Homing for VPLS Networks 1131
Figure 240 Example Topology 1133
Figure 241 Nodes Involved in BGP MH 1137
Figure 242 MAC Flush for BGP MH 1150
Figure 243 Access PE/CE Signaling 1151
Figure 244 Oper-Groups and BGP-MH 1154

BGP Virtual Private Wire Services 1165
Figure 245 Example Topology 1166
Figure 246 Single Homed BGP VPWS using Auto-Provisioned SDPs 1173
Figure 247 Single Homed BGP VPWS using Pre-Provisioned SDP 1180
Figure 248 Dual Homed BGP VPWS with Single Pseudowire 1184
Figure 249 Dual Homed BGP VPWS with Active/Standby Pseudowire 1190

BGP VPLS 1195
Figure 250 Network Topology 1196
Figure 251 BGP VPLS Using Auto-Provisioned SDPs 1202
Figure 252 BGP VPLS Using Pre-Provisioned SDP 1218

EVPN for MPLS Tunnels 1229
Figure 253 EVPN Route Types and NLRIs 1230
Figure 254 EVPN-MPLS for VPLS Services 1232
Figure 255 EVPN-MPLS All-Active Multi-Homing Concepts 1246
Figure 256 EVPN-MPLS Single-Active Multi-Homing: Mass-Withdraw, Backup Path 1261

EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS) 1283
Figure 257 Route Types and NLRIs for EVPN-VPWS 1284
Figure 258 EVPN-VPWS Example Topology 1286
Figure 259 Example Topology for EVPN-VPWS without Multi-Homing 1288
Figure 260 Example Topology EVPN-VPWS with Multi-Homing 1294

EVPN for PBB over MPLS (PBB-EVPN) 1313
Figure 261 EVPN Route Types 1315
Figure 262 PBB-EVPN Network without Multi-Homing 1316
Figure 263 PBB-EVPN — Flooding Lists 1319
Figure 264 PBB-EVPN Multi-homing 1332
Figure 265 The Use of es-bmac to Minimize CMAC Flush 1333
Figure 266 PBB-EVPN Single-Active Support for Epipes 1348

EVPN for VXLAN Tunnels (Layer 2) 1351
Figure 267 EVPN-VXLAN Example Topology 1354
Figure 268 BGP Adjacencies and Enabled Families 1357
Figure 269 EVPN MAC Mobility 1370

EVPN for VXLAN Tunnels (Layer 3) 1379
Figure 270 EVPN-VXLAN for R-VPLS Services 1381
Figure 271 BGP adjacencies and enabled families 1384
Figure 272 EVPN-VXLAN for IRB Backhaul R-VPLS Services 1389
Figure 273 EVPN-VXLAN in EVPN-tunnel R-VPLS Services 1398
Figure 274 Routing Policies for Egress EVPN Routes 1406
Figure 275 Routing Policies for Ingress EVPN Routes 1407
Figure 276 EVPN in Parallel R-VPLS Services 1411

EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services 1419
Figure 277 EVPN-MPLS Interconnect for EVPN-VXLAN - Example Topology 1422
Figure 278 EVPN Destinations Created on Multi-Homed Anycast DC GWs 1430
Figure 279 Use of Provider-tunnels Between Anycast DC GWs Create Packet Duplication 1442

Fully Dynamic VSD Integration Model 1445
Figure 280 Nuage VSP overview 1446
Figure 281 DC Gateway fully dynamic provisioning workflow 1448
Figure 282 F-D XMPP provisioning setup 1454

Inter-AS Model C for VLL 1497
Figure 283 Network Setup - Inter-AS Model C for VLL 1498
Figure 284 Inter-AS Model C for VLL 1498
Figure 285 Network Setup Configuration 1499

LDP VPLS using BGP-Auto Discovery 1517
Figure 286 Example Topology 1518
Figure 287 VPLS Instance with Auto-Provisioned SDPs 1526
Figure 288 VPLS Instance using Pre-Provisioned SDPs 1537

Multi-Chassis Endpoint for VPLS Active/Standby Pseudowire 1543
Figure 289 H-VPLS with STP 1544
Figure 290 VPLS Pseudowire Redundancy 1544
Figure 291 Multi-Chassis Endpoint with Mesh Resiliency 1545
Figure 292 Multi-Chassis Endpoint with Square Resiliency 1545
Figure 293 Example Topology 1546
Figure 294 Core Node Failure 1561
Figure 295 Multi-Chassis Node Failure 1563
Figure 296 Multi-Chassis Passive Mode 1566

Multi-Segment Pseudowire Routing 1569
Figure 297 FEC129 Structure 1570
Figure 298 AII Type 2 Format 1571
Figure 299 Pseudowire Routing NLRI (the AC ID is always zero) 1571
Figure 300 Configuration Flow Chart 1573
Figure 301 Intra-AS MS-PW Example Topology 1590
Figure 302 Inter-AS MS-PW Example Topology 1603

P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services 1619
Figure 303 P2MP mLDP Tree with Root Node PE-1 and Leaf Nodes PE-5, PE-6, and PE-7.....1620
Figure 304 BGP-EVPN Route Type 3 with PTA.....1621
Figure 305 PTA for Composite Tunnel IMET-P2MP-IR.....1623
Figure 306 P2MP mLPD in PBB-EVPN.....1638

PBB-Epipe 1645
Figure 307 Network Topology.....1646
Figure 308 Setup Detailed View.....1648
Figure 309 Virtual MEPs for Flooding Avoidance.....1656

PBB-VPLS 1665
Figure 310 Example Topology.....1666
Figure 311 MTU-4 and PE-1 Nodes as Configuration Examples.....1668
Figure 312 black-hole.....1680
Figure 313 Send Flush on BVPLS Failure Example.....1683
Figure 314 Inter-Domain B-VPLS and MMRP Policies/ISID-Based Filters Example.....1690

Shortest Path Bridging for MAC 1701
Figure 315 Basic SPBM Topology.....1703
Figure 316 Control and User B-VPLS Test Topology.....1714
Figure 317 Access Resiliency Test Topology.....1718
Figure 318 Access Resiliency Test Topology.....1722

Carrier Supporting Carrier IP VPNs.....1733
Figure 319 CSC Network Topology.....1734

Layer 3 VPN: VPRN Type Spoke.....1757
Figure 320 CE Hub and Spoke Data Path.....1758
Figure 321 CE Hub and Spoke Control Plane Isolation.....1759
Figure 322 Internal VPRN Logic on a PE Router.....1759
Figure 323 CE Hub and Spoke Topology and Addressing Scheme.....1761

NG-MVPN Configuration with PIM 1771
Figure 324 Network Topology.....1774
Figure 325 Network Topology for Anycast RP.....1796
Figure 326 IGMP and PIM Control Messaging Schematic.....1801
Figure 327 PIM SSM in Customer Signaling Plane.....1805

NG-MVPN Configuration with MPLS.....1819
Figure 328 Network Topology.....1822
Figure 329 VPRN 1 Topology used for mLDP.....1827
Figure 330 VPRN 2 Topology used for RSVP-TE P2MP.....1846
Figure 331 VPRN 2 Topology used for MVPN Source Redundancy.....1865
Figure 332 VPRN 3 Topology used for MDT SAFI.....1871

Rosen MVPN Core Diversity 1875
Figure 333 Core Diversity Schematic 1877
Figure 334 Core Diversity Network 1878
Figure 335 Core Diversity Network — Base OSPF 1879
Figure 336 Core Diversity Network - OSPF Instance 1 1880

Rosen MVPN Inter-AS Option B 1901
Figure 337 General Topology for Inter-AS MVPN 1902
Figure 338 Protocols Used for Inter-AS MVPN 1902
Figure 339 BGP Signaling Steps 1905
Figure 340 PIM-P Signaling Steps for Default MDT 1906
Figure 341 PIM-C Signaling 1907
Figure 342 PIM-P Signaling Steps for Data MDT 1908
Figure 343 Example Topology Details 1909
Figure 344 BGP Signaling Steps 1913
Figure 345 PIM-P Signaling Steps for Default MDT 1917
Figure 346 PIM-C Signaling 1920
Figure 347 PIM-P Signaling Steps for Data MDT 1923

NG-MVPN Sender-Only, Receiver-Only 1927
Figure 348 Default PMSI Hierarchy 1928
Figure 349 Optimized PMSI Structure 1928
Figure 350 Example Topology 1929
Figure 351 RSVP-Based BGP Message Flow Between PE-1 and PE-2 1937
Figure 352 RSVP-Based BGP Message Flow Between PE-1 and PE-3 1940
Figure 353 mLDP-Based BGP Message Flow Between PE-1 and PE-2 1955
Figure 354 mLDP-Based BGP Message Flow Between PE-1 and PE-3 1962

NG-MVPN Wildcard S-PMSI 1975
Figure 355 Multicast VPN 1977
Figure 356 Schematic Topology 1980
Figure 357 S-PMSI P2MP LSP Schematic 1999

NG-MVPN Source Redundancy 2005
Figure 358 Source Redundancy Example. 2006
Figure 359 Schematic Topology 2009

Spoke Termination for IPv6-6VPE 2039
Figure 360 Spoke Termination for IPv6 2040
Figure 361 IPv6 Addressing and IPv6 Prefixes 2041
Figure 362 MP-BGP VPN IPv6 2042
Figure 363 Spoke Termination for IPv6 Addressing 2044
Figure 364 PE-4 VPRN with SAP to CE-5 2049

Traffic Leaking from VPRN to GRT 2063
Figure 365 VPRN to GRT leak process 2064
Figure 366 Example Topology with IPv4 Addresses 2065
Figure 367 IPv4 VPRN to GRT route leaking for ISIS 2066

Figure 368 Example Topology with IPv6 Addresses 2076

VPRN Inter-AS VPN Model C 2085
Figure 369 Inter-AS VPN Model C....2087
Figure 370 Protocol Overview 2088

Class Fair Hierarchical Policing for SAPs 2103
Figure 371 Policer Token Bucket Model....2105
Figure 372 Peak Information Rate (PIR) Bucket 2106
Figure 373 Committed Information Rate (CIR) Bucket....2107
Figure 374 Fair Information Rate (FIR) Bucket....2108
Figure 375 Policer and Arbiter Hierarchy 2109
Figure 376 Parent Policer and Root Arbiter....2110
Figure 377 Configuration Example 2112
Figure 378 Post Policing Queues 2113
Figure 379 Parent Policer Thresholds 2119

Pseudowire QoS 2183
Figure 380 Ingress PW QoS....2184
Figure 381 Egress PW QoS 2185
Figure 382 Example Epipe Pseudowire Topology....2188

QoS Architecture and Basic Operation 2205
Figure 383 Service and Network QoS Policies....2208
Figure 384 Visualization of Default Network Policies 2226
Figure 385 Default Buffer Pools....2228
Figure 386 WRED Slope Characteristics....2234
Figure 387 Ingress Buffer Pools and Queue Sizing....2236
Figure 388 Egress Buffer Pools and Queue Sizing 2237
Figure 389 Scheduling (Dequeuing Packets from the Queue) 2239
Figure 390 IOM QoS Overview....2240
Figure 391 Ingress and Egress SAP Queue Statistics for an IES Service 2241
Figure 392 Ingress and Egress Network Port Queue Statistics....2242

Preface

About This Guide

The Advanced Configuration Guide is divided into two books, the Part I Guide and the Part II Guide.

Part I provides advanced configurations for basic systems, system management, interface configuration, router configuration, unicast routing protocols, MPLS, services overview, Layer 2 and EVPN services, Layer 3 services, and Quality of Service.

Part II provides advanced configurations for Multi-Service Integrated Service Adapter and Triple Play Service Delivery Architecture.

Parts I and II of the Advanced Configuration Guide supplement the user configuration guides listed below.

The guide is organized alphabetically within each chapter and provides feature and configuration explanations, CLI descriptions and overall solutions. The chapters in the Advanced Configuration Guide are written for and based on several releases, up to 14.0.R5. The Applicability section in each chapter specifies on which release the configuration is based.

Audience

This manual is intended for network administrators who are responsible for configuring the routers. It is assumed that the network administrators have a detailed understanding of networking principles and configurations.

List of Technical Publications

The 7x50 series documentation set also includes the following guides:

- Basic System Configuration Guide

  This guide describes CLI usage, file system management, and boot option file (BOF) configuration, as well as how to configure basic system management, node timing, and synchronization functions.

- System Management Guide

  This guide describes system security features, SNMP and NETCONF features, and event and accounting logs. It covers basic tasks such as configuring management access filters, passwords, and user profiles.

- Interface Configuration Guide

  This guide describes how to provision Input/Output Modules (IOMs), XMA Control Modules (XCMs), Media Dependent Adapters (MDAs), XRS Media Adapters (XMAs), and ports.

- Router Configuration Guide

  This guide describes logical IP routing interfaces and associated attributes such as IP addresses, as well as Redundancy Protocol (VRRP), and Cflowd.

- Unicast Routing Protocols Guide

  This guide provides an overview of unicast routing concepts and provides configuration examples for Routing Information Protocol (RIP), Open Shortest Path First (OSPF), Intermediate-system-to-intermediate-system (IS-IS), and Border Gateway Protocol (BGP) routing protocols, and for route policies.

- Multicast Routing Protocols Guide

  This guide provides an overview of multicast routing concepts and provides configuration examples for Internet Group Management Protocol (IGMP), Multicast Listener Discovery (MLD), Protocol Independent Multicast (PIM), Multicast Source Discovery Protocol (MSDP), Multipoint LDP, multicast extensions to BGP, and Multicast Connection Admission Control (MCAC).

- MPLS Guide

  This guide describes how to configure Multiprotocol Label Switching (MPLS), Resource Reservation Protocol (RSVP), Generalized Multiprotocol Label Switching (GMPLS), and Label Distribution Protocol (LDP).

- Services Overview Guide

  This guide provides a general overview of functionality provided by the routers and describes how to configure service parameters such as Service Access Points (SAPs), Service Distribution Points (SDPs), customer information, and user services.

- Layer 2 Services and EVPN Guide: VLL, VPLS, PBB, and EVPN

  This guide describes Layer 2 service and Ethernet Virtual Private Network (EVPN) functionality and provides examples to configure and implement Virtual Leased Lines (VLLs), Virtual Private LAN Service (VPLS), Provider Backbone Bridging (PBB), and EVPN.

- Layer 3 Services Guide: Internet Enhanced Services and Virtual Private Routed Network Services

This guide describes Layer 3 service functionality and provides examples to configure and implement Internet Enhanced Services (IES) and Virtual Private Routed Network (VPRN) services.

- Versatile Service Module Guide (for the 7750 SR)

  This guide describes how to configure service parameters for the Versatile Service Module (VSM).

- OAM and Diagnostics Guide

  This guide describes how to configure features such as service mirroring and Lawful Intercept (LI), and how to use the Operations, Administration and Management (OAM) and diagnostics tools.

- Log Events Guide

  This guide provides general information about log events.

- Triple Play Service Delivery Architecture Guide (for the 7450 ESS and 7750 SR)

  This guide describes the Triple Play Service Delivery Architecture (TPSDA) support and provides examples to configure and implement various protocols and services.

- Quality of Service Guide

  This guide describes how to configure Quality of Service (QoS) policy management.

- RADIUS Attributes Reference Guide (for the 7750 SR)

  This guide describes all supported RADIUS Authentication, Authorization, and Accounting attributes.

- Multiservice Integrated Service Adapter Guide (for the 7450 ESS and 7750 SR)

  This guide describes services provided by integrated service adapters, such as Application Assurance, IPSec, ad insertion (ADI), and Network Address Translation (NAT).

- Gx AVPs Reference Guide (for 7750 SR)

  This guide describes Gx Attribute Value Pairs (AVPs).

- Acronyms Reference Guide

  This guide lists acronyms used in the customer documentation for 7750 SR, 7450 ESS, and 7950 XRS.

Basic System

In this section

This section provides configuration information for the following topics:

- IEEE 1588 for Frequency, Phase, and Time Distribution
- Synchronous Ethernet

IEEE 1588 for Frequency, Phase, and Time Distribution

This chapter provides information about IEEE 1588 for frequency, phase, and time distribution.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This section is applicable to all of the 7750 SR and 7450 ESS series, except for the SR-1, ESS-1, and ESS-6/6v. It is not applicable to t.he 7710 SR nor the 7950 XRS series. Description and examples are based on release 12.0.R2. The only software pre-requisites are IP reachability between the node and neighboring 1588 clocks.

IEEE 1588 has several hardware dependencies both for the basic functionality as well as the 1588 port based timestamping necessary for high accuracy time distribution. Please consult the related Nokia documentation for the details of all the hardware requirements.

Overview

Defined in IEEE Std 1588™-2008 (1588v2), Precision Time Protocol (PTP) is a protocol that distributes frequency, phase and time over packet based networks.



Note: Many applications do not need time alignment but only phase alignment. However, phase is derived from time and so, for the remainder of the document, discussion refers to time but those references imply both time and phase.

The IEEE 1588 protocol has become the standard for distribution of high accuracy time. Following guidelines for specific network architectures allows the delivery of time to accuracies of one microsecond. This level of accuracy is required for mobile base stations using either Time Division Duplex technology and/or advanced LTE functions, as well as in the power industry for intelligent electronic device alignment.

More lenient architectures can still achieve 100 microseconds or better accuracies which can greatly enhance the usefulness of event logging and network one way delay measurements.

In addition, 1588 has been used to deliver a frequency reference for T1/E1 ports or for mobile base station frequency alignment. This is useful in environments where the transport network does not provide physical layer synchronization services.

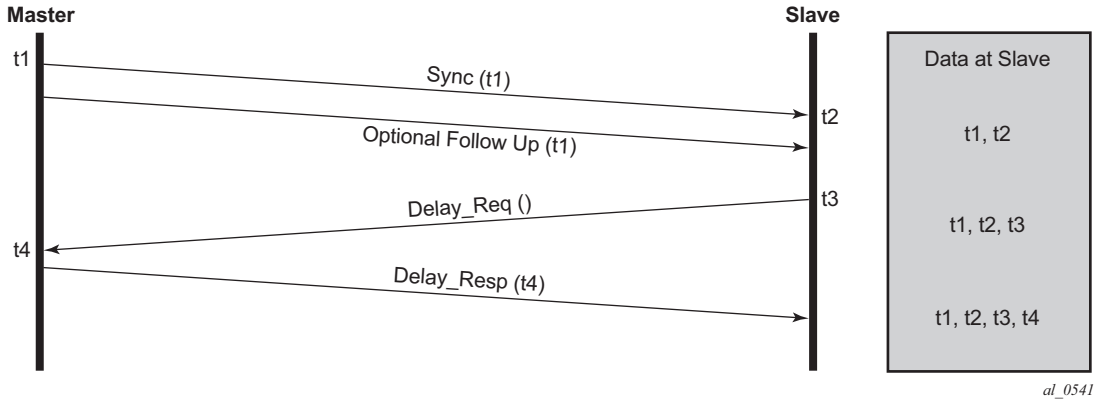
The following 1588 capabilities are provided within the 7750 SR and 7450 ESS nodes:

- CPM/CFM based 1588 master, boundary, and slave clock functionality
- Transport over Unicast UDP/IPv4 packets
- Access to 1588 process through base routing, IES, and VPRNs
- Port based timestamping of 1588 packets
- IEEE 1588 Profiles: 2008 standard default and ITU-T G.8265.1
- Utilization of 1588 derived time for NTP and System time.

PTP Basics

PTP uses an exchange of four timestamps between a reference clock (master port) and the clock to be synchronized (slave port). A simplified illustration of this mechanism is shown in Figure 1.

Figure 1 **PTP Messages and Timestamp Exchange**



The master sends a PTP Sync message containing a timestamp of when the Sync message is transmitted (t1) to the slave. In a two-step master clock, the t1 timestamp is sent in a Follow_Up message. The slave records the time it receives the Sync message (t2). At some point after receiving the Sync message, the slave sends a Delay_Req message back to the master. The slave records the time of transmission of the Delay_Req message (t3) locally. The master records the time it receives the Delay_Req message (t4) and sends this timestamp back to the slave in a Delay_Resp message.

Note: The Follow_Up message was defined to allow for implementations to generate a timestamp for the transmission of the Sync message but not have to try to insert that timestamp into the Sync message and update any frame checksums on the fly as it is in the process of transmission. While many recent implementations can perform the timestamping, update and checksum calculation on the fly, not all devices could perform this three step process with the desired accuracy. By using the Follow_Up message to transmit the timestamp of the Sync message, the master port can still provide extremely accurate timestamps for the transmission of the Sync message to the slave port. Apart from the extra message required, there is no detriment to a master port using one-step clock versus a two-step clock procedures. All PTP clocks that have slave port capability must accept timing information from both types of master port. There is no requirement to force a clock that is a one-step clock to use two-step clock procedures on its master ports. The nodes covered by this example all support one-step clock master port procedures.

After the four timestamp exchange the slave can calculate the mean path delay and the clock offset from master using the following two equations:

- mean_path_delay = [(t4-t1) – (t3-t2)] /2
- offset_from_master = [(t2 –t1) – mean_path_delay]

These calculations can occur on every message exchange or some initial packet selection can be performed so that only optimal message exchanges are used. The latter is useful if there is variable delay between the master and slave ports.

If only frequency is necessary, then the slave may use one or both pairs of timestamps (t1, t2) and (t3, t4). The slave can monitor the change in the perceived delay master-to-slave (t2 - t1) or slave-to-master (t4 - t3) over time. If the delay (t2 – t1) decreases over time, it means the t2 timestamps are not progressing quickly enough and the slave clock frequency needs to be increased.

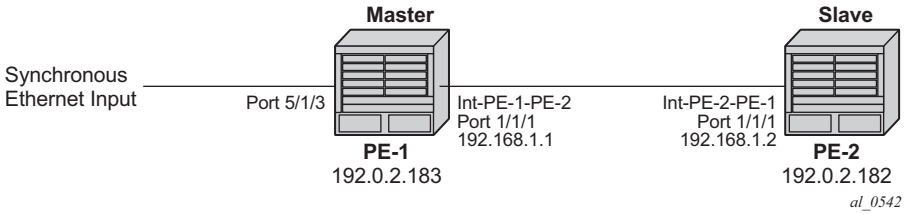
If time is necessary, then all four timestamps must be used. It is also important to note how the equation for offset uses the mean_path_delay. If the delays in the two directions are actually different, then the equation will introduce an error in the offset_from_master that is half of the difference of the two delays. The IEEE 1588 standard includes procedures to compensate for this asymmetry, if it is known, but if it is uncompensated it does introduce time error.

PTP Deployment Architectures

It is important to understand that there are very different topologies recommended for using 1588 for frequency distribution and using 1588 for time distribution.

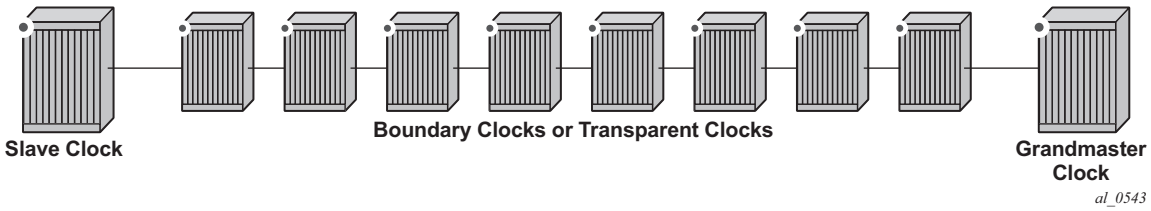
Frequency distribution was developed for an architecture where there are mobile providers who have points of presence at the mobile telephony switching offices (MTSOs) and the cell site locations which depend on other parties for the connectivity between the MTSOs and the cell site locations. The mobile providers wanted a solution that could span the transport networks with minimal dependence on that network. This can be achieved by placing a 1588 grandmaster at the MTSO and a slave in a cell site router or directly in the basestation and distributing the timestamped packets between the two, as shown in Figure 2. The transport network does introduce packet delay variation (PDV) to the 1588 messages which makes it more difficult to track the frequency of the grandmaster’s clock. However, the slaves have been designed to perform packet selection and noise filtering to allow for the recovery of a frequency within the required accuracies of the mobile basestations. This architecture and the performance requirements are covered by the ITU-T G.826x series of recommendations.

Figure 2 **1588 Topology for Frequency Distribution**



For time distribution, it has been recognized that the architecture used above is extremely unlikely to be successful. The fundamental reason is that the performance requirement is much tighter and the network introduces not only PDV but also potentially asymmetric delay which causes time error in the slave. The topology recommended for time distribution is what is sometimes referred to as “Full On-Path Support (OPS)”. Full OPS means that every network element between the grandmaster clock and the slave clock is either a 1588 boundary clock or a 1588 transparent clock, as shown in Figure 3. Boundary clocks and transparent clocks process the 1588 messages and remove the PDV noise that would be present in a non 1588 network element. By using network elements that have very tight constraints on the time error they introduced, the network can be built to guarantee time accuracy under all network traffic conditions. This architecture and the performance requirements are covered by the ITU-T G.827x series of recommendations.

Figure 3 **1588 Topology for Time Distribution**



PTP Profiles

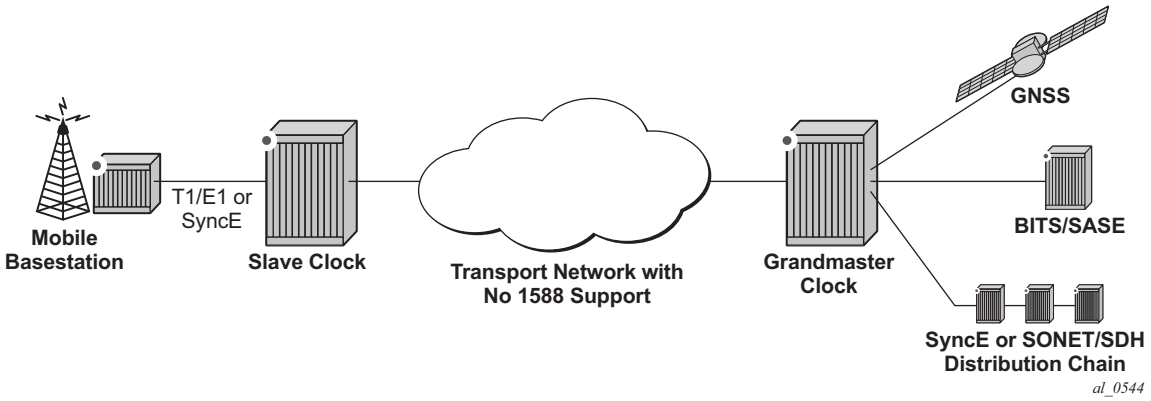
The 1588v2 standard includes the concept of a PTP profile. A PTP profile allows standardization bodies or industry groups to adapt the 1588v2 standard to a particular application. A profile defines which aspects of the 1588v2 standard are included or excluded, along with configurable ranges and defaults necessary for the application.

The 1588 standard itself includes a **default** profile that can be used for either time or frequency distribution. The default profile was defined principally for multicast operation. However, it can be used with the unicast sessions as described below. The default profile supports all 1588 clock types and includes the Best Master Clock Algorithm (BMCA) that automatically builds the synchronization distribution hierarchy amongst the PTP clocks. The SR OS only supports the unicast session version of the default profile.

In the telecommunications industry, the ITU-T is the body that develops these profiles. They have generated a profile for frequency distribution (G.8265.1) and a profile for time distribution (G.8275.1). The frequency profile permits only grandmaster and slave clocks and can be used to extended a traditional physical layer synchronization distribution (SONET/SDH, PDH, or SyncE) with a final leg of 1588 messages. The frequency source of the 1588 grandmaster could be a GPS receiver, a central office BITS or SASE device or it could use the frequency recovered from a Synchronous Ethernet or SONET/SDH interface. This is shown in Figure 4

Because a 1588 distribution system is significantly noisier than a physical layer distribution system, it should only be used as the final segment to connect the end application into the synchronization network. It should not be used to connect two Synchronous Ethernet or SONET/SDH islands.

Figure 4 **Frequency Distribution with 1588 as Last Mile**



The important features defined in the G.8265.1 profile are:

- Only master clocks and slave clocks are allowed.
- Unicast Message Negotiation using Signaling messages from the slave clocks toward the master clocks is used to establish communications.
- PTP messages are encapsulated over UDP over IPv4.

- PTP clock class values are based on a mapping of traditional quality levels from SSM/ESMC.

Note: SSM stands for Synchronization Status Messages and ESMC stands for Ethernet Synchronization Messaging Channel. These are two capabilities in SDH/SONET and Synchronous Ethernet respectively for the relaying of source clock quality information.

The slave clock uses an alternate BMCA to select the grandmaster clock from the available master clocks based on:

- Quality Level.
- Relative Priority.

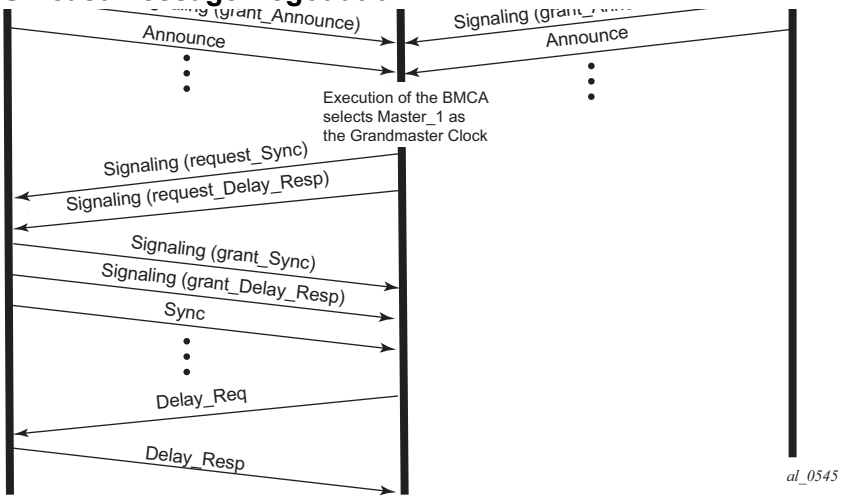
The ITU-T has defined the first time distribution profile in G.8275.1. It uses an architecture of a Global Navigation Satellite System (GNSS) based grandmaster clock distributing time through a chain of boundary clocks to a final slave device and end application. It includes the use of Synchronous Ethernet and 1588 at the same time for optimal performance. Physical layer Synchronous Ethernet is an excellent tool for the distribution of an accurate and stable frequency. This frequency can be used to advance time between offset adjustments made using the 1588 information.

Unicast Message Negotiation

The initial IEEE 1588-2002 standard defined a multicast messaging model. IEEE 1588-2008 introduced the option of using unicast messaging with unicast discovery to establish a message exchange between a master and slave.

The typical unicast message flow between a master and slave is illustrated in Figure 5.

Figure 5 **Unicast Message Negotiation**



A slave clock initiates unicast discovery by sending a Signaling message to one of its configured master clocks requesting the master send unicast Announce messages to the slave. The request includes the desired rate for the Announce messages and the duration over which the messages should be sent. If the master can support the request it replies with a Signaling message indicating that the session for unicast Announce messages has been granted.

From this point on, the master sends unicast Announce messages to the slave at the rate requested. A slave will generally establish an Announce message session with at least two master clocks.

The slave then uses the Announce messages it receives from all masters as input to the BMCA that determines which master clock is the best source for information. The selected master becomes the grandmaster clock to the slave. The slave then sends additional Signaling messages to the grandmaster to request unicast delivery of Sync and Delay_Resp messages. Assuming the grandmaster clock has sufficient resources, the request is granted and unicast Sync and Delay_Resp messages are sent from the grandmaster to the slave.

As with the Announce messages, the rate at which the Sync and Delay_Resp messages are sent and the duration of the unicast sessions is requested by the slave in the initial Signaling messages.

The unicast sessions for Announce, Sync and Delay Response messages have an expiry time. The slave renews all three sessions before this time is reached.

Network Limits

A common concern around 1588 is whether it will work on or over a specific customer network. For time distribution using full OPS as shown in Figure 3, there are well defined limits on the number of network elements allowed in the distribution chain (see below). However, for the frequency distribution using the architecture shown in Figure 2, it is a more difficult question to answer. There are so many different types of network elements and inter-node links that a simple limit on the number of network elements is not adequate. What has been specified is a limit to the noise that the network can introduce to the 1588 message flow between the grandmaster and slave clocks. This noise occurs as packet delay variation (PDV). The following sections provide some description of this PDV and a new metric that has been defined for PDV as well as the recommended limit to PDV for 1588 deployments.

Packet Delay Variation

If the packet delay through the packet network is constant, then it is relatively easy to use a series of timestamp exchanges to remove the delay as an unknown and track the master clock frequency. However, in most network technologies, the packet delay will be different for each individual packet. This PDV makes it more difficult to track the master clock since observations have both the master information and PDV noise included.

PDV is introduced when packets get placed in queues before they are forwarded. The time each packet sits in any one queue is influenced by multiple factors:

- the speed of the interface toward which the queue drains, for example 100Mbps versus 100Gbps,
- the traffic load on the interface, for example 20% versus 100% of line capacity,
- the distribution of packet sizes and priorities in the traffic load toward the interface, and
- the underlying physical technology used, xPON, xDSL, Ethernet, or microwave.

In addition, the load and packet distribution within the load will vary over time so the distribution of the PDV can shift rapidly such as when a network event triggers congestion or slowly, for example as end customers gradually come online over a period of several hours.

Also there are pipeline effects that can occur in a chain of queuing devices, where the small timing packets can catch up to a large packets moving across the network. Once behind such a packet, the timing packet can remain stuck behind that packet on all subsequent transmit queues.

QoS prioritization of packets helps reduce PDV significantly during congestion periods, but does not remove the PDV effects during lighter loading. This is due to the fact that a timing packet may be delivered to the egress queue for an interface while the interface is busy transmitting a packet. Pre-emption of packet transmissions is not used in today's networks.

Having stated all of the above, most of the time, the network will still present a percentage of packets that get across the network with minimal queuing delays. These are often referred to as 'lucky' or 'fastest' packets. Since these lucky packets are never waiting in queues or have minimal wait times, their transit across the network is relatively consistent. By running a selection filter on all 1588 packets to find these lucky packets, a level of variation of network delay can be removed or reduced significantly. Then the slave clocks have a much easier time determining the frequency of the grandmaster.

However, there will always be a limit to the amount of PDV that can be tolerated. The ITU has defined a metric to quantify the PDV, the limit of the PDV for a compliant network, and the required tolerance of a slave clock.

PDV Metrics

In order to know whether a particular timing-over-packet implementation will meet the performance targets in a given network deployment, it is desirable to both characterize the limits on the PDV that the implementation can tolerate and to measure the network against these limits. In 2012, the ITU-T published three documents that address these requirements:

- G.8260 defines the Floor Packet Percentage (FPP) metric.
- G.8261.1 defines a network limit for PDV in terms of FPP.
- G.8263 defines the input tolerance expected of a 1588 frequency slave in terms of FPP.

The Floor Packet Percentage (FPP) metric provides an indication of the guarantee that there are packets experiencing minimal delay across the network. The rationale behind this focus on 'fastest' packets is that many networks do provide good consistency of these packets in most operating conditions and because most slave clocks are capable of operating using only the information from these fastest packets.

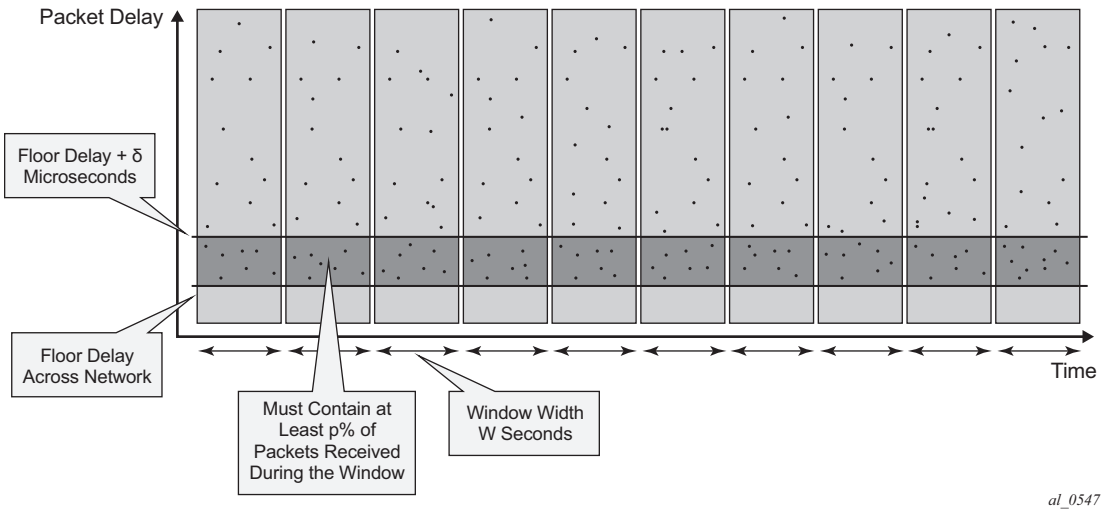
There are four parameters associated with the metric:

- **W** is the width of the windows used to monitor for the presence of fastest packets.

- **Floor Delay** is a value that is as close as possible to the absolute minimum transit delay across the network. Every actual delay measurement must be equal to or larger than this value.
- δ is the range above the floor to be analyzed for the presence of fastest packets.
- ρ is the percentage of all the packets received in a window whose delay must be within the range floor delay to floor delay + δ.

Figure 6 illustrates how these parameters and the metric work. First the delays of all individual 1588 packets are plotted over the period of observation. Next the observation period is broken down into a series of consecutive windows of width **W** seconds. Then for each window a count is made of all the 1588 packets whose delays are within the range **floor delay** to **floor delay + δ** and this count is compared with all the 1588 packets received during the window to turn the count into a percentage. Finally the percentage of each window is checked against the threshold percentage ρ. For the FPP metric to be met, every window must have a percentage greater or equal to the threshold. If even one single window does not meet this threshold then the metric condition is not met.

Figure 6 **Floor Packet Counting for FPP (n, W, δ)**



Note: This metric is not perfect as it does not take into account slaves that use other aspects of the packet delay distribution (such as average delay), nor does it discuss the impact of reroutes, nor do the limits discuss how to apply these limits to the forward and backward message exchanges at the same time. However, it was agreed that this metric was a good start for the definitions of the network and slave limits. Expect to see timing test equipment vendors providing the tools to generate 1588 PDV profiles providing FPP based distributions.

ITU-T Budget for Frequency

The network limit on PDV for frequency distribution is defined in G.8271.1 using the FPP metrics defined above.

In general most carrier grade networks with spans of up to 10 nodes and which do not exceed 80% load on their internode links should meet the requirement. However, very low (sub 50 Mbps) shaping or very long networks or last mile technologies such as xDSL or xPON may need to be studied to determine their acceptability.

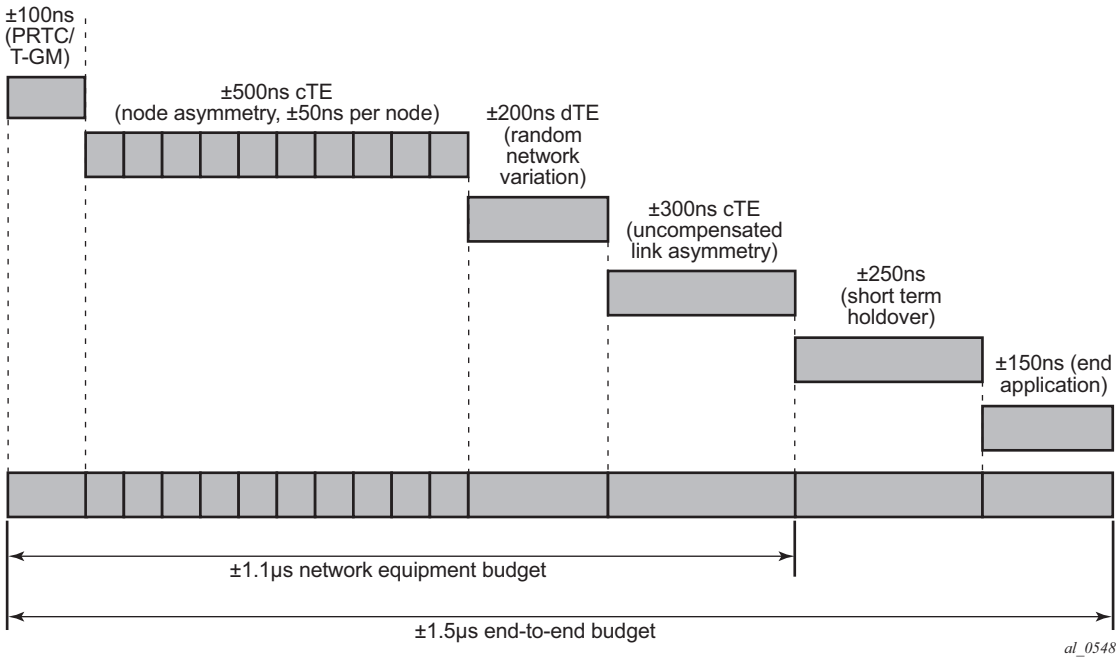
A general strategy for rolling out 1588 frequency distribution is to evaluate the specific grandmaster and slave pairing in a lab environment using a network emulator to introduce controlled PDV. Once the grandmaster and slave have passed the lab tests, then field trial locations should be identified. Ideally, the sites should include locations where the PDV of the network will likely be at its worst. This would be sites with the most intervening network elements between the grandmasters and the slaves and include segments of the network that have a high load. The slaves' clocks should be deployed and monitored over several days to ensure that their frequency recovery engines can maintain lock with the grandmasters. During the initial field trails, it is beneficial to use external frequency test equipment at the slave locations to accurately monitor the frequency generated out of the slaves and ensure it stays within limits. As more sites are evaluated and confidence in the PDV environment increases, more deployments can be rolled out. In the deployed network, PTP frequency recovery slave states can be monitored to ensure the solution continues to work.

There may be some locations in the network where the PDV will be too large preventing the slaves to achieve or maintain lock. If it is possible to utilize an alternate network interface to obtain a frequency such as a leased T1 or E1 interface then that could be used. A last resort would be the deployment of a GNSS receiver at the location to provide the frequency reference.

ITU-T Budget for Time

The ITU-T has defined a topology for time distribution based on a full OPS environment. This means that every network element in the time distribution chain is a 1588 clock of some type. Currently the work has defined an environment using Boundary Clocks, but this might be modified in the future to include transparent clocks. The ITU-T tackled the time distribution problem in a more traditional way when compared with the frequency distribution. The ITU-T first defined specific network element clock performance constraints and then defined a longest chain network permitted to ensure that the solution meets the end to end budget. The breakdown of the chain and the budget is shown in Figure 7.

Figure 7 **G.8271.1 Time Error Budget**



The overall end to end budget is defined as ±1.5 microseconds. From this the following allocations are made:

- ±100ns Time error due to the GNSS receiver and the 1588 grandmaster.
- ±500ns Constant Time error due to ten Telecom Boundary Clocks (50 ns limit per

  boundary clock).
- ±200ns Dynamic Time error presented at the end of the boundary clock chain into the

  end slave.
- ±300ns Time error due to errors in cable latency asymmetry compensation (see below).
- ±150ns Time error due to the end slave and any internals of the basestation between

  the recovery and the presentation on the air interface.
- ±200ns Time error in the end application during short term holdovers such as

  network topology re-arrangements.

Note there is discussion that some of these elements could be traded-off against each other. For example, if the link asymmetry needs a higher budget then the holdover budget would have to be less – implying a better end device or a shorter duration of holdover.

The link asymmetries are an important aspect of this budget. The network topology not only has to have the network elements that meet the clock specifications but it also needs to have links that meet certain requirements. As explained above, the time offset calculation makes the assumptions that the master–to-slave latency is the same from the slave–to-master latency. When the latency is not equal, an error is introduced. Some analysis of network intersite connections may need to be performed to determine the budget for the link asymmetries.

Configuration

IP Addressing for PTP Communication

The system supports communication to the PTP process on the CPM using any of the IPv4 local interface addresses or an IPv4 local loopback addresses. The system will record both the source and destination address information from the received Signaling message which establishes the unicast session. The system will then swap these addresses for use for the Sync and/or Delay_Req messages generated toward the external clock.

The IP address becomes more significant when 1588 port based timestamping is enabled. The port level functionality will filter received PTP packets for a known IP address. This ensures that only PTP messages intended for the node are modified and not PTP messages merely transiting the node.

If the 1588 nodes are directly connected or it is ensured that the PTP messages for a peer shall always enter/exit the system through a single interface, then the IP address of that interface can be used for the PTP message communication. If the PTP messages from a peer could enter through more than one interface, then it may be easier to utilize a loopback address for the PTP message communication.

If using a loopback address and 1588 port based timestamping is also to be used, then the specific loopback address must be assigned to PTP for use using the source-address command. An example is provided in the “Port Based Timestamping” section below. Note: When a source address is defined for the PTP process within a given routing context, then the source address for all Signaling messages originating out of the node within that routing context shall use that address.

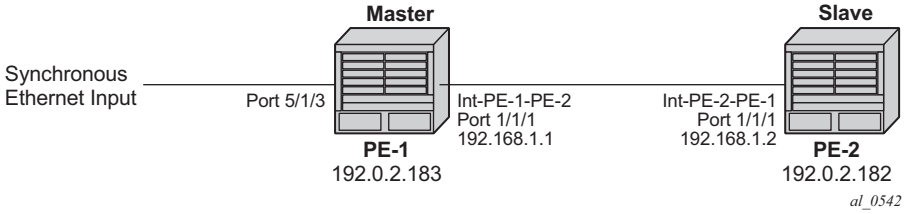
Note: The procedures to establish IP connectivity for the specific addresses used in these examples are not included.

Master and Slave Clocks for Frequency

A typical deployment scenario for a system configured as an ordinary master to distribute frequency to an external slave clock, often a cell site router or a base station, is shown in Figure 8. The central clock of the system is locked via its BITS ports or a Synchronous Ethernet port to an external source that is traceable to a primary reference. The frequency of the central clock is used to generate the timestamps contained in PTP event messages. The timestamps generated do not correlate to any standard epoch and therefore indicate an arbitrary timescale. As such it is only the rate of progression of the timestamps that has meaning.

The 7750 SR and the 7450 ESS can be configured as a 1588 slave clock for frequency recovery. In real deployments, it is more likely for the slave devices to be smaller cell site routers or basestations instead of another 7750 SR or 7450 ESS.

Figure 8 **Master and Slave Clocks for Frequency**



In the topology in Figure 8, the systems will most likely be configured with the ITU-T G.8265.1 Profile.

For this example, a loopback address is used for PTP communication between the nodes.

Ordinary Master Configuration

The steps to configure PE-1 as a PTP ordinary-clock master for frequency distribution using the G.8265.1 Telecom profile are outlined below:

Configure a /32 IPv4 system address on PE-1 and an interface to reach PE-2.

```
*A:PE-1#
        configure
            router
                interface "system"
                    address 192.0.2.183/32
                    no shutdown
                exit
                interface "int-PE-1-PE-2"
                    address 192.168.1.1/30
                    port 1/1/1
                    no shutdown
                exit
            exit
```

Configure an input reference for the central clock on PE-1. In this example, Synchronous Ethernet port 5/1/3 is used as the source for **ref2**.

```
*A:PE-1#
        configure
            port 5/1/3
                description "Sync-E reference for node"
                ethernet
                    ssm
                        code-type sonet
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            system
                sync-if-timing
                    begin
                    ql-selection
                    ref2
                        source-port 5/1/3
                        no shutdown
                    exit
                    commit
                exit
            exit
```

The default clock type is set to ordinary slave so that must be changed to ordinary master. The only other relevant configuration parameter for the master clock running the G.8265.1 profile is the network-type. The coding of the SSM/ESMC Quality Level into PTP clock Class must match the environment. The system supports both SONET and SDH networks. The default network-type is sdh but for this example, the system is configured for the North American market so the network-type is set to sonet.

```
*A:PE-1#
        configure
            system
                ptp
                    clock-type ordinary master
                    network-type sonet
                    no shutdown
                exit
            exit
```

Ordinary Slave Configuration

To configure PE-2 as a PTP ordinary slave for frequency distribution using the G.8265.1 Telecom profile, firstly configure a /32 IPv4 system address on PE-2 and an interface to reach PE-1.

```
*A:PE-2#
        configure
            router
                interface "system"
                    address 192.0.2.182/32
                    no shutdown
                exit
                interface "int-PE-2-PE-1"
                    address 192.168.1.2/30
                    port 1/1/1
                    no shutdown
                exit
            exit
```

As the default clock type is ordinary slave, PE-1 is configured as a peer clock, and the PTP process is enabled. In this example, the Quality Level encoding is also changed to sonet in order to match the North American market

```
*A:PE-1#
        configure
            system
                ptp
                    network-type sonet
                    peer 192.0.2.183 create
                        no shutdown
```

```
                exit
                no shutdown
            exit
        exit
```

Usually a 1588 slave has at least two peers configured in order to provide redundant sources.

Configure PTP as the reference for the central clock on PE-2.

```
*A:PE-2#
        configure
            system
                sync-if-timing
                    begin
                    ql-selection
                    ptp
                        no shutdown
                    exit
                    commit
                exit
            exit
```

Verification of Session Establishment

When PTP is set to no shutdown on PE-2, it initiates a PTP unicast session with PE-1. Correct session establishment can be verified by checking PTP related information as follows:

```
*A:PE-1# show system ptp unicast
===============================================================================
IEEE 1588/PTP Unicast Negotiation Information
===============================================================================
Router
  IP Address      Dir Type     Rate       Duration State    Time
-------------------------------------------------------------------------------
Base
  192.0.2.182     Tx  Announce 1 pkt/2 s  300      Granted  05/30/2014 06:08:38
  192.0.2.182     Tx  Sync     64 pkt/s   300      Granted  05/30/2014 06:08:43
  192.0.2.182     Rx  DelayReq 64 pkt/s   300      Granted  05/30/2014 06:08:43
  192.0.2.182     Tx  DelayRsp 64 pkt/s   300      Granted  05/30/2014 06:08:43
-------------------------------------------------------------------------------
PTP Peers               : 1
Total Packet Rate       : 192 packets/second
===============================================================================


*A:PE-2# show system ptp unicast
===============================================================================
IEEE 1588/PTP Unicast Negotiation Information
===============================================================================
```

```
Router
  IP Address      Dir Type     Rate       Duration State    Time
-------------------------------------------------------------------------------
Base
  192.0.2.183     Rx  Announce 1 pkt/2 s  300      Granted  05/30/2014 09:08:38
  192.0.2.183     Rx  Sync     64 pkt/s   300      Granted  05/30/2014 09:08:43
  192.0.2.183     Tx  DelayReq 64 pkt/s   300      Granted  05/30/2014 09:08:43
  192.0.2.183     Rx  DelayRsp 64 pkt/s   300      Granted  05/30/2014 09:08:43
-------------------------------------------------------------------------------
PTP Peers               : 1
Total Packet Rate       : 192 packets/second
===============================================================================
```

A **Pending** state indicates the system has sent a Unicast Request toward the peer but has not received a response. If the state remains **Pending**, then the IP connectivity between the systems should be verified.

To verify the slave frequency is operating properly, first check the high level information for PTP on PE-2. Note that the PTP Recovery State initially shows phase-tracking and then changes to locked. The time to achieve locked state varies based on the PDV.

```
*A:PE-2# show system ptp
===============================================================================
IEEE 1588/PTP Clock Information
===============================================================================
-------------------------------------------------------------------------------
Local Clock
-------------------------------------------------------------------------------
Clock Type        : ordinary,slave     PTP Profile        : ITU-T G.8265.1
Domain            : 4                  Network Type       : sonet
Admin State       : up                 Oper State         : up
Announce Interval : 1 pkt/2 s          Announce Rx Timeout: 3 intervals
Peer Limit        : none (Base Router)
Clock Id          : 00233efffe808250   Clock Class        : 255 (slave-only)
Clock Accuracy    : unknown            Clock Variance     : ffff (not computed)
Clock Priority1   : 128                Clock Priority2    : 128
PTP Port State    : slave              Last Changed       : 05/30/2014 09:08:42
PTP Recovery State: phase-tracking     Last Changed       : 05/30/2014 09:08:42
Frequency Offset  : -2.704 ppb
-------------------------------------------------------------------------------
Parent Clock
-------------------------------------------------------------------------------
IP Address        : 192.0.2.183        Router             : Base
Parent Clock Id   : 00233efffe69f250   Remote PTP Port    : 1
GM Clock Id       : 00233efffe69f250   GM Clock Class     : 80 (prs)
GM Clock Accuracy : unknown            GM Clock Variance  : ffff (not computed)
GM Clock Priority1: 128                GM Clock Priority2 : 128
-------------------------------------------------------------------------------
Time Information
-------------------------------------------------------------------------------
Timescale           : Arbitrary
Current Time        : 2014/05/30 14:12:52.9 (ARB)
Frequency Traceable : yes
Time Traceable      : no
Time Source         : other
```

```
===============================================================================
```

In addition PTP packet statistics can be checked to verify reception of the PTP messages and the execution of the frequency slave:

```
*A:PE-2# show system ptp statistics
===============================================================================
IEEE 1588/PTP Packet Statistics
===============================================================================
                                                                 Input     Output
-------------------------------------------------------------------------------
PTP Packets                                                       5506       2742
  Announce                                                          23          0
  Sync                                                            2740          0
  Follow Up                                                          0          0
  Delay Request                                                      0       2740
  Delay Response                                                  2740          0
  Signaling                                                          3          3
    Request Unicast Transmission TLVs                                0          3
      Announce                                                       0          1
      Sync                                                           0          1
      Delay Response                                                 0          1
    Grant Unicast Transmission (Accepted) TLVs                       3          0
      Announce                                                       1          0
      Sync                                                           1          0
      Delay Response                                                 1          0
    Grant Unicast Transmission (Denied) TLVs                         0          0
      Announce                                                       0          0
      Sync                                                           0          0
      Delay Response                                                 0          0
    Cancel Unicast Transmission TLVs                                 0          0
      Announce                                                       0          0
      Sync                                                           0          0
      Delay Response                                                 0          0
    Ack Cancel Unicast Transmission TLVs                             0          0
      Announce                                                       0          0
      Sync                                                           0          0
      Delay Response                                                 0          0
    Other TLVs                                                       0          0
  Other                                                              0          0
  Event Packets timestamped at port                                  0          0
  Event Packets timestamped at cpm                                2740       2740
Discards                                                             0          0
  Bad PTP domain                                                     0          0
  Alternate Master                                                   0          0
  Out Of Sequence                                                    0          0
  Peer Disabled                                                      0          0
  Other                                                              0          0
===============================================================================
===============================================================================
IEEE 1588/PTP Frequency Recovery State Statistics
===============================================================================
State                                                          Seconds
-------------------------------------------------------------------------------
Initial                                                              0
Acquiring                                                            0
Phase-Tracking                                                      43
Locked                                                               0
```

```
Hold-over                                                                      0
===============================================================================

===============================================================================
IEEE 1588/PTP Event Statistics
===============================================================================
Event                                                  Sync Flow Delay Flow
-------------------------------------------------------------------------------
Packet Loss                                                 0          0
Excessive Packet Loss                                       0          0
Excessive Phase Shift Detected                              0          0
Too Much Packet Delay Variation                             0          0
===============================================================================
*
```

Secondly, the central clock status on the system can be checked:

```
*A:PE-2# show system sync-if-timing
===============================================================================
System Interface Timing Operational Info
===============================================================================
System Status CPM B                : Master Locked
    Reference Input Mode           : Non-revertive
    Quality Level Selection        : Disabled
    Reference Selected             : ptp
    System Quality Level           : prs
    Current Frequency Offset (ppm) : +0

Reference Order                    : bits ref1 ref2 ptp

Reference Mate CPM
    Qualified For Use              : No
        Not Qualified Due To       :     LOS
    Selected For Use               : No
        Not Selected Due To        :     not qualified

Reference Input 1
    Admin Status                   : down
    Rx Quality Level               : unknown
    Quality Level Override         : none
    Qualified For Use              : No
        Not Qualified Due To       :     disabled
    Selected For Use               : No
        Not Selected Due To        :     disabled
    Source Port                    : None

Reference Input 2
    Admin Status                   : down
    Rx Quality Level               : unknown
    Quality Level Override         : none
    Qualified For Use              : No
        Not Qualified Due To       :     disabled
    Selected For Use               : No
        Not Selected Due To        :     disabled
    Source Port                    : None

Reference BITS B
    Input Admin Status             : down
```

```
    Rx Quality Level               : failed
    Quality Level Override         : none
    Qualified For Use              : No
        Not Qualified Due To       :     disabled
    Selected For Use               : No
        Not Selected Due To        :     disabled
    Interface Type                 : DS1
    Framing                        : ESF
    Line Coding                    : B8ZS
    Line Length                    : 0-110ft
    Output Admin Status            : down
    Output Source                  : line reference
    Output Reference Selected      : none
    Tx Quality Level               : N/A

Reference PTP
    Admin Status                   : up
    Rx Quality Level               : prs
    Quality Level Override         : none
    Qualified For Use              : Yes
    Selected For Use               : Yes
```

Optional Configuration Items for Ordinary Master or Slave Configuration

The G.8265.1 profile is the default PTP profile on the system and it uses domain number value of 4. The domain number must match at both ends of the communication path or the PTP messages will be dropped. Some very old 1588 devices, may have the domain number set to zero which is the value used by the IEEE1588 default profile. In this case, the system would need to have its domain number changed to match that of the external slave.

```
configure
    system
        ptp
            shutdown
            domain 0
            no shutdown
        exit
    exit
```

Note that the domain number can only be adjusted if PTP is shutdown and only one common domain number is allowed for all 1588 messages to and from the system.

When using the system as a 1588 slave for frequency distribution, it is strongly recommended to use the default message rate of 64 pps for Sync and Delay_Resp messages. If for some reason the parent 1588 peer cannot offer this rate, then the rate that the system requests must be adjusted. For example, if the maximum rate supported by the external 1588 grandmaster device (with an IP address of 192.0.2.166) only is 32 pps, then the system can be adjusted to request that rate as follows:

```
configure
    system
        ptp
            peer 192.0.2.166 create
                log-sync-interval -5
                no shutdown
            exit
        exit
    exit
```

Note that the Sync message rate can only be adjusted if the peer is shutdown.

The message rates are entered as the base 2 logarithm of the inter-message interval. So 32 pps has an inter message interval of 1/32 seconds and a log-sync-interval of -5.

The Announce message rate impact the speed at which PTP can detect communication failures and the speed at which the PTP topology is re-arranged. The default Announce rate is one message every two seconds and this should be adequate for networks with short chains of PTP clocks, for example G.8265.1 architectures. However, in network with longer chains of PTP clocks (for example, more than 5 boundary clocks), it may be desired to use a faster Announce message rate. In the following example, the slave is configured to request two Announce messages per second:

```
configure
    system
        ptp
            shutdown
            log-anno-interval -1
            no shutdown
        exit
    exit
```

Note that the Announce rate can only be adjusted if PTP is shutdown. In addition, there is one common Announce rate for all unicast sessions; it cannot be configured on an individual peer basis.

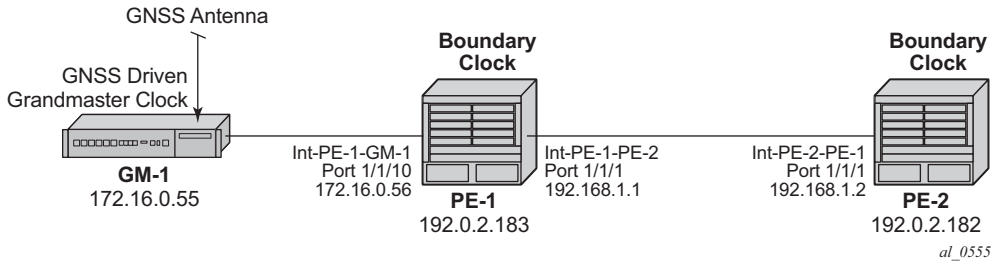
Boundary Clock

With the increase interest in high accuracy time distribution across networks, the system most likely takes on the role of a 1588 boundary clock. In this role, the system requests time from a GNSS driven grandmaster clock or from a neighboring boundary clock. The system only supports boundary clock configuration when the ptp profile is configured as the default profile.

In this mode of operation, it is strongly recommended to have Synchronous Ethernet physical layer frequency distribution configured at the same time.

The example in Figure 9 shows a boundary clock (PE-1) communicating directly with the GNSS driven grandmaster (GM-1) and a second boundary clock (PE-2) communicating with the first boundary clock.

Figure 9 **Boundary Clock**



The steps to configure the systems as boundary clocks running the IEEE default profile are:

On PE-1, configure a /32 IPv4 system address, an interface to reach PE-2, and an interface to reach GM-1.

```
*A:PE-1#
        configure
            router
                interface "system"
                    address 192.0.2.183/32
                    no shutdown
                exit
                interface "int-PE-1-PE-2"
                    address 192.168.1.1/30
                    port 1/1/1
                    no shutdown
                exit
                interface "int-PE-1-GM-1"
                    address 172.16.0.56/30
                    port 1/1/2
                    no shutdown
                exit
            exit
```

On PE-2, configure a /32 IPv4 system address and an interface to reach PE-1.

```
*A:PE-2#
        configure
            router
                interface "system"
                    address 192.0.2.182/32
                    no shutdown
                exit
                interface "int-PE-2-PE-1"
                    address 192.168.1.2/30
                    port 1/1/1
                    no shutdown
                exit
            exit
```

Configure both PE-1 and PE-2 to have physical layer frequency sources into their central clocks. PE-2 is configured to receive Synchronous Ethernet from PE-1 on the same port as is used for PTP. This commonality is not a requirement but might be common in the network topology.

On PE-1, configure the port toward PE-2 as a Synchronous Ethernet port. This will cause the port transmit timing to be sourced from the node timing. Also configure the port to transmit ssm codes using the sonet codes.

```
*A:PE-1#
        configure card 1 mda 1 sync-e
        configure port 1/1/1 ethernet
            code-type sonet
            no shutdown
        exit
```

On PE-2, configure the port on towards PE-1 as a Synchronous Ethernet port and to use sonet codes and to be the reference into the central clock of PE-2.

```
*A:PE-2#
        configure card 1 mda 1 sync-e
        configure port 1/1/1
            ethernet
                ssm
                    code-type sonet
                    no shutdown
                exit
            exit
        exit
        configure system sync-it-timing
            begin
            ql-selection
            ref1
                source-port 1/1/1
                no shutdown
            exit
            commit
```

```
        exit
```

Next configure PE-1 as a boundary clock requesting service from GM-1 using the default profile. In this example, the interface address of GM-1 is used for the PTP communication.

```
*A:PE-1#
        configure system ptp
            shutdown
            profile ieee1588-2008
            clock-type boundary
            peer 172.16.0.55 create
                no shutdown
            exit
            no shutdown
        exit
```

If it is desired to operate the network at the default for the G.8275.1 profile, then the Announce messages should be set to 8 pps and the Sync and Delay_Resp messages should be set to 16 pps.

```
*A:PE-1#
        configure system ptp
            shutdown
            log-anno-interval -3
            peer 172.16.0.55
                shutdown
                log-sync-interval -4
                no shutdown
            exit
            no shutdown
        exit
```

Configure PE-2 as a boundary clock using PE-1 as its parent clock and the same set of 1588 parameters. In this example, PE-2 uses a loopback address of PE-1 for communication.

```
*A:PE-2#
        configure system ptp
            shutdown
            profile ieee1588-2008
            clock-type boundary
            log-anno-interval -3
            peer 192.0.2.183 create
                shutdown
                log-sync-interval -4
                no shutdown
            exit
            no shutdown
        exit
```

On PE-1, validate the status of the PTP topology by checking the unicast sessions. Also validate the PTP process has elected GM-1 as both the parentClock and the grandmaster clock.

```
*A:PE-1# show system ptp unicast
===============================================================================
IEEE 1588/PTP Unicast Negotiation Information
===============================================================================
Router
  IP Address      Dir Type     Rate       Duration State    Time
-------------------------------------------------------------------------------
Base
  192.0.2.182     Tx  Announce 8 pkt/s    300      Granted  05/30/2014 07:02:36
  192.0.2.182     Tx  Sync     16 pkt/s   300      Granted  05/30/2014 07:02:37
  192.0.2.182     Rx  DelayReq 16 pkt/s   300      Granted  05/30/2014 07:02:37
  192.0.2.182     Tx  DelayRsp 16 pkt/s   300      Granted  05/30/2014 07:02:37
  172.16.0.55     Rx  Announce 8 pkt/s    300      Granted  05/30/2014 07:02:42
  172.16.0.55     Rx  Sync     16 pkt/s   300      Granted  05/30/2014 07:02:43
  172.16.0.55     Tx  DelayReq 16 pkt/s   300      Granted  05/30/2014 07:02:43
  172.16.0.55     Rx  DelayRsp 16 pkt/s   300      Granted  05/30/2014 07:02:43
-------------------------------------------------------------------------------
PTP Peers               : 2
Total Packet Rate       : 112 packets/second
===============================================================================


*A:PE-1# show system ptp
===============================================================================
IEEE 1588/PTP Clock Information
===============================================================================
-------------------------------------------------------------------------------
Local Clock
-------------------------------------------------------------------------------
Clock Type        : boundary           PTP Profile        : IEEE 1588-2008
Domain            : 0                  Network Type       : sdh
Admin State       : up                 Oper State         : up
Announce Interval : 8 pkt/s            Announce Rx Timeout: 3 intervals
Peer Limit        : none (Base Router)
Clock Id          : 00233efffe69f250   Clock Class        : 248 (default)
Clock Accuracy    : unknown            Clock Variance     : ffff (not computed)
Clock Priority1   : 128                Clock Priority2    : 128
PTP Recovery State: locked             Last Changed       : 05/30/2014 07:05:17
Frequency Offset  : +50.305 ppb
-------------------------------------------------------------------------------
Parent Clock
-------------------------------------------------------------------------------
IP Address        : 172.16.0.55        Router             : Base
Parent Clock Id   : 8887868584838281   Remote PTP Port    : 1
GM Clock Id       : 8887868584838281   GM Clock Class     : 7
GM Clock Accuracy : within 250 ns      GM Clock Variance  : 0x6400 (3.7E-09)
GM Clock Priority1: 128                GM Clock Priority2 : 128
-------------------------------------------------------------------------------
Time Information
-------------------------------------------------------------------------------
Timescale           : PTP
Current Time        : 2014/05/30 15:07:01.1 (UTC)
Frequency Traceable : yes
Time Traceable      : yes
```

```
Time Source         : GPS
```

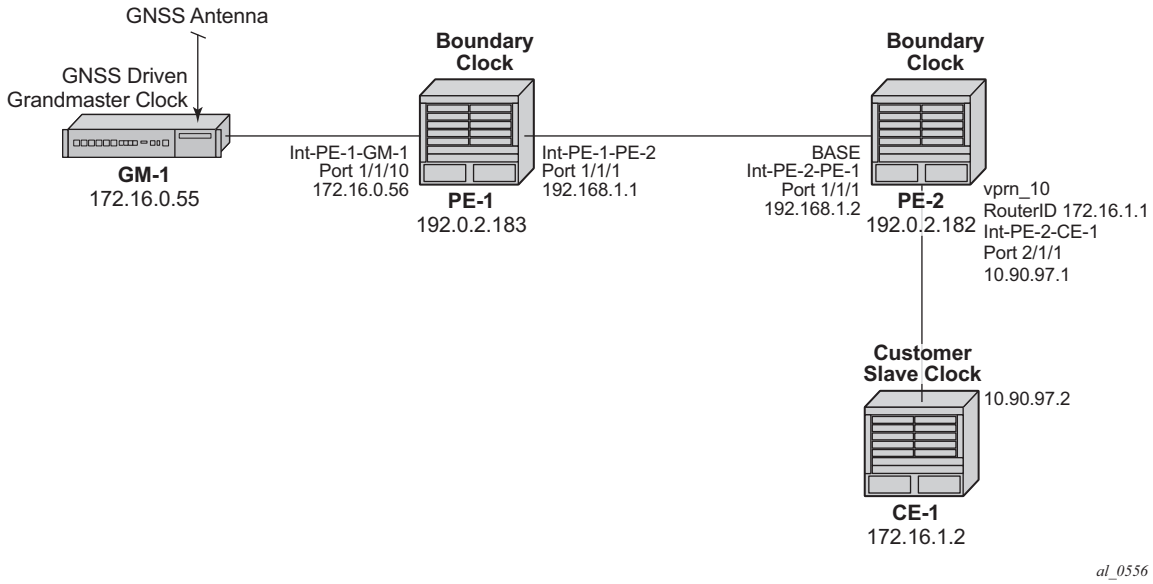
On PE-2, validate the PTP process has elected PE-1 as its parentClock and that the grandmaster clock is GM-1.

```
*A:PE-2# show system ptp
================================================================================
IEEE 1588/PTP Clock Information
================================================================================
--------------------------------------------------------------------------------
Local Clock
--------------------------------------------------------------------------------
Clock Type        : boundary           PTP Profile        : IEEE 1588-2008
Domain            : 0                  Network Type       : sdh
Admin State       : up                 Oper State         : up
Announce Interval : 8 pkt/s            Announce Rx Timeout: 3 intervals
Peer Limit        : none (Base Router)
Clock Id          : 00233efffe808250   Clock Class        : 248 (default)
Clock Accuracy    : unknown            Clock Variance     : ffff (not computed)
Clock Priority1   : 128                Clock Priority2    : 128
PTP Recovery State: locked             Last Changed       : 05/30/2014 10:02:36
Frequency Offset  : -9.345 ppb
--------------------------------------------------------------------------------
Parent Clock
--------------------------------------------------------------------------------
IP Address        : 192.0.2.183        Router             : Base
Parent Clock Id   : 00233efffe69f250   Remote PTP Port    : 1
GM Clock Id       : 8887868584838281   GM Clock Class     : 7
GM Clock Accuracy : within 250 ns      GM Clock Variance  : 0x6400 (3.7E-09)
GM Clock Priority1: 128                GM Clock Priority2 : 128
--------------------------------------------------------------------------------
Time Information
--------------------------------------------------------------------------------
Timescale           : PTP
Current Time        : 2014/05/30 15:09:26.5 (UTC)
Frequency Traceable : yes
Time Traceable      : yes
Time Source         : GPS
================================================================================
```

Boundary Clock with VPRN Access

The system supports access to the 1588 process through Base routing, IES, and VPRN contexts. This permits the system 1588 topology to be created and managed in one context with access for edge distribution through other contexts. For example, building on top of the base routing distribution shown in the previous example, access can be given to the 1588 process on PE-2 via a VPRN existing on that node. This allows the VPRN customer to have access to the high accuracy time available within the system in the customer edge equipment connecting into that node.

Figure 10 **Boundary Clocks with Edge VPRN Access**



For the example shown in Figure 10, it is assumed that a VPRN service is already configured and operational on PE-2 providing connectivity between PE-2 and CE-1:

```
*A:PE-2#
        configure service vprn 10 customer 1 create
            router-id 176.16.1.1
            autonomous-system 64496
            route-distinguisher 64496:10
            interface "int-PE-2-CE-1" create
                address 10.90.97.1/30
                sap 2/1/1 create
                exit
            exit
            no shutdown
        exit
```

To enable access to the PTP process via VPRN 10 in PE-2, PTP must be enabled within the VPRN context. To ensure that no more than 10 external clocks access the system PTP through this VPRN at any one time, a peer-limit may be defined.

```
*A:PE-2#
        configure service vprn 10
            peer-limit 10
            ptp no shutdown
        exit
```

To confirm PTP access with the VPRN, the PTP information with the VPRN context can be queried. Either of the following two commands can be used:

```
*A:PE-2# show system ptp unicast router 10
```

or

```
*A:PE-2# show service id 10 ptp unicast
```

These two commands provide the same information as shown below.

```
*A:PE-2# show system ptp unicast router 10
===============================================================================
IEEE 1588/PTP Unicast Negotiation Information
===============================================================================
Router
  IP Address      Dir Type     Rate       Duration State    Time
-------------------------------------------------------------------------------
10
  172.16.1.2      Tx  Announce 1 pkt/2 s  300      Granted  05/30/2014 12:40:53
  172.16.1.2      Tx  Sync     64 pkt/s   300      Granted  05/30/2014 12:40:59
  172.16.1.2      Rx  DelayReq 64 pkt/s   300      Granted  05/30/2014 12:40:59
  172.16.1.2      Tx  DelayRsp 64 pkt/s   300      Granted  05/30/2014 12:40:59
-------------------------------------------------------------------------------
PTP Peers               : 1
Total Packet Rate       : 192 packets/second
===============================================================================
```

Port Based Timestamping

As described above, optimal performance is achieved when the 1588 port based timestamping (PBT) feature is used. This feature is not available on all hardware and the interfaces for PTP should be planned in advance if this feature is to be used.

Since 1588 messages ingress and egress the node through router interfaces, the configuration of the 1588 PBT feature is enabled within the router interface context. In the previous examples, if 1588 PBT is to be enabled on all the PTP interfaces the following commands are required.

On PE-1, enable 1588 PBT on the interface toward GM-1 and PE-2.

```
*A:PE-1#
        configure
            router
                interface "int-PE-1-PE-2"
                    ptp-hw-assist
                exit
                interface "int-PE-1-GM-1"
                    ptp-hw-assist
                exit
            exit
```

On PE-2, enable 1588 PBT on the interface toward PE-1 and CE-1.

```
*A:PE-2#
        configure
            router
                interface "int-PE-2-PE-1"
                    ptp-hw-assist
                exit
            exit
        exit
        configure service vprn 10 customer 1
            interface "int-PE-2-CE-1"
                ptp-hw-assist
            exit
        exit
```

To verify 1588 PBT is active on the 1588 messages to the peers, check the timestamp point for the specific peer. It now indicates *port* rather than cpm.

On PE-1 for the CE-1 communication:

```
*A:PE-1# show system ptp peer 172.16.0.55
===============================================================================
IEEE 1588/PTP Peer Information
===============================================================================
Router            : Base
IP Address        : 172.16.0.55        Announce Direction : rx
Admin State       : up                 G.8265.1 Priority  : n/a
Sync Interval     : 16 pkt/s
Local PTP Port    : 2                  PTP Port State     : slave
Clock Id          : 8887868584838281   Remote PTP Port    : 1
GM Clock Id       : 8887868584838281   GM Clock Class     : 7
GM Clock Accuracy : within 250 ns      GM Clock Variance  : 0x6400 (3.7E-09)
GM Clock Priority1: 128                GM Clock Priority2 : 128
Steps Removed     : 0                  Parent Clock       : yes
Tx Timestamp Point: port               Rx Timestamp Point : port
Last Tx Port      : 5/1/1              Last Rx Port       : 5/1/1
===============================================================================
```

On PE-1 the communication with the PE-2 will still be CPM timestamping since the port has not been configured to watch for the ‘system’ loopback address.

```
*A:PE-1# show system ptp peer 192.0.2.182
===============================================================================
IEEE 1588/PTP Peer Information
===============================================================================
Router            : Base
IP Address        : 192.0.2.182        Announce Direction : tx
Admin State       : n/a                G.8265.1 Priority  : n/a
Sync Interval     : n/a
Local PTP Port    : 3                  PTP Port State     : master
Clock Id          : 00233efffe808250   Remote PTP Port    : 4
Tx Timestamp Point: cpm                Rx Timestamp Point : cpm
Last Tx Port      : 5/1/2              Last Rx Port       : 5/1/2
===============================================================================
```

In order to configure the **system** loopback address for PTP, enter the following on PE-1:

```
*A:PE-1#
        configure
            system security
                source-address application ptp "system"
                exit
            exit
```

Now the timestamp point on PE-1 will be the port.

```
*A:PE-1# show system ptp peer 192.0.2.182
===============================================================================
IEEE 1588/PTP Peer Information
===============================================================================
Router            : Base
IP Address        : 192.0.2.182        Announce Direction : tx
Admin State       : n/a                G.8265.1 Priority  : n/a
Sync Interval     : n/a
Local PTP Port    : 3                  PTP Port State     : master
Clock Id          : 00233efffe808250   Remote PTP Port    : 4
Tx Timestamp Point: port               Rx Timestamp Point : port
Last Tx Port      : 5/1/2              Last Rx Port       : 5/1/2
===============================================================================
```

Repeat this configuration of system address for the base routing context on PE-2

```
*A:PE-2#
        configure
            system security
                source-address application ptp "system"
                exit
            exit
```

Now the timestamp point on PE-2 will be the port.

```
*A:PE-2# show system ptp peer 192.0.2.183
===============================================================================
IEEE 1588/PTP Peer Information
===============================================================================
Router            : Base
IP Address        : 192.0.2.183        Announce Direction : rx
Admin State       : up                 G.8265.1 Priority  : n/a
Sync Interval     : 16 pkt/s
Local PTP Port    : 4                  PTP Port State     : slave
Clock Id          : 00233efffe69f250   Remote PTP Port    : 3
GM Clock Id       : 8887868584838281   GM Clock Class     : 6
GM Clock Accuracy : within 100 ns      GM Clock Variance  : 0x6400 (3.7E-09)
GM Clock Priority1: 128                GM Clock Priority2 : 128
Steps Removed     : 1                  Parent Clock       : yes
Tx Timestamp Point: port               Rx Timestamp Point : port
Last Tx Port      : 1/1/2              Last Rx Port       : 1/1/2
===============================================================================
```

On PE-2, a loopback address must assigned for PTP communication as follows:

```
*A:PE-2#
configure service vprn 10
    interface "ptp_loopback"
        address 172.16.1.1/32
        loopback
    exit
    source-address
        application ptp "ptp_loopback"
            exit
        exit
```

1588 as NTP Local Clock (server)

If the system is configured as a boundary clock or slave clock then the time recovered from the 1588 slave port can be used as the source of system time on the node. This allows for higher accuracy and better stability in the timebase when compared to NTP. To enable this, PTP must be made the preferred server in the NTP context in the node.

Note that if the system is acting as an NTP server or peer to other NTP clocks, then turning on this feature will impact the existing NTP topology. The system shall advertise itself as an NTP Stratum 1 server to external clients and peers. Given the much higher accuracies achievable with PTP time distribution, this change in topology does not degrade the time in the clients and peers.

```
*A:PE-1#
        configure system time ntp
            server ptp prefer
        exit
```

To validate PTP is now being used for NTP time and system time, use the following command:

```
*A:PE-1# show system ntp all
===============================================================================
NTP Status
===============================================================================
Configured         : Yes                Stratum              : 1
Admin Status       : up                 Oper Status          : up
Server Enabled     : No                 Server Authenticate  : No
Clock Source       : ptp
Auth Check         : Yes
Current Date & Time: 2014/05/30 17:53:11 UTC
===============================================================================
===============================================================================
NTP Active Associations
===============================================================================
```

```
State                    Reference ID    St Type  A  Poll Reach     Offset(ms)
    Remote
-------------------------------------------------------------------------------
chosen                   PTP             0  srvr  -  64   ......YY  0.000
    ptp
===============================================================================
===============================================================================
NTP Clients
===============================================================================
vRouter                                                    Time Last Request Rx
    Address
-------------------------------------------------------------------------------
===============================================================================
```

Conclusion

The systems provide support for IEEE 1588 frequency and time distribution for the synchronization applications of the mobile networks. They can be configured as frequency distribution grandmasters and slave clocks or time distribution boundary and slave clocks.

Synchronous Ethernet

This chapter provides information about Synchronous Ethernet (SyncE).

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

This information and configuration in this chapter is based on SR OS release 8.0.R7. There are no software prerequisites for this configuration, however, the hardware requires the use of Synchronous Ethernet capable MDA-XPs/CMA-XPs or the IMMs.

In addition, Synchronous Ethernet is only supported on optical interfaces. It is not supported on 10/100/1000 base copper interfaces.

Summary

Synchronous Ethernet (SyncE) is the ability to provide PHY-level frequency distribution through an Ethernet port. It is one of the building blocks of Next Generation Networks (NGNs).

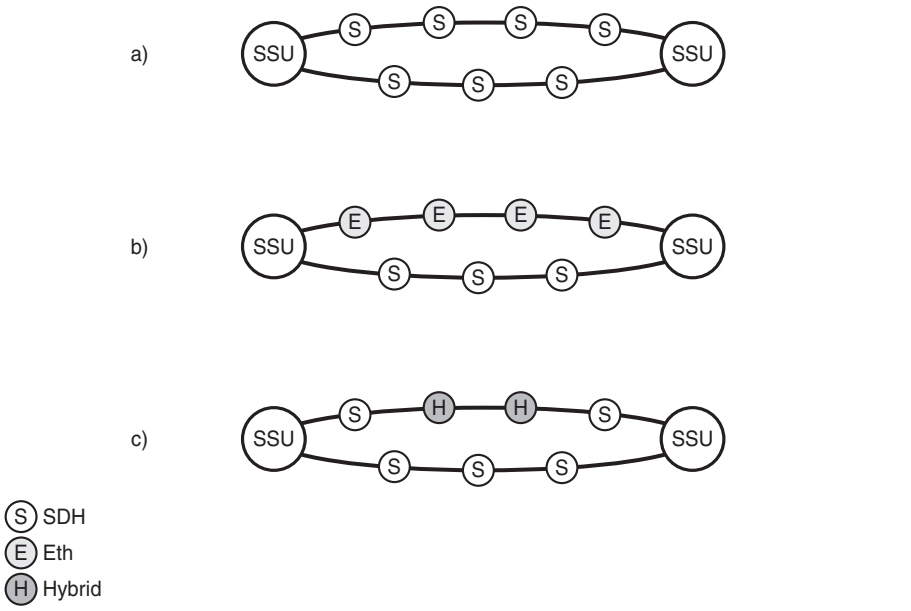
Overview

Synchronous Ethernet

Traditionally, Ethernet based networks employ the physical layer transmitter clock to be derived from an inexpensive +/-100ppm crystal oscillator and the receiver locks onto it. There is no need for long term frequency stability as the data is packetized and can be buffered. For the same reason, there is no need for consistency between the frequencies of different links. However, one could elect to derive the physical layer transmitter clock from a high quality frequency reference by replacing the crystal with a frequency source traceable to a primary reference clock. This would not affect the operation of any of the Ethernet layers, for which this change would be transparent. The receiver at the far end of the link would lock onto the physical layer clock of the received signal, and thus itself gain access to a highly accurate and stable frequency reference. Then, in a manner analogous to conventional hierarchical master-slave network synchronization, this receiver could lock the transmission clock of its other ports to this frequency reference and a fully time synchronous network could be established.

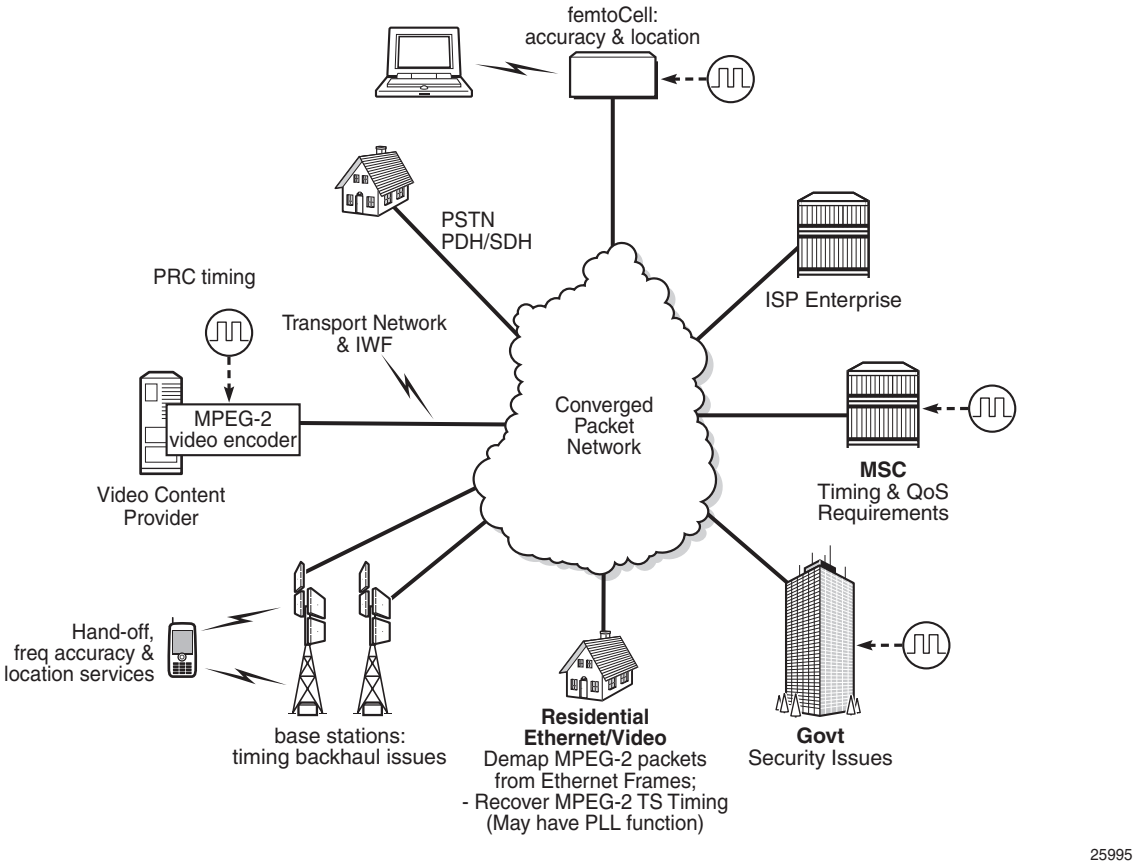
The advantage of using SyncE, as compared to methods relying on sending timing information in packets over an unlocked physical layer, is that SyncE is not influenced by impairments introduced by the higher levels of the networking technology (packet loss, packet delay variation). Therefore, the frequency accuracy and stability may be expected to exceed those of networks with unsynchronized physical layers. In addition, SyncE was designed to integrate into any existing SONET/SDH synchronization distribution architecture to allow for the easy migration from the traditional to the new synchronous interfaces. SyncE includes the concept of a hybrid switch which supports the interworking of synchronization distribution through SONET/SDH and the SyncE interfaces at the same time.

Figure 11 **SyncE Hypothetical Reference Network Architecture**



Many Tier 1 carriers are looking to migrate their synchronization infrastructure to a familiar and manageable model. In order to enable rapid migration of these networks, SyncE may be the easiest to deploy in order to ensure robust frequency synchronization.

Figure 12 **Packet Based Network Timing Infrastructure**

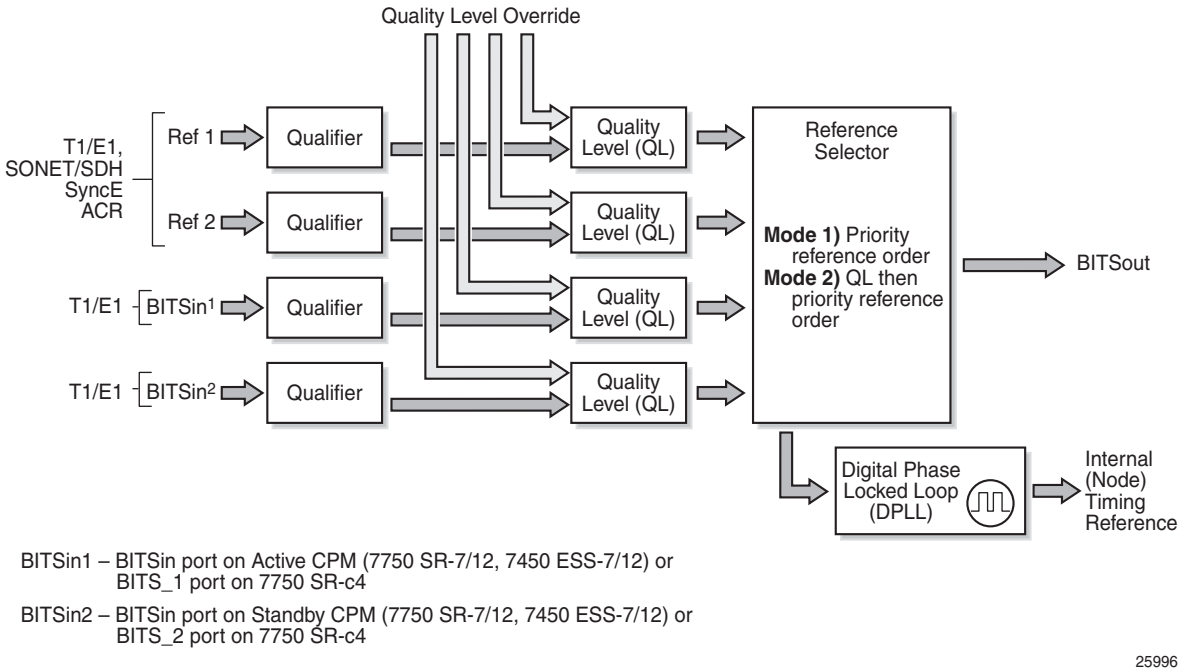


Central Synchronization Sub-System

The timing subsystem for the SR/ESS platforms has a central clock located on the Control Processor Module (CPM). The timing subsystem performs many of the duties of the network element clock as defined by Telcordia (GR-1244) and ITU-T G.781.

The system can select from up to four timing inputs to train the local oscillator. The priority order of these references must be specified. This is a simple ordered list of inputs: {BITS [Building Integrated Timing Source], ref1, ref2}. The CPM clock output has the ability to drive the clocking for all line cards in the system. The SR/ESS supports selection of the node reference using Quality Level (QL) indications.

Figure 13 **Current 7x50 Timing Sub-System Architecture**



The recovered clock is able to derive its timing from any of the following:

- OC3/STM1, OC12/STM4, OC48/STM16, OC192/STM64 ports
- T1/E1 CES channel (adaptive clocking)
- SyncE ports
- T1/E1 ports
- BITS port on a channelized OC3/STM1 CES CMA (7710 SR-c4, 7710 SR-c12, and the 7750 SR-c12)
- BITS port on the CPM or CFM module

On 7750 SR-12 and 7750 SR-7 systems with redundant CPMs, the system has two BITS input ports (one per CPM). On the 7750 SRc-4 systems, there are two BITS input ports on the chassis front plate. These BITS input ports provide redundant synchronization inputs from an external BITS/SSU. Note: the 7750 SR-c12 does not support BITS input port redundancy or BITS out.

All settings of the signal characteristics for the BITS input apply to both ports. When the active CPM considers the BITS input as a possible reference, it will consider first the BITS input port on the active CPM followed the BITS input port on the standby CPM in that relative priority order. This relative priority order is in addition to the user definable ref-order. For example, a ref-order of 'bits-ref1-ref2' would actually be BITS in (active CPM) followed by BITS in (standby CPM) followed by ref1 followed by ref2. When ql-selection is enabled, then the QL of each BITS input port is viewed independently. The higher QL source is chosen.

On the 7750 SR-c4 platform CFM, there are two BITS input ports and two BITS output ports on this one module. These two ports are provided for BITS redundancy for the chassis. All settings of the signal characteristics for the BITS input apply to both ports. This includes the ql-override setting. When the CFM considers the BITS input as a possible reference, it will consider first the BITS input port "bits1" followed the BITS input port "bits2" in that relative priority order. This relative priority order is in addition to the user definable ref-order. For example, a ref-order of 'bits-ref1-ref2' would actually be "bits1" followed by "bits2" followed by ref1 followed by ref2. When ql-selection is enabled, then the QL of each BITS input port is viewed independently. The higher QL source is chosen.

The BITS output ports are provided to deliver a unfiltered recovered line clock from a SR/ESS port directly to a dedicated timing device in the facility (BITS or Standalone Synchronization Equipment (SASE) device). The signal selected will be one of ref1 or ref2. It cannot be the BITS input port signal nor can it be the output of the central clock.

When QL selection mode is disabled, then the reversion setting controls when the central clock can re-select a previously failed reference.

Table 1 **Revertive, Non-Revertive Timing Reference Switching Operation**

Status of Reference A	Status of Reference B	Active Reference Non-revertive Case	Active Reference Revertive Case
OK	OK	A	A
Failed	OK	B	B
OK	OK	B	A
OK	Failed	A	A
OK	OK	A	A
Failed	Failed	holdover	holdover
OK	Failed	A	A
Failed	Failed	holdover	holdover

Table 1 **Revertive, Non-Revertive Timing Reference Switching Operation (Continued)**

Status of Reference A	Status of Reference B	Active Reference Non-revertive Case	Active Reference Revertive Case
Failed	OK	B	B
Failed	Failed	holdover	holdover
OK	OK	A or B	A

Synchronization Status Messages (SSM)

SSM provides a mechanism to allow the synchronization distribution network to both determine the quality level of the clock sourcing a given synchronisation trail and to allow a network element to select the best of multiple input synchronization trails. Synchronization Status messages have been defined for various transport protocols including SONET/SDH, T1/E1, and SyncE, for interaction with office clocks, such as BITS or SSUs (synchronisation supply unit) and embedded network element clocks.

SSM allows equipment to autonomously provision and reconfigure (by reference switching) their synchronization references, while helping to avoid the creation of timing loops. These messages are particularly useful to allow synchronization reconfigurations when timing is distributed in both directions around a ring.

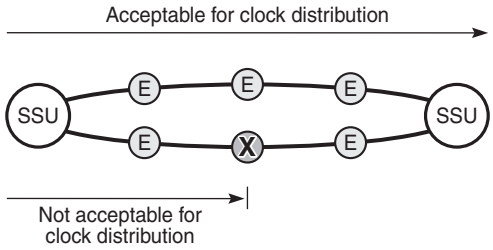
In SyncE, the SSM is provided through the Ethernet Synchronization Messaging Channel (ESMC). This mechanism uses Ethernet OAM PDU to exchange the Quality Level values over the SyncE link.

SyncE Chains

Transmission of a reference clock through a chain of Ethernet equipment requires that all of the equipment support SyncE.

A single piece of equipment not capable of SyncE breaks the chain as shown in Figure 14. Ethernet frames will still get through but downstream device will recognize that the signal is out of pull-in range and not use it for reference.

Figure 14 **Network Considerations for Ethernet Timing Distribution**



Configuration

Configuration 1

The following example shows the configuration options for SyncE when ql-selection mode is disabled. Generally, North American SONET networks do not use the automatic reference selection mechanisms. If SyncE is being added into such a network, it would likely have ql-selection set to disabled.

```
A:PE-1>config# card 1 mda 1
A:PE-1>config>card>mda#
     access          + Configure access MDA parameters
     egress          + Configure egress MDA parameters
[no] hi-bw-mcast-src - Enable/disable allocation of resources for high
                       bandwidth multicast streams
     ingress         + Configure ingress MDA parameters
[no] mda-type        - Provisions/de-provisions an MDA to/from the device
                       configuration for the slot
     named-pool-mode + Enable/Disable named pool mode
     network         + Configure network MDA parameters
[no] shutdown        - Administratively shut down an mda
[no] sync-e          - Enable/Disable Synchronous Ethernet
A:PE-1>config>card>mda# sync-e
*A:PE-1>config>card>mda# info detail
----------------------------------------------
           mda-type m20-1gb-xp-sfp
           sync-e
           named-pool-mode
               ingress
                   no named-pool-policy
               exit
               egress
                   no named-pool-policy
               exit
           exit
           ingress
```

```
                no hsmda-pool-policy
                no scheduler-policy
            exit
            egress
                no hsmda-pool-policy
            exit
            network
                ingress
                    pool default
        ...
```

```
*A:PE-1>config>system# sync-if-timing
*A:PE-1>config>system>sync-if-timing#
      abort           - Discard the changes that have been made to sync
                        interface timing during a session
      begin           - Switch to edit mode for sync interface timing - use
                        commit to save or abort to discard the changes made in
                        a session
      bits            + Configure parameters for the Building Integrated Timing
                        Supply(BITS)
      commit          - Save the changes made to sync interface timing during a
                        session
 [no] ql-selection    - Enable/disable reference selection based on
                        quality-level
 [no] ref-order       - Priority order of timing references
      ref1            + Configure parameters for the first timing reference
      ref2            + Configure parameters for the second timing reference
 [no] revert          - Revert/do not revert to a higher priority re-validated
                        reference source
*A:PE-1>config>system>sync-if-timing# begin
*A:PE-1>config>system>sync-if-timing# ref-order bits ref1
*A:PE-1>config>system>sync-if-timing# bits input no shutdown
*A:PE-1>config>system>sync-if-timing# bits interface-type ds1 esf
*A:PE-1>config>system>sync-if-timing# revert
*A:PE-1>config>system>sync-if-timing# ref1
*A:PE-1>config>system>sync-if-timing>ref1#
 [no] ql-override     - Override the quality level of a timing reference
 [no] shutdown        - Administratively shutdown the timing reference
 [no] source-port     - Configure the source port for the first timing reference
*A:PE-1>config>system>sync-if-timing>ref1# source-port 1/1/2
*A:PE-1>config>system>sync-if-timing>ref1# no shutdown
*A:PE-1>config>system>sync-if-timing>ref1# exit
*A:PE-1>config>system>sync-if-timing# commit
*A:PE-1>config>system>sync-if-timing# info detail
----------------------------------------------
           no ql-selection
           ref-order bits ref1 ref2
           ref1
               source-port 1/1/2
               no shutdown
               no ql-override
           exit
           ref2
               shutdown
               no source-port
               no ql-override
```

```
            exit
            bits
                interface-type ds1 esf
                no ql-override
                input
                    no shutdown
                exit
                output
                    shutdown
                    line-length 110
                exit
            exit
            revert
```

The following output displays the associated show information.

```
*A:PE-1>show>system# sync-if-timing
===============================================================================
System Interface Timing Operational Info
===============================================================================
System Status CPM A                : Master Locked
    Reference Input Mode           : Revertive
    Quality Level Selection        : Disabled
    Reference Selected             : ref1
    System Quality Level           : unknown
    Current Frequency Offset (ppm) : -5

Reference Order                    : bits ref1 ref2

Reference Mate CPM
    Qualified For Use              : No
        Not Qualified Due To       :     LOS
    Selected For Use               : No
        Not Selected Due To        :     not qualified

Reference Input 1
    Admin Status                   : up
    Rx Quality Level               : unknown
    Quality Level Override         : none
    Qualified For Use              : Yes
    Selected For Use               : Yes
    Source Port                    : 1/1/2

Reference Input 2
    Admin Status                   : down
    Rx Quality Level               : unknown
    Quality Level Override         : none
    Qualified For Use              : No
        Not Qualified Due To       :     disabled
    Selected For Use               : No
        Not Selected Due To        :     disabled
    Source Port                    : None

Reference BITS A
    Input Admin Status             : up
    Rx Quality Level               : failed
    Quality Level Override         : none
```

```
    Qualified For Use              : No
        Not Qualified Due To       :     LOS
    Selected For Use               : No
        Not Selected Due To        :     not qualified
    Interface Type                 : DS1
    Framing                        : ESF
    Line Coding                    : B8ZS
===============================================================================
*A:PE-1>show>system#
```

Configuration 2

The following example shows the configuration options for SyncE when ql-selection mode is enabled.

This is the normal case for European SDH networks.

```
A:PE-1>config# card 1 mda 1
A:PE-1>config>card>mda#
     access          + Configure access MDA parameters
     egress          + Configure egress MDA parameters
[no] hi-bw-mcast-src - Enable/disable allocation of resources for high
                       bandwidth multicast streams
     ingress         + Configure ingress MDA parameters
[no] mda-type        - Provisions/de-provisions an MDA to/from the device
                       configuration for the slot
     named-pool-mode + Enable/Disable named pool mode
     network         + Configure network MDA parameters
[no] shutdown        - Administratively shut down an mda
[no] sync-e          - Enable/Disable Synchronous Ethernet
A:PE-1>config>card>mda# sync-e
A:PE-1>config>card>mda# info detail
-----------------------------------------------
           mda-type m20-1gb-xp-sfp
           sync-e
           named-pool-mode
               ingress
                   no named-pool-policy
               exit
               egress
                   no named-pool-policy
               exit
           exit
           ingress
               no hsmda-pool-policy
               no scheduler-policy
           exit
           egress
               no hsmda-pool-policy
           exit
           network
               ingress
                   pool default
```

```
     ...


A:PE-1>config# port 1/1/2 ethernet ssm
A:PE-1>config>port>ethernet>ssm#
A:PE-1>config>port>ethernet>ssm#
 [no] code-type       - Set the SSM channel to either use sonet or sdh
 [no] shutdown        - Enable/Disable SSM
 [no] tx-dus          - Enable/disable always transmit 0xF (dus/dnu) in SSM messaging
                        channel

A:PE-1>config>port>ethernet>ssm# code-type sdh
*A:PE-1>config>port>ethernet>ssm# no shutdown
*A:PE-1>config>port>ethernet>ssm# info detail
----------------------------------------------
               code-type sdh
               no tx-dus
               no shutdown
----------------------------------------------
*A:PE-1>config>port>ethernet>ssm#



*A:PE-1>config>system# sync-if-timing
*A:PE-1>config>system>sync-if-timing#
      abort           - Discard the changes that have been made to sync
                        interface timing during a session
      begin           - Switch to edit mode for sync interface timing - use
                        commit to save or abort to discard the changes made in
                        a session
      bits            + Configure parameters for the Building Integrated Timing
                        Supply(BITS)
      commit          - Save the changes made to sync interface timing during a
                        session
 [no] ql-selection    - Enable/disable reference selection based on
                        quality-level
 [no] ref-order       - Priority order of timing references
      ref1            + Configure parameters for the first timing reference
      ref2            + Configure parameters for the second timing reference
 [no] revert          - Revert/do not revert to a higher priority re-validated
                        reference source
*A:PE-1>config>system>sync-if-timing# begin
*A:PE-1>config>system>sync-if-timing# ref-order bits ref1
*A:PE-1>config>system>sync-if-timing# ql-selection
*A:PE-1>config>system>sync-if-timing# bits input no shutdown
*A:PE-1>config>system>sync-if-timing# bits interface-type ds1 esf
*A:PE-1>config>system>sync-if-timing# bits ql-override prc
*A:PE-1>config>system>sync-if-timing# revert
*A:PE-1>config>system>sync-if-timing# ref1
*A:PE-1>config>system>sync-if-timing>ref1#
 [no] ql-override     - Override the quality level of a timing reference
 [no] shutdown        - Administratively shutdown the timing reference
 [no] source-port     - Configure the source port for the first timing reference
*A:PE-1>config>system>sync-if-timing>ref1# source-port 1/1/2
*A:PE-1>config>system>sync-if-timing>ref1# no shutdown
*A:PE-1>config>system>sync-if-timing>ref1# exit
*A:PE-1>config>system>sync-if-timing# commit
*A:PE-1>config>system>sync-if-timing# info detail
----------------------------------------------
           ql-selection
```

```
            ref-order bits ref1 ref2
            ref1
                source-port 1/1/2
                no shutdown
                no ql-override
            exit
            ref2
                shutdown
                no source-port
                no ql-override
            exit
            bits
                interface-type ds1 esf
                ql-override prc
                input
                    no shutdown
                exit
                output
                    shutdown
                    line-length 110
                exit
            exit
            revert
```

The following output displays the associated show information.

```
*A:PE-1>show>system# sync-if-timing
===============================================================================
System Interface Timing Operational Info
===============================================================================
System Status CPM A                : Master Locked
    Reference Input Mode           : Revertive
    Quality Level Selection        : Enabled
    Reference Selected             : ref1
    System Quality Level           : prc
    Current Frequency Offset (ppm) : -5

Reference Order                    : bits ref1 ref2

Reference Mate CPM
    Qualified For Use              : No
        Not Qualified Due To       :     LOS
    Selected For Use               : No
        Not Selected Due To        :     not qualified

Reference Input 1
    Admin Status                   : up
    Rx Quality Level               : prc
    Quality Level Override         : none
    Qualified For Use              : Yes
    Selected For Use               : Yes
    Source Port                    : 1/1/2

Reference Input 2
    Admin Status                   : down
    Rx Quality Level               : unknown
    Quality Level Override         : none
```

```
    Qualified For Use              : No
        Not Qualified Due To       :     disabled
    Selected For Use               : No
        Not Selected Due To        :     disabled
    Source Port                    : None

Reference BITS A
    Input Admin Status             : up
    Rx Quality Level               : failed
    Quality Level Override         : none
    Qualified For Use              : No
        Not Qualified Due To       :     LOS
    Selected For Use               : No
        Not Selected Due To        :     not qualified
    Interface Type                 : DS1
    Framing                        : ESF
    Line Coding                    : B8ZS
===============================================================================
*A:PE-1>show>system#
```

Conclusion

With the world rapidly transitioning to IP/MPLS-based NGNs with Ethernet as the transport medium of choice, there is an increasing need to enhance services and capabilities while still leveraging existing infrastructure, thereby easing the transition while continuing to increase revenue and reduce the Total Cost of Ownership (TCO). In areas such as mobile backhaul, TDM CES etc., these requirements create a need for SONET/SDH-like frequency synchronization capability in the inherently asynchronous Ethernet network.

SyncE, natively supported on the Nokia SR OS routers, is an ITU-T standardized PHY-level way of transmitting frequency synchronization across Ethernet packet networks that fulfills that need in a reliable, secure, scalable, efficient, and cost-effective manner. It allows service providers to keep existing revenue streams alive and create new ones while simplifying the network design and reducing TCO.

System Management

In This Section

This section provides configuration information for the following topics:

- Distributed CPU Protection
- Event Handling System

Distributed CPU Protection

This chapter describes Distributed CPU Protection (DCP) configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This Distributed CPU Protection (DCP) configuration example was created using the 7750 SR-c12 platform but is equally applicable to the following platforms: 7750 SR-7/12, 7450 ESS-6/7/12, 7750 SR-c4/c12 and 7950 XRS. DCP is not supported on the 7750 SR-1, 7450 ESS-1 or 7710 SR platforms.

DCP operates on the line cards and requires line cards with the FP2 or greater hardware (for example, IOM3-XP, IMMs and C-XMAs).

The configuration was tested on release 11.0R1.

Overview

SR OS provides several rate limiting mechanisms to protect the CPM/CFM processing resources of the router:

- CPU Protection: A centralized rate limiting function that operates on the CPM to limit traffic destined to the CPUs.
- Distributed CPU Protection: A control traffic rate limiting protection mechanism for the CPM/CFM that operates on the line cards (hence ‘distributed’). CPU protection protects the CPU of the node that it is configured on from a DOS attack by limiting the amount of traffic coming in from one of its ports and destined to the CPM (to be processed by its CPU) using a combination of the configurable limits.

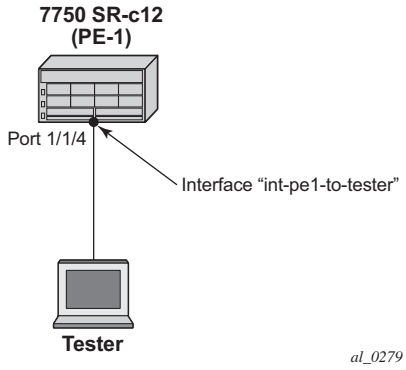
The goal of this example is to familiarize the reader with the configuration and use of Distributed CPU Protection. A simple and controlled setup is used to illustrate how the protection behaves and how to use the tools provided for the feature.

External testing equipment (“tester”) is used to send control traffic of various protocols at various rates to the router in order to exercise DCP. Log events and show routines are examined to explain the indications that the router provides to an operator.

Configuration

The test topology is shown in Figure 15. A Gigabit Ethernet link is used between the Tester and the router.

Figure 15 **Test Topology**



Step 1. The basic configuration of the mda, port, interface and a security event log on the router is shown below.

```
*A:PE-1# configure card 1 mda 1
*A:PE-1>config>card>mda# info
----------------------------------------------
            mda-type m5-1gb-sfp-b
            no shutdown
----------------------------------------------
*A:PE-1>config>card>mda# exit all
*A:PE-1# configure port 1/1/4
*A:PE-1>config>port# info
----------------------------------------------
        ethernet
        exit
        no shutdown
----------------------------------------------
*A:PE-1>config>port# exit all
```

```
*A:PE-1# configure router interface "int-pe1-to-tester"
*A:PE-1>config>router>if# info
----------------------------------------------
            address 192.168.10.1/24
            port 1/1/4
            no shutdown
----------------------------------------------
*A:PE-1>config>router>if# exit all
*A:PE-1# configure log log-id 15
*A:PE-1>config>log>log-id# info
----------------------------------------------
            from security
            to memory 1024
----------------------------------------------
```

This example was developed on a 7750 SR-c12 platform but it is equally applicable to other platforms such as the 7750 SR-7/12. If other platforms, such as the 7750 SR-7/12 that support centralized CPU Protection, are used to explore Distributed CPU Protection then the centralized CPU Protection should be disabled (for the purposes of this example) so that it does not interfere with reproducing the same results as described below. In a normal production network CPU Protection and DCP are complimentary and can be used together. To disable centralized CPU Protection for the purposes of reproducing the results below please ensure that:

- **protocol-protection** is disabled.
- All rates in all polices (including any default polices) are configure to **max**.

Step 2. In order to activate DCP a policy is created and assigned to the interface.

The first policy that is used in this example is used to simply count protocol packets to see that they are indeed flowing from the tester to the router and being extracted and indentified.

The *dcp-policy-count* policy is configured as follows:

```
*A:PE-1# configure system security dist-cpu-protection
*A:PE-1>config>sys>security>dist-cpu-protection# info
----------------------------------------------
                policy "dcp-policy-count" create
                    description "Static policers with rate 0 for counting packets"
                    static-policer "sp-arp" create
                        rate packets 0 within 1
                    exit
                    static-policer "sp-icmp" create
                        rate packets 0 within 1
                    exit
                    static-policer "sp-igmp" create
                        rate packets 0 within 1
                    exit
                    protocol arp create
                        enforcement static "sp-arp"
```

```
        exit
        protocol icmp create
            enforcement static "sp-icmp"
        exit
        protocol igmp create
            enforcement static "sp-igmp"
        exit
    exit
```

For the *dcp-policy-count policy* configuration:

- The policy contains three static policers: s*p-arp*, *sp-icmp* and *sp-igmp*. These policers are then used by the three configured protocols that are part of the policy: *arp*, *icmp* and *igmp*.
- The list of protocols that are applicable to DCP are as follows: arp, dhcp, http-redirect, icmp, igmp, mld, ndis, pppoe-pppoa, all-unspecified, mpls-ttl, bfd-cpm, bgp, eth-cfm, isis, ldp, ospf, pim and rsvp. The all-unspecified protocol is a special “catch-all”. Please see the 7750 SR OS System Management Guide for more details.
- This policy instantiates three permanent (static) policers for every object (for example, interface) that the policy is associated with.
- The three protocols each reference their own static policer so each protocol will be independently rate limited. A single static policer can also be used to rate limit multiple protocols but that capability is not used in this example.
- The rate is set to 0 which means all packets will be considered as non-conformant to the policer. This configuration is used to provide counters of protocol packets. The DCP counters provide the count of packets exceeding the policing parameters since the given policer was previously declared as conformant or newly instantiated. A rate of zero ensures that the policer will never be declared as conformant and hence will never reset the counters.
- The exceed-action is not configured and takes the default value of *none.* The *log-events* parameter is not configured and is enabled by default. That means the policer will notify the operator when the first packet arrives but will not discard or mark any packets.

Step 3. Assign the *dcp-policy-count* to the interface:

```
*A:PE-1# configure router interface "int-pe1-to-tester"
*A:PE-1>config>router>if# dist-cpu-protection "dcp-policy-count"
```

Step 4. Examine some log and status on the router to get a baseline (no traffic is flowing from the tester to the router at this point). Notice that the cpu utilization is fairly low with an overall Idle of 96% and no task groups at more than 5% capacity usage. Future example output from this show routine will be snipped to only show relevant and interesting lines.

```
*A:PE-1# show system cpu
===============================================================================
CPU Utilization (Sample period: 1 second)
===============================================================================
Name                                  CPU Time       CPU Usage         Capacity
                                        (uSec)                           Usage
-------------------------------------------------------------------------------
BFD                                          0           0.00%           0.00%
BGP                                     28,779           0.32%           0.47%
BGP PE-CE                                    0           0.00%           0.00%
CFLOWD                                   7,384           0.08%           0.38%
Cards & Ports                           65,941           0.73%           5.35%
DHCP Server                                 55          ~0.00%          ~0.00%
ICC                                      1,195           0.01%           0.06%
IGMP/MLD                                 1,883           0.02%           0.12%
IMSI Db Appl                               120          ~0.00%          ~0.00%
IOM                                    132,522           1.47%           3.11%
IP Stack                                 7,666           0.08%           0.39%
IS-IS                                    1,415           0.01%           0.07%
ISA                                     11,988           0.13%           0.43%
LDP                                        496          ~0.00%           0.04%
Logging                                    185          ~0.00%           0.01%
MBUF                                         0           0.00%           0.00%
MPLS/RSVP                                6,219           0.06%           0.48%
MSCP                                         0           0.00%           0.00%
MSDP                                         0           0.00%           0.00%
Management                               4,077           0.04%           0.13%
OAM                                     10,311           0.11%           0.44%
OSPF                                       661          ~0.00%           0.05%
PIM                                          0           0.00%           0.00%
RIP                                          0           0.00%           0.00%
RTM/Policies                                 0           0.00%           0.00%
Redundancy                               7,641           0.08%           0.51%
SNMP Daemon                                  0           0.00%           0.00%
Services                                 3,965           0.04%           0.09%
Stats                                        0           0.00%           0.00%
Subscriber Mgmt                          7,437           0.08%           0.44%
System                                  57,081           0.63%           3.49%
Traffic Eng                                  0           0.00%           0.00%
VRRP                                     1,918           0.02%           0.09%
WEB Redirect                                77          ~0.00%          ~0.00%
-------------------------------------------------------------------------------
Total                                8,965,427         100.00%
   Idle                              8,605,657          95.98%
   Usage                               359,770           4.01%
Busiest Core Utilization               134,481          13.49%
===============================================================================
```

The DCP feature is reporting no violations for interfaces on card 1.

```
*A:PE-1# tools dump security dist-cpu-protection violators enforcement interface card
1
===============================================================================
Distributed Cpu Protection Current Interface Enforcer Policer Violators
===============================================================================
Interface                       Policer/Protocol                    Hld Rem
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Violators on Slot-1 Fp-1
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
[S]-Static [D]-Dynamic [M]-Monitor
-------------------------------------------------------------------------------
===============================================================================
```

There are no security log events.

```
*A:PE-1# show log log-id 15
===============================================================================
Event Log 15
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=1024   next event=1  (not wrapped)]
```

The detailed DCP status for the interface shows all three policers are currently in the conform state.

```
*A:PE-1# show router interface "int-pe1-to-tester" dist-cpu-protection
===============================================================================
Interface "int-pe1-to-tester" (Router: Base)
===============================================================================
Distributed CPU Protection Policy :  dcp-policy-count
-------------------------------------------------------------------------------
Statistics/Policer-State Information
===============================================================================
-------------------------------------------------------------------------------
Static Policer
-------------------------------------------------------------------------------
Policer-Name        : sp-arp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : arp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none

Policer-Name        : sp-icmp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : icmp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none

Policer-Name        : sp-igmp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : igmp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Local-Monitoring Policer
```

```
-------------------------------------------------------------------------------
No entries found
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Dynamic-Policer (Protocol)
-------------------------------------------------------------------------------
No entries found
-------------------------------------------------------------------------------
===============================================================================
```

Step 5. Configure the tester to send ARP, ICMP and IGMP traffic to the router using the following rates:

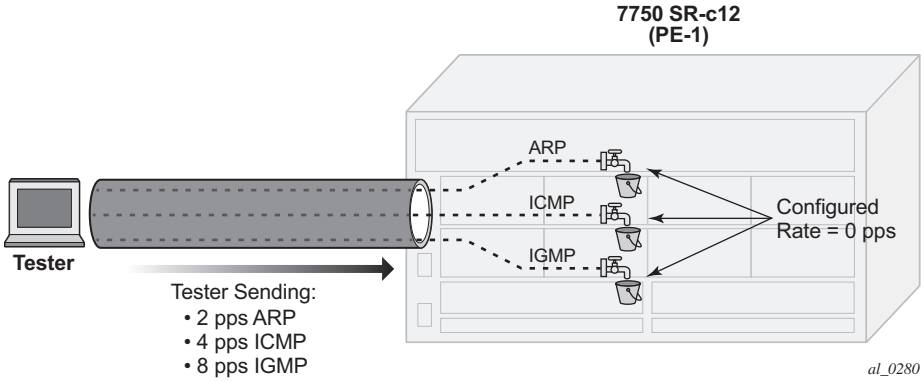
- ARP: 2 packets per second (pps)
- ICMP: 4 pps
- IGMP: 8 pps

Here are some tips for how to configure the tester to send protocol packets that will be recognized by the router:

- ARP:
  - Set the MAC destination address to FF-FF-FF-FF-FF-FF
  - Use an ARP Request format
- ICMP:
  - Use an icmp type of 8 (echo request, such as **ping**).
  - Set the MAC destination address equal to the MAC address of the receiving port. The MAC address of port 1/1/4 can be seen in the output of show port 1/1/4 as the Configured Address.
  - Set the IP destination address to 192.168.10.1 and the IP source address to 192.168.10.2.
- IGMP:
  - Set the MAC destination address equal to the MAC address of the receiving port. The MAC address of port 1/1/4 can be seen in the output of show port 1/1/4 as the Configured Address.
  - Set the IP destination address to 224.0.0.2 and the IP source address to 0.0.0.0.
  - Set the IGMP version to 2, make the IGMP message type a Membership Query to Group 0.

Also ensure that the tester interleaves the three streams of protocol packets such that it schedules them independently in an interleaved fashion, not serially.

Figure 16 **Count Traffic with DCP Policy Count**



Step 6. Notice that DCP now reports some violations of the policy against the interface.

```
*A:PE-1# tools dump security dist-cpu-protection violators enforcement interface card
1
===============================================================================
Distributed Cpu Protection Current Interface Enforcer Policer Violators
===============================================================================
Interface                       Policer/Protocol                    Hld Rem
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Violators on Slot-1 Fp-1
-------------------------------------------------------------------------------
int-pe1-to-tester               sp-arp                           [S] none
int-pe1-to-tester               sp-icmp                          [S] none
int-pe1-to-tester               sp-igmp                          [S] none
-------------------------------------------------------------------------------
[S]-Static [D]-Dynamic [M]-Monitor
-------------------------------------------------------------------------------
===============================================================================
```

After a few seconds the DCP exceed-count values can be seen incrementing.

Note the following details:

- Exceed-Count is non-zero. This will continue incrementing and will never reset since the rate configured in the DCP policy is zero.
- The Policer-State is Exceed. The policers have detected that the protocol is non-conformant to the configured rate.
- Detec. Time Remain stays at 29 seconds. This countdown timer is automatically reset to 30 seconds every time a policer is detected as non-conformant (which will be continually when the rate is set to 0 and packets of that protocol are being received).

```
*A:PE-1# show router interface "int-pe1-to-tester" dist-cpu-protection
===============================================================================
```

```
Interface "int-pe1-to-tester" (Router: Base)
===============================================================================
Distributed CPU Protection Policy :  dcp-policy-count
-------------------------------------------------------------------------------
Statistics/Policer-State Information
===============================================================================
-------------------------------------------------------------------------------
Static Policer
-------------------------------------------------------------------------------
Policer-Name        : sp-arp
Card/FP             : 1/1               Policer-State       : Exceed
Protocols Mapped    : arp
Exceed-Count        : 72
Detec. Time Remain  : 29 seconds        Hold-Down Remain.   : none

Policer-Name        : sp-icmp
Card/FP             : 1/1               Policer-State       : Exceed
Protocols Mapped    : icmp
Exceed-Count        : 144
Detec. Time Remain  : 29 seconds        Hold-Down Remain.   : none

Policer-Name        : sp-igmp
Card/FP             : 1/1               Policer-State       : Exceed
Protocols Mapped    : igmp
Exceed-Count        : 290
Detec. Time Remain  : 29 seconds        Hold-Down Remain.   : none
-------------------------------------------------------------------------------
[snip]
```

Step 7. Keep the tester running.

Now a DCP policy that enforces protocol rates using static policers will be applied to the interface. First, the policy is created:

```
*A:PE-1# configure system security dist-cpu-protection
*A:PE-1>config>sys>security>dist-cpu-protection# policy "dcp-static-policy-1" create
                    description "Static policers for arp, icmp and igmp"
                    static-policer "sp-arp" create
                        rate packets 10 within 1
                        exceed-action discard
                    exit
                    static-policer "sp-icmp" create
                        rate packets 20 within 1
                        exceed-action discard
                    exit
                    static-policer "sp-igmp" create
                        rate packets 10 within 1
                        exceed-action discard
                    exit
                    protocol arp create
                        enforcement static "sp-arp"
                    exit
                    protocol icmp create
                        enforcement static "sp-icmp"
                    exit
                    protocol igmp create
                        enforcement static "sp-igmp"
```

```
            exit
        exit
```

For the dcp-static-policy-1 policy configuration, note that a few parameters are different than in the previously created dcp-policy-cou*nt* policy:

- The rates are set to low (but non-zero) values.
- The exceed-action is configured such that packets are dropped once the rate is exceeded.

Now assign the policy to the test interface:

```
*A:PE-1# configure router interface "int-pe1-to-tester"
*A:PE-1>config>router>if# dist-cpu-protection "dcp-static-policy-1"
*A:PE-1>config>router>if# exit all
*A:PE-1# show system security dist-cpu-protection policy "dcp-static-policy-1"
association
===============================================================================
Distributed CPU Protection Policy
===============================================================================
Policy Name : dcp-static-policy-1
Description : Static policers for arp, icmp and igmp

-------------------------------------------------------------------------------
Associations
-------------------------------------------------------------------------------

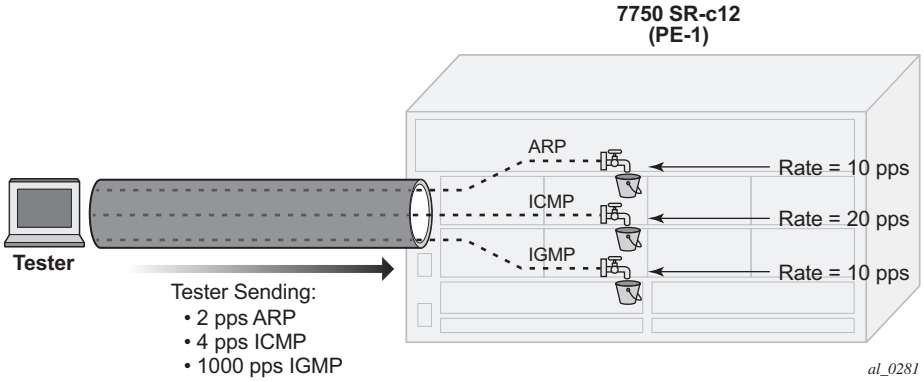
SAP associations
-------------------------------------------------------------------------------
  None

Managed SAP associations
-------------------------------------------------------------------------------
  None

Interface associations
-------------------------------------------------------------------------------
Router-Name : Base
  int-pe1-to-tester
-------------------------------------------------------------------------------
Number of interfaces : 1
===============================================================================
```

Step 8. Increase the rate of IGMP packets that the tester is sending to 1000pps (keep ARP and ICMP at 2pps and 4pps).

Figure 17 **Limit Traffic with dcp-static-policy-1**



Step 9. Notice that the system has identified a violation of the DCP rates for the igmp policer.

```
*A:PE-1# tools dump security dist-cpu-protection violators enforcement interface card
1
===============================================================================
Distributed Cpu Protection Current Interface Enforcer Policer Violators
===============================================================================
Interface                       Policer/Protocol                    Hld Rem
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Violators on Slot-1 Fp-1
-------------------------------------------------------------------------------
int-pe1-to-tester             sp-igmp                           [S] none
-------------------------------------------------------------------------------
[S]-Static [D]-Dynamic [M]-Monitor
-------------------------------------------------------------------------------
===============================================================================
```

After a few minutes the violation will be indicated as a log event. This delay is due to the design of DCP. In order to support large scale operation of DCP, and also to avoid overload conditions, a polling process is used to monitor state changes in the policers and to gather violations. This means there can be a delay between when an event occurs in the data plane and when the relevant state change or event notification occurs towards an operator, but in the meantime the policers are still operating and protecting the control plane.

```
*A:PE-1# show log log-id 15
===============================================================================
Event Log 15
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=1024   next event=11  (not wrapped)]

10 2013/04/18 17:31:54.58 EDT WARNING: SECURITY #2066 Base DCPUPROT
"Non conformant network_if "int-pe1-to-tester" on fp 1/1 detected at 04/18/2013
```

```
17:31:33. Policy "dcp-static-policy-1". Policer="sp-igmp"(static). Excd count=135"
… [snip] …
```

The status of DCP on the interface also shows the igmp policer as being in an Exceed state:

```
*A:PE-1# show router interface "int-pe1-to-tester" dist-cpu-protection
===============================================================================
Interface "int-pe1-to-tester" (Router: Base)
===============================================================================
Distributed CPU Protection Policy :  dcp-static-policy-1
-------------------------------------------------------------------------------
Statistics/Policer-State Information
===============================================================================
-------------------------------------------------------------------------------
Static Policer
-------------------------------------------------------------------------------
Policer-Name        : sp-arp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : arp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none

Policer-Name        : sp-icmp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : icmp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none

Policer-Name        : sp-igmp
Card/FP             : 1/1               Policer-State       : Exceed
Protocols Mapped    : igmp
Exceed-Count        : 19031
Detec. Time Remain  : 29 seconds        Hold-Down Remain.   : none
-------------------------------------------------------------------------------
…[snip]…
```

The CPU utilization of the IGMP task group is not impacted since DCP is discarding packets that are non-conformant to the configure rate.

```
*A:PE-1# show system cpu
===============================================================================
CPU Utilization (Sample period: 1 second)
===============================================================================
Name                                    CPU Time       CPU Usage        Capacity
                                          (uSec)                          Usage
-------------------------------------------------------------------------------
BFD                                            0           0.00%           0.00%
…[snip]…
IGMP/MLD                                   1,883           0.02%       0.12%
IMSI Db Appl                                 120          ~0.00%          ~0.00%
IOM                                        132,522           1.47%           3.11%
IP Stack                                     7,666           0.08%           0.39%
IS-IS                                        1,415           0.01%           0.07%
ISA                                         11,988           0.13%           0.43%
LDP                                            496          ~0.00%           0.04%
```

```
…[snip]…
WEB Redirect                                    77          ~0.00%          ~0.00%
-------------------------------------------------------------------------------
Total                                    8,965,427         100.00%
   Idle                                  8,605,657          95.98%
   Usage                                   359,770           4.01%
Busiest Core Utilization                   134,481          13.49%
===============================================================================
```

Step 10. Remove the DCP policy from the interface and see the CPU utilization goes up for the IGMP task group.

```
*A:PE-1# configure router interface "int-pe1-to-tester"
*A:PE-1>config>router>if# no dist-cpu-protection
*A:PE-1>config>router>if# /show system cpu
===============================================================================
CPU Utilization (Sample period: 1 second)
===============================================================================
Name                                    CPU Time       CPU Usage        Capacity
                                          (uSec)                           Usage
-------------------------------------------------------------------------------
BFD                                             0           0.00%           0.00%
…[snip]…
IGMP/MLD                                   82,142           0.91%           8.14%
IMSI Db Appl                                   98          ~0.00%          ~0.00%
IOM                                       129,851           1.45%           3.15%
IP Stack                                  196,549           2.19%          19.35%
IS-IS                                       1,484           0.01%           0.07%
ISA                                        11,765           0.13%           0.42%
LDP                                           449          ~0.00%           0.04%
…[snip]…
WEB Redirect                                  102          ~0.00%           0.01%
-------------------------------------------------------------------------------
Total                                    8,948,806         100.00%
   Idle                                  8,259,903          92.30%
   Usage                                   688,903           7.69%
Busiest Core Utilization                   210,435          21.16%
===============================================================================
```

Step 11. Increase the rate of IGMP traffic from the tester to 5000 pps. See the CPU utilization increase further.

```
*A:PE-1# show system cpu
===============================================================================
CPU Utilization (Sample period: 1 second)
===============================================================================
Name                                    CPU Time       CPU Usage        Capacity
                                          (uSec)                           Usage
-------------------------------------------------------------------------------
BFD                                             0           0.00%           0.00%
…[snip]…
IGMP/MLD                                  417,124           4.65%          41.78%
IMSI Db Appl                                   82          ~0.00%          ~0.00%
IOM                                       133,029           1.48%           2.92%
```

```
IP Stack                                935,491          10.43%          93.45%
IS-IS                                     1,343           0.01%           0.06%
ISA                                      12,350           0.13%           0.45%
LDP                                         394          ~0.00%           0.03%
…[snip]…
WEB Redirect                                116          ~0.00%           0.01%
-------------------------------------------------------------------------------
Total                                 8,966,128         100.00%
   Idle                               6,972,962          77.77%
   Usage                              1,993,166          22.22%
Busiest Core Utilization                484,748          48.65%
===============================================================================
```

Step 12. Reinstall the DCP policy to the interface and see the CPU utilization drop.

```
*A:PE-1# configure router interface "int-pe1-to-tester"
*A:PE-1>config>router>if# dist-cpu-protection "dcp-static-policy-1"
*A:PE-1>config>router>if# exit all
*A:PE-1# show system cpu
===============================================================================
CPU Utilization (Sample period: 1 second)
===============================================================================
Name                                    CPU Time       CPU Usage        Capacity
                                          (uSec)                          Usage
-------------------------------------------------------------------------------
BFD                                            0           0.00%           0.00%
…[snip]…
IGMP/MLD                                   2,058           0.02%           0.10%
IMSI Db Appl                                  48          ~0.00%          ~0.00%
IOM                                      135,148           1.50%           3.04%
IP Stack                                   7,851           0.08%           0.47%
IS-IS                                      1,398           0.01%           0.07%
ISA                                       11,730           0.13%           0.43%
LDP                                          299          ~0.00%           0.02%
…[snip]…
WEB Redirect                                  71          ~0.00%          ~0.00%
-------------------------------------------------------------------------------
Total                                  8,975,262         100.00%
   Idle                                8,611,593          95.94%
   Usage                                 363,669           4.05%
Busiest Core Utilization                 136,669          13.70%
===============================================================================
```

Step 13. Stop the tester from sending packets, wait a few minutes and then note the status of the system.

There are no longer any violations of any enforcement policers on any interfaces on card 1.

```
*A:PE-1# tools dump security dist-cpu-protection violators enforcement interface card
1
===============================================================================
Distributed Cpu Protection Current Interface Enforcer Policer Violators
===============================================================================
Interface                       Policer/Protocol                    Hld Rem
```

```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Violators on Slot-1 Fp-1
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
[S]-Static [D]-Dynamic [M]-Monitor
-------------------------------------------------------------------------------
===============================================================================
```

The IGMP policer is indicated as conformant in the log events.

```
*A:PE-1# show log log-id 15
===============================================================================
Event Log 15
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=1024   next event=7  (not wrapped)]

…[snip]…

12 2013/04/18 17:42:12.43 EDT WARNING: SECURITY #2072 Base DCPUPROT
"Network_if "int-pe1-to-tester" on fp 1/1 newly conformant at 04/18/2013 17:41:57:27.
Policy "dcp-static-policy-1". Policer="sp-igmp"(static). Excd count=316418"

…[snip]…
```

The interface DCP details show all policers as conformant.

```
*A:PE-1# show router interface "int-pe1-to-tester" dist-cpu-protection
===============================================================================
Interface "int-pe1-to-tester" (Router: Base)
===============================================================================
Distributed CPU Protection Policy :  dcp-static-policy-1
-------------------------------------------------------------------------------
Statistics/Policer-State Information
===============================================================================
-------------------------------------------------------------------------------
Static Policer
-------------------------------------------------------------------------------
Policer-Name        : sp-arp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : arp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none

Policer-Name        : sp-icmp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : icmp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none

Policer-Name        : sp-igmp
Card/FP             : 1/1               Policer-State       : Conform
Protocols Mapped    : igmp
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none
-------------------------------------------------------------------------------
```

```
…[snip]…
```

An optional hold-down can be used in the configuration of the exceed-action of the policers in order to apply the exceed-action for a defined period (even if the policer goes conformant again during that period). The hold-down could be used, for example, to discard all packets associated with a policer for one hour after a violation is detected. An “indefinite” period is also supported which enforces discard or marking until the operator clears the policer with the **tools perform security dist-cpu-protection release-hold-down** command.

Step 14. The next scenario explored in this example is the use of DCP dynamic enforcement.

In order to use dynamic enforcement policers, a number of dynamic policers must be allocated to the DCP pool for the particular card being used.

```
*A:PE-1# configure card 1 fp dist-cpu-protection
*A:PE-1>config>card>fp>d-cpu-prot# info
----------------------------------------------
                dynamic-enforcement-policer-pool 1000
----------------------------------------------
```

The number allocated should be greater than the maximum number of dynamic policers expected to be in use on the card at one time. A conservative (large) number could be selected at first, and then the following show command can give data to help tune the pool to a smaller size over time:

```
*A:PE-1# show card 1 fp 1 dist-cpu-protection
===============================================================================
Card : 1 Forwarding Plane(FP) : 1
===============================================================================
Dynamic Enforcement Policer Pool : 1000
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Statistics Information
-------------------------------------------------------------------------------
Dynamic-Policers Currently In Use      : 0
Hi-WaterMark Hit Count                 : 0
Hi-WaterMark Hit Time                  : 04/20/2013 08:16:24 UTC
Dynamic-Policers Allocation Fail Count : 0
-------------------------------------------------------------------------------
===============================================================================
```

If the dynamic-enforcement-policer-pool is too small then when a local-monitoring-policer detects violating traffic, the dynamic enforcement policers will not be able to be instantiated. A log event will warn the operator when the pool is nearly exhausted.

A sample dynamic enforcement policy is created as follows:

```
*A:PE-1# configure system security dist-cpu-protection
*A:PE-1>config>sys>security>dist-cpu-protection# policy "dcp-dynamic-policy-1" create
                    description "Dynamic policing policy"
                    local-monitoring-policer "local-mon" create
                        description "Monitor for arp, icmp, igmp
                                     and all-unspecified"
                        rate packets 100 within 10
                    exit
                    protocol arp create
                        enforcement dynamic "local-mon"
                        dynamic-parameters
                            rate packets 20 within 10
                            exceed-action discard
                        exit
                    exit
                    protocol icmp create
                        enforcement dynamic "local-mon"
                        dynamic-parameters
                            rate packets 20 within 10
                            exceed-action discard
                        exit
                    exit
                    protocol igmp create
                        enforcement dynamic "local-mon"
                        dynamic-parameters
                            rate packets 20 within 10
                            exceed-action discard
                        exit
                    exit
                    protocol all-unspecified create
                        enforcement dynamic "local-mon"
                        dynamic-parameters
                            rate packets 100 within 10
                            exceed-action discard
                        exit
                    exit
```

For the *dcp-dynamic-policy-1* policy configuration:

- The policy contains no static policers. Per-protocol enforcement policers will be instantiated dynamically but only if triggered by a violation of the local-monitoring-policer.
- A local-monitoring-policer is configured for the policy. The configured rate determines the rate of arriving protocol packets at which the policy will trigger the automatic instantiation of dynamic per-protocol policers for the interface.
- Four protocols are configured and they are all associated with the local-monitoring-policer. The all-unspecified protocol will include all other extracted control packets on the interface.

– Each protocol has its own configured dynamic rates that will be used by the dynamic enforcement policers if they are instantiated. Note these rates are lower than previous scenarios (the **within** parameter is 10 seconds instead of 1 second).

– When this DCP policy is associated with an interface, only a single policer (the local-monitoring-policer) will be instantiated (statically/ permanently). The per-protocol dynamic policers are only instantiated when there is a violation of the local-monitoring-policer.

The policy is then associated with the interface:

```
*A:PE-1# configure router interface "int-pe1-to-tester"
*A:PE-1>config>router>if# dist-cpu-protection "dcp-dynamic-policy-1"
```

Step 15. Configure the tester to send:

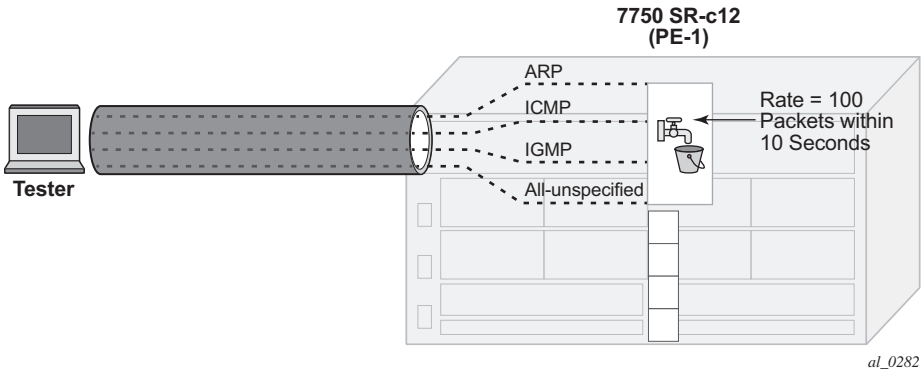
– 1pps of ARP

– 4pps of ICMP

– 1000pps of IGMP

Start the tester.

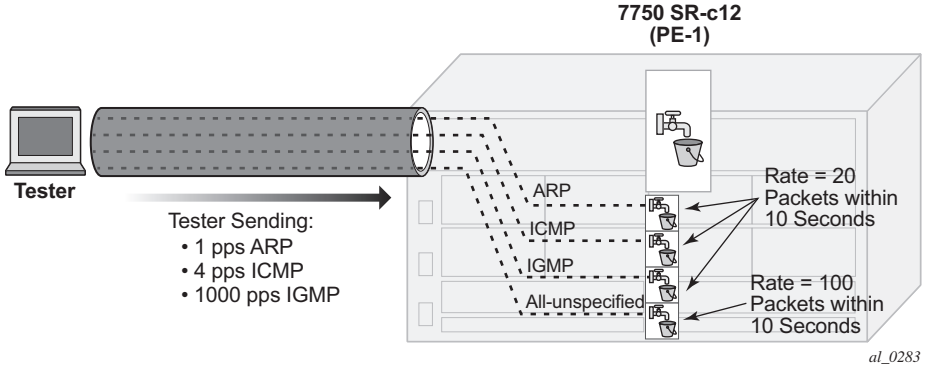
Figure 18 **Dynamic Policing – Local Monitor**



In Figure 18, the dynamic policers have not been instantiated yet.

Step 16. The local-monitoring-policer will become non-conforming since the aggregate arrival rate of arp+icmp+igmp+all-unspecified packets is greater than the configured local-monitoring-policer rate of 100 packets within 10 seconds. Dynamic enforcement policers will then be instantiated.

Figure 19 **Dynamic Policers Instantiated**



The ICMP and IGMP dynamic policers will see violations since their dynamic rates are being exceeded.

```
*A:PE-1# tools dump security dist-cpu-protection violators enforcement interface card
1
===============================================================================
Distributed Cpu Protection Current Interface Enforcer Policer Violators
===============================================================================
Interface                       Policer/Protocol                    Hld Rem
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Violators on Slot-1 Fp-1
-------------------------------------------------------------------------------
int-pe1-to-tester              icmp                             [D] none
int-pe1-to-tester              igmp                             [D] none
-------------------------------------------------------------------------------
[S]-Static [D]-Dynamic [M]-Monitor
-------------------------------------------------------------------------------
===============================================================================
```

The arp and all-unspecified dynamic policers were instantiated but will be counting down their detection time (if this show command is issued within 30 seconds of the attack starting).

```
*A:PE-1# show router interface "int-pe1-to-tester" dist-cpu-protection
===============================================================================
Interface "int-pe1-to-tester" (Router: Base)
===============================================================================
Distributed CPU Protection Policy :  dcp-dynamic-policy-1
-------------------------------------------------------------------------------
Statistics/Policer-State Information
===============================================================================
-------------------------------------------------------------------------------
Static Policer
-------------------------------------------------------------------------------
No entries found
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Local-Monitoring Policer
-------------------------------------------------------------------------------
```

```
Policer-Name        : local-mon
Card/FP             : 1/1               Policer-State       : Exceed
Protocols Mapped    : arp, icmp, igmp, all-unspecified
Exceed-Count        : 1097
All Dyn-Plcr Alloc. : True
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Dynamic-Policer (Protocol)
-------------------------------------------------------------------------------
Protocol(Dyn-Plcr)  : arp
Card/FP             : 1/1               Protocol-State      : Conform
Exceed-Count        : 0
Detec. Time Remain  : 5 seconds         Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True

Protocol(Dyn-Plcr)  : icmp
Card/FP             : 1/1               Protocol-State      : Exceed
Exceed-Count        : 31
Detec. Time Remain  : 28 seconds        Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True

Protocol(Dyn-Plcr)  : igmp
Card/FP             : 1/1               Protocol-State      : Exceed
Exceed-Count        : 23867
Detec. Time Remain  : 29 seconds        Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True

Protocol(Dyn-Plcr)  : all-unspecified
Card/FP             : 1/1               Protocol-State      : Conform
Exceed-Count        : 0
Detec. Time Remain  : 5 seconds         Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True
-------------------------------------------------------------------------------
===============================================================================
```

After 30 seconds have passed, the “Detec. Time Remain” for arp and all-unspecified will simply read 0 (zero).

After a few minutes the log events will be collected indicating a non-conformance was seen.

```
*A:PE-1# show log log-id 15
===============================================================================
Event Log 15
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=1024   next event=3  (not wrapped)]

2 2013/04/20 08:56:59.37 EDT WARNING: SECURITY #2067 Base DCPUPROT
"Non conformant network_if "int-pe1-to-tester" on fp 1/1 detected at 04/20/2013
08:52:28. Policy "dcp-dynamic-policy-1". Policer="icmp"(dynamic). Excd count=2"

1 2013/04/20 08:56:59.37 EDT WARNING: SECURITY #2067 Base DCPUPROT
"Non conformant network_if "int-pe1-to-tester" on fp 1/1 detected at 04/20/2013
08:52:22. Policy "dcp-dynamic-policy-1". Policer="igmp"(dynamic). Excd count=27"
```

Step 17. Stop the tester.

The dynamic policer detection timers will start counting down since they are no longer seeing violating packets.

```
*A:PE-1# show router interface "int-pe1-to-tester" dist-cpu-protection
===============================================================================
Interface "int-pe1-to-tester" (Router: Base)
===============================================================================
Distributed CPU Protection Policy :  dcp-dynamic-policy-1
-------------------------------------------------------------------------------
Statistics/Policer-State Information
===============================================================================
-------------------------------------------------------------------------------
Static Policer
-------------------------------------------------------------------------------
No entries found
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Local-Monitoring Policer
-------------------------------------------------------------------------------
Policer-Name        : local-mon
Card/FP             : 1/1               Policer-State       : Exceed
Protocols Mapped    : arp, icmp, igmp, all-unspecified
Exceed-Count        : 1097
All Dyn-Plcr Alloc. : True
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Dynamic-Policer (Protocol)
-------------------------------------------------------------------------------
Protocol(Dyn-Plcr)  : arp
Card/FP             : 1/1               Protocol-State      : Conform
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True

Protocol(Dyn-Plcr)  : icmp
Card/FP             : 1/1               Protocol-State      : Exceed
Exceed-Count        : 511
Detec. Time Remain  : 14 seconds        Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True

Protocol(Dyn-Plcr)  : igmp
Card/FP             : 1/1               Protocol-State      : Exceed
Exceed-Count        : 345550
Detec. Time Remain  : 18 seconds        Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True

Protocol(Dyn-Plcr)  : all-unspecified
Card/FP             : 1/1               Protocol-State      : Conform
Exceed-Count        : 0
Detec. Time Remain  : 0 seconds         Hold-Down Remain.   : none
Dyn-Policer Alloc.  : True
-------------------------------------------------------------------------------
===============================================================================
```

After 30 seconds there are no more violators.

```
*A:PE-1# tools dump security dist-cpu-protection violators enforcement interface card
1
===============================================================================
Distributed Cpu Protection Current Interface Enforcer Policer Violators
===============================================================================
Interface                       Policer/Protocol                    Hld Rem
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Violators on Slot-1 Fp-1
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
[S]-Static [D]-Dynamic [M]-Monitor
-------------------------------------------------------------------------------
===============================================================================
```

The dynamic policer pool Hi-WaterMark for card 1 fp 1 shows 4 since the highest number of dynamic policers allocated at any one time on the card/fp was 4.

```
*A:PE-1# show card 1 fp 1 dist-cpu-protection
===============================================================================
Card : 1 Forwarding Plane(FP) : 1
===============================================================================
Dynamic Enforcement Policer Pool : 1000
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Statistics Information
-------------------------------------------------------------------------------
Dynamic-Policers Currently In Use      : 0
Hi-WaterMark Hit Count                 : 4
Hi-WaterMark Hit Time                  : 04/20/2013 08:52:22 UTC
Dynamic-Policers Allocation Fail Count : 0
-------------------------------------------------------------------------------
===============================================================================
```

A few minutes later the log events indicate that the flood has ended.

```
*A:PE-1# show log log-id 15
===============================================================================
Event Log 15
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=1024   next event=5  (not wrapped)]

4 2013/04/20 09:01:59.39 EDT WARNING: SECURITY #2073 Base DCPUPROT
"Network_if "int-pe1-to-tester" on fp 1/1 newly conformant at 04/20/2013 08:58:39.
Policy "dcp-dynamic-policy-1". Policer="igmp"(dynamic). Excd count=345550"

3 2013/04/20 09:01:59.39 EDT WARNING: SECURITY #2073 Base DCPUPROT
"Network_if "int-pe1-to-tester" on fp 1/1 newly conformant at 04/20/2013 08:58:35.
Policy "dcp-dynamic-policy-1". Policer="icmp"(dynamic). Excd count=511"
```

Conclusion

Distributed CPU Protection (DCP) offers a powerful rate limiting function for control protocol traffic that is extracted from the data path and sent to the CPM.

This example has demonstrated how to configure DCP on an interface and what indications SR OS provides to the operator during a potential attack or misconfiguration.

DCP can also be deployed in scenarios where per-SAP-per-protocol rate limiting is useful, such as for subscriber management in a subscriber per-vlan scenario. A DCP policy can be assigned to an MSAP policy on a Broadband Network Gateway, for example, to limit traffic related to certain protocols and to discard certain protocols. When deployed in a subscriber management scenario, DCP can help isolate SAPs (subscribers) from each other and even isolate protocols from each other within an individual SAP (subscriber). Many of the same concepts introduced in this example are applicable when DCP is deployed in a subscriber management application.

Event Handling System

This chapter provides information about Event Handling Systems (EHS).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 13.0.R3. The CLI in the current edition is based on SR OS release 14.0.R5.

SR OS release 13.0.R1 introduced Event Handling System (EHS).

SR OS release 14.0.R4 introduced EHS script enhanced capabilities, but the examples in this chapter do not include these enhancements, such as static variables, advanced syntax (shell scripting commands), and so on. A known issue in SR OS 14.0.R4 is that the results file is empty for a successful script run.

Overview

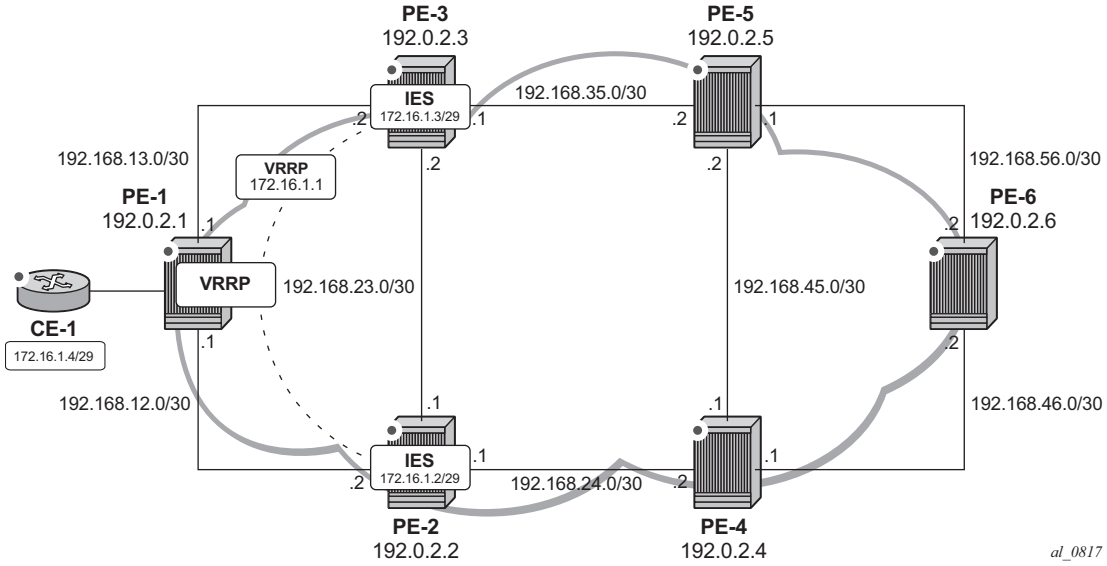
The Event Handling System (EHS) is a tool available in SR OS that allows operators to configure user-defined actions on the router in response to an event. The event is referred to as the trigger, and can be all or part of any event message generated by the event-control framework. The user-defined action is controlled by the script-control function. This script-control function references one or more scripts that are able to execute any command available in CLI when the trigger event occurs.

This feature allows for customized automated event management based on specific operator requirements.

Configuration

The topology shown in Figure 20 provides an example of an EHS configuration. All routers within the example topology participate in the same IS-IS Level-2 area and run LDP. All routers are BGP speakers and form part of Autonomous System 64496, exchanging routes for IPv4 address family only.

Figure 20 **Example Topology**



PE-1 has a connected CE router, CE-1, indexed into a VPLS service. The VPLS has spoke-SDPs to an IES instance on both PE-2 and PE-3, which provide a redundant default gateway to CE-1 using the Virtual Router Redundancy Protocol (VRRP). The subnet used for this redundant gateway connectivity between PE-2 and PE-3 is 172.16.1.0/29. The configuration at PE-3 is shown in the following output. The configuration at PE-2 is similar; the exception being IP addressing and VRRP priority, which is 254.

```
configure
    service
        ies 1 customer 1 create
            interface "redundant-interface" create
                address 172.16.1.3/29
                ip-mtu 1500
                vrrp 1
                    backup 172.16.1.1
                    priority 253
                    ping-reply
                exit
                spoke-sdp 31:1 create
                    no shutdown
                exit
```

```
            exit
            no shutdown
        exit
```

The objective of this configuration example is to ensure that both upstream and downstream traffic are always routed through the same PE router. That is, if PE-3 is VRRP Master, it will attract upstream traffic from CE-1 using the VRRP virtual IP/MAC, but PE-3 should also be the transit PE for downstream traffic destined toward CE-1. When both upstream and downstream traffic transit through the same PE router, it can simplify troubleshooting, QoS configuration, and reconciliation of ingress/egress statistics.

In normal operation, PE-2 is the VRRP master and advertises the BGP prefix 172.16.1.0/29 with a local preference of 100 (default value). Similarly, PE-3 is the VRRP backup and advertises the BGP prefix 172.16.1.0/29 with a local preference of 50, using a BGP export policy (route-policy). Therefore, upstream and downstream traffic normally transit through PE-2.

```
*A:PE-3# show router vrrp instance
===============================================================================
VRRP Instances
===============================================================================
Interface Name                   VR Id Own Adm  State       Base Pri   Msg Int
                                 IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
redundant-interface              1     No  Up   Backup       253       1
                                 IPv4      Up   n/a         253       No
  Backup Addr: 172.16.1.1
-------------------------------------------------------------------------------
Instances : 1
===============================================================================


*A:PE-3# show router bgp routes 172.16.1.0/29 hunt | match expression
"Network|Nexthop|To|Local Pref"
Network        : 172.16.1.0/29
Nexthop        : 192.0.2.3
To             : 192.0.2.6
Res. Nexthop   : n/a
Local Pref.    : 50                      Interface Name : NotAvailable
```

When PE-3 transitions from backup to master, it must modify its local preference attribute for prefix 172.16.1.0/29 to a value of 150 to attract downstream traffic destined toward CE-1. Similarly, when PE-3 reverts to backup, it must advertise the prefix with a local preference of 50.

Script Control

The first step in configuring event handling is to configure a script containing the CLI commands to be executed when the event is triggered. This script can be stored locally on the compact flash, or it can be stored off-node at a defined remote URL, where it can be accessed using FTP or TFTP. When the script is stored locally on the compact flash and the router is equipped with redundant CPMs, the script must be manually saved on the same compact flash on both CPMs, because it is not synchronized automatically.

The first requirement is to modify the local preference of the prefix 172.16.1.0/29 to 150 on transition to VRRP master. The script, which in this example is held locally on CF1:/, therefore contains the following commands (where the policy-statement, redundant-interface, is the name of the export policy used to advertise the 172.16.1.0/29 prefix):

```
*A:PE-3# file type cf1:/vrrp-master.txt
File: vrrp-master.txt
-------------------------------------------------------------------------------
exit all
config router policy-options
begin
policy-statement redundant-interface
entry 10
action accept
local-preference 150
exit
exit
exit
commit
exit all
```

The script file does not support variables or conditions. Output modifiers such as “| match” and “>” redirect are not supported. If these modifiers are present, the script will continue to run but the pipe/match or redirect will not work. Similarly, there is no syntax checking when the script file is created; instead, the script will fail with a command error. Also, transactional CLI (candidate edit) cannot be used in the script, and will fail with a command error.

Within the **system>script-control** context, the script is assigned a name and reference is made to its location. It is then put in the no shutdown state. When the script has been defined, a **script-policy** is configured that calls the previously configured script. The script-policy also specifies a location and filename for a results file that records the successful or unsuccessful conclusion of each script run and each command executed during that run. However, in release 14.0.R5, the result file is empty for a successful policy run (known issue in Release Notes - internal reference 234641-MI). Each time the script is run, the results are recorded in a file

with the name specified for results, followed by an underscore and the date and time that the script was run. A results file must be specified in order for the script to successfully run. The results file can be on the local compact flash, or a remote URL can be specified. As with the script, the script-policy must also be put in the no shutdown state.

```
configure
    system
        script-control
            script "vrrp-master-script"
                location "cf1:/vrrp-master.txt"
                no shutdown
            exit
            script-policy "vrrp-master-policy"
                results "cf1:/script-results.txt"
                script "vrrp-master-script"
                max-completed 4
                expire-time 3600
                lifetime forever
                no shutdown
            exit
        exit
```

The optional lifetime command specifies the maximum time that the script may run. The max-completed command specifies the maximum number of script run history status entries to be retained. An optional expire-time command specifies the maximum time that the system keeps the run history status (default is 1 h). The system maintains the script run history table, which has a maximum size of 255 entries. Entries are removed from this table when the max-completed or expire-time thresholds are crossed. If the table reaches the maximum value, subsequent script launch requests are not run until older run history entries expire (due to expire-time), or entries are manually cleared. To manually clear entries, the following command is used:

```
clear system script-control script-policy completed <script-policy-name>
```

The script run history status information can be viewed using the following command:

```
*A:PE-3# show system script-control script-policy "vrrp-master-policy"

===============================================================================
Script-policy Information
===============================================================================
Script-policy               : vrrp-master-policy
Script-policy Owner         : TiMOS CLI
Administrative status       : enabled
Operational status          : enabled
Script                      : vrrp-master-script
Script owner                : TiMOS CLI
Script source location      : cf1:/vrrp-master.txt
Script results location     : cf1:/script-results.txt
Max running allowed         : 1
Max completed run histories  : 4
```

```
Max lifetime allowed         : 248d 13:13:56 (21474836 seconds)
Completed run histories      : 1
Executing run histories      : 0
Initializing run histories   : 0
Max time run history saved   : 0d 01:00:00 (3600 seconds)
Script start error           : N/A
Last change                  : 2016/10/27 16:35:07
Max row expire time          : never
Last application             : event-script
Last auth. user account      : not-specified

===============================================================================
Script Run History Status Information
-------------------------------------------------------------------------------
Script Run #4
-------------------------------------------------------------------------------
Start time    : 2016/10/27 16:36:09     End time      : 2016/10/27 16:36:19
Elapsed time  : 0d 00:00:10             Lifetime      : 0d 00:00:00
State         : terminated              Run exit code : noError
Result time   : 2016/10/27 16:36:19     Keep history  : 0d 00:59:24
Error time    : never
Results file  : cf1:/script-results.txt_20161027-153607.281824.out
Run exit      : Success
Error         : N/A
Application   : event-script            Auth. user ac*: not-specified
* indicates that the corresponding row element may have been truncated.
===============================================================================
```

Event Handler

The second step in configuring event handling is to assign actions to be performed as a result of the event-trigger. These actions are typically one or more configured scripts defined as entries in an action-list. In the following output, the event-handler is assigned the name event-handler-1, and the action-list consists of a single entry. This entry calls the previously configured script-policy vrrp-master-policy (which in turn references the previously defined script vrrp-master-script). If multiple actions are required based on a single event-trigger, they can be configured in the action-list with subsequent entries, which are run in sequence (up to 1500 action-list entries are supported).

For this example, only a single entry is required; therefore, there is a one to one relationship between the event-handler and the action-list entry. Both the entry within the action-list and the handler should be put in the no shutdown state.

```
configure
    log
        event-handling
            handler "event-handler-1"
                action-list
                    entry 10
                        script-policy "vrrp-master-policy"
```

```
                    no shutdown
                exit
            exit
            no shutdown
        exit
    exit
```

Event Trigger

The final step in configuring event handling is to configure the event-trigger. The event-trigger defines the event that triggers the running of the script. The event-trigger is based on any event generated by the event-control framework, and can match against the application and event number (event_id). Log filters can also be used to match against specific events using the subject and/or message fields. Regular expressions can be used where required. EHS will not use any message that is suppressed through event-control configuration, or any event message that is throttled.

The general format for an event in an event log is as follows:

```
nnnn YYYY/MM/DD HH:MM:SS.SS Zone <severity>:<application> # <event_id> <router-name>
<subject> description
```

Where:

```
nnnnThe log entry sequence number
YYYY/MM/DDThe UTC date stamp for the log entry:
YYYY - Year
MM - Month
DD - Date
HH:MM:SS.SSThe UTC time stamp for the event
HH - Hours (24 hour format)
MM - Minutes
SS.SS - Seconds
ZoneTimezone
<severity>The severity level name of the event
<application>The application generating the log message
<event_id>The application’s event ID number for the event
<subject>The subject/affected object for the event
<message>A textual description of the event
```

In the example, the following event message is generated when PE-3 becomes VRRP Master:

```
2 2016/10/27 08:23:37.54 UTC MINOR: VRRP #2001 Base Becoming Master
"VRRP virtual router instance 1 on interface redundant-interface (primary address
172.16.1.3) changed state to master"
```

Therefore, the event-trigger configuration is based on an application of VRRP and an event number of 2001 (vrrptrapNewMaster). In the following output, vrrp 2001 is configured as the event. The trigger-entry is defined as 1, and in this example, there is only one trigger event. Up to 1500 trigger-entries can be included, each of which can act as a potential trigger event. The trigger-entry also references the previously configured event-handler-1. (Recall that the event-handler references the script-control, which in turn references the script that should be run.)

Finally, there is a reference to log-filter 1. Without more explicit filtering, event handling will be triggered on any event with the application of VRRP and event number 2001. There may be multiple VRRP instances running on this router, but the requirement is that event handling should only be triggered when the VRRP instance running on redundant-interface transitions to master at PE-3. Therefore, log-filter 1 is used to define a more explicit match using the message field, which contains an explicit reference to the interface. Both the trigger-entry and the event-handler should be put in the no shutdown state.

```
configure
    log
        filter 1
            default-action drop
            entry 10
                action forward
                match
                    message eq pattern "interface redundant-interface (primary
                                address 172.16.1.3) changed state to master"
                exit
            exit
        exit


configure
    log
        event-trigger
            event "vrrp" 2001
                trigger-entry 1
                    event-handler "event-handler-1"
                    log-filter 1
                    no shutdown
                exit
                no shutdown
            exit
        exit
    exit
```

The configuration of the example event handling for the failure event (PE-3 transitions to VRRP master) is now complete. By shutting down the spoke-SDP between PE-1 and PE-2, it is possible to simulate a failure event where the VRRP message path is broken. Therefore, four events are generated.

- The first indicates that PE-3 has become VRRP master for the interface named redundant-interface.

- The second indicates that EHS handler event-handler-1 was invoked by a CLI user.
- The third indicates that a script file has initiated an attempt to execute CLI commands contained in script file vrrp-master.txt.
- The fourth indicates that the attempt to execute those CLI commands was successful (in release 14.0.R5, result “none” is displayed for a successful script run; in release 13.0.R3, the result is “success”).

```
15 2016/10/27 09:13:59.03 UTC MINOR: VRRP #2001 Base Becoming Master
"VRRP virtual router instance 1 on interface redundant-interface (primary address
172.16.1.3) changed state to master"

16 2016/10/27 09:13:59.03 UTC MINOR: SYSTEM #2069 Base EHS script
"Ehs handler :"event-handler-1" with the description : "" was invoked by the cli-
user account "not-specified"."

17 2016/10/27 09:14:01.64 UTC MAJOR: SYSTEM #2052 Base CLI 'exec'
"A CLI user has initiated an 'exec' operation to process the commands in the SROS CLI
file cf1:\vrrp-master.txt"

18 2016/10/27 09:14:10.73 UTC MAJOR: SYSTEM #2053 Base CLI 'exec'
"The CLI user initiated 'exec' operation to process the commands in the SROS CLI
 file cf1:\vrrp-master.txt has completed with the result of none"
```

The results file is empty for a successful script run in release 14.0.R5 (known issue). However, in release 13.0.R3, the last message would indicate that the result is “success” and the result file would contain all the commands contained in the script. In release 13.0.R3, an example of the results file configured in the script-policy is as follows. A successful script run shows the commands contained in the script, followed by an indication that the commands were executed.

```
*A:PE-3# file type script-results.txt_20150608-160703.402470.out
File: script-results.txt_20150608-160703.402470.out
-------------------------------------------------------------------------------
*A:PE-3# exit all
*A:PE-3# config router policy-options
*A:PE-3# begin
*A:PE-3# policy-statement redundant-interface
*A:PE-3# entry 10
*A:PE-3# action accept
*A:PE-3# local-preference 150
*A:PE-3# exit
*A:PE-3# exit
*A:PE-3# exit
*A:PE-3# commit
*A:PE-3# exit all
Executed 12 lines in 8.2 seconds from file cf1:\vrrp-master.txt
```

The following outputs confirm that PE-3 is VRRP master, and that the local preference attribute for prefix 172.16.1.0/29 has changed to a value of 150. The result of this action is that PE-3 will now be the transit router for both upstream and downstream traffic.

```
*A:PE-3# show router vrrp instance
===============================================================================
VRRP Instances
===============================================================================
Interface Name                  VR Id Own Adm  State       Base Pri   Msg Int
                                IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
redundant-interface             1     No  Up   Master       253       1
                                IPv4      Up   n/a          253       No
  Backup Addr: 172.16.1.1
-------------------------------------------------------------------------------
Instances : 1
===============================================================================


*A:PE-3# show router bgp routes 172.16.1.0/29 hunt | match expression
"Network|Nexthop|To|Local Pref"
Network        : 172.16.1.0/29
Nexthop        : 192.0.2.3
To             : 192.0.2.6
Res. Nexthop   : n/a
Local Pref.    : 150                    Interface Name : NotAvailable
```

The event-handler indicates that the referenced script was triggered and run using the command shown in the following output. The Action-List Entry Execution Statistics window provides statistics on the number of times an action (script) was queued to run, and the number of times an error was experienced, both during launch and due to a non-operational admin status. The remainder of the fields in the output are self-explanatory.

```
*A:PE-3# show log event-handling handler "event-handler-1"
===============================================================================
Event Handling System - Handlers
===============================================================================
===============================================================================
Handler          : event-handler-1
===============================================================================
Description      : (Not Specified)
Admin State      : up                                 Oper State : up

-------------------------------------------------------------------------------
Handler Execution Statistics
  Success        : 11
  Err No Entry   : 0
  Err Adm Status : 0
Total            : 11

-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Handler Action-List Entry
-------------------------------------------------------------------------------
Entry-id         : 10
Description      : (Not Specified)
Admin State      : up                                 Oper State : up
Script
  Policy Name    : vrrp-master-policy
  Policy Owner   : TiMOS CLI
```

```
Min Delay        : 0
Last Exec        : 10/27/2016 15:41:52
-------------------------------------------------------------------------------
Handler Action-List Entry Execution Statistics
  Enqueued       : 11
  Err Launch     : 0
  Err Adm Status : 0
Total            : 11
===============================================================================
```

The example includes an event-trigger and script to meet the requirements of a fail-forward where PE-3 becomes VRRP master. Now, configuration is needed for when PE-3 reverts to VRRP backup. Without another event-trigger and script, PE-3 will continue to advertise the prefix 172.16.1.0/29 with a local preference of 150 and upstream/downstream traffic will be asymmetric through PE-1/PE-3 respectively.

As before, a script is required. Because PE-2 advertises the prefix with a local preference of 100 (default), PE-3 needs to advertise the same prefix with a lower value (50 in the following output), so that PE-2 is the preferred next hop.

```
*A:PE-3# file type cf1:/vrrp-backup.txt
File: vrrp-backup.txt
-------------------------------------------------------------------------------
exit all
config router policy-options
begin
policy-statement redundant-interface
entry 10
action accept
local-preference 50
exit
exit
exit
commit
exit all
```

The script must then be configured within the script-control context, and subsequently referenced in a script-policy as vrrp-backup-policy.

```
configure
    system
        script-control
            script "vrrp-backup-script"
                location "cf1:/vrrp-backup.txt"
                no shutdown
            exit
            script-policy "vrrp-backup-policy"
                results "cf1:/script-revert-results.txt"
                script "vrrp-backup-script"
                max-completed 4
                lifetime forever
                no shutdown
            exit
        exit
```

The event-handler acts as the interface between the configured script-policy and event-trigger. Therefore, a second event-handler is configured with an action-list consisting of a single entry referencing the newly configured vrrp-backup-policy.

```
configure
    log
        event-handling
            handler "event-handler-2"
                action-list
                    entry 10
                        script-policy "vrrp-backup-policy"
                        no shutdown
                    exit
                exit
                no shutdown
            exit
        exit
```

Finally, the event-trigger is configured. To revert to VRRP Backup, the application is VRRP and the event number is 2006 (tmnxVrrpBecameBackup). The configuration is filtered on the message field, as before, using log-filter 2, so that it is specific to the interface named redundant-interface.

```
configure
    log
        filter 2
            default-action drop
            entry 10
                action forward
                match
                    message eq pattern "interface redundant-interface changed
                                        state to backup"
                exit
            exit
        exit

configure
    log
        event-trigger
            event "vrrp" 2006
                trigger-entry 1
                    event-handler "event-handler-2"
                    log-filter 2
                    no shutdown
                exit
                no shutdown
            exit
        exit
    exit
```

The configuration of the example event handling for the revertive failure event (PE-3 transitions to VRRP backup) is now complete. By re-enabling the spoke-SDP between PE-1 and PE-2, the VRRP message path is restored, and PE-2 again becomes the VRRP master. The following events are generated. The first indicates that PE-3 has become VRRP backup for the interface named redundant-interface. The second indicates that EHS handler event-handler-2 was invoked by a CLI user. The third indicates that a script file has initiated an attempt to execute CLI commands contained in script file vrrp-backup.txt. The fourth indicates that the attempt to execute those CLI commands was successful (in release 14.0.R5, result “none” is displayed for a successful script run; in release 13.0.R3, the result is “success”).

```
7 2016/10/27 16:17:53.61 UTC MINOR: VRRP #2006 Base Becoming Backup
"VRRP virtual router instance 1 on interface redundant-interface changed state to
backup - current master is 172.16.1.2"

8 2016/10/27 16:17:53.61 UTC MINOR: SYSTEM #2069 Base EHS script
"Ehs handler :"event-handler-2" with the description : "" was invoked by the cli-user
account "not-specified"."

9 2016/10/27 16:17:56.17 UTC MAJOR: SYSTEM #2052 Base CLI 'exec'
"A CLI user has initiated an 'exec' operation to process the commands in the SR OS CLI
file cf1:\vrrp-backup.txt"

10 2016/10/27 16:18:04.64 UTC MAJOR: SYSTEM #2053 Base CLI 'exec'
"The CLI user initiated 'exec' operation to process the commands in the SROS CLI
 file cf1:\vrrp-backup.txt has completed with the result of none"
```

The following outputs confirm that PE-3 is VRRP backup, and that the local preference attribute for prefix 172.16.1.0/29 has changed to a value of 50. The result of this action is that PE-2 will now be the transit router for both upstream and downstream traffic.

```
*A:PE-3# show router vrrp instance
===============================================================================
VRRP Instances
===============================================================================
Interface Name                   VR Id Own Adm  State       Base Pri   Msg Int
                                 IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
redundant-interface              1     No  Up   Backup       253       1
                                 IPv4      Up   n/a         253        No
  Backup Addr: 172.16.1.1
-------------------------------------------------------------------------------
Instances : 1
===============================================================================


*A:PE-3# show router bgp routes 172.16.1.0/29 hunt | match expression
"Network|Nexthop|To|Local Pref"
Network        : 172.16.1.0/29
Nexthop        : 192.0.2.3
To             : 192.0.2.6
Res. Nexthop   : n/a
Local Pref.    : 50                      Interface Name : NotAvailable
```

Conclusion

EHS allows operators to configure user-defined actions on the router when an event occurs. The event trigger can be anything that is generated by the event-control framework, and explicit filtering is possible using regular expressions. A user-defined action typically runs a script that allows any CLI commands to be executed. Multiple actions are permitted, running multiple scripts if required.

Interface Configuration

In This Section

This section provides interface configuration information for the following topics:

- Multi-Chassis APS and Pseudowire Redundancy Interworking
- Multi-Chassis LAG and Pseudowire Redundancy Interworking

Multi-Chassis APS and Pseudowire Redundancy Interworking

This chapter describes multi-chassis APS and pseudowire redundancy interworking.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 6.0.R2, but the CLI in the current edition is based on SR OS release 14.0.R5.The configuration in this chapter includes the use of the ATM ports. See the Release Notes for information about support of ATM (and other) MDAs on various platforms as well as MC-APS restrictions.

Overview

MC-APS

MC-APS is an extension to the APS feature to provide not only link redundancy but also node level redundancy. It can protect against nodal failure by configuring the working circuit of an APS group on one node while configuring the protect circuit of the same APS group on a different node.

The two nodes connect to each other with an IP link that is used to establish a signaling path between them. The relevant APS groups in both the working and protection routers must have the same group ID and working circuit, and the protect circuit must have compatible configurations (such as the same speed, framing, and port-type). Signaling is provided using the direct connection between the two service routers. A heartbeat protocol can be used to add robustness to the interaction between the two routers.

Signaling functionality includes support for:

- APS group matching between service routers.
- Verification that one side is configured as a working circuit and the other side is configured as the protect circuit. In case of a mismatch, a trap (incompatible-neighbor) is generated.
- Change in working circuit status is sent from the working router to keep the protection router in sync.
- Protection router, based on K1/K2 byte data, member circuit status, and external request, selects the active circuit and informs the working router to activate or de-activate the working circuit.

Pseudowire Redundancy

Pseudowire (PW) redundancy provides the ability to protect a pseudowire with a pre-provisioned pseudowire and to switch traffic over to the secondary standby pseudowire in case of a SAP and/or network failure condition. Normally, pseudowires are redundant by the virtue of the SDP redundancy mechanism. For instance, if the SDP is an RSVP LSP and is protected by a secondary standby path and/or by Fast-Reroute paths, the pseudowire is also protected.

However, there are a few applications in which SDP redundancy does not protect the end-to-end pseudowire path when there are two different destination 7x50 PE nodes for the same VLL service. The main use case is the provisioning of dual-homing of a CPE or access node to two 7x50 PE nodes located in different POPs. The other use case is the provisioning of a pair of active and standby BRAS nodes, or active and standby links to the same BRAS node, to provide service resiliency to broadband service subscribers.

Example Topology

The setup in this section contains two access nodes and 4 PE nodes. The access nodes can be any ATM switches that support 1+1 bi-directional APS. Figure 21 shows the physical topology of the setup. Figure 23 shows the use of both MC-APS in the access network and pseudowire redundancy in the core network to provide a resilient end-to-end VLL service.

Figure 21 **MC-APS Network Topology**

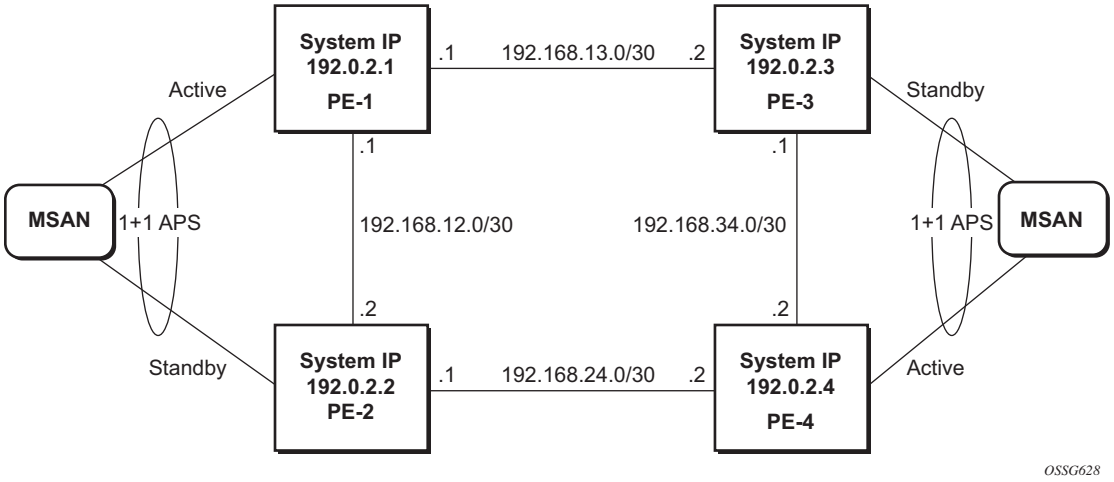


Figure 22 **Access Node and Network Resilience (Part 1)**

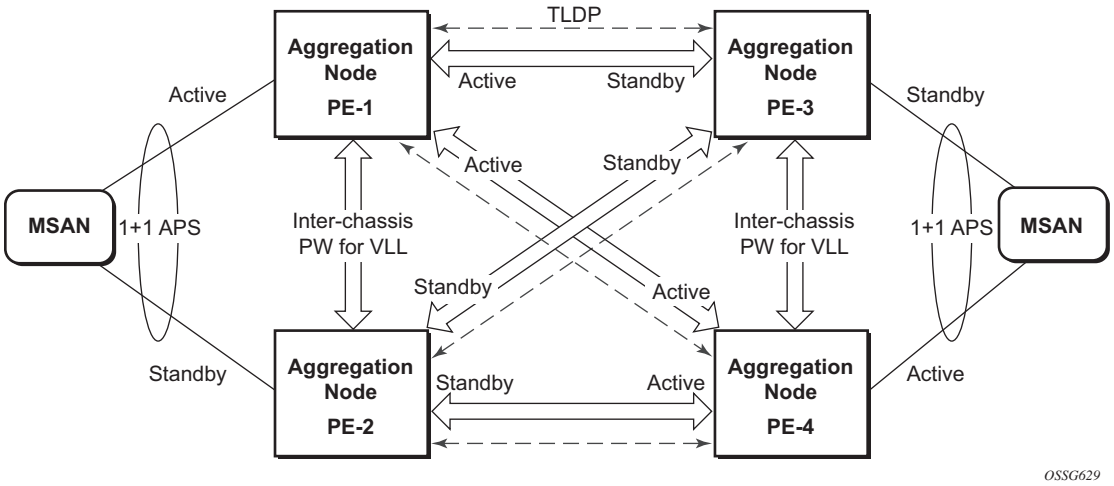
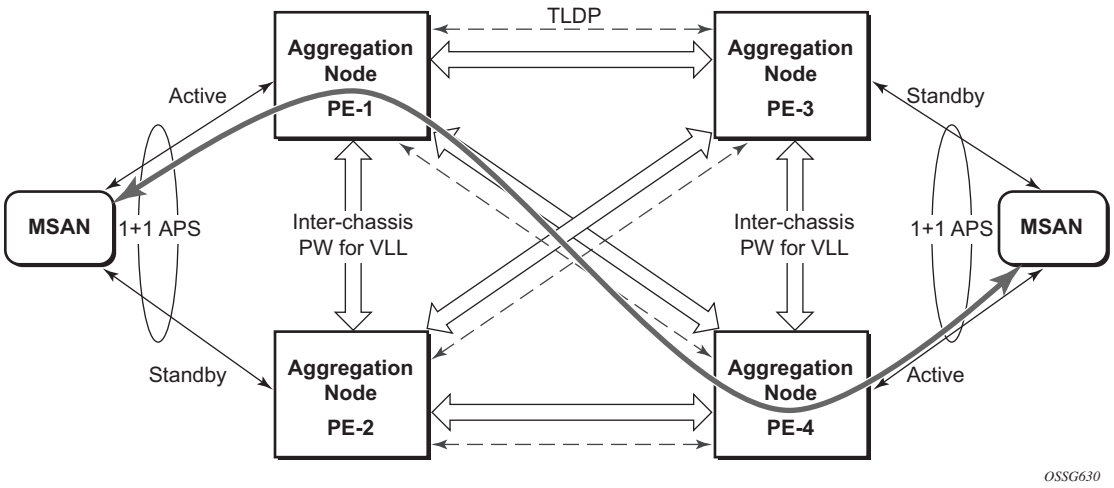


Figure 23 **Access Node and Network Resilience (Part 2)**



Configuration

The following configuration should be completed on the PEs before configuring MC-APS:

- Cards, MDAs and ports
- Interfaces
- IGP configured and converged
- MPLS
- SDPs configured between all PE routers

For the IGP, OSPF or IS-IS can be used. MPLS or GRE can be used for the transport tunnels. For MPLS, LDP or RSVP protocols can be used for signaling MPLS labels. In this example, OSPF and LDP are used. The following commands can be used to check if OSPF has converged and to make sure the SDPs are up (for example, on PE-1):

```
*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     00h02m12s  0
       system                                                       0
192.0.2.2/32                                    Remote  OSPF      00h01m17s  10
```

```
       192.168.12.2                                                    10
192.0.2.3/32                                  Remote  OSPF      00h01m05s  10
       192.168.13.2                                                    10
192.0.2.4/32                                  Remote  OSPF      00h01m08s  10
       192.168.12.2                                                    20
192.168.12.0/30                               Local   Local     00h02m13s  0
       int-PE-1-PE-2                                                   0
192.168.13.0/30                               Local   Local     00h02m12s  0
       int-PE-1-PE-3                                                   0
192.168.24.0/30                               Remote  OSPF      00h01m17s  10
       192.168.12.2                                                    20
192.168.34.0/30                               Remote  OSPF      00h01m05s  10
       192.168.13.2                                                    20
-------------------------------------------------------------------------------
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#

*A:PE-1# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
12     0       1556    192.0.2.2        Up   Up          MPLS    L     TLDP
13     0       1556    192.0.2.3        Up   Up          MPLS    L     TLDP
14     0       1556    192.0.2.4        Up   Up          MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
===============================================================================
*A:PE-1#
```

Step 1. APS configuration on MSANs

The access nodes can be any ATM switches that support 1+1 bi-directional APS. Here is an example on 7670RSP (Routing Switching Platform).

```
RSP> configure
RSP> port 1-6-1-1
RSP> options protection type 1+1
RSP> options protection switching bidirect
RSP> options protection
                                                                    (Standby)
#                  Type        Status           Name
1-6-1-1            STM1_IR8    OK
Protection Group Contains:
  Protection Port     : 1-6-1-1     (Standby)
  Working Port        : 1-5-1-1
Protection Type       : 1+1
Switching Type        : Non-Revertive
Switching Mode        : Bi-directional
```

```
Wait-To-Restore Timer :  5 minute(s)
```

Step 2. MC-APS configuration on PE-1 and PE-2

Assuming the link between MSAN and PE-1 is working circuit and the link between MSAN and PE-2 is protection circuit.

Configure APS on the PE-1 port. Specify the system IP address of neighbor node (PE-2) and working-circuit.

```
*A:PE-1# configure
    port 1/2/1
        sonet-sdh
        exit
        no shutdown


*A:PE-1# configure
    port aps-1
        aps
            neighbor 192.0.2.2
            working-circuit 1/2/1
        exit
        sonet-sdh
            path
                atm
                exit
                no shutdown
            exit
        exit
        no shutdown
```

Configure APS on the PE-2 port. Specify the system IP address of neighbor node (PE-1) and protect-circuit instead of working-circuit.

```
*A:PE-2# configure
    port 1/2/1
        sonet-sdh
        exit
        no shutdown


*A:PE-2# configure port aps-1
        aps
            neighbor 192.0.2.1
            protect-circuit 1/2/1
        exit
        sonet-sdh
            path
                atm
                exit
                no shutdown
            exit
        exit
        no shutdown
```

The following parameters can be configured under APS optionally.

- advertise-interval — This command specifies the time interval, in 100s of milliseconds, between 'I am operational' messages sent by both protect and working circuits to their neighbor for multi-chassis APS.
- hold-time — This command specifies how much time can pass, in 100s of milliseconds, without receiving an advertise packet from the neighbor before the multi-chassis signaling link is considered not operational.
- revert-time — This command configures the revert-time timer to determine how long to wait before switching back to the working circuit after that circuit has been restored into service.
- switching-mode — This command configures the switching mode for the APS port which can be bi-directional or uni-directional.

Step 3. Verify the APS status on PE-1.

```
*A:PE-1# show port aps-1

===============================================================================
SONET/SDH Interface
===============================================================================
Description        : APS Group
Interface          : aps-1                   Speed                : oc3
Admin Status       : up                      Oper Status          : up
Physical Link      : Yes                     Loopback Mode        : none
Single Fiber Mode  : No
Clock Source       : node                    Framing              : sonet
Last State Change  : 10/25/2016 13:45:58     Port IfIndex         : 1358987264
Configured Address : 02:15:ff:00:02:49
Hardware Address   : 02:15:ff:00:02:49
Last Cleared Time  : N/A
J0 String          : 0x01                    Section Trace Mode   : byte
Rx S1 Byte         : 0x00 (stu)              Rx K1/K2 Byte        : 0x00/0x00
Tx S1 Byte         : 0x0a (st3)              Tx DUS/DNU           : Disabled
Rx J0 String (Hex) : 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
Cfg Alarm          : loc lrdi lb2er-sf slof slos
Alarm Status       :
BER SD Threshold   : 6                       BER SF Threshold     : 3
Hold time up       : 500 milliseconds        Reset On Path Down   : Disabled
Hold time down     : 200 milliseconds

Transceiver Data

Transceiver Status : not-equipped

===============================================================================
Port Statistics
===============================================================================
                                                        Input                Output
-------------------------------------------------------------------------------
Packets                                                     0                     0
Discards                                                    0                     0
Unknown Proto Discards                                      0
===============================================================================
*A:PE-1#
```

Step 4. Verify the MC-APS status and parameters on PE-1 and PE-2

Detailed parameters of the APS configuration on PE-1 can be verified, as follows. The admin/oper status of APS group 1 shows up/up. K1/K2 byte shows N/A as APS 1+1 exchanges that information through the protection circuit.

The admin/oper status of the working circuit (the link between MSAN and PE-1) is up/up.

```
*A:PE-1# show aps detail

===============================================================================
APS Group: aps-1
===============================================================================
Description        : APS Group
Group Id           : 1                     Active Circuit       : 1/2/1
Admin Status       : Up                    Oper Status          : Up
Working Circuit    : 1/2/1                 Protection Circuit   : N/A
Switching-mode     : Bi-directional        Switching-arch       : 1+1(sig-only)
Annex B            : No
Revertive-mode     : Non-revertive         Revert-time (min)    :
Rx K1/K2 byte      : N/A
Tx K1/K2 byte      : N/A
Current APS Status : OK
Multi-Chassis APS  : Yes
Neighbor           : 192.0.2.2
Control link state : Up
Advertise Interval : 1000 msec             Hold Time            : 3000 msec
Mode mismatch Cnt  : 0                     Channel mismatch Cnt : 0
PSB failure Cnt    : 0                     FEPL failure Cnt     : 0
-------------------------------------------------------------------------------
 APS Working Circuit - 1/2/1
-------------------------------------------------------------------------------
Admin Status       : Up                    Oper Status          : Up
Current APS Status : OK                    No. of Switchovers   : 0
Last Switchover    : None                  Switchover seconds   : 0
Signal Degrade Cnt : 0                     Signal Failure Cnt   : 0
Last Switch Cmd    : N/A                   Last Exercise Result : N/A
Tx L-AIS           : None
===============================================================================
*A:PE-1#
```

Detailed parameters of the APS configuration on PE-2 can be verified, as follows. The admin/oper status of APS group 1 shows up/up. Both Rx and Tx of the K1/K2 byte are in the status of 0x00/0x05 (No-Req on Protect) as there is no failure or force-switchover request.

The admin/oper status of the protection circuit (the link between MSAN and PE-2) is up/up.

```
*A:PE-2# show aps detail
===============================================================================
APS Group: aps-1
===============================================================================
Description        : APS Group
Group Id           : 1                     Active Circuit       : N/A
Admin Status       : Up                    Oper Status          : Up
```

```
Working Circuit    : N/A                  Protection Circuit   : 1/2/1
Switching-mode     : Bi-directional       Switching-arch       : 1+1(sig-only)
Annex B            : No
Revertive-mode     : Non-revertive        Revert-time (min)    :
Rx K1/K2 byte      : 0x00/0x05 (No-Req on Protect)
Tx K1/K2 byte      : 0x00/0x05 (No-Req on Protect)
Current APS Status : OK
Multi-Chassis APS  : Yes
Neighbor           : 192.0.2.1
Control link state : Up
Advertise Interval : 1000 msec            Hold Time            : 3000 msec
Mode mismatch Cnt  : 0                    Channel mismatch Cnt : 0
PSB failure Cnt    : 0                    FEPL failure Cnt     : 1
-------------------------------------------------------------------------------
 APS Working Circuit - Neighbor
-------------------------------------------------------------------------------
Admin Status       : N/A                  Oper Status          : N/A
Current APS Status : OK                   No. of Switchovers   : 0
Last Switchover    : None                 Switchover seconds   : 0
Signal Degrade Cnt : 0                    Signal Failure Cnt   : 1
Last Switch Cmd    : No Cmd               Last Exercise Result : Unknown
Tx L-AIS           : None
-------------------------------------------------------------------------------
 APS Protection Circuit - 1/2/1
-------------------------------------------------------------------------------
Admin Status       : Up                   Oper Status          : Up
Current APS Status : OK                   No. of Switchovers   : 0
Last Switchover    : None                 Switchover seconds   : 0
Signal Degrade Cnt : 0                    Signal Failure Cnt   : 0
Last Switch Cmd    : No Cmd               Last Exercise Result : Unknown
Tx L-AIS           : None
===============================================================================
*A:PE-2#
```

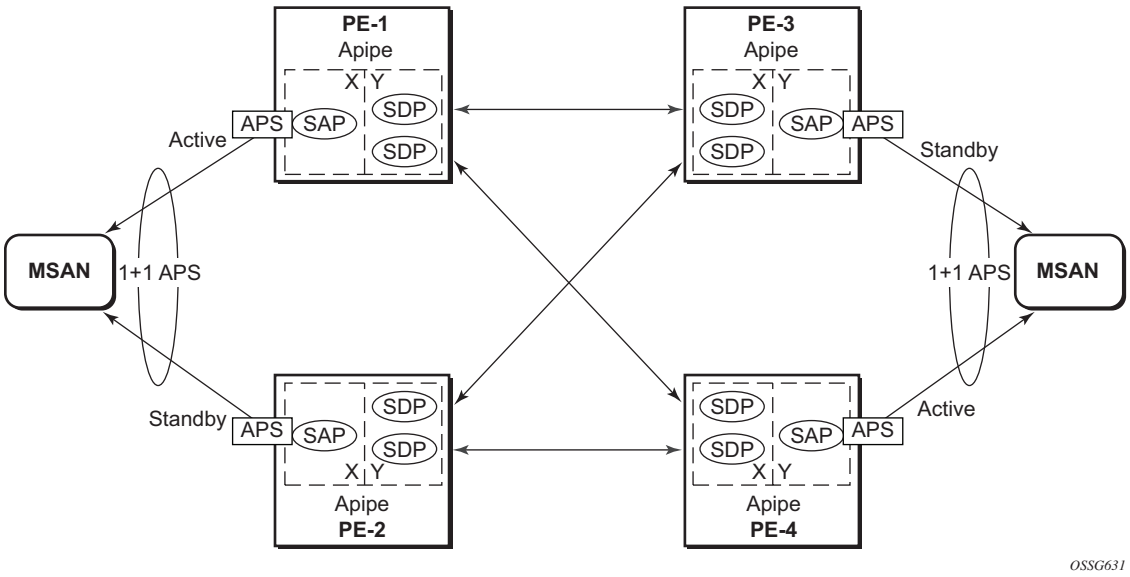
Step 5. MC-APS configuration on PE-3 and PE-4

The MC-APS configuration on PE-3 and PE-4 is similar to the configuration on PE-1 and PE-2. Configure the working circuit on PE-4 and the protection circuit on PE-3.

Step 6. Pseudowire configuration

Configure an Apipe service on every PE and create endpoints X and Y. Associate the SAPs and spoke SDPs with the endpoints, as shown in Figure 24.

Figure 24 **Association of SAPs/SDPs and Endpoints**




```
*A:PE-1# configure
    service
        apipe 1 customer 1 create
            endpoint "X" create
            exit
            endpoint "Y" create
            exit
            sap aps-1:0/32 endpoint "X" create
            exit
            spoke-sdp 13:1 endpoint "Y" create
            exit
            spoke-sdp 14:1 endpoint "Y" create
            exit
            no shutdown
        exit
```

Syntax aps-1:0/32 specifies the APS group and VPI/VCI of the ATM circuit (aps-id:vpi/vci).

Likewise, an Apipe service, with endpoints, SAPs and spoke SDPs must be configured on the other PE routers.

Step 7. Pseudowire verification

The Apipe service is up in PE-1 (MC-APS working circuit), as follows:

```
*A:PE-1# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Up   1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
```

```
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The Apipe service is down in PE-2 (MC-APS protect circuit), as follows:

```
*A:PE-2# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Down 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

The Apipe service is down in PE-3 (MC-APS protect circuit), as follows:

```
*A:PE-3# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Down 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

The Apipe service is up in PE-4 (MC-APS working circuit), as follows:

```
*A:PE-4# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Up   1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-4#
```

Note: After configuring ICB spoke-SDPs, the Apipe will be up on all PEs.

Step 8. Verify SDP status

The status of SDP 23:1 on PE-2 can be verified as follows.

Peer Pw Bits shows the status of the pseudowire on the peer node. In this example, both the local node (PE-2) as the remote node (PE-3) are sending the **lacIngressFault lacEgressFault** and **pwFwdingStandby** flags. This is because the Apipe service on these nodes is down because the MC-APS is in protection status.

```
*A:PE-2# show service id 1 sdp 23:1 detail

===============================================================================
Service Destination Point (Sdp Id : 23:1) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 23:1  -(192.0.2.3)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id            : 23:1                      Type              : Spoke
Spoke Descr     : (Not Specified)
Split Horiz Grp   : (Not Specified)
VC Type           : AAL5SDU                   VC Tag            : 0
Admin Path MTU    : 0                         Oper Path MTU     : 1556
Delivery          : MPLS
Far End           : 192.0.2.3
Tunnel Far End    : 192.0.2.3                 LSP Types         : LDP

Admin State       : Up                        Oper State        : Up
Acct. Pol         : None                      Collect Stats     : Disabled
Ingress Label     : 262139                    Egress Label      : 262138
Ingr Mac Fltr-Id  : n/a                       Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id   : n/a                       Egr IP Fltr-Id    : n/a
Admin ControlWord : Preferred                 Oper ControlWord  : True
Admin BW(Kbps)    : 0                         Oper BW(Kbps)     : 0
BFD Template      : None
BFD-Enabled       : no                        BFD-Encap         : ipv4
Last Status Change : 10/25/2016 13:48:30      Signaling         : TLDP
Last Mgmt Change  : 10/25/2016 13:48:22
Endpoint          : Y                         Precedence        : 4
PW Status Sig     : Enabled
Class Fwding State : Down
Flags             : None
Local Pw Bits     : lacIngressFault lacEgressFault pwFwdingStandby
Peer Pw Bits      : lacIngressFault lacEgressFault pwFwdingStandby
Peer Fault Ip     : None
Peer Vccv CV Bits  : lspPing bfdFaultDet
Peer Vccv CC Bits  : pwe3ControlWord mplsRouterAlertLabel

Ingress Qos Policy : (none)                   Egress Qos Policy : (none)
Ingress FP QGrp   : (none)                    Egress Port QGrp  : (none)
Ing FP QGrp Inst  : (none)                    Egr Port QGrp Inst: (none)

KeepAlive Information :
Admin State       : Disabled                  Oper State        : Disabled
Hello Time        : 10                        Hello Msg Len     : 0
Max Drop Count    : 3                         Hold Down Time    : 10
```

```
---snip---
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

In case of failure, the access link can be protected by MC-APS. An MPLS network failure can be protected by pseudowire redundancy. Node failure can be protected by the combination of MC-APS and pseudowire redundancy.

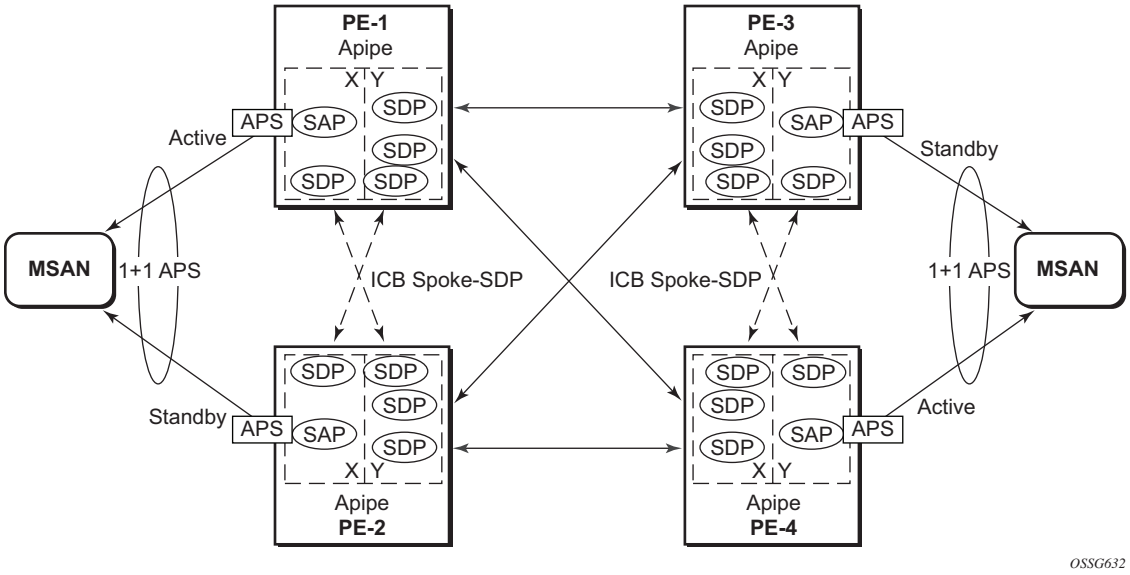
Step 9. Inter-Chassis Backup (ICB) pseudowire configuration.

Configuring Inter-Chassis Backup (ICB) is optional. It can reduce traffic impact by forwarding traffic on ICB spoke SDPs during MC-APS switchover. The ICB spoke SDP cannot be added to the endpoint if the SAP is not part of an MC-APS (or MC-LAG) instance. Conversely, a SAP which is not part of a MC-APS (or MC-LAG) instance cannot be added to an endpoint which already has an ICB spoke SDP. Forwarding between ICBs is blocked on the same node. The user has to explicitly indicate the spoke SDP is actually an ICB at creation time. Figure 5 shows some setup examples where ICBs are required.

After configuring ICB spoke SDPs, the Apipe will be in admin/oper up/up status on all PE routers.

ICB SDPs are configured and associated to endpoints, as shown in Figure 25.

Figure 25 **ICB Spoke SDPs and Association with the Endpoints**



Two ICB spoke SDPs must be configured in the Apipe service on each PE router, one in each endpoint. The same SDP IDs can be used for the ICBs since the far-end will be the same. However, the vc-id must be different. The ICB spoke SDPs must cross, meaning one end should be associated with endpoint X and the other end (on the other PE) should be associated with endpoint Y.

An ICB is always the last forwarding resort. Only one spoke SDP will be forwarding. If there is an ICB and an MC-APS SAP in an endpoint, the ICB will only forward if the SAP goes down. If an ICB resides in an endpoint together with other spoke SDPs the ICB will only forward if there is no other active spoke SDP.

The following shows the additional configuration for ICB on each PE:

```
*A:PE-1# configure
    service
        apipe 1
            spoke-sdp 12:1 endpoint "X" icb create
            exit
            spoke-sdp 12:2 endpoint "Y" icb create
            exit


*A:PE-2# configure
    service
        apipe 1
            spoke-sdp 21:1 endpoint "Y" icb create
            exit
            spoke-sdp 21:2 endpoint "X" icb create
            exit


*A:PE-3# configure
    service
        apipe 1
            spoke-sdp 34:1 endpoint "X" icb create
            exit
            spoke-sdp 34:2 endpoint "Y" icb create
            exit


*A:PE-4# configure
    service
        apipe 1
            spoke-sdp 43:1 endpoint "Y" icb create
            exit
            spoke-sdp 43:2 endpoint "X" icb create
            exit
```

Step 10. Verification of active objects for each endpoint

The following command shows which objects are configured for each endpoint and which is the active object at this moment:

```
*A:PE-1# show service id 1 endpoint

===============================================================================
Service 1 endpoints
===============================================================================
Endpoint name               : X
Description                 : (Not Specified)
Creation Origin             : manual
Revert time                 : 0
Act Hold Delay              : 0
Tx Active                   : aps-1:0/32
Tx Active Up Time           : 0d 00:02:58
Revert Time Count Down      : N/A
Tx Active Change Count      : 1
Last Tx Active Change       : 10/25/2016 13:48:15
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
SAP      : aps-1:0/32                           Oper Status: Up
Spoke-sdp: 12:1 Prec:4 (icb)                    Oper Status: Up
===============================================================================
Endpoint name               : Y
Description                 : (Not Specified)
Creation Origin             : manual
Revert time                 : 0
Act Hold Delay              : 0
Tx Active (SDP)             : 14:1
Tx Active Up Time           : 0d 00:02:35
Revert Time Count Down      : N/A
Tx Active Change Count      : 2
Last Tx Active Change       : 10/25/2016 13:48:38
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 12:2 Prec:4 (icb)                    Oper Status: Up
Spoke-sdp: 13:1 Prec:4                          Oper Status: Up
Spoke-sdp: 14:1 Prec:4                          Oper Status: Up
===============================================================================
===============================================================================
*A:PE-1#
```

On PE-1, both the SAP and the spoke SDP 14:1 are active. The other objects do not forward traffic.

Step 11. Other types of setups

The following figures show other setups that combine MC-APS and pseudowire redundancy.

Figure 26 **Additional Setup Example 1 (Part 1)**

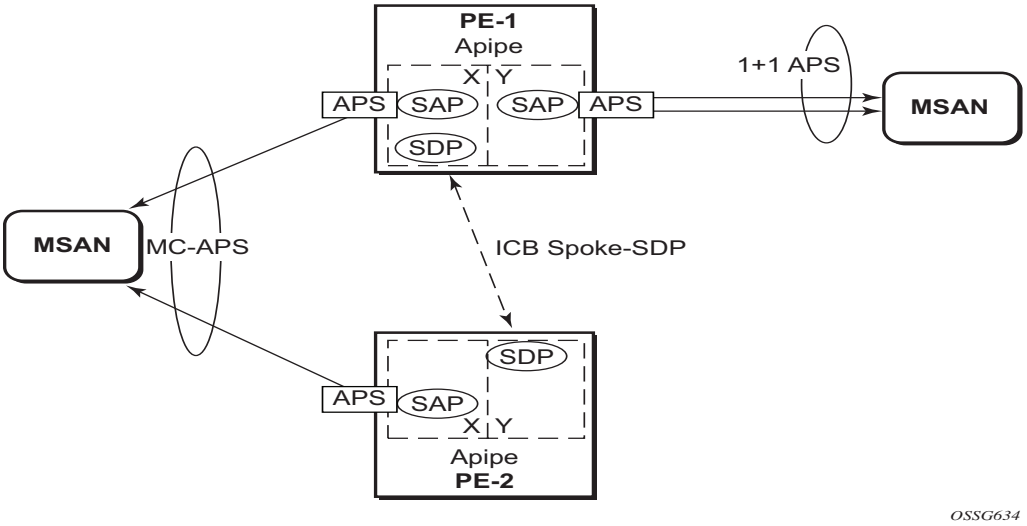


Figure 27 **Additional Setup Example 1 (Part 2)**

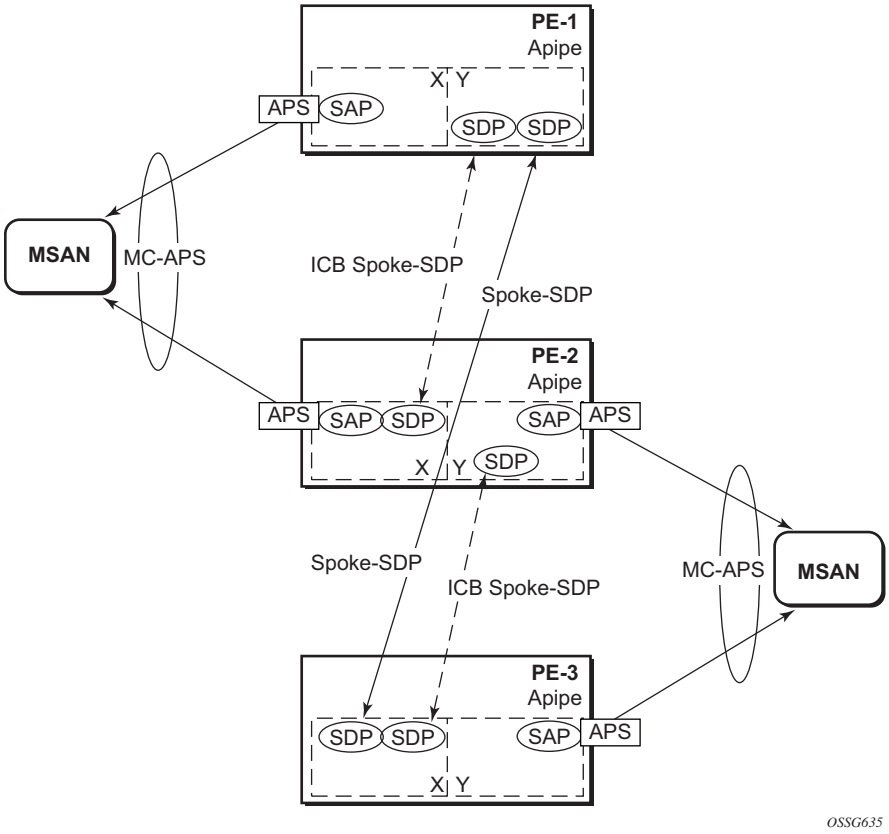


Figure 28 **Additional Setup Example 2 (Part 1a)**

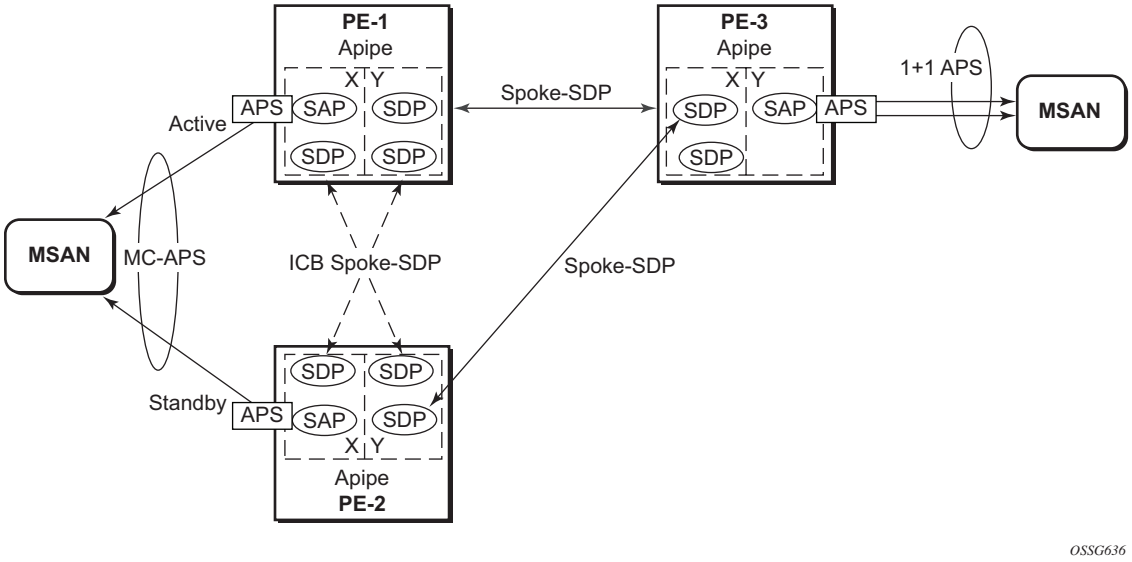


Figure 29 **Additional Setup Example 2 (Part 1b)**

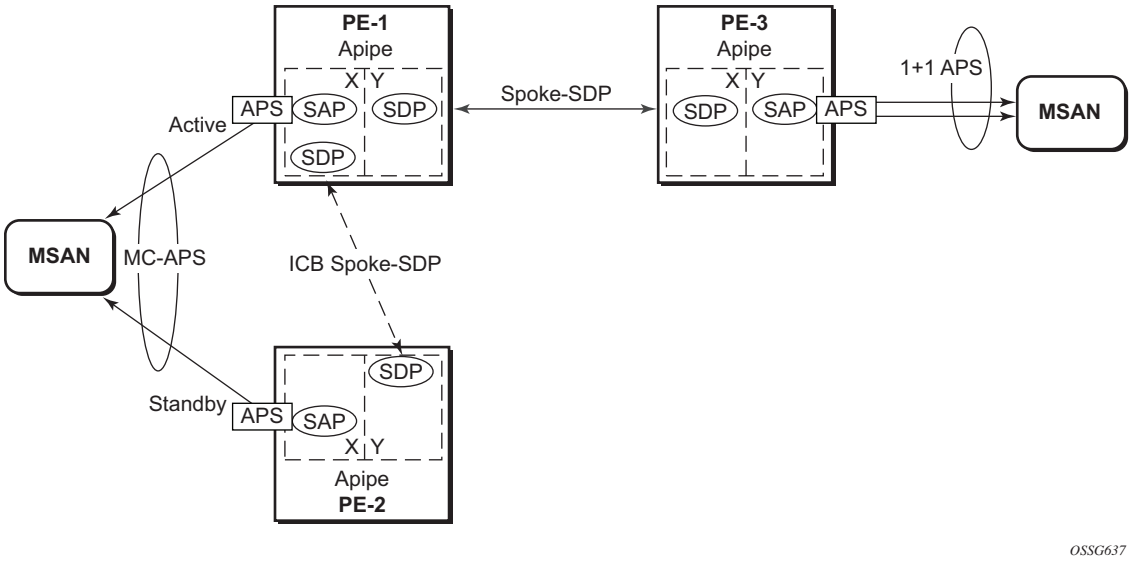
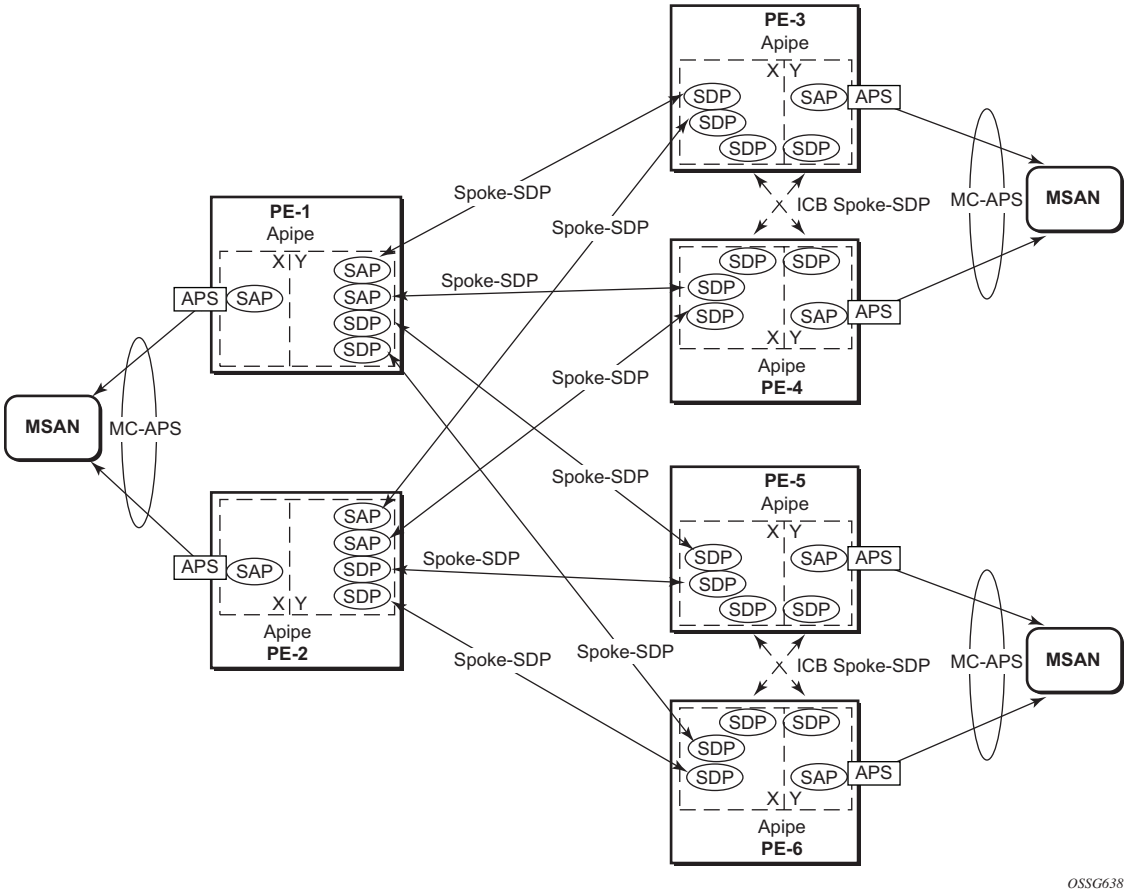


Figure 30 **Additional Setup Example 2 (Part 2)**



Forced Switchover

MC-APS convergence can be forced with the **tools perform aps** command:

```
*A:PE-1# tools perform aps force
  - force <aps-id> {protect|working} [number <number>]

 <aps-id>             : aps-<group-id>
                          aps            - keyword
                          group-id       - [1..128]
 <protect|working>    : keyword
 <number>             : [1-2]
```

After the forced switchover, it is important to clear the forced switchover:

```
*A:PE-1# tools perform aps clear
  - clear <aps-id> {protect|working} [number <number>]
```

```
<aps-id>              : aps-<group-id>
                          aps            - keyword
                          group-id       - [1..128]
<protect|working>     : protect|working
<number>              : [1-2]
```

Conclusion

In addition to Multi-Chassis LAG, Multi-Chassis APS provides a solution for both network redundancy and access node redundancy. It supports ATM VLL and Ethernet VLL with ATM SAP. Access links and PE nodes are protected by APS and the MPLS network is protected by pseudowire redundancy/FRR. With this feature, Nokia can provide resilient end-to-end solutions.

Multi-Chassis LAG and Pseudowire Redundancy Interworking

This chapter provides information about Multi-Chassis Link Aggregation (MC-LAG) and pseudowire redundancy interworking.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

MC-LAG is supported only on Ethernet MDAs, and this only for access ports, because the LAG group must be in access mode.

This chapter was initially written for SR OS release 7.0.R5. However, the CLI in the current edition is based on SR OS release 14.0.R5.

Overview

MC-LAG

MC-LAG is an extension to the LAG feature to provide not only link redundancy but also node-level redundancy. This feature provides a Nokia added value solution which is not defined in any IEEE standard.

A proprietary messaging system between redundant-pair nodes supports coordinating the LAG switchover.

Multi-chassis LAG supports LAG switchover coordination: one node connected to two redundant-pair peer nodes with the LAG. During the LACP negotiation, the redundant-pair peer nodes act like a single node using active/stand-by signaling to ensure that only links of one peer node are used at a time.

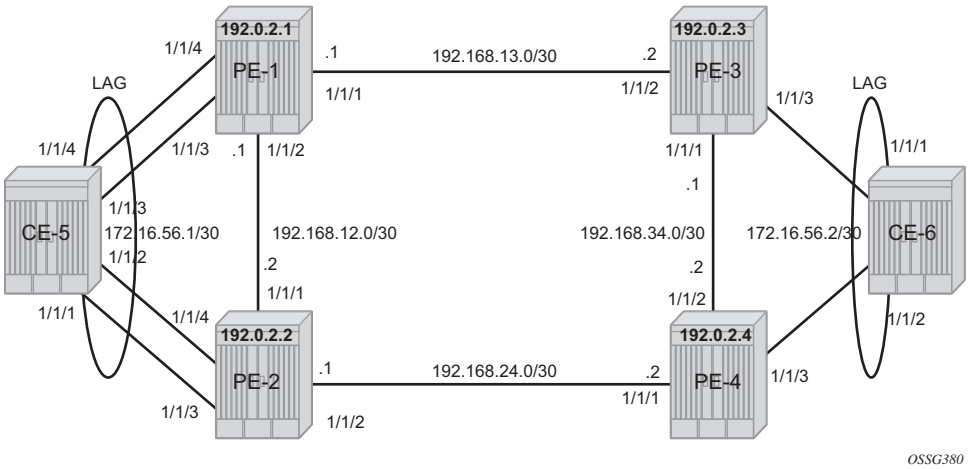
Pseudowire Redundancy

Pseudowire (PW) redundancy provides the ability to protect a pseudowire with a pre-provisioned pseudowire and to switch traffic over to the secondary standby pseudowire in case of a SAP and/or network failure condition. Normally, pseudowires are redundant by the virtue of the SDP redundancy mechanism. For instance, if the SDP is an RSVP LSP and is protected by a secondary standby path and/or by Fast-Reroute paths, the pseudowire is also protected.

However, there are a few applications in which SDP redundancy does not protect the end-to-end pseudowire path when there are two different destination 7x50 PE nodes for the same VLL service. The main use case is the provision of dual-homing of a CPE or access node to two 7x50 PE nodes located in different POPs. The other use case is the provisioning of a pair of active and standby BRAS nodes, or active and standby links to the same BRAS node, to provide service resiliency to broadband service subscribers.

Example Topology

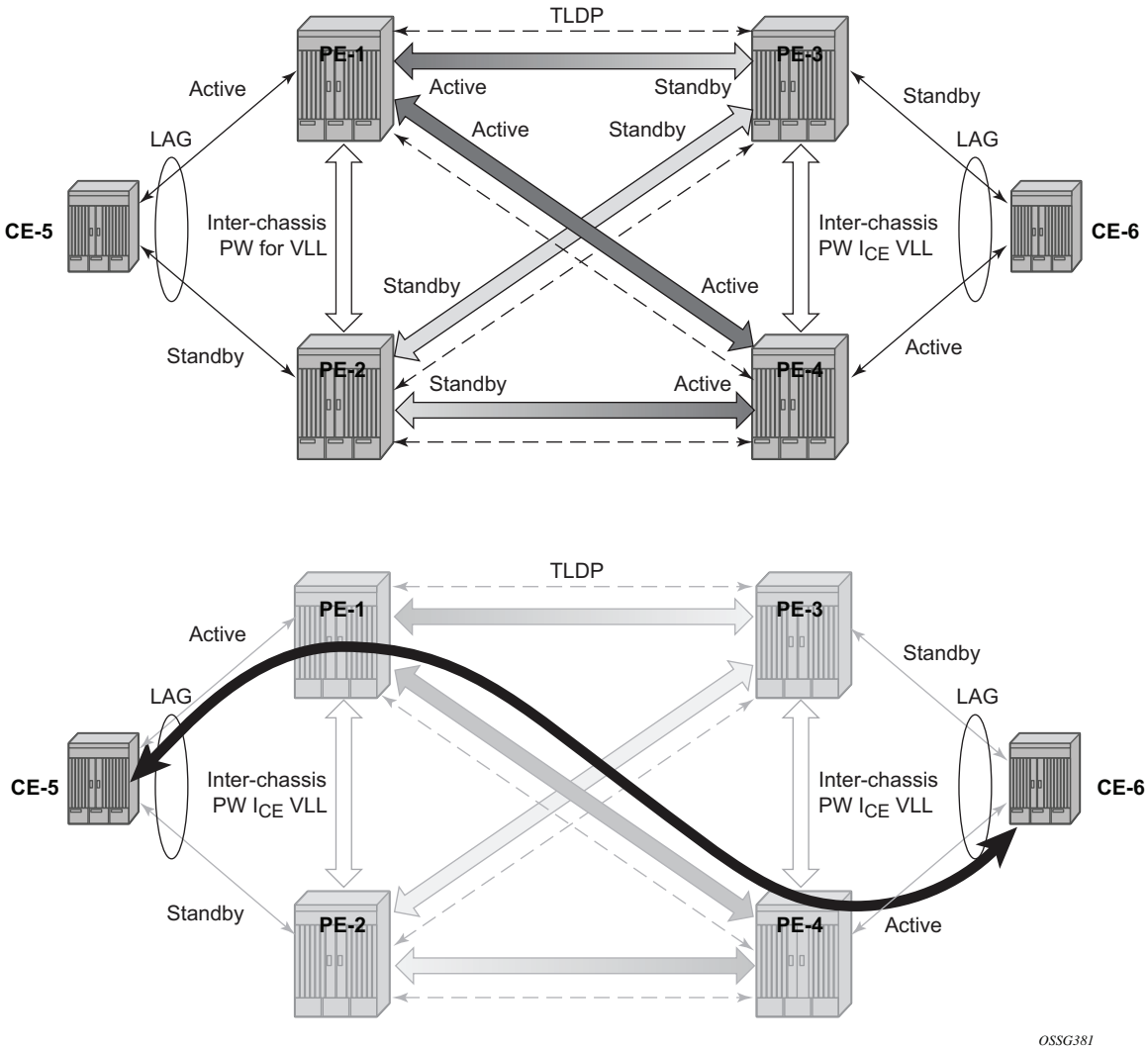
Figure 31 **MC-LAG Example Topology**



This section describes a setup which contains 2 CE and 4 PE nodes. The CE nodes can be any routing/switching device that support the OUT_OF_SYNC signaling as described in IEEE Standard 802.3-2005 section 3 section 43.6.1. Figure 31 shows the physical topology of the setup.

Figure 32 shows the use of both MC-LAG in the access network and pseudowire redundancy in the core network to provide a resilient end-to-end VLL service between CE-5 and CE-6.

Figure 32 **Network Resiliency**



When an SDP is in standby, it sends the pseudowire status bit pwFwdingStandby to its peer.

Configuration

It is assumed that the following base configuration has been implemented on the PEs:

- Cards, MDAs and ports

- Interfaces
- IGP configured and converged
- MPLS
- SDPs configured between all PE routers

Either OSPF or IS-IS can be used as the IGP. Both LDP or RSVP can be used for signaling the transport MPLS labels. Alternatively, GRE can be used for the transport tunnels.

It does not matter if the SDPs are using LDP, RSVP, or GRE. In this example, OSPF and LDP are used.

The following command is used to check if OSPF has converged (for example, on PE-1):

```
*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     01h10m55s  0
       system                                                        0
192.0.2.2/32                                    Remote  OSPF      01h10m01s  10
       192.168.12.2                                                  100
192.0.2.3/32                                    Remote  OSPF      01h09m55s  10
       192.168.13.2                                                  100
192.0.2.4/32                                    Remote  OSPF      01h09m05s  10
       192.168.12.2                                                  200
192.168.12.0/30                                 Local   Local     01h10m55s  0
       int-PE-1-PE-2                                                 0
192.168.13.0/30                                 Local   Local     01h10m55s  0
       int-PE-1-PE-3                                                 0
192.168.24.0/30                                 Remote  OSPF      01h10m01s  10
       192.168.12.2                                                  200
192.168.34.0/30                                 Remote  OSPF      01h09m55s  10
       192.168.13.2                                                  200
-------------------------------------------------------------------------------
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

The following command shows that the SDPs are up:

```
*A:PE-1# show service sdp
===============================================================================
Services: Service Destination Points
```

```
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr          Del     LSP   Sig
-------------------------------------------------------------------------------
12     0       1556    192.0.2.2        Up   Up           MPLS    L     TLDP
13     0       1556    192.0.2.3        Up   Up           MPLS    L     TLDP
14     0       1556    192.0.2.4        Up   Up           MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
===============================================================================
*A:PE-1#
```

Step 1. MC-LAG configuration on CEs.

The LAG configuration on CEs is only included for completeness of the example; any CE device could be used.

Auto-negotiation must be switched off or set to limited on all ports that will be included into the LAG in order to guarantee a specific port speed.

Note: Disabling autonegotiation on Gigabit ports is not allowed because the IEEE 802.3 specification for Gigabit Ethernet requires autonegotiation to be enabled for far end fault detection.

Configure LACP on the LAG with at least one side of the LAG in **active** mode.

```
*A:CE-5# configure port 1/1/[1..4] ethernet autonegotiate limited
*A:CE-5# configure port 1/1/[1..4] no shutdown
*A:CE-5# configure lag 1
        port 1/1/1 1/1/2 1/1/3 1/1/4
        lacp active
        no shutdown
```

Step 2. LAG configuration on PEs.

The PE ports connected to the CEs must be configured as access ports because they will be used in the redundant pseudowire service. The LAG must also be configured in access mode.

The LAG encapsulation type (null | dot1q | qinq) must match the port encapsulation type of the LAG members.

Auto-negotiation must be switched off or configured to limited.

Configure LACP on the LAG. At least 1 side of the LAG (PE or CE) must be configured in **active** mode.

```
*A:PE-1# configure port 1/1/[3..4] ethernet mode access
*A:PE-1# configure port 1/1/[3..4] ethernet autonegotiate limited
*A:PE-1# configure port 1/1/[3..4] no shutdown
```

```
*A:PE-1# configure lag 1
        mode access
        port 1/1/3 1/1/4
        lacp active
        no shutdown
```

Step 3. MC-LAG configuration on PE-1 and PE-2

The redundant PEs must act as one virtual node toward the CE. They have to be able to communicate the same LACP parameters to the CE side.

The following parameters uniquely identify a LAG instance:

- lacp-key
- system-id
- system-priority

These three parameters must be configured with the same value on both redundant PEs.

Multi-chassis redundancy with a peering session (which operates by an IP connection using UDP destination port 1025) is configured toward the redundant PE system address and MC-LAG redundancy is enabled, as follows. The peering session can be configured with MD5 authentication.

```
*A:PE-1# configure
    redundancy
        multi-chassis
            peer 192.0.2.2 create
                authentication-key "441dO/0RgDhHgzYwpOCTK9zbKjv4GZ/z" hash2
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:00:01 system-priority 100
                    no shutdown
                exit
                no shutdown
            exit
        exit


*A:PE-2# configure
    redundancy
        multi-chassis
            peer 192.0.2.1 create
                authentication-key "441dO/0RgDg2CA0JlyzVNQBoRc327b1j" hash2
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:00:01 system-priority 100
                    no shutdown
                exit
                no shutdown
            exit
        exit
```

Step 4. MC-LAG verification.

Verify MC peers showing that the authentication and admin state are enabled.

```
*A:PE-1# show redundancy multi-chassis sync
```

```
===============================================================================
Multi-chassis Peer Table
===============================================================================
Peer
-------------------------------------------------------------------------------
Peer IP Address         : 192.0.2.2
Description             : (Not Specified)
Authentication          : Enabled
Source IP Address       : 192.0.2.1
Admin State             : Enabled
Warm standby            : No
Remote warm standby     : No
-------------------------------------------------------------------------------
Sync: Not-configured
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
*A:PE-1#
```

Step 5. Verify MC-LAG peer status and LAG parameters.

```
*A:PE-1# show redundancy multi-chassis mc-lag peer 192.0.2.2
===============================================================================
Multi-Chassis MC-Lag Peer 192.0.2.2
===============================================================================
Last State chg  : 10/25/2016 07:40:53
Admin State     : Up                 Oper State           : Up
KeepAlive       : 10 deci-seconds    Hold On Ngbr Failure : 3
-------------------------------------------------------------------------------
Lag Id Lacp   Remote Source Oper   System Id          Sys   Last State Changed
       Key    Lag Id MacLSB MacLSB                    Prio
-------------------------------------------------------------------------------
1      1      1      Def    n/a    00:00:00:00:00:01  100   10/25/2016 07:40:54
-------------------------------------------------------------------------------
Number of LAGs : 1
===============================================================================
*A:PE-1#
```

There is a fixed keepalive timer of 1 second. The **hold-on-neighbor-failure multiplier** command indicates the interval that the standby node will wait for packets from the active node before assuming a redundant-neighbor failure. The **hold-on-neighbor-failure multiplier** command is configurable in the **config>redundancy>multi-chassis>peer>mc-lag** context. The standby node will also assume a redundant-neighbor failure when there is no route available to the redundant-neighbor.

```
*A:PE-1# configure
    redundancy
        multi-chassis
            peer 192.0.2.2
                mc-lag
                    hold-on-neighbor-failure 10
```

In this example, the *lag-id* is 1 on both redundant PEs. This is not mandatory. If the *lag-id* on PE-2 is, for example 2, the following should be configured on PE-1:

```
*A:PE-1# configure
    redundancy
        multi-chassis
            peer 192.0.2.2
                mc-lag
                    lag 1 remote-lag 2 lacp-key 1 system-id 00:00:00:00:00:01
                                                  system-priority 100
```

Step 6. Verify MC-LAG

```
*A:PE-1# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      up      No       0         2        active
===============================================================================
*A:PE-1#

*A:PE-2# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      down    No       0         0        standby
===============================================================================
*A:PE-2#
```

In this case, the LAG on PE-1 is active (operationally up) whereas the LAG on PE-2 is standby (operationally down).

The selection criteria by default is highest number of links and priority. In this example, the number of links and the priority of the links is the same on both redundant PEs. Whichever PE's LAG gets the operational status **up** first, will be the active.

LAG ports of one PE could be preferred over the other PE by configuring port priority. For example, the following command lowers the priority of the LAG ports on PE-1, thus giving this LAG higher preference. The default priority is 32768, but it is modified to a value of 10, as follows:

```
*A:PE-1# configure lag 1 port 1/1/3 1/1/4 priority 10
```

The selection criteria can be configured as highest-count, highest-weight or best-port (the default is highest count).

```
*A:PE-1# configure lag 1 selection-criteria
  - selection-criteria [best-port|highest-count|highest-weight] [slave-to-partner]
[subgroup-hold-time <hold-time>]
```

```
  - no selection-criteria

 <best-port|highest*> : keywords
 <slave-to-partner>   : keyword
 <hold-time>          : [0..2000] tenths of a second | infinite
```

If highest-weight is configured, the sum of the weights of the LAG members is considered. The weight of an individual LAG member is calculated as priority 65535 (the default is 32768).

Step 7. Verify detailed MC-LAG status on PE-1

```
*A:PE-1# show lag 1 detail
===============================================================================
LAG Details
===============================================================================
Description        : N/A
-------------------------------------------------------------------------------
Details
-------------------------------------------------------------------------------
Lag-id              : 1                     Mode                : access
Adm                 : up                    Opr                 : up
Thres. Exceeded Cnt : 2                     Port Threshold      : 0
Thres. Last Cleared : 10/25/2016 07:42:51    Threshold Action    : down
Dynamic Cost        : false                 Encap Type          : null
Configured Address  : 4a:c4:ff:00:01:41     Lag-IfIndex         : 1342177281
Hardware Address    : 4a:c4:ff:00:01:41     Adapt Qos (access)  : distribute
Hold-time Down      : 0.0 sec               Port Type           : standard
Per-Link-Hash       : disabled
Include-Egr-Hash-Cfg: disabled              Forced              : -
Per FP Ing Queuing  : disabled              Per FP Egr Queuing  : disabled
Per FP SAP Instance : disabled
Access Bandwidth    : not-applicable        Access Booking Factor: 100
Access Available BW : 0
Access Booked BW    : 0
LACP                : enabled               Mode                : active
LACP Transmit Intvl : fast                  LACP xmit stdby     : enabled
Selection Criteria  : highest-count         Slave-to-partner    : disabled
MUX control         : coupled
Subgrp hold time    : 0.0 sec               Remaining time      : 0.0 sec
Subgrp selected     : 1                     Subgrp candidate    : -
Subgrp count        : 1
System Id           : 4a:c4:ff:00:00:00     System Priority     : 32768
Admin Key           : 32768                 Oper Key            : 1
Prtr System Id      : 4a:c8:ff:00:00:00     Prtr System Priority : 32768
Prtr Oper Key       : 32768
Standby Signaling   : lacp
Port weight speed   : 0 gbps                Number/Weight Up    : 2
Weight Threshold    : 0                     Threshold Action    : down

MC Peer Address     : 192.0.2.2             MC Peer Lag-id      : 1
MC System Id        : 00:00:00:00:00:01     MC System Priority  : 100
MC Admin Key        : 1                     MC Active/Standby   : active
MC Lacp ID in use   : true                  MC extended timeout : false
MC Selection Logic  : local master decided
MC Config Mismatch  : no mismatch

-------------------------------------------------------------------------------
Port-id        Adm     Act/Stdby Opr     Primary   Sub-group     Forced  Prio
```

```
-------------------------------------------------------------------------------
1/1/3          up      active    up      yes       1             -       10
1/1/4          up      active    up                1             -       10

-------------------------------------------------------------------------------
Port-id        Role      Exp   Def   Dist  Col   Syn   Aggr  Timeout  Activity
-------------------------------------------------------------------------------
1/1/3          actor     No    No    Yes   Yes   Yes   Yes   Yes      Yes
1/1/3          partner   No    No    Yes   Yes   Yes   Yes   Yes      Yes
1/1/4          actor     No    No    Yes   Yes   Yes   Yes   Yes      Yes
1/1/4          partner   No    No    Yes   Yes   Yes   Yes   Yes      Yes
===============================================================================
```

After changing the LAG port priorities from default (32768) to 10, the LAG on PE-1 is in up/up state and the ports are in up/active/up status. This show command also displays actor and partner bits set in the LACP messages.

Step 8. MC-LAG configuration on PE-3 and PE-4.

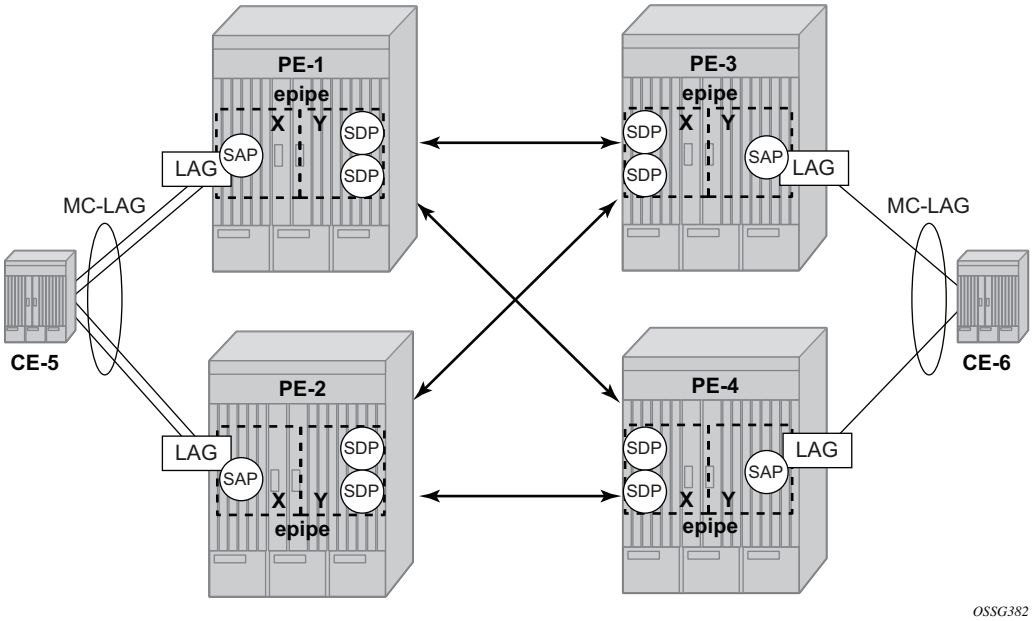
The MC-LAG configuration on PE-3 and PE-4 is similar to the configuration on PE-1 and PE-2. In this case, the priority of the LAG port on PE-4 is lowered to obtain the behavior in Figure 32 where the LAGs on PE-1 and PE-4 are active.

Step 9. Pseudowire configuration.

Configure an Epipe service on every PE and create endpoints **x** and **y** (the endpoint names can be any text string). Traffic can only be forwarded between two endpoints, for example, it is not possible for objects associated with the same endpoint to forward traffic to each other.

Associate the SAPs and spoke SDPs with the endpoints as shown in Figure 33.

Figure 33 **Association of SAPs/SDPs and Endpoints**




```
*A:PE-1# configure
    service
        epipe 1 customer 1 create
            endpoint "X" create
            exit
            endpoint "Y" create
            exit
            sap lag-1 endpoint "X" create
            exit
            spoke-sdp 13:1 endpoint "Y" create
            exit
            spoke-sdp 14:1 endpoint "Y" create
            exit
            no shutdown
        exit
```

Likewise, an Epipe service, endpoints, SAPs and spoke SDPs must be configured on the other PE routers.

Step 10. Pseudowire verification.

```
*A:PE-1# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
```

```
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#

*A:PE-2# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Down 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#

*A:PE-3# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Down 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#

*A:PE-4# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-4#
```

The Epipe service on PE-2 and PE-3 is down and up on PE-1 and PE-4. This reflects the standby behavior shown in Figure 32. However, after configuring ICB spoke SDPs (described later in this document), the Epipe will be in up/up status on all PE routers.

Step 11. Verify SDP status

Local pseudowire bits indicate the status of the pseudowire on the PE node. These pseudowire bits will be sent to the peer. Peer pseudowire bits indicate the status of the pseudowire on the peer, as sent by the peer. The following example is taken on PE-2:

```
*A:PE-2# show service id 1 sdp 23:1 detail

===============================================================================
Service Destination Point (Sdp Id : 23:1) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 23:1  -(192.0.2.3)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id          : 23:1                      Type            : Spoke
Spoke Descr     : (Not Specified)
VC Type         : Ether                     VC Tag          : n/a
Admin Path MTU  : 0                         Oper Path MTU   : 1556
Delivery        : MPLS
Far End         : 192.0.2.3
Tunnel Far End  : 192.0.2.3                 LSP Types       : LDP
Hash Label      : Disabled                  Hash Lbl Sig Cap  : Disabled
Oper Hash Label : Disabled
Entropy Label   : Disabled

Admin State     : Up                        Oper State      : Up
MinReqd SdpOperMTU : 1514
Acct. Pol       : None                      Collect Stats   : Disabled
Ingress Label   : 262139                    Egress Label    : 262138
Ingr Mac Fltr-Id  : n/a                     Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id   : n/a                     Egr IP Fltr-Id    : n/a
Ingr IPv6 Fltr-Id : n/a                     Egr IPv6 Fltr-Id  : n/a
Admin ControlWord : Not Preferred           Oper ControlWord  : False
Admin BW(Kbps)  : 0                         Oper BW(Kbps)   : 0
BFD Template    : None
BFD-Enabled     : no                        BFD-Encap       : ipv4
Last Status Change : 10/25/2016 07:45:49    Signaling       : TLDP
Last Mgmt Change   : 10/25/2016 07:45:39
Endpoint        : Y                         Precedence      : 4
PW Status Sig   : Enabled
Force Vlan-Vc   : Disabled                  Force Qinq-Vc   : Disabled
Class Fwding State : Down
Flags           : None
Local Pw Bits   : lacIngressFault lacEgressFault pwFwdingStandby
Peer Pw Bits    : lacIngressFault lacEgressFault pwFwdingStandby
Peer Fault Ip   : None
Peer Vccv CV Bits  : lspPing bfdFaultDet
Peer Vccv CC Bits  : mplsRouterAlertLabel


---snip---

-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

In this example, the remote side of the SDP is sending lacIngressFault lacEgressFault pwFwdingStandby flags. This is because the Epipe service on PE-3 is down because the MC-LAG is in standby/down status.

Link and node protection can be tested. The access links are protected by the MC-LAG, the PE routers are protected by the combination of MC-LAG/ pseudowire redundancy. The SDPs can be protected by FRR in the case of RSVP-TE or LDP.

Revertive behavior is expected when different MC-LAG port priorities are configured or if the number of MC-LAG ports is different on the MC-LAG peers: convergence takes place when the active PE fails and convergence takes place again when that PE is online again.

In case of revertive behavior, MC-LAG convergence might take less time than the setup of the spoke SDPs, thus creating a temporary black-hole. To avoid this situation, it is best to configure **hold-time up** on the LAG ports. In that case, the ports are kept in a down state for a configured period of time after the node has rebooted. This is done to ensure that the SDPs are operationally up when the MC-LAG convergence takes place. The **hold-time up** is expressed in seconds.

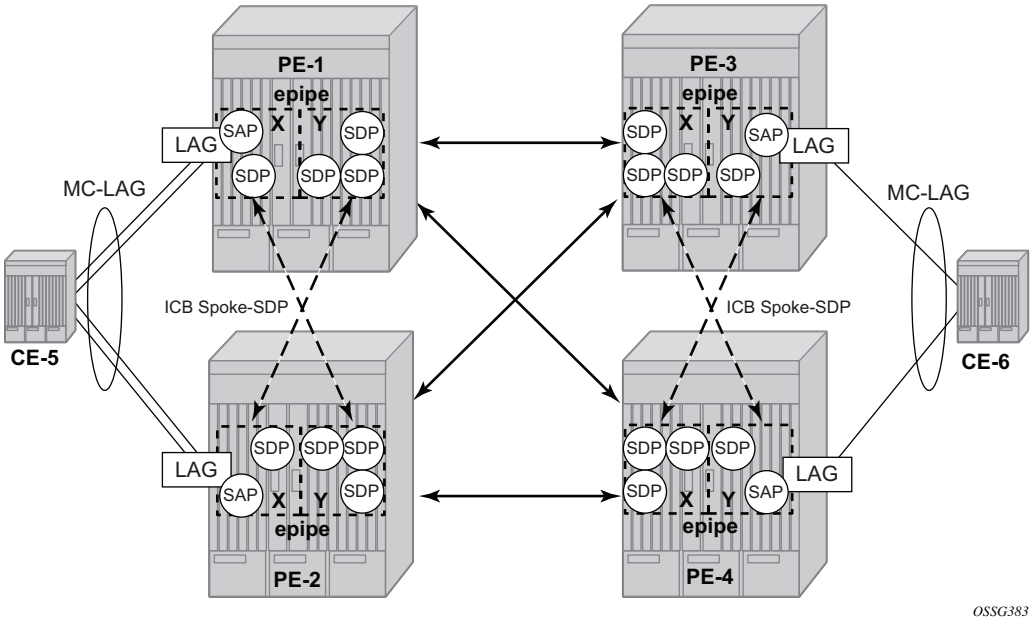
```
*A:PE-1# configure port 1/1/3 ethernet hold-time up 50
*A:PE-1# configure port 1/1/4 ethernet hold-time up 50
```

Step 12. Inter-Chassis Backup (ICB) pseudowire configuration.

In this setup, the configuration of ICBs is optional. It can be used to speed up convergence by forwarding in-flight packets during MC-LAG transition. Figure 35 shows some setup examples where ICBs are required. ICBs cannot be configured at endpoints where the other object is a standard SAP, only MC-LAG SAPs and pseudowires are allowed with ICBs.

ICB SDPs and associated to endpoints as shown in Figure 34.

Figure 34 **ICB Spoke SDPs and Their Association with the Endpoints**



Two ICB spoke SDPs must be configured in the Epipe service on each PE router, one in each endpoint. Different SDP IDs can be used for the ICBs (as opposed to the regular pseudowires), but this is not necessary, because the far-end will be the same. The *vc-id* must be different however.

The ICB spoke SDPs must cross, one end should be associated with endpoint **x** and the other end (on the other PE) should be associated with endpoint **y**. After configuring the ICB spoke SDPs, the Epipe service will be up/up on all four PE routers.

Only one spoke SDP will be forwarding. If there is an ICB and a MC-LAG SAP in an endpoint, the ICB will only forward if the SAP goes down. If an ICB resides in an endpoint together with other spoke SDPs, the ICB will only forward if there is no other active spoke SDP.

The following output shows the additional Epipe service configuration on each PE:

```
*A:PE-1# configure
    service
        epipe 1
            spoke-sdp 12:1 endpoint "X" icb create
            exit
            spoke-sdp 12:2 endpoint "Y" icb create
            exit


*A:PE-2# configure
    service
```

```
        epipe 1
            spoke-sdp 21:1 endpoint "Y" icb create
            exit
            spoke-sdp 21:2 endpoint "X" icb create
            exit

*A:PE-3# configure service
        epipe 1
            spoke-sdp 34:1 endpoint "X" icb create
            exit
            spoke-sdp 34:2 endpoint "Y" icb create
            exit

*A:PE-4# configure service
        epipe 1
            spoke-sdp 43:1 endpoint "Y" icb create
            exit
            spoke-sdp 43:2 endpoint "X" icb create
            exit
```

Step 13. Verification of active objects for each endpoint.

The following command shows which objects are configured for each endpoint and which is the active object at this moment:

```
*A:PE-1# show service id 1 endpoint

===============================================================================
Service 1 endpoints
===============================================================================
Endpoint name               : X
Description                 : (Not Specified)
Creation Origin             : manual
Revert time                 : 0
Act Hold Delay              : 0
Standby Signaling Master    : false
Standby Signaling Slave     : false
Tx Active                   : lag-1
Tx Active Up Time           : 0d 00:04:27
Revert Time Count Down      : N/A
Tx Active Change Count      : 1
Last Tx Active Change       : 10/25/2016 07:44:51
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
SAP      : lag-1                                    Oper Status: Up
Spoke-sdp: 12:1 Prec:4 (icb)                        Oper Status: Up
===============================================================================
Endpoint name               : Y
Description                 : (Not Specified)
Creation Origin             : manual
Revert time                 : 0
Act Hold Delay              : 0
Standby Signaling Master    : false
Standby Signaling Slave     : false
Tx Active (SDP)             : 14:1
Tx Active Up Time           : 0d 00:03:47
Revert Time Count Down      : N/A
```

```
Tx Active Change Count       : 2
Last Tx Active Change        : 10/25/2016 07:45:31
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 12:2 Prec:4 (icb)                        Oper Status: Up
Spoke-sdp: 13:1 Prec:4                              Oper Status: Up
Spoke-sdp: 14:1 Prec:4                              Oper Status: Up
===============================================================================
===============================================================================
*A:PE-1#
```

On PE-1, the SAP and the spoke SDP 14:1 are active. The other objects do not forward traffic.

Step 14. Other types of setups.

Figure 35 and Figure 36 show other setups that combine MC-LAG and pseudowire redundancy.

Figure 35 **Additional Setup Example 1**

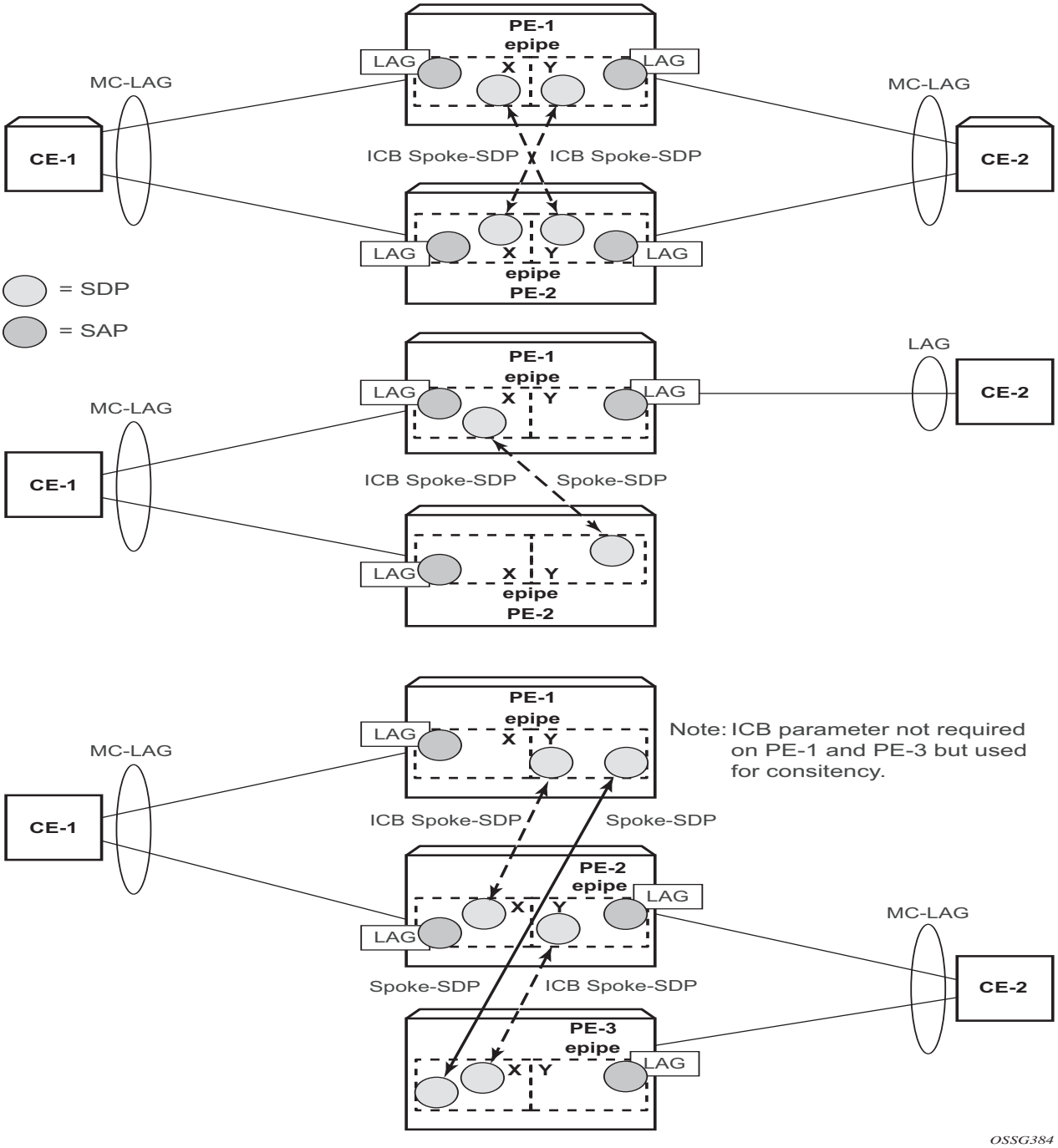
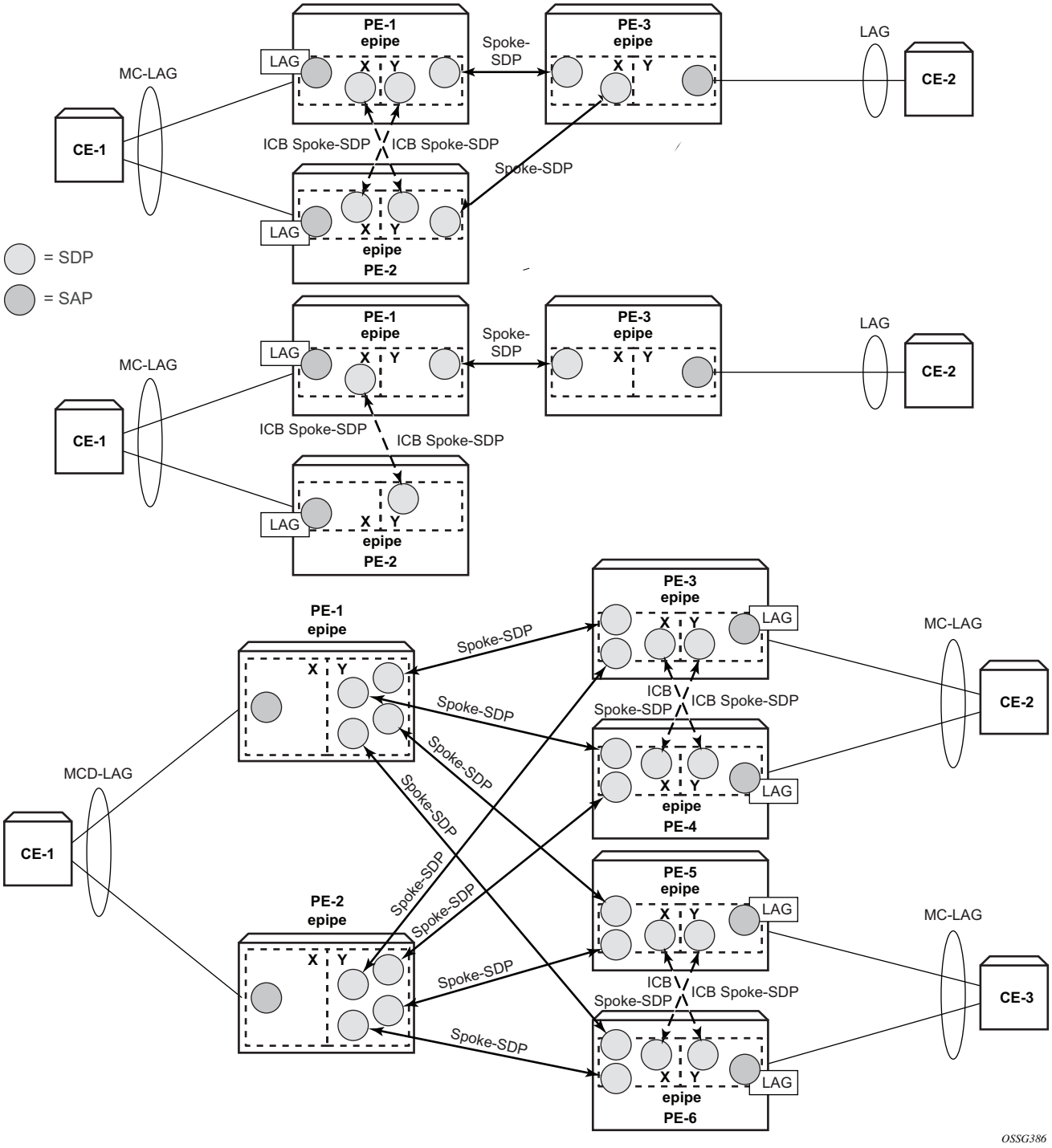


Figure 36 **Additional Setup Example 2**



MC-LAG in VPLS Services

MC-LAG can also be configured in VPLS services. When the MC-LAG converges, the PE that transitions to standby state for the MC-LAG will send out an LDP address withdrawal message to all peers configured in the VPLS service. Both types of SDPs (spoke and mesh) support this feature. The PE peers will then flush all the MAC addresses learned through the PE that sent the LDP MAC address withdrawal message.

Because a VPLS service is a multipoint service, pseudowire redundancy is not required. The MC-LAG redundancy configuration is identical.

Forced Switchover

MC-LAG convergence can be forced with **tools perform lag** command:

```
*A:PE-1# tools perform lag force
  - force all-mc {active|standby}
  - force lag-id <lag-id> [sub-group <sub-group-id>] {active|standby}
  - force peer-mc <ip-address> {active|standby}

 <lag-id>             : [1..800]
 <sub-group-id>       : [1..16]
 <all-mc>             : keyword
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <active|standby>     : keywords

*A:PE-1# tools perform lag force lag-id 1 standby

*A:PE-1# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      down    No       0         0        standby
===============================================================================
*A:PE-1#
```

After the forced switchover, it is important to clear the forced switchover:

```
*A:PE-1# tools perform lag clear-force
  - clear-force all-mc
  - clear-force lag-id <lag-id> [sub-group <sub-group-id>]
  - clear-force peer-mc <ip-address>
```

```
 <lag-id>             : [1..800]
 <sub-group-id>       : [1..16]
 <all-mc>             : keyword
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D

*A:PE-1# tools perform lag clear-force lag-id 1
```

Conclusion

MC-LAG is a Nokia added value redundancy feature that offers fast access link convergence in Epipe and VPLS services for CE devices that support standard LACP. PE node convergence for VPLS services is enhanced by using LDP address withdrawal messages to flush the FDB on the PE peers. PE node convergence for Epipes is guaranteed by using pseudowire redundancy.

Router Configuration

In This Section

This section provides configuration information for the following topics:

- Aggregate Route Indirect Next-Hop Option
- Bi-Directional Forwarding Detection
- LFA Policies Using OSPF as IGP

Aggregate Route Indirect Next-Hop Option

This chapter provides information about aggregate route indirect next-hop option configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

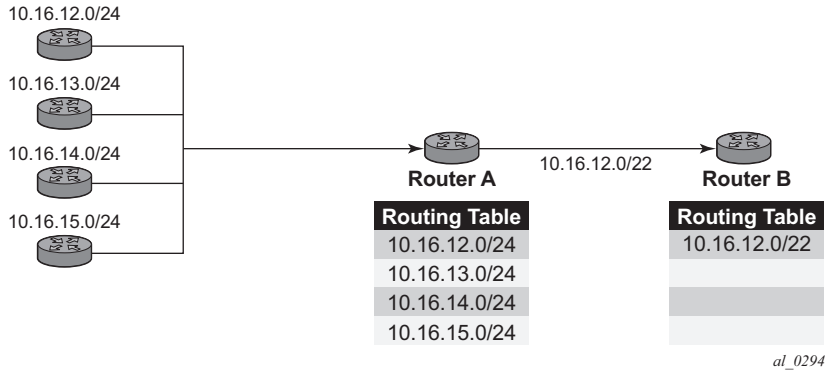
Applicability

This chapter was initially written for SR OS release 11.0.R1. The CLI in this edition corresponds to release 14.0.R2.

Overview

The 7x50s have for many releases supported the ability to configure IPv4 and IPv6 aggregate routes. A configured aggregate route that has the best preference for the prefix is added to the routing table (activated) when it has at least one contributing route and removed when there are no longer any more contributing routes. A contributing route is any route installed in the forwarding table that is a more-specific match of the aggregate. (10.16.12.0/24 is a contributing route to the aggregate route 10.16.12.0/22, but for this same aggregate 10.16.0.0/16 and 10.0.0.0/8 are not contributing routes).

Figure 37 **Aggregate Routes**



In Figure 37, Router A could choose to advertise all the four routes or one aggregate route. By aggregating the four routes, fewer updates are sent on the link between routers A and B, router B needs to maintain a smaller routing table resulting in better convergence and router B saves on computational resources by evaluating fewer entries in its routing table.

Release 11.0.R1 added the ability to configure an indirect hop for aggregate routes. The indirect next hop specifies where packets will be forwarded if they match the aggregate route, but not a more specific route in the IP forwarding table.

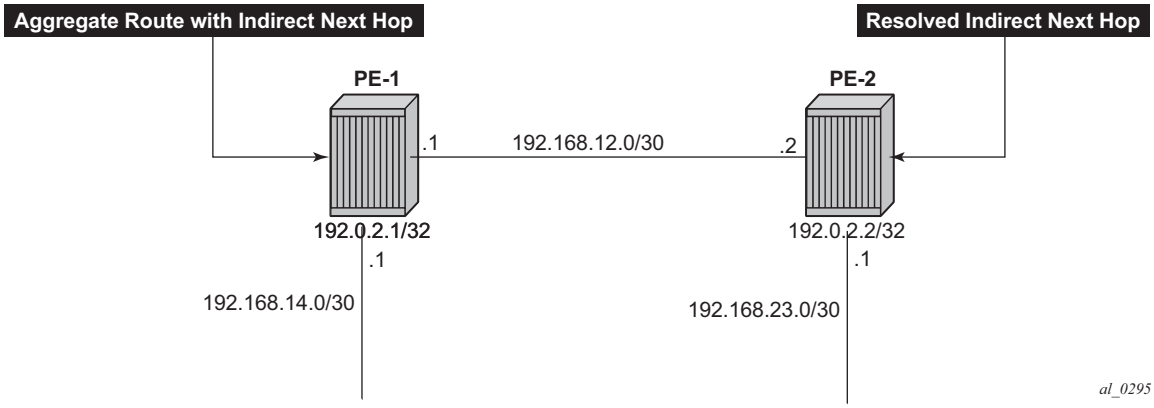
Different network operators have different requirements on how to forward a packet that matches an aggregate route but not any of the more-specific routes in the forwarding table that activated the aggregate. In general, there are three different options:

a. The packet can be forwarded according to the next-most specific route, ignoring the aggregate route. This can lead to routing loops in some topologies.
b. The packet can be discarded.
c. The packet can be forwarded toward an indirect next-hop address that is configured by the operator. The indirect next-hop could be the address of a threat management server that analyzes the packets it receives for security threats. This option requires the aggregate route to be installed in the forwarding table with a resolved next-hop interface determined from a route lookup of the indirect next-hop address.

Configuration

The example topology is shown in Figure 38.

Figure 38 **Example topology**



Initial Configuration

The nodes have the following basic configuration:

- Cards, MDAs
- Ports
- Router interfaces

The router interfaces on PE-1 are configured as follows:

```
*A:PE-1# configure router
      interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/1
        exit
        interface "int-PE-1-PE-4"
            address 192.168.14.1/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.1/32
        exit
```

The configuration on the other nodes is similar. The IP addresses are shown in Figure 38. In this example, static routes will be configured. There is no need for an IGP, but could be configured.

Aggregate Route with Indirect Next Hop Option

This feature adds a keyword **indirect** and an associated IP address parameter to the aggregate command in these configuration contexts:

- **config>router**
- **config>service>vprn**

The aggregate route configuration commands are as follows:

```
configure [router | vprn <vprn-id> ] aggregate
  - no aggregate <ip-prefix/ip-prefix-length>
  - aggregate <ip-prefix/ip-prefix-length> [summary-only] [as-set]
    [aggregator <as-number:ip-address>] [black-hole [generate-icmp]]
    [community <comm-id>][description <description>]
  - aggregate <ip-prefix/ip-prefix-length> [summary-only] [as-set]
    [aggregator <as-number:ip-address>] [community <comm-id>]
    [indirect <ip-address>][description <description>]

 <ip-prefix/ip-pref*> : ipv4-prefix    - a.b.c.d (host bits must be 0)
                        ipv4-prefix-le - [0..32]
                        ipv6-prefix    - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
                        ipv6-prefix-le - [0..128]
 <summary-only>       : keyword
 <as-set>             : keyword
 <as-number:ip-addr*> : as-number      - [1..4294967295]
                        ip-address     - a.b.c.d
 <black-hole>         : keyword
 <comm-id>            : [72 chars max]
                        <2 byte-asnumber:comm-val>|<well-known-comm>
                        2 byte-asnumber - [0..65535]
                        comm-val        - [0..65535]
                        well-known-comm - no-export|no-export-subconfed|
                                          no-advertise
 <ip-address>         : [64 chars max]
 <generate-icmp>      : keyword
 <description-string> : [80 chars max]
```

Parameters:

- **indirect** — This indicates that the aggregate route has an indirect address. The indirect option is mutually exclusive with the black-hole option. To change the next-hop type of an aggregate route (for example from black-hole to indirect) the route must be deleted and then re-added with the new next-hop type (however other configuration attributes can generally be changed dynamically).

- <ip-address> — Installing an aggregate route with an indirect next-hop is supported for both IPv4 and IPv6 prefixes. However if the aggregate prefix is IPv6 the indirect next-hop must be an IPv6 address and if the aggregate prefix is IPv4 the indirect next-hop must be an IPv4 address.

An indirect next-hop address of an aggregate route may be resolved by any of the following route types:

- Direct/local route
- Static route with regular next-hop, black-hole next-hop or an indirect next-hop
- OSPFv2 or RIP IPv4 route (applicable only to IPv4 aggregate routes)
- LDP shortcut route (applicable only to IPv4 aggregate routes])
- OSPFv2 or IS-IS shortcut route (IPv4 route with an LDP/RSVP or RSVP tunnel next-hop) (applicable only to IPv4 aggregate routes)
- OSPFv3 or IS-IS route
- BGP route resolved by an IGP route
- BGP route resolved by a BGP route
- BGP labeled-IPv4 route resolved by an LDP or RSVP tunnel (applicable only to IPv4 aggregate routes
- 6PE route resolved by an LDP tunnel or static route with black-hole next-hop (applicable only to IPv6 aggregate routes)
- BGP-VPN route resolved by a BGP labeled-IPv4 route, LDP tunnel, or RSVP tunnel (applicable only to aggregate routes configured in a VPRN context)

If an indirect next-hop is not resolved, the aggregate route will show up as black-hole.

The actual configuration of the aggregate route 10.16.12.0/22 is as follows:

```
*A:PE-1# configure router aggregate 10.16.12.0/22 community 64496:64498 indirect
192.168.11.11
```

This creates an aggregate route, but there are no contributing routes that are more specific defined yet. Therefore the aggregate route remains inactive.

```
*A:PE-1# show router aggregate
===============================================================================
Legend: G - generate-icmp enabled
===============================================================================
Aggregates (Router: Base)
===============================================================================
Prefix                                          Aggr IP-Address   Aggr AS
   Summary                                         AS Set          State
     NextHop                                         Community     NextHopType
-------------------------------------------------------------------------------
10.16.12.0/22                                   0.0.0.0           0
   False                                           False           Inactive
    192.168.11.11                                   64496:64498       Indirect
```

```
-------------------------------------------------------------------------------
No. of Aggregates: 1
===============================================================================
*A:PE-1#
```

The inactive aggregate route does not appear in the routing table:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     00h01m57s   0
       system                                                          0
192.168.12.0/30                                 Local   Local     00h01m52s   0
       int-PE-1-PE-2                                                   0
192.168.14.0/30                                 Local   Local     00h01m52s   0
       int-PE-1-PE-4                                                   0
-------------------------------------------------------------------------------
No. of Routes: 3
```

Configure Contributing Routes

The aggregate route remains inactive as long as there is no contributing route which is more specific than the aggregate route. The following contributing routes are statically configured on PE-1:

```
*A:PE-1# configure router
        static-route-entry 10.16.12.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
        static-route-entry 10.16.13.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
        static-route-entry 10.16.14.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
        static-route-entry 10.16.15.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
```

As a result, the aggregate route becomes active:

```
*A:PE-1# show router aggregate
===============================================================================
Legend: G - generate-icmp enabled
===============================================================================
Aggregates (Router: Base)
===============================================================================
Prefix                                        Aggr IP-Address   Aggr AS
   Summary                                       AS Set          State
     NextHop                                       Community     NextHopType
-------------------------------------------------------------------------------
10.16.12.0/22                                 0.0.0.0           0
   False                                         False           Active
    192.168.11.11                                 64496:64498       Indirect
-------------------------------------------------------------------------------
No. of Aggregates: 1
===============================================================================
*A:PE-1#
```

The active aggregate route is added to the route table, as well as the contributing routes:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.16.12.0/22                                 Blackh* Aggr      00h00m00s  130
       Black Hole                                                 0
10.16.12.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
10.16.13.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
10.16.14.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
10.16.15.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
192.0.2.1/32                                  Local   Local     00h04m10s  0
       system                                                     0
192.168.12.0/30                               Local   Local     00h04m05s  0
       int-PE-1-PE-2                                              0
192.168.14.0/30                               Local   Local     00h04m05s  0
       int-PE-1-PE-4                                              0
-------------------------------------------------------------------------------
No. of Routes: 8
```

The aggregate route is black-holed since the next hop is not resolved. There is no route to 192.168.11.0/24.

Configure Route to Indirect Next Hop

A static route is configured on PE-1to the indirect next hop, as follows:

```
*A:PE-1# configure router
        static-route-entry 192.168.11.0/24
            next-hop 192.168.12.2
                no shutdown
            exit
        exit
```

In the route table, the aggregate route is no longer black-holed. The next hop for the indirect next hop is 192.168.12.2 (PE-2).

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.16.12.0/22                                   Remote  Aggr      00h02m05s  130
       192.168.12.2                                                    0
10.16.12.0/24                                   Remote  Static    00h02m05s  5
       192.168.14.2                                                    1
10.16.13.0/24                                   Remote  Static    00h02m05s  5
       192.168.14.2                                                    1
10.16.14.0/24                                   Remote  Static    00h02m05s  5
       192.168.14.2                                                    1
10.16.15.0/24                                   Remote  Static    00h02m05s  5
       192.168.14.2                                                    1
192.0.2.1/32                                    Local   Local     00h06m15s  0
       system                                                          0
192.168.11.0/24                                 Remote  Static    00h02m05s  5
       192.168.12.2                                                    1
192.168.12.0/30                                 Local   Local     00h06m10s  0
       int-PE-1-PE-2                                                   0
192.168.14.0/30                                 Local   Local     00h06m10s  0
       int-PE-1-PE-4                                                   0
-------------------------------------------------------------------------------
No. of Routes: 9
```

In this example, PE-2 is the resolved indirected next hop and it has a route for prefix 10.16.12.0/22:

```
*A:PE-2# configure router
        static-route-entry 10.16.12.0/22
            next-hop 192.168.23.2
                no shutdown
            exit
        exit
```

The route table on PE-2 looks as follows:

```
*A:PE-2# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
```

```
-------------------------------------------------------------------------------
10.16.12.0/22                                 Remote  Static    00h00m00s  5
       192.168.23.2                                                    1
192.0.2.2/32                                  Local   Local     00h13m44s  0
       system                                                          0
192.168.12.0/30                               Local   Local     00h13m44s  0
       int-PE-2-PE-1                                                   0
192.168.23.0/30                               Local   Local     00h13m44s  0
       int-PE-2-PE-3                                                   0
-------------------------------------------------------------------------------
No. of Routes: 4
```

Conclusion

Aggregate routes offer several advantages, the key being reduction in the routing table size and overcoming routing loops, among other things. Aggregate routes with indirect next hop option helps in faster network convergence by decreasing the number of route table changes. This example shows how to configure aggregate routes with indirect next hop option.

Bi-Directional Forwarding Detection

This chapter provides information about bi-directional forwarding (BFD) detection.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to the 7x50 series. BFD timing differences among platforms will be indicated. Note that the centralized cpm-np type is only supported by 7750/7450s equipped with SF/CPM 2 or higher.

This chapter was originally written for and tested on release 8.0.R4. The CLI now corresponds to release 13.0.R6.

Overview

Bi-Directional Forwarding Detection (BFD) is a light-weight protocol which provides rapid path failure detection between two systems. It has been published as a series of RFCs (RFC 5880, *Bidirectional Forwarding Detection (BFD)*, to RFC 5884, *Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)*.

If a system running BFD stops receiving BFD messages on an interface, it will determine that there has been a failure in the path and notifies other protocols associated with the interface. BFD is useful in situations where two nodes are interconnected through either an optical Dense Wavelength Division Multiplexing (DWDM) or Ethernet network. In both cases, the physical network has numerous extra devices which are not part of the Layer 3 network and therefore, the Layer 3 nodes are incapable of detecting failures which occur in the physical network on spans to which the Layer 3 devices are not directly connected.

BFD protocol provides rapid link continuity checking between network devices, and the state of BFD can be propagated to IP routing protocols to drastically reduce convergence time in cases where a physical network error occurs in a transport network.

RFC 5880 defines two modes of operation for BFD:

- Asynchronous mode (supported) — Uses periodic BFD control messages to test the path between systems
- Demand mode (not supported)

In addition to the two operational modes, an echo function is defined (SR OS routers only supports response sending, this is looping back received BFD messages to the original sender).

BFD is running between two peers and supported for the following scenarios:

- BFD for ISIS
- BFD for OSPF
- BFD for PIM
- BFD for Static route
- BFD for RSVP

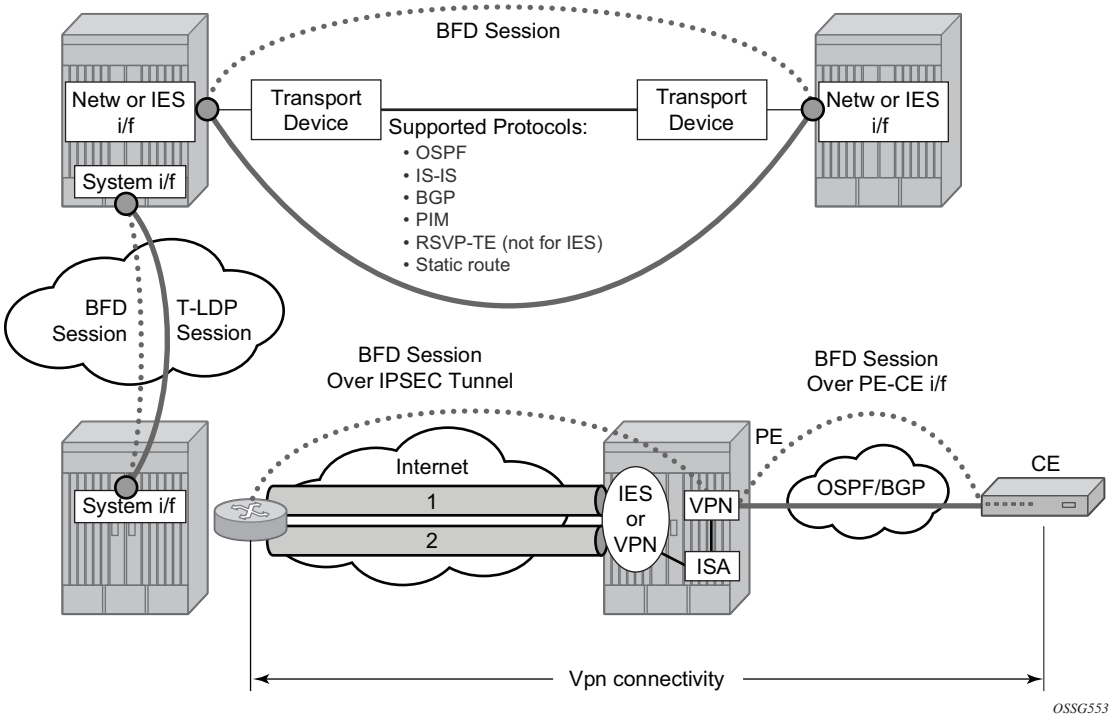
  BFD for I-LDP
- BFD for T-LDP

  BFD for MPLS-TP
- BFD for OSPF CE-PE adjacencies
- BFD for IPSec
- BFD for VRRP

  BFD for SRRP

The scenarios shown in Figure 39 are described in this chapter.

Figure 39 **BFD Multi-Scenarios**



Configuration

BFD packets are processed both locally (processed on the IOM CPU) and centrally (processed on the CPM).

Starting with release 8, the CPM is able to centrally generate the BFD packets at a sub second interval as low as 10 msec. However it should be noted that the BFD state machine is still implemented in software. It is the BFD packet generation that can be now selectively delegated to CPM hardware as needed. This is applicable when sub second operations or exceeding the IOM scaling limits is required.

Centralized sessions are processed:

- in software by 7x50 SR-1 and ESS-1, and 7x50 equipped with SF/CPM 1.
- in hardware by 7x50 equipped with SF/CPM 2 or higher.

Minimum transmitting and receiving intervals are as follows:

- Centralized sessions:

- Minimum 300 ms in 7x50 SR-1 and ESS-1
- Minimum 100 ms in 7x50 equipped with SF/CPM 1 and in every 7x50 up to Release 7.0
- Minimum 10 ms in 7x50 equipped with SF/CPM 2 or higher

• Local sessions:

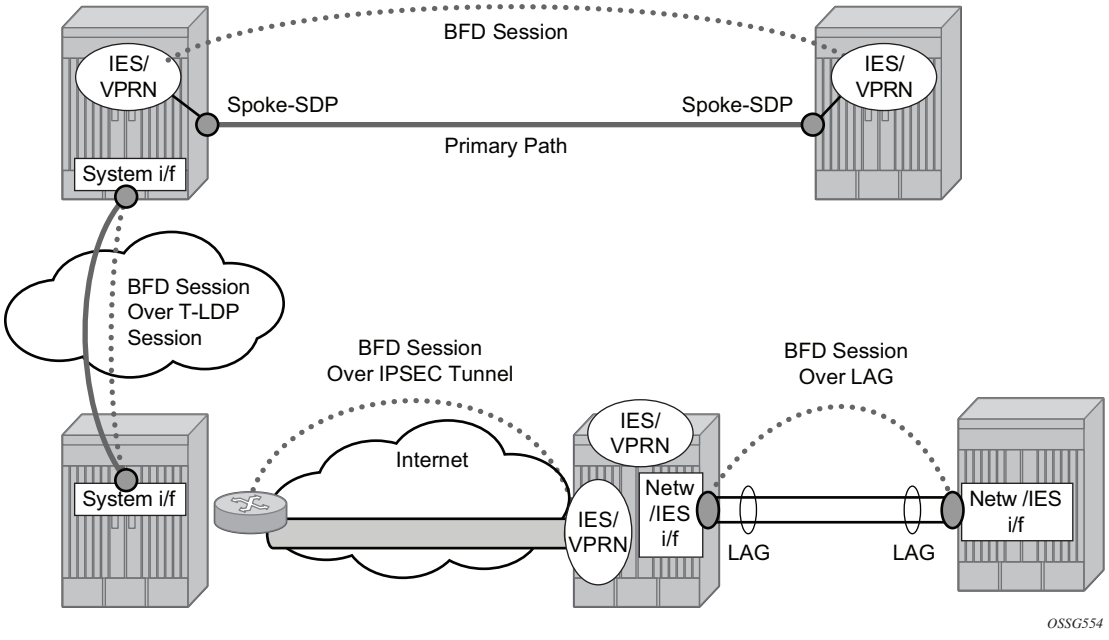
- Minimum 100 ms

The following applications require BFD to run centrally on the SF/CPM and a centralized session will be created independently of the type explicitly declared by the user:

- BFD for IES/VPRN over Spoke SDP
- BFD for LAG and VSM Interfaces
- Protocol associations using loopback and system interfaces (e.g. BFD for T-LDP)
- BFD for IPSec sessions
- BFD sessions associated with multi-hop peering (BGP)

Figure 40 shows the most relevant scenarios where centralized BFDsessions are used.

Figure 40 **BFD Centralized Sessions**



On the other end, when the two peers are directly connected, the BFD session is local by default, but in a 7x50 equipped with SF/CPM 2 or higher, the user can choose what session type (local or centralized) to implement.

As general rule, the following steps are required to configure and enable a BFD session when peers are directly connected:

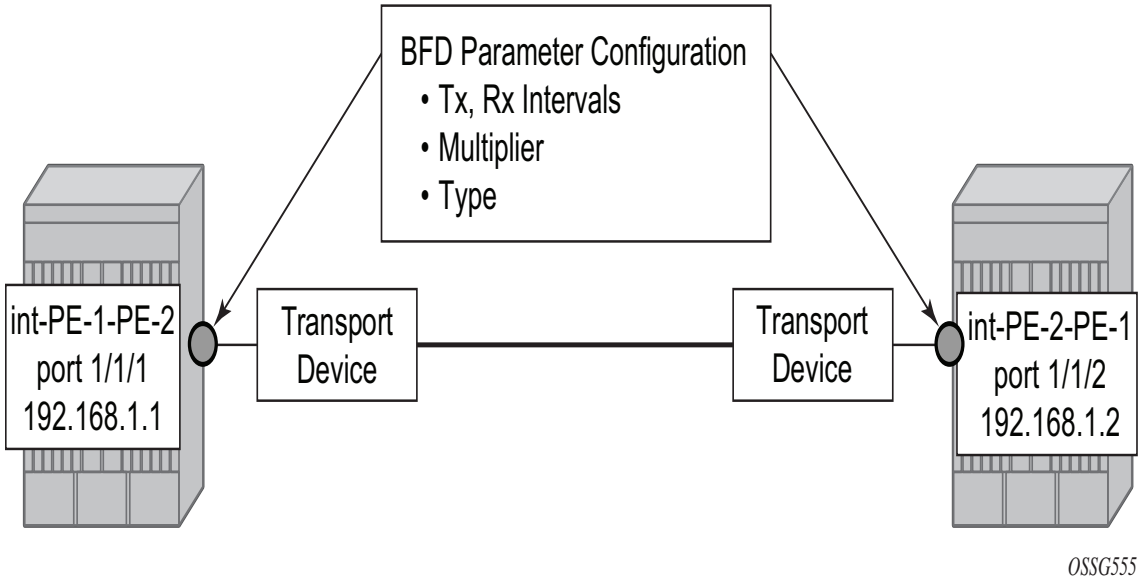
1. Configure BFD parameters on the peering interfaces.
2. Check that the Layer 3 protocol, that is to be bound to BFD, is up and running.
3. Enable BFD under the Layer 3 protocol interface.

Since most of the following procedures share the same first step, it is described only once in the next paragraph and then referred to in the following paragraphs.

BFD Base Parameter Configuration and Troubleshooting

The reference topology for the generic configuration of BFD over two local peers is shown in Figure 41.

Figure 41 **BFD Interface Configuration**



To configure BFD between two peers, the user should firstly enable base level BFD on interfaces between PE-1 and PE-2.

On PE1:

```
configure
    router
        interface "int-PE-1-PE-2"
            address 192.168.1.1/30
            port 1/1/1
            bfd 100 receive 100 multiplier 3
            no shutdown
        exit
    exit
exit
```

On PE2:

```
configure
    router
        interface "int-PE-2-PE-1"
            address 192.168.1.2/30
            port 1/1/2
            bfd 100 receive 100 multiplier 3
            no shutdown
        exit
    exit
exit
```

The following **show** commands are used to verify the BFD configuration on the router interfaces on PE1 and PE2.

On PE1:

```
*A:PE-1# show router bfd interface
===============================================================================
BFD Interface
===============================================================================
Interface name                      Tx Interval    Rx Interval    Multiplier
-------------------------------------------------------------------------------
int-PE-1-PE-2                       100            100            3
-------------------------------------------------------------------------------
No. of BFD Interfaces: 1
===============================================================================
*A:PE-1#
```

On PE2:

```
*A:PE-2# show router bfd interface
===============================================================================
BFD Interface
===============================================================================
Interface name                      Tx Interval    Rx Interval    Multiplier
-------------------------------------------------------------------------------
int-PE-2-PE-1                       100            100            3
-------------------------------------------------------------------------------
No. of BFD Interfaces: 1
===============================================================================
```

```
*A:PE-2#
```

Note that, BFD being an asynchronous protocol, it is possible to configure different tx and rx intervals on the two peers. This is because BFD rx/tx interval values are signaled in the BFD packets while establishing the BFD session.

In 7x50s equipped with SF/CPM 2 or higher, configurable BFD parameters are as follows:

```
bfd <transmit-interval> [receive <receive-interval>] [multiplier <multiplier>]
                                  [echo-receive <echo-interval>] [type <cpm-np>]
no bfd

 <transmit-interval>  : [10..100000] in milliseconds
 <receive-interval>   : [10..100000] in milliseconds
 <multiplier>         : [3..20]
 <echo-interval>      : [100..100000] in milliseconds
 <cpm-np>             : keyword - use CPM network processor
```

Note that it is possible to force the BFD session to be centrally managed by the CPM hardware.

As regards the echo function, it is possible to set the minimum echo receive interval, in milliseconds, for the BFD session. The default value is 100 ms.

If a BFD session is running, it is possible to modify its parameters, but to change its type, the session must be previously shut down manually. Note that this causes the upper layer protocols bound to it to be brought down as well.

```
configure
    router
        interface "int-PE-1-PE-2"
            bfd 100 receive 100 multiplier 3
        exit
    exit
exit
```

Forcing a centralized session in the case of directly connected peers can be useful when:

- Lower Tx and Rx intervals are desired (up to 10 ms instead of 100 ms supported by local sessions)
- There are no more available local (IOM) sessions
- Max limit of 500 packet per second per IOM has been reached

The instructions illustrated in following paragraphs are required to complete the configuration and enable BFD.

The BFD session should come up. To verify it, execute a **show router bfd session** command (bound to OSPF in the following example).

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                 Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols             Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-1-PE-2                 Up                    100       100       3
   192.168.1.2                ospf2                 114       96        iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

If the command gives a negative output, troubleshoot it by firstly checking that the protocol that is bound to it is up: for instance, check the OSPF neighbor adjacency as shown in following example.

```
*A:PE-1# show router ospf neighbor
===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-PE-2                    192.0.2.2       Full       1    0       34
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
*A:PE-1#
```

Then check whether a BFD resource limit has been reached (maximum number of local/centralized sessions or maximum number of packet per second per IOM).

If the overloaded limit is the maximum supported number of sessions, the cause is shown in log-id 99 (maxSessionsPerSlot).

In this case, when one of the running sessions is manually removed or goes down, then the additional configured session will come up. If the IOM limit is reached, it is possible to bring up the session by changing the session type to centralized.

To check if IOM CPU is able to start more local BFD sessions, execute a **show router bfd session summary** command.

If the **show router bfd session** command reports that the BFD session is down, then check the BFD peer's configuration and state.

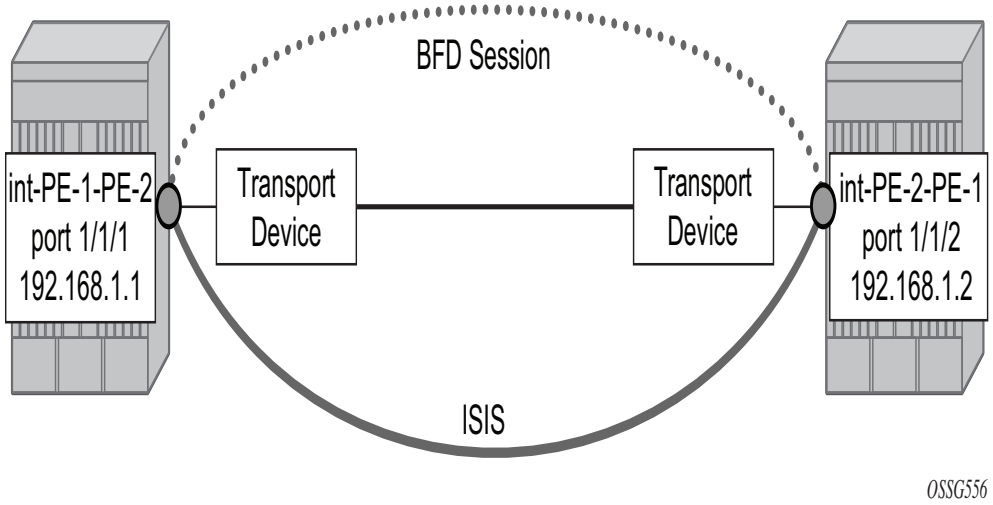
The **show router bfd session src <*ip-address*> detail** command can help debugging the BFD session. The sent and received counters are not supported for cpm-np type sessions.

```
*A:PE-1# show router bfd session src 192.168.1.1 detail
===============================================================================
BFD Session
===============================================================================
Remote Address : 192.168.1.2
Admin State    : Up                        Oper State       : Up
Protocols      : ospf2
Rx Interval    : 100                       Tx Interval      : 100
Multiplier     : 3                         Echo Interval    : 0
Recd Msgs      : 996                       Sent Msgs        : 1031
Up Time        : 0d 00:00:07               Up Transitions   : 2
Down Time      : None                      Down Transitions : 1
                                           Version Mismatch : 0
Forwarding Information
Local Discr    : 1                         Local State      : Up
Local Diag     : 0 (None)                  Local Mode       : Async
Local Min Tx   : 100                       Local Mult       : 3
Last Sent      : 12/09/2015 10:19:58       Local Min Rx     : 100
Type           : iom
Remote Discr   : 2                         Remote State     : Up
Remote Diag    : 0 (None)                  Remote Mode      : Async
Remote Min Tx  : 100                       Remote Mult      : 3
Last Recv      : 12/09/2015 10:19:58       Remote Min Rx    : 100
===============================================================================
===============================================================================
*A:PE-1#
```

BFD for IS-IS

The goal of this section is to configure BFD on a network interlink between two 7750 nodes that are IS-IS peers. The topology used is shown in Figure 42.

Figure 42 **BFD for ISIS**



For the base BFD configuration, please refer to BFD Base Parameter Configuration and Troubleshooting.

Apply BFD to the ISIS Interfaces.

On PE1:

```
configure
    router
        isis
            interface "int-PE-1-PE-2"
                bfd-enable ipv4
            exit
        exit
    exit
exit
```

On PE2:

```
configure
    router
        isis
            interface "int-PE-2-PE-1"
                bfd-enable ipv4
            exit
        exit
    exit
exit
```

Finally, verify that the BFD session is operational between PE1 and PE2.

On PE1:

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-1-PE-2                 Up                   100       100       3
   192.168.1.2                isis                 144       143       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

On PE2:

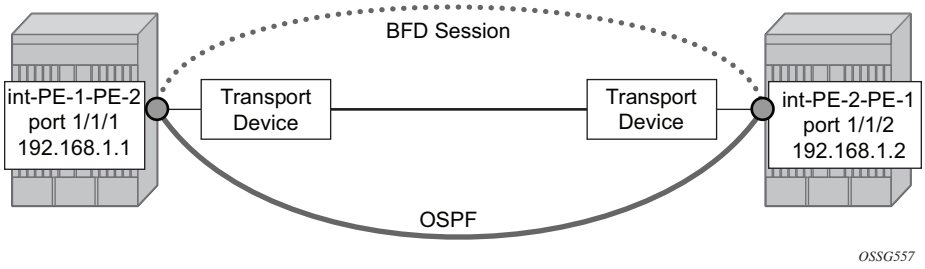
```
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-2-PE-1                 Up                   100       100       3
   192.168.1.1                isis                 159       159       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

BFD for OSPF

The goal of this section is to configure BFD on a network interlink between two 7750 nodes that are OSPF peers.

For this scenario, the topology is shown in Figure 43.

Figure 43 **BFD for OSPF**



For the base BFD configuration, refer to BFD Base Parameter Configuration and Troubleshooting.

Apply BFD on the OSPF Interfaces.

On PE1:

```
configure
    router
        ospf
            area 0
                interface "int-PE-1-PE-2"
                    bfd-enable
                exit
            exit
        exit
    exit
exit
```

On PE2:

```
configure
    router
        ospf
            area 0
                interface "int-PE-2-PE-1"
                    bfd-enable
                exit
            exit
        exit
    exit
exit
```

Verify that the BFD session is operational between PE1 and PE2.

On PE1:

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
```

```
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-1-PE-2                 Up                   100       100       3
   192.168.1.2                ospf2                174       175       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

On PE2:

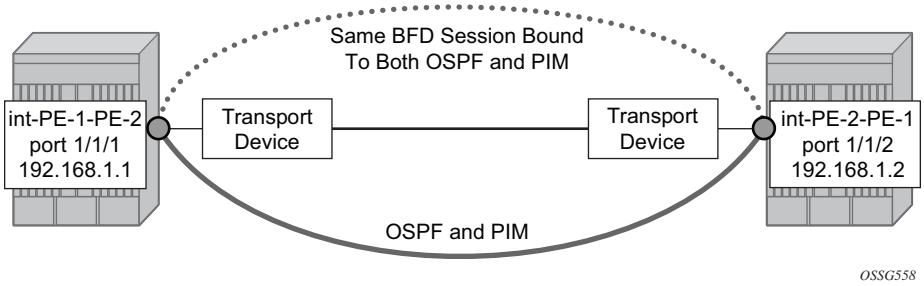
```
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-2-PE-1                 Up                   100       100       3
   192.168.1.1                ospf2                136       136       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

BFD for PIM

Since the implementation of PIM uses an Interior Gateway Protocol (IGP) in order to determine its Reverse Path Forwarding (RPF) tree, BFD configuration to support PIM will require BFD configuration of both the IGP protocol and the PIM protocol. Let's assume that IGP protocol is OSPF and that the starting configuration is as described in the previous section.

In this paragraph, configure and enable BFD for PIM on the same interfaces that were previously configured with BFD for OSPF, in reference to the topology shown in Figure 44.

Figure 44 **BFD for OSPF and PIM**



Since BFD has been already configured on the router interfaces, let's start by applying BFD on the PIM Interface.

On PE1:

```
configure
    router
        pim
            interface "int-PE-1-PE-2"
                bfd-enable ipv4
            exit
        exit
    exit
exit
```

On PE2:

```
configure
    router
        pim
            interface "int-PE-2-PE-1"
                bfd-enable ipv4
            exit
        exit
    exit
exit
```

The final step is to verify whether the BFD Session is operational between PE1 and PE2 for PIM.

On PE1:

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                 Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols             Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
```

```
-------------------------------------------------------------------------------
int-PE-1-PE-2                 Up                    100       100       3
   192.168.1.2                ospf2 pim             614       615       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

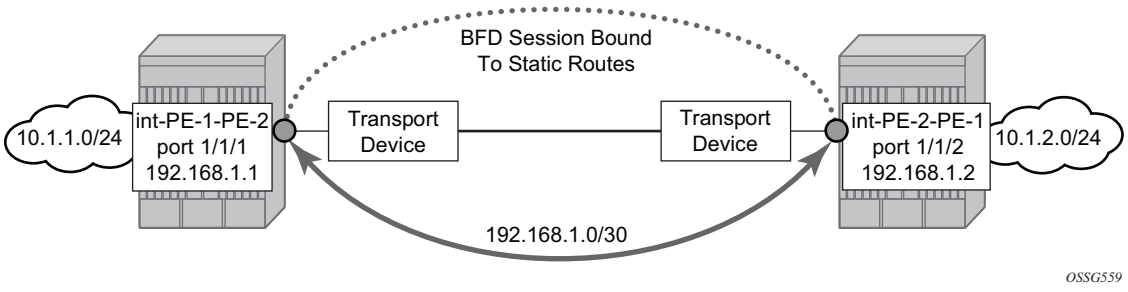
On PE2:

```
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                 Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols             Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-2-PE-1                 Up                    100       100       3
   192.168.1.1                ospf2 pim             617       617       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

BFD for Static Routes

The following procedures will go through the necessary steps to configure the base level BFD configuration and then apply BFD to the static routes between PE1 and PE2, using the topology shown in Figure 45.

Figure 45 **BFD for Static Routes**



First, create the static routes for the remote networks both in PE-1 and PE-2.

On PE1:

```
configure
```

```
    router
        static-route 10.1.2.0/24 next-hop 192.168.1.2
    exit
exit
```

On PE2:

```
configure
    router
        static-route 10.1.1.0/24 next-hop 192.168.1.1
    exit
exit
```

Next, verify that static routes are populated in the routing table.

On PE1:

```
*A:PE-1# show router route-table protocol static
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.1.2.0/24                                   Remote  Static    00h00m15s  5
       192.168.1.2                                                1
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

On PE2:

```
*A:PE-2# show router route-table protocol static
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.1.1.0/24                                   Remote  Static    00h00m14s  5
       192.168.1.1                                                1
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

The next step is to configure the base level BFD on PE1 and PE2.

Refer to paragraph BFD Base Parameter Configuration and Troubleshooting.

Then apply BFD to the static routing entries using the BFD interfaces as next-hop.

On PE1:

```
configure
    router
        static-route 10.1.2.0/24 next-hop 192.168.1.2 bfd-enable
    exit
exit
```

On PE2:

```
configure
    router
        static-route 10.1.1.0/24 next-hop 192.168.1.1 bfd-enable
    exit
exit
```

Note that BFD cannot be enabled if the next hop is indirect or the **blackhole** keyword is specified.

Finally, show the BFD session status.

On PE1:

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-1-PE-2                 Up                   100       100       3
   192.168.1.2                static               257       258       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

On PE2:

```
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
```

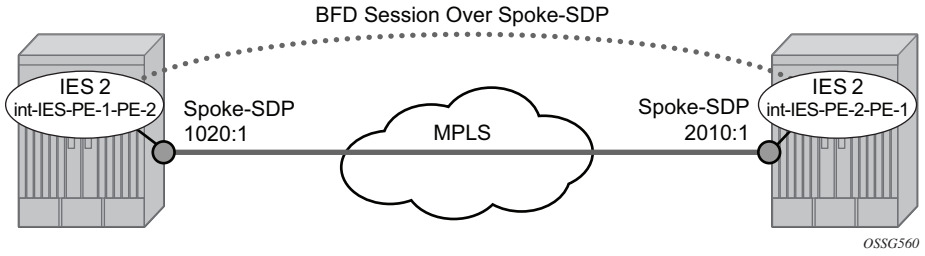
```
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-2-PE-1                 Up                   100       100       3
   192.168.1.1                static               243       243       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

BFD for IES

The goal of this section is to configure BFD for one IES service over a spoke SDP.

The IES service is configured in both 7750 nodes, PE1 and PE2, and their interfaces are connected by spoke SDP's. The topology is shown in Figure 46.

Figure 46 **BFD for IES over Spoke SDP**



Note that in this scenario BFD is run between the IES interfaces independent of the SDP/LSP paths.

The first step is to configure the IES service on both nodes.

On PE-1:

```
configure
    service
        ies 2 customer 1 create
            interface int-IES-PE-1-PE-2 create
                address 192.168.3.1/30
                spoke-sdp 1020:1 create
                exit
            exit
            no shutdown
        exit
    exit
exit
```

On PE-2:

```
configure
    service
        ies 2 customer 1 create
            interface int-IES-PE-2-PE-1 create
                address 192.168.3.2/30
                spoke-sdp 2010:1 create
                exit
            exit
            no shutdown
        exit
    exit
exit
```

The next step is to add the IES interfaces to the OSPF area domain.

On PE-1:

```
configure
    router
        ospf
            area 0
                interface "int-IES-PE-1-PE-2"
                exit
            exit
        exit
    exit
exit
```

On PE-2:

```
configure
    router
        ospf
            area 0
                interface "int-IES-PE-2-PE-1"
                exit
            exit
        exit
    exit
exit
```

Then verify that OSPF and the services are up using show commands on both routers.

On PE-1:

```
*A:PE-1# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
```

```
Service Type      : IES
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 12/09/2015 10:25:21
Last Mgmt Change  : 12/09/2015 10:25:08
Admin State       : Up                  Oper State        : Up
SAP Count         : 0                   SDP Bind Count    : 1

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:1020:1 S(192.0.2.2)                 Spok         0       1556    Up   Up
===============================================================================
*A:PE-1#
*A:PE-1# show router ospf neighbor
===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-PE-2                    192.0.2.2       Full       1    0       35
   0.0.0.0
int-IES-PE-1-PE-2                192.0.2.2       Full       1    0       31
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 2
===============================================================================
*A:PE-1#
```

On PE-2:

```
*A:PE-2# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : IES
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 12/09/2015 10:25:22
Last Mgmt Change  : 12/09/2015 10:25:22
Admin State       : Up                  Oper State        : Up
SAP Count         : 0                   SDP Bind Count    : 1

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:2010:1 S(192.0.2.1)                 Spok         0       1556    Up   Up
===============================================================================
*A:PE-2#
```

```
*A:PE-2# show router ospf neighbor
===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-2-PE-1                    192.0.2.1       Full       1    0       32
   0.0.0.0
int-IES-PE-2-PE-1                192.0.2.1       Full       1    0       38
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 2
===============================================================================
*A:PE-2#
```

Then configure BFD on the IES interfaces.

On PE-1:

```
configure
    service
        ies 2
            interface "int-IES-PE-1-PE-2"
                bfd 100 receive 100 multiplier 3
            exit
        exit
    exit
exit
```

On PE-2:

```
configure
    service
        ies 2
            interface "int-IES-PE-2-PE-1"
                bfd 100 receive 100 multiplier 3
            exit
        exit
    exit
exit
```

Finally, enable BFD on the interfaces under OSPF area 0.

On PE-1:

```
*A:PE-1# configure router ospf area 0.0.0.0 interface int-IES-PE-1-PE-2 bfd-enable
```

On PE-2:

```
*A:PE-2# configure router ospf area 0.0.0.0 interface int-IES-PE-2-PE-1 bfd-enable
```

Note that in case of BFD over spoke SDP, a centralized BFD session is created even if a physical link exists between the two nodes. In fact, the next output shows that BFD session type is cpm-np. This is because the spoke SDP is terminated at the CPM. This is also true for BFD running over LAG bundles.

The cpm-np type only exists in 7x50 SR/ESS systems equipped with SF/CPM 2 or higher. When other network elements run centralized BFD sessions like this one, the BFD type is shown as **central**.

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-IES-PE-1-PE-2             Up                   100       100       3
   192.168.3.2                ospf2                N/A       N/A       cpm-np
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-IES-PE-2-PE-1             Up                   100       100       3
   192.168.3.1                ospf2                N/A       N/A       cpm-np
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

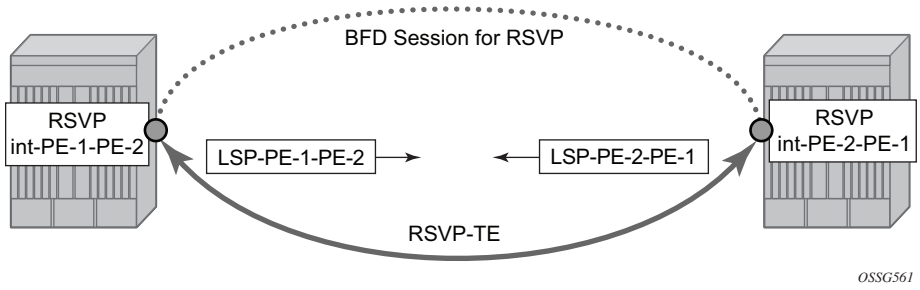
Note that in the case of centralized BFD sessions, transmitted and received packet counters are not shown.

BFD for RSVP

The goal of this section is to configure BFD between two RSVP interfaces configured in two 7750 nodes.

For this scenario, the topology is shown in Figure 47.

Figure 47 **BFD for RSVP**



To enable the BFD session between the two RSVP peers, the user should follow these steps:

First, configure BFD on interfaces between PE-1 and PE-2 as described in BFD Base Parameter Configuration and Troubleshooting.

Next, configure MPLS, creating the path, the LSP and the interfaces within MPLS (and RSVP).

On PE-1:

```
configure
    router
        mpls
            interface "system"
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                no shutdown
            exit
        exit
        rsvp
            interface "system"
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                no shutdown
            exit
            no shutdown
        exit
        mpls
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                cspf
                primary "dyn"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

```
    exit
exit
```

On PE-2:

```
configure
    router
        mpls
            interface "system"
                no shutdown
            exit
            interface "int-PE-2-PE-1"
                no shutdown
            exit
        exit
        rsvp
            interface "system"
                no shutdown
            exit
            interface "int-PE-2-PE-1"
                no shutdown
            exit
            no shutdown
        exit
        mpls
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-2-PE-1"
                to 192.0.2.1
                cspf
                primary "dyn"
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
exit
```

Next, verify that the RSVP session is up.

```
*A:PE-1# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.2       1      50176 LSP-PE-1-PE-2::dyn            Up
192.0.2.2       192.0.2.1       1      45568 LSP-PE-2-PE-1::dyn            Up
-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
*A:PE-1#
```

Then, apply BFD on the RSVP Interfaces.

On PE1:

```
configure
    router
        rsvp
            interface "int-PE-1-PE-2"
                bfd-enable
            exit
        exit
    exit
exit
```

On PE2:

```
configure
    router
        rsvp
            interface "int-PE-2-PE-1"
                bfd-enable
            exit
        exit
    exit
exit
```

Finally, verify that the BFD session is operational between PE1 and PE2.

On PE1:

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-1-PE-2                 Up                   100       100       3
   192.168.1.2                rsvp                 118       108       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

On PE2:

```
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
```

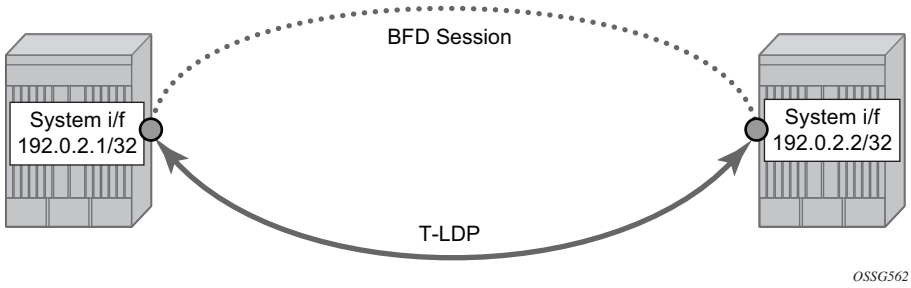
```
-------------------------------------------------------------------------------
int-PE-2-PE-1                  Up                    100       100       3
   192.168.1.1                 rsvp                  259       259       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

BFD for T-LDP

BFD tracking of an LDP session associated with a T-LDP adjacency allows for faster detection of the liveliness of the session by registering the transport address of an LDP session with a BFD session.

The goal of this paragraph is to configure BFD for T-LDP, referring to the scheme shown in Figure 48.

Figure 48 **BFD for T-LDP**



The parameters used for the BFD session are configured under the loopback interface corresponding to the LSR-ID (by default, the LSR-ID matches the system interface address).

```
configure
    router
        interface "system"
            bfd 300 receive 3000 multiplier 3
        exit
    exit
exit
```

By enabling BFD for a selected targeted session, the state of that session is tied to the state of the underlying BFD session between the two nodes.

When using BFD over other links with the ability to reroute, such as spoke-SDPs, the interval and multiplier values configuring BFD should be set to allow sufficient time for the underlying network to re-converge before the associated BFD session expires. A general rule of thumb should be that the expiration time (interval * multiplier) is three times the convergence time for the IGP network between the two endpoints of the BFD session.

Before enabling BFD, ensure that the T-LDP session is up.

On PE-1:

```
*A:PE-1# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Both      Established   676       680       0d 00:24:51
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 1
===============================================================================
===============================================================================
LDP IPv6 Sessions
===============================================================================
Peer LDP Id
 Adj Type            State          Msg Sent       Msg Recv       Up Time
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
```

On PE-2:

```
*A:PE-2# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.1:0         Both      Established   687       687       0d 00:25:10
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 1
===============================================================================
===============================================================================
LDP IPv6 Sessions
===============================================================================
Peer LDP Id
 Adj Type            State          Msg Sent       Msg Recv       Up Time
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-2#
```

Then, enable the BFD session.

```
configure
    router
        ldp
            targeted-session
                peer 192.0.2.2
                    bfd-enable
                exit
            exit
        exit
    exit
exit
```

Note that the loopback interface can be used to source BFD sessions to many peers in the network.

Finally, check that the BFD session is up.

On PE-1:

```
*A:PE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
system                        Up                   3000      3000      3
   192.0.2.2                  ldp                  N/A       N/A       cpm-np
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

On PE-2:

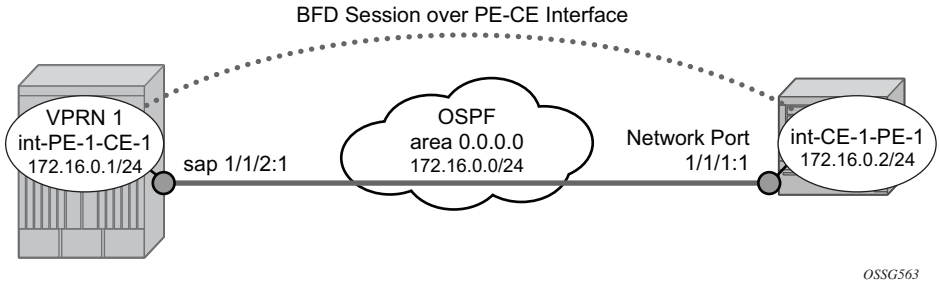
```
*A:PE-2# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
system                        Up                   3000      3000      3
   192.0.2.1                  ldp                  N/A       N/A       cpm-np
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-2#
```

When the T-LDP session comes up, a centralized BFD session is always created (cpm-np) even if the local interface has a direct link to the peer.

BFD for OSPF PE-CE Adjacencies

This feature extends BFD support to OSPF within a VPRN context when OSPF is used as the PE-CE protocol. In this section, the topology used is shown in Figure 49.

Figure 49 **BFD for OSPF PE-CE I/F**



First, configure the VPRN service interface int-PE-1-CE-1 on PE-1 with BFD parameters.

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 1:1
            vrf-target target:1:1
            interface "int-PE-1-CE-1" create
                address 172.16.0.1/24
                bfd 100 receive 100 multiplier 3
                sap 1/1/2:1 create
                exit
            exit
            ospf
                area 0.0.0.0
                    interface "int-PE-1-CE-1"
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
exit
```

Next, configure the router interface on CE-1 and add it to the OSPF area 0 domain.

```
configure
    router
        interface "int-CE-1-PE-1"
            address 172.16.0.2/24
            port 1/1/1:1
            bfd 100 receive 100 multiplier 3
            no shutdown
        exit
        ospf
            area 0
                interface int-CE-1-PE-1
                exit
            exit
        exit
    exit
exit
```

Then, ensure that OSPF adjacency is up.

On PE-1:

```
*A:PE-1# show router 1 ospf neighbor
===============================================================================
Rtr vprn1 OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-CE-1                    192.0.2.5       Full       1    0       37
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
*A:PE-1#
```

On CE-1:

```
*A:CE-1# show router ospf neighbor
===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-CE-1-PE-1                    192.0.2.1       Full       1    0       36
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
*A:CE-1#
```

Then, enable BFD on the PE-1-CE-1 interface on PE-1.

```
configure service vprn 1 ospf area 0 interface int-PE-1-CE-1 bfd-enable
```

Enable BFD on the CE-1-PE-1 interface on CE-1.

```
configure router ospf area 0 interface int-CE-1-PE-1 bfd-enable
```

Finally, check that the BFD sessions are up in both PE-1 and CE-1.

```
*A:PE-1# show router 1 bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-PE-1-CE-1                 Up                   100       100       3
   172.16.0.2                 ospf2                81        70        iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
*A:CE-1# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
int-CE-1-PE-1                 Up                   100       100       3
   172.16.0.1                 ospf2                161       162       iom
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:CE-1#
```

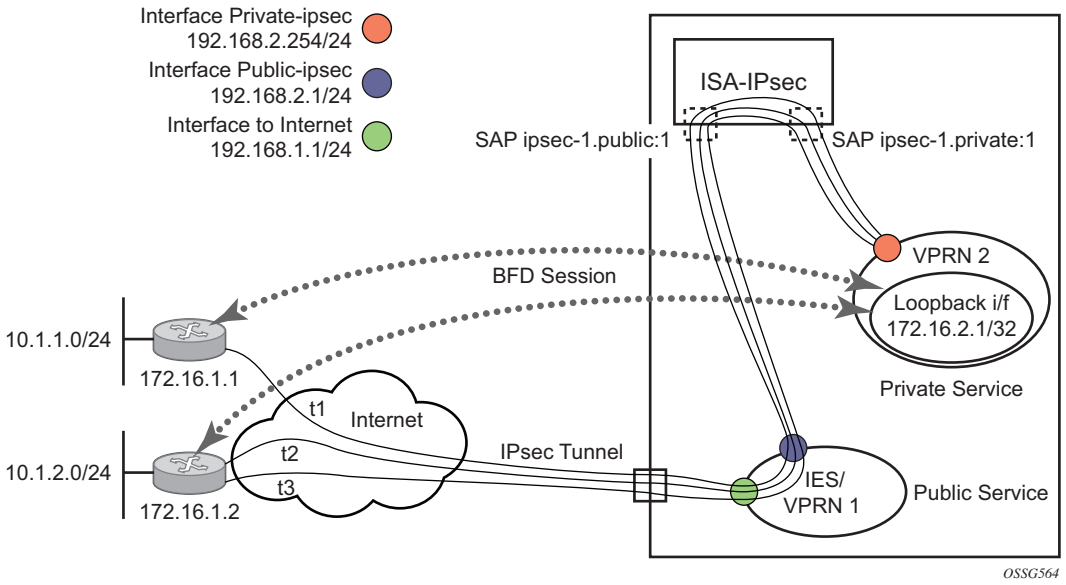
BFD within IPSec Tunnels

The ability to assign a BFD session to a given static LAN-to-LAN IPSec tunnel that provides heart-beat mechanism for fast failure detection has been introduced in Release.8.0.

IPSec needs a Multi-service Integrated Service Adapter (MS-ISA) installed, so this scenario is only applicable to 7750 SR-7/12 equipped with IOM-2 or 3.

The topology used is shown in Figure 50.

Figure 50 **BFD Sessions within IPSec Tunnels**



The first step is to configure MS-ISA card as **type isa-tunnel**.

```
configure
    card 1
        card-type iom3-xp
        mda 1
            mda-type m4-10gb-xp-xfp
        exit
        mda 2
            mda-type isa-tunnel
        exit
        no shutdown
    exit
exit
```

Next, instantiate the tunnels t1, t2 and t3 from the private service (in this example, VPRN 2) to the peers passing through the public service (in this example VPRN 1, but it could be instead an IES).

Since the configuration of IPSec tunnels is out of the scope of this section, only relevant command lines are reported to configure the interfaces shown in Figure 50.

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 1:1
            interface "toInternet" create
                address 192.168.1.1/24
                sap 1/1/2 create
                exit
            exit
```

```
            interface "public-ipsec" create
                address 192.168.2.1/24
                sap tunnel-1.public:1 create
                exit
            exit
            no shutdown
        exit
    exit
exit
```

```
configure
    service
        vprn 2 customer 1 create
            ipsec
                security-policy 1 create
                    entry 10 create
                        local-ip 192.168.3.1/32
                        remote-ip any
                    exit
                exit
            exit
            route-distinguisher 1:2
            interface "private-ipsec" tunnel create
                sap tunnel-1.private:1 create
                    ipsec-tunnel "t1" create
                        security-policy 1
                        local-gateway-address 192.168.2.254 peer 172.16.1.1
                                                            delivery-service 1
                        dynamic-keying
                            pre-shared-key "vpn1secret"
                        exit
                    exit
                    ipsec-tunnel "t2" create
                        security-policy 1
                        local-gateway-address 192.168.2.254 peer 172.16.1.2
                                                            delivery-service 1
                        dynamic-keying
                            pre-shared-key "vpn1secret"
                        exit
                    exit
                    ipsec-tunnel "t3" create
                        security-policy 1
                        local-gateway-address 192.168.2.254 peer 172.16.1.2
                                                            delivery-service 1
                        dynamic-keying
                            pre-shared-key "vpn1secret"
                        exit
                    exit
                exit
            exit
            interface "loop" create
                address 172.16.2.1/32
                loopback
            exit
            static-route 10.1.1.0/24 ipsec-tunnel "t1"
```

```
            static-route 10.1.2.0/24 ipsec-tunnel "t2" metric 1
            static-route 10.1.2.0/24 ipsec-tunnel "t3" metric 5
            no shutdown
        exit
    exit
exit
```

Then configure the BFD parameters within loopback interface loop (refer to BFD Base Parameter Configuration and Troubleshooting).

```
configure
    service
        vprn 2
            interface "loop"
                bfd 100 receive 100 multiplier 3
            exit
        exit
    exit
exit
```

And finally enable BFD within the tunnels.

```
configure
    service
        vprn 2
            interface "private-ipsec" tunnel create
                sap tunnel-1.private:1 create
                    ipsec-tunnel "t1" create
                        bfd-enable service 2 interface "loop" dst-ip 172.16.1.1
                    exit
                    ipsec-tunnel "t2" create
                        bfd-designate
                        bfd-enable service 2 interface "loop" dst-ip 172.16.1.2
                    exit
                    ipsec-tunnel "t3" create
                        bfd-enable service 2 interface "loop" dst-ip 172.16.1.2
                    exit
                exit
            exit
        exit
    exit
exit
```

The BFD-enable parameters are as follows:

- **service** <*service-id*> — Specifies the service-id where the BFD session resides.
- **interface** <*interface-name*> — Specifies the name of the interface used by the BFD session.
- **dst-ip** <*ip-address*> — Specifies the destination address to be used for the BFD session.

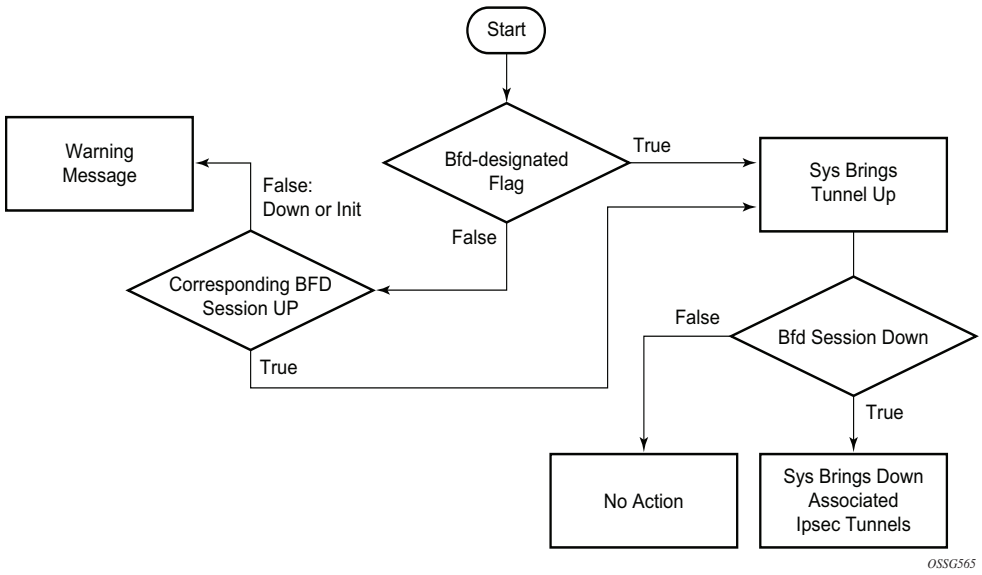
The following statements are to be taken into consideration to correctly configure BFD in this environment:

- BFD over IPSec sessions are centralized, managed by the hardware on the CPM.
- Only BFD over static lan-to-lan tunnel is supported in Release 8.0 (not dynamic).
- Only one BFD session is allowed between a given source/destination address pair.
- Each tunnel can be associated to only one BFD session but multiple tunnels can be associated to the same BFD session.
- In case multiple tunnels share the same BFD session, one IPSec tunnel carries BFD traffic: the BFD-DESIGNATED tunnel.

Referring to Figure 50 and to the above configuration, the tunnels t2 and t3 share the same BFD-session. Tunnel t2 is the bfd-designated tunnel, the BFD session runs within it and the other tunnel t3 shares its BFD session. If the BFD session goes down, the system will bring down both the designated tunnel t2 and the associated tunnel t3.

The state machine in Figure 51 shows the decision process in case of shared BFD sessions.

Figure 51 **Logic for Shared BFD Sessions**

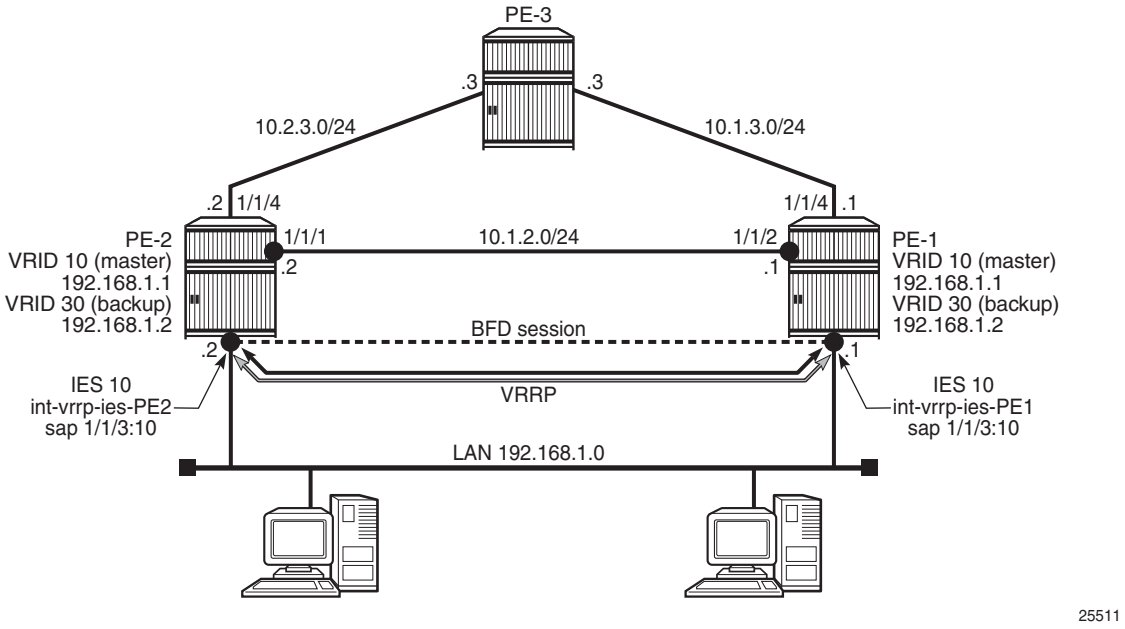


BFD for VRRP

This feature assigns a BFD session to provide a heart-beat mechanism for the given VRRP instance. It should be noted that there can be only one BFD session assigned to any given VRRP instance, but there can be multiple VRRP sessions using the same BFD session.

In this section, the topology is shown in Figure 52.

Figure 52 **BFD for VRRP**



First, create the LAN subnet. Two PE routers are connected by IES or VPRN services (in following examples IES 10 is created in both routers).

On PE-1:

```
configure
    service
        ies 10 customer 1 create
            interface "int-vrrp-ies-PE-1" create
                address 192.168.1.1/24
                sap 1/1/3:10 create
                exit
            exit
            no shutdown
        exit
    exit
exit
```

On PE-2:

```
configure
    service
        ies 10 customer 1 create
            interface "int-vrrp-ies-PE-2" create
                address 192.168.1.2/24
                sap 1/1/3:10 create
                exit
            exit
            no shutdown
        exit
    exit
exit
```

Verify that the IES services are operational (**show service service-using**) and verify that you can ping the remote interface IP address.

Next, configure the VRRP parameters for both PE-1 and PE-2, enable VRRP on the IES interface that connects to the 192.168.1.0/24 subnet.

In this section, the configurations are shown for the VRRP owner mode for master but any other scenario for VRRP can be configured (non owner mode for master).

In the following examples two VRRP instances are created on the 192.168.1.0/24 subnet:

```
     VRID = 10  Master (owner)  = PE-1
                Backup          = PE-2
                VRRP IP =  192.168.1.1
     VRID = 30  Master (owner)  = PE-2
                Backup          = PE-1
                VRRP IP =  192.168.1.2
```

Host 1 is configured with default gateway = 192.168.1.1, and host 2 is configured with default gateway = 192.168.1.2.

On PE-1:

```
configure service ies 10 customer 1 create
            interface "int-vrrp-ies-PE-1" create
                vrrp 10 owner
                    backup 192.168.1.1
                exit
                vrrp 30
                    backup 192.168.1.2
                    ping-reply
                    telnet-reply
                    ssh-reply
                exit
            exit
```

```
            no shutdown
        exit
    exit
exit
```

On PE-2:

```
configure service ies 10 customer 1 create
            interface "int-vrrp-ies-PE-2" create
                vrrp 10
                    backup 192.168.1.1
                    ping-reply
                    telnet-reply
                    ssh-reply
                exit
                vrrp 30 owner
                    backup 192.168.1.2
                exit
            exit
            no shutdown
        exit
    exit
exit
```

To bind the VRRP instances with a BFD session, add the following command under any VRRP instance: **bfd-enable** *service-id* **interface** *interface-name* **dst-ip** *ip-address*.

Note that the IES service-id must be declared where the interface is configured.

On PE-1:

```
configure service ies 10 customer 1 create
            interface "int-vrrp-ies-PE-1" create
                vrrp 10 owner
                    bfd-enable 10 interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                exit
                vrrp 30
                    bfd-enable 10 interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                exit
            exit
            no shutdown
        exit
    exit
exit
```

On PE-2:

```
configure service ies 10 customer 1 create
            interface "int-vrrp-ies-PE-2" create
                vrrp 10
                    bfd-enable 10 interface "int-vrrp-ies-PE-2" dst-ip 192.168.1.1
                exit
                vrrp 30 owner
                    bfd-enable 10 interface "int-vrrp-ies-PE-2" dst-ip 192.168.1.1
```

```
            exit
          exit
          no shutdown
        exit
    exit
exit
```

The parameters used for the BFD are set by the BFD command under the IP interface.

Note that unlike the previous scenarios, the user can enter the commands above, enabling the BFD session, even if the specified interface (vrrp_ies_PE1) has not been configured with BFD parameters.

If it has not been configured yet, the BFD session will be initiated only after the following configuration.

On PE-1:

```
configure service ies 10 interface "int-vrrp-ies-PE-1" bfd 1000
                                          receive 1000 multiplier 10
```

On PE-2:

```
configure service ies 10 interface "int-vrrp-ies-PE-2" bfd 1000
                                          receive 1000 multiplier 10
```

Finally, verify that the BFD session is up (for instance on PE-1):

```
*A:PE-1# show router bfd session src 192.168.1.1 detail
===============================================================================
BFD Session
===============================================================================
Remote Address : 192.168.1.2
Admin State    : Up                       Oper State       : Up
Protocols      : vrrp
Rx Interval    : 1000                     Tx Interval      : 1000
Multiplier     : 10                       Echo Interval    : 0
Recd Msgs      : 655                      Sent Msgs        : 668
Up Time        : 0d 00:00:11              Up Transitions   : 3
Down Time      : None                     Down Transitions : 2
                                          Version Mismatch : 0

Forwarding Information

Local Discr    : 1                        Local State      : Up
Local Diag     : 0 (None)                 Local Mode       : Async
Local Min Tx   : 1000                     Local Mult       : 10
Last Sent      : 12/09/2015 10:53:37      Local Min Rx     : 1000
Type           : iom
Remote Discr   : 5                        Remote State     : Up
Remote Diag    : 0 (None)                 Remote Mode      : Async
Remote Min Tx  : 1000                     Remote Mult      : 10
Last Recv      : 12/09/2015 10:53:37      Remote Min Rx    : 1000
```

```
===============================================================================
===============================================================================
*A:PE-1#
```

This session is shared by all the VRRP instances configured between the specified interfaces.

When BFD is configured in a VRRP instance, the following command gives details of BFD related to every instance:

```
*A:PE-1# show router vrrp instance interface "int-vrrp-ies-PE-1"
===============================================================================
VRRP Instances for interface "int-vrrp-ies-PE-1"
===============================================================================
-------------------------------------------------------------------------------
VRID 10
-------------------------------------------------------------------------------
Owner               : Yes                 VRRP State        : Master
Primary IP of Master: 192.168.1.1 (Self)
Primary IP          : 192.168.1.1         Standby-Forwarding: Disabled
VRRP Backup Addr    : 192.168.1.1
Admin State         : Up                  Oper State        : Up
Up Time             : 12/09/2015 10:42:33 Virt MAC Addr     : 00:00:5e:00:01:0a
Auth Type           : None
Config Mesg Intvl   : 1                   In-Use Mesg Intvl : 1
Base Priority       : 255                 In-Use Priority   : 255
Init Delay          : 0                   Init Timer Expires: 0.000 sec
Creation State      : Active

-------------------------------------------------------------------------------
BFD Interface
-------------------------------------------------------------------------------
Service ID          : 10
Interface Name      : int-vrrp-ies-PE-1
Src IP              : 192.168.1.1
Dst IP              : 192.168.1.2
Session Oper State  : connected

-------------------------------------------------------------------------------
Master Information
-------------------------------------------------------------------------------
Primary IP of Master: 192.168.1.1 (Self)
Addr List Mismatch  : No                  Master Priority   : 255
Master Since        : 12/09/2015 10:42:33

-------------------------------------------------------------------------------
Masters Seen (Last 32)
-------------------------------------------------------------------------------
Primary IP of Master   Last Seen             Addr List Mismatch     Msg Count
-------------------------------------------------------------------------------
192.168.1.1            12/09/2015 10:42:33   No                            0

-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
Become Master       : 1                   Master Changes    : 1
Adv Sent            : 683                 Adv Received      : 0
```

```
Pri Zero Pkts Sent  : 0                   Pri Zero Pkts Rcvd: 0
Preempt Events      : 0                   Preempted Events  : 0
Mesg Intvl Discards : 0                   Mesg Intvl Errors : 0
Addr List Discards  : 0                   Addr List Errors  : 0
Auth Type Mismatch  : 0                   Auth Failures     : 0
Invalid Auth Type   : 0                   Invalid Pkt Type  : 0
IP TTL Errors       : 0                   Pkt Length Errors : 0
Total Discards      : 0

-------------------------------------------------------------------------------
VRID 30
-------------------------------------------------------------------------------
Owner               : No                  VRRP State        : Backup
Primary IP of Master: 192.168.1.2 (Other)
Primary IP          : 192.168.1.1         Standby-Forwarding: Disabled
VRRP Backup Addr    : 192.168.1.2
Admin State         : Up                  Oper State        : Up
Up Time             : 12/09/2015 10:42:33 Virt MAC Addr     : 00:00:5e:00:01:1e
Auth Type           : None
Config Mesg Intvl   : 1                   In-Use Mesg Intvl : 1
Master Inherit Intvl: No
Base Priority       : 100                 In-Use Priority   : 100
Policy ID           : n/a                 Preempt Mode      : Yes
Ping Reply          : Yes                 Telnet Reply      : Yes
SSH Reply           : Yes                 Traceroute Reply  : No
Init Delay          : 0                   Init Timer Expires: 0.000 sec
Creation State      : Active

-------------------------------------------------------------------------------
BFD Interface
-------------------------------------------------------------------------------
Service ID          : 10
Interface Name      : int-vrrp-ies-PE-1
Src IP              : 192.168.1.1
Dst IP              : 192.168.1.2
Session Oper State  : connected

-------------------------------------------------------------------------------
Master Information
-------------------------------------------------------------------------------
Primary IP of Master: 192.168.1.2 (Other)
Addr List Mismatch  : No                  Master Priority   : 255
Master Since        : 12/09/2015 10:48:02
Master Down Interval: 3.609 sec (Expires in 2.950 sec)

-------------------------------------------------------------------------------
Masters Seen (Last 32)
-------------------------------------------------------------------------------
Primary IP of Master   Last Seen             Addr List Mismatch     Msg Count
-------------------------------------------------------------------------------
192.168.1.1            12/09/2015 10:48:02   No                            0
192.168.1.2            12/09/2015 10:53:54   No                          673

-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
Become Master       : 2                   Master Changes    : 4
Adv Sent            : 8                   Adv Received      : 673
Pri Zero Pkts Sent  : 0                   Pri Zero Pkts Rcvd: 1
```

```
Preempt Events      : 0                   Preempted Events  : 2
Mesg Intvl Discards : 0                   Mesg Intvl Errors : 0
Addr List Discards  : 0                   Addr List Errors  : 0
Auth Type Mismatch  : 0                   Auth Failures     : 0
Invalid Auth Type   : 0                   Invalid Pkt Type  : 0
IP TTL Errors       : 0                   Pkt Length Errors : 0
Total Discards      : 0

===============================================================================
*A:PE-1#
```

Finally, for troubleshooting: it could be that the BFD session between the two IP interfaces is up but (in one or both peers) the command **show router vrrp instance interface** *interface-name* gives the following output regarding BFD for one or more VRID's.

```
*A:PE-1# show router vrrp instance interface "int-vrrp-ies-PE-1"
---snipped ---
-------------------------------------------------------------------------------
BFD Interface
-------------------------------------------------------------------------------
Service ID          : None
Interface Name      : int-vrrp-ies-PE-1
Src IP              :
Dst IP              : 192.168.1.2
Session Oper State  : notConfigured

---snipped ---
*A:PE-1#
```

To fix this, check that BFD has been correctly configured for the VRRP istances.

For instance, in the following example, the cause of the misconfiguration is that the IES service-id is not declared in the bfd-enable command:

```
configure service ies 10
            interface int-vrrp-ies-PE-1
                vrrp 10
                    bfd-enable 10 interface int-vrrp-ies-PE-1 dst-ip 192.168.1.2
                exit
            exit
        exit
    exit
exit
```

Conclusion

BFD is a light-weight protocol which provides rapid path failure detection between two systems and it is useful in situations where the physical network has numerous intervening devices which are not part of the Layer 3 network.

BFD is linked to a protocol state. For BFD session to be established, the prerequisite condition is that the protocol to which the BFD is linked must be operationally active. Once the BFD session is established, the state of the protocol to which BFD is tied to is then determined based on the BFD session's state. This means that if the BFD session goes down, the corresponding protocol will be brought down.

In this section several scenarios where BFD could be implemented have been described, including the configuration, show output and troubleshooting hints.

LFA Policies Using OSPF as IGP

This chapter provides information about LFA policies using OSPF as IGP.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 12.0.R4, but the CLI in this edition corresponds to release 14.0.R2.

Overview

Loop Free Alternate (LFA) is a local control plane feature. When multiple LFAs exist, RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*, chooses the LFA providing the best coverage of the failure cases. In general, this means that node LFA has preference above link LFA. In some deployments, however, this can lead to suboptimal LFA. For example, an aggregation router (typically using lower bandwidth links) protecting a core node/link (typically using high bandwidth links) is potentially undesirable.

For this reason, the operator wants to have more control in the LFA next-hop selection algorithm. This is achieved by the introduction of LFA Shortest Path First (SPF) policies.

LFA policies can work in combination with IP fast reroute (FRR) and/or LDP FRR.

Implementation

The SROS LFA policy implementation is built around the concept of route-next-hop (NH) templates which are applied to IP interfaces. A route-next-hop template specifies criteria which influence the selection of an LFA backup NH for either:

- a set of prefixes in a prefix-list or
- a set of prefixes which resolve to a specific primary NH

See http://tools.ietf.org/html/draft-litkowski-rtgwg-lfa-manageability for further information. Two powerful methods which can be used as criteria inside a route-next-hop template are IP admin groups and IP shared risk link groups (SRLGs). IP SRLG and IP admin group criteria are applied before running the LFA NH algorithm. IP admin groups and IP SRLGs work in a similar way as the well-known MPLS admin groups and MPLS SRLGs.

For example, when one or more IP admin groups/SRLGs are applied to an IP interface, the same MPLS admin group/SRLG rules apply:

- IP interfaces which do not include one or more of the admin groups defined in the **include** statements are pruned before computing the LFA next-hop.
- IP interfaces which belong to admin groups which have been explicitly excluded using the **exclude** statement are pruned before computing the LFA next-hop.
- IP interfaces which belong to the SRLGs used by the primary NH of a prefix are pruned before computing the LFA next-hop.

For more information about MPLS admin groups, see chapter RSVP Point-to-Point LSPs; for SRLGs, see chapter Shared Risk Link Groups for RSVP-Based LSP.

For compatibility reasons with the existing MPLS, admin groups and SRLGs, a single set of admin groups and SRLGs are defined within the **configure router if-attribute** context .

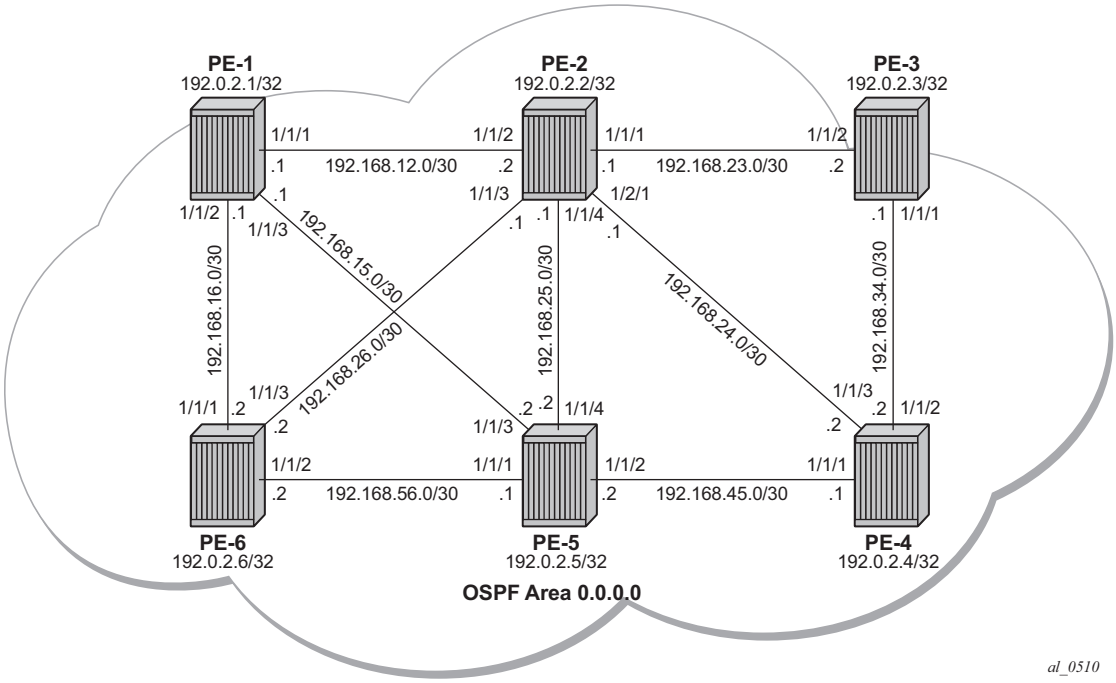
Once one or more admin groups or SRLGs have been defined, theyare applied to an MPLS interface and/or an IP interface.

In the SR OS implementation, IP admin groups and SRLGs are locally significant, meaning they are not advertised by the IGP.

The well-known MPLS admin groups and SRLGs are advertised in TE link TLVs and sub-TLVs when the traffic-engineering option is enabled in the IGP protocol.

Other selection criteria which can be configured inside a route-next-hop template are protection type preference and NH type preference. More details on these parameters are provided later in this chapter.

Figure 53 **Example Topology**



Configuration

Step 1. Configuring an IP/MPLS network with LDP FRR enabled on PE-7.

Since the focus is not on how to set up an IP/MPLS network, only summary bullets are provided.

- The system and IP interface addresses are configured according to Figure 53.
- OSPF area 0 is selected as the interior gateway protocol (IGP) to distribute routing information between all PEs. All OSPF interfaces are set up as **type point-to-point** to avoid running the designated router/ backup designated router (DR/BDR) election process. All links have the default metric, which is 10 in this example, except for interface “int-PE-2-PE-5” on PE-2, which is configured with a metric of 20.
- Enable link LDP on all interfaces. This establishes a full mesh of LDP LSPs between all PE system interfaces. As an example, the tunnel-table on PE-2 looks as follows:

```
*A:PE-2# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
```

```
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65537     9        192.168.12.1   10
192.0.2.3/32      ldp       MPLS  65538     9        192.168.23.2   10
192.0.2.4/32      ldp       MPLS  65539     9        192.168.24.2   10
192.0.2.5/32      ldp       MPLS  65540     9        192.168.12.1   20
192.0.2.6/32      ldp       MPLS  65541     9        192.168.26.2   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-2#
```

The LDP LSP metric follows the IGP cost.

– Enable LDP FRR on PE-2. This is a two-fold configuration command: first the IGP needs to be triggered to do LFA NH computation, and secondly, FRR needs to be enabled within the LDP context. This is configured as follows:

```
*A:PE-2# configure router ospf loopfree-alternate
*A:PE-2# show router ospf status | match LFA
LFA                          : Enabled
Remote-LFA                   : Disabled
```

Remote LFA can be enabled for segment routing, but this is beyond the scope of this chapter.

```
*A:PE-2# configure router ldp fast-reroute
*A:PE-2# show router ldp status | match FRR
FRR                : Enabled             Mcast Upstream FRR   : Disabled
```

Multicast upstream FRR is for multicast LDP and is beyond the scope of this chapter.

After issuing these two CLI commands, the software pre-computes both a primary and a backup Next-hop Label Forwarding Entry (NHLFE) for each LDP FEC in the network and downloads them into the IOM/IMM. The primary NHLFE corresponds to the label of the FEC received from the primary NH as per standard LDP resolution of the FEC prefix in the Routing Table Manager (RTM). The backup NHLFE corresponds to the label received for the same FEC from an LFA NH. The **show router route-table alternative** command adds an LFA flag to the associated alternative NH for a specific destination prefix. Other useful IGP related show commands are **show router ospf lfa-coverage** and **show router ospf routes alternative detail**.

```
*A:PE-2# show router route-table alternative

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
```

```
      Alt-NextHop                                                  Alt-
                                                                  Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                 Remote  OSPF      00h11m32s  10
       192.168.12.1                                                 10
       192.168.26.2 (LFA)                                           20
192.0.2.2/32                                 Local   Local     00h11m44s  0
       system                                                       0
192.0.2.3/32                                 Remote  OSPF      00h11m18s  10
       192.168.23.2                                                 10
       192.168.24.2 (LFA)                                           20
192.0.2.4/32                                 Remote  OSPF      00h11m12s  10
       192.168.24.2                                                 10
       192.168.23.2 (LFA)                                           20
192.0.2.5/32                                 Remote  OSPF      00h11m02s  10
       192.168.12.1                                                 20
       192.168.24.2 (LFA)                                           20
192.0.2.6/32                                 Remote  OSPF      00h10m54s  10
       192.168.26.2                                                 10
       192.168.12.1 (LFA)                                           20
192.168.12.0/30                              Local   Local     00h11m44s  0
       int-PE-2-PE-1                                                0
192.168.15.0/30                              Remote  OSPF      00h11m32s  10
       192.168.12.1                                                 20
       192.168.26.2 (LFA)                                           30
192.168.16.0/30                              Remote  OSPF      00h11m32s  10
       192.168.12.1                                                 20
       192.168.26.2 (LFA)                                           30
192.168.23.0/30                              Local   Local     00h11m44s  0
       int-PE-2-PE-3                                                0
192.168.24.0/30                              Local   Local     00h11m44s  0
       int-PE-2-PE-4                                                0
192.168.25.0/30                              Local   Local     00h11m44s  0
       int-PE-2-PE-5                                                0
192.168.26.0/30                              Local   Local     00h11m44s  0
       int-PE-2-PE-6                                                0
192.168.34.0/30                              Remote  OSPF      00h11m18s  10
       192.168.23.2                                                 20
       192.168.24.2 (LFA)                                           30
192.168.45.0/30                              Remote  OSPF      00h11m12s  10
       192.168.24.2                                                 20
       192.168.23.2 (LFA)                                           30
192.168.56.0/30                              Remote  OSPF      00h10m54s  10
       192.168.26.2                                                 20
       192.168.12.1 (LFA)                                           30
-------------------------------------------------------------------------------
No. of Routes: 16
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

Displaying the Label Forwarding Information Base (LFIB) on PE-2 shows the available alternate NHs; displayed with the BU flag.

```
*A:PE-2# show router ldp bindings active prefixes ipv4
```

```
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Push              --      262142BU
192.168.26.2                            1/1/3

192.0.2.1/32                            Swap            262142    262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Swap            262142    262142BU
192.168.26.2                            1/1/3

192.0.2.2/32                            Pop             262143      --
  --                                      --

192.0.2.3/32                            Push              --      262143
192.168.23.2                            1/1/1

192.0.2.3/32                            Push              --      262140BU
192.168.24.2                            1/2/1

192.0.2.3/32                            Swap            262141    262143
192.168.23.2                            1/1/1

192.0.2.3/32                            Swap            262141    262140BU
192.168.24.2                            1/2/1

192.0.2.4/32                            Push              --      262143
192.168.24.2                            1/2/1

192.0.2.4/32                            Push              --      262140BU
192.168.23.2                            1/1/1

192.0.2.4/32                            Swap            262140    262143
192.168.24.2                            1/2/1

192.0.2.4/32                            Swap            262140    262140BU
192.168.23.2                            1/1/1

192.0.2.5/32                            Push              --      262139
192.168.12.1                            1/1/2
```

```
192.0.2.5/32                                    Push              --      262139BU
192.168.24.2                                    1/2/1

192.0.2.5/32                                    Swap           262139     262139
192.168.12.1                                    1/1/2

192.0.2.5/32                                    Swap           262139     262139BU
192.168.24.2                                    1/2/1

192.0.2.6/32                                    Push              --      262143
192.168.26.2                                    1/1/3

192.0.2.6/32                                    Push              --      262138BU
192.168.12.1                                    1/1/2

192.0.2.6/32                                    Swap           262138     262143
192.168.26.2                                    1/1/3

192.0.2.6/32                                    Swap           262138     262138BU
192.168.12.1                                    1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 21
===============================================================================
*A:PE-2#
```

Finally, a synchronization timer is enabled between the IGP and LDP protocol when LDP FRR is enabled. From the moment that the interface for the previous primary NH is restored, the IGP may re-converge back to that interface before LDP has completed the FEC exchange with its neighbor over that interface. This may cause LDP to de-program the LFA NH from the FEC and blackhole the traffic. In this example a timer of 10 seconds is used. This is configured as follows:

```
*A:PE-x# configure router interface <itf-name> ldp-sync-timer 10
```

When this timer is set, on restoring a failed interface, the IGP advertises this link into the network with an infinite metric for the duration of this timer. When the failed link is restored, the **ldp-sync-timer** is started, and LDP adjacencies are brought up over the restored link and a label exchange is completed between the peers. After the **ldp-sync-timer expires**, the normal metric is advertised into the network again.

At this point, everything is in place to start creating LFA policies to influence the calculated LFA NHs.

Step 2. Create a route-next-hop policy template.

This is a mandatory step in the context of LFA policies. The route-next-hop template name is 32 characters at maximum. Creating a route-next-hop policy is done in the following way:

```
*A:PE-x# configure router route-next-hop-policy template <template name>
```

Commands within a **route-next-hop** policy template follow the **begin-abort-commit** model. After a **commit**, the IGP re-evaluates the template and schedules a new LFA SPF to re-compute the LFA NH for the prefixes associated with this template.

Step 3. Configure admin group constraints in route-next-hop policy.

This is an optional step in the context of LFA policies. Firstly, configure a group-name and a group-value for each admin group locally on the router. This is configured as follows:

```
*A:PE-x# configure router if-attribute admin-group <group-name> value <group-value>
```

Second, configure the admin group membership of the IP interface(s) (network, IES or VPRN). Up to five admin groups can be applied to an IP interface in one command but the command can be applied multiple times. The configured IP admin group membership applies to all levels/areas the interface is participating in, as follows:

```
*A:PE-x# configure router interface <itf-name> if-attribute admin-group <group-name>
[ <group-name> ... (upto 5 max)]
*A:PE-x# configure service vprn <svc-id> interface <itf-name> if-attribute admin-group
<group-name> [ <group-name> ... (upto 5 max)]
*A:PE-x# configure service ies <svc-id> interface <itf-name> if-attribute admin-group
<group-name> [ <group-name> ... (upto 5 max)]
```

Third, add the IP admin group constraints to the route-next-hop policy template one by one. The **include-group** statement instructs the LFA SPF selection algorithm to select a subset of LFA NHs among the links which belong to one or more of the specified admin groups. A link which does not belong to at least one of the admin groups is excluded. The **pref** option is used to provide a relative preference for the admin group selection. A lower preference value means that LFA SPF will first attempt to select an LFA backup NH which is a member of the corresponding admin group. If none is found, then the admin group with the next higher preference value is evaluated. If no preference is configured, then it is the least preferred (default preference value is 255).

When evaluating multiple **include-group** statements having the same preference, any link which belongs to one or more of the included admin groups can be selected as an LFA next-hop. There is no relative preference based on how many of those included admin groups the link is a member.

The **exclude-group** command simply prunes all links belonging to the specified admin group before making the LFA backup NH selection for a prefix. If the same group name is part of both **include** and **exclude** statements, the exclude statement will takes precedence. In other words, the **exclude** statement can be viewed as having an implicit preference value of 0.

This is configured as follows:

```
*A:PE-x# configure router route-next-hop-policy template <template-name> exclude-
group <group-name>
*A:PE-x# configure router route-next-hop-policy template <template-name> include-
group <group-name> [pref <preference>]
```

Step 4. Configure SRLG constraints in route-next-hop policy.

This is an optional step in the context of LFA policies. Firstly, configure a group-name and group-value, of each SRLG group locally on the router. This is configured as follows:

```
*A:PE-x# configure router if-attribute srlg-group <group-name> value <group-value>
[penalty-weight <penalty-weight>]
```

Second, configure the SRLG group membership of the IP interfaces (network, IES or VPRN). Up to five SRLG groups can be applied to an IP interface in one command but the command can be applied multiple times. The configured IP SRLG group membership is applied in all levels/areas the interface is participating in. This is configured as follows:

```
*A:PE-x# configure router interface <itf-name> if-attribute srlg-group <group-name> [
<group-name> ... (upto 5 max)]
*A:PE-x# configure service vprn <svc-id> interface <itf-name> if-attribute srlg-group
<group-name> [ <group-name> ... (upto 5 max)]
*A:PE-x# configure service ies <svc-id> interface <itf-name> if-attribute srlg-group
<group-name> [ <group-name> ... (upto 5 max)]
```

Third, add IP SRLG group constraints to the route-next-hop policy template. When this command is applied to a prefix, the LFA SPF attempts to select an LFA NH which uses an outgoing interface that does not participate in any of the SRLGs of the outgoing interface used by the primary NH. This is configured as follows:

```
*A:PE-x# configure router route-next-hop-policy template <template-name> srlg-enable
```

Step 5. Configure the protection type in route-next-hop policy.

This is an optional step in the context of LFA policies. With the use of LFA policies, the user can also select if link protection or node protection is preferred for IP prefixes and LDP FEC prefixes protected by a backup LFA NH. By default, node protection is chosen. The implementation falls back to link protection if no LFA NH is found for node protection. This is configured as follows:

```
*A:PE-x# configure router route-next-hop-policy template <template-name> protection-
type {link|node}
```

Step 6. Configure the NH preference type in route-next-hop policy.

This is an optional step in the context of LFA policies. With the use of LFA policies, the user can also select if tunnel backup NH or IP backup NH is preferred for IP prefixes and LDP FEC prefixes protected by a backup LFA NH. By default, IP backup NH is chosen. The implementation falls back to the other type (tunnel) if no LFA NH of the preferred type is found. This is configured as follows:

```
*A:PE-x# configure router route-next-hop-policy template <template-name> nh-type
{ip|tunnel}
```

Step 7. Apply the route-next-hop policy template to an IP interface.

When the route-next-hop policy is applied to an IP interface, all prefixes using this interface as primary NH take the selection criteria specified in Step 3, Step 4, Step 5 and Step 6 into account. This is configured as follows:

```
*A:PE-x# configure router ospf area interface lfa-policy-map route-nh-template
<template-name>
*A:PE-x# configure router ospf3 area interface lfa-policy-map route-nh-template
<template-name>
*A:PE-x# configure service vprn ospf area interface lfa-policy-map route-nh-template
<template-name>
*A:PE-x# configure service vprn ospf3 area interface lfa-policy-map route-nh-template
<template-name>
```

LFA Policy Examples

All of the examples focus on providing another LFA NH for LDP FEC prefix 192.0.2.1/32 and 192.0.2.6/32 (the system IP addresses of PE-1 and PE-6), with PE-2 being the Point of Local Repair (PLR).

See Example Topology for the example topology.

The default LFA NH (without policy) for LDP FEC prefix192.0.2.1/32 is as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
```

```
===============================================================================
Prefix                                    Op             IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                              Push             --      262143
192.168.12.1                              1/1/2

192.0.2.1/32                              Push             --      262142BU
192.168.26.2                              1/1/3

192.0.2.1/32                              Swap           262142    262143
192.168.12.1                              1/1/2

192.0.2.1/32                              Swap           262142    262142BU
192.168.26.2                              1/1/3

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-2#
```

The default LFA NH forLDP FEC prefix 192.0.2.6/32 is as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op             IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                              Push             --      262143
192.168.26.2                              1/1/3

192.0.2.6/32                              Push             --      262138BU
192.168.12.1                              1/1/2

192.0.2.6/32                              Swap           262138    262143
192.168.26.2                              1/1/3

192.0.2.6/32                              Swap           262138    262138BU
192.168.12.1                              1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-2#
```

This default LFA NH can be changed by adding specific selection criteria inside a route-next-hop policy template.

Example 1: LFA Policy with Admin Group Constraint

The objective is to force the LFA NH for both LDP FEC prefixes to use the path between PE-2 and PE-5.

Define admin group ‘red’ with value ‘1’ and apply it to the IP interfaces PE-2 to PE-1 and PE-2 to PE-6.

```
*A:PE-2# configure router if-attribute admin-group "red" value 1
*A:PE-2# configure router interface "int-PE-2-PE-1" if-attribute admin-group "red"
*A:PE-2# configure router interface "int-PE-2-PE-6" if-attribute admin-group "red"
```

Define a route-next-hop policy template ‘LFA_NH_exclRed’, which excludes IP admin group ‘red’.

```
*A:PE-2# configure
    router
        route-next-hop-policy
            begin
            template "LFA_NH_exclRed"
                exclude-group "red"
            exit
            commit
```

The policy is applied ot the OSPF interfaces toward PE-1 and PE-6, as follows:

```
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-1" lfa-policy-map route-
nh-template "LFA_NH_exclRed"
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-6" lfa-policy-map route-
nh-template "LFA_NH_exclRed"
```

From the moment that the route-next-hop policy template ‘LFA_NH_exclRed’ is applied to the OSPF interfaces toward PE-1 and PE-6, the LFA NHs for both LDP FEC prefixes change. They now both point to the IP interface from PE-2 to PE-5 as LFA backup NH:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
```

```
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Push              --      262142BU
192.168.25.2                            1/1/4

192.0.2.1/32                            Swap            262142    262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Swap            262142    262142BU
192.168.25.2                            1/1/4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-2#

*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push              --      262143
192.168.26.2                            1/1/3

192.0.2.6/32                            Push              --      262138BU
192.168.25.2                            1/1/4

192.0.2.6/32                            Swap            262138    262143
192.168.26.2                            1/1/3

192.0.2.6/32                            Swap            262138    262138BU
192.168.25.2                            1/1/4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
```

```
*A:PE-2#
```

Example 2: LFA Policy with SRLG Constraint

The objective is to force the LFA NH for both LDP FEC prefixes to use the path from PE-2 to PE-5.

Define SRLG group ‘blue’ with value ‘2’ and apply it to the IP interfaces PE-2 to PE-5 and PE-2 to PE-6.

```
*A:PE-2# configure router if-attribute srlg-group "blue" value 2

*A:PE-2# configure router interface "int-PE-2-PE-1" if-attribute srlg-group "blue"
*A:PE-2# configure router interface "int-PE-2-PE-6" if-attribute srlg-group "blue"
```

Define a route-next-hop policy template ‘LFA_NH_SRLG’, where SRLG is enabled, as follows:

```
*A:PE-2# configure
    router
        route-next-hop-policy
            begin
            template "LFA_NH_SRLG"
                srlg-enable
            exit
            commit
```

The policy is applied to the OSPF interface toward PE-1 and PE-6, as follows:

```
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-1" lfa-policy-map route-
nh-template "LFA_NH_SRLG"
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-6" lfa-policy-map route-
nh-template "LFA_NH_SRLG"
```

Only one LFA policy mapping is allowed on an OSPF interface at a time. The new LFA policy mapping replaces the previous one.

The LFA NHs for both LDP FEC prefixes will both point now to the interface from PE-2 to PE-5 as LFA backup NH, as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
```

```
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Push              --      262142BU
192.168.25.2                            1/1/4

192.0.2.1/32                            Swap            262142    262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Swap            262142    262142BU
192.168.25.2                            1/1/4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================


*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push              --      262143
192.168.26.2                            1/1/3

192.0.2.6/32                            Push              --      262138BU
192.168.25.2                            1/1/4

192.0.2.6/32                            Swap            262138    262143
192.168.26.2                            1/1/3

192.0.2.6/32                            Swap            262138    262138BU
192.168.25.2                            1/1/4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-2#
```

The LFA policy mapping is removed from the OSPF interfaces as follows:

```
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-1" no lfa-policy-map
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-6" no lfa-policy-map
```

Example 3: LFA Policy with NH-type Constraint

The objective is to force the LFA NH for IP prefix 192.0.2.6/32 to use an RSVP tunnel.

Enable IP FRR as follows:

```
*A:PE-2# configure router ip-fast-reroute
```

Set up an RSVP LSP tunnel toward 192.0.2.6 with a strict MPLS path going over PE-2 to PE-4 to PE-5 to PE-6.



Note: Since an RSVP LSP is set up between PE-2 and PE-6, MPLS/RSVP protocols also need to be enabled on all the corresponding IP interfaces along the MPLS path.

```
*A:PE-2# configure
    router
        mpls
            interface "int-PE-2-PE-4"
            exit
            path "path-PE-2-PE-4-PE-5-PE-6"
                hop 10 192.168.24.2 strict
                hop 20 192.168.45.2 strict
                hop 30 192.168.56.2 strict
                no shutdown
            exit
            lsp "LSP-PE-2-PE-6-strict"
                to 192.0.2.6
                primary "path-PE-2-PE-4-PE-5-PE-6"
                exit
                no shutdown
            exit
            no shutdown
```

Enable RSVP shortcut within the IGP on PE-2 and indicate that the newly created RSVP LSP is a possible shortcut candidate for LFA backup NH only.

```
*A:PE-2# configure router ospf rsvp-shortcut
*A:PE-2# configure router mpls lsp "LSP-PE-2-PE-6-strict" igp-shortcut lfa-only
```

Displaying the tunnel-table on PE-2 shows that an LDP LSP and an RSVP LSP is available toward PE-6:

```
*A:PE-2# show router tunnel-table 192.0.2.6

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.6/32      rsvp      MPLS  1         7        192.168.24.2   16777215
192.0.2.6/32      ldp       MPLS  65541     9        192.168.26.2   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-2#
```

The RSVP tunnel with tunnel ID 1 corresponds to the RSVP LSP “LSP-PE-2-PE-6-strict”, as can be seen as follows:

```
*A:PE-2# show router mpls lsp

===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                            To               Tun     Fastfail  Adm  Opr
                                                     Id      Config
-------------------------------------------------------------------------------
LSP-PE-2-PE-6-strict                192.0.2.6        1       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-2#
```

By default, the preferred NH type is IP, not tunnel. Therefore, the RSVP tunnel will not be used for the LFA backup, as can be seen in the following output:

```
*A:PE-2# show router route-table alternative 192.0.2.6/32

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
      Alt-NextHop                                                  Alt-
                                                                   Metric
-------------------------------------------------------------------------------
192.0.2.6/32                                    Remote  OSPF      00h03m21s  10
       192.168.26.2                                                  10
       192.168.12.1 (LFA)                                            20
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

Define a route-next-hop policy template **"LFA_NH_Tunnel"**, where nh-type is set to **tunnel**.

```
*A:PE-2# configure
    router
        route-next-hop-policy
            begin
            template "LFA_NH_Tunnel"
                nh-type tunnel
            exit
            commit
```

Apply the policy template to the interface toward PE-6, as follows:

```
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-6" lfa-policy-map route-
nh-template "LFA_NH_Tunnel"
```

The LFA NH uses the RSVP tunnel. The reference to the RSVP tunnel-ID (1) in the following show output corresponds with the tunnel-ID shown in the preceding **show router tunnel-table 192.0.2.6** output:

```
*A:PE-2# show router route-table alternative 192.0.2.6/32

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
      Alt-NextHop                                                 Alt-
                                                                  Metric
-------------------------------------------------------------------------------
192.0.2.6/32                                    Remote  OSPF      00h10m35s  10
       192.168.26.2                                                 10
       192.0.2.6 (LFA) (tunneled:RSVP:1)                            65535
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#


*A:PE-2# show router fib 1 nh-table-usage

===============================================================================
FIB Next-Hop Summary
===============================================================================
IPv4/IPv6                    Active                   Available
-------------------------------------------------------------------------------
IP Next-Hop                  9                        16383
Tunnel Next-Hop              1                        993279
ECMP Next-Hop                0                        512000
ECMP Tunnel Next-Hop         0                        261120
===============================================================================
*A:PE-2#
```

Example 4: Exclude Prefix from LFA Policy

The objective is to force no LFA NH for LDP FEC prefix 192.0.2.1/32 where PE-2 is the PLR.

The IP/LDP FRR implementation in SR OS allows to exclude an IGP interface, IGP area (OSPF), or IGP level (IS-IS) from the LFA SPF computation. The user also has the ability to exclude specific prefixes from the LFA SPF by using well-known prefix-lists and policy statements.

This is configured as follows:

```
*A:PE-2# configure
    router
        policy-options
            begin
            prefix-list "lo0-PE-1"
                prefix 192.0.2.1/32 exact
            exit
            policy-statement "LFA_Exclude_PE-1"
                entry 10
                    from
                        prefix-list "lo0-PE-1"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

The configured policy statement is applied to the IGP protocol, as follows:

```
*A:PE-2# configure router ospf loopfree-alternate-exclude prefix-policy
"LFA_Exclude_PE-1"
```

From the moment that it is applied, the existing LFA NH entries for LDP FEC prefix 192.0.2.5/32 disappear instantly (compare with the preceding Example 1):

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
```

```
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                              Push              --      262143
192.168.12.1                              1/1/2

192.0.2.1/32                              Swap            262142    262143
192.168.12.1                              1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-2#
```

Conclusion

In production MPLS networks where IP FRR and/or LDP FRR are deployed, it is possible that the existing calculated LFA NHs are not always taking the most optimal or desirable paths.

With LFA policies, operators have better control on the way in which LFA backup NHs are computed.

Different selection criteria can be part of the route-next-hop policy: IP admin groups, IP SRLG groups, protection type preference and NH type preference.

Unicast Routing Protocols

In This Section

This section provides configuration information for the following topics:

- Associating Communities with Static and Aggregate Routes
- BGP FlowSpec for IPv4 and IPv6
- BGP Multipath
- BGP Route Leaking
- IS-IS Link Bundling
- Policy Chaining and Logical Expressions
- Separate BGP RIBs for Labeled Routes

Associating Communities with Static and Aggregate Routes

This chapter provides information about associating communities with static and aggregate routes configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 11.0.R3, but the CLI in this edition corresponds to release 14.0.R2. There are no prerequisites for this configuration.

Introduction

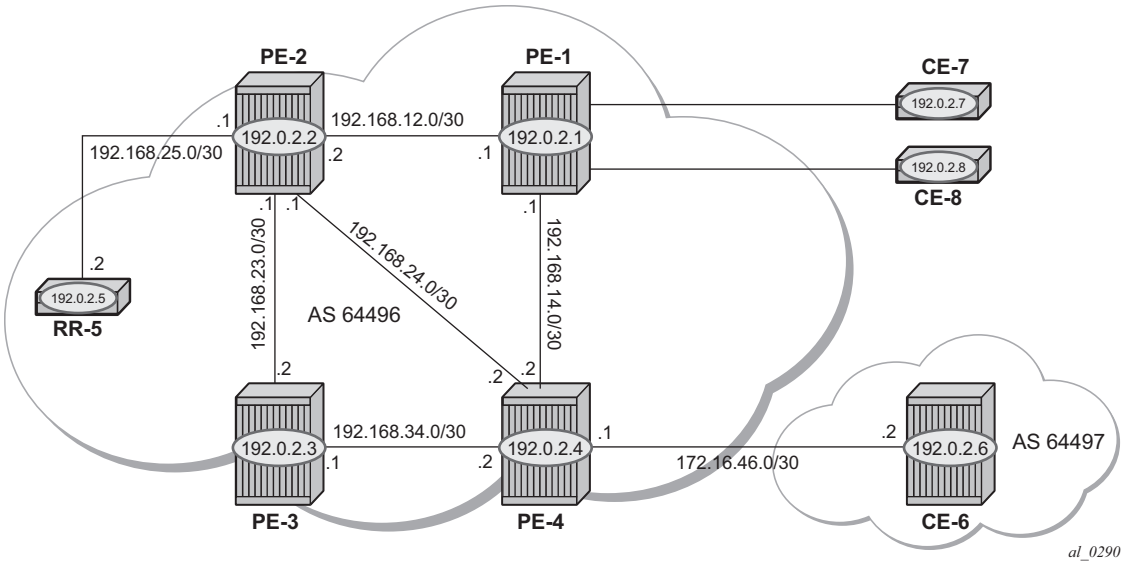
Border gateway protocol (BGP) communities are optional, transitive attributes attached to BGP route prefixes to carry additional information about that route prefix. A number of route prefixes can have the same community attached such that it can be matched by a route policy. As a result, the presence of a community value can be used to influence and control route policy.

A BGP community is a 32-bit value that is written as two separate 16 bit numbers separated by a colon. The first number usually represents the autonomous system (AS) number that defines or originates the community whilst the second is set by the network administrator.

Knowledge of RFC 4271 (BGP-4) and RFC 1997 (BGP Communities Attribute) is assumed throughout this document, as well as knowledge of multi-protocol BGP (MP-BGP) and RFC 4364 (BGP/MPLS IP VPNs).

Overview

Figure 54 **Example Topology**



The example topology is displayed in Figure 54. The setup uses 7750 Service Router (SR) nodes. PE-1 to PE-4 and the Route Reflector (RR-5) are located in the same Autonomous System (AS); AS64496. CE-6 is in a separate AS 64497 and peers using eBGP with its directly connected neighbor, PE-4.

The objectives are:

- To configure static-routes in a VPRN in PE-1 with various community values – including well-known communities – export them to other PEs within the same AS, and then via eBGP to CE-6. During this process, the community values for each route will be examined to ensure that the transitive nature of the attribute is maintained.
- To associate a community with an aggregate route that represents a larger number of composite prefixes. The aggregate will be advertised in place of the composite prefixes.

The following configuration tasks should be completed as a pre-requisite:

- Full mesh IS-IS or OSPF between all of the PE routers and the route reflector.
- iBGP between the RR and all PEs.
- eBGP between PE-4 and CE-6.
- Link-layer LDP between each PE.

Associating Communities with Static and Aggregate Routes

It is possible to add a single community value to a static and aggregate route without using a route policy.

The community value can be in the 4-byte format comprising of a 2-byte AS value, followed by a 2 byte decimal value, separated by a colon. It can also be the name of a well-known standard community; no-export, no-advertise, no-export-subconfed.

Any community added can be matched using a route policy.

The purpose of this example is to provision static and aggregate IPv4 route prefixes and associate a community with each route. These routes are then redistributed into the BGP protocol and advertised to other BGP speakers.

This is shown for IPv4 routes within a VPRN. Well-known, standard communities will also be configured to show that the correct behavior is observed.

Configuration

The first step is to configure an iBGP session between each of the PEs and the Route Reflector (RR). The address family negotiated between peers is vpn-ipv4.

The configuration for all PEs is as follows:

```
configure
    router
        autonomous-system 64496
        bgp
            group "internal"
                family vpn-ipv4
                peer-as 64496
                neighbor 192.0.2.5
                exit
            exit
```

The IP addresses can be derived from Figure 54.

The configuration for RR-5 is:

```
configure
    router
        autonomous-system 64496
        bgp
            cluster 0.0.0.1
```

```
            group "RR-clients"
                family vpn-ipv4
                peer-as 64496
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
```

On RR-5, show that BGP sessions with each PE are established, and have correctly negotiated the VPN IPv4 address family capability.

```
*A:RR-5# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.5        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 4
Total BGP Paths         : 6           Total Path Memory           : 1104
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

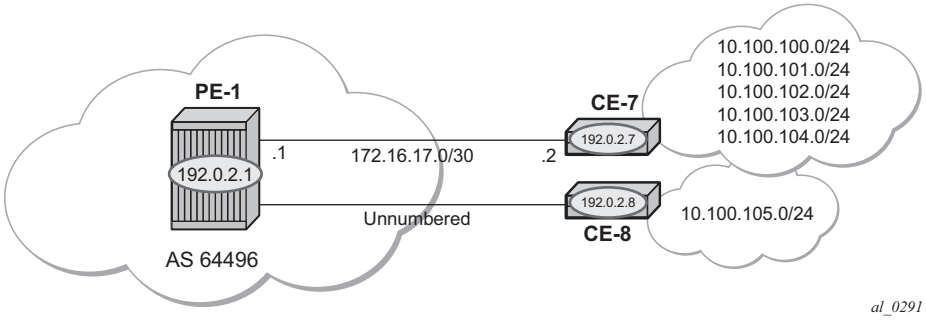
Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
```

```
                  AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                     PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
                64496        3    0 00h00m05s 0/0/0 (VpnIPv4)
                             3    0
192.0.2.2
                64496        3    0 00h00m05s 0/0/0 (VpnIPv4)
                             3    0
192.0.2.3
                64496        3    0 00h00m05s 0/0/0 (VpnIPv4)
                             3    0
192.0.2.4
                64496        3    0 00h00m05s 0/0/0 (VpnIPv4)
                             3    0
-------------------------------------------------------------------------------
*A:RR-5#
```

VPRN: IPv4

Figure 55 **CE Connections for Next-Hops**



The VPRN configuration for PE-1 is as follows:

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-1-CE-7" create
                address 172.16.17.1/30
                sap 1/2/1:1.0 create
                exit
            exit
            interface "loop1" create
```

```
                address 192.0.2.100/32
                loopback
            exit
            interface "int-PE-1-CE-8" create
                unnumbered "loop1"
                sap 1/2/2:1.0 create
            exit
        exit
        no shutdown
```

For unnumbered interfaces, an IP address is borrowed from a loopback interface, for example from the system interface, see MPLS chapter Unnumbered Interfaces in RSVP-TE and LDP.

LDP is used as the label-switching protocol for next-hop resolution.

PE-4 is configured with an interface toward CE-6 that supports eBGP. The following export policy is configured:

```
configure
    router
        policy-options
            begin
            policy-statement "BGP-VPN-accept"
                entry 10
                    from
                        protocol bgp-vpn
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

The configuration of the VPRN service is as follows:

```
configure
    service
        vprn 1 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-4-CE-6" create
                address 172.16.46.1/30
                sap 1/2/1:1 create
                exit
            exit
            bgp
                group "VPRN1-external"
```

```
                export "BGP-VPN-accept"
                peer-as 64497
                neighbor 172.16.46.2
                exit
            exit
        exit
        no shutdown
```

Static Routes with Communities

A static route has a number of next-hop options: direct connected IP address, black-hole, indirect IP address, and interface-name.

Figure 55 shows a pair of Customer Edge (CE) routers connected to PE1. The link to CE-7 is a numbered link. The link to CE-8 is an unnumbered link. The loopback interface address is used as a reference address for the unnumbered Ethernet interface.

Beyond CE-7 are a number of /24 subnets. Static routes to these individual subnets are created on PE-1 using a static route with a next-hop type of “interface address” or an “indirect address”. The indirect address is learned using a static route.

Beyond CE-8 is a single /24 subnet. A static route to this subnet is created with an interface-name as the next-hop.

There are a number of well-known, standard communities:

- no-export: the route is not advertised to any external peer. This route should be present in the route tables of all BGP speakers in the originating AS, but not in those in neighboring ASs.
- no-advertise: the route is not advertised to any peer. This route should not be present in any router as BGP-learned route.

The requirement for each subnet is:

- 10.100.100.0/24 must not be advertised outside of the AS. This must be associated with the standard, well-known community no-export. The community value is encoded as 65535:65281 (0xFFFFFF01), but the CLI requires the keyword **no-export**.

```
configure service vprn 1
            static-route-entry 10.100.100.0/24
                next-hop 172.16.17.2
                    community no-export
                    no shutdown
                exit
```

- 10.100.101.0/24 must be advertised with a community of 64496:101

```
configure service vprn 1
            static-route-entry 10.100.101.0/24
                next-hop 172.16.17.2
                    community 64496:101
                    no shutdown
                exit
```

- 10.100.102.0/24 must not be advertised to any BGP peer. This must be associated with the standard, well-known community **no-advertise**. The community value is encoded as 65535:65282 (0xFFFFFF02), but the CLI requires the keyword **no-advertise**.

```
configure service vprn 1
            static-route-entry 10.100.102.0/24
                next-hop 172.16.17.2
                    community no-advertise
                    no shutdown
                exit
```

- 10.100.103.0/24 must be advertised with a community of 64496:103 and a route tag of 10.

```
configure service vprn 1
            static-route-entry 10.100.103.0/24
                next-hop 172.16.17.2
                    community 64496:103
                    tag 10
                    no shutdown
                exit
            exit
```

- 10.100.104.0/24 must be advertised with a community of 64496:104. It is reachable via 192.0.2.7 which, in turn, is reachable via 172.16.17.2. For more information about this configuration, see chapter Associating Communities with Static and Aggregate Routes. This is using a static route which does not need to be advertised – hence it is associated with the **no-advertise** community.

```
configure service vprn 1
            static-route-entry 10.100.104.0/24
                indirect 192.0.2.7
                    community 64496:104
                    no shutdown
                exit
            exit
            static-route-entry 192.0.2.7/32
                next-hop 172.16.17.2
                    community no-advertise
                    no shutdown
                exit
            exit
```

- 10.100.105.0/24 must be advertised with a community of 64496:105. It is reachable via the unnumbered interface to CE-8.

```
configure service vprn 1
            static-route-entry 10.100.105.0/24
                next-hop "int-PE-1-CE-8"
                    community 64496:105
                    no shutdown
                exit
            exit
```

On PE-1, configure static routes that match the static routes from Figure 55, and the preceding conditions.

The default behavior of a VPRN is to export all static and connected routes into a BGP labeled route with the appropriate route-target extended community configured in the vrf-target statement. A single community string can be added using the static-route community commands shown above. If multiple communities are required, then a VRF-export policy should be used. This is outside the scope of this chapter.

Examine the BGP table of PE-1 to establish that routes have been exported correctly in VPN IPv4 toward RR-5.

```
*A:PE-1# show router bgp neighbor 192.0.2.5 advertised-routes vpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
i     64496:1:10.100.100.0/24                          100         None
-------------------------------------------------------------------------------
      192.0.2.1                                        None        262139
      No As-Path
i     64496:1:10.100.101.0/24                          100         None
      192.0.2.1                                        None        262139
      No As-Path
i     64496:1:10.100.103.0/24                          100         None
      192.0.2.1                                        None        262139
      No As-Path
i     64496:1:10.100.104.0/24                          100         None
      192.0.2.1                                        None        262139
      No As-Path
i     64496:1:10.100.105.0/24                          100         None
      192.0.2.1                                        None        262139
      No As-Path
i     64496:1:172.16.17.0/30                           100         None
      192.0.2.1                                        None        262139
      No As-Path
i     64496:1:192.0.2.100/32                           100         None
      192.0.2.1                                        None        262139
      No As-Path
```

```
-------------------------------------------------------------------------------
Routes : 7
===============================================================================
*A:PE-1#
```

There are only seven exported routes. The route prefixes associated with the **no-advertise** community are not present, as expected.

Examining the BGP table of PE-4 shows the presence of the expected routes, with the correct community values.

The prefix 10.100.100.0/24 is a member of community **no-export**. This is not advertised to PE-4.

In release 14.0.R1, the **show router bgp routes** command has been restructured to support more consistent parameter ordering. In most cases, a family filter must be specified before all other filtering parameters, except for the prefix.

```
*A:PE-4# show router bgp routes 10.100.100.0/24 vpn-ipv4 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.100.100.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262139
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-4-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : no-export target:64496:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Add Paths Send : Default
Source Class   : 0                      Dest Class     : 0
Last Modified  : 01h16m07s
```

```
VPRN Imported  :  1
---snip---
```

The following command shows all members of the community no-report:

```
*A:PE-4# show router bgp routes vpn-ipv4 community no-export
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  64496:1:10.100.100.0/24                              100         None
      192.0.2.1                                            None        262139
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-4#
```

Because the community no-export is encoded as community 65535:65281, the same output can be retrieved as follows:

```
*A:PE-4# show router bgp routes vpn-ipv4 community 65535:65281
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  64496:1:10.100.100.0/24                              100         None
      192.0.2.1                                            None        262139
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-4#
```

The prefix 10.100.101.0/24 is a member of community 64496:101. This is correctly advertised to PE-4.

```
*A:PE-4# show router bgp routes 10.100.101.0/24 vpn-ipv4 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.100.101.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262139
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-4-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : 64496:101 target:64496:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h34m23s
VPRN Imported  :  1
---snip---
```

The prefix 10.100.103.0/24 is a member of community 64496:103. This is correctly advertised to PE-4.

```
*A:PE-4# show router bgp routes 10.100.103.0/24 vpn-ipv4 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Original Attributes
```

```
Network        : 10.100.103.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262139
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-4-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : 64496:103 target:64496:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h26m24s
VPRN Imported  :  1
---snip---
```

The prefix 10.100.104.0/24 is a member of community 64496:104. This is correctly advertised to PE-4.

```
*A:PE-4# show router bgp routes 10.100.104.0/24 vpn-ipv4 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.100.104.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262139
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-4-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : 64496:104 target:64496:1
Cluster        : 0.0.0.1
```

```
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h20m45s
VPRN Imported  :  1
---snip---
```

The prefix 10.100.105.0/24 is a member of community 64496:105. This is correctly advertised to PE-4.

```
*A:PE-4# show router bgp routes 10.100.105.0/24 vpn-ipv4 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.100.105.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262139
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-4-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : 64496:105 target:64496:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h18m11s
VPRN Imported  :  1
---snip---
```

Examine the route table of VPRN 1 on PE-4 – looking specifically at the BGP-learned routes, the same seven routes are present as valid routes.

```
*A:PE-4# show router 1 route-table protocol bgp-vpn

===============================================================================
Route Table (Service: 1)
===============================================================================
      Next Hop[Interface Name]                                    Metric
Dest Prefix[Flags]                          Type    Proto     Age         Pref
-------------------------------------------------------------------------------
10.100.100.0/24                             Remote  BGP VPN   01h54m30s  170
       192.0.2.1 (tunneled)                                        0
10.100.101.0/24                             Remote  BGP VPN   01h46m55s  170
       192.0.2.1 (tunneled)                                        0
10.100.103.0/24                             Remote  BGP VPN   01h37m47s  170
       192.0.2.1 (tunneled)                                        0
10.100.104.0/24                             Remote  BGP VPN   01h30m18s  170
       192.0.2.1 (tunneled)                                        0
10.100.105.0/24                             Remote  BGP VPN   01h26m58s  170
       192.0.2.1 (tunneled)                                        0
172.16.17.0/30                              Remote  BGP VPN   01h54m30s  170
       192.0.2.1 (tunneled)                                        0
192.0.2.100/32                              Remote  BGP VPN   01h54m30s  170
       192.0.2.1 (tunneled)                                        0
-------------------------------------------------------------------------------
No. of Routes: 7
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-4#
```

Examine the route table of CE-6 – looking specifically at the BGP-learned routes, six routes are present as valid routes, as expected.

```
*A:CE-6# show router route-table protocol bgp

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                          Type    Proto     Age         Pref
      Next Hop[Interface Name]                                  Metric
-------------------------------------------------------------------------------
10.100.101.0/24                             Remote  BGP       00h04m31s  170
       172.16.46.1                                                 0
10.100.103.0/24                             Remote  BGP       00h04m31s  170
       172.16.46.1                                                 0
10.100.104.0/24                             Remote  BGP       00h04m31s  170
       172.16.46.1                                                 0
10.100.105.0/24                             Remote  BGP       00h04m31s  170
       172.16.46.1                                                 0
172.16.17.0/30                              Remote  BGP       00h04m31s  170
       172.16.46.1                                                 0
192.0.2.100/32                              Remote  BGP       00h04m31s  170
       172.16.46.1                                                 0
```

```
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:CE-6#
```

The prefix 10.100.100.0/24 is not received from PE-4 as it is a member of the **no-export** community.

```
*A:CE-6# show router bgp routes 10.100.100.0/24 detail

===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
No Matching Entries Found
===============================================================================
*A:CE-6#
```

Static route 10.100.101.0/24 is received with the correct community 64496:101.

```
*A:CE-6# show router bgp routes community 64496:101
===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  10.100.101.0/24                                  None        None
      172.16.46.1                                      None        -
      64496
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CE-6#
```

Static route 10.100.103.0/24 is received with the correct community 64496:103.

```
*A:CE-6# show router bgp routes community 64496:103

===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  10.100.103.0/24                                      None        None
      172.16.46.1                                          None        -
      64496
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CE-6#
```

Static route 10.100.104.0/24 is received with the correct community 64496:104.

```
*A:CE-6# show router bgp routes community 64496:104
===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
 Legend -
===============================================================================
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  10.100.104.0/24                                      None        None
      172.16.46.1                                          None        -
      64496
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CE-6#
```

Static route 10.100.105.0/24 is received with the correct community 64496:105.

```
*A:CE-6# show router bgp routes community 64496:105
===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
===============================================================================
```

```
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
 Legend -
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  10.100.105.0/24                                      None        None
      172.16.46.1                                          None        -
      64496
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CE-6#
```

Aggregate Routes with Communities

An aggregate route can be configured to represent a larger number of prefixes. For example, a set of prefixes 10.101.0.0/24 to 10.101.7.0/24 can be represented as a single aggregate prefix of 10.101.0.0/21.

This is due to the fact that the third octet in the range 0 to 7 can be represented by the 8 bits 00000000 to 00000111. The first 5 bits of this octet are common, along with the previous 2 octets, giving a prefix where the first 21 bits are common. Therefore, the aggregate can be written as 10.101.0.0/21.

In order to illustrate the configuration of an aggregate, consider following.

Figure 56 **CE-7 Connectivity**

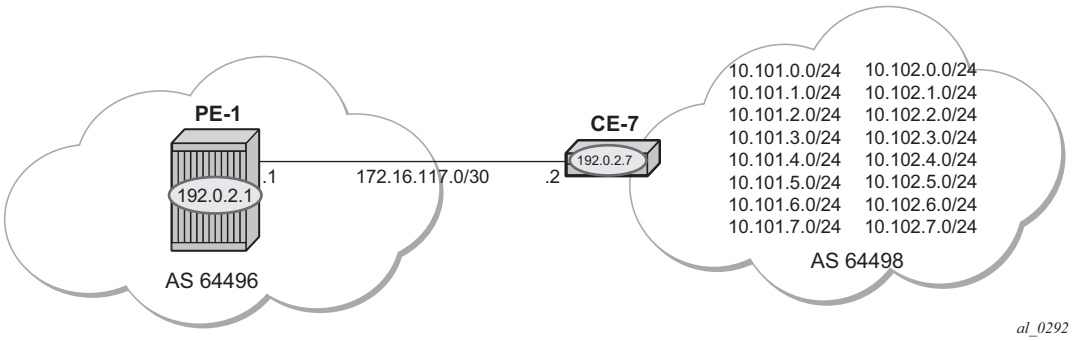


Figure 56 shows a CE router (CE-7), in AS 64498, that advertises a series of contiguous prefixes via BGP.

- 10.101.0.0/24 to 10.101.7.0/24
- 10.102.0.0/24 to 10.102.7.0/24

Instead of advertising all of these prefixes out of the VPRN towards an external CE individually, an aggregate route can be configured that summarizes each set of eight prefixes and a community can be directly associated with each aggregate route.

The configuration for a VPRN on PE-1, including the external BGP configuration is as follows:

```
*A:PE-1# configure
    service
        vprn 2 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:2
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:2
            interface "int-PE-1-CE-7_2nd" create
                address 172.16.117.1/30
                sap 1/2/1:2.0 create
                exit
            exit
            bgp
                group "external"
                    peer-as 64498
                    neighbor 172.16.117.2
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The neighbor relationship shows:

```
*A:PE-1# show router 2 bgp neighbor

===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                  : 172.16.117.2
Description           : (Not Specified)
Group                 : external
-------------------------------------------------------------------------------
Peer AS               : 64498            Peer Port            : 49673
Peer Address          : 172.16.117.2
Local AS              : 64496            Local Port           : 179
Local Address         : 172.16.117.1
Peer Type             : External         Dynamic Peer         : No
```

```
State                : Established      Last State           : Established
Last Event           : recvKeepAlive
Last Error           : Cease (Connection Collision Resolution)
Local Family         : IPv4
Remote Family        : IPv4
Hold Time            : 90               Keep Alive           : 30
Min Hold Time        : 0
Active Hold Time     : 90               Active Keep Alive    : 30
Cluster Id           : None
Preference           : 170              Num of Update Flaps  : 0
Recd. Paths          : 1
IPv4 Recd. Prefixes  : 16               IPv4 Active Prefixes : 16
IPv4 Suppressed Pfxs : 0                VPN-IPv4 Suppr. Pfxs : 0
VPN-IPv4 Recd. Pfxs  : 0                VPN-IPv4 Active Pfxs : 0
Mc IPv4 Recd. Pfxs.  : 0                Mc IPv4 Active Pfxs. : 0
Mc IPv4 Suppr. Pfxs  : 0                IPv6 Suppressed Pfxs : 0
IPv6 Recd. Prefixes  : 0                IPv6 Active Prefixes : 0
VPN-IPv6 Recd. Pfxs  : 0                VPN-IPv6 Active Pfxs : 0
VPN-IPv6 Suppr. Pfxs : 0
Mc IPv6 Recd. Pfxs.  : 0                Mc IPv6 Active Pfxs. : 0
Mc IPv6 Suppr. Pfxs  : 0                L2-VPN Suppr. Pfxs   : 0
L2-VPN Recd. Pfxs    : 0                L2-VPN Active Pfxs   : 0
MVPN-IPv4 Suppr. Pfxs: 0                MVPN-IPv4 Recd. Pfxs : 0
MVPN-IPv4 Active Pfxs: 0                MDT-SAFI Suppr. Pfxs : 0
MDT-SAFI Recd. Pfxs  : 0                MDT-SAFI Active Pfxs : 0
Flow-IPv4 Suppr. Pfxs: 0                Flow-IPv4 Recd. Pfxs : 0
Flow-IPv4 Active Pfxs: 0                Rte-Tgt Suppr. Pfxs  : 0
Rte-Tgt Recd. Pfxs   : 0                Rte-Tgt Active Pfxs  : 0
Backup IPv4 Pfxs     : 0                Backup IPv6 Pfxs     : 0
Mc Vpn Ipv4 Recd. Pf*: 0                Mc Vpn Ipv4 Active P*: 0
Mc Vpn Ipv4 Suppr. P*: 0
Backup Vpn IPv4 Pfxs : 0                Backup Vpn IPv6 Pfxs : 0
Input Queue          : 0                Output Queue         : 0
i/p Messages         : 6                o/p Messages         : 6
i/p Octets           : 228              o/p Octets           : 232
i/p Updates          : 1                o/p Updates          : 1
MVPN-IPv6 Suppr. Pfxs: 0                MVPN-IPv6 Recd. Pfxs : 0
MVPN-IPv6 Active Pfxs: 0
Flow-IPv6 Suppr. Pfxs: 0                Flow-IPv6 Recd. Pfxs : 0
Flow-IPv6 Active Pfxs: 0
Evpn Suppr. Pfxs     : 0                Evpn Recd. Pfxs      : 0
Evpn Active Pfxs     : 0
MS-PW Suppr. Pfxs    : 0                MS-PW Recd. Pfxs     : 0
MS-PW Active Pfxs    : 0
TTL Security         : Disabled         Min TTL Value        : n/a
Graceful Restart     : Disabled         Stale Routes Time    : n/a
Restart Time         : n/a
Advertise Inactive   : Disabled         Peer Tracking        : Disabled
Advertise Label      : None
Auth key chain       : n/a
Disable Cap Nego     : Disabled         Bfd Enabled          : Disabled
Flowspec Validate    : Disabled         Default Route Tgt    : Disabled
Aigp Metric          : Disabled         Split Horizon        : Disabled
Damp Peer Oscillatio*: Disabled         Update Errors        : 0
GR Notification      : Disabled         Fault Tolerance      : Disabled
Rem Idle Hold Time   : 00h00m00s
Next-Hop Unchanged   : None
Local Capability     : RtRefresh MPBGP 4byte ASN
Remote Capability    : RtRefresh MPBGP 4byte ASN
```

```
Routes Resolve To St*: Disabled
Local AddPath Capabi*: Disabled
Remote AddPath Capab*: Send - None
                     : Receive - None
Import Policy        : None Specified / Inherited
Export Policy        : None Specified / Inherited
Origin Validation    : N/A
EBGP Link Bandwidth  : n/a
IPv4 Rej. Pfxs       : 0               IPv6 Rej. Pfxs       : 0
VPN-IPv4 Rej. Pfxs   : 0               VPN-IPv6 Rej. Pfxs   : 0
Mc IPv4 Rej. Pfxs    : 0               Mc IPv6 Rej. Pfxs    : 0
MVPN-IPv4 Rej. Pfxs  : 0               MVPN-IPv6 Rej. Pfxs  : 0
Flow-IPv4 Rej. Pfxs  : 0               Flow-IPv6 Rej. Pfxs  : 0
L2-VPN Rej. Pfxs     : 0               MDT-SAFI Rej. Pfxs   : 0
Rte-Tgt Rej. Pfxs    : 0               MS-PW Rej. Pfxs      : 0
Mc Vpn Ipv4 Rej. Pfxs: 0               Evpn Rej. Pfxs       : 0
-------------------------------------------------------------------------------
Neighbors : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The following output shows that 16 BGP routes are received by PE-1.

```
*A:PE-1# show router 2 bgp routes
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  10.101.0.0/24                                        None        None
      172.16.117.2                                         None        -
      64498
u*>i  10.101.1.0/24                                        None        None
      172.16.117.2                                         None        -
      64498
u*>i  10.101.2.0/24                                        None        None
      172.16.117.2                                         None        -
      64498
u*>i  10.101.3.0/24                                        None        None
      172.16.117.2                                         None        -
      64498
u*>i  10.101.4.0/24                                        None        None
      172.16.117.2                                         None        -
      64498
u*>i  10.101.5.0/24                                        None        None
      172.16.117.2                                         None        -
      64498
```

```
u*>i  10.101.6.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.101.7.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.0.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.1.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.2.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.3.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.4.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.5.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.6.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
u*>i  10.102.7.0/24                                     None        None
      172.16.117.2                                      None        -
      64498
-------------------------------------------------------------------------------
Routes : 16
===============================================================================
*A:PE-1#
```

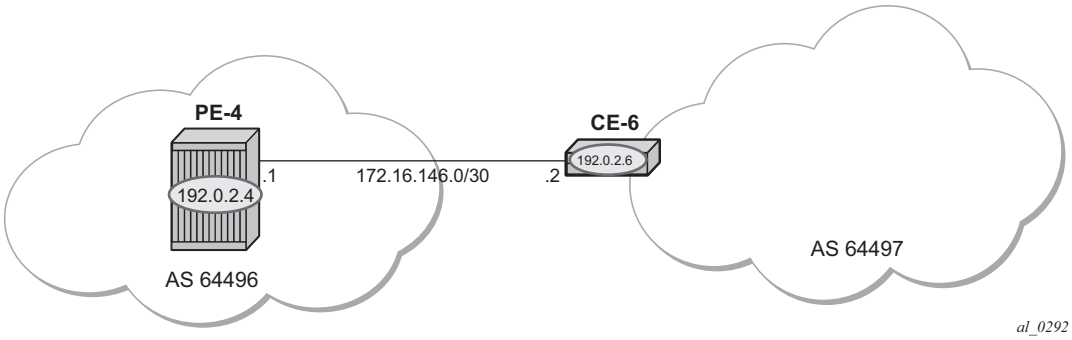
PE-4 also has a VPRN 2 instance configured, so that it will receive the imported BGP routes. The configuration for PE-4 is:

```
*A:PE-4# configure
    service
        vprn 2 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:2
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:2
            interface "int-PE-4-CE-6_2nd" create
                address 172.16.146.1/30
                sap 1/2/1:2 create
                exit
            exit
            bgp
                group "VPRN2-external"
                    peer-as 64497
```

```
                neighbor 172.16.146.2
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

Figure 57 shows the connectivity between PE-4 and CE-6. PE-4 will only forward a summarizing aggregate route toward CE-6.

Figure 57 **CE-6 Connectivity**



PE-4 receives labeled BGP route prefixes from PE-1 via the route reflector and installs them in the FIB for router instance 2:

```
*A:PE-4# show router 2 route-table

===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
10.101.0.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.1.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.2.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.3.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.4.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.5.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.6.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.101.7.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.0.0/24                                   Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.1.0/24                                   Remote  BGP VPN   00h01m07s  170
```

```
       192.0.2.1 (tunneled)                                          0
10.102.2.0/24                                 Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.3.0/24                                 Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.4.0/24                                 Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.5.0/24                                 Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.6.0/24                                 Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
10.102.7.0/24                                 Remote  BGP VPN   00h01m07s  170
       192.0.2.1 (tunneled)                                          0
172.16.117.0/30                               Remote  BGP VPN   00h02m41s  170
       192.0.2.1 (tunneled)                                          0
172.16.146.0/30                               Local   Local     00h02m42s  0
       int-PE-4-CE-6_2nd                                             0
-------------------------------------------------------------------------------
No. of Routes: 18
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-4#
```

The CE-6 configuration for an interface toward PE-4 is as follows:

```
*A:CE-6# configure
    service
        ies 2 customer 1 create
            interface "int-CE-6-PE-4_2nd" create
                address 172.16.146.2/30
                sap 1/1/1:2 create
                exit
            exit
            no shutdown
```

The BGP configuration of CE-6:

```
*A:CE-6# configure
    router
        bgp
            group "external-toVPRN2onPE-4"
                peer-as 64496
                neighbor 172.16.146.1
                exit
            exit
            no shutdown
```

The BGP neighbor state for PE-4:

```
*A:PE-4# show router 2 bgp neighbor 172.16.146.2

===============================================================================
BGP Neighbor
```

```
===============================================================================
-------------------------------------------------------------------------------
Peer                 : 172.16.146.2
Description          : (Not Specified)
Group                : VPRN2-external
-------------------------------------------------------------------------------
Peer AS              : 64497            Peer Port            : 51154
Peer Address         : 172.16.146.2
Local AS             : 64496            Local Port           : 179
Local Address        : 172.16.146.1
Peer Type            : External         Dynamic Peer         : No
State                : Established      Last State           : Established
Last Event           : recvKeepAlive
Last Error           : Cease (Connection Collision Resolution)
Local Family         : IPv4
Remote Family        : IPv4
Hold Time            : 90               Keep Alive           : 30
Min Hold Time        : 0
Active Hold Time     : 90               Active Keep Alive    : 30
Cluster Id           : None
Preference           : 170              Num of Update Flaps  : 0
Recd. Paths          : 0
IPv4 Recd. Prefixes  : 0                IPv4 Active Prefixes : 0
IPv4 Suppressed Pfxs : 0                VPN-IPv4 Suppr. Pfxs : 0
VPN-IPv4 Recd. Pfxs  : 0                VPN-IPv4 Active Pfxs : 0
Mc IPv4 Recd. Pfxs.  : 0                Mc IPv4 Active Pfxs. : 0
Mc IPv4 Suppr. Pfxs  : 0                IPv6 Suppressed Pfxs : 0
IPv6 Recd. Prefixes  : 0                IPv6 Active Prefixes : 0
VPN-IPv6 Recd. Pfxs  : 0                VPN-IPv6 Active Pfxs : 0
VPN-IPv6 Suppr. Pfxs : 0
Mc IPv6 Recd. Pfxs.  : 0                Mc IPv6 Active Pfxs. : 0
Mc IPv6 Suppr. Pfxs  : 0                L2-VPN Suppr. Pfxs   : 0
L2-VPN Recd. Pfxs    : 0                L2-VPN Active Pfxs   : 0
MVPN-IPv4 Suppr. Pfxs: 0                MVPN-IPv4 Recd. Pfxs : 0
MVPN-IPv4 Active Pfxs: 0                MDT-SAFI Suppr. Pfxs : 0
MDT-SAFI Recd. Pfxs  : 0                MDT-SAFI Active Pfxs : 0
Flow-IPv4 Suppr. Pfxs: 0                Flow-IPv4 Recd. Pfxs : 0
Flow-IPv4 Active Pfxs: 0                Rte-Tgt Suppr. Pfxs  : 0
Rte-Tgt Recd. Pfxs   : 0                Rte-Tgt Active Pfxs  : 0
Backup IPv4 Pfxs     : 0                Backup IPv6 Pfxs     : 0
Mc Vpn Ipv4 Recd. Pf*: 0                Mc Vpn Ipv4 Active P*: 0
Mc Vpn Ipv4 Suppr. P*: 0
Backup Vpn IPv4 Pfxs : 0                Backup Vpn IPv6 Pfxs : 0
Input Queue          : 0                Output Queue         : 0
i/p Messages         : 160              o/p Messages         : 36
i/p Octets           : 577              o/p Octets           : 558
i/p Updates          : 0                o/p Updates          : 0
MVPN-IPv6 Suppr. Pfxs: 0                MVPN-IPv6 Recd. Pfxs : 0
MVPN-IPv6 Active Pfxs: 0
Flow-IPv6 Suppr. Pfxs: 0                Flow-IPv6 Recd. Pfxs : 0
Flow-IPv6 Active Pfxs: 0
Evpn Suppr. Pfxs     : 0                Evpn Recd. Pfxs      : 0
Evpn Active Pfxs     : 0
MS-PW Suppr. Pfxs    : 0                MS-PW Recd. Pfxs     : 0
MS-PW Active Pfxs    : 0
TTL Security         : Disabled         Min TTL Value        : n/a
Graceful Restart     : Disabled         Stale Routes Time    : n/a
Restart Time         : n/a
Advertise Inactive   : Disabled         Peer Tracking        : Disabled
```

```
Advertise Label      : None
Auth key chain       : n/a
Disable Cap Nego     : Disabled          Bfd Enabled          : Disabled
Flowspec Validate    : Disabled          Default Route Tgt    : Disabled
Aigp Metric          : Disabled          Split Horizon        : Disabled
Damp Peer Oscillatio*: Disabled          Update Errors        : 0
GR Notification      : Disabled          Fault Tolerance      : Disabled
Rem Idle Hold Time   : 00h00m00s
Next-Hop Unchanged   : None
Local Capability     : RtRefresh MPBGP 4byte ASN
Remote Capability    : RtRefresh MPBGP 4byte ASN
Routes Resolve To St*: Disabled
Local AddPath Capabi*: Disabled
Remote AddPath Capab*: Send - None
                     : Receive - None
Import Policy        : None Specified / Inherited
Export Policy        : None Specified / Inherited
Origin Validation    : N/A
EBGP Link Bandwidth  : n/a
IPv4 Rej. Pfxs       : 0                 IPv6 Rej. Pfxs       : 0
VPN-IPv4 Rej. Pfxs   : 0                 VPN-IPv6 Rej. Pfxs   : 0
Mc IPv4 Rej. Pfxs    : 0                 Mc IPv6 Rej. Pfxs    : 0
MVPN-IPv4 Rej. Pfxs  : 0                 MVPN-IPv6 Rej. Pfxs  : 0
Flow-IPv4 Rej. Pfxs  : 0                 Flow-IPv6 Rej. Pfxs  : 0
L2-VPN Rej. Pfxs     : 0                 MDT-SAFI Rej. Pfxs   : 0
Rte-Tgt Rej. Pfxs    : 0                 MS-PW Rej. Pfxs      : 0
Mc Vpn Ipv4 Rej. Pfxs: 0                 Evpn Rej. Pfxs       : 0
-------------------------------------------------------------------------------
Neighbors : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-4#
```

In order to advertise a summarizing aggregate route with an associated community string, an aggregate route is required. In this case, the 10.101.x.0/24 group of prefixes will be associated with community 64496:101. The 10.102.x.0/24 group of prefixes will be associated with the standard community n**o-export**, so that it will not be advertised to any external peer.

The configuration required on PE-4 is as follows:

```
configure
    service
        vprn 2
            aggregate 10.101.0.0/21 community 64496:101
            aggregate 10.102.0.0/21 community no-export
        exit
```

An export policy is required on PE-4 to allow the advertising of the aggregate route. No community is applied using this policy.

```
configure
    router
        policy-options
            begin
```

```
            policy-statement "PE-4-VPN-Agg"
                entry 10
                    from
                        protocol aggregate
                    exit
                    action accept
                    exit
                exit
            commit
            exit
```

This is applied as an export policy within the group context of the BGP configuration of the VPRN.

```
configure
    service
        vprn 2
            bgp
                group "VPRN2-external"
                    export "PE-4-VPN-Agg"
                exit
```

The aggregate route 10.101.0.0/21 is received at CE-6 via BGP. The community that was associated with this prefix is seen: 64496:101. The route is seen as an aggregate, with PE-4 as the aggregating router (192.0.2.4). The “Atomic Aggregate” attribute is present, meaning that PE-4 has not advertised any details of the AS Paths of the composite routes.

```
*A:CE-6# show router bgp routes 10.101.0.0/21 hunt
===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.101.0.0/21
Nexthop        : 172.16.146.1
Path Id        : None
From           : 172.16.146.1
Res. Nexthop   : 172.16.146.1
Local Pref.    : None                  Interface Name : int-CE-6-PE-4_2nd
Aggregator AS  : 64496                 Aggregator     : 192.0.2.4
Atomic Aggr.   : Atomic                MED            : None
AIGP Metric    : None
Connector      : None
Community      : 64496:101
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 192.0.2.4
```

```
Fwd Class      : None                    Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64496
Route Tag      : 0
Neighbor-AS    : 64496
Orig Validation: NotFound
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h02m07s
---snip---
```

The aggregate route 10.102.0.0/21 is not received at CE-6, as PE-4 does not advertise it, due to the fact that it is associated with the “no-export” community.

```
*A:CE-6# show router bgp routes 10.102.0.0/21 hunt
===============================================================================
 BGP Router ID:192.0.2.6        AS:64497       Local AS:64497
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
No Matching Entries Found
===============================================================================
*A:CE-6#
```

Conclusion

Community strings can be added to static and aggregate routes. This example shows the configuration of communities with both static and aggregate routes, together with the associated show outputs which can be used to verify and troubleshoot them.

BGP FlowSpec for IPv4 and IPv6

This chapter provides information about BGP FlowSpec for IPv4 and IPv6.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to 7750 SR-7/12, 7750 SRc4/c12, 7750 SR-12E, 7750 SR-a, 7750 SR-e, XRS-20/40/16c, and 7450 ESS-7/12 in mixed-mode. Only interfaces supported on IOM3-XP, IOM4-e, XMA/C-XMA, IOM-a, and IMM cards can be configured for FlowSpec filtering. Also, all network interfaces must be on IOM3-XP, IOM4-e, XMA/C-XMA, IOM-a, or IMM cards in order to enable FlowSpec on any spoke-SDP interface.

Overview

The base BGP Flow Specification (FlowSpec) is defined in RFC 5575 (Flow Specification) and describes a method of encoding IPv4 flow specification information into Network Layer Reachability Information (NLRI). Draft-ietf-idr-flow-spec-v6 followed, and was considered an extension of RFC 5575 to include the IPv6 address family. The flow specification is an n-tuple consisting of one or more matching criteria, which can be applied to IP traffic. The FlowSpec NLRI is encoded into Multiprotocol BGP using MP_REACH_NLRI and MP_UNREACH_NLRI attributes.

As well as the flow specification defining match criteria, extended community attributes are defined to provide traffic filtering actions for the specified flow specification. Therefore, a FlowSpec route (MP_REACH_NLRI) contains a description of the traffic to be matched (using FlowSpec NLRI), and the filtering action to be taken with that traffic (using traffic filtering action extended communities). RFC 7674 provided an update to the original RFC 5575 specification to clarify the formatting of some of these traffic actions, notably redirect to VRF.

Initially, the primary use for FlowSpec was to dynamically distribute traffic filtering rules for mitigating distributed denial of service (DDoS) attacks. A router receiving a FlowSpec update can dynamically create IP filters to prevent both intra-AS and inter-AS DDoS attacks. Prevention is typically implemented by dropping traffic at the ingress point of the network (or nearest possible point toward the source of the DDoS attack) or by redirecting traffic to a separate routing context for forwarding (off-ramping) to a traffic-cleansing device. The ability to redirect traffic led to FlowSpec being considered for software defined networking (SDN)-driven applications or network re-optimization tools. In those cases, a subset of traffic needs to be forced (redirected) into a specific routing context or tunnel/label switched path (LSP) for network capacity optimization or to meet a service level agreement (SLA).

BGP FlowSpec uses AFI 1 (IPv4) or AFI 2 (IPv6) with SAFI 133 (IPv4 dissemination of flow specification rules) or SAFI 134 (VPNv4 dissemination of flow specification rules). Currently, only IPv4 and IPv6 are supported in SR OS.

The FlowSpec NLRI may consist of several components that form the flow specification. A packet only matches the flow specification when it matches all of the components in the NLRI. Table 1 lists the component types that are currently defined, their type values, and their support in SR OS. Flow specification components must follow strict ordering. If present in the specification, a component must precede any other component of higher type value.

Table 2 **FlowSpec Component Types**

Type Value	Component Type	SR OS Support
1	Destination Prefix	Yes
2	Source Prefix	Yes
3	IP Protocol	Partial (TCP/UDP only)
4	Port	Yes (specific number or single contiguous range)
5	Destination Port	Yes
6	Source Port	Yes
7	ICMP Type	Yes
8	ICMP Code	Yes
9	TCP Flags	SYN/ACK only
10	Packet Length	No
11	DiffServ Code Point	Yes

Table 2 **FlowSpec Component Types**

Type Value	Component Type	SR OS Support
12	Fragment	Yes (non-first only, first-only, fragment true, fragment false. Last fragment matching is not supported).

The traffic filtering action for a flow specification uses a number of extended community attributes. The attributes standardized in RFC 5575 are listed in Table 2, with their types and support in SR OS. The traffic rate extended community specifies the rate in bytes per second, where a rate of zero specifies a drop action. The traffic action extended community consists of six bytes; only the two least significant bits of the last byte are currently defined. The terminal action (T-bit), when set to 1, indicates that subsequent filtering rules should be applied (like a next-entry action). When this bit is set to zero, and this action is applied, the evaluation of the traffic filter stops. The sample bit (S-bit), when set to 1, enables traffic sampling and logging for this flow specification. The redirect to VRF and mark traffic class extended communities are self-explanatory, with a route-target value being used to define the target redirect VRF.

Table 3 **FlowSpec Extended Community Attributes**

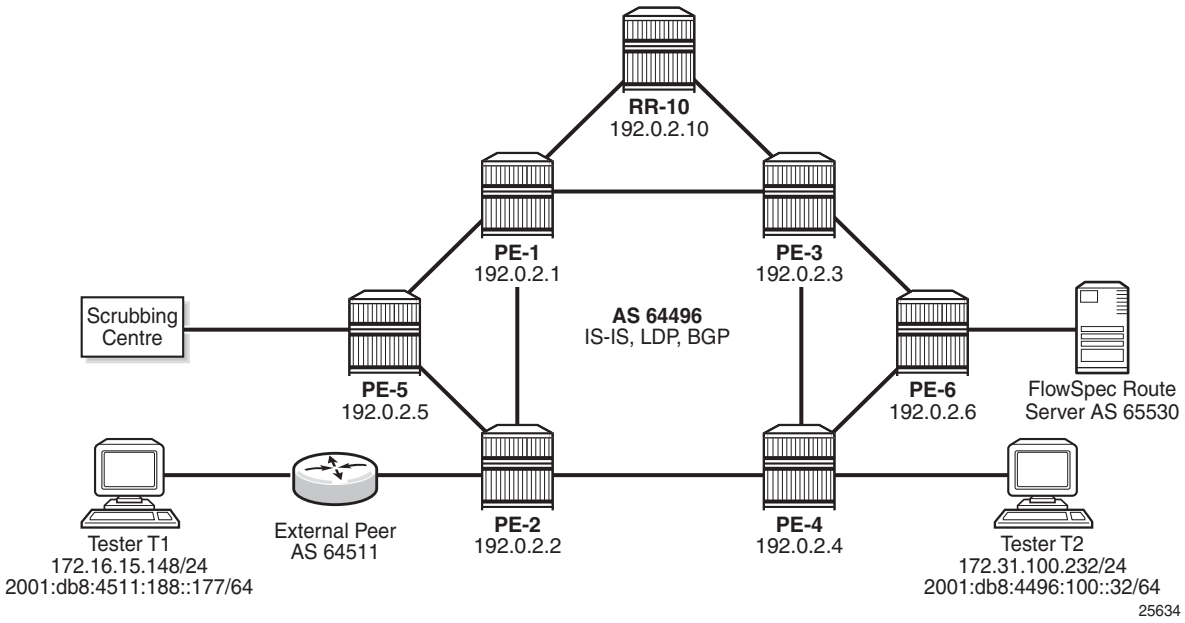
Type	Extended Community	SR OS Support
0x8006	Traffic Rate	0 (= drop)
0x8007	Traffic Action S-bit	Yes (uses default filter log 101)
0x8007	Traffic Action T-bit	No
0x8008	Redirect to VRF	Yes
0x8009	Mark Traffic Class	No

Example Topology

The example topology used throughout this chapter is shown in Figure 1. PE-1 through PE-6 and RR-10 participate in IS-IS Level-2 and LDP. All devices form part of network AS 64496, with all PE routers peering in IBGP with the Route-Reflector RR-10 for address families IPv4, IPv6, VPN-IPv4, VPN-IPv6, Flow-IPv4, and Flow-IPv6.

To allow for tunneling of IPv4 traffic in MPLS, all PE routers are configured for BGP next-hop-resolution using LDP shortcut-tunnels. To allow for tunneling of IPv6 traffic in MPLS, all PE routers are configured to run 6PE, with peering sessions configured to advertise labels for the IPv6 address family. Therefore, all IPv4 and IPv6 traffic is tunneled in MPLS.

Figure 58 **Example Topology**



To demonstrate FlowSpec, the following items are connected to AS 64496:

- PE-2 is connected to an external peer in AS 64511, which advertises the IPv4 prefix 172.16.0.0/20 and the IPv6 prefix 2001:db8:4511::/48 in EBGP. Both prefixes are advertised within AS 64496 by PE-2.
- PE-4 advertises IPv4 prefix 172.31.100.0/24 and IPv6 prefix 2001:db8:4496::/48 into IBGP, which PE-2 subsequently advertises in EBGP to AS 64511.
- Tester T1 is connected to the external peer in AS 64511 and sources and sinks traffic from IPv4 address 172.16.15.148 and IPv6 address 2001:db8:4511:188::177. Tester T2 is connected to PE-4 and sources and sinks traffic from IPv4 address 172.31.100.232 and IPv6 address 2001:db8:4496:100:32.
- PE-6 externally peers with a FlowSpec route server belonging to AS 65530.
- PE-5 connects to a DDoS scrubbing center with two interfaces:

– A "dirty" interface for forwarding of mitigated traffic toward the scrubbing center for cleansing. This interface is connected to an off-ramp VPRN configured on PE-5 and PE-2. PE-5 has static IPv4/IPv6 default routes toward the scrubbing center, which are subsequently advertised into the off-ramp VPRN. This provides sufficient routing information to attract redirected traffic from PE-2 toward the scrubbing center for cleansing.
– A "clean" interface for traffic received from the scrubbing center after it has been cleansed. This interface is connected to an IES service and is therefore routed toward its destination using the Global Routing Table (GRT).

Configuration

As an example of FlowSpec configuration, the following output shows the BGP configuration at PE-5. Similar configurations are applied on all other PE routers. All PE routers within AS 64496 peer as clients with RR-10 for the address families IPv4, IPv6, VPN-IPv4, VPN-IPv6, Flow-IPv4, and Flow-IPv6. The latter two are required for propagating the FlowSpec routes, and represent the only part of the BGP configuration required by FlowSpec.

The next-hop-resolution context allows for IPv4 traffic advertised by unlabeled BGP routes to use shortcut tunnels using the best tunnel type, based on tunnel-table preference. Because AS 64496 is only configured for LDP, LDP-based LSPs will be used for these shortcut tunnels. Similarly, to permit tunneling of IPv6 traffic in MPLS, the neighbor context contains the advertise-label ipv6 command to enable 6PE functionality.

```
bgp
    next-hop-resolution
        shortcut-tunnel
            family ipv4
                resolution any
            exit
        exit
    exit
    group "IBGP"
        family ipv4 ipv6 vpn-ipv4 vpn-ipv6 flow-ipv4 flow-ipv6
        peer-as 64496
        neighbor 192.0.2.10
            authentication-key <password> hash2
            advertise-label ipv6
        exit
    exit
    no shutdown
exit
```

PE-2 uses an IES to externally peer with AS 64511. The peering is for the IPv4 and IPv6 address families, and a separate IPv4/IPv6 BGP session is maintained for each address family. This external peering point is the point where ingress FlowSpec filters are instantiated. In the following output, the flowspec and flowspec-ipv6 parameters are configured within the SAP ingress context, to enable FlowSpec for IPv4 and IPv6, respectively. As well as SAPs, FlowSpec can also be enabled on spoke-SDPs within routed interfaces, and is supported within the base and VPRN routing instances.

```
ies 1 customer 1 create
    interface "to-AS64511" create
        address 192.168.2.1/30
        ipv6
            address 2001:db8:1b0c:2121::2/127
        exit
        sap 1/1/4:10.* create
            ingress
                flowspec
                flowspec-ipv6
            exit
        exit
    exit
    service-name "FlowSpec-test"
    no shutdown
exit
```

FlowSpec Operation

With FlowSpec enabled, FlowSpec routes can be advertised to dynamically trigger the instantiation of ingress filters. When valid FlowSpec routes are received, the system creates these filters on all interfaces where FlowSpec is enabled.

This section demonstrates the use of FlowSpec for traffic black-holing and traffic redirection for both IPv4 and IPv6.

IPv4 FlowSpec

To validate the instantiation of ingress filters based on IPv4 FlowSpec routes, a bidirectional traffic stream is commenced between Tester T1 (172.16.15.148) in AS 64511 and Tester T2 (172.31.100.232) in AS 64496. In the Tester T1 to Tester T2 direction, the destination port is TCP port 4191.

An IPv4 FlowSpec route is subsequently established to black-hole/drop traffic with a source address of 172.16.15.148 (Tester T1) and a destination address of 172.31.100.232 (Tester T2), for any destination ports in the range 4190-4199. The following output shows the route as received at PE-2.

```
1 2016/02/01 17:08:51.80 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.10
"Peer 1: 192.0.2.10: UPDATE
Peer 1: 192.0.2.10 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family FLOW_IPV4
        NLRI len: 22
          dest_pref   172.31.100.232/32
          src_pref    172.16.15.148/32
          ip_proto    [ == 6 ]
          dest_port   [ >4190 ] and [ <4199 ]
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 65530 >
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.10
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        rate-limit: 0:0
"
```

The route is shown as an MP_REACH_NLRI for address family Flow-IPv4 (AFI 1 SAFI 133). The NLRI uses the source and destination prefixes, the IP protocol, and the destination-port components to describe the flow and create the filter match criteria. The traffic rate extended community is then used to define a rate of 0, which is the filter drop action.

SR OS does not support a non-zero traffic rate. Although the route will be accepted, no filters will be instantiated.

Unlike other address families, there is no strict requirement for the Next-Hop attribute to be present in the MP_REACH_NLRI. The Length of Next-Hop in the Address field can optionally be set to zero and should be ignored on receipt.

The received FlowSpec route can also be verified in the RIB, which provides a concise output of the flow attributes and traffic filtering function, as follows.

```
A:PE-2# show router bgp routes flow-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
```

```
BGP FLOW IPV4 Routes
===============================================================================
Flag  Network             Nexthop                 LocalPref       MED
      As-Path
-------------------------------------------------------------------------------
u*>i  --                  0.0.0.0                 100             None
      65530

      Community Action:  rate-limit: 0:0
      NLRI Subcomponents:
      Dest Pref : 172.16.15.148/32
      Src Pref  : 172.31.100.232/32
      Ip Proto  : [ == 6 ]
      Dest Port : [ >4190 ] and [ <4199 ]
-------------------------------------------------------------------------------
Routes : 1
```

The dynamically created FlowSpec IPv4 ingress filter is shown in the following output, using the filter identifier "fSpec-0", created for the base routing instance. The entry within the filter is 32767 and indicates that it has been inserted by BGP FlowSpec. If a static ingress IP filter is also in use, FlowSpec generated filter entries are processed after all static filter entries (also referred to as user-defined filter entries), but prior to the default static filter action. The output shows a count of ingress matches, which are dropped. This is verified with the loss of traffic from Tester T1 to Tester T2, but not vice versa.

```
A:PE-2# show filter ip "fSpec-0"
===============================================================================
IP Filter
===============================================================================
Filter Id    : fSpec-0                          Applied        : Yes
Scope        : Template                         Def. Action    : Forward
System filter: Unchained
Radius Ins Pt: n/a
CrCtl. Ins Pt: n/a
RadSh. Ins Pt: n/a
PccRl. Ins Pt: n/a
Entries      : 1 (insert By Bgp)
Description  : IPv4 BGP FlowSpec filter for the Base router
-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry        : 32767 - inserted by BGP FlowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 172.16.15.148/32
Src. Port    : n/a
Dest. IP     : 172.31.100.232/32
Dest. Port   : 4191..4198
Protocol     : 6                                Dscp           : Undefined
ICMP Type    : Undefined                        ICMP Code      : Undefined
Fragment     : Off                              Src Route Opt  : Off
Sampling     : Off                              Int. Sampling  : On
IP-Option    : 0/0                              Multiple Option: Off
TCP-syn      : Off                              TCP-ack        : Off
Option-pres  : Off
```

```
Egress PBR   : Disabled
Match action : Drop
Ing. Matches : 7234 pkts (723400 bytes)
Egr. Matches : 0 pkts
```

When the route is withdrawn and PE-2 receives an MP_UNREACH_NLRI for the same FlowSpec NLRI, the dynamically created filter entries are removed and all associated hardware resources (TCAM entries) are released.

Instead of dropping traffic at the ingress point to the network, an alternative option is to redirect the mitigated traffic to a traffic-cleansing device, if this infrastructure exists. FlowSpec has the Redirect to VRF extended community for this purpose, with the process of forwarding traffic toward a scrubbing center frequently referred to as off-ramping. At PE-2, a VPRN is configured to off-ramp traffic toward the scrubbing center connected to PE-5, as shown in the following output.

In the case of FlowSpec, traffic redirection is half-duplex. That is, traffic is forwarded from PE-2 toward PE-5, but not from PE-5 toward PE-2. This is because when the traffic has been cleansed, it re-enters the network at PE-5 within an IES, and is therefore routed toward its destination using the GRT. This process is frequently referred to as on-ramping. As a result of this half-duplex traffic flow, only a vrf-target import statement is created. There is no requirement to export any routes from PE-2.

```
vprn 2 customer 1 create
    route-distinguisher 64496:2
    auto-bind-tunnel
        resolution any
    exit
    vrf-target import target:64496:2
    service-name "FlowSpec-OffRamp-VRF"
    no shutdown
exit
```

The VPRN used for the purpose of off-ramping is also configured at PE-5 and has a single SAP toward the scrubbing center, as shown in the following output. Static IPv4 and IPv4 default routes are configured with next hops of the scrubbing center and these are advertised into VPN-IPv4/VPN-IPv6 using route-policy. There is no requirement for PE-5 to import any BGP-VPN routes.

```
vprn 2 customer 1 create
    vrf-export "vrf2-export"
    route-distinguisher 64496:2
    interface "OffRamp-to-Scrubbing-Center" create
        address 192.168.2.5/30
        ipv6
            address 2001:db8:1b0c:2121::4/127
        exit
        sap 1/1/3:12 create
        exit
    exit
    static-route 0.0.0.0/0 next-hop 192.168.2.6
    static-route ::/0 next-hop 2001:db8:1b0c:2121::5
```

```
            service-name "FlowSpec-OffRamp-VRF"
            no shutdown
        exit
```

The following output shows the on-ramp IES at PE-5. This is the point where cleansed traffic re-enters the network and is forwarded toward its destination using the GRT.

```
 ies 3 customer 1 create
            interface "OnRamp-from-Scrubbing-Center" create
                address 192.168.2.9/30
                ipv6
                    address 2001:db8:1b0c:2121::6/127
                exit
                sap 1/1/3:13 create
                exit
            exit
            service-name "FlowSpec-OnRamp-IES"
            no shutdown
        exit
```

To validate the instantiation of the redirection filter, the same bidirectional traffic stream is started between Tester T1 (172.16.15.148) in AS 64511 and Tester T2 (172.31.100.232) in AS 64496. In the Tester T1 to Tester T2 direction, the destination port is TCP port 4191. When the IPv4 FlowSpec route is received at PE-2, the NLRI shows the same traffic match criteria previously used for the black-hole/drop scenario. However, the extended community has changed to redirect-to-vrf with a route-target value of 64496:2, as shown in the following output.

```
1 2016/02/02 10:48:07.81 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.10
"Peer 1: 192.0.2.10: UPDATE
Peer 1: 192.0.2.10 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family FLOW_IPV4
        NLRI len: 22
          dest_pref   172.31.100.232/32
          src_pref    172.16.15.148/32
          ip_proto    [ == 6 ]
          dest_port   [ >4190 ] and [ <4199 ]
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 65530 >
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.10
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        redirect-to-vrf:64496:2
"
```

The FlowSpec ingress filter in the following output uses the filter identifier "fSpec-0" and the entry 32767 inserted by BGP FlowSpec. The output shows that there are ingress packet/byte matches against the defined match criteria received in the FlowSpec NLRI. The match action field indicates an action of forward, and the forwarding router/service ID ("Fwd Rtr/SvcId") is service ID 2 (the off-ramp VPRN).

Traffic is correctly received in the Tester T1 to Tester T2 direction, and also in the opposite Tester T2 to Tester T1 direction. However, traffic in the Tester T1 to Tester T2 direction is redirected by PE-2 toward the scrubbing center attached to PE-5, before being forwarded to its destination at PE-4.

```
A:PE-2# show filter ip "fSpec-0"
===============================================================================
IP Filter
===============================================================================
Filter Id    : fSpec-0                          Applied        : Yes
Scope        : Template                         Def. Action    : Forward
System filter: Unchained
Radius Ins Pt: n/a
CrCtl. Ins Pt: n/a
RadSh. Ins Pt: n/a
PccRl. Ins Pt: n/a
Entries      : 1 (insert By Bgp)
Description  : IPv4 BGP FlowSpec filter for the Base router
-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry        : 32767 - inserted by BGP FlowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 172.16.15.148/32
Src. Port    : n/a
Dest. IP     : 172.31.100.232/32
Dest. Port   : 4191..4198
Protocol     : 6                                Dscp           : Undefined
ICMP Type    : Undefined                        ICMP Code      : Undefined
Fragment     : Off                              Src Route Opt  : Off
Sampling     : Off                              Int. Sampling  : On
IP-Option    : 0/0                              Multiple Option: Off
TCP-syn      : Off                              TCP-ack        : Off
Option-pres  : Off
Egress PBR   : Disabled
Match action : Forward
Fwd Rtr/SvcId: 2
PBR Down Act : Drop (entry-default)
Ing. Matches : 1252 pkts (125200 bytes)
Egr. Matches : 0 pkts
```

IPv6 FlowSpec

To validate the instantiation of ingress filters based on IPv6 FlowSpec routes, a bidirectional traffic stream is commenced between Tester T1 (2001:db8:4511:188::177) in AS 64511 and Tester T2 (2001:db8:4496:100::32) in AS 64496. In the Tester T1 to Tester T2 direction, the destination port is TCP port 4191.

An IPv6 FlowSpec route is then established to black-hole/drop traffic with a source address of 2001:db8:4511:188::177 (Tester T1) and a destination address of 2001:db8:4496:100::32 (Tester T2), for any destination ports in the range 4190-4199. The following output shows the route as received at PE-2.

```
1 2016/02/04 14:08:37.80 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.10
"Peer 1: 192.0.2.10: UPDATE
Peer 1: 192.0.2.10 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 110
    Flag: 0x90 Type: 14 Len: 54 Multiprotocol Reachable NLRI:
        Address Family FLOW_IPV6
        NLRI len: 48
          dest_pref   2001:db8:4496:100::32/128 offset 0
          src_pref    2001:db8:4511:188::177/128 offset 0
          ip_proto    [ == 6 ]
          dest_port   [ >4190 ] and [ <4199 ]
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 65530 >
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.10
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        rate-limit: 0:0
"
```

The route is shown as an MP_REACH_NLRI for address family Flow-IPv6 (AFI 2 SAFI 133). As with the FlowSpec IPv4 example, the NLRI uses the source and destination prefixes, the IP protocol, and the destination-port components to describe the flow and create the filter match criteria. The traffic rate extended community is then used to define a rate of 0, which is equivalent to a filter drop action.

The dynamically created FlowSpec IPv6 ingress filter is shown in the following output. Although this is an IPv6 filter, the filter identifier "fSpec-0" is still used for the base routing instance. The filter entry is 32767 and indicates that it has been inserted by BGP FlowSpec. The output shows a count of ingress matches, which are dropped in accordance with the Match action field. This is verified with the loss of traffic from Tester T1 to Tester T2, but not vice versa.

```
A:PE-2# show filter ipv6 "fSpec-0"
===============================================================================
IPv6 Filter
===============================================================================
Filter Id    : fSpec-0                          Applied        : Yes
Scope        : Template                         Def. Action    : Forward
System filter: Unchained
Radius Ins Pt: n/a
CrCtl. Ins Pt: n/a
RadSh. Ins Pt: n/a
PccRl. Ins Pt: n/a
Entries      : 1 (insert By Bgp)
Description  : IPv6 BGP FlowSpec filter for the Base router
-------------------------------------------------------------------------------
Filter Match Criteria : IPv6
-------------------------------------------------------------------------------
Entry        : 32767 - inserted by BGP FlowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 2001:db8:4511:188::177/128
Src. Port    : n/a
Dest. IP     : 2001:db8:4496:100::32/128
Dest. Port   : 4191..4198
Next Header  : 6                                Dscp           : Undefined
ICMP Type    : Undefined                        ICMP Code      : Undefined
Sampling     : Off                              Int. Sampling  : On
TCP-syn      : Off                              TCP-ack        : Off
Fragment     : Off
HopByHop Opt : Off                              Routing Type0  : Off
Auth Hdr     : Off                              ESP header     : Off
Flow-label   : n/a                              Flow-label Mask: n/a
Match action : Drop
Ing. Matches : 2996 pkts (305592 bytes)
Egr. Matches : 0 pkts
```

The FlowSpec IPv6 route with the drop action is subsequently withdrawn, restoring the traffic flow between Tester T1 and Tester T2.

To off-ramp IPv6 traffic toward the scrubbing center, the same redirect infrastructure is used as in the IPv4 example:

- PE-2 and PE-5 use the same off-ramp VPRN (VPRN 2), which transports both VPN-IPv4 and VPN-IPv6 traffic.
- PE-5 uses the same on-ramp (IES). When traffic is returned from the scrubbing center, PE-5 routes packets toward their destination using the GRT.

An IPv6 FlowSpec route with a redirect-to-VRF extended community is then sourced by the FlowSpec route generator. When the route is received at PE-2, the NLRI shows the same traffic match criteria previously used for the IPv6 black-hole/drop scenario. However, the extended community has changed to *redirect-to-vrf* with a route-target value of 64496:2, as shown in the following output.

```
1 2016/02/04 17:00:28.74 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.10
"Peer 1: 192.0.2.10: UPDATE
```

```
Peer 1: 192.0.2.10 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 110
    Flag: 0x90 Type: 14 Len: 54 Multiprotocol Reachable NLRI:
        Address Family FLOW_IPV6
        NLRI len: 48
          dest_pref   2001:db8:4496:100::32/128 offset 0
          src_pref    2001:db8:4511:188::177/128 offset 0
          ip_proto    [ == 6 ]
          dest_port   [ >4190 ] and [ <4199 ]
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 65530 >
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.10
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        redirect-to-vrf:64496:2
"
```

The FlowSpec ingress IPv6 filter in the following output uses the filter identifier "fSpec-0" and the entry 32767 inserted by BGP FlowSpec. The output shows that there are ingress packet/byte matches against the defined match criteria received in the FlowSpec NLRI. The match action field indicates an action of forward, and the forwarding router/service ID is service ID 2 (the off-ramp VPRN).

Traffic is correctly received in the Tester T1 to Tester T2 direction, and also in the opposite Tester T2 to Tester T1 direction. However, traffic in the Tester T1 to Tester T2 direction is redirected by PE-2 toward the scrubbing center attached to PE-5 before being forwarded to its destination at PE-4.

```
A:PE-2-# show filter ipv6 "fSpec-0"
===============================================================================
IPv6 Filter
===============================================================================
Filter Id    : fSpec-0                          Applied        : Yes
Scope        : Template                         Def. Action    : Forward
System filter: Unchained
Radius Ins Pt: n/a
CrCtl. Ins Pt: n/a
RadSh. Ins Pt: n/a
PccRl. Ins Pt: n/a
Entries      : 1 (insert By Bgp)
Description  : IPv6 BGP FlowSpec filter for the Base router
-------------------------------------------------------------------------------
Filter Match Criteria : IPv6
-------------------------------------------------------------------------------
Entry        : 32767 - inserted by BGP FlowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 2001:db8:4511:188::177/128
Src. Port    : n/a
Dest. IP     : 2001:db8:4496:100::32/128
```

```
Dest. Port    : 4191..4198
Next Header   : 6                                  Dscp           : Undefined
ICMP Type     : Undefined                          ICMP Code      : Undefined
Sampling      : Off                                Int. Sampling  : On
TCP-syn       : Off                                TCP-ack        : Off
Fragment      : Off
HopByHop Opt  : Off                                Routing Type0  : Off
Auth Hdr      : Off                                ESP header     : Off
Flow-label    : n/a                                Flow-label Mask: n/a
Match action  : Forward
Fwd Rtr/SvcId: 2
PBR Down Act  : Drop (entry-default)
Ing. Matches  : 1952 pkts (199104 bytes)
Egr. Matches  : 0 pkts
```

FlowSpec Validation

FlowSpec routes are typically originated and contained within the administrative domain of an operator; particularly when used for DDoS mitigation purposes. This approach means applying ingress filters at the point where traffic enters the autonomous system (AS), such as an external peering point.

For these filters to be optimal, they should be instantiated as close as possible to the source of the attack traffic, even if that means applying filters within another operator's domain. This means that FlowSpec routes must be exchanged between autonomous systems, which requires that:

- A trust relationship exists between those autonomous systems.
- A method of validating FlowSpec routes advertised between autonomous systems exists. This is referred to as FlowSpec validation.

When FlowSpec validation is enabled, it introduces an additional step in the BGP path selection process. That is, a FlowSpec NLRI is only considered feasible if the originator of the FlowSpec route matches the originator of the best-match unicast route for the destination prefix embedded in the FlowSpec NLRI. The originator is defined as being either the BGP originator path attribute, if Route Reflection is being used, or the transport address of the BGP peer, if this attribute is not present.

FlowSpec validation is enabled at the BGP, group, or neighbor level, using the flowspec-validate command, as shown in the following output, and can be supported within the base and VPRN routing contexts. Within the test topology, it is enabled on all IBGP sessions within AS 64496, and also at PE-2's EBGP peering session to AS 64511.

```
router Base
        group "EBGP"
            neighbor 192.168.2.2
                family ipv4 flow-ipv4
```

```
                peer-as 64511
                flowspec-validate
                split-horizon
            exit
        exit
```

To verify the use of FlowSpec validation, the external peer in AS 64511 will attempt to instantiate an IPv4 FlowSpec filter at PE-4. The FlowSpec NLRI uses Tester T2's IP address (172.31.100.232) as the source prefix and Tester T1's IP address (172.16.15.148) as the destination prefix.

Initially, the external peer in AS 64511 originates only the IPv4 FlowSpec route, but does not advertise the IPv4 prefix 172.16.0.0/20, which would be the destination prefix contained in the FlowSpec NLRI. Therefore, PE-2 cannot match the originator of the FlowSpec IPv4 route with the originator of the best-match unicast route for the destination prefix in the FlowSpec NLRI, and the route is not validated. That is, because AS 64511 does not advertise the prefix 172.16.0.0/20, there is no unicast route for the destination prefix in the FlowSpec NLRI. This is shown in the following output, where the FlowSpec route from AS 64511 is held in RIB-In, but is not flagged as used, valid, or best. Therefore, the route remains inactive and is not propagated any further by PE-2.

```
*A:PE-2# show router bgp routes flow-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP FLOW IPV4 Routes
===============================================================================
Flag  Network             Nexthop                 LocalPref       MED
      As-Path
-------------------------------------------------------------------------------
i     --                  0.0.0.0                 n/a             None
      64511

      Community Action:  rate-limit: 0:0
      NLRI Subcomponents:
      Dest Pref : 172.16.15.148/32
      Src Pref  : 172.31.100.232/32
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The external peer in AS 64511 now advertises the prefix 172.16.0.0/20, together with the same IPv4 FlowSpec route with NLRI, containing a source prefix of 172.16.15.148/32 and destination prefix of 172.31.100.232/32. When the route is received at PE-2, the route can be validated because the best-match unicast route for the destination prefix in the FlowSpec NLRI, and the actual FlowSpec route, have a common originator. Therefore, the route is installed as valid, used, and best, and is advertised internally by PE-2 to AS 64496, as shown in the following output.

The originator in this case is the BGP transport address because no Route-Reflection exists on the external peering between AS 64511 and AS 64496.

```
*A:PE-2>config>router>bgp# show router bgp routes flow-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP FLOW IPV4 Routes
===============================================================================
Flag  Network             Nexthop                 LocalPref       MED
      As-Path
-------------------------------------------------------------------------------
u*>i  --                  0.0.0.0                 n/a             None
      64511

      Community Action:  rate-limit: 0:0
      NLRI Subcomponents:
      Dest Pref : 172.16.15.148/32
      Src Pref  : 172.31.100.232/32
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The FlowSpec route is reflected by RR-10 and received at PE-4, where FlowSpec validation again takes place. Because Route-Reflection exists within AS 64496, both the IPv4 prefix 172.16.0.0/20 and the FlowSpec IPv4 route have a common originator ID (inserted by RR-10) of 192.0.2.2 (PE-2). Therefore, the route is validated, and again installed as valid, used, and best, as shown in the following output.

```
*A:PE-4# show router bgp routes flow-ipv4
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP FLOW IPV4 Routes
===============================================================================
```

```
Flag  Network             Nexthop                 LocalPref       MED
      As-Path
-------------------------------------------------------------------------------
u*>i  --                  0.0.0.0                 100             None
      64511

      Community Action:  rate-limit: 0:0
      NLRI Subcomponents:
      Dest Pref : 172.16.15.148/32
      Src Pref  : 172.31.100.232/32
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-4 also has FlowSpec enabled on the SAP toward Tester T2, and because it has received a valid FlowSpec route, it dynamically instantiates the FlowSpec IPv4 filter using the filter "fSpec-0", as shown in the following output.

```
*A:PE-4-vm# show filter ip "fSpec-0"
===============================================================================
IP Filter
===============================================================================
Filter Id    : fSpec-0                          Applied        : Yes
Scope        : Template                         Def. Action    : Forward
System filter: Unchained
Radius Ins Pt: n/a
CrCtl. Ins Pt: n/a
RadSh. Ins Pt: n/a
PccRl. Ins Pt: n/a
Entries      : 1 (insert By Bgp)
Description  : IPv4 BGP FlowSpec filter for the Base router
-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry        : 32767 - inserted by BGP FlowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 172.31.100.232/32
Src. Port    : n/a
Dest. IP     : 172.16.15.148/32
Dest. Port   : n/a
Protocol     : Undefined                        Dscp           : Undefined
ICMP Type    : Undefined                        ICMP Code      : Undefined
Fragment     : Off                              Src Route Opt  : Off
Sampling     : Off                              Int. Sampling  : On
IP-Option    : 0/0                              Multiple Option: Off
TCP-syn      : Off                              TCP-ack        : Off
Option-pres  : Off
Egress PBR   : Disabled
Match action : Drop
Ing. Matches : 0 pkts
Egr. Matches : 0 pkts
```

Resource Consumption

Similar to static filters consuming hardware resources, also dynamically instantiated FlowSpec filters consume hardware resources (TCAM entries) on the associated linecards. Therefore, resources must be checked and monitored to ensure that the system operates within its scaling boundaries.

Before the activation of any FlowSpec routes, there are two ingress ACL/QoS entries consumed for IPv4 and another two entries for IPv6, as shown in the following output.

```
*A:PE-2# tools dump system-
resources | match expression "Hardware Resource|Ingress ACL/QoS Entries" post-
lines 1
Hardware Resource Usage for Slot #1, CardType iom-xp-b, Cmplx #0:
                                |   Total   | Allocated |    Free
         Ingress ACL/QoS Entries |      32768|          2|      32766
        Ingress IPv6 ACL Entries |      14336|          2|      14334
```

A FlowSpec IPv4 route matches on a source/destination IP address only. When the filter is dynamically instantiated, a single ingress ACL/QoS entry is consumed, as shown in the following output.

```
*A:PE-2# tools dump system-
resources | match expression "Hardware Resource|Ingress ACL/QoS Entries" post-
lines 1
Hardware Resource Usage for Slot #1, CardType iom-xp-b, Cmplx #0:
                                |   Total   | Allocated |    Free
         Ingress ACL/QoS Entries |      32768|          3|      32765
        Ingress IPv6 ACL Entries |      14336|          2|      14334
```

Similarly, a single entry is consumed if the source/destination IP address is used with a source/destination port. However, if ranges are used for port definition, a number of entries are consumed. A FlowSpec IPv4 route matching the source/destination IP address and a destination port range of 4190 to 4199 consumes four entries, as shown in the following output.

```
*A:PE-2# tools dump system-
resources | match expression "Hardware Resource|Ingress ACL/QoS Entries" post-
lines 1
Hardware Resource Usage for Slot #1, CardType iom-xp-b, Cmplx #0:
                                |   Total   | Allocated |    Free
         Ingress ACL/QoS Entries |      32768|          6|      32762
        Ingress IPv6 ACL Entries |      14336|          2|      14334
```

TCAM entries are not consumed on a per-interface basis. When TCAM entries are consumed on a linecard for a FlowSpec NLRI match criteria, the same criteria can be used for filtering across multiple IP interfaces on the same linecard without consuming additional TCAM entries.

Conclusion

FlowSpec IPv4 and IPv6 provide a dynamic way to activate (and tear down) ingress filters to mitigate against DDoS attacks. SR OS supports a wide range of match criteria (FlowSpec NLRI) coupled with the ability to either drop or redirect mitigated traffic. This offers flexibility not only in what traffic is matched, but also in traffic treatment, depending on the availability of a traffic-cleansing infrastructure.

The ability of FlowSpec to dynamically create and remove filters has some immediate benefits:

- Reduces the likelihood of configuration errors on one or more devices
- Allows for temporary use of hardware resources, which are released when the threat has passed
- Allows for a push configuration from a single point to a potentially large number of network devices, without having to visit each one to configure filters manually.

BGP Multipath

This chapter provides information about BGP Multipath.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4.

In release 14.0.R1, BGP multipath has been improved by offering more control over the selection of the eligible multipaths. This chapter provides information about these enhancements.

Overview

When BGP multipath is enabled, traffic can be forwarded to an IP prefix destination over multiple BGP paths that are considered equal by the BGP decision process. BGP multipath is supported in base router and VPRNs. The **multipath** command specifies the maximum number of BGP paths that each BGP RIB can submit to the route table for an IP prefix. The equal cost multipath (ECMP) limit defines how many paths are selected for installation in the forwarding information base (FIB). Traffic in the data path that matches the IP prefix is load-balanced across the ECMP next hops on a per-packet hash calculation.



Note: As described in chapter Separate BGP RIBs for Labeled Routes, labeled routes and unlabeled routes do not mix.

The BGP multipath enhancements in SR OS release 14.0.R1 are the following:

1. The **multipath** command in the base router and VPRN BGP contexts contains the following options:

   multipath *max-paths* [**ebgp** *ebgp-max-paths*] [**ibgp** *ibgp-max-paths*] [**restrict** {**same-neighbor-as** | **exact-as-path**}]

   - **max-paths** is the default maximum number of paths. It is overruled by **ebgp-max-paths** and **ibgp-max-paths**. However, if there is no maximum set for the number of eBGP paths or iBGP paths, then the maximum number of paths is set by **max-paths**.
   - **ebgp-max-paths** specifies the maximum number of paths that can be used when the best path is eBGP.
   - **ibgp-max-paths** specifies the maximum number of paths that can be used when the best path is iBGP.
   - **restrict same-neighbor-as** forces multipaths to have the same AS path length (unless **as-path-ignore** is configured) and the same neighbor AS.
   - **restrict exact-as-path** forces multipaths to have the exact same AS paths.

```
*A:PE-5# configure router bgp multipath
  - multipath <max-paths> [ebgp <ebgp-max-paths>] [ibgp <ibgp-max-paths>] [restrict
    {same-neighbor-as|exact-as-path}]
  - no multipath

 <max-paths>          : [1..16]
 <ebgp-max-paths>     : [1..16]
 <ibgp-max-paths>     : [1..16]
```

2. The **ebgp-ibgp-equal** command is added to the **best-path-selection** contexts in base router and VPRN BGP contexts. When this command is configured, as follows, the BGP decision process skips the step that prefers eBGP over iBGP. This enables load-balancing between eBGP and iBGP paths. The earlier **eibgp-loadbalance** command, from previous releases, can still be used, but it is only applicable in a VPRN, not in the base router.

```
*A:PE-5# configure router bgp best-path-selection
  - best-path-selection

 [no] always-compare* - Determine how the Multi-Exit Discriminator (MED) path
                        attribute is used in the BGP route selection process
 [no] as-path-ignore  - Determine whether the AS Path is used in determining the
                        best BGP route
 [no] compare-origin* - Enable/Disable compare validation state
 [no] deterministic-* - Enable/Disable deterministic Multi-Exit Discriminator
 [no] ebgp-ibgp-equal - Determine whether EBGP and IBGP learned paths are
                        considered equal
 [no] ignore-nh-metr* - Enable/Disable ignore next-hop metric
 [no] ignore-router-* - Enable/Disable ignore router-id
 [no] origin-invalid* - Enable/Disable origin invalid unusable routes.
```

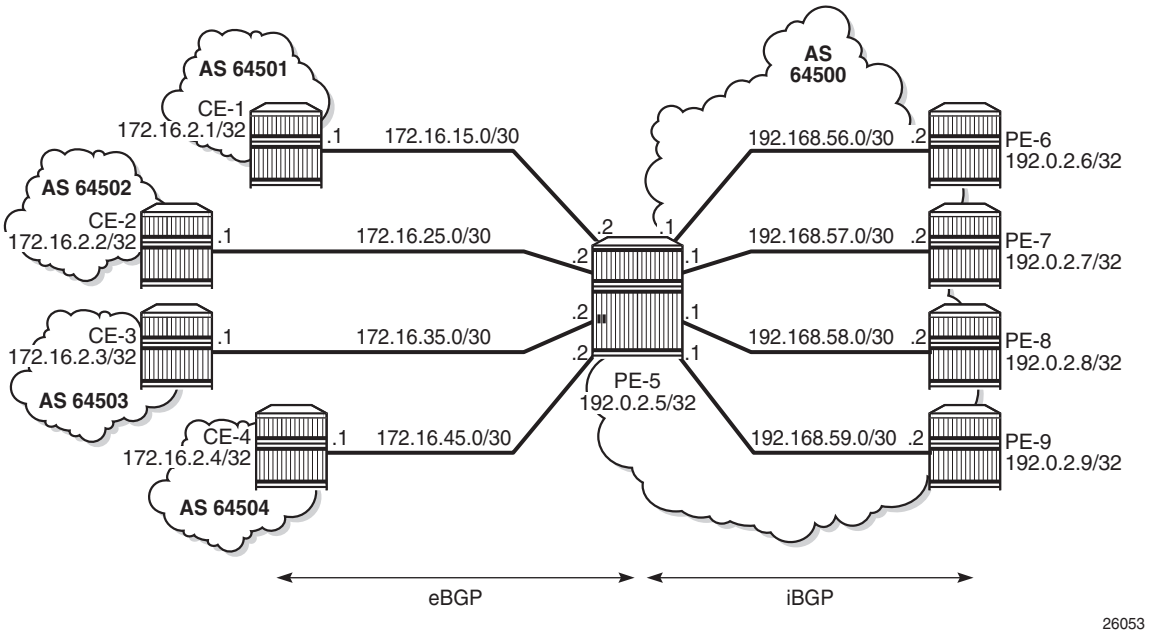
The **eibgp-loadbalance** command in a VPRN is used to provide ECMP over BGP-VPN (imported routes) and BGP routes. It is called eibgp-loadbalance because, in such scenarios, BGP-VPN is typically used between iBGP peers and BGP is used between eBGP peers. However, this is not always the case, so the name can be misleading.

Configuration

The examples in this section show the BGP configuration in the base router. For BGP multipath in a VPRN, the configuration is similar.

Figure 59 shows the example configuration with the used IP addresses. PE-5 has eBGP sessions with CEs in different autonomous systems (ASs) and iBGP sessions with PE-6, PE-7, PE-8, and PE-9.

Figure 59 **Example Topology**



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS in AS 64500
- LDP in AS 64500

- BGP on all nodes (eBGP between CEs and PE-5; iBGP between PEs)
- Export policy "export-bgp" accepting routes from protocol direct on all nodes

The BGP configuration on CE-1 is as follows:

```
configure
    router
        autonomous-system 64501
        bgp
            min-route-advertisement 1
            group "eBGP"
                export "export-bgp"
                peer-as 64500
                neighbor 172.16.15.2
                exit
            exit
        exit
    exit
exit
```

The BGP configuration on the other nodes that advertise routes to PE-5 is similar.

The BGP configuration on PE-5 is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            min-route-advertisement 1
            group "eBGP"
                neighbor 172.16.15.1
                    peer-as 64501
                exit
                neighbor 172.16.25.1
                    peer-as 64502
                exit
                neighbor 172.16.35.1
                    peer-as 64503
                exit
                neighbor 172.16.45.1
                    peer-as 64504
                exit
            exit
            group "iBGP"
                peer-as 64500
                neighbor 192.0.2.6
                exit
                neighbor 192.0.2.7
                exit
                neighbor 192.0.2.8
                exit
                neighbor 192.0.2.9
                exit
            exit
        exit
    exit
```

```
exit
```

The following will be configured and verified:

- BGP multipath with different limits for eBGP and iBGP paths
- BGP multipath with equal treatment for eBGP and iBGP paths
- BGP multipath restricted to the same neighbor AS
- BGP multipath restricted to the exact AS path
- EIBGP load-balancing in a VPRN

BGP Multipath with Different eBGP and iBGP Limits

On PE-5, BGP multipath is configured as follows:

```
*A:PE-5# configure router bgp multipath 8 ebgp 2 ibgp 3
```

It is mandatory to specify a maximum for BGP multipaths, as follows, but that is overruled by the individual limits for eBGP and iBGP. It is optional to configure limits for eBGP and iBGP.

```
*A:PE-5# configure router bgp multipath ebgp 2 ibgp 3
                                        ^
Error: Missing parameter
```

It is allowed to specify a lower value for multipath than for either eBGP or iBGP, as follows:

```
*A:PE-5# configure router bgp multipath 1 ebgp 2 ibgp 3
```

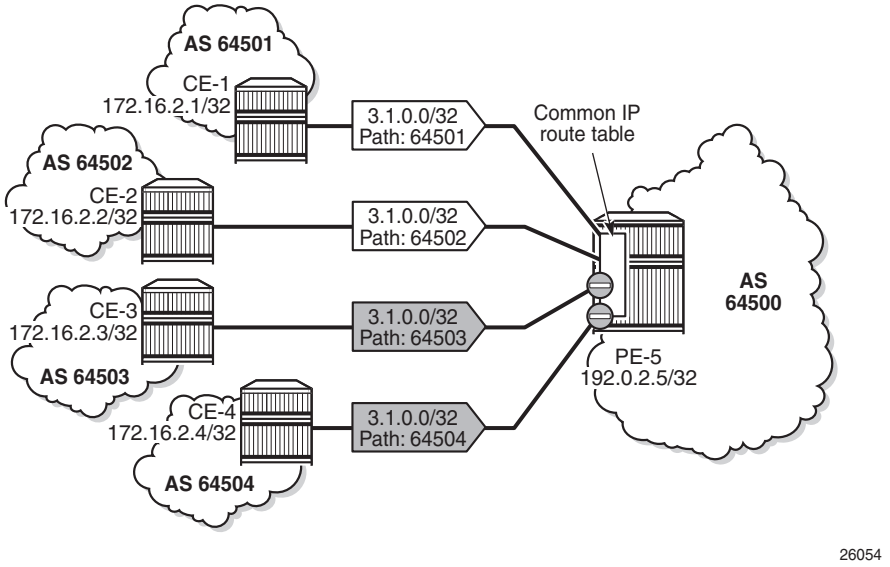
With this configuration, regardless of the value of multipath, there can be two eBGP routes for the same prefix and three iBGP routes for the same prefix. If the best route is eBGP, the multipath value is 2; if the best route is iBGP, the multipath value is 3. The value for multipath (1) is never used when limits for both eBGP and iBGP are configured.

```
*A:PE-5# configure router bgp multipath 3 ebgp 2
```

If the best route is eBGP, the multipath value is 2, and if the best route is iBGP, the multipath value is 3.

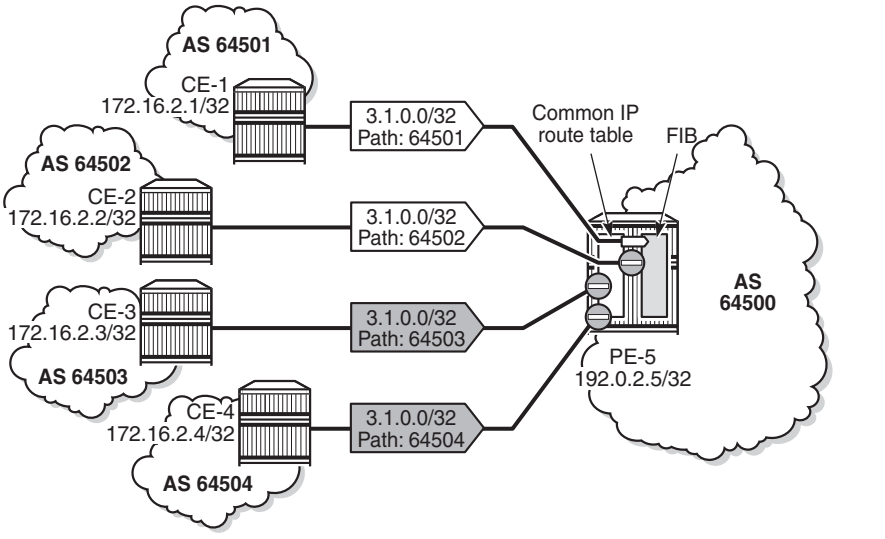
In this example, all four eBGP neighbors advertise prefix 3.1.0.0/32 to PE-5 and all four iBGP neighbors advertise prefix 3.2.0.0/32 to PE-5. PE-5 receives four eBGP routes for prefix 3.1.0.0/32, but only two will be added to the common IP route table, as shown in Figure 60.

Figure 60 **BGP Multipath with eBGP Limit 2**



These routes can only be added to the FIB if ECMP is configured to a value at least equal to the number of routes allowed in BGP multipath. By default, ECMP is disabled and only one route is added to the FIB, as shown in Figure 61.

Figure 61 **eBGP Multipath with Limit 2 and ECMP Disabled**



With ECMP disabled, only one of the four paths is active for prefix 3.1.0.0/32, as follows:

```
*A:PE-5# show router bgp routes 3.1.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.1.0.0/32                                           None        None
      172.16.15.1                                          None        -
      64501
*i    3.1.0.0/32                                           None        None
      172.16.25.1                                          None        -
      64502
*i    3.1.0.0/32                                           None        None
      172.16.35.1                                          None        -
      64503
*i    3.1.0.0/32                                           None        None
      172.16.45.1                                          None        -
      64504
-------------------------------------------------------------------------------
Routes : 4
```

In the remainder of the chapter, ECMP will be configured with a value of eight, implying that the routes added to the common IP route table will be added to the FIB as well. ECMP is configured on PE-5 as follows:

```
*A:PE-5# configure router ecmp 8
```

With ECMP configured with a limit of eight, two eBGP paths are active for prefix 3.1.0.0/32.

The first two of the following BGP routes for prefix 3.1.0.0/32 are used on PE-5:

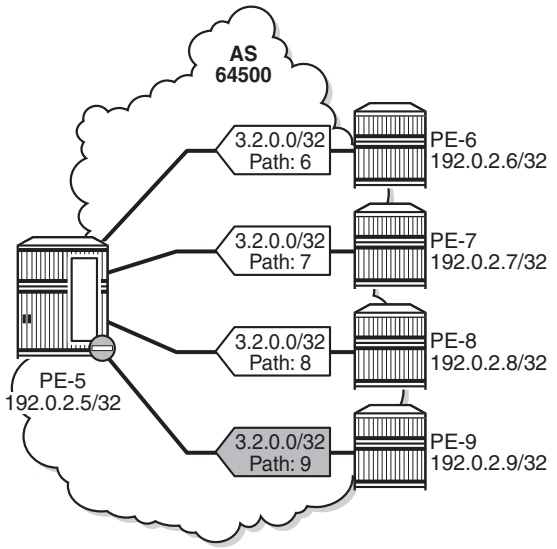
```
*A:PE-5# show router bgp routes 3.1.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
```

```
BGP IPv4 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.1.0.0/32                                       None        None
      172.16.15.1                                      None        -
      64501
u*>i  3.1.0.0/32                                       None        None
      172.16.25.1                                      None        -
      64502
*>i   3.1.0.0/32                                       None        None
      172.16.35.1                                      None        -
      64503
*>i   3.1.0.0/32                                       None        None
      172.16.45.1                                      None        -
      64504
-------------------------------------------------------------------------------
Routes : 4
```

The four iBGP neighbors of PE-5 advertise prefix 3.2.0.0/32 to PE-5. BGP multipath has a limit of three for iBGP routes. Consequently, three BGP routes will be added to the common IP route table and to the FIB, as shown in Figure 62.

Figure 62 **BGP Multipath with iBGP Limit 3 and ECMP Limit 8**



Three iBGP paths are active for prefix 3.2.0.0/32, as follows:

```
*A:PE-5# show router bgp routes 3.2.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.2.0.0/32                                           100         None
      192.0.2.6                                            None        -
      6
u*>i  3.2.0.0/32                                           100         None
      192.0.2.7                                            None        -
      7
u*>i  3.2.0.0/32                                           100         None
      192.0.2.8                                            None        -
      8
*>i   3.2.0.0/32                                           100         None
      192.0.2.9                                            None        -
      9
-------------------------------------------------------------------------------
Routes : 4
```

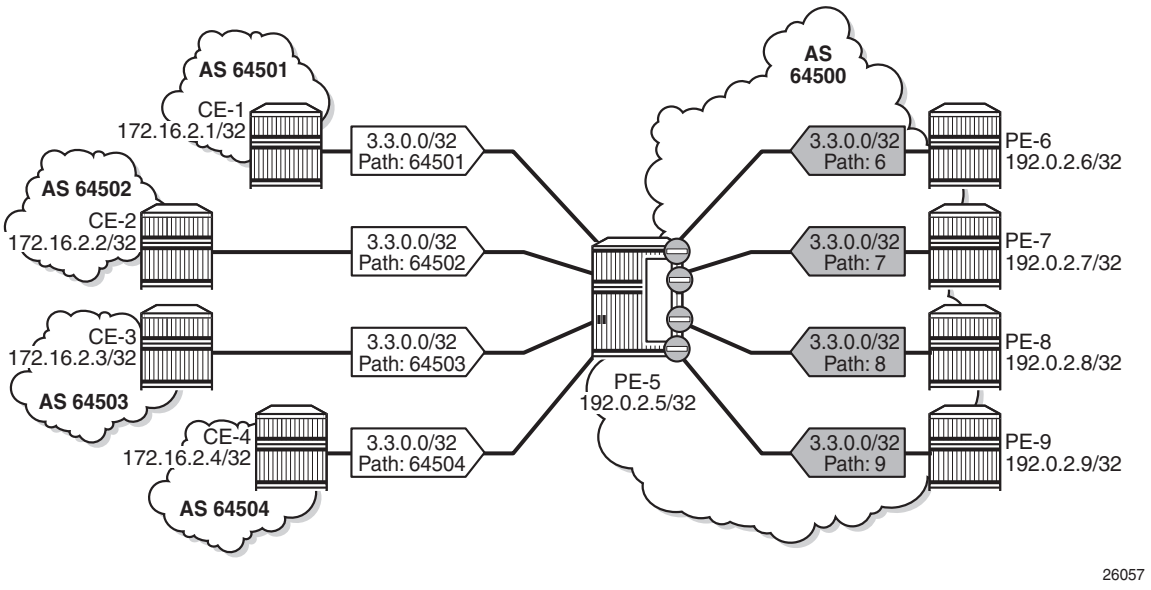
BGP Multipath with eBGP Equal to iBGP

It is optional to specify limits for eBGP and iBGP; an overall multipath limitation is sufficient, such as:

```
*A:PE-5# configure router bgp multipath 6
```

With this configuration, there can be six routes for the same prefix. These routes can be eBGP or iBGP routes. By default, eBGP routes are preferred and, therefore, only the four eBGP routes will be imported in the common IP route table, as shown in Figure 63.

Figure 63 **BGP Multipath with Limit 6 and eBGP Preferred**



Only the four eBGP paths are active, as follows:

```
*A:PE-5# show router bgp routes 3.3.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                                LocalPref   MED
      Nexthop (Router)                                       Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.3.0.0/32                                             None        None
      172.16.15.1                                            None        -
      64501
u*>i  3.3.0.0/32                                             None        None
      172.16.25.1                                            None        -
      64502
u*>i  3.3.0.0/32                                             None        None
      172.16.35.1                                            None        -
      64503
u*>i  3.3.0.0/32                                             None        None
      172.16.45.1                                            None        -
      64504
*i    3.3.0.0/32                                             100         None
      192.0.2.6                                              None        -
      6
```

```
*i    3.3.0.0/32                                          100         None
      192.0.2.7                                           None        -
      7
*i    3.3.0.0/32                                          100         None
      192.0.2.8                                           None        -
      8
*i    3.3.0.0/32                                          100         None
      192.0.2.9                                           None        -
      9
-------------------------------------------------------------------------------
Routes : 8
```

The BGP decision process prefers eBGP over iBGP, but this step can be skipped by configuring the following:

```
*A:PE-5# configure router bgp best-path-selection ebgp-ibgp-equal ipv4
```

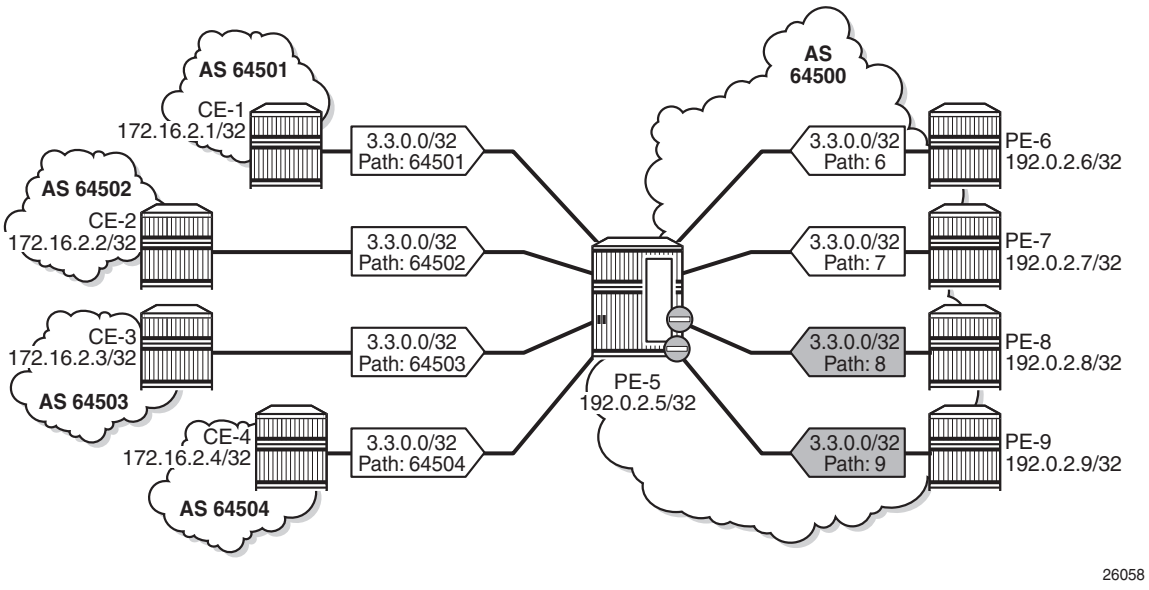
This configuration only skips one step in the BGP decision process. If the best route is still eBGP, the eBGP multipath limit applies; if the best route is iBGP, the iBGP multipath limit applies.

Optionally, other best path selection criteria can also be configured, such as ignore-nh-metric, as follows:

```
*A:PE-5# configure router bgp best-path-selection ignore-nh-metric
```

However, when all other path options are identical (such as local preference, MED, IGP cost, and other criteria from the BGP decision process), or when the best-path-selection is configured to ignore specific path options, and the only differentiator is an originator ID, the remaining steps in the BGP decision process do not exclude any routes. In that case, six of the eight eligible BGP paths will be included in the BGP multipath, as shown in Figure 64.

Figure 64 **BGP Multipath with Limit 6, eBGP Equal to iBGP, and Other Path Options Identical**



From the eight advertised BGP routes for prefix 3.3.0.0/32, six paths will be active, as follows:

```
*A:PE-5# show router bgp routes 3.3.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.3.0.0/32                                           None        None
      172.16.15.1                                          None        -
      64501
u*>i  3.3.0.0/32                                           None        None
      172.16.25.1                                          None        -
      64502
u*>i  3.3.0.0/32                                           None        None
      172.16.35.1                                          None        -
      64503
u*>i  3.3.0.0/32                                           None        None
      172.16.45.1                                          None        -
      64504
u*>i  3.3.0.0/32                                           100         None
```

```
      192.0.2.6                                                None        -
      6
u*>i  3.3.0.0/32                                               100         None
      192.0.2.7                                                None        -
      7
*>i   3.3.0.0/32                                               100         None
      192.0.2.8                                                None        -
      8
*>i   3.3.0.0/32                                               100         None
      192.0.2.9                                                None        -
      9
-------------------------------------------------------------------------------
Routes : 8
```

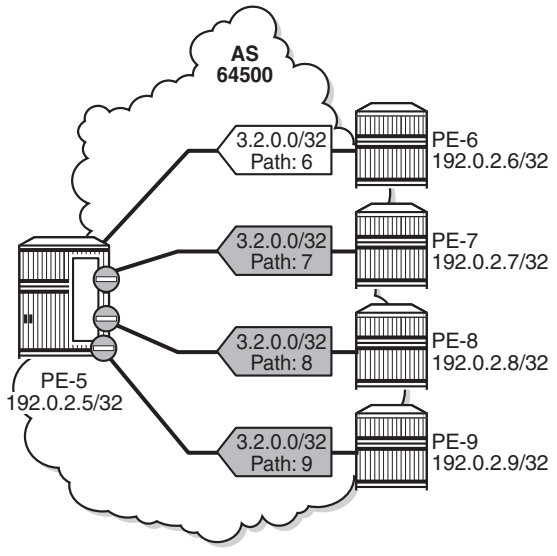
BGP Multipath Restricted to the Same Neighbor AS

BGP multipath can be configured with the restriction that the neighbor AS must be the same for all the active paths. When all routes have a different neighbor AS, only one path will be active. This can be shown for prefix 3.2.0.0/32 that is advertised by the iBGP neighbors. The BGP multipath configuration on PE-5 is as follows:

```
*A:PE-5# configure router bgp multipath 8 ebgp 2 ibgp 3 restrict same-neighbor-as
```

Figure 65 shows that with the restriction to the same neighbor AS, only one path is active because all BGP routes have a different neighbor AS.

Figure 65 **BGP Multipath Configured with Restriction to the Same Neighbor AS**



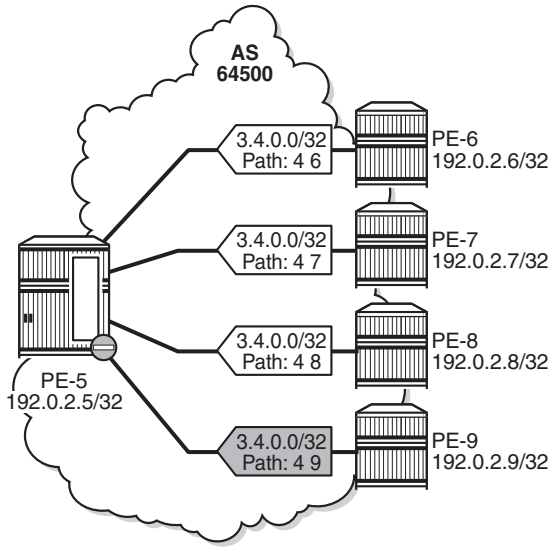
Only one BGP path will be active, because all the other routes have a different neighbor AS, as follows:

```
*A:PE-5# show router bgp routes 3.2.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                                LocalPref   MED
      Nexthop (Router)                                       Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.2.0.0/32                                             100         None
      192.0.2.6                                              None        -
      6
*>i   3.2.0.0/32                                             100         None
      192.0.2.7                                              None        -
      7
*>i   3.2.0.0/32                                             100         None
      192.0.2.8                                              None        -
      8
*>i   3.2.0.0/32                                             100         None
      192.0.2.9                                              None        -
      9
-------------------------------------------------------------------------------
Routes : 4
```

Figure 66 shows that the iBGP neighbors also advertise prefix 3.4.0.0/32 with a different AS path, but the AS path is equally long and the neighbor AS is the same. Three of these BGP paths will be active.

Figure 66 **BGP Multipath Restricted to the Same Neighbor AS: AS Paths with Same Length**



All iBGP neighbors have the same neighbor AS and an AS path of equal length. Three of the iBGP paths for prefix 3.4.0.0/32 are active, as follows:

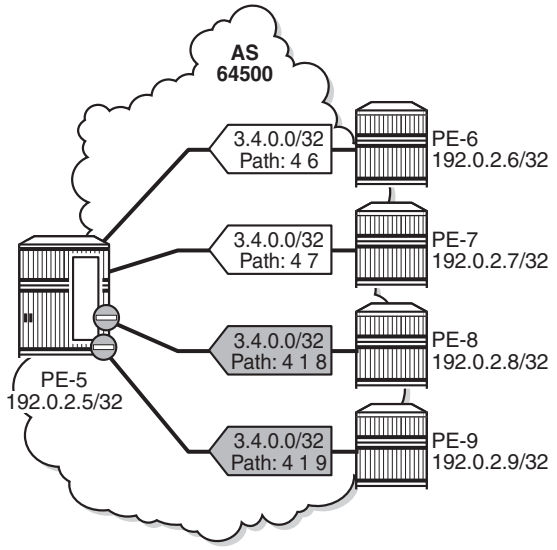
```
*A:PE-5# show router bgp routes 3.4.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.4.0.0/32                                           100         None
      192.0.2.6                                            None        -
      4 6
u*>i  3.4.0.0/32                                           100         None
      192.0.2.7                                            None        -
      4 7
u*>i  3.4.0.0/32                                           100         None
      192.0.2.8                                            None        -
      4 8
*>i   3.4.0.0/32                                           100         None
      192.0.2.9                                            None        -
```

```
      4 9
------------------------------------------------------------------------------
Routes : 4
```

The restriction that the neighbor AS must be the same does not overrule the BGP selection criterion that the shorter AS path is preferred. When the AS path is longer for the routes advertised by neighbors 192.0.2.8 and 192.0.2.9, only the BGP paths with the shorter AS path will be active, as shown in Figure 67.

Figure 67 **BGP Multipath Restricted to the Same Neighbor AS: AS Paths of Different Lengths**



All BGP routes advertised by the iBGP neighbors have the same neighbor AS, but the AS path is longer for neighbors 192.0.2.8 and 192.0.2.9. The routes advertised by these neighbors will not be selected as best path and will not be added to the route table. Only the two BGP routes with the shorter AS path will be active, as follows:

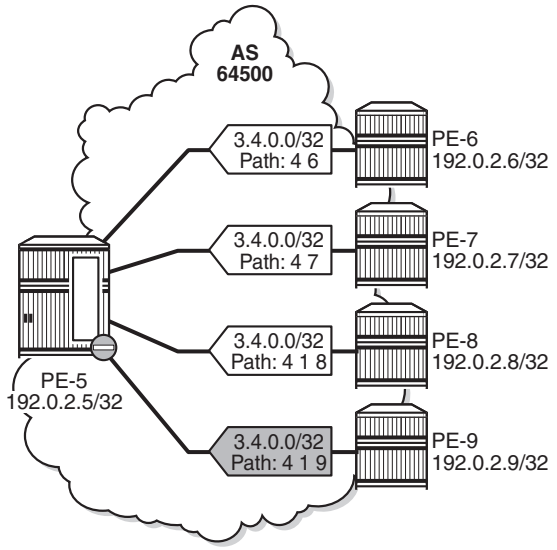
```
*A:PE-5# show router bgp routes 3.4.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
```

```
      As-Path
-------------------------------------------------------------------------------
u*>i  3.4.0.0/32                                          100         None
      192.0.2.6                                         None        -
      4 6
u*>i  3.4.0.0/32                                          100         None
      192.0.2.7                                         None        -
      4 7
*i    3.4.0.0/32                                          100         None
      192.0.2.8                                         None        -
      4 1 8
*i    3.4.0.0/32                                          100         None
      192.0.2.9                                         None        -
      4 1 9
-------------------------------------------------------------------------------
Routes : 4
```

When the best path selection is configured to ignore the AS path, three paths will be active again, as shown in Figure 68.

Figure 68 **BGP Multipath Restricted to the Same Neighbor AS: AS Paths of Different Lengths, AS Path Ignored**



The best path selection is reconfigured as follows:

```
*A:PE-5# configure router bgp best-path-selection as-path-ignore
```

Three of the four eligible BGP routes will be active, as follows:

```
*A:PE-5# show router bgp routes 3.4.0.0/32
===============================================================================
```

```
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.4.0.0/32                                       100         None
      192.0.2.6                                        None        -
      4 6
u*>i  3.4.0.0/32                                       100         None
      192.0.2.7                                        None        -
      4 7
u*>i  3.4.0.0/32                                       100         None
      192.0.2.8                                        None        -
      4 1 8
*>i   3.4.0.0/32                                       100         None
      192.0.2.9                                        None        -
      4 1 9
-------------------------------------------------------------------------------
Routes : 4
```

The best selection path settings are restored as follows:

```
*A:PE-5# configure router bgp best-path-selection no as-path-ignore
```

BGP Multipath Restricted to the Exact AS Path

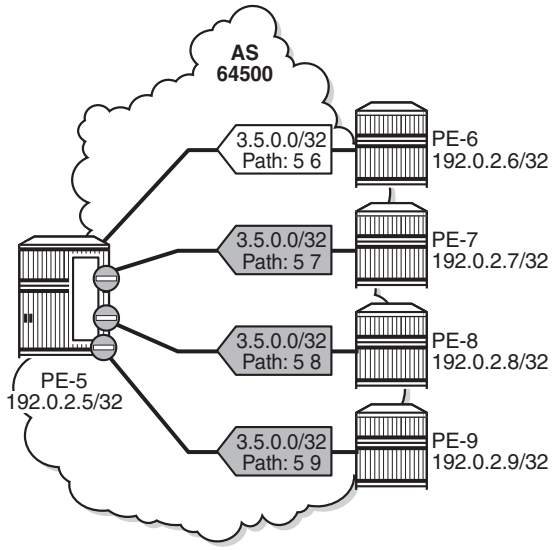
The BGP multipath configuration on PE-5 restricts BGP to only use identical AS paths, as follows:

```
*A:PE-5# configure router bgp multipath 8 ebgp 2 ibgp 3 restrict exact-as-path
```

The four iBGP neighbors advertise prefixes 3.5.0.0/32 and 3.6.0.0/32 to PE-5, see Figure 11 and Figure 12. The AS paths for prefix 3.5.0.0/32 are not identical, but the neighbor AS is the same, and the AS path is of equal length. The AS paths for prefix 3.6.0.0/32 are identical.

The BGP multipath configuration specifies that the AS paths must be identical, which is not the case for the received BGP routes for prefix 3.5.0.0/32. Only one BGP route will be imported in the route table, as shown in Figure 69.

Figure 69 **BGP Multipath Restricted to Exact Same AS. All AS Paths are Different.**



All the BGP routes for prefix 3.5.0.0/32 have a different AS path. Only the BGP route advertised by neighbor 192.0.2.6 will be active, as follows:

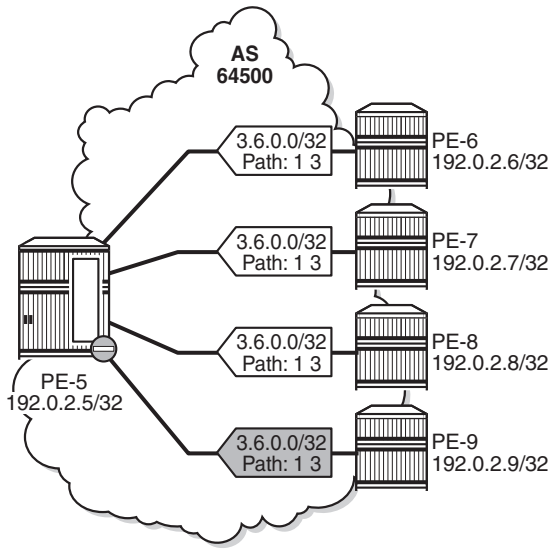
```
*A:PE-5# show router bgp routes 3.5.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.5.0.0/32                                           100         None
      192.0.2.6                                            None        -
      5 6
*>i   3.5.0.0/32                                           100         None
      192.0.2.7                                            None        -
      5 7
*>i   3.5.0.0/32                                           100         None
      192.0.2.8                                            None        -
      5 8
*>i   3.5.0.0/32                                           100         None
      192.0.2.9                                            None        -
```

```
      5 9
------------------------------------------------------------------------------
Routes : 4
```

However, all the received BGP routes for prefix 3.6.0.0/32 have the same AS path. Three of these BGP paths will become active, as shown in Figure 70.

Figure 70 **BGP Multipath Restricted to Exact Same AS. All AS Paths are Identical**



Three of the four received BGP routes for prefix 3.6.0.0/32 are used, as follows:

```
*A:PE-5# show router bgp routes 3.6.0.0/32
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.6.0.0/32                                       100         None
      192.0.2.7                                        None        -
      1 3
u*>i  3.6.0.0/32                                       100         None
```

```
      192.0.2.8                                              None        -
      1 3
u*>i  3.6.0.0/32                                             100         None
      192.0.2.9                                              None        -
      1 3
*i    3.6.0.0/32                                             100         None
      192.0.2.6                                              None        -
      1 3
-------------------------------------------------------------------------------
Routes : 4
```

EIBGP Load-Balancing in a VPRN

The **eibgp-loadbalance** command is used to perform ECMP over BGP-VPN (imported routes) and BGP routes, not to make eBGP routes equal to iBGP routes. Typically, BGP-VPN routes are used between iBGP peers and BGP routes are used between eBGP peers. This explains the naming of the command, but it can be misleading. Not all BGP-VPN routes are iBGP routes and not all BGP routes are eBGP routes. Four different types of routes are received by PE-5 in the following example:

- two BGP routes learned from eBGP peers CE-1 and CE-2
- two BGP-VPN routes learned from eBGP peers CE-3 and CE-4
- two BGP routes learned from iBGP peers PE-6 and PE-7
- two BGP-VPN routes learned from iBGP peers PE-8 and PE-9

VPRN 1 is configured on all nodes. On PE-5, ECMP must be configured in this VPRN to a value of 8, as follows:

```
*A:PE-5# configure service vprn 1 ecmp 8
```

In this example, the best path selection criteria will be such that iBGP routes are treated as equal to eBGP routes, as explained in section BGP Multipath with eBGP Equal to iBGP. This is configured for BGP routes and for VPN-BGP routes, as follows:

```
*A:PE-5# configure router bgp best-path-selection ebgp-ibgp-equal vpn-ipv4
*A:PE-5# configure service vprn 1 bgp best-path-selection ebgp-ibgp-equal ipv4
```

Optionally, other best path selection criteria can also be configured, such as ignore-nh-metric and as-path-ignore, as follows:

```
*A:PE-5# configure router bgp best-path-selection as-path-ignore vpn-ipv4
*A:PE-5# configure service vprn 1 bgp best-path-selection as-path-ignore ipv4

*A:PE-5# configure router bgp best-path-selection ignore-nh-metric
*A:PE-5# configure service vprn 1 bgp best-path-selection ignore-nh-metric
```

BGP multipath is configured with an eBGP multipath limit of 4 and an iBGP multipath limit of 4, as follows:

```
*A:PE-5# configure service vprn 1 bgp multipath 4 ebgp 4 ibgp 4
```

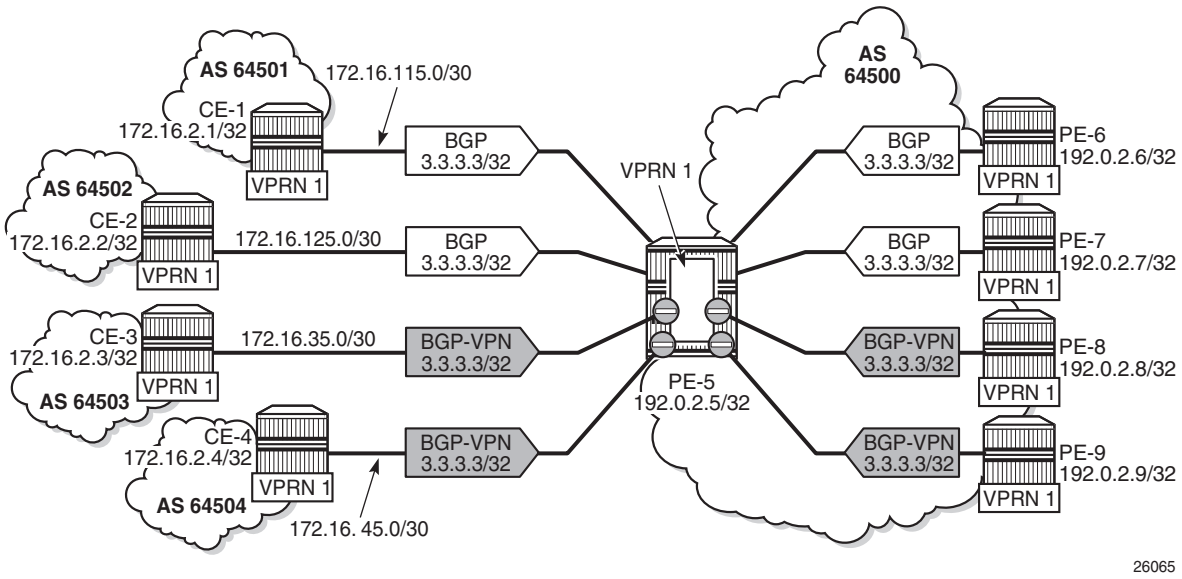
The **eibgp-loadbalance** command is not configured yet. By default, the BGP decision process prefers BGP routes over BGP-VPN routes, and therefore, the routing table only contains BGP routes and no BGP-VPN routes, as follows:

```
*A:PE-5# show router 1 route-table 3.3.3.3/32
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
3.3.3.3/32                                      Remote  BGP       00h01m11s  170
       172.16.115.1                                               0
3.3.3.3/32                                      Remote  BGP       00h01m11s  170
       172.16.125.1                                               0
3.3.3.3/32                                      Remote  BGP       00h01m11s  170
       192.168.156.2                                              0
3.3.3.3/32                                      Remote  BGP       00h01m11s  170
       192.168.157.2                                              0
-------------------------------------------------------------------------------
No. of Routes: 4
```

The first two routes in this routing table are eBGP routes learned from CE-1 and CE-2. The latter two routes are iBGP routes learned from PE-6 and PE-7. EBGP routes and iBGP routes are treated as equal, but BGP routes are still preferred to BGP-VPN routes.

Figure 71 shows that the BGP routes are preferred. Even though it is allowed to have a maximum of four eBGP routes and four iBGP routes, only two eBGP routes are included and only two iBGP routes are included. No BGP-VPN routes are used, as follows.

Figure 71 **EBGP Equal to IBGP: No EIBGP Load-Balancing**



BGP-VPN routes will be treated equally to BGP routes after configuring the following command:

```
*A:PE-5# configure service vprn 1 bgp eibgp-loadbalance
```

When BGP-VPN routes are treated equal to BGP routes, the routing table for prefix 3.3.3.3/32 contains eight entries, as follows:

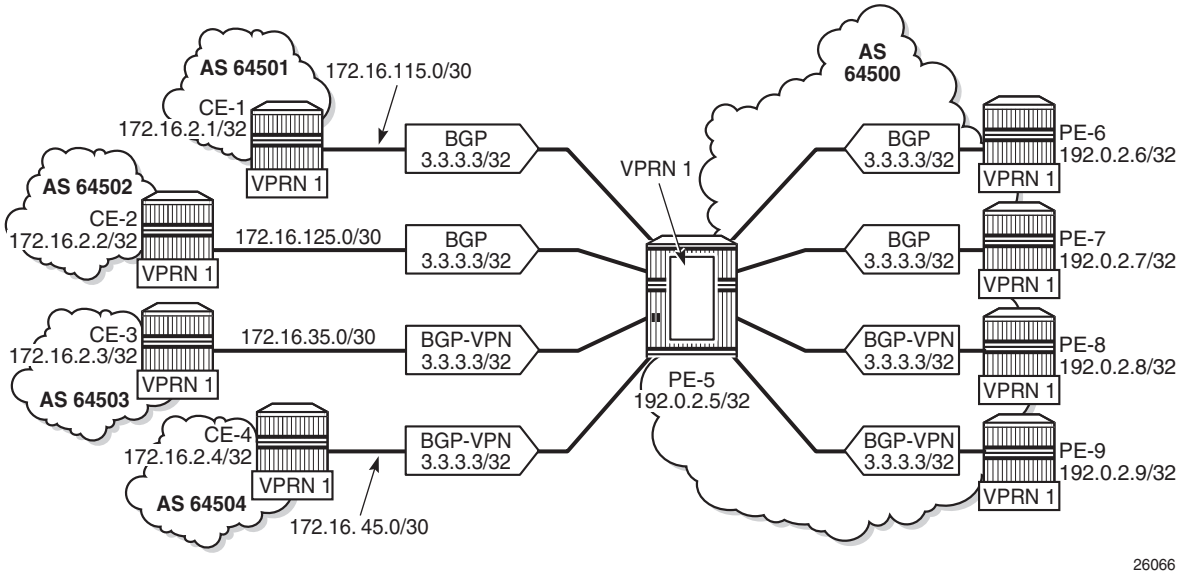
```
*A:PE-5# show router 1 route-table 3.3.3.3/32
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
3.3.3.3/32                                      Remote  BGP VPN   00h00m04s  170
       172.16.35.1 (tunneled)                                          0
3.3.3.3/32                                      Remote  BGP VPN   00h00m04s  170
       172.16.45.1 (tunneled)                                          0
3.3.3.3/32                                      Remote  BGP       00h00m04s  170
       172.16.115.1                                                    0
3.3.3.3/32                                      Remote  BGP       00h00m04s  170
       172.16.125.1                                                    0
3.3.3.3/32                                      Remote  BGP VPN   00h00m04s  170
       192.0.2.8 (tunneled)                                            0
3.3.3.3/32                                      Remote  BGP VPN   00h00m04s  170
       192.0.2.9 (tunneled)                                            0
3.3.3.3/32                                      Remote  BGP       00h00m04s  170
       192.168.156.2                                                   0
3.3.3.3/32                                      Remote  BGP       00h00m04s  170
       192.168.157.2                                                   0
```

```
-------------------------------------------------------------------------------
No. of Routes: 8
```

The first two BGP-VPN entries are eBGP routes learned from CE-3 and CE-4 and the other two BGP-VPN entries are iBGP routes learned from PE-8 and PE-9.

Figure 72 shows that when EIBGP load-balancing is configured, the BGP-VPN routes are equal to the BGP routes. All other selection criteria are also equal in this example and all eight routes are used in the routing table, as follows.

Figure 72 **EBGP Equal to IBGP: EIBGP Load-Balancing Enabled**



This example shows that the **eibgp-loadbalance** command can be used in combination with **best-path-selection ebgp-ibgp-equal**, because their scope is different.

Conclusion

BGP multipath allows the IP routing table to have multiple BGP paths to the same destination. Different path limits can be applied for eBGP and iBGP paths. It is possible to treat eBGP and iBGP routes as equal. Restrictions can be imposed related to AS path.

BGP Route Leaking

This chapter provides information about BGP route leaking.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4.

Overview

Route leaking refers to the process of copying a route from one router context to another.

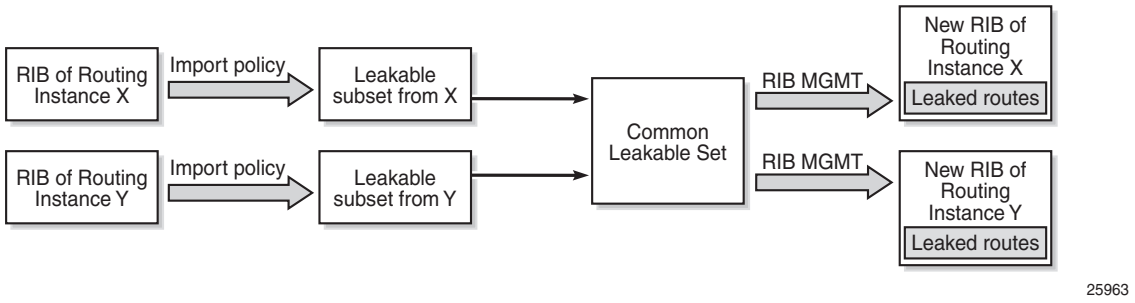
Network administrators may need to leak routes between routing instances in the same SR OS router. BGP route leaking is an alternative to using import/export policies based on communities to exchange routes between virtual router and forwarders (VRFs).

It is possible to leak a copy of a BGP route (including all its path attributes) from one routing instance to another in the same SR OS router. This BGP route leaking capability applies to IPv4, IPv6, and label-IPv4 routes. Leaking is supported from the GRT to a VPRN, from one VPRN to another VPRN, and from a VPRN to the GRT. Any valid BGP route for an IPv4 or IPv6 prefix can be leaked. A BGP route does not have to be the best path or used for forwarding in the source instance in order to be leaked, but it does have to be valid (that is, the next-hop must be resolved; the AS PATH must not exhibit a loop, for example).

An IPv4 or IPv6 BGP route becomes a candidate for leaking to another instance when it is specially marked by a BGP import policy. This special marking is achieved by accepting the route with a **bgp-leak** action in the route policy. Routes that are candidates for leaking to other instances show a **leakable** flag in the output of various show router bgp commands.

To copy a leakable BGP route from a source instance into the BGP RIB of a target instance, the target instance must be configured with a leak-import policy that matches and accepts the leakable route. There are separate leak-import policies for IPv4 and IPv6 routes. Up to 15 leak-import policies can be chained together for more complex examples. In the target instance, the **show router bgp routes** command displays leaked BGP RIB-IN routes in addition to direct RIB-IN routes learned from neighbors of the routing instance. A **leaked** flag is added to the leaked RIB-IN entries. Figure 73shows the process of BGP route leaking.

Figure 73 **BGP Route Leaking Process**



Leaked BGP routes can be advertised to BGP neighbors (peers) of the target routing instance. The BGP next hop of a leaked route is automatically reset to self whenever it is advertised to a peer of the target instance. Normal route advertisement rules apply: by default, the leaked route is advertised if it is the overall best path that is used as the active route to the destination and it is not blocked by the IBGP-to-IBGP split-horizon rule.

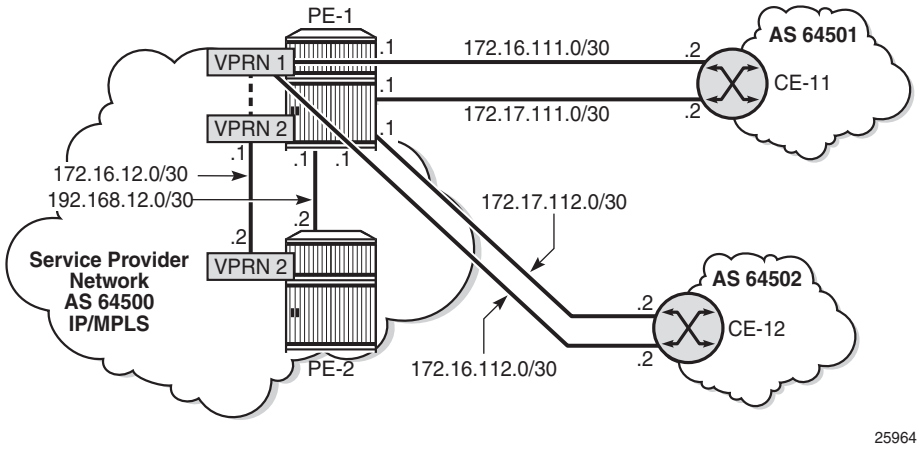
A BGP route leaked into a VPRN can be exported from the VPRN as a VPN-IPv4/v6 route if it matches the VRF export policy. Normal VPN export rules apply: by default, the leaked route is exported if it is the overall best path and it is used as the active route to the destination.

This chapter describes BGP route leaking only. For other routes, such as IS-IS, OSPF, RIP, and static routes, VPRN route leaking mechanisms apply that are protocol independent, see chapter Traffic Leaking from VPRN to GRT.

Configuration

Figure 74 shows the example topology used in this chapter, including the IPv4 addresses. For each of the examples, a dedicated figure will show the specific topology, which is a subset of the topology in Figure 2. The interfaces also have IPv6 addresses, which will be shown in Figure 6 and Figure 7. VPRN 2 also has CEs attached, but for simplicity, these are not shown on the figures and no CLI will be shown for any CE.

Figure 74 **Example Topology**



The following examples will be explained:

- Example 1 - BGP IPv4 Route Leaking between VPRNs. Global BGP Policy
- Example 2 - BGP IPv4 Route Leaking between VPRNs per Neighbor
- Example 3 - BGP IPv4 Route Leaking from VPRN to GRT per BGP Group
- Example 4 - BGP IPv4 Route Leaking from GRT to VPRN per Neighbor
- Example 5 - BGP IPv6 Route Leaking between VPRNs. Global VPRN BGP Configuration
- Example 6 - BGP IPv6 Route Leaking from GRT to VPRN and from VPRN to VPRN

Initial Configuration

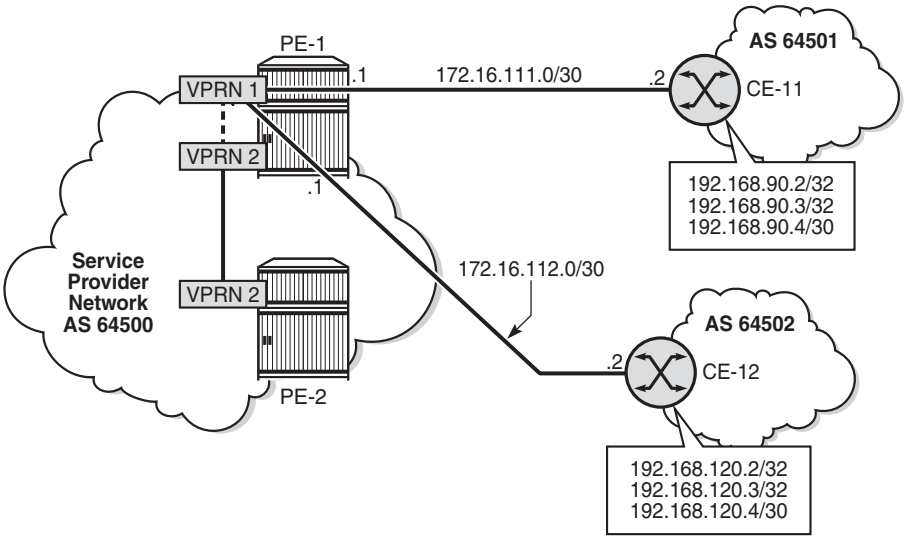
The nodes in the example topology have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces
- IGP (IS-IS or OSPF) between the PEs
- LDP between the PEs
- VPRN 1 on PE-1; VPRN 2 on PE-1 and PE-2
- BGP (IBGP between the PEs; EBGP between PE-1 and the CEs)
  - On the PEs, BGP is configured in the base router and in the VPRNs.
- Loopback addresses and black-hole static routes in the CEs. Different routes are exported to GRT and VPRN 1 on PE-1

Example 1 - BGP IPv4 Route Leaking between VPRNs. Global BGP Policy

Figure 75 shows the topology for this example. CE-11 exports routes such as 192.168.90.2/32 to VPRN 1 on PE-1, and CE-12 exports routes such as 192.168.120.2/32 to VPRN 1 on PE-1.

Figure 75 **BGP IPv4 Route Leaking Between VPRNs**



BGP leaking is disabled by default. The routing table for VPRN 1 on PE-1 includes routes that are learned from CE-11 and CE-12, as shown:

```
*A:PE-1# show router 1 route-table
```

```
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.1/32                                   Local   Local     00h45m08s  0
       system                                                       0
172.16.111.0/30                                 Local   Local     00h45m08s  0
       int-PE-1-CE-11                                               0
172.16.112.0/30                                 Local   Local     00h45m08s  0
       int-PE-1-CE-12                                               0
192.168.90.2/32                                 Remote  BGP       00h22m40s  170
       172.16.111.2                                                 0
192.168.90.3/32                                 Remote  BGP       00h22m40s  170
       172.16.111.2                                                 0
192.168.90.4/30                                 Remote  BGP       00h22m40s  170
       172.16.111.2                                                 0
192.168.120.2/32                                Remote  BGP       00h20m57s  170
       172.16.112.2                                                 0
192.168.120.3/32                                Remote  BGP       00h20m57s  170
       172.16.112.2                                                 0
192.168.120.4/32                                Remote  BGP       00h20m57s  170
       172.16.112.2                                                 0
-------------------------------------------------------------------------------
No. of Routes: 9
```

These BGP routes are not leakable, by default, as follows:

```
*A:PE-1# show router 1 bgp routes ipv4 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
```

The routing table for VPRN 2 does not include any of these routes because BGP route leaking is disabled by default:

```
*A:PE-1# show router 2 route-table
===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
```

```
-------------------------------------------------------------------------------
172.16.2.1/32                                 Local   Local     00h47m19s  0
       system                                                        0
172.16.12.0/30                                Local   Local     00h47m19s  0
       int-PE-1-PE-2_VPN2                                            0
-------------------------------------------------------------------------------
No. of Routes: 2
```

To configure BGP route leaking, an import policy is required in VPRN 1. The BGP route leaking policy is configured on PE-1, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "BGP-Leak-Policy"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

By adding the action accept **bgp-leak**, BGP routes are imported and marked as BGP leakable, meaning they are available to be copied - with their complete set of BGP path attributes - to the BGP RIB-IN of another routing instance.

The BGP route leaking policy can be applied in VPRN 1 in the general BGP configuration (as is the case here), in the group context, or per neighbor:

```
configure
    service
        vprn 1
            bgp
                import "BGP-Leak-Policy"
            exit
        exit
    exit
exit
```

With the preceding configuration, SR OS is marking all the BGP routes imported into the VPRN as leakable. The BGP routes originate from CE-11 or CE-12 in this example.

The following command shows which BGP routes in VPRN 1 are marked as leakable:

```
*A:PE-1# show router 1 bgp routes ipv4 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.168.90.2/32                                   None        None
      172.16.111.2                                      None        -
      64501
u*>i  192.168.90.3/32                                   None        None
      172.16.111.2                                      None        -
      64501
u*>?  192.168.90.4/30                                   None        None
      172.16.111.2                                      None        -
      64501
u*>i  192.168.120.2/32                                  None        None
      172.16.112.2                                      None        -
      64502
u*>i  192.168.120.3/32                                  None        None
      172.16.112.2                                      None        -
      64502
u*>?  192.168.120.4/32                                  None        None
      172.16.112.2                                      None        -
      64502
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

The routes learned from CE-11 and CE-12 are leakable. The detailed output for any route in the preceding list shows the flag "leakable". The route source is external because the routes are imported (from CE-11 or CE-12):

```
*A:PE-1# show router 1 bgp routes 192.168.90.2/32 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 192.168.90.2/32
Nexthop        : 172.16.111.2
Path Id        : None
From           : 172.16.111.2
Res. Nexthop   : 172.16.111.2
---snip---
```

```
Flags          : Used  Valid  Best  IGP  Leakable
Route Source   : External
---snip---
```

BGP leakable routes can be imported into another VPRN. Prefix lists can be used to filter specific routes for BGP leaking, but that is not configured in this example. The following import policy is configured on PE-1 to import BGP leakable routes:

```
configure
    router
        policy-options
            begin
            policy-statement "Import-Leakable-Routes"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

In each of the examples, the same import policy will be used. The import policy to import BGP leakable routes is applied in the VPRN 2 on PE-1 as follows:

```
configure
    service
        vprn 2
            bgp
                rib-management
                    ipv4
                        leak-import "Import-Leakable-Routes"
                    exit
                exit
            exit
        exit
    exit
exit
```

The following command shows that VPRN 2 imported leaked BGP routes from VPRN 1. The status code "l" indicates that the route is leaked.

```
*A:PE-1# show router 2 bgp routes ipv4 leaked
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
```

```
      Nexthop (Router)                                        Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>li 192.168.90.2/32                                         100         None
      172.16.111.2 (VPRN 1)                                   None        -
      64501
u*>li 192.168.90.3/32                                         100         None
      172.16.111.2 (VPRN 1)                                   None        -
      64501
u*>l? 192.168.90.4/30                                         100         None
      172.16.111.2 (VPRN 1)                                   None        -
      64501
u*>li 192.168.120.2/32                                        100         None
      172.16.112.2 (VPRN 1)                                   None        -
      64502
u*>li 192.168.120.3/32                                        100         None
      172.16.112.2 (VPRN 1)                                   None        -
      64502
u*>l? 192.168.120.4/32                                        100         None
      172.16.112.2 (VPRN 1)                                   None        -
      64502
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

The flags in the detailed output for a particular leaked BGP route from the preceding list include the flag "leaked". The route source for this leaked route is VPRN 1 and all BGP attributes are preserved, as shown:

```
*A:PE-1# show router 2 bgp routes 192.168.90.2/32 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 192.168.90.2/32
Nexthop        : 172.16.111.2 (VPRN 1)
Path Id        : None
From           : ::
Res. Nexthop   : 172.16.111.2
Local Pref.    : 100                    Interface Name : int-PE-1-CE-11
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 0.0.0.0
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP  Leaked
```

```
Route Source   : Leaked from VPRN 1
AS-Path        : 64501
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h00m07s
---snip---
```

The route-table for VPRN 2 in the neighbor PE-2 contains the leaked routes, as shown:

```
*A:PE-2# show router 2 route-table
===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.2.1/32                                   Remote  BGP VPN   00h40m26s  170
       192.0.2.1 (tunneled)                                          0
172.16.2.2/32                                   Local   Local     00h40m58s  0
       system                                                        0
172.16.12.0/30                                  Local   Local     00h40m58s  0
       int-PE-2-PE-1_VPN2                                            0
192.168.90.2/32                                 Remote  BGP       00h00m00s  170
       172.16.12.1                                                   0
192.168.90.3/32                                 Remote  BGP       00h00m00s  170
       172.16.12.1                                                   0
192.168.90.4/30                                 Remote  BGP       00h00m00s  170
       172.16.12.1                                                   0
192.168.120.2/32                                Remote  BGP       00h00m00s  170
       172.16.12.1                                                   0
192.168.120.3/32                                Remote  BGP       00h00m00s  170
       172.16.12.1                                                   0
192.168.120.4/32                                Remote  BGP       00h00m00s  170
       172.16.12.1                                                   0
-------------------------------------------------------------------------------
No. of Routes: 9
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

Example 2 - BGP IPv4 Route Leaking between VPRNs per Neighbor

The topology used for this example is the same as for Example 1; see Figure 75. Both CEs export the same routes as in the preceding example, and the BGP route leaking policy is identical:

```
configure
    router
        policy-options
            begin
            policy-statement "BGP-Leak-Policy"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

In the preceding example, the BGP route leaking policy was applied in the global BGP context in VPRN 1 and consequently, it applied to routes from all neighbors. In this example, the BGP route leaking policy is applied in VPRN 1 for neighbor CE-11 only, as follows:

```
configure
    service
        vprn 1
            bgp
                group "EBGP_64500to64501_IPv4"
                    neighbor 172.16.111.2
                        import "BGP-Leak-Policy"
                    exit
                exit
            exit
        exit
    exit
exit
```

This import policy implies that only routes learned from CE-11 will be leakable. The following command shows all the BGP routes learned in VPRN 1 on PE-1. Not all of these are leakable.

```
*A:PE-1# show router 1 bgp routes
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.168.90.2/32                                   None        None
      172.16.111.2                                      None        -
      64501
u*>i  192.168.90.3/32                                   None        None
      172.16.111.2                                      None        -
      64501
u*>?  192.168.90.4/30                                   None        None
      172.16.111.2                                      None        -
      64501
u*>i  192.168.120.2/32                                  None        None
      172.16.112.2                                      None        -
      64502
u*>i  192.168.120.3/32                                  None        None
      172.16.112.2                                      None        -
      64502
u*>?  192.168.120.4/32                                  None        None
      172.16.112.2                                      None        -
      64502
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

Some routes are learned from CE-11 and other routes are learned from CE-12, and only the routes imported from CE-11 are leakable. The following command shows which routes are marked as leakable in the RT of the VPRN:

```
*A:PE-1# show router 1 bgp routes ipv4 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.168.90.2/32                                   None        None
      172.16.111.2                                      None        -
      64501
u*>i  192.168.90.3/32                                   None        None
      172.16.111.2                                      None        -
      64501
u*>?  192.168.90.4/30                                   None        None
```

```
      172.16.111.2                                          None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The BGP leakable routes can be imported into another VPRN instance. The import policy is the same as for Example 1:

```
configure
    router
        policy-options
            begin
            policy-statement "Import-Leakable-Routes"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

This import policy is applied in VPRN 2 in the same way as in Example 1:

```
configure
    service
        vprn 2
            bgp
                rib-management
                    ipv4
                        leak-import "Import-Leakable-Routes"
                    exit
                exit
            exit
        exit
    exit
exit
```

The following command shows the leaked routes in VPRN 2. Each of these routes is leaked from VPRN 1, as indicated between brackets in the following output. Only routes learned from CE-11 in VPRN 1 are leaked to VPRN 2.

```
*A:PE-1# show router 2 bgp routes ipv4 leaked
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
```

```
Flag  Network                                                 LocalPref   MED
      Nexthop (Router)                                        Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>li 192.168.90.2/32                                       100         None
      172.16.111.2 (VPRN 1)                                 None        -
      64501
u*>li 192.168.90.3/32                                       100         None
      172.16.111.2 (VPRN 1)                                 None        -
      64501
u*>l? 192.168.90.4/30                                       100         None
      172.16.111.2 (VPRN 1)                                 None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The detailed output for any of these BGP routes shows that the flag "leaked" is set and that the route source corresponds to VPRN 1, as shown for route 192.168.90.2/32:

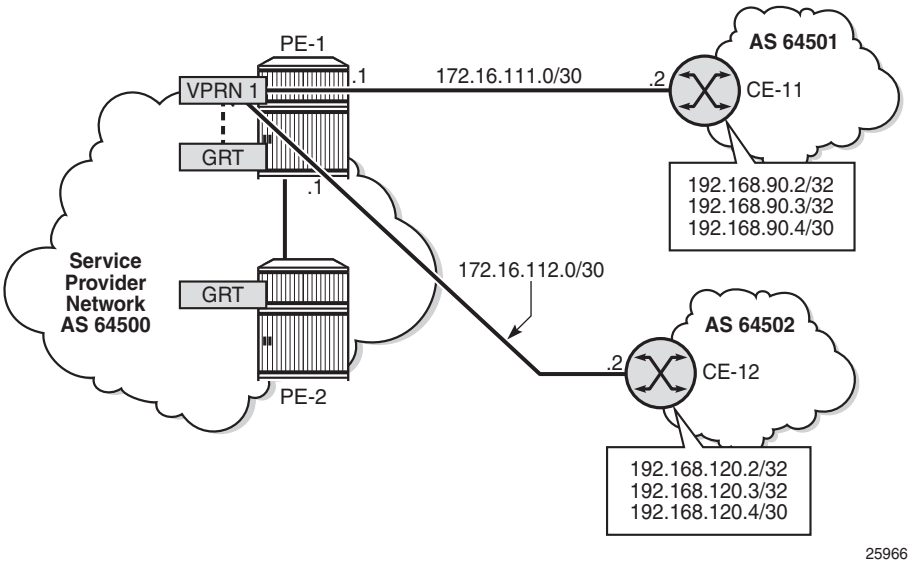
```
*A:PE-1# show router 2 bgp routes 192.168.90.2/32 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 192.168.90.2/32
Nexthop        : 172.16.111.2 (VPRN 1)
---snip---
Flags          : Used  Valid  Best  IGP  Leaked
Route Source   : Leaked from VPRN 1
```

Example 3 - BGP IPv4 Route Leaking from VPRN to GRT per BGP Group

Figure 76 shows the topology for this example. CE-11 and CE-12 export the same routes to VPRN 1. These routes will be marked as leakable and leaked to the GRT.

Figure 76 **BGP IPv4 Route Leaking from VPRN to GRT**



The routing table for VPRN 1 in PE-1 contains the BGP routes exported by CE-11 and CE-12, as follows:

```
*A:PE-1# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                        Metric
-------------------------------------------------------------------------------
172.16.1.1/32                                     Local   Local     01h03m55s  0
       system                                                          0
172.16.111.0/30                                   Local   Local     01h03m55s  0
       int-PE-1-CE-11                                                  0
172.16.112.0/30                                   Local   Local     01h03m55s  0
       int-PE-1-CE-12                                                  0
192.168.90.2/32                                   Remote  BGP       00h00m34s  170
       172.16.111.2                                                    0
192.168.90.3/32                                   Remote  BGP       00h00m34s  170
       172.16.111.2                                                    0
192.168.90.4/30                                   Remote  BGP       00h00m34s  170
       172.16.111.2                                                    0
192.168.120.2/32                                  Remote  BGP       00h32m53s  170
       172.16.112.2                                                    0
192.168.120.3/32                                  Remote  BGP       00h32m53s  170
       172.16.112.2                                                    0
192.168.120.4/32                                  Remote  BGP       00h32m53s  170
       172.16.112.2                                                    0
-------------------------------------------------------------------------------
No. of Routes: 9
```

The routing table of the base router does not include any of the BGP routes exported by the CEs, as follows:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.17.111.0/30                                 Local   Local     01h16m32s  0
       int-PE-1-CE-11                                               0
172.17.112.0/30                                 Local   Local     01h16m32s  0
       int-PE-1-CE-12                                               0
192.0.2.1/32                                    Local   Local     01h16m32s  0
       system                                                       0
192.0.2.2/32                                    Remote  ISIS      01h16m14s  15
       192.168.12.2                                                 10
192.168.12.0/30                                 Local   Local     01h16m32s  0
       int-PE-1-PE-2                                                0
-------------------------------------------------------------------------------
No. of Routes: 5
```

The BGP routes are marked as leakable after applying the following configuration:

```
configure
    router
        policy-options
            begin
            policy-statement "BGP-Leak-Policy"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

This BGP route leaking policy can be applied in the general BGP configuration of VPRN 1, or per BGP group (as is the case here), or per BGP neighbor:

```
configure
    service
        vprn 1
            bgp
                group "EBGP_64500to64501_IPv4"
                    import "BGP-Leak-Policy"
                exit
            exit
        exit
    exit
exit
```

The following command shows the leakable BGP routes in VPRN 1:

```
*A:PE-1# show router 1 bgp routes ipv4 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.168.90.2/32                                      None        None
      172.16.111.2                                         None        -
      64501
u*>i  192.168.90.3/32                                      None        None
      172.16.111.2                                         None        -
      64501
u*>?  192.168.90.4/30                                      None        None
      172.16.111.2                                         None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The leakable BGP routes in VPRN 1 can be imported into the GRT. The import policy is identical to the import policy in the preceding examples, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "Import-Leakable-Routes"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

This import policy is applied in the base router, as follows:

```
configure
    router
        bgp
            rib-management
                ipv4
                    leak-import "Import-Leakable-Routes"
```

```
            exit
          exit
        exit
    exit
exit
```

As a result, the leakable BGP routes in VPRN 1 are leaked to the GRT, as follows:

```
*A:PE-1# show router bgp routes ipv4 leaked
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>li 192.168.90.2/32                                   100         None
      172.16.111.2 (VPRN 1)                             None        -
      64501
u*>li 192.168.90.3/32                                   100         None
      172.16.111.2 (VPRN 1)                             None        -
      64501
u*>l? 192.168.90.4/30                                   100         None
      172.16.111.2 (VPRN 1)                             None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The detailed information for any of these leaked routes shows that the flag "leaked" is present and that the route source is VPRN 1, as follows:

```
*A:PE-1# show router bgp routes 192.168.90.2/32 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 192.168.90.2/32
Nexthop        : 172.16.111.2 (VPRN 1)
---snip---
Flags          : Used  Valid  Best  IGP  Leaked
```

```
Route Source   : Leaked from VPRN 1
---snip---
```

The GRT includes the leaked routes, as follows:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.17.111.0/30                                 Local   Local     01h34m47s   0
       int-PE-1-CE-11                                               0
172.17.112.0/30                                 Local   Local     01h34m47s   0
       int-PE-1-CE-12                                               0
192.0.2.1/32                                    Local   Local     01h34m47s   0
       system                                                       0
192.0.2.2/32                                    Remote  ISIS      01h34m29s   15
       192.168.12.2                                                 10
192.168.12.0/30                                 Local   Local     01h34m47s   0
       int-PE-1-PE-2                                                0
192.168.90.2/32                                 Remote  BGP       00h06m38s   170
       172.16.111.2                                                 0
192.168.90.3/32                                 Remote  BGP       00h06m38s   170
       172.16.111.2                                                 0
192.168.90.4/30                                 Remote  BGP       00h06m38s   170
       172.16.111.2                                                 0
-------------------------------------------------------------------------------
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

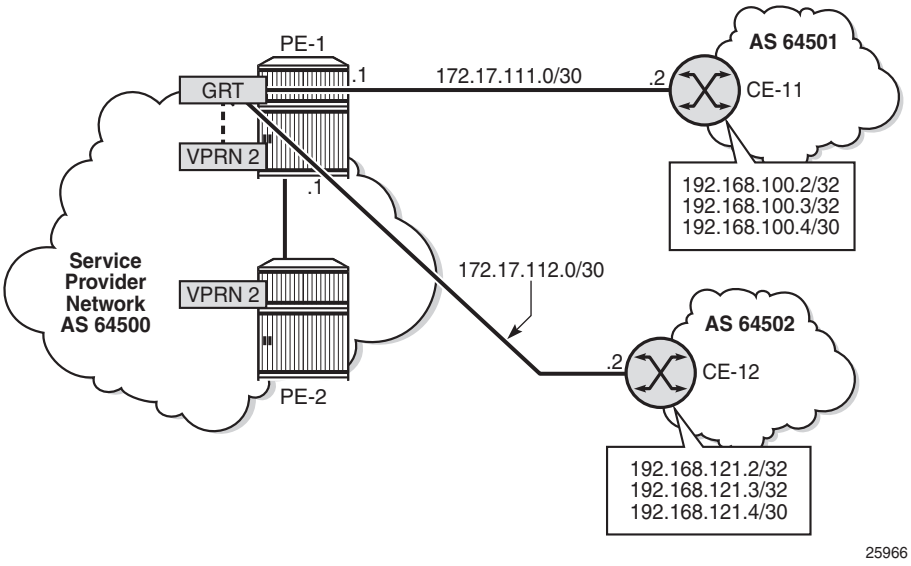
The GRT on neighbor PE-2 also includes the leaked routes, as follows:

```
*A:PE-2# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
---snip---
192.168.90.2/32                                 Remote  BGP       00h00m23s   170
       192.168.12.1                                                 0
192.168.90.3/32                                 Remote  BGP       00h00m23s   170
       192.168.12.1                                                 0
192.168.90.4/30                                 Remote  BGP       00h00m23s   170
       192.168.12.1                                                 0
-------------------------------------------------------------------------------
No. of Routes: 6
```

Example 4 - BGP IPv4 Route Leaking from GRT to VPRN per Neighbor

Figure 77 shows the topology for this example, and the corresponding IP addresses. CE-11 exports routes such as 192.168.100.2/32 to the base router and CE-12 exports routes such as 192.168.121.2/32 to the base router. The routes will be leaked from the base router to VPRN 2.

Figure 77 **BGP IPv4 Route Leaking from GRT to VPRN**



The GRT in PE-1 includes BGP routes learned from CE-11 and CE-12, as follows:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                       Metric
-------------------------------------------------------------------------------
172.17.111.0/30                                   Local   Local     01h14m11s  0
       int-PE-1-CE-11                                                 0
172.17.112.0/30                                   Local   Local     01h14m11s  0
       int-PE-1-CE-12                                                 0
192.0.2.1/32                                      Local   Local     01h14m10s  0
       system                                                         0
192.0.2.2/32                                      Remote  ISIS      01h13m40s  15
       192.168.12.2                                                   10
192.168.12.0/30                                   Local   Local     01h14m10s  0
       int-PE-1-PE-2                                                  0
192.168.100.102/32                                Remote  BGP       00h02m37s  170
```

```
       172.17.111.2                                                    0
192.168.100.103/32                          Remote  BGP       00h02m37s  170
       172.17.111.2                                                    0
192.168.100.104/30                          Remote  BGP       00h02m37s  170
       172.17.111.2                                                    0
192.168.121.2/32                            Remote  BGP       00h12m52s  170
       172.17.112.2                                                    0
192.168.121.3/32                            Remote  BGP       00h12m52s  170
       172.17.112.2                                                    0
192.168.121.4/30                            Remote  BGP       00h12m52s  170
       172.17.112.2                                                    0
-------------------------------------------------------------------------------
No. of Routes: 11
```

The BGP leaking policy is the same as in the preceding examples:

```
configure
    router
        policy-options
            begin
            policy-statement "BGP-Leak-Policy"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

The BGP route leaking policy is applied on the base router for neighbor CE-11 only, as follows:

```
configure
    router
        bgp
            group "EBGP_64500to64501_IPv4"
                neighbor 172.17.111.2
                    import "BGP-Leak-Policy"
                exit
            exit
        exit
    exit
exit
```

The following command shows that only the routes imported from neighbor CE-11 are marked as leakable in the GRT:

```
*A:PE-1# show router bgp routes ipv4 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.168.100.102/32                                None        None
      172.17.111.2                                      None        -
      64501
u*>i  192.168.100.103/32                                None        None
      172.17.111.2                                      None        -
      64501
u*>?  192.168.100.104/30                                None        None
      172.17.111.2                                      None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The leakable BGP routes in the GRT can be imported into VPRN 2. The import policy is identical to the import policy in the preceding examples, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "Import-Leakable-Routes"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

This import policy is applied in VPRN 2, as follows:

```
configure
    service
        vprn 2
            bgp
                rib-management
                    ipv4
                        leak-import "Import-Leakable-Routes"
                    exit
                exit
            exit
        exit
    exit
exit
```

The following command shows the imported leaked BGP routes in VPRN 2. The source of these leaked routes is the base router, not a VPRN.

```
*A:PE-1# show router 2 bgp routes ipv4 leaked
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>li 192.168.100.102/32                                100         None
      172.17.111.2 (Base)                               None        -
      64501
u*>li 192.168.100.103/32                                100         None
      172.17.111.2 (Base)                               None        -
      64501
u*>l? 192.168.100.104/30                                100         None
      172.17.111.2 (Base)                               None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

Any of these leaked BGP routes has the flag "leaked", and the route source is the base router (leaked from base), as shown:

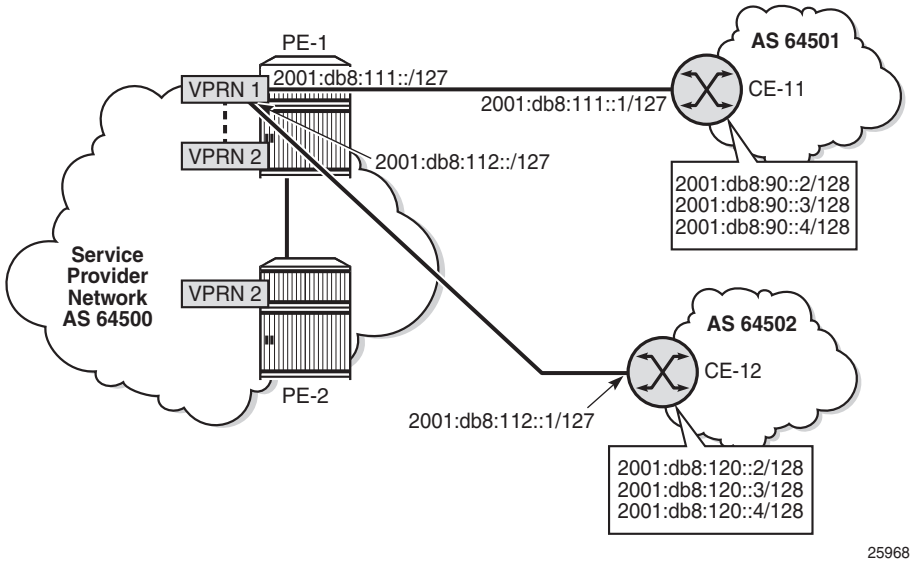
```
*A:PE-1# show router 2 bgp routes 192.168.100.102/32 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 192.168.100.102/32
Nexthop        : 172.17.111.2 (Base)
---snip---
Flags          : Used  Valid  Best  IGP  Leaked
Route Source   : Leaked from Base
---snip---
```

Example 5 - BGP IPv6 Route Leaking between VPRNs. Global VPRN BGP Configuration

Figure 6 shows the topology and the IP addresses used for this example. CE-11 exports routes such as 2001:db8:90::2/128 to VPRN 1 on PE-1, and CE-12 exports routes such as 2001:db8:120::2/128 to VPRN 1 on PE-1.

Figure 78 **BGP IPv6 Route Leaking between VPRNs**




```
*A:PE-1# show router 1 route-table ipv6
===============================================================================
IPv6 Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8::1:1/128                               Local   Local     01h41m56s  0
       system                                                        0
2001:db8:90::2/128                              Remote  BGP       00h02m19s  170
       2001:db8:111::1                                               0
2001:db8:90::3/128                              Remote  BGP       00h02m19s  170
       2001:db8:111::1                                               0
2001:db8:90::4/126                              Remote  BGP       00h02m19s  170
       2001:db8:111::1                                               0
2001:db8:111::/127                              Local   Local     01h41m55s  0
       int-PE-1-CE-11                                                0
2001:db8:112::/127                              Local   Local     01h41m55s  0
       int-PE-1-CE-12                                                0
2001:db8:120::2/128                             Remote  BGP       00h02m05s  170
       2001:db8:112::1                                               0
2001:db8:120::3/128                             Remote  BGP       00h02m05s  170
```

```
       2001:db8:112::1                                                  0
2001:db8:120::4/126                          Remote  BGP       00h02m05s  170
       2001:db8:112::1                                                  0
-------------------------------------------------------------------------------
No. of Routes: 9
```

The BGP route leaking policy is the same as for IPv4 routes:

```
configure
    router
        policy-options
            begin
            policy-statement "BGP-Leak-Policy"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

This import policy is applied in the BGP context of VPRN 1, as follows:

```
configure
    service
        vprn 1
            bgp
                import "BGP-Leak-Policy"
            exit
        exit
    exit
exit
```

With the preceding configuration, all the routes imported into the VPRN using BGP are marked as leakable.

The following command shows which BGP IPv6 routes are marked as leakable in VPRN 1:

```
*A:PE-1# show router 1 bgp routes ipv6 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
```

```
      As-Path
-------------------------------------------------------------------------------
u*>i  2001:db8:90::2/128                                  None        None
      2001:db8:111::1                                     None        -
      64501
u*>i  2001:db8:90::3/128                                  None        None
      2001:db8:111::1                                     None        -
      64501
u*>?  2001:db8:90::4/126                                  None        None
      2001:db8:111::1                                     None        -
      64501
u*>i  2001:db8:120::2/128                                 None        None
      2001:db8:112::1                                     None        -
      64502
u*>i  2001:db8:120::3/128                                 None        None
      2001:db8:112::1                                     None        -
      64502
u*>?  2001:db8:120::4/126                                 None        None
      2001:db8:112::1                                     None        -
      64502
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

The BGP leakable routes can be imported into VPRN 2 when the following import policy is configured and applied in VPRN 2:

```
configure
    router
        policy-options
            begin
            policy-statement "Import-Leakable-Routes"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

The only difference from IPv4 routes is that the policy is applied to the IPv6 context of the RIB management:

```
configure
    service
        vprn 2
            bgp
                rib-management
                    ipv6
                        leak-import "Import-Leakable-Routes"
                    exit
                exit
            exit
        exit
```

```
    exit
exit
```

The following command shows that the VPRN is importing the leaked BGP IPv6 routes from another VPRN instance:

```
*A:PE-1# show router 2 bgp routes ipv6 leaked
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>li 2001:db8:90::2/128                                   100         None
      2001:db8:111::1 (VPRN 1)                             None        -
      64501
u*>li 2001:db8:90::3/128                                   100         None
      2001:db8:111::1 (VPRN 1)                             None        -
      64501
u*>l? 2001:db8:90::4/126                                   100         None
      2001:db8:111::1 (VPRN 1)                             None        -
      64501
u*>li 2001:db8:120::2/128                                  100         None
      2001:db8:112::1 (VPRN 1)                             None        -
      64502
u*>li 2001:db8:120::3/128                                  100         None
      2001:db8:112::1 (VPRN 1)                             None        -
      64502
u*>l? 2001:db8:120::4/126                                  100         None
      2001:db8:112::1 (VPRN 1)                             None        -
      64502
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

The BGP routes have the flag "leaked" and the route source is VPRN 1, as follows:

```
*A:PE-1# show router 2 bgp routes 2001:db8:90::2/128 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
```

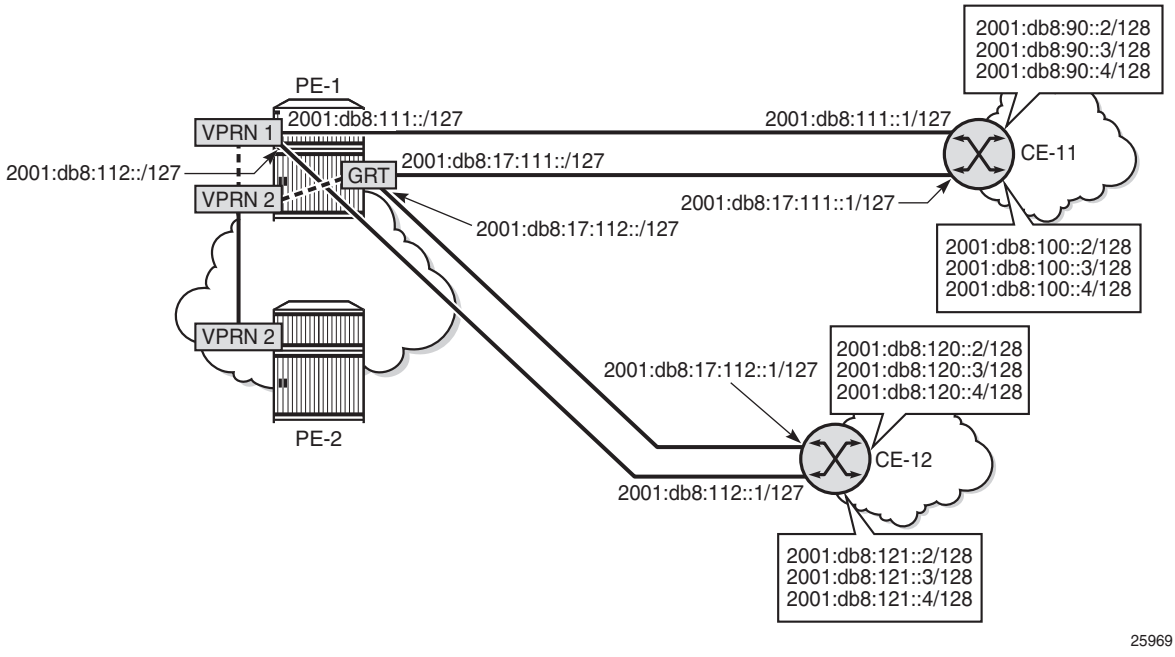
```
Original Attributes

Network        : 2001:db8:90::2/128
Nexthop        : 2001:db8:111::1 (VPRN 1)
---snip---
Flags          : Used  Valid  Best  IGP  Leaked
Route Source   : Leaked from VPRN 1
---snip---
```

Example 6 - BGP IPv6 Route Leaking from GRT to VPRN and from VPRN to VPRN

Figure 7 shows the topology and the IPv6 addresses used in this example. CE-11 exports IPv6 routes such as 2001:db8:90::2/128 to VPRN 1 and IPv6 routes such as 2001:db8:100::2/128 to the GRT. CE-12 exports IPv6 routes such as 2001:db8:120::2/128 to VPRN 1 and IPv6 routes such as 2001:db8:121::2/128 to the GRT.

Figure 79 **BGP IPv6 Route Leaking from GRT and VPRN to VPRN**



The IPv6 routing table in the GRT contains routes exported by CE-11 and CE-12, as follows:

```
*A:PE-1# show router route-table ipv6
===============================================================================
```

```
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
2001:db8::1/128                                 Local   Local     00h10m57s  0
       system                                                        0
2001:db8::2/128                                 Remote  ISIS      00h10m45s  15
       fe80::48c5:1ff:fe01:1-"int-PE-1-PE-2"                         10
2001:db8:12::/126                               Local   Local     00h10m56s  0
       int-PE-1-PE-2                                                 0
2001:db8:17:111::/127                           Local   Local     00h10m56s  0
       int-PE-1-CE-11                                                0
2001:db8:17:112::/127                           Local   Local     00h10m57s  0
       int-PE-1-CE-12                                                0
2001:db8:100::2/128                             Remote  BGP       00h06m52s  170
       2001:db8:17:111::1                                            0
2001:db8:100::3/128                             Remote  BGP       00h06m52s  170
       2001:db8:17:111::1                                            0
2001:db8:100::4/126                             Remote  BGP       00h06m52s  170
       2001:db8:17:111::1                                            0
2001:db8:121::2/128                             Remote  BGP       00h09m40s  170
       2001:db8:17:112::1                                            0
2001:db8:121::3/128                             Remote  BGP       00h09m40s  170
       2001:db8:17:112::1                                            0
2001:db8:121::4/126                             Remote  BGP       00h09m40s  170
       2001:db8:17:112::1                                            0
-------------------------------------------------------------------------------
No. of Routes: 11
```

The IPv6 routing table for VPRN 1 also contains routes exported by CE-11 and CE-12, as follows:

```
*A:PE-1# show router 1 route-table ipv6
===============================================================================
IPv6 Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
2001:db8::1:1/128                               Local   Local     00h12m14s  0
       system                                                        0
2001:db8:90::2/128                              Remote  BGP       00h08m13s  170
       2001:db8:111::1                                               0
2001:db8:90::3/128                              Remote  BGP       00h08m13s  170
       2001:db8:111::1                                               0
2001:db8:90::4/126                              Remote  BGP       00h08m13s  170
       2001:db8:111::1                                               0
2001:db8:111::/127                              Local   Local     00h12m14s  0
       int-PE-1-CE-11                                                0
2001:db8:112::/127                              Local   Local     00h12m14s  0
       int-PE-1-CE-12                                                0
2001:db8:120::2/128                             Remote  BGP       00h11m12s  170
       2001:db8:112::1                                               0
2001:db8:120::3/128                             Remote  BGP       00h11m12s  170
       2001:db8:112::1                                               0
2001:db8:120::4/126                             Remote  BGP       00h11m12s  170
       2001:db8:112::1                                               0
```

```
-------------------------------------------------------------------------------
No. of Routes: 9
```

The policy to mark imported BGP routes as leakable can be identical to the policy used in the preceding examples. However, in this case, prefix-lists are added as a filter. The base router may accept routes such as 2001:db8:100::2/128 and 2001:db8:121::2/128.

```
configure
    router
        policy-options
            begin
            prefix-list "2001:db8:90::"
                prefix 2001:db8:90::/100 longer
            exit
            prefix-list "2001:db8:120::"
                prefix 2001:db8:120::/100 longer
            exit
            policy-statement "BGP-Leak-Policy_90_120"
                entry 10
                    from
                        protocol bgp
                        prefix-list "2001:db8:90::"
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        prefix-list "2001:db8:120::"
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

This import policy is applied in the general BGP settings for VPRN 1, as follows:

```
configure
    service
        vprn 1
            bgp
                import "BGP-Leak-Policy_90_120"
            exit
        exit
    exit
exit
```

In a similar way, the base router may accept routes such as 2001:8db:100::2/128 and 2001:8db:121::2/128:

```
configure
    router
        policy-options
            begin
            prefix-list "2001:db8:100::"
                prefix 2001:db8:100::/100 longer
            exit
            prefix-list "2001:db8:121::"
                prefix 2001:db8:121::/100 longer
            exit
            policy-statement "BGP-Leak-Policy_100_121"
                entry 10
                    from
                        protocol bgp
                        prefix-list "2001:db8:100::"
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        prefix-list "2001:db8:121::"
                    exit
                    action accept
                        bgp-leak
                    exit
                exit
            exit
            commit
```

This BGP leaking policy is applied for neighbor CE-11 in the base router, as follows. The routes exported by CE-12 will not be marked as leakable.

```
configure
    router
        bgp
            group "EBGP_64500to64501_IPv6"
                neighbor 2001:db8:17:111::1
                    import "BGP-Leak-Policy_100_121"
                exit
            exit
        exit
    exit
exit
```

The following command shows which routes are marked as leakable in the GRT:

```
*A:PE-1# show router bgp routes ipv6 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
```

```
===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  2001:db8:100::2/128                               None        None
      2001:db8:17:111::1                                None        -
      64501
u*>i  2001:db8:100::3/128                               None        None
      2001:db8:17:111::1                                None        -
      64501
u*>?  2001:db8:100::4/126                               None        None
      2001:db8:17:111::1                                None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The following command shows which routes are marked as leakable in VPRN 1:

```
*A:PE-1# show router 1 bgp routes ipv6 leakable
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  2001:db8:90::2/128                                None        None
      2001:db8:111::1                                   None        -
      64501
u*>i  2001:db8:90::3/128                                None        None
      2001:db8:111::1                                   None        -
      64501
u*>?  2001:db8:90::4/126                                None        None
      2001:db8:111::1                                   None        -
      64501
u*>i  2001:db8:120::2/128                               None        None
      2001:db8:112::1                                   None        -
      64502
u*>i  2001:db8:120::3/128                               None        None
      2001:db8:112::1                                   None        -
      64502
u*>?  2001:db8:120::4/126                               None        None
      2001:db8:112::1                                   None        -
      64502
-------------------------------------------------------------------------------
Routes : 6===============================================================================
```

```
=====
*A:PE-1#
```

On PE-1, a policy is created to import the BGP leakable routes (the same as in the preceding examples), as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "Import-Leakable-Routes"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

This import policy is configured for IPv6 routes in VPRN2, as follows:

```
configure
    service
        vprn 2
            bgp
                rib-management
                    ipv6
                        leak-import "Import-Leakable-Routes"
                    exit
                exit
            exit
        exit
    exit
exit
```

The following command shows the leaked IPv6 routes in VPRN 2:

```
*A:PE-1# show router 2 bgp routes ipv6 leaked
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>li 2001:db8:90::2/128                                100         None
      2001:db8:111::1 (VPRN 1)                          None        -
```

```
      64501
u*>li 2001:db8:90::3/128                                  100         None
      2001:db8:111::1 (VPRN 1)                            None        -
      64501
u*>l? 2001:db8:90::4/126                                  100         None
      2001:db8:111::1 (VPRN 1)                            None        -
      64501
u*>li 2001:db8:100::2/128                                 100         None
      2001:db8:17:111::1 (Base)                           None        -
      64501
u*>li 2001:db8:100::3/128                                 100         None
      2001:db8:17:111::1 (Base)                           None        -
      64501
u*>l? 2001:db8:100::4/126                                 100         None
      2001:db8:17:111::1 (Base)                           None        -
      64501
u*>li 2001:db8:120::2/128                                 100         None
      2001:db8:112::1 (VPRN 1)                            None        -
      64502
u*>li 2001:db8:120::3/128                                 100         None
      2001:db8:112::1 (VPRN 1)                            None        -
      64502
u*>l? 2001:db8:120::4/126                                 100         None
      2001:db8:112::1 (VPRN 1)                            None        -
      64502
-------------------------------------------------------------------------------
Routes : 9
===============================================================================
*A:PE-1#
```

Some of these routes are leaked from the base router and some routes are leaked from VPRN 1. The detailed information for any of these leaked routes shows that the flag "leaked" is present. For route 2001:db8:100::2/128, the route source is the base router, as follows:

```
*A:PE-1# show router 2 bgp routes 2001:db8:100::2/128 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
Original Attributes

Network        : 2001:db8:100::2/128
Nexthop        : 2001:db8:17:111::1 (Base)
---snip---
Flags          : Used  Valid  Best  IGP  Leaked
Route Source   : Leaked from Base
---snip---
```

For route 2001:db8:90::2/128, the route source is VPRN 1, as follows:

```
*A:PE-1# show router 2 bgp routes 2001:db8:90::2/128 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv6 Routes
===============================================================================
Original Attributes

Network        : 2001:db8:90::2/128
Nexthop        : 2001:db8:111::1 (VPRN 1)
---snip---
Flags          : Used  Valid  Best  IGP  Leaked
Route Source   : Leaked from VPRN 1
---snip---
```

Conclusion

BGP provides many ways to manipulate routes. In this example, IPv4/IPv6 routes learned from BGP neighbors could be marked as "leakable" and imported into other routing instances (VPRN to VPRN, VPRN to GRT, GRT to VPRN) without the use of communities in the network policy.

IS-IS Link Bundling

This chapter provides information about IS-IS link bundling.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

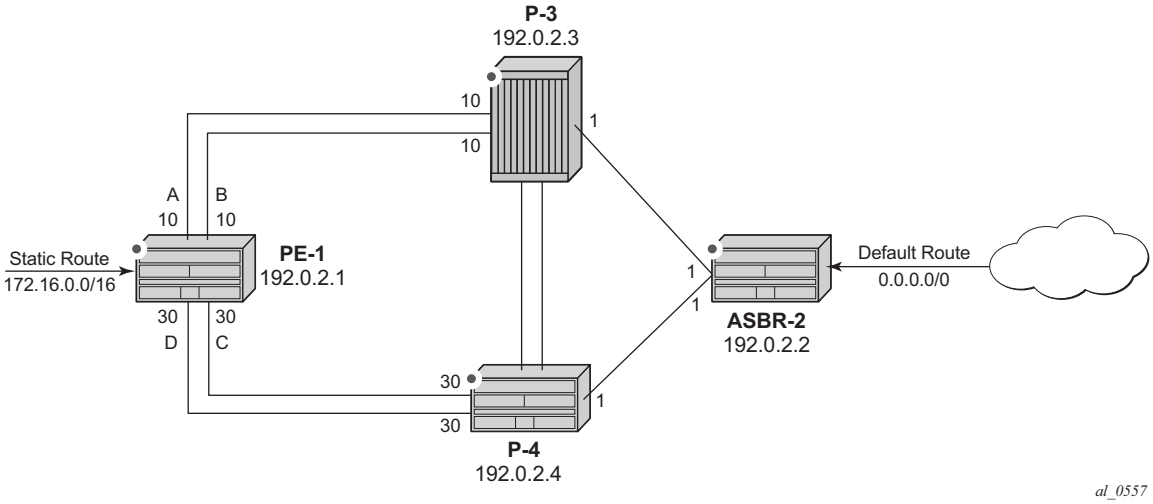
Applicability

The chapter was initially written for SR OS release 11.0.R6. However, the CLI in the current edition is based on release 14.0.R5.

Overview

Intermediate System to Intermediate System (IS-IS) Link Bundling allows for the grouping of a number of IS-IS interfaces into a single virtual link, called an IS-IS link group. It is used in conjunction with Equal Cost Multipath (ECMP) to dynamically change the metric of parallel IS-IS links if one or more links fail or suffer some sort of performance degradation.

Figure 80 **Link Bundle Schematic**



Consider the network in Figure 80, where a Provider Edge router PE-1 connects to a core network comprised of two Provider (P) routers and a single Autonomous System Border Router (ASBR). The links between PE-1 and P-3, and PE-1 and P-4 are 10 Gigabit Ethernet links. The links between ASBR-2 and P-3 and P-4 are both 100 Gigabit links. The link metrics are as shown in Figure 80.

In order to maximize the use of link bandwidth, ECMP is enabled on all routers and set to a value of 2, so that IP traffic flowing between PE1 and P-3, and PE-1 and P-4, is load balanced across the two links.

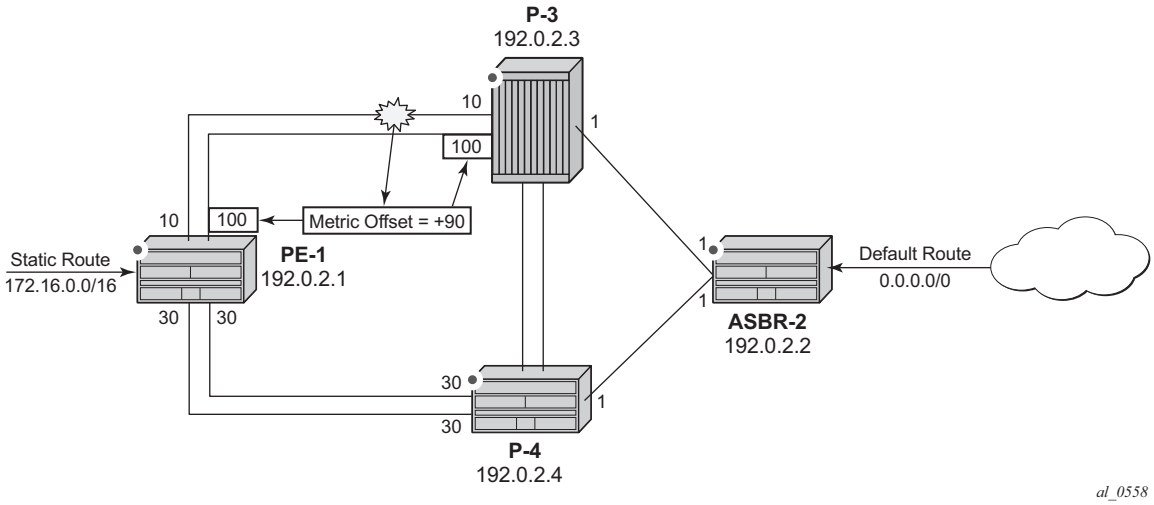
A default route is injected into the ASBR-2 router and re-distributed via a policy statement into IS-IS, so that traffic flowing from PE-1 to the ASBR is resolved by this route. Traffic flows between PE-1 and ASBR-2, using the path with the lowest IS-IS metric, via P-3 with a metric of 11. The second path PE-1 to ASBR-2 via P-4 has the same bandwidth, but a higher IS-IS metric of 31.

Traffic in the reverse direction flows toward a user subnet described by a static route configured on PE-1, which is redistributed into IS-IS using a policy statement. Once again, the shortest path between ASBR-2 and PE-1 is via P-3, so the bi-directional traffic flow is symmetric.

If one of the links between PE-1 and P-3 fails, traffic still flows via P-3, because the IS-IS metric is unchanged, but this now has less bandwidth than the second path via P-4. It is desirable to make use of the additional bandwidth of the second path, but this requires a change in metric. This can be achieved using IS-IS link bundling.

IS-IS link bundling allows for the creation of a group of IS-IS links, where the failure of a member link allows the metric of the remaining members of the link group to be increased by an offset value.

Figure 81 **Effect of Single Link Failure on Bundle Group**



Using Figure 81 as an example, the links between PE-1 and P-3 are included in a bundle group. To illustrate the change in metrics, a default static route is configured on ASBR-2 and re-distributed into IS-IS, and the path to this route is monitored at PE-1. Similarly, a static route to subnet 172.16.0.0/16 is configured on PE-1 and redistributed into IS-IS and viewed on ASBR-2.

Should one of the links between PE-1 and P-3 fail, the metric of the remaining members can be increased by an offset, for example 90, so that the metric of the remaining link becomes 10+90 = 100. The IS-IS metric between PE-1 and ASBR-2 via P-3 is now 101. The metric offset is applied to each remaining IS-IS interface individually and is advertised within the IS-IS database as the default cost in the TE-IS neighbors Type Length Variable (TLV).

The path between PE-1 and ASBR-2 via P-4 now has the lowest IS-IS metric, and any affected routers within the IS-IS area will try and re-route the traffic based on the new metric.

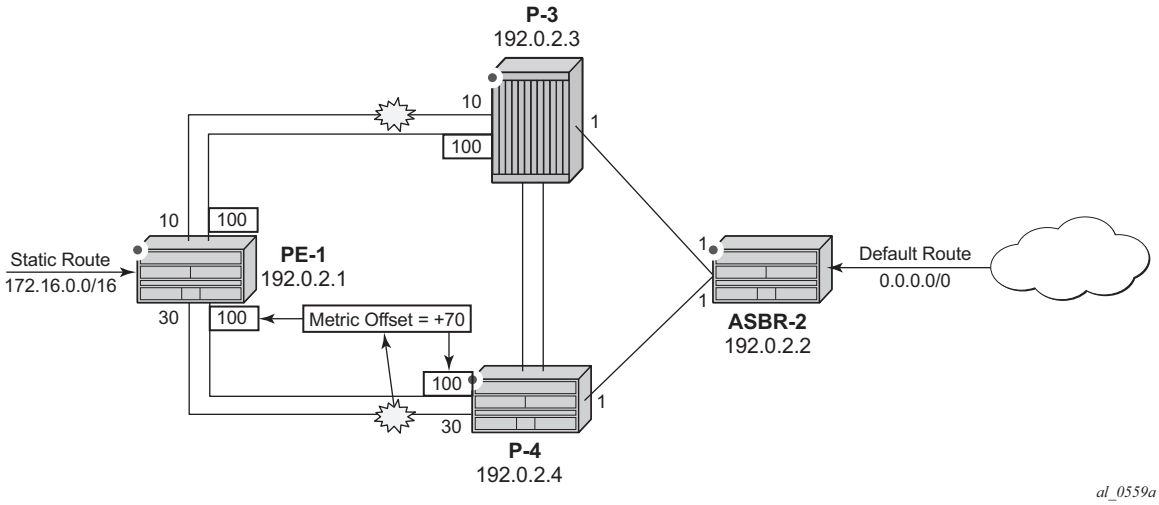
The fundamentals of this feature are:

- The treatment of all member links in a link group bundle as a single virtual interface.
- The increase in metric by a specific offset value of each remaining individual link within the group when a failure of one or more links occurs.

– The application of the offset occurs when the number of active links drops below a configured threshold.
– The offset is removed when the number of active links within the link group bundle reaches the configured reversion threshold.
– A link bundle is required on a router for the thresholds and offsets to apply.

Consider a second and subsequent failure where a link between PE-1 and P-4 also fails, so that there is only one active IS-IS interface between PE-1 and each of its neighboring P routers. This is shown in Figure 82.

Figure 82 **Double Link Failure**



In this case, the metric for the remaining link between PE-1 and P-4 can be increased by an offset value of +70 so that the IS-IS metric PE-1 to P-4 becomes 100, the same as that between PE-1 and P-3 when a link has failed.

PE-1 now sees two equal cost paths to the default route – one via P-3 and one via P-4, so there are still two 10Gigabit Ethernet links across which the traffic can be load shared.

This can be summarized using the following table, where ABCD are the 4 links as per Figure 80 and link status is Up (U) or Down (D).

Table 4 **Status of the links A, B, C, and D**

ABCD Status	A (metric,status)	B (metric,status)	C (metric,status)	D (metric,status)
UUUU	10 Transmit	10 Transmit	30 Idle	30 Idle
UDUU	100 Idle	Down	30 Transmit	30 Transmit

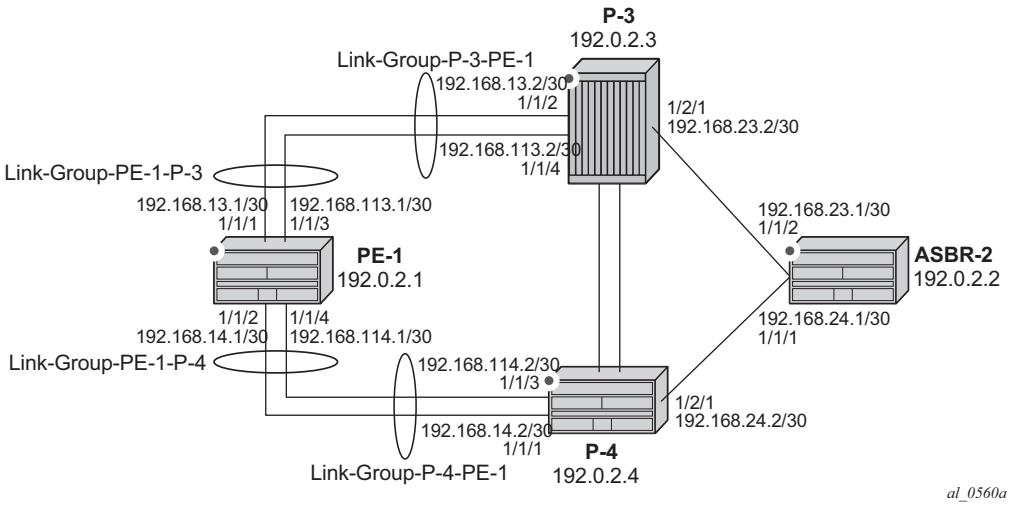
Table 4 **Status of the links A, B, C, and D**

ABCD Status	A (metric,status)	B (metric,status)	C (metric,status)	D (metric,status)
UDUD	100 Transmit	Down	100 Transmit	Down
UUUD	10 Transmit	10 Transmit	100 Idle	Down

Configuration

The example topology is shown in Figure 83.

Figure 83 **Example Topology**



The PE-1 router configuration commands are as follows.

```
*A:PE-1# configure
    router
        interface "int-PE-1-P-3-1"
            address 192.168.13.1/30
            port 1/1/1
        exit
        interface "int-PE-1-P-3-2"
            address 192.168.113.1/30
            port 1/1/3
        exit
        interface "int-PE-1-P-4-1"
            address 192.168.14.1/30
            port 1/1/2
        exit
```

```
        interface "int-PE-1-P-4-2"
            address 192.168.114.1/30
            port 1/1/4
        exit
        interface "system"
            address 192.0.2.1/32
        exit
        ecmp 2
```

The IP router configuration for the remaining routers can be derived from Figure 83.

The IS-IS network is a level 1 network.

The IS-IS configuration for PE-1, including the interface metrics is as follows:

```
*A:PE-1# configure
    router
        isis
            level-capability level-1
            area-id 49.0001
            advertise-passive-only
            level 1
                wide-metrics-only
            exit
            interface "system"
                passive
            exit
            interface "int-PE-1-P-3-1"
                interface-type point-to-point
                level 1
                    metric 10
                exit
            exit
            interface "int-PE-1-P-3-2"
                interface-type point-to-point
                level 1
                    metric 10
                exit
            exit
            interface "int-PE-1-P-4-1"
                interface-type point-to-point
                level 1
                    metric 30
                exit
            exit
            interface "int-PE-1-P-4-2"
                interface-type point-to-point
                level 1
                    metric 30
                exit
            exit
```

The IS-IS configuration for the remaining routers can be derived from Figure 83.

The following configuration is for the static route and export policy on ASBR-2. The configuration of the static route on PE-1 is similar.

```
*A:ASBR-2# configure router static-route-entry 0.0.0.0/0 black-hole no shutdown

*A:ASBR-2# configure router
        policy-options
            begin
            policy-statement "STATIC-ISIS"
                entry 10
                    from
                        protocol static
                    exit
                    to
                        level 1
                    exit
                    action accept
                        metric set igp
                    exit
                exit
            exit
            commit
        exit

*A:ASBR-2# configure router isis export "STATIC-ISIS"
```

Link Group Configuration

PE-1 contains 2 link groups. The first link group contains the IS-IS interfaces toward P-3. The second contains the interfaces toward P-4.

Each link-group is configured using a unique name, which is unique per router, and the IS-IS interface names are configured within the group as group members.

The metric offset value is the amount by which the IS-IS metric of active member links are increased when the number of links drops below a configured threshold.

The IS-IS link group configuration for PE-1 for the interfaces toward P-3 is as follows:

```
*A:PE-1# configure
    router
        isis
            link-group "Link-Group-PE-1-P-3"
                level 1
                    ipv4-unicast-metric-offset 90
                    member "int-PE-1-P-3-1"
                    member "int-PE-1-P-3-2"
                    revert-members 2
                    oper-members 2
                exit
            exit
        exit
```

Similarly, the IS-IS link group for PE-1 for the interfaces toward P-4 is:

```
*A:PE-1# configure
    router
        isis
            link-group "Link-Group-PE-1-P-4"
                level 1
                    ipv4-unicast-metric-offset 70
                    member "int-PE-1-P-4-1"
                    member "int-PE-1-P-4-2"
                    revert-members 2
                    oper-members 2
                exit
            exit
        exit
```

Within the link-group, two thresholds are configured:

- oper-members threshold
- revert-members threshold

If the number of operational links in the link-group drops below the oper-members value, then all interfaces associated with that IS-IS link group have their interface metric increased by the configured offset value. As a result, IS-IS then tries to reroute traffic over lower cost paths.

If the number of operational links in the link-group equals the revert-members threshold value, then all interfaces associated with that IS-IS link group have their interface metric decreased by the configured offset value.

In this configuration, there is a requirement to increase the metric of each interface within a link-group when a single interface fails. This means that the oper-members value is set to 2. In normal working circumstances, when both interfaces are active, the metric used is the configured interface metric. This means that the revert-members value must also be set to 2.

It is not possible to set the oper-members threshold to a value higher than that of the revert-members.

For completeness, the IS-IS configuration of each P-router is as follows.

P-3

```
*A:P-3# configure
    router
        isis
            level-capability level-1
            area-id 49.0001
            advertise-passive-only
            level 1
                wide-metrics-only
            exit
            interface "system"
                passive
```

```
            exit
            interface "int-P-3-PE-1-1"
                interface-type point-to-point
                level 1
                    metric 10
                exit
            exit
            interface "int-P-3-PE-1-2"
                interface-type point-to-point
                level 1
                    metric 10
                exit
            exit
            interface "int-P-3-ASBR-2"
                interface-type point-to-point
                level 1
                    metric 1
                exit
            exit
            link-group "Link-Group-P-3-PE-1"
                level 1
                    ipv4-unicast-metric-offset 90
                    member "int-P-3-PE-1-1"
                    member "int-P-3-PE-1-2"
                    revert-members 2
                    oper-members 2
                exit
            exit
        exit
```

P-4

```
*A:P-4# configure
    router
        isis
            level-capability level-1
            area-id 49.0001
            advertise-passive-only
            level 1
                wide-metrics-only
            exit
            interface "system"
                passive
            exit
            interface "int-P-4-PE-1-1"
                interface-type point-to-point
                level 1
                    metric 30
                exit
            exit
            interface "int-P-4-PE-1-2"
                interface-type point-to-point
                level 1
                    metric 30
                exit
            exit
            interface "int-P-4-ASBR-2"
                interface-type point-to-point
```

```
            level 1
                metric 1
            exit
        exit
        link-group "Link-Group-P-4-PE-1"
            level 1
                ipv4-unicast-metric-offset 70
                member "int-P-4-PE-1-1"
                member "int-P-4-PE-1-2"
                revert-members 2
                oper-members 2
            exit
        exit
    exit
```

An overview of all of the link groups can be shown using the following commands, in this case on node PE-1.

The link group status is as follows:

```
*A:PE-1# show router isis link-group-status
===============================================================================
Router Base ISIS Instance 0 Link-Group Status
===============================================================================
Link-group              Mbrs   Oper   Revert Active Level   State
                               Mbr     Mbr    Mbr
-------------------------------------------------------------------------------
Link-Group-PE-1-P-3     2      2      2      2      L1      normal
Link-Group-PE-1-P-4     2      2      2      2      L1      normal
===============================================================================
*A:PE-1#
```

The output for the individual link group members is as follows:

For "Link-Group-PE-1-P-3" at PE-1:

```
*A:PE-1# show router isis link-group-member-status level 1 "Link-Group-PE-1-P-3"
===============================================================================
Router Base ISIS Instance 0 Link-Group Member
===============================================================================
Link-group            I/F name               Level      State
-------------------------------------------------------------------------------
Link-Group-PE-1-P-3   int-PE-1-P-3-1         L1         Up
Link-Group-PE-1-P-3   int-PE-1-P-3-2         L1         Up
-------------------------------------------------------------------------------
Legend: BER = bitErrorRate
===============================================================================
*A:PE-1#
```

For "Link-Group-PE-1-P-4" at PE-1:

```
*A:PE-1# show router isis link-group-member-status level 1 "Link-Group-PE-1-P-4"
===============================================================================
Router Base ISIS Instance 0 Link-Group Member
===============================================================================
```

```
Link-group            I/F name                Level      State
-------------------------------------------------------------------------------
Link-Group-PE-1-P-4   int-PE-1-P-4-1          L1         Up
Link-Group-PE-1-P-4   int-PE-1-P-4-2          L1         Up
-------------------------------------------------------------------------------
Legend: BER = bitErrorRate
===============================================================================
*A:PE-1#
```

For P-3, the link group status is as follows:

```
*A:P-3# show router isis link-group-status
===============================================================================
Router Base ISIS Instance 0 Link-Group Status
===============================================================================
Link-group              Mbrs   Oper   Revert Active Level   State
                               Mbr     Mbr    Mbr
-------------------------------------------------------------------------------
Link-Group-P-3-PE-1     2      2      2      2      L1      normal
===============================================================================
*A:P-3#
```

For P-3, the link group member status is as follows:

```
*A:P-3# show router isis link-group-member-status level 1 "Link-Group-P-3-PE-1"
===============================================================================
Router Base ISIS Instance 0 Link-Group Member
===============================================================================
Link-group            I/F name                Level      State
-------------------------------------------------------------------------------
Link-Group-P-3-PE-1   int-P-3-PE-1-1          L1         Up
Link-Group-P-3-PE-1   int-P-3-PE-1-2          L1         Up
-------------------------------------------------------------------------------
Legend: BER = bitErrorRate
===============================================================================
*A:P-3#
```

Routing Table PE-1

In a normal working state, the routing table for PE-1 contains the default route for forwarding traffic toward ASBR-2. Because ECMP is set to a value of 2, two entries are available with next-hops pointing toward P-3, as follows. The metric for each path is 11.

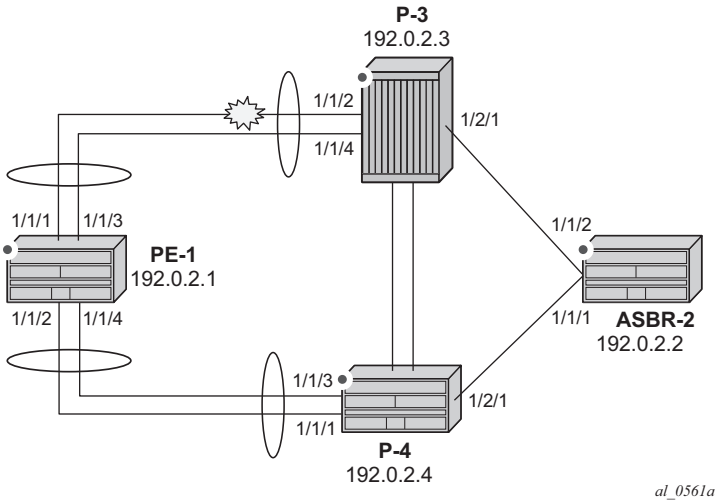
```
*A:PE-1# show router route-table 0.0.0.0/0

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                       Remote  ISIS      00h02m27s  15
       192.168.13.2                                                    11
0.0.0.0/0                                       Remote  ISIS      00h02m27s  15
```

```
       192.168.113.2                                                    11
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

Failure of link member PE-1 to P-3

Figure 84 **Link Failure**



One of the links between PE-1 and P-3 is put into a failed state by shutting down port 1/1/2 on P-3, as per Figure 84.

```
*A:P-3# configure port 1/1/2 shutdown
```

The route-table on PE-1 shows that the metric for the default route prefix, 0.0.0.0/0, has increased from 11 to 31, and the next-hops are now interface addresses on P-4, as follows:

```
A:PE-1# show router route-table 0.0.0.0/0
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                       Remote  ISIS      00h01m14s  15
       192.168.14.2                                                 31
0.0.0.0/0                                       Remote  ISIS      00h01m14s  15
       192.168.114.2                                                31
-------------------------------------------------------------------------------
```

```
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

The link-group status shows that the number of active members has fallen below the oper-members threshold and as a result, the metric offset has been applied.

```
*A:PE-1# show router isis link-group-status
===============================================================================
Router Base ISIS Instance 0 Link-Group Status
===============================================================================
Link-group              Mbrs   Oper   Revert Active Level   State
                               Mbr     Mbr    Mbr
-------------------------------------------------------------------------------
Link-Group-PE-1-P-3     2      2      2      1      L1      Offset-Applied
Link-Group-PE-1-P-4     2      2      2      2      L1      normal
===============================================================================
*A:PE-1#
```

Finally, the status of an individual link group member is as follows:

```
*A:PE-1# show router isis link-group-member-status "Link-Group-PE-1-P-3"
===============================================================================
Router Base ISIS Instance 0 Link-Group Member
===============================================================================
Link-group            I/F name                Level      State
-------------------------------------------------------------------------------
Link-Group-PE-1-P-3   int-PE-1-P-3-1          L1         If-Down
Link-Group-PE-1-P-3   int-PE-1-P-3-2          L1         Up
-------------------------------------------------------------------------------
Legend: BER = bitErrorRate
===============================================================================
*A:PE-1#
```

By examining the IS-IS database on PE-1, it can be seen that the link metric (TE-IS neighbor) toward P-3 has a metric of 100, comprised of the original metric of 10 plus the offset of 90.

```
*A:PE-1# show router isis database PE-1 detail
===============================================================================
Router Base ISIS Instance 0 Database
===============================================================================

Displaying Level 1 database
-------------------------------------------------------------------------------
LSP ID    : PE-1.00-00                                  Level     : L1
Sequence  : 0x9                  Checksum  : 0x9312     Lifetime  : 1016
Version   : 1                    Pkt Type  : 18         Pkt Ver   : 1
Attributes: L1                   Max Area  : 3
SysID Len : 6                    Used Len  : 156        Alloc Len : 1492

TLVs :
```

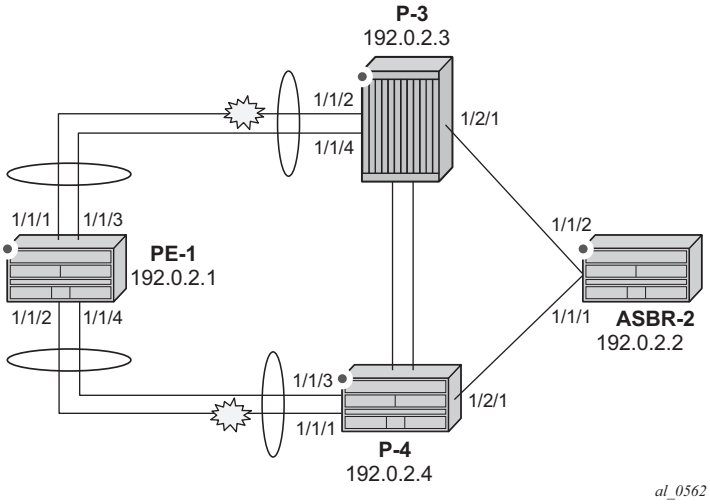
```
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols     : IPv4
  IS-Hostname   : PE-1
  Router ID   :
    Router ID   : 192.0.2.1
  I/F Addresses :
    I/F Address   : 192.0.2.1
    I/F Address   : 192.168.13.1
    I/F Address   : 192.168.14.1
    I/F Address   : 192.168.113.1
    I/F Address   : 192.168.114.1
  TE IS Nbrs   :
    Nbr   : P-3.00
    Default Metric  : 100
    Sub TLV Len     : 12
    IF Addr   : 192.168.113.1
    Nbr IP    : 192.168.113.2
  TE IS Nbrs   :
    Nbr   : P-4.00
    Default Metric  : 30
    Sub TLV Len     : 12
    IF Addr   : 192.168.14.1
    Nbr IP    : 192.168.14.2
  TE IS Nbrs   :
    Nbr   : P-4.00
    Default Metric  : 30
    Sub TLV Len     : 12
    IF Addr   : 192.168.114.1
    Nbr IP    : 192.168.114.2
  TE IP Reach   :
    Default Metric  : 0
    Control Info:    , prefLen 16
    Prefix   : 172.16.0.0
    Default Metric  : 0
    Control Info:    , prefLen 32
    Prefix   : 192.0.2.1

Level (1) LSP Count : 1

Displaying Level 2 database
-------------------------------------------------------------------------------
Level (2) LSP Count : 0
===============================================================================
*A:PE-1#
```

Failure of link member PE-1 to P-4:

Figure 85 **Second Link Failure**



If a link between PE-1 and P-4 now fails, simulated by shutting down port 1/1/1 on P-4, then the metric offset is applied to the link groups on PE-1 and P-4 as the number of active links has dropped below the oper-members threshold for the link groups Link-Group-PE-1-P-4 on PE-1 and Link-Group-P-4-PE-1on P-4.

```
*A:P-4# configure port 1/1/1 shutdown
```

The routing table for PE-1 now shows that there are still two equal cost paths for the default route prefix advertised by ASBR-2, as follows:

```
*A:PE-1# show router route-table 0.0.0.0/0

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                        Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                         Remote  ISIS      00h01m16s  15
       192.168.113.2                                                  101
0.0.0.0/0                                         Remote  ISIS      00h01m16s  15
       192.168.114.2                                                  101
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

The metric for each routing table entry is 101, comprising of a cost of 100 for the PE-1 to P router link, where the link-group offset has been applied, and the cost of 1 for the P router to ASBR-2 router link.

By examining the IS-IS database on the PE-1 router, the updated metric for the link to neighbors P-3 and P-4 can be seen with the offset applied. These are seen in the “TE-IS Nbrs” TLV in the following output.

```
*A:PE-1# show router isis database PE-1 detail
===============================================================================
Router Base ISIS Instance 0 Database
===============================================================================
Displaying Level 1 database
-------------------------------------------------------------------------------
LSP ID    : PE-1.00-00                                  Level     : L1
Sequence  : 0xa                    Checksum  : 0x5ebc   Lifetime  : 1105
Version   : 1                      Pkt Type  : 18       Pkt Ver   : 1
Attributes: L1                     Max Area  : 3
SysID Len : 6                      Used Len  : 131      Alloc Len : 1492

TLVs :
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols     : IPv4
  IS-Hostname   : PE-1
  Router ID   :
    Router ID   : 192.0.2.1
  I/F Addresses :
    I/F Address   : 192.0.2.1
    I/F Address   : 192.168.13.1
    I/F Address   : 192.168.14.1
    I/F Address   : 192.168.113.1
    I/F Address   : 192.168.114.1
  TE IS Nbrs   :
    Nbr   : P-3.00
    Default Metric  : 100
    Sub TLV Len     : 12
    IF Addr   : 192.168.113.1
    Nbr IP    : 192.168.113.2
  TE IS Nbrs   :
    Nbr   : P-4.00
    Default Metric  : 100
    Sub TLV Len     : 12
    IF Addr   : 192.168.114.1
    Nbr IP    : 192.168.114.2
  TE IP Reach   :
    Default Metric  : 0
    Control Info:    , prefLen 16
    Prefix   : 172.16.0.0
    Default Metric  : 0
    Control Info:    , prefLen 32
    Prefix   : 192.0.2.1

Level (1) LSP Count : 1

Displaying Level 2 database
```

```
-------------------------------------------------------------------------------
Level (2) LSP Count : 0
===============================================================================
*A:PE-1#
```

Conclusion

IS-IS link bundling allows service providers to configure multiple IS-IS interfaces as a single link group for ECMP purposes and allow link metric increases if an interface within the bundle group fails. This example provides the configuration for IS-IS link bundling, together with the associated commands and outputs which can be used for verifying and troubleshooting.

Policy Chaining and Logical Expressions

This chapter provides information about policy statement chaining and logical expressions.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4. In SR OS releases earlier than 14.0.R1, only policy chaining was supported. SR OS release 14.0.R1 introduced support for route policy logical expressions using the logical operators AND, OR, and NOT. The **action drop** replaces the **action reject** in SR OS release 14.0.R4, and later.

Overview

Multiple policies can be chained together for sequential evaluation. For more complex evaluation logic, logical expressions (with operators AND, OR, and NOT) can be used. A logical expression can be included in a larger policy chain. Route policy logical expressions are supported in the following contexts:

- BGP export
- BGP import
- BGP leak-import (RIB leaking)
- VRF import
- VRF export
- GRT export (GRT leaking)

Table 5 shows a comparison between examples of policy chaining and policy logical expressions.

Table 5 **Policy Chaining Versus Policy Logical Expressions**

Policy chaining example	Policy logical expressions example
configure router bgp import "A" "B" "C"	configure router bgp import "[A] OR [B]"
For each route, policy A is evaluated first. • If policy A matches the route with **action next-policy**, then apply any route modifications and continue to evaluate policy B, and so on. • When the route is matched in a policy with **action accept** or **drop/reject**, the evaluation is completed.	Several logical operators can be used. This example shows an OR relationship between policy A and policy B. For each route, policy A is evaluated first. A true/false result is determined for policy A: • If true, the logical expression with operator OR is true already, and the evaluation is completed. • If false, then policy B is evaluated to determine the final true/false result. The final result is mapped back to a policy action (**accept**, **next-policy**, and so on).

Note: In SR OS release 14.0.R4, and later, the **action drop** replaces the **action reject**. The difference is that with **action drop**, it is possible to modify attributes in the same way as for **action accept**. This behavior is useful when a NOT operation makes a false expression true and the attributes are required. In a similar way, it is possible that an OR operation is true, even though the first policies that were evaluated were false. Routes are accepted when the final result is true and all policies that were evaluated will modify the route attributes. Examples of this behavior are included in the Configuration section.

To configure policy chaining - which may or may not include a policy logical expression - the syntax is the following:

```
*A:PE-1# configure router bgp export
  - export <plcy-or-long-expr> [<plcy-or-expr> [<plcy-or-expr>...(upto 14 max)]]
  - no export
```

```
<plcy-or-long-expr>  : <policy-name> | <long-expr>
                       <policy-name>  - [64 chars max]
                       <long-expr>    - [255 chars max]
<plcy-or-expr>       : <policy-name> | <expr>
                       <policy-name>  - [64 chars max]
                       <expr>         - [64 chars max]
```

A policy chain is 15 policies at maximum, one of which can be a logical expression. A logical expression in the policy chain can contain a maximum of 16 policies and can be anywhere in the command: the first operand can contain up to 255 characters, while the remaining operands can contain a maximum of 64 characters.

Policies in a logical expression need to be enclosed in square brackets. Like single policies, the logical expression can be enclosed in double quotes or not, as follows:

```
*A:PE-1# configure router bgp import [C]AND[A] "B"
*A:PE-1# configure router bgp import "[C]AND[A]" "B"
```

The logical expression can be anywhere in the policy chain, as follows:

```
*A:PE-1# configure router bgp import "B" [C]AND[A]
```

When quotes are used, spaces are allowed in the logical expression, as follows:

```
*A:PE-1# configure router bgp import "[A] AND [B]"
```

Without quotes, the logical operators are interpreted as policy names, as follows:

```
*A:PE-1# configure router bgp import [A] AND [B]
WARNING: CLI Policy "AND" does not exist.
```

The following message is raised when more than one logical expression is included in the policy chain:

```
*A:PE-1# configure router bgp import [C]AND[A] [C]AND[B]
INFO: BGP #1001 Configuration failed because of inconsistent values -
 BGP [VR 1] Policy stmts nbr logical expressions exceeded
```

The following message is raised when the logical expression is duplicated in the policy chain:

```
*A:PE-2# configure router bgp import [C]AND[A] "B" [C]AND[A]
INFO: BGP #1001 Configuration failed because of inconsistent values -
 BGP [VR 1] Policy stmts should be unique and set in order!
```

The following message is raised when the logical operators are not in uppercase:

```
*A:PE-1# configure router bgp import [A]and[B]
INFO: BGP #1001 Configuration failed because of inconsistent values -
 BGP [VR 1] Policy stmts expression format error
```

The following message is raised when more than 16 policies are in a policy expression:

```
*A:PE-1# configure router bgp export
[1]AND[2]AND[3]AND[4]AND[5]AND[6]AND[7]AND[8]AND[9]AND[10]AND[11]AND[12]AND[13]AND[1
4]AND[15]AND[16]


*A:PE-1# configure router bgp export [1]AND[2]AND[3]AND[4]AND[5]AND[6]AND[7]AND[8]
AND[9]AND[10]AND[11]AND[12]AND[13]AND[14]AND[15]AND[16]AND[17]
INFO: BGP #1001 Configuration failed because of inconsistent values -
 BGP [VR 1] Policy stmts expression format error
```

Route Policy Logical Expressions

Logical expressions are evaluated to be true or false. Table 6 shows the mapping of policy actions to Boolean values.

Table 6 **Boolean Values for the Policy Actions**

Policy action	Boolean value
Accept	True
Next-entry	True
Next-policy	True
Reject	False
Drop	False

Note: : The policy **action drop** replaces **action reject** in release 14.0.R4, and later. The **action drop** supports route attribute modifications while **action reject** does not. SR OS automatically converts reject actions to drop actions.

Table 7 shows the evaluation actions for the logical operators NOT, OR, and AND.

Table 7 **Actions for the Logical Operators**

Logical Operator	Action
NOT <expr>	Swaps the true/false result of the expression.

Table 7 **Actions for the Logical Operators (Continued)**

Logical Operator	Action
<expr1> OR <expr2>	If expr1 is true, the result is true and expr2 need not be evaluated. If expr1 is false, expr2 must be evaluated. The final result is true if either expression is true; otherwise, it is false.
<expr1> AND <expr2>	If expr1 is false, the result is false and expr2 need not be evaluated. If expr1 is true, expr2 must be evaluated. The final result is true only if both expressions are true.

Table 8 shows the mapping of the final result of an expression to a policy action. Routes will be dropped when the entire expression is false.

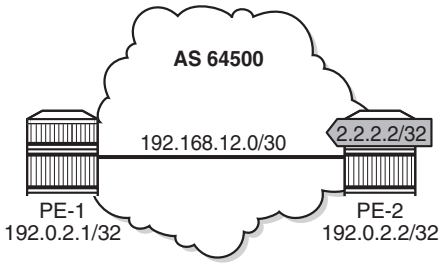
Table 8 **Mapping Final Result of an Expression to a Policy Action**

Final result	Action
True	**accept**, **next-entry**, or **next-policy** (depending on the last entry evaluated)
False	**drop/reject**

Configuration

Figure 1 shows the example topology including the advertised route.

Figure 86 **Example Topology**



The initial configuration of the routers includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS
- LDP
- BGP
- Export policy "export-bgp" accepting routes for prefix 2.2.2.2/32 on PE-2.

It is possible to configure VPRNs and assign policies to BGP in the VPRN, but in this chapter, all examples are for BGP in the base router.

Policy Chaining and Policy Logical Expressions

In this section, three route policies are configured that will add a community and set the local preference (LP): only policy C does not set LP. Policy C has **action next-policy**, and policies A and B have **action accept**. The configuration is as follows:

```
configure
    router
        policy-options
            begin
            community "A" members "1:1"
            community "B" members "2:2"
            community "C" members "3:3"
            policy-statement "A"
                entry 10
                    action accept
                        community add "A"
                        local-preference 110
                    exit
                exit
            exit
            policy-statement "B"
                entry 10
                    action accept
                        community add "B"
                        local-preference 200
                    exit
                exit
            exit
            policy-statement "C"
                entry 10
                    action next-policy
                        community add "C"
                    exit
                exit
            exit
            commit
```

Initially, policy chaining will be configured without a logical expression. Subsequently, policy chaining will be configured with only one policy logical expression and no other policies in the chain, as described in the following sections.

Policy Chaining without Logical Expression

Policy chaining may include one logical expression, but in this example, there is no policy logical expression in the chain.

Policy chaining is configured as follows on PE-1:

```
*A:PE-1# configure router bgp import "C" "A" "B"
```

PE-1 receives route 2.2.2.2/32 from PE-2. For each route, PE-1 evaluates policy C first. This policy adds community C (3:3) and has **action next-policy**, which implies that the next policy also needs to be evaluated. Policy A adds community A (1:1) and sets the LP to a value of 110 (by default, the local preference equals 100). Policy A has **action accept** and, therefore, the evaluation is completed. The local preference and the community are shown in the following output:

```
*A:PE-1# show router bgp routes hunt brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.12.2
Local Pref.    : 110                    Interface Name : int-PE-1-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : 1:1 3:3
---snip---
```

Policy Logical Expressions with Two Policies

In the following examples, the policy chain contains only a policy logical expression. When both policy A and policy B need to be executed, the logical operator to be used is AND. The sequence is important in this case, because both policies A and B set the LP and the last executed policy will set the final value for the LP. The following import policy expression is configured on PE-1:

```
*A:PE-1# configure router bgp import "[A]AND[B]"
```

Policy A is evaluated first and it adds community A (1:1) and sets LP 110. Then, policy B is evaluated, which adds community B (2:2) and sets LP 200.

```
*A:PE-1# show router bgp routes hunt brief | match "Local Pref."
Local Pref.    : 200                    Interface Name : int-PE-1-PE-2

*A:PE-1# show router bgp routes hunt brief | match "Community"
Community      : 2:2 1:1
```

When the policy expression is [B]AND[A], the order is reversed. First, policy B sets LP 200, then policy A sets LP 110, as follows:

```
*A:PE-1# show router bgp routes hunt brief | match "Local Pref."
Local Pref.    : 110                    Interface Name : int-PE-1-PE-2

*A:PE-1# show router bgp routes hunt brief | match "Community"
Community      : 1:1 2:2
```

When the policy expression contains operator OR instead of AND, the first true expression results in a completed evaluation. Because both policy A and policy B result in a true expression, whichever policy is evaluated first is executed and the second one is skipped. When policy A is evaluated first and the result is true, policy B will be skipped. Therefore, the community will be A (1:1) and the LP 110, as follows:

```
*A:PE-1# configure router bgp import "[A]OR[B]"

*A:PE-1# show router bgp routes hunt brief | match "Local Pref."
Local Pref.    : 110                    Interface Name : int-PE-1-PE-2

*A:PE-1# show router bgp routes hunt brief | match "Community"
Community      : 1:1
```

Likewise, when policy B is evaluated first and the result is true, policy A will be skipped. The added community will be B (2:2) and the LP 200, as follows:

```
*A:PE-1# configure router bgp import "[B]OR[A]"
```

```
*A:PE-1# show router bgp routes hunt brief | match "Local Pref."
Local Pref.    : 200                    Interface Name : int-PE-1-PE-2

*A:PE-1# show router bgp routes hunt brief | match "Community"
Community      : 2:2
```

The logical operator NOT swaps the result from true to false, and vice versa. When policy A is evaluated as true, NOT[A] is false. A false expression in an AND relationship leads to a false result. The next policy in the logical expression need not be evaluated. No communities will be added and no LP will be set (the default value for LP is 100). The route will be rejected as invalid, as follows:

```
*A:PE-1# configure router bgp import "NOT[A]AND[B]"

*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
---snip---
Community      : No Community Members
---snip---
Flags          : Invalid  IGP  Rejected
---snip---
```

However, a false NOT[A] expression in an OR relation may still lead to the expression being evaluated to true, as follows:

```
*A:PE-1# configure router bgp import "NOT[A]OR[B]"

*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 200                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 2:2 1:1
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

Policy B is evaluated as true for the route and, therefore, the entire logical expression "NOT[A]OR[B]" is true, and the route is accepted. Every policy in the expression that was evaluated, before the entire logical expression was recognized to be true, is executed, including policy A. This implies that policy A adds community A (1:1) to the route and sets LP to a value of 110. Then, policy B adds community B (2:2) to the route and overwrites the LP to a value of 200.

The import policy "[B] OR NOT[A]" is true after the first policy is evaluated as true. Only policy B is executed and the assigned community is B (2:2) and the LP is 200, as follows:

```
*A:PE-1# configure router bgp import "[B] OR NOT[A]"

*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 200                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 2:2
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

Table 9 summarizes the results for these different scenarios.

Table 9 **Assigned LP and Communities for the Import Logical Expressions**

Import Logical Expression	Assigned LP	Assigned Community
import "[A] AND [B]"	200	2:2 1:1
import "[B] AND [A]"	110	1:1 2:2
import "[A] OR [B]"	110	1:1
import "[B] OR [A]"	200	2:2
import "NOT[A] AND [B]"	None	None (Route rejected)
import "NOT[A] OR [B]"	200	2:2 1:1
Import "[B] OR NOT[A]"	200	2:2

Policy Logical Expressions with Three Policies

In policy chaining, the next policy in the chain will be evaluated when the action is **next-policy**. In policy logical expressions, the next policy will be evaluated depending on the logical operator and the Boolean value for the previous policies in the expression.

Policy C has **action next-policy** instead of **accept** and adds community C (3:3), but does not set the LP.

Several logical expressions can be made with policies A, B, and C. The following import policy has all three policies in an AND relationship. The expression is evaluated as true and all policies are executed: three communities are added and the LP is set.

```
*A:PE-1# configure router bgp import "[C]AND[A]AND[B]"
```

The first policy adds community C (3:3), the second policy adds community A (1:1) and sets LP 110, and the third policy adds community B (2:2) and sets LP 200, as follows:

```
*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 200                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 2:2 1:1 3:3
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

The import policy "[C]AND[A]OR[B]" results in the first two being executed. Policy C is evaluated as true, and the logical operation is AND. Therefore, the next policy needs to be evaluated too. Policy A is also evaluated as true and the next operation is OR. The final result is evaluated as true without evaluating policy B. The communities added are C and A (3:3 and later 1:1) and the LP is 110.

```
*A:PE-1# configure router bgp import "[C]AND[A]OR[B]"

*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
```

```
---snip---
Local Pref.    : 110                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 1:1 3:3
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

The import policy "[C]OR[A]OR[B]" is evaluated as true after the first policy is evaluated as true. Even though the action in policy C is **next-policy**, the next policy in this expression does not need to be evaluated, because the expression is true. Only policy C is executed and it adds the community C (3:3), but does not configure the LP, as follows:

```
*A:PE-1# configure router bgp import "[C]OR[A]OR[B]"

*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 3:3
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

However, if the policy chain contains not only a logical expression, but also single policies, the action next-policy ensures that a following policy in the chain is executed; for example, policy D in the following policy chain:

```
*A:PE-1# configure router bgp import "[C]OR[A]OR[B]" "D"
```

The expression "[C]OR[A]OR[B]" is true after policy C has been evaluated, but policy C has as **action next-hop** and policy D is the next policy to be evaluated.

The import policy "[C]OR[A]AND[B]" evaluates policy C as true. Policy C has an OR relation with policy A in the logical expression {[C]OR[A]}, and therefore, policy A need not be evaluated. There is an AND relation with policy B and policy B is evaluated as true. Therefore, the entire logical expression "[C]OR[A]AND[B]" is true and the route will be accepted. Both policy C and B are executed. First, policy C adds community C (3:3), then policy B adds community B (2:2) and sets LP 200, as follows:

```
*A:PE-1# configure router bgp import "[C]OR[A]AND[B]"

*A:PE-1# show router bgp routes hunt brief
```

```
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 200                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 2:2 3:3
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

Table 10 summarizes the results for these different scenarios.

Table 10 **Assigned LP and Communities for the Import Logical Expressions**

Import Logical Expression	Assigned LP	Assigned Community
import "[C] AND [A] AND [B]"	200	2:2 1:1 3:3
import "[C] AND [A] OR [B]"	110	1:1 3:3
import "[C] OR [A] OR [B]"	None	3:3
import "[C] OR [A] AND [B]"	200	2:2 3:3

Combinations of Policy Logical Operations Using Brackets

For this section, the following communities and policies are configured on PE-1. All these policies have a **from** condition that matches a community (D, E, F, G). Besides these policies, there are also export policies on PE-2 that add one or more communities (D, E, F, G) to the advertised routes. On PE-1, incoming route 2.2.2.2/32 will have one or more communities that may or may not match the **from** condition in the following route policies.

```
configure
    router
        policy-options
            begin
```

```
community "D" members "4:4"
community "E" members "5:5"
community "F" members "6:6"
community "G" members "7:7"
policy-statement "D"
    entry 10
        from
            community "D"
        exit
        action accept
            local-preference 4
        exit
    exit
    default-action drop
    exit
exit
policy-statement "E"
    entry 10
        from
            community "E"
        exit
        action accept
            local-preference 5
        exit
    exit
    default-action drop
    exit
exit
policy-statement "F"
    entry 10
        from
            community "F"
        exit
        action accept
            local-preference 6
        exit
    exit
    default-action drop
    exit
exit
policy-statement "G"
    entry 10
        from
            community "G"
        exit
        action accept
            local-preference 7
        exit
    exit
    default-action drop
    exit
exit
commit
```

The received routes have community E (5:5) present. The following import policy is configured on PE-1:

```
*A:PE-1# configure router bgp import "([D]AND[E])OR([F]AND[G])"
```

The first policy that is evaluated requires community D (4:4) to be present. This is not the case and the expression between brackets, ([D]AND[E]), is false. Policy E need not be evaluated. The next policy to be evaluated is F and it requires community F (6:6), which is not present. The second expression between brackets, ([F]AND[G]), is therefore also false and policy G need not be evaluated. The entire policy logical expression is false and the route will be rejected.

The following commands show what policy evaluation caused the route to be rejected. For the entire logical expression "([D]AND[E])OR([F]AND[G])", the last policy that was evaluated, and that caused the route to be rejected, was policy F, as follows:

```
*A:PE-1# show router bgp policy-test "([D]AND[E])OR([F]AND[G])" family ipv4
prefix 0.0.0.0/0 longer display-rejects brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network
-------------------------------------------------------------------------------
Rejected by Logical expression last policy F Default action
      2.2.2.2/32
-------------------------------------------------------------------------------
 Total Routes : 1 Routes rejected : 1
```

For the logical expression "[D]AND[E]", the last policy that was evaluated, and that led to the conclusion that the expression was false, was policy D, as follows:

```
*A:PE-1# show router bgp policy-test "[D]AND[E]" family ipv4 prefix 0.0.0.0/0
longer display-rejects brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network
-------------------------------------------------------------------------------
Rejected by Logical expression last policy D Default action
      2.2.2.2/32
-------------------------------------------------------------------------------
 Total Routes : 1 Routes rejected : 1
```

For the logical expression "[F]AND[G]", the last policy that was evaluated, and that led to the conclusion that the expression was false, was policy F, as follows:

```
*A:PE-1# show router bgp policy-test "[F]AND[G]" family ipv4 prefix 0.0.0.0/0
longer display-rejects brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network
-------------------------------------------------------------------------------
Rejected by Logical expression last policy F Default action
      2.2.2.2/32
-------------------------------------------------------------------------------
 Total Routes : 1 Routes rejected : 1
```

The logical expression "([D]AND[E])OR([F]AND[G])" is false and, therefore, the route is rejected, as follows. No LP will be set. Community E (5:5) was already present in the incoming route.

```
*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 5:5
---snip---
Flags          : Invalid  IGP  Rejected
---snip---
```

In the second example, the incoming route contains communities D (4:4) and E (5:5). The same policy logical expression "([D]AND[E])OR([F]AND[G])" is evaluated as true because both policy D and policy E are true. There is an OR relationship with the rest of the expression and, therefore, the entire logical expression is true. Policy E is the last policy to be evaluated, as follows:

```
*A:PE-1# show router bgp policy-test "([D]AND[E])OR([F]AND[G])" family ipv4
prefix 0.0.0.0/0 longer display-rejects brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network
-------------------------------------------------------------------------------
Accepted by Logical expression last policy E Entry 10
      2.2.2.2/32
-------------------------------------------------------------------------------
Routes : 1
```

The route is accepted as valid and gets LP 5. The communities D (4:4) and E (5:5) were already present for the incoming route. The first policy that was executed, was policy D and it set the LP to a value of 4. Policy E was the second and last policy that was executed and it set the LP to a value of 5, as follows:

```
*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 5                      Interface Name : int-PE-1-PE-2
---snip---
Community      : 5:5 4:4
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
---snip---
```

For the third example, the incoming route contains communities D (4:4) and E (5:5). The logical expression "([D]OR[E])AND([F]OR[G])" will be evaluated as false and the route will be rejected, follows:

```
*A:PE-1# configure router bgp import "([D]OR[E])AND([F]OR[G])"
```

First, policy D is evaluated as true because community D (4:4) is present. Policy D has an OR relationship with policy E, which will be true without the need to evaluate policy E. The next policy to be evaluated is F. Policy F requires the community F (6:6) to be present, which is not the case. The logical expression [F]OR[G] can only be true if policy G is true. Policy G requires community G (7:7) to be present, which is false. The last policy that was evaluated before the route was rejected was policy G, as follows:

```
*A:PE-1# show router bgp policy-test "([D]OR[E])AND([F]OR[G])" family ipv4
prefix 0.0.0.0/0 longer display-rejects brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network
-------------------------------------------------------------------------------
Rejected by Logical expression last policy G Default action
      2.2.2.2/32
-------------------------------------------------------------------------------
 Total Routes : 1 Routes rejected : 1
```

The route was rejected and, therefore, no policy was executed. The LP kept its default value of 100, as follows:

```
*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 5:5 4:4
---snip---
Flags          : Invalid  IGP  Rejected
---snip---
```

For the fourth example, the incoming route has communities E (5:5) and G (7:7). The logical expression "([D]OR[E])AND([F]OR[G])" will be evaluated as true and the route will be accepted. First, policy D is evaluated as false. Policy D has an OR relationship with policy E, which will be evaluated as true. Consequently, the expression [D]OR[E] is true. This expression has an AND relationship with the expression [F]OR[G].

The next policy to be evaluated is F. Policy F requires the community F (6:6) to be present, which is false. The logical expression [F]OR[G] can only be true if policy G is true. Policy G requires community G (7:7) to be present, which is true. This makes [F]OR[G] true and also the entire expression "([D]OR[E])AND([F]OR[G])".

The last policy that was evaluated before the route was accepted was policy G, as follows:

```
*A:PE-1# show router bgp policy-test "([D]OR[E])AND([F]OR[G])" family ipv4
prefix 0.0.0.0/0 longer display-rejects brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
```

```
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network
-------------------------------------------------------------------------------
Accepted by Logical expression last policy G Entry 10
      2.2.2.2/32
-------------------------------------------------------------------------------
Routes : 1
```

The route was accepted and has the changes of all policies that were evaluated: initially, policy D set the LP to 4. This value was overwritten by policy E to 5, by policy F to 6, and finally by policy G to a value of 7, as follows:

```
*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 7                      Interface Name : int-PE-1-PE-2
---snip---
Community      : 5:5 7:7
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

Table 11 summarizes the results for these different scenarios.

Table 11 **Assigned LP for the Import Logical Expressions**

Ingress Community	Import Logical Expression	Assigned LP
5:5	import "([D]AND[E])OR([F]AND[G])"	Prefix rejected
5:5 4:4	import "([D]AND[E])OR([F]AND[G])"	5
5:5 4:4	import "([D]OR[E])AND([F]OR[G])"	Prefix rejected
5:5 7:7	import "([D]OR[E])AND([F]OR[G])"	7

Modification of Attributes While Processing

During the policy evaluation process, some prefix attributes can be modified while processing, and these modified attributes can be used as criteria for other policies in the logical expression.

In the following example, two route policies are configured:

- Policy X adds a new community Y (11:11) to the incoming route update.
- Policy Y uses community Y (11:11) as the only match criterion and removes communities X and Y. Policy Y also sets the LP to a value of 9, which is used here as an indication that policy Y was executed.

An export policy on PE-2 adds community X (10:10) to prefix 2.2.2.2/32 (not shown here).

Route policies X and Y are configured as follows on PE-1:

```
configure
    router
        policy-options
            begin
            community "X" members "10:10"
            community "Y" members "11:11"
            policy-statement "X"
                entry 10
                    from
                        community "X"
                    exit
                    action accept
                        community add "Y"
                    exit
                exit
            exit
            policy-statement "Y"
                entry 10
                    from
                        community "Y"
                    exit
                    action accept
                        community remove "X" "Y"
                        local-preference 9
                    exit
                exit
            exit
            commit
```

When no import policy is applied on PE-1, the received route 2.2.2.2/32 has community 10:10 and the default LP, as follows:

```
*A:PE-1# show router bgp routes hunt brief
---snip---
-------------------------------------------------------------------------------
```

```
RIB In Entries
------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
---snip---
Community      : 10:10
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

The import policy "[X]AND[Y]" is configured on PE-1, as follows:

```
*A:PE-1# configure router bgp import "[X]AND[Y]"
```

The route update contains community X (10:10) and policy X is evaluated as true. Policy X adds community Y (11:11) to the route. Policy Y requires this community and is evaluated as true. Therefore, the entire logical expression "[X]AND[Y]" is true and the route is accepted. Policy Y removes communities X (10:10) and Y (11:11), and sets the LP to a value of 9, as follows:

```
*A:PE-1# show router bgp routes hunt brief
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
------------------------------------------------------------------------------
RIB In Entries
------------------------------------------------------------------------------
Network        : 2.2.2.2/32
Nexthop        : 192.0.2.2
---snip---
Local Pref.    : 9                      Interface Name : int-PE-1-PE-2
---snip---
Community      : No Community Members
---snip---
Flags          : Used  Valid  Best  IGP
---snip---
```

Conclusion

Route policy chaining and logical expressions allow complex route processing logic to be broken into smaller components. These policy components are reusable and facilitate the process of updating route control logic. Logical expressions offer more flexible combinations of policy statements.

Separate BGP RIBs for Labeled Routes

This chapter provides information about separate border gateway protocol (BGP) route information bases (RIBs) for labeled-unicast routes.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4.

Release 14.0.R4 introduced separate BGP RIBs for labeled-unicast routes.

Overview

In SR OS releases earlier than 14.0.R4, BGP maintained one RIB for both unlabeled IPv4 and labeled-unicast IPv4 routes and another RIB for both unlabeled IPv6 and labeled-unicast IPv6 routes. In SR OS release 14.0.R4, and later, the labeled-unicast and unlabeled routes are held in separate RIBs: IPv4, label-IPv4, IPv6, and label-IPv6.

Table 12 compares the previous BGP RIB architecture to the current one.

Table 12 **Comparison of the BGP RIB Architectures**

	SR OS earlier than 14.0.R4	SR OS 14.0.R4 and later
BGP peer with both labeled-unicast (SAFI 4) and unlabeled IP (SAFI 1) routes	SAFI 1 and SAFI 4 are mutually exclusive. Received routes with the non-negotiated SAFI are treated as route withdrawals.	Both labeled and unlabeled routes can be advertised to the same neighbor.

Table 12 **Comparison of the BGP RIB Architectures (Continued)**

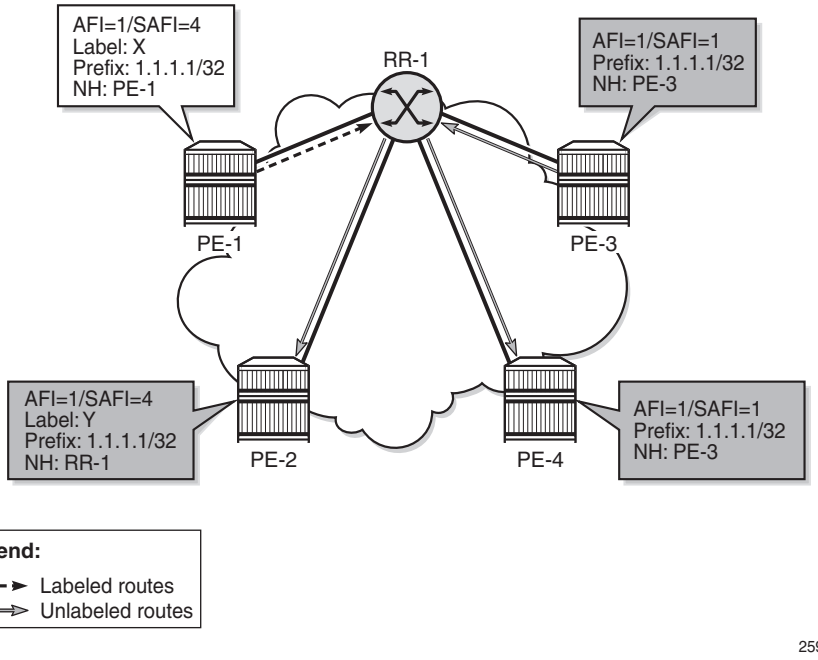
	SR OS earlier than 14.0.R4	SR OS 14.0.R4 and later
RR default behavior	BGP RR re-advertises all types of routes to all its clients, including labeled routes to unlabeled peers and vice versa.	BGP RR only re-advertises labeled-unicast routes to labeled-unicast peers and unlabeled routes to unlabeled IP peers, unless route-table-import policies are used to allow route leaking.
Route-table import policies	No route-table import policies supported. All active routes are automatically imported to BGP RIB.	Route-table import policies can save BGP memory by blocking the import of unnecessary IP routes. All unlabeled and all labeled BGP-owned routes are by default imported into both BGP RIBs.

BGP Common IPv4 RIB Implementation

In SR OS releases earlier than 14.0.R4, unlabeled IP routes and labeled IP routes for the same prefix were considered different paths to the same destination. SR OS could not advertise both labeled and unlabeled routes to the same neighbor. If the advertise-label ipv4 command was configured on a session, labeled routes were advertised, and unlabeled routes were not expected and not sent. For labeled routes, scarce MPLS datapath resources were consumed. In many networks, advertising a labeled route was only required for certain prefixes and not the entire IPv4 or IPv6 routing table.

Another issue with the common IPv4 RIB architecture is shown in Figure 87.

Figure 87 **RR-1 with Common IPv4 RIB**



Route reflector RR-1 has two clients sending labeled-IPv4 routes (PE-1, PE-2) and two other clients sending unlabeled IPv4 routes (PE-3, PE-4). The RR is intended to be a control plane RR and not configured to perform next-hop-self toward any of the clients. When both PE-1 (labeled) and PE-3 (unlabeled) send a route for IPv4 prefix 1.1.1.1/32, the RR uses the BGP decision process to select the best path for 1.1.1.1/32. Assume that PE-3 sends the best path. This path will be advertised to clients PE-2 (labeled) and PE-4 (unlabeled).

The route received from PE-3 is unlabeled, and can be advertised as an unlabeled route to peer PE-4. However, toward PE-2, the route must get a label Y, according to the configured label in the **advertise-label ipv4** command, but the next hop cannot remain PE-3, because PE-3 does not expect any labels. Therefore, RR-1 must set next-hop-self and data traffic between PE-2 and PE-3 will flow via RR-1 where label Y will be pushed or popped. This is not a desirable situation.

BGP Separate Labeled-IPv4 RIB Implementation

In SR OS release 14.0.R4, and later, a separate RIB is used for labeled-IPv4 routes. With this implementation, client PE-2 learns the best labeled-IPv4 route and client PE-4 learns the best unlabeled IPv4 route. RR-1 does not need to set next-hop-self and traffic can be sent directly from PE-2 to PE-1 and from PE-4 to PE-3. The RR is used only for control traffic, as intended. Figure 88 shows how RR-1 sends a labeled-IPv4 route to PE-2 with label X and next hop PE-1.

Figure 88 **RR-1 with Separate Labeled-IPv4 RIB Implementation**

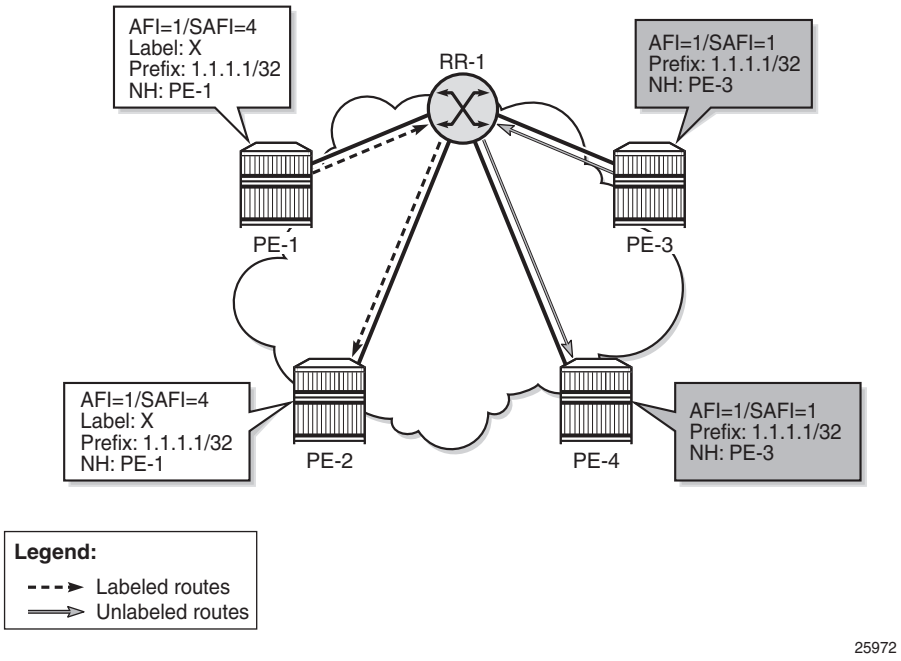
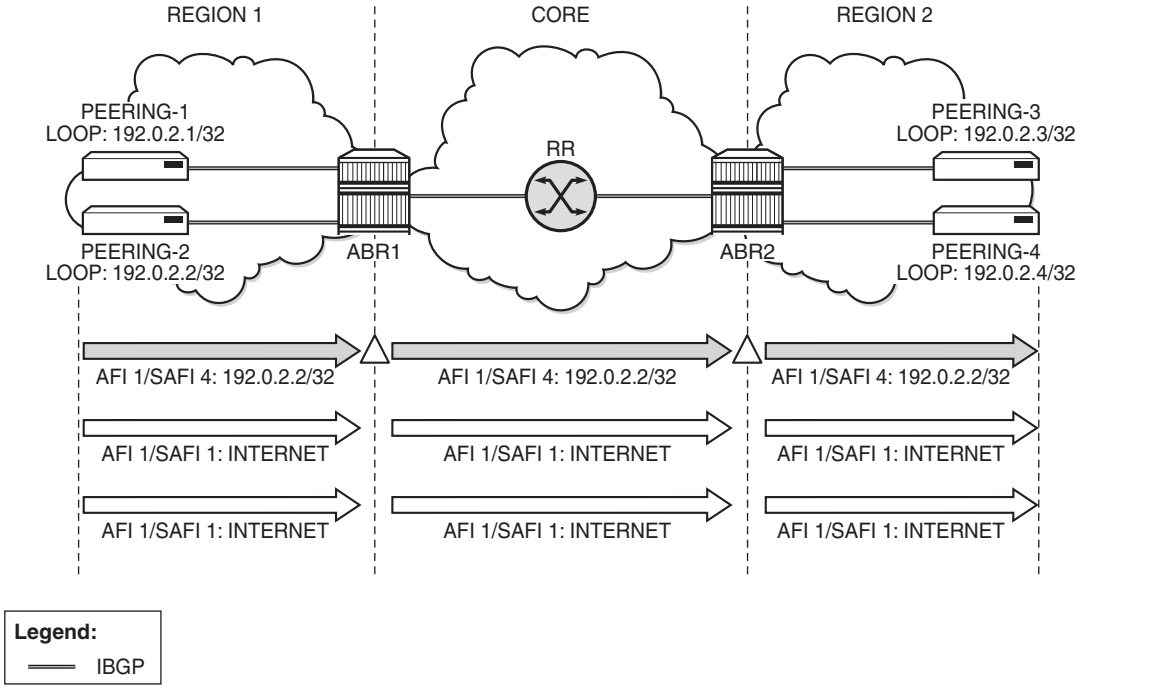


Figure 89 shows a seamless MPLS use case, which is a good example of the coexistence of labeled (AFI 1/SAFI 4) and unlabeled (AFI 1/SAFI 1) BGP sessions.

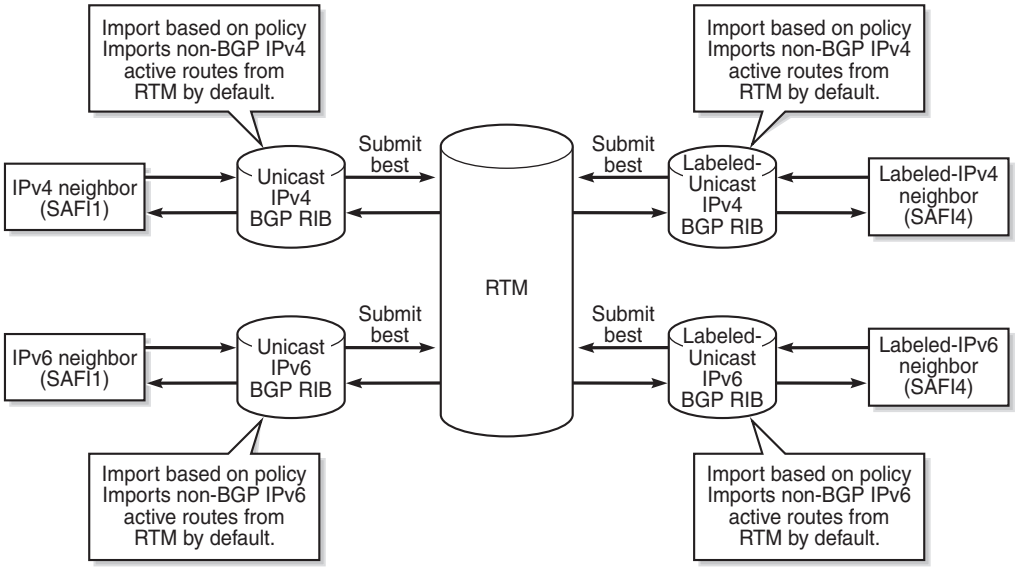
Figure 89 **Seamless MPLS - Separate Labeled-IPv4 Implementation**



RIB Architecture

Figure 90 shows the system architecture with four separate RIBs for IPv4 and IPv6 routes.

Figure 90 **System Architecture with Separate RIBs for Labeled-Unicast and Unlabeled Routes**



Labeled-unicast routes from peers are stored in a labeled RIB and unlabeled routes from the same or different peers are stored in a non-labeled RIB. Both labeled and unlabeled routes can be sent and received to and from the same peer. Different sets of routes can be advertised to labeled/unlabeled peers. Labeled and unlabeled BGP sessions are using the common equal cost multipath (ECMP) and multipath limit.

More user control is provided over the RTM route import process. By default, a RIB imports all non-BGP active routes from RTM, but a user-defined route policy can be applied. Route policies can be used to reduce BGP memory usage.

The configuration for labeled IP routes has changed in SR OS release 14.0.R4:

- The **advertise-label** command is deprecated.
- Address families mapped to the RIBs are: **ipv4, label-ipv4, ipv6, label-ipv6**.
- In route policies, protocol types **bgp** and **bgp-label** can be used.
- The default RTM preference for labeled IP routes is configurable (**label-preference**) in the BGP context of the base router or a VPRN. The default preference is 170.

The consequences of having four separate RIBs instead of two are the following:

- Increased BGP memory usage is possible because the same route may be present in two RIBs. This can be avoided by a good route-table-import policy.

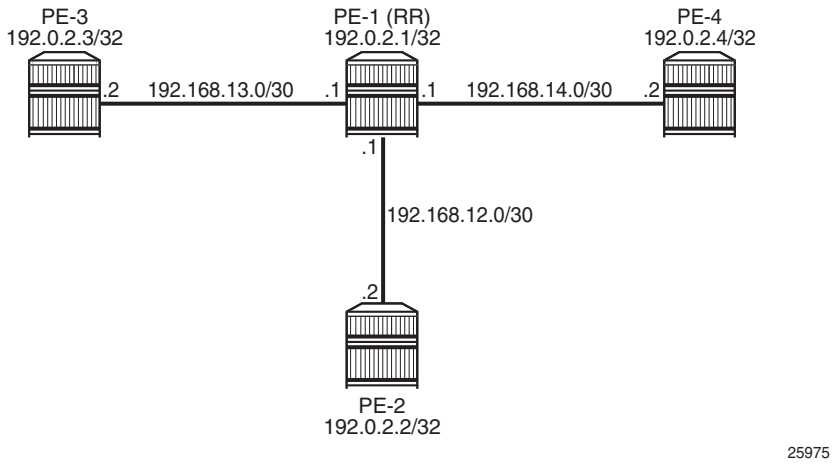
- By default, BGP unlabeled routes are no longer advertised to BGP label peers and vice versa. However, non-default route-table-import policies can accept BGP routes of the other type.
- The paths that are advertised to a peer for a prefix may be different, even if full routes are leaked between labeled and unlabeled RIBs. Leaking is done through the RTM and each RIB submits its best path. However, by default, the same route will be sent to both the labeled and unlabeled peer if the route-table-import policy allows the import to the BGP/BGP-Label owned routes.
- Care must be taken during upgrade.

Configuration

All the examples are based on labeled and unlabeled IPv4 addresses. For IPv6, the configuration is similar.

Figure 91 shows the example topology using IPv4 addresses.

Figure 91 **Example IPv4 Topology**



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS in AS 64500 (PE-1, PE-2, PE-4)
- LDP in AS 64500
- Loopback addresses 3.3.3.3/32 in PE-3 and 4.4.4.4/32 in PE-4

- Export policy "export-bgp" accepting routes from protocol direct on all nodes

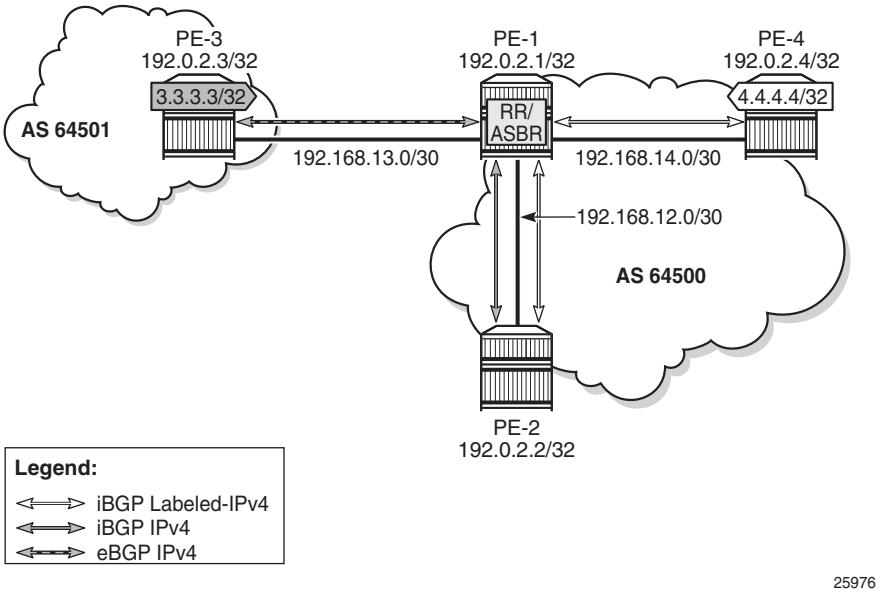
The following will be configured and verified:

1. Coexistence of labeled and unlabeled address families for BGP
2. Applying next-hop-self
3. Export Policy to Advertise Route as Labeled/Unlabeled
4. Behavior of RR with a mix of labeled and unlabeled iBGP sessions

Coexistence of Labeled and Unlabeled Address Families for BGP

Figure 92 shows the eBGP and iBGP sessions that are established between the nodes and the routes advertised for the loopback addresses.

Figure 92 **BGP Sessions**



PE-1 acts as RR for PE-2 and PE-4, and it is an autonomous system border router (ASBR) toward PE-3. PE-1 has two single-family connections: unlabeled IPv4 to PE-3 and labeled IPv4 to PE-4. PE-1 also has one dual-family connection to PE-2. The BGP configuration on PE-1 is as follows:

```
configure
```

```
    router
        autonomous-system 64500
        bgp
            min-route-advertisement 1
            split-horizon
            group "eBGP"
                peer-as 64501
                neighbor 192.168.13.2
                    family ipv4
                exit
            exit
            group "iBGP"
                cluster 192.0.2.1
                export "export-bgp"
                peer-as 64500
                neighbor 192.0.2.2
                    family ipv4 label-ipv4
                exit
                neighbor 192.0.2.4
                    family label-ipv4
                exit
            exit
        exit
    exit
exit
```

The BGP configuration on PE-2 is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            min-route-advertisement 1
            split-horizon
            group "iBGP"
                export "export-bgp"
                peer-as 64500
                neighbor 192.0.2.1
                    family ipv4 label-ipv4
                exit
            exit
        exit
    exit
exit
```

The BGP configuration on PE-3 in AS 64501 is as follows:

```
configure
    router
        autonomous-system 64501
        bgp
            min-route-advertisement 1
            split-horizon
            group "eBGP"
                export "export-bgp"
                peer-as 64500
                neighbor 192.168.13.1
```

```
                family ipv4
            exit
        exit
    exit
exit
```

The BGP configuration on PE-4 is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            min-route-advertisement 1
            split-horizon
            group "iBGP"
                export "export-bgp"
                peer-as 64500
                neighbor 192.0.2.1
                    family label-ipv4
                exit
            exit
        exit
    exit
exit
```

The BGP summary on PE-1 shows that there is a dual-family connection with PE-2: IPv4 and Lbl-IPv4. PE-1 has an Lbl-IPv4 connection with PE-4 and an IPv4 connection with PE-3.

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 2           Total Peers                 : 3
Total BGP Paths         : 14          Total Path Memory           : 2808
Total IPv4 Remote Rts   : 5           Total IPv4 Rem. Active Rts  : 2
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
```

```
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 5           Total LblIpv4 Rem. Act Rts  : 1
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
               64500        5    0 00h00m25s 2/0/6 (IPv4)
                            8    0           2/0/5 (Lbl-IPv4)
192.0.2.4
               64500        5    0 00h00m32s 3/1/4 (Lbl-IPv4)
                            7    0
192.168.13.2
               64501        5    0 00h00m43s 3/2/0 (IPv4)
                            5    0
-------------------------------------------------------------------------------
*A:PE-1#
```

The unlabeled IPv4 routes on PE-1 include unlabeled routes imported from PE-2 and PE-3, including the loopback address 3.3.3.3/32 advertised by PE-3, as follows:

```
*A:PE-1# show router bgp routes ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.3.3.3/32                                           None        None
      192.168.13.2                                         None        -
      64501
*i    192.0.2.2/32                                         100         None
      192.0.2.2                                            None        -
      No As-Path
u*>i  192.0.2.3/32                                         None        None
```

```
      192.168.13.2                                          None        -
      64501
*i    192.168.12.0/30                                       100         None
      192.0.2.2                                             None        -
      No As-Path
*i    192.168.13.0/30                                       None        None
      192.168.13.2                                          None        -
      64501
-------------------------------------------------------------------------------
Routes : 5
===============================================================================
*A:PE-1#
```

The labeled-unicast IPv4 routes on PE-1 include labeled routes imported from PE-2 and PE-4, including the loopback address 4.4.4.4/32 advertised by PE-4, as follows:

```
*A:PE-1# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
===============================================================================
Flag  Network                                               LocalPref   MED
      Nexthop (Router)                                      Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  4.4.4.4/32                                            100         None
      192.0.2.4                                             None        262139
      No As-Path
*i    192.0.2.2/32                                          100         None
      192.0.2.2                                             None        262139
      No As-Path
*i    192.0.2.4/32                                          100         None
      192.0.2.4                                             None        262140
      No As-Path
*i    192.168.12.0/30                                       100         None
      192.0.2.2                                             None        262140
      No As-Path
*i    192.168.14.0/30                                       100         None
      192.0.2.4                                             None        262138
      No As-Path
-------------------------------------------------------------------------------
Routes : 5
===============================================================================
*A:PE-1#
```

PE-2 imports the prefix 3.3.3.3/32 in its unlabeled RIB, as follows:

```
*A:PE-2# show router bgp routes 3.3.3.3/32
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
```

```
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     3.3.3.3/32                                           100         None
      192.168.13.2                                         None        -
      64501
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

PE-2 imports the prefix 4.4.4.4/32 in its labeled RIB, as follows:

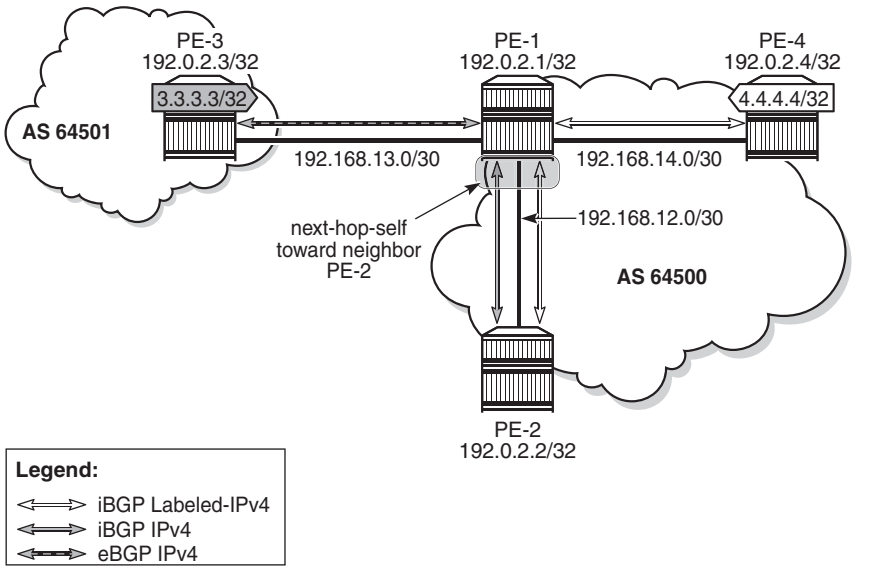
```
*A:PE-2# show router bgp routes 4.4.4.4/32 label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  4.4.4.4/32                                           100         None
      192.0.2.4                                            None        262139
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

As expected, the prefixes from address family label-ipv4 are advertised independently from the prefixes from address family ipv4.

Applying Next-Hop-Self

Figure 93 shows that PE-1 applies next-hop-self for BGP updates toward PE-2.

Figure 93 **PE-1 Applies Next-Hop-Self toward Neighbor PE-2**



On PE-1, next-hop-self is enabled for neighbor PE-2 only, as follows:

```
*A:PE-1# configure router bgp group "iBGP" neighbor 192.0.2.2 next-hop-self
```

This applies to both address families. The next hop for unlabeled route 3.3.3.3/32 will be 192.0.2.1, as follows:

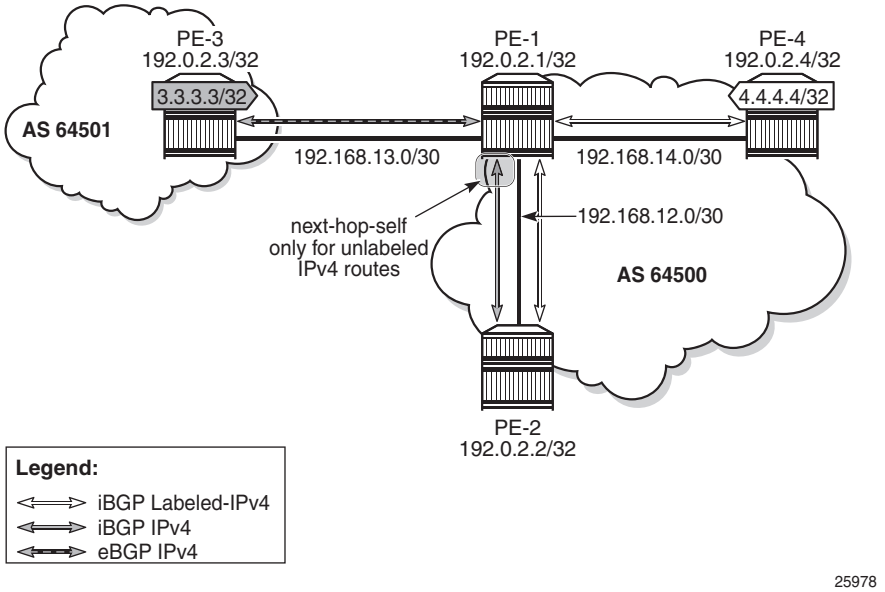
```
*A:PE-2# show router bgp routes 3.3.3.3/32
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.3.3.3/32                                           100         None
      192.0.2.1                                            None        -
      64501
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

The labeled-unicast route 4.4.4.4/32 also has next hop 192.0.2.1, as follows:

```
*A:PE-2# show router bgp routes 4.4.4.4/32 label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  4.4.4.4/32                                           100         None
      192.0.2.1                                            None        262136
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

Figure 94 shows that next-hop-self is applied to unlabeled IPv4 routes only.

Figure 94 **Applying Next-Hop-Self to Unlabeled IP-4 Routes to Neighbor PE-2**



An export policy is configured to ensure that next-hop-self is only applied for address family ipv4. The route policy is configured as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "export-nhs"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp-label
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

The next-hop-self configuration for neighbor PE-2 is replaced by export policy "export-nhs", as follows:

```
*A:PE-1# configure router bgp group "iBGP" neighbor 192.0.2.2 no next-hop-self
*A:PE-1# configure router bgp group "iBGP" neighbor 192.0.2.2 export "export-nhs"
```

With this export policy, only the unlabeled route 3.3.3.3/32 will have next hop 192.0.2.1, while the labeled-unicast route 4.4.4.4/32 will have next hop 192.0.2.4, as follows:

```
*A:PE-2# show router bgp routes 3.3.3.3/32
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.3.3.3/32                                           100         None
      192.0.2.1                                            None        -
      64501
-------------------------------------------------------------------------------
Routes : 1
```

```
===============================================================================
*A:PE-2# show router bgp routes 4.4.4.4/32 label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  4.4.4.4/32                                           100         None
      192.0.2.4                                            None        262139
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

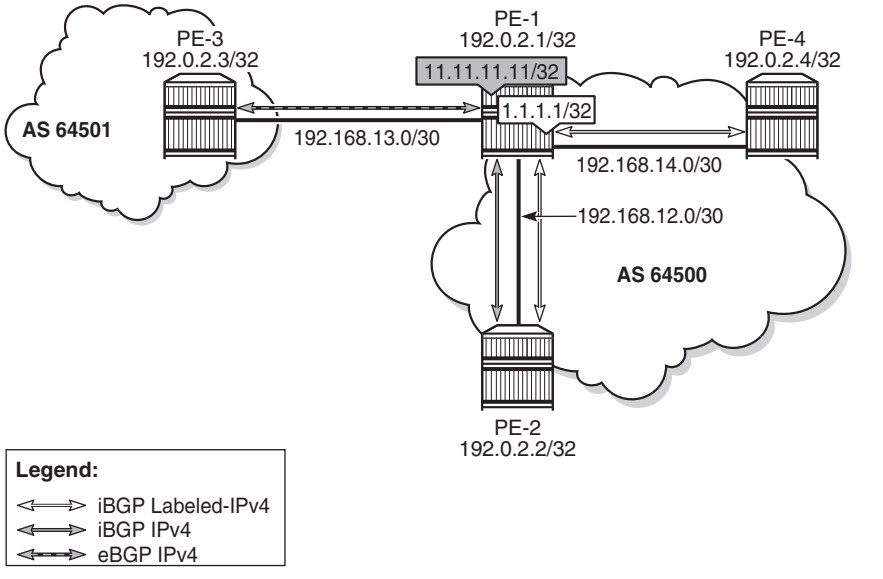
The export policy "export-nhs" toward neighbor PE-2 is removed as follows:

```
*A:PE-1# configure router bgp group "iBGP" neighbor 192.0.2.2 no export
```

Export Policy to Advertise Route as Labeled/ Unlabeled

Figure 95 shows that two loopback addresses are configured in PE-1 to be advertised: prefix 1.1.1.1/32 and 11.11.11.11/32. Initially, there is no route policy applied for a selective export as labeled or unlabeled route.

Figure 95 **PE-1 Advertises Prefixes 1.1.1.1/32 and 11.11.11.11/32**



By default, these prefixes will be advertised as both labeled and unlabeled routes toward dual-family neighbor PE-2. On PE-2, the unlabeled IPv4 RIB contains prefixes 1.1.1.1/32 and 11.11.11.11/32, as follows:

```
*A:PE-2# show router bgp routes
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  1.1.1.1/32                                           100         None
      192.0.2.1                                            None        -
      No As-Path
---snip---
u*>i  11.11.11.11/32                                       100         None
      192.0.2.1                                            None        -
      No As-Path
---snip---
```

The labeled-IPv4 RIB on PE-2 also contains prefixes 1.1.1.1/32 and 11.11.11.11/32, as follows:

```
*A:PE-2# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
===============================================================================
Flag  Network                                                LocalPref   MED
      Nexthop (Router)                                       Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
*>i   1.1.1.1/32                                           100         None
      192.0.2.1                                            None        262136
      No As-Path
---snip---
*>i   11.11.11.11/32                                       100         None
      192.0.2.1                                            None        262135
      No As-Path
---snip---
```

In many cases, it is not required to advertise both a labeled route and an unlabeled route. The following policy is configured to advertise prefix 1.1.1.1/32 as a labeled-IPv4 route and prefix 11.11.11.11/32 as an unlabeled IPv4 route:

```
configure
    router
        policy-options
            begin
            prefix-list "1.1.1.1/32"
                prefix 1.1.1.1/32 exact
            exit
            prefix-list "11.11.11.11/32"
                prefix 11.11.11.11/32 exact
            exit
            policy-statement "export-bgp1"
                entry 10
                    from
                        prefix-list "1.1.1.1/32"
                    exit
                    to
                        protocol bgp-label
                    exit
                    action accept
                    exit
                exit
                entry 30
                    from
                        prefix-list "11.11.11.11/32"
                    exit
                    to
```

```
                    protocol bgp
                exit
                action accept
                exit
            exit
            default-action drop
            exit
        exit
        commit
```

This policy is applied on PE-1 as an export policy for neighbor PE-2, as follows:

```
*A:PE-1# configure router bgp group "iBGP" neighbor 192.0.2.2 export "export-bgp1"
```

Prefix 11.11.11.11/32 is received as an unlabeled route on PE-2 and stored in the unlabeled RIB, but prefix 1.1.1.1/32 is not, as follows:

```
*A:PE-2# show router bgp routes
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  3.3.3.3/32                                           100         None
      192.168.13.2                                         None        -
      64501
u*>i  11.11.11.11/32                                       100         None
      192.0.2.1                                            None        -
      No As-Path
u*>i  192.0.2.3/32                                         100         None
      192.168.13.2                                         None        -
      64501
-------------------------------------------------------------------------------
Routes : 3
```

On PE-2, prefix 1.1.1.1/32 is received as a labeled route and stored in the labeled-IPv4 RIB, but prefix 11.11.11.11/32 is not, as follows:

```
*A:PE-2# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
```

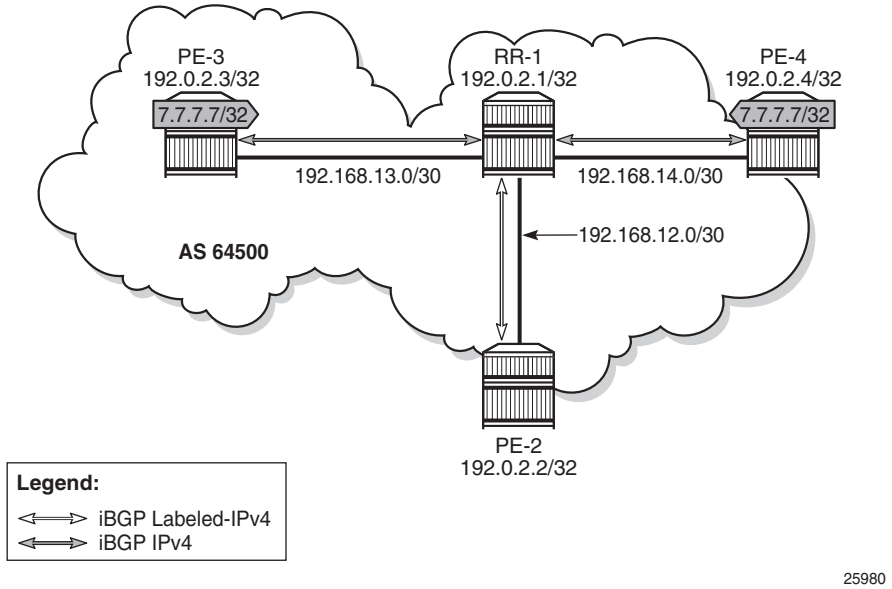
```
BGP Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  1.1.1.1/32                                       100         None
      192.0.2.1                                        None        262136
      No As-Path
u*>i  4.4.4.4/32                                       100         None
      192.0.2.4                                        None        262138
      No As-Path
-------------------------------------------------------------------------------
Routes : 2
```

This selective route advertisement from PE-1 reduces the memory usage for the RIBs on PE-2.

RR Behavior with a Mix of Labeled and Unlabeled BGP Sessions

Figure 96 shows a slightly different setup, with all PEs in the AS 64500 and RR-1 acting as the RR for all PEs. There are no dual-family connections. PE-3 and PE-4 have an unlabeled BGP session with RR-1 and PE-2 has a labeled BGP connection with RR-1. RR-1 has add-path=2 capability configured for neighbor PE-2. RR-1 receives the same prefix 7.7.7.7/32 from two neighbors: PE-3 and PE-4.

Figure 96 **RR with Labeled and Unlabeled BGP Sessions**



On RR-1, BGP is configured as follows:

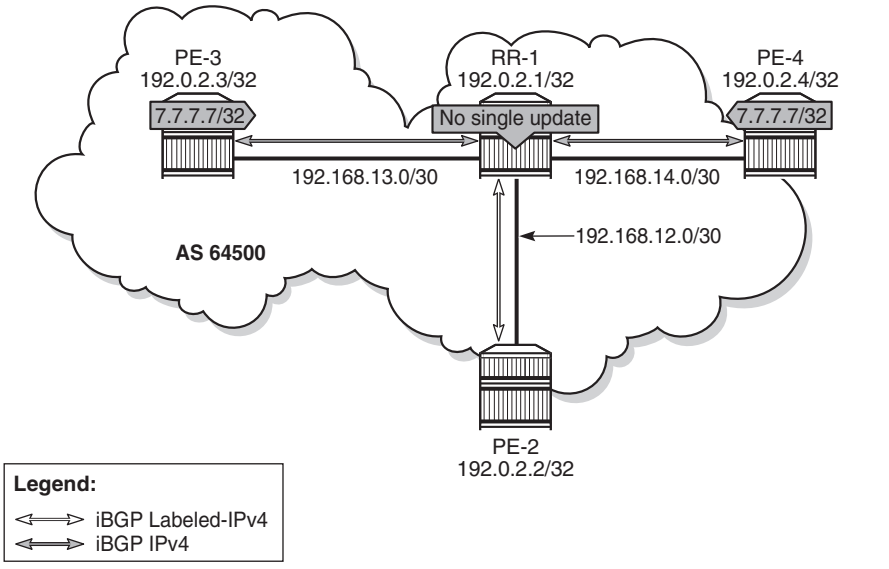
```
configure
    router
        bgp
            min-route-advertisement 1
            split-horizon
            group "iBGP"
                cluster 192.0.2.1
                export "export-bgp"
                peer-as 64500
                neighbor 192.0.2.2
                    family label-ipv4
                    add-paths
                        label-ipv4 send 2 receive
                    exit
                exit
                neighbor 192.0.2.3
                    family ipv4
                exit
                neighbor 192.0.2.4
                    family ipv4
                exit
            exit
        exit
    exit
exit
```

RR-1 receives the prefix 7.7.7.7/32 from neighbors PE-3 and PE-4, as follows:

```
*A:RR-1# show router bgp routes 7.7.7.7/32
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  7.7.7.7/32                                           100         None
      192.0.2.3                                            None        -
      No As-Path
*i    7.7.7.7/32                                           100         None
      192.0.2.4                                            None        -
      No As-Path
-------------------------------------------------------------------------------
Routes : 2
```

Both routes are unlabeled and BGP updates from unlabeled sessions are by default not exported to a labeled-IPv4 session, as shown in Figure 97.

Figure 97 **Updates from Unlabeled Sessions Not Propagated to Labeled Sessions (Default)**



PE-2 will not receive prefix 7.7.7.7/32, neither as unlabeled route, nor as labeled route, as follows:

```
*A:PE-2# show router bgp routes 7.7.7.7/32 ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-2#
*A:PE-2# show router bgp routes 7.7.7.7/32 label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
```

```
BGP Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-2#
```

A route policy is created on RR-1 to accept both labeled and unlabeled routes, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "import-all"
                entry 10
                    from
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
                entry 20
                    from
                        protocol bgp-label
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

This policy accepts all routes, labeled and unlabeled. For route 7.7.7.7/32 to be advertised to the labeled peer PE-2, it is sufficient to have a policy with only entry 10 that says from protocol bgp action accept. However, the preceding policy can also be used to import labeled routes to be advertised to unlabeled peers.

The following policy is applied as route-table-import policy in BGP RIB management, both for unlabeled IPv4 routes and labeled-IPv4 routes on RR-1:

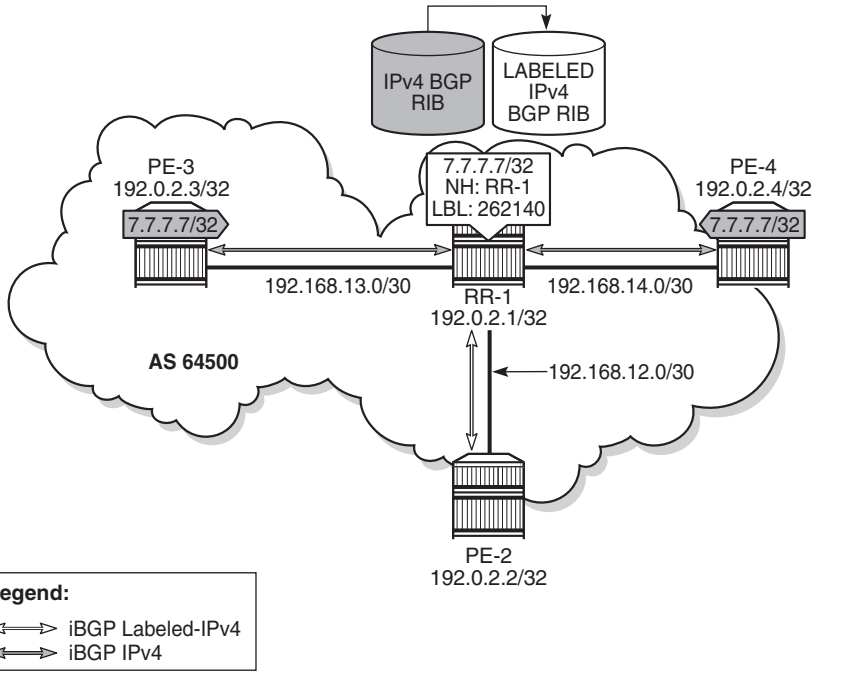
```
configure
    router
        bgp
            rib-management
                ipv4
                    route-table-import "import-all"
                exit
                label-ipv4
                    route-table-import "import-all"
                exit
            exit
        exit
```

```
    exit
exit
```

For allowing unlabeled route 7.7.7.7/32 to be advertised on a labeled session, it is sufficient to have a route-table-import for labeled-IPv4 only. However, the configuration allows for RIB leaking in both ways: from unlabeled IPv4 BGP RIB to labeled-IPv4 BGP RIB and vice versa.

Figure 98 shows this RIB leaking process.

Figure 98 **RIB Leaking from IPv4 BGP RIB to Labeled-IPv4 BGP RIB**



After applying this RIB leaking, RR-1 will advertise prefix 7.7.7.7/32 to PE-2. Therefore, RR-1 needs to add a label to the route and RR-1 needs to set next-hop-self. RR-1 advertises only one labeled route for prefix 7.7.7.7/32, with next hop 192.0.2.1, as follows:

```
*A:RR-1# show router bgp neighbor 192.0.2.2 label-ipv4 advertised-routes
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
```

```
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     7.7.7.7/32                                        100         None
      192.0.2.1                                         None        262140
      No As-Path
---snip---
```

The BGP update message is as follows:

```
9 2016/09/07 11:44:01.86 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 49
    Flag: 0x90 Type: 14 Len: 17 Multiprotocol Reachable NLRI:
        Address Family LBL-IPV4-Labeled
        NextHop len 4 NextHop 192.0.2.1
        7.7.7.7/32 Label 262140
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.1
"
```

On PE-2, the following labeled BGP route is imported:

```
*A:PE-2# show router bgp routes 7.7.7.7/32 label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  7.7.7.7/32                                        100         None
      192.0.2.1                                         None        262140
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
```

Conclusion

The BGP RIB architecture in SR OS release 14.0.R4, and later, with separate RIBs for unlabeled and labeled-unicast routes supports unlabeled sessions and labeled sessions in parallel. By default, labeled routes are not advertised to unlabeled sessions and vice versa. Route-table import policies for RIB management allow route leaking between separate RIBs: unlabeled BGP RIB and labeled-unicast BGP RIB.

MPLS

In This Section

This section provides MPLS configuration information for the following topics:

- Automatic Bandwidth Adjustment in P2P LSPs
- Automatic Creation of RSVP-TE LSPs
- Class-Based Forwarding
- DiffServ Traffic Engineering
- Entropy Label
- IGP Shortcuts
- Inter-Area TE Point-to-Point LSPs
- LDP FEC to BGP Label Route Stitching
- LDP over RSVP Using OSPF as IGP
- LDP Point-to-Point LSPs
- LDP-SR Stitching for IPv4 Prefixes (IS-IS)
- MPLS LDP FRR using ISIS as IGP
- MPLS Transport Profile
- Multicast Label Distribution Protocol
- RSVP Point-to-Point LSPs
- RSVP Signaled Point-to-Multipoint LSPs
- Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP
- Segment Routing with IS-IS Control Plane
- Shared Risk Link Groups for RSVP-Based LSP
- Static Point-to-Point LSPs
- Tunneling of ICMP Reply Packets over MPLS LSPs
- Unnumbered Interfaces in RSVP-TE and LDP

Automatic Bandwidth Adjustment in P2P LSPs

This chapter provides information about automatic bandwidth adjustment in P2P LSPs.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This note is applicable to all 7x50 chassis. At a minimum chassis mode C must be used (and for 7450 ESS, at least chassis mode D). The feature was first introduced in 8.0.R4. Initially, this chapter was written for release 13.0.R2, but the CLI in the current edition corresponds to release 14.0.R1.

Overview

Automatic bandwidth adjustment refers to the capability of an ingress label edge router (iLER) to dynamically adjust the bandwidth of a resource reservation protocol (RSVP) label switched path (LSP) tunnel based on active measurement of the traffic rate into the tunnel. The bandwidth assigned to an RSVP LSP tunnel is taken into account by the control plane, to verify that sufficient bandwidth is available for a new LSP or for an increase or decrease in bandwidth for an existing LSP. The actual bandwidth in the data plane is not capped by this setting. QoS mechanisms can be set up to filter and police the traffic in the data plane, but that is beyond the scope of this chapter.

Auto-bandwidth adjustment uses the existing LSP egress statistics feature to track the bandwidth on a specific LSP. When egress statistics are enabled, the control processing module (CPM) collects statistics from all IOMs forwarding traffic belonging to the LSP (whether the traffic is currently leaving the ingress LER via the primary path, a secondary path, or an FRR detour/bypass path). The egress statistics have counts for the number of packets and bytes forwarded per LSP on a per-forwarding class, per priority (in-profile versus out-of-profile) basis.

For the actual bandwidth adjustment, Make-Before-Break (MBB) is used. No traffic interruption is noticed. If an auto-bandwidth attempt fails, there will be 5 retries and, if they all fail, the bandwidth remains unchanged. The next attempt may occur with the next trigger.

Retries follow the retry-limit (5 in this case), retry-timer (by default 30s), and exponential back-off timer, if enabled in MPLS.

Auto-bandwidth adjustment can be triggered in four different ways:

1. Periodic trigger

   The iLER determines at the end of each adjust-interval whether to attempt an auto-bandwidth adjustment.
2. Overflow or underflow trigger

   The measured bandwidth of an LSP has increased or decreased significantly since the start of the current adjust-interval. It may be preferable to adjust the bandwidth of the LSP after a number of overflow/underflow samples, rather than wait for the adjust-interval to end (default: 24 h).
3. Manual trigger

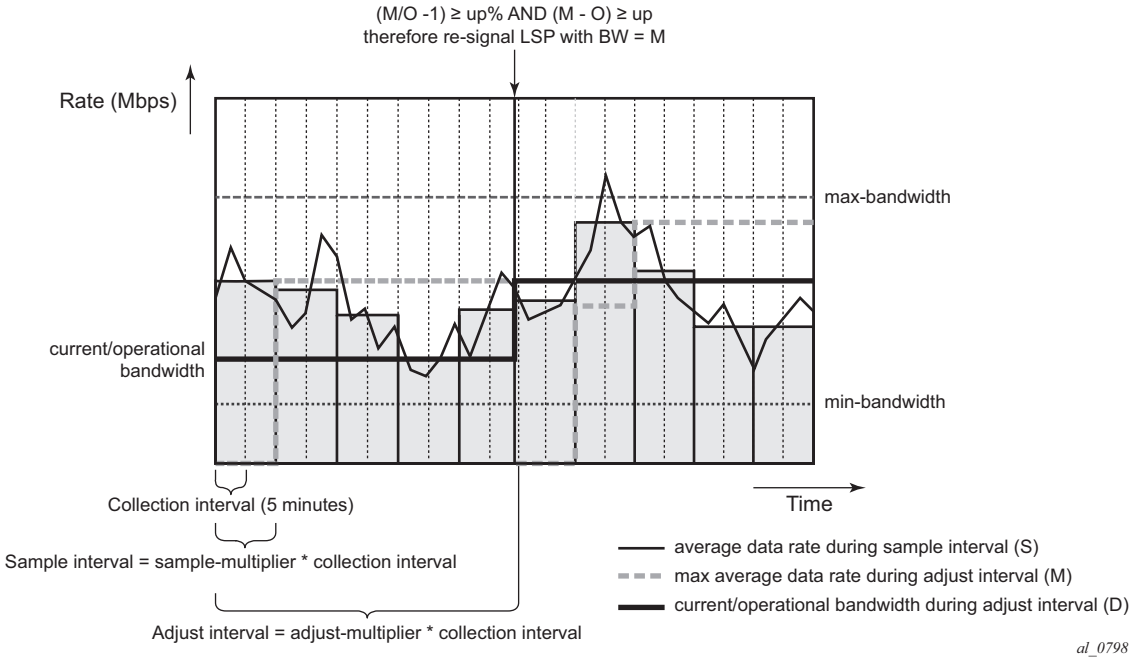
   An operator launches a tools command to trigger an auto-bandwidth adjustment.
4. Active path change

   The LSP has a primary and one or more secondary paths. When there is a change from the primary path to a secondary path without the LSP going down, an auto-bandwidth MBB is triggered. When the primary path becomes active again, another auto-bandwidth MBB is triggered.

Periodic Trigger

Figure 99 shows the different time intervals and bandwidths defined in the auto-bandwidth adjustment implementation. In this example, there will be an auto-bandwidth attempt when the adjust-interval elapses (periodic trigger). If the auto-bandwidth algorithm is met, the current bandwidth is increased. The parameters are explained after the figure.

Figure 99 **Auto-Bandwidth Adjustment Implementation**



The time intervals are:

- Collection interval in minutes. This is a global parameter to be set in an accounting policy. Range: 5 to 120 minutes. Default: 5 minutes.

  For this kind of record type, the minimum interval is 5 minutes. For policies using a record type of SAA or PM, the minimum is 1 minute.

```
*A:PE-1# configure log accounting-policy 10 collection-interval
  - collection-interval <minutes>
  - no collection-interval

 <minutes>            : [1..120]

*A:PE-1# configure log accounting-policy 10 collection-interval 1
MAJOR: LOG #1076 Except for policies using a record type of SAA or PM the minimum
interval is 5 mins
```

  - Sample interval: sample-multiplier * collection interval
  - Sample-multiplier is configurable globally in the MPLS context or per LSP. Default value: 1. In Figure 99, the sample multiplier equals 2 for a sample interval of 2 * 5 minutes = 10 minutes.

- Adjust-interval: adjust-multiplier * collection interval
  - Nokia recommends that the adjust-multiplier is an integer multiple of the sample-multiplier.

– Adjust-multiplier is configurable globally in the MPLS context or per LSP. Default value: 288 (288 * 5 minutes = 1440 minutes = 24 h). In Figure 99, the adjust multiplier equals 10 for an adjust-interval of 10 * 5 minutes = 50 minutes.

```
*A:PE-1# configure router mpls auto-bandwidth-multipliers
  - auto-bandwidth-multipliers sample-multiplier <number1> adjust-multiplier <number2>
  - no auto-bandwidth-multipliers

 <number1>            : [1..511]
 <number2>            : [1..16383]


*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth multipliers
  - multipliers sample-multiplier <num1> adjust-multiplier <num2>
  - no multipliers

 <num1>               : [1..511]
 <num2>               : [1..16383]
```

The different bandwidths are:

- Minimum bandwidth: configured minimum bandwidth in Mbps that the auto-bandwidth adjustment can signal for the LSP. Granularity: 1 Mbps. Default: 0 Mbps.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth min-bandwidth
  - min-bandwidth <mbps>
  - no min-bandwidth

 <mbps>               : [0..100000]
```

- Maximum bandwidth: configured maximum bandwidth in Mbps that the auto-bandwidth adjustment can signal for the LSP. Granularity: 1 Mbps. Default: 100 Mbps.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth max-bandwidth
  - max-bandwidth <mbps>
  - no max-bandwidth

 <mbps>               : [0..100000]
```

- Current bandwidth or operational bandwidth (O): currently reserved bandwidth in Mbps for the LSP in the control plane. This is the operational bandwidth that is maintained in the Management Information Base (MIB) and is the bandwidth that will be auto-adjusted. Granularity: 1 Mbps.
- Sampled bandwidth (S): average data rate for the last sample interval.

- Measured bandwidth (M): maximum averaged (per sample interval) data rate in the current adjust-interval. The SR OS keeps track of the maximum average data rate of each LSP since the last reset of the adjust-count.
- Signaled bandwidth: bandwidth in Mbps that is provided to the CSPF algorithm and signaled in the RSVP SENDER_TSPEC and FLOWSPEC objects, when an auto-bandwidth adjustment is attempted. Granularity: 1 Mbps.

The other auto-bandwidth parameters for periodically triggered auto-bandwidth adjustment are:

- Up% (adjust-up in percent): minimum increase in bandwidth from current to measured bandwidth, expressed as a percentage of the current bandwidth. Default: 5%.
- Up (adjust-up bw): minimum increase in bandwidth as absolute bandwidth in Mbps. Up = measuredBW – currentBW. Granularity: 1 Mbps. Default: 0 Mbps.

```
*A:PE-1>config>router>mpls>lsp>auto-bandwidth# adjust-up
  - adjust-up <percent> [bw <mbps>]
  - no adjust-up

 <percent>            : [0..100]
 <mbps>               : [0..100000]
```

- Down% (adjust-down in percent): minimum decrease in bandwidth from current to measured bandwidth, expressed as a percentage of the current bandwidth. Default: 5%.
- Down (adjust-down bw): minimum decrease in bandwidth as absolute bandwidth in Mbps. Down = currentBW – measuredBW. Granularity: 1 Mbps. Default: 0 Mbps.

```
*A:PE-1>config>router>mpls>lsp>auto-bandwidth# adjust-down
  - adjust-down <percent> [bw <mbps>]
  - no adjust-down

 <percent>            : [0..100]
 <mbps>               : [0..100000]
```

In Figure 99, the minimum and maximum bandwidths mark the bandwidth range where auto-bandwidth adjustments are allowed. The sample interval is two collection intervals long (2 * 5 minutes = 10 minutes). The adjust-interval is 10 collection intervals long (10 * 5 minutes = 50 minutes). Initially, the current bandwidth (O) equals the configured bandwidth for the primary path. It is good practice to give that same value to the minimum bandwidth for auto-bandwidth. The system doesn’t confirm this and these bandwidths are independent from each other.

In this example, the sampled bandwidth exceeds the current bit rate in most of the sample intervals. The maximum sampled bandwidth in the current adjust-interval corresponds to the measured bandwidth (M). When auto-bandwidth adjustment is triggered at the end of the adjust-interval, this measured bandwidth will be signaled and, after a successful adjustment, will be the new current bandwidth. After the auto-bandwidth adjustment, a new adjust-interval starts and the measured bandwidth is reset to 0. As long as the first sample interval of the new adjust-interval is not finished, the measured bandwidth equals 0 and auto-adjustment would be impossible even when triggered manually.

The auto-bandwidth attempt follows these rules:

- When measuredBW ≥ currentBW
  - if {(measuredBW / currentBW – 1) ≥ up%} &&{(measuredBW – currentBW) ≥ up
    then signaledBW = max{(min(measuredBW, maxBW)), minBW}
- When measuredBW ≤ currentBW
  - if {(1 – measuredBW/currentBW) ≥ down%} && {(currentBW – measuredBW) ≥ down}
    then signaledBW = min{(max(measuredBW, minBW)), maxBW}

CLI configured bandwidths have a granularity of 1 Mbps, while the threshold calculations with measured bandwidth are performed at full precision. This means that the signaled bandwidth in the RSVP message is rounded up to the nearest integer multiple of 1 Mbps.

Overflow/Underflow Trigger

Auto-bandwidth adjustment can also be triggered by overflow or underflow. When the bandwidth changes drastically, the bandwidth can be auto-adjusted after a number of consecutive overflow/underflow samples. In this case, there is no need to wait for the adjust-interval to end (by default: 24 h).

The parameters used in case of overflow are:

- Overflow sample: a sample interval counts as an overflow sample if the sampled bandwidth is higher than the current bandwidth by at least the configured overflow thresholds.
- Overflow-limit/overflow-count: an auto-bandwidth adjustment occurs after this number of consecutive overflow samples.
- Threshold%: minimum difference between sampled bandwidth and current bandwidth, expressed as a percentage of the current bandwidth.

- Threshold bw: minimum difference between sampled bandwidth and current bandwidth in Mbps. Default value: 0.

```
A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth overflow-limit
  - overflow-limit <number> threshold <percent> [bw <mbps>]
  - no overflow-limit

 <number>             : [1..10]
 <percent>            : [0..100]
 <mbps>               : [0..100000]
```

The rules for overflow-triggered auto-bandwidth adjustment are as follows:

- Overflow sample: {(sampledBW / currentBW – 1) ≥ threshold%} && {(sampledBW – currentBW) ≥ thresholdBW}
- The signaled bandwidth will be:
  - if (measuredBW ≥ maxBW) then signaledBW = maxBW
  - if (measuredBW ≤ minBW) then signaledBW = minBW
  - else signaledBW = measuredBW

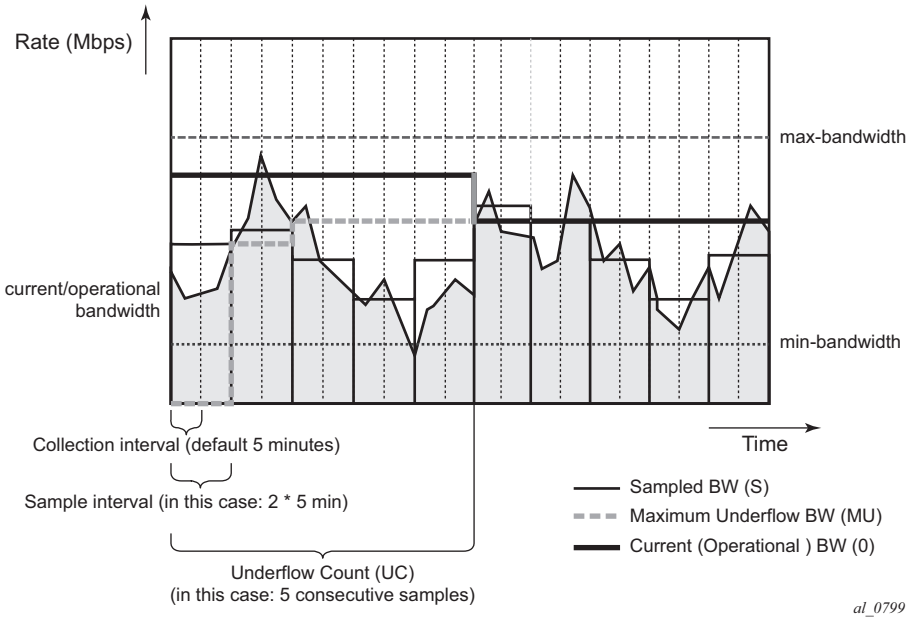
Underflow triggers were introduced in 12.0.R1. The parameters used in case of underflow are:

- Underflow sample: a sample interval counts as an underflow sample if the sampled bandwidth is lower than the current bandwidth by at least the configured underflow thresholds.
- Underflow-limit/underflow-count: an auto-bandwidth adjustment occurs after this number of consecutive underflow samples.
- Threshold%: minimum difference between current bandwidth and sampled bandwidth, expressed as a percentage of the current bandwidth.
- Threshold bw: minimum difference between current bandwidth and sampled bandwidth in Mbps. Default value: 0.
- Maximum underflow bandwidth (MU): maximum sampled bandwidth in the consecutive underflow samples.

```
A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth underflow-limit
  - underflow-limit <number> threshold <percent> [bw <mbps>]
  - no underflow-limit

 <number>             : [1..10]
 <percent>            : [0..100]
 <mbps>               : [0..100000]
```

Figure 100 **Underflow-Triggered Auto-Bandwidth Implementation**



In Figure 100, the adjust-interval is not displayed. It is assumed to be the default of 288 collection intervals (24 h). The figure only shows five consecutive underflow samples. The underflow-limit equals 5. In each of the samples, the sample bandwidth is below the underflow threshold. The maximum sampled bandwidth of these five samples corresponds to the maximum underflow bandwidth. This bandwidth will be signaled when auto-bandwidth adjustment is triggered because the underflow count is reached.

The rules for underflow-triggered auto-bandwidth adjustment are as follows:

- Underflow sample:
  - {(1 - sampledBW / currentBW) ≥ threshold%} && {(currentBW – sampledBW) ≥ thresholdBW}
- Underflow count/underflow limit: after that many consecutive underflow samples, an auto-bandwidth adjustment is triggered.
- The signaled bandwidth will be:
  - if (maxUnderflowBW ≥ maxBW) then signaledBW = maxBW
  - if (maxUnderflowBW ≤ minBW) then signaledBW = minBW
    else signaledBW = maxUnderflowBW

If the adjustment is successful, the sample counter within the adjust-interval is reset, along with other parameters, such as the maximum underflow bandwidth, the measured bandwidth, and the underflow count. The next adjust-interval will elapse in 24 h.

If the adjustment fails, there will be 5 retries. If they all fail, only the underflow count and the maximum underflow bandwidth are reset. The current adjust-interval continues.

Manual Trigger

Besides the periodic trigger and the overflow/underflow trigger, an operator can launch a tools command to trigger an auto-bandwidth adjustment.

```
A:PE-1# tools perform router mpls adjust-autobandwidth
```

This tools command can be launched with or without explicit LSP name. In the latter case, all active LSPs are attempted for auto-bandwidth.

```
A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
  - adjust-autobandwidth [lsp <lsp-name> [force [bandwidth <mbps>]]]

 <lsp-name>           : [64 chars max]
 <force>              : keyword
 <mbps>               : [0..100000]

A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
```

This command (without the keyword force) triggers a new auto-bandwidth calculation according to the rules of periodic triggered type. If the LSP already has the correct reserved bandwidth, the following message is returned.

```
A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
MINOR: CLI lsp LSP-PE-1-PE-2 active path is already at the requested value 12 Mbps.
```

If the keyword force is added without a specific value for the bandwidth, there is no threshold checking. The bandwidth can also be adjusted if the difference in bandwidth is below the thresholds. The granularity remains 1 Mbps.

```
A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
```

The rules for the signaled bandwidth are unchanged:

- if (measuredBW ≥ maxBW) then signaledBW = maxBW
- if (measurealedBdBW ≤ minBW) then signaledBW = minBW
  else signW = measuredBW

If the keyword force with bandwidth (in Mbps) option is given, the signaled bandwidth is set to this configured bandwidth, even if it is a value below the minimum or higher than the maximum bandwidth.

```
A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
bandwidth 30
```

After a manually triggered auto-bandwidth MBB, no counters are reset. The ongoing adjust-interval is not aborted.

A clear command resets all counters and timers associated with auto-bandwidth adjustment on a specified LSP.

```
A:PE-1# clear router mpls lsp-autobandwidth "LSP-PE-1-PE-2"
```

This command clears the parameters that are shown in bold in the following example. The parameters will be explained in the configuration section.

```
A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 2 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                                AB Samp Cnt    : 0
AB Last Adj : 04/28/2015 08:10:19              AB Next Adj    : 1440 Mins
ABMaxAvgRt  : 0 Mbps                           AB Lst AvgRt   : 0 Mbps
AB Ovfl Lmt : 1                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: underflow                        AB Monitor BW  : False
Be Weight   : 50 percent                       Af Weight      : 80 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
A:PE-1#
```

Passive Monitoring

The system offers the option to measure the bandwidth of an LSP without taking any action to adjust the bandwidth reservation.

```
A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth monitor-bandwidth
```

Auto-Bandwidth Based on Forwarding Class

From 11.0.R4 onward, the bandwidth can be calculated as a weighted sum of all the traffic in the eight forwarding classes. By default, all forwarding classes have the same weight: 100%, but that sampling weight is configurable.

```
A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc
  - fc <fc-name> sampling-weight <sampling-weight>
  - no fc <fc-name>

 <fc-name>            : be|l2|af|l1|h2|ef|h1|nc
 <sampling-weight>    : [0..100]
```

Active Path Change

From 12.0.R4 onward, auto-bandwidth adjustment is also supported on LSPs that have secondary paths. If the secondary path is standby, an auto-bandwidth MBB can be triggered when the active path changes from primary to secondary. The secondary/standby path is only initialized at its configured bandwidth when it is established, and the bandwidth is adjusted only when it becomes active. This happens when the primary path goes down or becomes degraded. When another path becomes active, the bandwidth used to signal the auto-bandwidth MBB will be the operational bandwidth of the previous path.

The definition for current bandwidth is modified for this feature:

- Current bandwidth: last known reserved bandwidth for the LSP. This may be for a different path than the active one.

  Auto-bandwidth adjustment will only take place on the active path. When the active path changes, the current bandwidth is updated to the operational bandwidth of the new active path.

- For a secondary path that is signaled as standby, if the active path for an LSP changes without the LSP going down, an auto-bandwidth MBB is triggered on the new active path. The signaled bandwidth is the operational bandwidth of the previous path. The reserved bandwidth of the new active path will be its configured bandwidth until the MBB succeeds.
- For a secondary path where the active path goes down, the LSP will go down temporarily until the secondary path is set up. When the LSP goes down, all statistics and counters are cleared, so the previous path operational bandwidth is lost. There will be no immediate bandwidth adjustment on the secondary path.

The following rules apply to determine the signaled bandwidth of the new active path.

- For a path that is operationally down, signaledBW = configuredBW.
- For the first 5 MBB attempts on the path that just became active,
  signaledBW = currentBW (operational bandwidth of the previous path).

For the remaining MBB attempts, signaledBW = operationalBW.

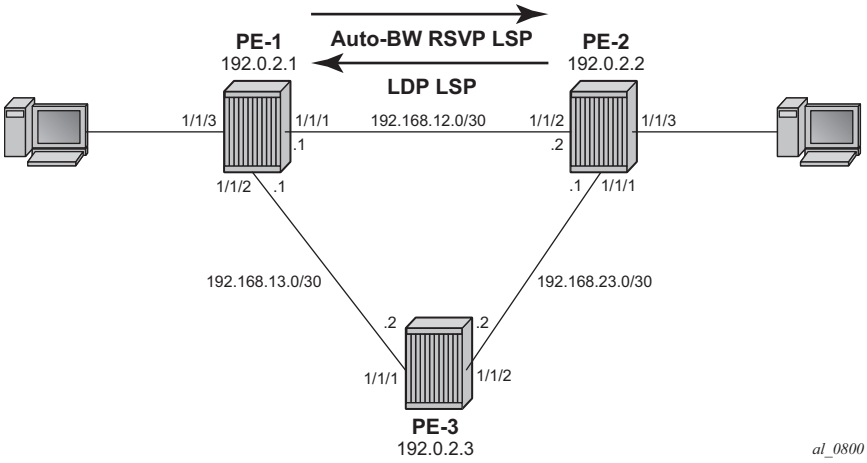
- For all MBBs other than auto-bandwidth MBB on the active path,
  MBB signaledBW = operationalBW.
- For an MBB on the inactive (standby) path, MBB signaledBW = configuredBW.

When the system reverts from a secondary standby path to the primary path, a Delayed Retry MBB is attempted to bring the bandwidth on the standby path back to the configured bandwidth. MBB is attempted once, and if it fails, the standby is torn down. A Delayed Retry MBB has the highest priority among MBBs, so it will take precedence over any other MBB in progress on the standby path, such as configuration change or pre-emption.

Configuration

Figure 101 shows the example setup. The focus will be on the RSVP LSP from PE-1 to PE-2.

Figure 101 **Example Setup for Auto-Bandwidth Point-to-Point LSPs**



Base Configuration

The cards, MDAs and ports need to be configured.

Configure the interfaces on all nodes. For PE-1:

```
configure router
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/1
        exit
        interface "int-PE-1-PE-3"
            address 192.168.13.1/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.1/32
        exit
```

As an IGP, OSPF or IS-IS can be used. In this example, OSPF is configured. Traffic engineering should be enabled. For PE-1:

```
configure router
        ospf
            traffic-engineering
            area 0.0.0.0
                interface "system"
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-1-PE-3"
```

```
                    interface-type point-to-point
                exit
            exit
        exit
```

Optionally, enable LDP on all interfaces. Link-layer LDP is not a pre-requisite for using auto-bandwidth RSVP LSPs. In this example, the SDP from PE-2 to PE-1 uses an LDP LSP, but it could have been an RSVP-TE LSP instead. For PE-2:

```
configure router ldp
            interface-parameters
                interface "int-PE-2-PE-1"
                exit
                interface "int-PE-2-PE-3"
                exit
            exit
            targeted-session
            exit
```

Enable MPLS and RSVP on all nodes.

```
configure router mpls no shutdown
configure router rsvp no shutdown
```

Add all interfaces to the MPLS context. They will automatically be added to the RSVP context. For PE-1:

```
configure router
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-3"
            exit
        exit
```

Configure a path with no explicitly defined hops and LSP LSP-PE-1-PE-2 on PE-1:

```
configure router mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "loose"
                exit
                no shutdown
            exit
```

In the example, traffic needs to be injected into the LSP tunnel. For that, a VPLS service is created. For PE-1, an SDP using the RSVP LSP to PE-2 is created.

```
configure service
        sdp 212 mpls create
            description "SDP-PE-1-PE-2-overRSVP-TE"
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
```

On PE-2, an SDP using LDP is created toward PE-1.

```
configure service
        sdp 121 mpls create
            description "SDP-PE-2-PE-1-overLDP"
            far-end 192.0.2.1
            ldp
            no shutdown
        exit
```

On PE-1 and PE-2, a VPLS is created. For PE-1:

```
configure service vpls 100 customer 1 create
            sap 1/1/3 create
            exit
            spoke-sdp 212:100 create
            exit
            no shutdown
            exit
```

The configuration on PE-2 is similar.

Pre-requisites for Auto-Bandwidth LSP Configuration

Enable constrained shortest path first (CSPF) on the LSP by adding the keyword **cspf**.

```
configure router mpls lsp “LSP-PE-1-PE-2” cspf
```

The bandwidth of the LSP will be adjusted in a make-before-break (MBB) manner. Enable MBB on the LSP by adding the keyword adaptive to the primary path.

```
configure router mpls lsp “LSP-PE-1-PE-2” primary “loose” adaptive
```

Enter a value for the bandwidth in Mbps for the primary path. It is good practice to configure the same value as for the minimum bandwidth in the auto-bandwidth settings.

```
configure router mpls lsp "LSP-PE-1-PE-2" primary "loose" bandwidth 2
```

Auto-Bandwidth LSP Configuration

MPLS auto-bandwidth adjustment allows the ingress LER to dynamically adjust the bandwidth of an RSVP tunnel based on active measurements of the traffic rate into the tunnel. Therefore, LSP egress statistics need to be enabled on the iLER.

Auto-bandwidth adjustment requires an accounting policy to be defined and operational. The accounting policy specifies the collection interval for LSP statistics collection, which is fundamental to the auto-bandwidth algorithm. The minimum interval for this type of collection is 5 minutes, which is the default value.

```
configure log
        accounting-policy 10
            record combined-mpls-lsp-egress
            to no-file
            no shutdown
       exit
```

An accounting policy of record type **combined-mpls-lsp-egress** doesn't need a reference to a specific file ID.

From the moment auto-bandwidth is enabled with an LSP context, the record combined-mpls-lsp-egress inside the accounting policy will also take bandwidth measurements.

```
configure log accounting-policy 10 to no-file
```

When the no-file is configured, no LSP statistics are stored anymore. The MPLS auto-bandwidth feature retrieves it LSP statistics information directly from the statistics module.

However, the accounting policy can reference a file and, therefore, a CF card. An additional CF card may be required in each node as a storage location. For releases prior to 10.0.R4, this is mandatory.

```
configure log
        file-id 66
            location cf1:
            rollover 15 retention 1
```

```
        exit
        accounting-policy 66
            record combined-mpls-lsp-egress
            to file 66
            no shutdown
        exit
```

In the remainder of the example, the accounting policy will reference to no-file.

After the accounting policy has been created, egress statistics can be enabled on the LSP.

```
configure router mpls lsp "LSP-PE-1-PE-2"
         egress-statistics
            no shutdown
            collect-stats
            accounting-policy 10
         exit
```

The system does not verify whether egress statistics have been enabled on the LSP. When a user configures auto-bandwidth adjustment, but without enabling egress statistics, no auto-bandwidth measurements and adjustments are performed. The operational state of auto-bandwidth (AB OpState) is down.

Enable auto-bandwidth with default settings by adding the keyword auto-bandwidth to the LSP.

```
configure router mpls lsp LSP-PE-1-PE-2 auto-bandwidth
```

The actual values are shown in the following output. They are explained after the output.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 0 Mbps                           Auto BW Max    : 100000 Mbps
AB Up Thresh: 5 percent                        AB Down Thresh : 5 percent
AB Up BW    : 0 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 2 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288+                             AB Samp Mul    : 1+
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                                AB Samp Cnt    : 0
AB Last Adj : n/a                              AB Next Adj    : 1440 Mins
ABMaxAvgRt  : 0 Mbps                           AB Lst AvgRt   : 0 Mbps
AB Ovfl Lmt : 0                                AB Ovfl Cnt    : 0
```

```
ABOvflThres : 0 percent                          AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                                  AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent                          AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: none                               AB Monitor BW  : False
Be Weight   : 100 percent                        Af Weight      : 100 percent
L1 Weight   : 100 percent                        L2 Weight      : 100 percent
Nc Weight   : 100 percent                        Ef Weight      : 100 percent
H1 Weight   : 100 percent                        H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

The plus sign (+) indicates that the value is inherited from the global MPLS settings (AB Adj Mul: 288+ and AB Samp Mul: 1+). The sample-multiplier and the adjust-multiplier can both be configured globally in the MPLS context or overruled by the settings per LSP. In this example, nothing has been configured in the MPLS context or in the LSP. Therefore, the default values as defined in the MPLS context are applicable.

Auto-Bandwidth – Periodic Trigger (Normal)

The default collection interval is 5 minutes. The sample-multiplier is 1, by default. The sample interval equals 1 * 5 minutes = 5 minutes. The adjust-multiplier is 288, by default 288. The adjustment interval equals 288 * 5 minutes = 1440 minutes (24 hours).

The auto-bandwidth settings for the LSP are modified as follows:

```
configure router mpls lsp LSP-PE-1-PE-2
                auto-bandwidth
                    multipliers sample-multiplier 1 adjust-multiplier 3
                    adjust-up 10 bw 1
                    adjust-down 5 bw 0  ## default
                    max-bandwidth 20
                    min-bandwidth 2
                exit
```

In the example, the bandwidth of the LSP can be auto-adjusted every 15 minutes (after 3 intervals of 5 minutes). For a decrease in bandwidth, the default settings apply and no explicit command is required in this example. That means that the current bandwidth will be reduced when the difference in bandwidth is at least 5%. There is no absolute decrease (in Mbps) defined. For an increase in bandwidth, there

will only be an adjustment when the increase is at least 10% and at least 1 Mbps. The minimum bandwidth is 2 Mbps. This equals the bandwidth set in the path in the LSP (recommended). The maximum bandwidth equals 20 Mbps. The system will not compare the minimum or maximum bandwidth to the configured bandwidth for the path.

Display the actual auto-bandwidth data after 5, 10, and 15, minutes.

There are different bandwidths displayed:

- The AB Curr BW is the operational bandwidth during the adjustment interval. It is initially the configured bandwidth of the path in the LSP, but it can be auto-adjusted. This bandwidth is taken into account in the control plane when an LSP is set up or modified in case of MBB. The real data rate in the data plane may exceed this operational bandwidth.
- The ABMaxAvgRt is the measured bandwidth, meaning the maximum averaged bandwidth (calculated every sample interval of 5 minutes) in the adjustment interval of 15 minutes (AB Adj Time: 15 Min).
- The AB Lst AvgRt is the sampled bandwidth, averaged over the latest sample interval of 5 minutes (AB Samp Intv: 5 Mins).

After 5 minutes, one collection interval has elapsed within the adjust-interval (AB Adj Cnt = 1) and the next adjustment time is in 10 minutes (AB Next Adj = 10 Min). The current bandwidth equals 2 Mbps, while the measured and the sampled bandwidths are much higher: 12 Mbps.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 2 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 3                                AB Samp Mul    : 1
AB Adj Time : 15 Mins                          AB Samp Time   : 5 Mins
AB Adj Cnt  : 1                                AB Samp Cnt    : 0
AB Last Adj : n/a                              AB Next Adj    : 10 Mins
ABMaxAvgRt  : 12 Mbps                          AB Lst AvgRt   : 12 Mbps
AB Ovfl Lmt : 0                                AB Ovfl Cnt    : 0
ABOvflThres : 0 percent                        AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                                AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent                        AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: none                             AB Monitor BW  : False
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
```

```
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

After 10 minutes, another collection interval has elapsed in the adjust-interval (**AB Adj Cnt** = 2) and the next adjustment time is in 5 minutes (**AB Next Adj** = 5 Min).

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 2 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 3                                AB Samp Mul    : 1
AB Adj Time : 15 Mins                          AB Samp Time   : 5 Mins
AB Adj Cnt  : 2                                AB Samp Cnt    : 0
AB Last Adj : n/a                              AB Next Adj    : 5 Mins
ABMaxAvgRt  : 12 Mbps                          AB Lst AvgRt   : 12 Mbps
AB Ovfl Lmt : 0                                AB Ovfl Cnt    : 0
ABOvflThres : 0 percent                        AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                                AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent                        AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: none                             AB Monitor BW  : False
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

After 15 minutes, auto-bandwidth adjustment occurs. **AB Adj Cause** is normal for periodically triggered adjustments. The next adjustment interval will elapse in 15 minutes. The measured bandwidth **ABMaxAvgRt** is reset to 0 after a successful adjustment.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
```

```
AB Up Thresh: 10 percent                      AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                          AB Down BW     : 0 Mbps
AB Curr BW  : 13 Mbps                         AB Samp Intv   : 5 Mins
AB Adj Mul  : 3                               AB Samp Mul    : 1
AB Adj Time : 15 Mins                         AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                               AB Samp Cnt    : 0
AB Last Adj : 04/24/2015 12:57:25             AB Next Adj    : 15 Mins
ABMaxAvgRt  : 0 Mbps                          AB Lst AvgRt   : 13 Mbps
AB Ovfl Lmt : 0                               AB Ovfl Cnt    : 0
ABOvflThres : 0 percent                       AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                               AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent                       AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: normal                          AB Monitor BW  : False
Be Weight   : 100 percent                     Af Weight      : 100 percent
L1 Weight   : 100 percent                     L2 Weight      : 100 percent
Nc Weight   : 100 percent                     Ef Weight      : 100 percent
H1 Weight   : 100 percent                     H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

The periodic trigger type rules for auto-bandwidth are:

- When measuredBW ≥ currentBW
  - if {(measuredBW / currentBW – 1) ≥ up%} &&{(measuredBW – currentBW) ≥ up

    then signaledBW = max{(min(measuredBW, maxBW)), minBW}

In this case, the measuredBW (13 Mbps) is greater than the currentBW (2 Mbps). The increase is at least 10% (up%) and at least 1 Mbps (up). The bandwidth will be adjusted. The new bandwidth that will be signaled is calculated as follows:

```
signaledBW= max{(min(measuredBW, maxBW)), minBW}
signaledBW= max {(min (13 Mbps, 20 Mbps)), 2 Mbps}
signaledBW= max {13 Mbps, 2 Mbps}
signaledBW= 13 Mbps
```

Whenever an auto-bandwidth adjustment is performed, a message is stored in log 99.

```
*A:PE-1# show log log-id 99 application "mpls"
===============================================================================
Event Log 99
===============================================================================
Description : Default System Log
Memory Log contents  [size=500   next event=62  (not wrapped)]

61 2016/03/23 12:39:43.18 UTC WARNING: MPLS #2014 Base VR 1:
"LSP path LSP-PE-1-PE-2::loose resignaled as result of autoBandwidth MBB"
```

When the maximum bandwidth is modified to a value that is lower than the current bandwidth, an adjustment occurs at the end of the adjustment interval.

```
configure router mpls lsp LSP-PE-1-PE-2
               auto-bandwidth
                   max-bandwidth 10
               exit
```

The current bandwidth will be reduced to 10 Mbps (for a measured bandwidth of 11 Mbps).

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 10 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 10 Mbps                          AB Samp Intv   : 5 Mins
AB Adj Mul  : 3                                AB Samp Mul    : 1
AB Adj Time : 15 Mins                          AB Samp Time   : 5 Mins
AB Adj Cnt  : 1                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 12:54:43              AB Next Adj    : 10 Mins
ABMaxAvgRt  : 11 Mbps                          AB Lst AvgRt   : 11 Mbps
AB Ovfl Lmt : 0                                AB Ovfl Cnt    : 0
ABOvflThres : 0 percent                        AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                                AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent                        AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: normal                           AB Monitor BW  : False
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================*
```

Auto-Bandwidth - Passive Monitoring

When passive monitoring is enabled, no automatic bandwidth adjustments occurs. When the maximum bandwidth is again raised to 20 Mbps, the bandwidth will not be auto-adjusted even if the measured bandwidth is high enough.

```
configure router mpls lsp LSP-PE-1-PE-2 auto-bandwidth max-bandwidth 20
configure router mpls lsp LSP-PE-1-PE-2 auto-bandwidth monitor-bandwidth
```

The system monitors the bandwidth, but without taking action at the end of the adjust-interval.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 10 Mbps                          AB Samp Intv   : 5 Mins
AB Adj Mul  : 3                                AB Samp Mul    : 1
AB Adj Time : 15 Mins                          AB Samp Time   : 5 Mins
AB Adj Cnt  : 2                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 12:54:43              AB Next Adj    : 5 Mins
ABMaxAvgRt  : 11 Mbps                          AB Lst AvgRt   : 11 Mbps
AB Ovfl Lmt : 0                                AB Ovfl Cnt    : 0
ABOvflThres : 0 percent                        AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                                AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent                        AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: normal                           AB Monitor BW  : True
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

The value for the parameter **AB Monitor BWis True**

For the remainder of the chapter, there is no passive monitoring. The settings are restored to normal:

```
configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth no monitor-bandwidth
```

Auto-Bandwidth – Overflow and Underflow Trigger Type

With default settings, the adjustment interval is 24 hours. If the bandwidth changes significantly since the start of the current adjust-interval, overflow and underflow triggers can be used. This will speed up the auto-bandwidth adjustment. Overflow triggers are supported from 8.0.R4 onward and underflow triggers from 12.0.R1 onward.

Stop auto-bandwidth in order to force a MBB attempt toward the configured primary path bandwidth (2Mbps in this example).

```
configure router mpls lsp "LSP-PE-1-PE-2" no auto-bandwidth
```

Check the operational bandwidth of the LSP.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" detail
===============================================================================
MPLS LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
LSP Type    : RegularLsp                        LSP Tunnel ID  : 1
LSP Index   : 1                                 TTM Tunnel Id  : 1
From        : 192.0.2.1                         To             : 192.0.2.2
Adm State   : Up                                Oper State     : Up
LSP Up Time : 0d 01:04:24                       LSP Down Time  : 0d 00:00:00
Transitions : 1                                 Path Changes   : 1
Retry Limit : 0                                 Retry Timer    : 30 sec
Signaling   : RSVP                              Resv. Style    : SE
Hop Limit   : 255                               Negotiated MTU : 1564
Adaptive    : Enabled                           ClassType      : 0
FastReroute : Disabled                          Oper FR        : Disabled
Egress Stats: Enabled                           Egress Oper St*: N/A
CSPF        : Enabled                           ADSPEC         : Disabled
Metric      : N/A                               Use TE metric  : Disabled
Load Bal Wt : N/A                               ClassForwarding: Disabled
Include Grps:                                   Exclude Grps   :
None                                            None
Least Fill  : Disabled
BFD Template: None                              BFD Ping Intvl : 60
BFD Enable  : False

Revert Timer: Disabled                          Next Revert In : N/A
EntropyLbl  : Inherited
Auto BW     : Disabled
LdpOverRsvp : Enabled
VprnAutoBind: Enabled
IGP Shortcut: Enabled                           BGP Shortcut   : Enabled
IGP LFA     : Disabled                          IGP Rel Metric : Disabled
BGPTransTun : Enabled
Oper Metric : 10
Prop Adm Grp: Disabled

Primary(a)  : loose                             Up Time        : 0d 01:04:24
Bandwidth   : 2 Mbps
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

Enable auto-bandwidth with similar settings as before and add overflow and underflow triggers. The multipliers are default. Therefore, a periodically triggered auto-adjustment will only take place once every 24 hours.

```
configure router mpls lsp LSP-PE-1-PE-2 auto-bandwidth
                    multipliers sample-multiplier 1 adjust-multiplier 288
                    adjust-up 10 bw 1
                    max-bandwidth 20
                    min-bandwidth 2
                    overflow-limit 1 threshold 10 bw 2
                    underflow-limit 3 threshold 10 bw 2
                exit
```

The overflow count indicates the number of consecutive times that the overflow condition is detected at the end of a sample interval. Auto-bandwidth adjustment occurs after that number of overflow samples is reached, in this case, after the first overflow sample (overflow-limit = 1). The conditions for an overflow sample are:

{(sampledBW / currentBW – 1) ≥ threshold%} && {(sampledBW – currentBW) ≥ thresholdBW}

{(11 Mbps/2Mbps – 1) ≥ 0,1} && {(11Mbps – 2Mbps) ≥ 2Mbps}

The signaled bandwidth will be:

- if (measuredBW ≥ maxBW) then signaledBW = maxBW
- if (measuredBW ≤ minBW) then signaledBW = minBW
  else signaledBW = measuredBW
- if (11 Mbps ≥ 20 Mbps) then signaledBW = 20 Mbps;
- if (11 Mbps ≤ 2 Mbps) then signaledBW = 2 Mbps;
  else signaledBW = 11 Mbps

Display the auto-bandwidth data. The **AB Adj Cause** is now **overflow**. The overflow limit is the configured value of 1 (**AB Ovfl Lmt**). The overflow count has been reset after the auto-bandwidth was adjusted (**AB Ovfl Cnt = 0**), along with the **ABMaxAvgRt** and **the AB Adj Cnt**. This is the start of a new adjust-interval of 24 hours.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
```

```
AB Up BW    : 1 Mbps                          AB Down BW     : 0 Mbps
AB Curr BW  : 11 Mbps                         AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                             AB Samp Mul    : 1
AB Adj Time : 1440 Mins                       AB Samp Time   : 5 Mins
AB Adj Cnt  : 5                               AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 13:29:42             AB Next Adj    : 1415 Mins
ABMaxAvgRt  : 11 Mbps                         AB Lst AvgRt   : 11 Mbps
AB Ovfl Lmt : 1                               AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                      AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                               AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                      AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: overflow                        AB Monitor BW  : False
Be Weight   : 100 percent                     Af Weight      : 100 percent
L1 Weight   : 100 percent                     L2 Weight      : 100 percent
Nc Weight   : 100 percent                     Ef Weight      : 100 percent
H1 Weight   : 100 percent                     H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

In the following example, the current bandwidth is 13 Mbps, but the bandwidth dropped from 10 Mbps and the conditions for underflow are met. At the end of a sample interval, the sampled bandwidth is reduced by at least 10% and at least 2 Mbps, and this becomes an underflow sample. The conditions for an underflow sample are:

{(1 - sampledBW / currentBW) ≥ threshold%} && {(currentBW – sampledBW) ≥ thresholdBW}

{(1 - 10 Mbps / 13 Mbps) ≥ 0,1} && {(13Mbps – 10Mbps) ≥ 2Mbps}

Since the underflow limit equals 3, an auto-bandwidth adjustment can only take place after the third consecutive underflow sample. The new bandwidth will equal the **maximum sampled underflow bandwidth (ABMaxUndflBW)**. This is the maximum sampled bandwidth in the three consecutive underflow samples.

The signaled bandwidth will be:

- if (maxUnderflowBW ≥ maxBW) then signaledBW = maxBW
- if (maxUnderflowBW≤ minBW) then signaledBW = minBW
  else signaledBW = maxUnderflowBW
- if (10 Mbps≥ 20 Mbps) then signaledBW = 20 Mbps;
- if (10 Mbps ≤ 2 Mbps) then signaledBW = 2 Mbps;
  else signaledBW = 10 Mbps

The following output shows the auto-bandwidth data after two consecutive underflow samples (**AB Underfl Cnt: 2**). The maximum sampled underflow bandwidth equals 10 Mbps. No bandwidth adaptation can take place until there are three consecutive underflow samples (**AB UndflLmt: 3**).

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 13 Mbps                          AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 3                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 13:29:25              AB Next Adj    : 1425 Mins
ABMaxAvgRt  : 12 Mbps                          AB Lst AvgRt   : 10 Mbps
AB Ovfl Lmt : 1                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 2
ABUndflThrs : 10 percent                       AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 10 Mbps
AB Adj Cause: overflow                         AB Monitor BW  : False
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

After a successful auto-bandwidth adjustment, the **ABMaxUndflBW** is reset, along with the AB Adj **Cnt**, AB **Underfl Cnt** and **ABMaxAvgRt**.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 10 Mbps                          AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 14:24:25              AB Next Adj    : 1440 Mins
ABMaxAvgRt  : 0 Mbps                           AB Lst AvgRt   : 7 Mbps
AB Ovfl Lmt : 1                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: underflow                        AB Monitor BW  : False
```

```
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

If the overload or underload trigger condition is met at the end of an adjust-interval, the auto-bandwidth adjustment is normal, based on the periodic trigger. Overflow and underflow auto-bandwidth adjustments only take place when the adjust-interval is not yet completed.

Auto-Bandwidth – Manual Trigger Type

The auto-bandwidth adjustment can be triggered manually at all times by the following command (with or without the keyword force).

```
tools perform router mpls adjust-autobandwidth
tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
```

When no specific LSP is referred to, auto-bandwidth will be attempted on all LSPs. If the LSP already has the requested bandwidth, the following output is returned.

```
*A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
MINOR: CLI No Thresholds crossed for lsp LSP-PE-1-PE-2.
```

By adding the keyword **force**, there is no check whether the thresholds are crossed. However, the granularity is 1 Mbps. In this case, it is not possible to signal a bandwidth that is at least 1 Mbps different, so the following error message is returned.

```
*A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
MINOR: CLI lsp LSP-PE-1-PE-2 active path is already at the requested value 13 Mbps.
```

If the first sample interval has not yet expired, the following error message is returned.

```
*A:PE-1# tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
MINOR: CLI No Autobandwidth Averages computed for lsp LSP-PE-1-PE-2.
```

If the tools command is launched after the first sample interval has expired (**ABMaxAvgRt** is filled in), the overflow trigger rules can be applied. In this example, the traffic is reduced and the thresholds are crossed.

The **AB Adj Cause** is manual.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 4 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 1                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 14:45:36              AB Next Adj    : 1435 Mins
ABMaxAvgRt  : 4 Mbps                           AB Lst AvgRt   : 4 Mbps
AB Ovfl Lmt : 1                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: manual                           AB Monitor BW  : False
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

The counters are not reset after a manually triggered auto-bandwidth adjustment. The adjust-interval is not interrupted, the measured bandwidth and the maximum underflow bandwidth are not reset, and the overflow and underflow count are not reset.

Launch the tools command with the keyword **force** and a bandwidth value. This will set the current bandwidth to this value, even if the value is not within the allowed range between the minimum and maximum bandwidth.

```
tools perform router mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force bandwidth 30
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
```

```
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 30 Mbps                          AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 1                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 14:50:42              AB Next Adj    : 1435 Mins
ABMaxAvgRt  : 4 Mbps                           AB Lst AvgRt   : 4 Mbps
AB Ovfl Lmt : 1                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: manual                           AB Monitor BW  : False
Be Weight   : 100 percent                      Af Weight      : 100 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

Manually triggered auto-bandwidth adjustments are also performed using MBB procedures.

Auto-Bandwidth Adjustment Based on Forwarding Class Subset

With the configuration applied so far, there is no distinction between traffic from different forwarding classes (FCs). The average data rate is the sum of the traffic from all eight FCs. From 11.0.R4 onward, it is possible to provide a sampling weight for each Forwarding Class (FC) for each auto-bandwidth LSP. The average data rate is now the weighted sum of the traffic from all FCs.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc
  - fc <fc-name> sampling-weight <sampling-weight>
  - no fc <fc-name>

 <fc-name>            : be|l2|af|l1|h2|ef|h1|nc
 <sampling-weight>    : [0..100]

configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc be sampling-weight 50
configure router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc af sampling-weight 80

*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
```

```
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 2 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                                AB Samp Cnt    : 0
AB Last Adj : 03/23/2016 14:52:25              AB Next Adj    : 1440 Mins
ABMaxAvgRt  : 0 Mbps                           AB Lst AvgRt   : 0 Mbps
AB Ovfl Lmt : 1                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 2 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 2 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: underflow                        AB Monitor BW  : False
Be Weight   : 50 percent                       Af Weight      : 80 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

The sampling-weight values can be changed while auto-bandwidth is enabled. The auto-bandwidth algorithm will be reset on the LSP at the end of the current collection interval. At that time, the current bandwidth will not be adjusted and the following counters will be reset to 0: sample count, adjust count, overflow count, underflow count, max average data rate, and max average underflow data rate.

Auto-Bandwidth on LSPs with Secondary Paths

This feature is supported from 12.0.R4 onward.

When the active path goes down or becomes degraded, the bandwidth used to signal the auto-bandwidth MBB will be the operational bandwidth of the previous active path. The parameter current-bandwidth requires a modified definition:

Current-bandwidth — The last known reserved bandwidth for the LSP (this may be for a different path than the active one).

When the active path changes, the current bandwidth is updated to the operational bandwidth of the new active path. While the auto-bandwidth MBB on the active path is in progress, a statistics sample might be triggered because the intervals aren't reset when the active path changes. It is possible that an auto-adjustment is needed. The in-progress auto-bandwidth MBB will be restarted with retry attempts to 0 and signaled bandwidth equal to the new measured bandwidth. If after five attempts, auto-bandwidth MBB fails, the current bandwidth and secondary **oper-bw** remain unchanged.

For a secondary/standby path, if the active path changes without the LSP going down, an auto-bandwidth MBB is triggered for the new active path. The bandwidth used to signal the MBB is the operational bandwidth of the previous active path (current bandwidth).

If the primary path is not currently active, but it has not gone down, then any MBB should use the configured bandwidth for the primary path.

Create two new strict paths and assign them to the LSP. The primary path is the direct strict path from PE-1 to PE-2. There are two secondary paths: **path-PE-1-PE-3-PE-2** and **loose**. The first one is standby, the latter is not.

```
configure router mpls
            path path-PE-1-PE-2
                hop 10 192.0.2.2 strict
                no shutdown
            exit
            path path-PE-1-PE-3-PE-2
                hop 10 192.0.2.3 strict
                hop 20 192.0.2.2 strict
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                cspf
                fast-reroute facility
                    no node-protect
                exit
                primary loose shutdown
                no primary loose
                primary path-PE-1-PE-2
                    adaptive
                    bandwidth 2
                exit
                secondary path-PE-1-PE-3-PE-2
                    adaptive
                    bandwidth 2
                    standby
                exit
                secondary loose
                    adaptive
                    bandwidth 2
                exit
                no shutdown
            exit
```

Stop the auto-bandwidth MBB to have the current bandwidth equal to the bandwidth configured for the primary path (2 Mbps).

```
configure router mpls lsp "LSP-PE-1-PE-2" no auto-bandwidth
```

Configure auto-bandwidth with the following settings:

```
configure router mpls lsp "LSP-PE-1-PE-2"
               auto-bandwidth
                   multipliers sample-multiplier 1 adjust-multiplier 288
                   adjust-up 10 bw 1
                   max-bandwidth 20
                   min-bandwidth 2
                   overflow-limit 2 threshold 10
                   underflow-limit 3 threshold 10
                   fc be sampling-weight 50
                   fc af sampling-weight 80
               exit
```

Initially, the current bandwidth is the configured bandwidth of the primary path: 2 Mbps, but in case of overflow, it will be increased after two overflow samples (10 minutes).

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 5 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                                AB Samp Cnt    : 0
AB Last Adj : 03/24/2016 07:39:42              AB Next Adj    : 1440 Mins
ABMaxAvgRt  : 0 Mbps                           AB Lst AvgRt   : 5 Mbps
AB Ovfl Lmt : 2                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 0 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: overflow                         AB Monitor BW  : False
Be Weight   : 50 percent                       Af Weight      : 80 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

Shutdown port 1/1/1 on PE-1 to initiate a failure on the primary path.

```
*A:PE-1# configure port 1/1/1 shutdown
```

Verify that the secondary/standby path is now active.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" activepath
===============================================================================
MPLS LSP: LSP-PE-1-PE-2 (active paths)
===============================================================================
Legend :
    #  - Manually switched path
    #F - Manually forced switched path
===============================================================================
LSP Name     : LSP-PE-1-PE-2                    LSP Id       : 17452
Path Name    : path-PE-1-PE-3-PE-2              Active Path  : Standby
To           : 192.0.2.2

===============================================================================
*A:PE-1#
```

Check the auto-bandwidth data on the LSP. The current bandwidth for the LSP is the same as it used to be for the primary path. **AB Adj Cause** = activePathChange.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 5 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 1                                AB Samp Cnt    : 0
AB Last Adj : 03/24/2016 07:46:21              AB Next Adj    : 1435 Mins
ABMaxAvgRt  : 6 Mbps                           AB Lst AvgRt   : 6 Mbps
AB Ovfl Lmt : 2                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 0 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: activePathChange                 AB Monitor BW  : False
Be Weight   : 50 percent                       Af Weight      : 80 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

The original situation is restored.

```
configure port 1/1/1 no shutdown
```

The primary path comes up again.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" activepath
===============================================================================
MPLS LSP: LSP-PE-1-PE-2 (active paths)
===============================================================================
Legend :
    #  - Manually switched path
    #F - Manually forced switched path
===============================================================================
LSP Name     : LSP-PE-1-PE-2                    LSP Id       : 17456
Path Name    : path-PE-1-PE-2                   Active Path  : Primary
To           : 192.0.2.2

===============================================================================
*A:PE-1#
```

The auto-bandwidth data again shows **AB Adj Cause**: activePathChange, but with a different timestamp.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
===============================================================================
MPLS LSP (Auto Bandwidth)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-2
Auto BW     : Enabled                          AB OpState     : Up
Auto BW Min : 2 Mbps                           Auto BW Max    : 20 Mbps
AB Up Thresh: 10 percent                       AB Down Thresh : 5 percent
AB Up BW    : 1 Mbps                           AB Down BW     : 0 Mbps
AB Curr BW  : 5 Mbps                           AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                              AB Samp Mul    : 1
AB Adj Time : 1440 Mins                        AB Samp Time   : 5 Mins
AB Adj Cnt  : 2                                AB Samp Cnt    : 0
AB Last Adj : 03/24/2016 07:54:20              AB Next Adj    : 1430 Mins
ABMaxAvgRt  : 6 Mbps                           AB Lst AvgRt   : 6 Mbps
AB Ovfl Lmt : 2                                AB Ovfl Cnt    : 0
ABOvflThres : 10 percent                       AB Ovfl BW     : 0 Mbps
AB UndflLmt : 3                                AB Undrfl Cnt  : 0
ABUndflThrs : 10 percent                       AB Undrfl BW   : 0 Mbps
ABMaxUndflBW: 0 Mbps
AB Adj Cause: activePathChange                 AB Monitor BW  : False
Be Weight   : 50 percent                       Af Weight      : 80 percent
L1 Weight   : 100 percent                      L2 Weight      : 100 percent
Nc Weight   : 100 percent                      Ef Weight      : 100 percent
H1 Weight   : 100 percent                      H2 Weight      : 100 percent
===============================================================================
*A:PE-1#
```

Conclusion

Auto-bandwidth adjustment can be enabled on point-to-point LSPs in order to make a realistic bandwidth reservation, based on active iLER traffic monitoring. A user has control over how the bytes count for the different FCs by providing a sampling-weight factor. This can influence the average data rate over the sample interval.

The bandwidth is taken into account in the control plane when LSPs are established or when they change their bandwidth using MBB. The bandwidth in the data plane is not restricted by this setting.

Auto-bandwidth adjustment can be triggered in different ways: periodically, in case of overflow/underflow, manually, and in case of an active path change. It is also possible to have passive monitoring where no adjustment is done.

Automatic Creation of RSVP-TE LSPs

This chapter provides information about automatic creation of resource reservation protocol with traffic engineering (RSVP-TE) label switched paths (LSPs).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This feature is applicable to 7750 SR-7/12, 7750 SR-12e, 7450 ESS-7/12, XRS-20/16c, and 7750-c4/12 with no hardware constraints as this is a control-plane feature only.

This chapter was originally written for and tested on SR OS release 11.0.R6, but the CLI in this edition corresponds to release 14.0.R1.

Overview

Automatic creation of RSVP-TE LSPs enables the automated creation of point-to-point RSVP-TE LSPs within a single interior gateway protocol (IGP) intermediate system to intermediate system (IS-IS) level or open shortest path first (OSPF) area that can subsequently be used by services and/or IGP shortcuts. The feature is divided into two components: creation of an RSVP-TE LSP mesh, and creation of single-hop RSVP-TE LSPs. Although both can be used simultaneously, it is likely that one or the other is used.

When creating an RSVP-TE LSP mesh, the mesh can be full or partial, the extent of which is governed by a prefix-list containing the system addresses of all nodes that should form part of the mesh. When using single-hop RSVP-TE LSPs, point-to-point LSPs are established to all directly connected neighbors. The purpose of these single-hop LSPs is to allow for equal cost multi-path (ECMP) load-balancing of traffic using LDP-over-RSVP, which is not possible using native RSVP LSPs.

The use of automatically created RSVP-TE LSPs avoids manual configuration of RSVP-TE LSP meshes. Even when provisioning tools (such as 5620 SAM) are used to automatically provision these LSPs, auto-mesh still provides a benefit by avoiding increased configuration file sizes.

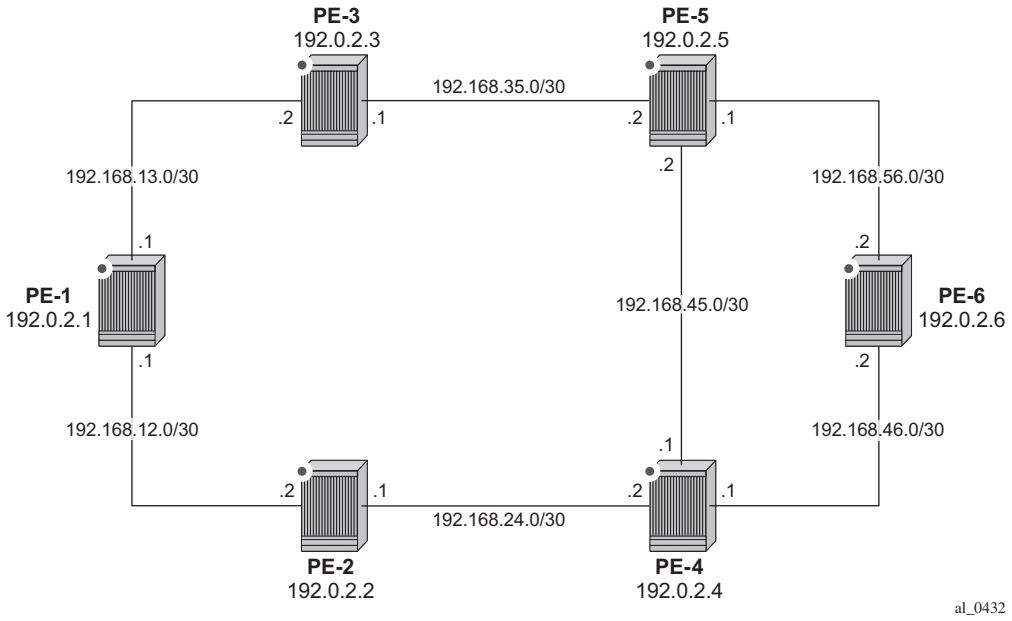
The use of automatically created targeted label distribution protocol (LDP) sessions is also described when using the automatically created RSVP LSPs for Layer 2 services.

Configuration

Example Topology

The example topology is shown in Figure 102. All routers participate in a single IS-IS Level-2 area that has traffic engineering enabled. Multiprotocol label switching (MPLS) and RSVP are enabled on every interface, but no LSPs are initially provisioned. All routers are border gateway protocol (BGP) speakers and form part of autonomous system (AS) 64496. PE-5 is a route reflector and the remaining routers are IBGP clients for the VPN IPv4 and L2-VPN address families. The objective of this example is to demonstrate how to automatically create transport LSPs using RSVP or LDP-over-RSVP, and then create services that utilize those LSPs. The exchange of BGP routes is needed for those services.

Figure 102 **Example Topology**



Automatic Creation of an RSVP-TE LSP Mesh

To start the process of automatically creating an RSVP-TE LSP mesh, the user must create a route policy referencing a prefix-list. This prefix-list contains the system addresses of all nodes that are required to be in the mesh, and can be entered as a series of /32 addresses, or simply as a range as shown configured below. This range encompasses all of the system addresses of the nodes in the test topology as the requirement is to make a full-mesh.

```
configure
    router
        policy-options
            begin
            prefix-list "System-Addresses"
                prefix 192.0.2.0/24 prefix-length-range 32-32
            exit
            policy-statement "Remote-PEs"
                entry 10
                    from
                        prefix-list "System-Addresses"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

After the route policy is created, the user must create an **lsp-template** containing the common parameters which are used to establish all point-to-point LSPs within the mesh. For an RSVP-TE LSP mesh, the **lsp-template** must be suffixed with the creation-time attribute **mesh-p2p**. Upon creation of the template, CSPF is automatically enabled (and cannot be disabled), and the template must reference a **default-path** before it can be placed in a **no shutdown** state. In the example contained in the following output, the template refers to a path named “loose” that has no strict or loose hops defined, meaning the system will dynamically calculate the path whilst considering other specified constraints. The **lsp-template** in this output also stipulates **fast-reroute facility** bypass protection. The default behavior is no node-protect, therefore this configuration requests link protection only. FRR one-to-one protection is not supported for automatically created RSVP-TE LSPs; therefore facility bypass is the only form of protection supported. Finally the template is placed in a no shutdown state.

Next, the user must associate the lsp-template with the previously defined route-policy, and this is accomplished using the **auto-lsp lsp-template** command. In this example, the lsp-template “Full-Mesh” is associated with the policy-statement “Remote-PEs” that in turn references a prefix-list containing all system addresses in the test topology. Up to five policies can be associated with a given lsp-template at the same time. If a policy associated with an lsp-template is modified in order to add or remove prefixes, the system immediately re-evaluates the policy and the prefix-list to determine if one or more LSPs need to be established, or one or more LSPs need to be torn down.

```
configure
    router
        mpls
            path "loose"
                no shutdown
            exit
            lsp-template "Full-Mesh" mesh-p2p
                default-path "loose"
                cspf
                fast-reroute facility
                exit
                no shutdown
            exit
            auto-lsp lsp-template "Full-Mesh" policy "Remote-PEs"
            no shutdown
        exit
```

Once the **auto-lsp lsp-template** command is entered, the system commences the process of establishing the point-to-point LSPs. The prefixes defined in the prefix-list are checked, and if a prefix corresponds to a router-id that is present in the traffic engineering database, the system instantiates a CSPF computed primary path to that prefix using the parameters specified in the lsp-template. With the previously

defined configuration applied at PE-6, the existence of point-to-point RSVP LSPs to every node in the test topology can be verified as shown in the following output. The LSP name is automatically constructed as TemplateName-DestIPv4Address-TunnelId. The LSP name signaled in the Session Attribute object concatenates the LSP name with the path name (for example Full-Mesh-192.0.2.1-61441::loose).

```
*A:PE-6# show router mpls lsp
===============================================================================

MPLS LSPs (Originating)
===============================================================================
LSP Name                            To              Tun     Fastfail  Adm  Opr
                                                    Id      Config
-------------------------------------------------------------------------------
Full-Mesh-192.0.2.1-61441           192.0.2.1       61441   Yes       Up   Up
Full-Mesh-192.0.2.2-61442           192.0.2.2       61442   Yes       Up   Up
Full-Mesh-192.0.2.3-61443           192.0.2.3       61443   Yes       Up   Up
Full-Mesh-192.0.2.4-61444           192.0.2.4       61444   Yes       Up   Up
Full-Mesh-192.0.2.5-61445           192.0.2.5       61445   Yes       Up   Up
-------------------------------------------------------------------------------
LSPs : 5
===============================================================================
*A:PE-6#
```

The lsp-template requests fast-reroute link protection. At PE-6, this protection can be verified by querying each primary LSP. In the following output, the primary LSP to PE-1 (Full-Mesh-192.0.2.1-61441) is signaled through PE-5 (192.0.2.5) and PE-3 (192.0.2.3), and the presence of the @ indicator after each hop denotes that link protection is available to the primary path.

```
*A:PE-6# show router mpls lsp path "Full-Mesh-192.0.2.1-61441" detail | match
expression "LSP Name|Actual Hops" post-lines 4
LSP Name         : Full-Mesh-192.0.2.1-61441
Path LSP ID      : 31232
From             : 192.0.2.6            To                  : 192.0.2.1
Admin State      : Up                   Oper State          : Up
Path Name        : loose                Path Type           : Primary
Actual Hops      :
    192.168.56.2 (192.0.2.6) @                  Record Label        : N/A
 -> 192.168.56.1 (192.0.2.5) @                  Record Label        : 131065
 -> 192.168.35.1 (192.0.2.3) @                  Record Label        : 131070
 -> 192.168.13.1 (192.0.2.1)                    Record Label        : 131067
*A:PE-6#
```

Finally, it can be verified that the signaled LSPs are placed in the tunnel table and made available to the tunnel table manager so they can be used by applications and services.

```
*A:PE-6# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
```

```
-------------------------------------------------------------------------------
192.0.2.1/32      rsvp      MPLS  61441     7        192.168.56.1   300
192.0.2.2/32      rsvp      MPLS  61442     7        192.168.46.1   200
192.0.2.3/32      rsvp      MPLS  61443     7        192.168.56.1   200
192.0.2.4/32      rsvp      MPLS  61444     7        192.168.46.1   100
192.0.2.5/32      rsvp      MPLS  61445     7        192.168.56.1   100
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-6#
```

Once the lsp-template is in use and LSPs are instantiated, it is necessary to place the template into a shutdown state to change any parameters that cannot be handled as a *Make-Before-Break* (MBB). This essentially includes all LSP parameters with the exception of bandwidth and fast-reroute without node-protection. Modification of any other parameters requires a shutdown of the lsp-template and a re-signal of the LSP once the lsp-template is placed in the no shutdown state again. MBB is supported for timer-based and manual re-signaling of the automatically created LSPs.

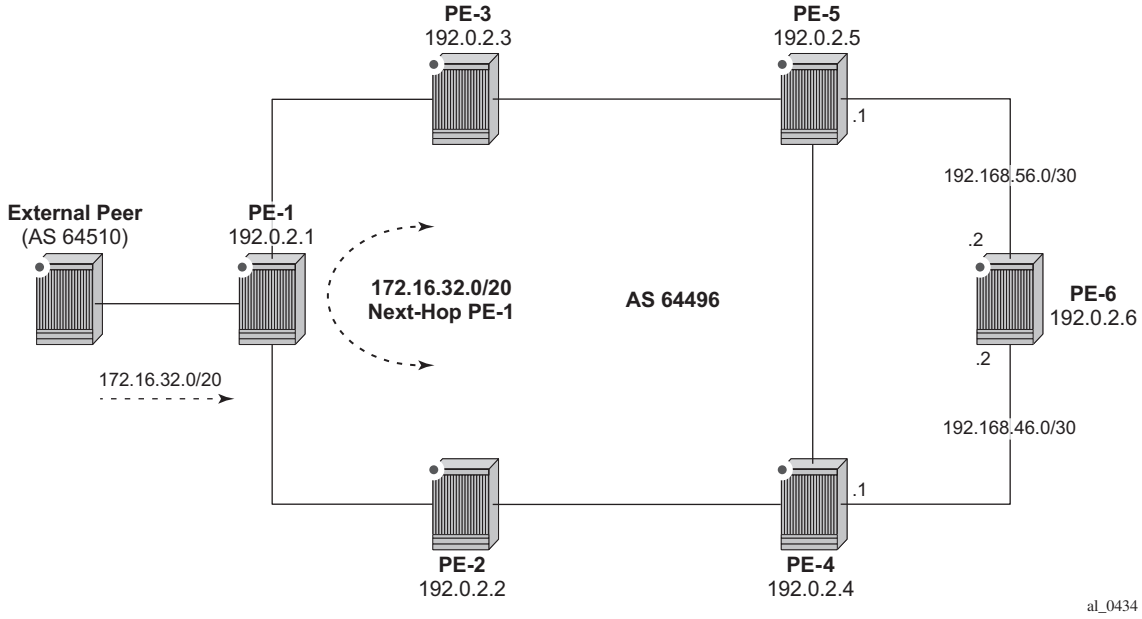
Service and Application Verification

With the RSVP-TE LSP mesh in place, it is now possible to create services and applications to utilize those LSPs. These applications and services include Layer 2 and Layer 3 VPNs, resolution of BGP labeled routes and resolution of BGP, IGP, and static routes. However, the automatically created LSPs are not available for explicit binding in a statically provisioned service distribution point (SDP).

IGP Shortcuts

Figure 103 demonstrates the use of IGP shortcuts. Prefix 172.16.32.0/20 is advertised to PE-1 from an external peer in AS 64510, which PE-1 subsequently advertises into IBGP, imposing Next-Hop-Self in the process. For more details on IGP shortcuts, see chapter IGP Shortcuts.

Figure 103 **IGP Shortcuts with RSVP-TE Auto-Mesh**



The objective is for PE-6 to use the automatically created LSP to PE-1 as an IGP shortcut (typically implemented in order to maintain a “BGP-free” core). IGP shortcuts for BGP are enabled under the main BGP context using the command **next-hop-resolution shortcut-tunnel** with options for **rsvp**, **ldp** or **bgp**. Since the example topology only has (automatically created) RSVP-TE LSPs, this option is selected. In fact, there are three more options: sr-isis, sr-ospf, and sr-te. These are related to segment routing, which is beyond the scope of this chapter.

```
configure router bgp next-hop-resolution shortcut-tunnel family
  - family <family>

 <family>             : ipv4


 [no] disallow-igp    - Allow/Disallow IGP shortcuts
     resolution      - Configure resolution state of BGP unlabelled routes to tunnels
      resolution-fil* + Configure specific tunnels to be used for resolving BGP
unlabelled routes


configure router bgp next-hop-resolution shortcut-tunnel family ipv4 resolution-filter
  - resolution-filter

 [no] bgp             - Use BGP tunnelling for next hop resolution
 [no] ldp             - Use LDP tunnelling for next hop resolution
 [no] rsvp            - Use RSVP tunnelling for next hop resolution
 [no] sr-isis         - Use sr-isis for next hop resolution
 [no] sr-ospf         - Use sr-ospf for next hop resolution
 [no] sr-te           - Use sr-te for next hop resolution
```

```
configure
    router
        bgp
            next-hop-resolution
                shortcut-tunnel
                    family ipv4
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                exit
            exit
```

Once the shortcuts are enabled, the route-table (and FIB) can be validated to ensure that the programmed next hop is the advertising BGP speaker (as opposed to the IGP next hop), and that traffic is tunneled to that next hop through an RSVP LSP. In this case the RSVP LSP is the LSP with Tunnel-Id 61441, which is the LSP to PE-1.

```
*A:PE-6# show router route-table 172.16.32.0/20
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
172.16.32.0/20                                  Remote  BGP       00h08m16s  170
       192.0.2.1 (tunneled:RSVP:61441)                               0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Layer 3 VPN

Layer 3 VPNs can utilize the automatically created LSPs by using the **auto-bind-tunnel** feature configured with the **rsvp** option (possibly in combination with LDP). The option to include both rsvp and ldp, allows the system to use an RSVP LSP if one exists, and if not, to revert to an LDP-based LSP. A virtual private routed network (VPRN) is configured in this manner at PE-1 and PE-6. PE-1 is configured with a loopback address of 172.16.1.1/24 and advertises the VPN IPv4 prefix 172.16.1.0/24 into IBGP, whilst PE-6 is configured with a loopback address of 172.16.6.1/24 and advertises the VPN IPv4 prefix 172.16.6.0/24 into IBGP. The next output illustrates the configuration at PE-6. The only difference at PE-1 is the IP address assigned to the loopback interface.

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "loopback" create
                address 172.16.6.1/24
                loopback
            exit
            no shutdown
        exit
```

Before a VPN IPv4 prefix is considered **valid**, the receiving SR OS PE router must be able to resolve the BGP next hop to an LSP in the tunnel table (if not, the prefix is held in routing information base RIB-In and flagged as **invalid**). At PE-6, it is possible to verify that the VPN IPv4 prefix 172.16.1.0/24 received from PE-1 is correctly resolved by looking at the VPRN-specific route table. In the following output, the VPN IPv4 prefix 172.16.1.0/24 with a next hop of PE-1 (192.0.2.1) is correctly resolved to an RSVP LSP with a tunnel ID of 61441.

```
*A:PE-6# show router 1 route-table 172.16.1.0/24
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                   Remote  BGP VPN   01h51m58s  170
       192.0.2.1 (tunneled:RSVP:61441)                              0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Layer 2 VPN

As previously described, automatically created RSVP LSPs cannot be referenced by statically provisioned SDPs. Without the ability for SDPs to explicitly reference automatically created RSVP LSPs, there is little value in manually defining SDPs within Layer 2 service constructs (there is little point in referring to an SDP that cannot bind to the underlying RSVP mesh). Therefore, in order to deliver Layer 2

services there is a requirement to adopt a model within the service construct that permits automatic creation of SDP bindings, and this is achieved using a pseudowire-template dictating the characteristics of the SDP. The secondary effect of using pseudowire-templates to dynamically create SDPs is that these automatically created SDPs can currently only use LDP or BGP as a transport tunnel, not RSVP. The solution is to enable LDP-over-RSVP.

This can be implemented using static provisioning of peers as shown in the next output, or it can be done using automatic creation of T-LDP sessions. Regardless of the method, a reciprocal configuration must exist at both peer endpoints. The static per-peer configuration is applied in the **targeted-session** context specifying the remote peer system IP address, and the keyword **tunneling,** which enables tunneling of LDP FECs over RSVP LSPs with a far-end address matching that of the T-LDP peer. At a global level, the **prefer-tunnel-in-tunnel** command is shown, but is only required when a next hop router advertises a FEC over link-level LDP and T-LDP. In this case, by default the system would prefer the link-level LDP tunnel, so the **prefer-tunnel-in-tunnel** instructs the system to prefer an LDP-over-RSVP tunnel if it is available. Although link-layer LDP is not present in the test topology, the command is included because the presence of link-layer LDP is common.

```
configure
    router
        ldp
            prefer-tunnel-in-tunnel
            interface-parameters
            exit
            targeted-session
                peer 192.0.2.1
                    tunneling
                    exit
                exit
            exit
            no shutdown
        exit
```

The next output provides an example demonstrating the automatic creation of T-LDP sessions. Here, no explicit reference is made to specific peers, but rather a **peer-template** is configured containing the parameters which apply to all T-LDP sessions spawned by this template. In this example, only the **tunneling** command is required. A **peer-template-map** is then used to create a mapping between the **peer-template** (TLDP-Mesh) and a **policy** defining the IP addresses of remote nodes to which T-LDP sessions should be established. In this example, the policy “Remote-PEs” is the same policy previously used by the auto-created RSVP LSP mesh.

```
configure
    router
        ldp
            prefer-tunnel-in-tunnel
            interface-parameters
            exit
            targeted-session
```

```
            peer-template "TLDP-Mesh"
                tunneling
            exit
            peer-template-map peer-template "TLDP-Mesh" policy "Remote-PEs"
        exit
        no shutdown
    exit
```

Regardless of whether T-LDP sessions are explicitly provisioned, or dynamically created using a peer-template, the result is that a targeted LDP session is established which can be used for advertising address and service FECs, and which is capable of tunneling LDP over RSVP.

```
*A:PE-6# show router ldp targ-peer 192.0.2.1 detail
===============================================================================
LDP IPv4 Targeted Peers
===============================================================================
-------------------------------------------------------------------------------
192.0.2.1
-------------------------------------------------------------------------------
Admin State        : Up              Oper State           : Up
Last Oper Chg      : 0d 00:05:15
Hold Time          : 45              Hello Factor         : 3
Oper Hold Time     : 45
Hello Reduction    : Disabled        Hello Reduction Fact*: 3
Keepalive Timeout  : 40              Keepalive Factor     : 4
Active Adjacencies : 1               Last Modified        : Never
Auto Created       : Yes
Creator            : template        Template Name        : TLDP-Mesh
Tunneling          : Enabled
Lsp Name           : None
Local LSR          : None
BFD Status         : Disabled
===============================================================================
No. of IPv4 Targeted Peers: 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-6#
```

To create VPLS services using dynamically-created SDPs, BGP auto-discovery (BGP-AD) must be used together with LDP (or BGP) pseudowire signaling, for more details see LDP VPLS using BGP-Auto Discovery.

In the following output PE-6 uses BGP-AD and LDP signaling. The same configuration is applied at PE-1. The **vpls-id** is configured in the **bgp-ad** context. The **vpls-id** is a network-wide identifier assigned to all VPLS switch instances (VSIs) belonging to the same VPLS, and is carried in VPLS network layer reachability information (NLRI) as an extended community attribute. A second parameter used for BGP-AD and carried in the VPLS NLRI is the VSI-ID, which uniquely identifies

each VSI. The VSI-ID is automatically derived from the global ASN, the VPLS service ID, and the system IP address. To automatically create SDPs, the **bgp** context of the VPLS service refers to a **pw-template** defining the parameters of the pseudowire. In this example, the use of the hash (entropy) label is enabled in the pseudowire template, and a **split-horizon-group,** SHG, is applied.

```
configure
    service
        pw-template 2 create
            hash-label
            split-horizon-group "SHG"
            exit
        exit
        vpls 2 customer 1 create
            bgp
                pw-template-binding 2
                exit
            exit
            bgp-ad
                vpls-id 64496:2
                no shutdown
            exit
            sap 1/1/4:2 create
            exit
            no shutdown
        exit
```

The following service information provides the BGP and BGP-AD operational parameters, and shows that an SDP of type **BgpAd** (17407:4294967295) has been automatically created. Both the SDP and the SAP are operationally up.

```
*A:PE-6# show service id 2 bgp

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 64496:2
Oper RD Type         : derivedVpls
Rte-Target Import    : None                  Rte-Target Export: None
Oper RT Imp Origin   : derivedVpls           Oper RT Import   : 64496:2
Oper RT Exp Origin   : derivedVpls           Oper RT Export   : 64496:2
PW-Template Id       : 2                     PW-Template SHG  : None
Oper Group           : None
Mon Oper Group       : None
BFD Template         : None
BFD-Enabled          : no                    BFD-Encap        : ipv4
Import Rte-Tgt       : None
-------------------------------------------------------------------------------
===============================================================================
*A:PE-6#

*A:PE-6# show service id 2 bgp-ad
```

```
-------------------------------------------------------------------------------
BGP Auto-discovery Information
-------------------------------------------------------------------------------
Admin State       : Up
Vpls Id           : 64496:2
Prefix            : 192.0.2.6
-------------------------------------------------------------------------------
*A:PE-6#

*A:PE-6# show service id 2 base | match "Service Access" post-lines 10
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2                               q-tag        1518    1518    Up   Up
sdp:17407:4294967295 SB(192.0.2.1)        BgpAd        0       1548    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-6#
```

To create Epipe services using dynamically created SDPs, two options exist. Either LDP FEC 129 signaling can be used, which in turn dictates the presence of pseudowire routing information, or BGP-VPWS based signaling can be used, for more details see BGP Virtual Private Wire Services. This example illustrates the use of BGP VPWS, but in either case, only single-segment pseudowires are supported. The next output shows the configuration requirements for a basic BGP-based Epipe service at PE-6. Once again a **pw-template** is used to define the characteristics of the pseudowire, and this template is referenced in the **bgp** context of the Epipe service. The **bgp** context is also where the **route-distinguisher** and **route-target** values are configured, which are carried in the VPWS NLRI and extended communities respectively. The **ve-name**, **ve-id**, and **remote-ve-name** are all configured in the **bgp-vpws** context. The **ve-id** is carried in the VPWS NLRI, and when a PE router receives a VPWS NLRI to try to establish an Epipe service, the **ve-id** from the NLRI is validated against the **ve-id** configured in the **remote-ve-name**. These must match before the Epipe becomes operational.

```
configure
    service
        pw-template 3 create
            hash-label
        exit
        epipe 3 customer 1 create
            bgp
                route-distinguisher 64496:3
                route-target export target:64496:3 import target:64496:3
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-6"
                    ve-id 6
                exit
                remote-ve-name "PE-1"
```

```
                ve-id 1
            exit
            no shutdown
        exit
        sap 1/1/4:3 create
        exit
        no shutdown
    exit
```

The basic service information is truncated to show only the relevant information in order to verify that the service is operational. SDP (17406:4294967294) has been automatically created and is of type **BgpVpws**. Both the SDP and the SAP are operationally up.

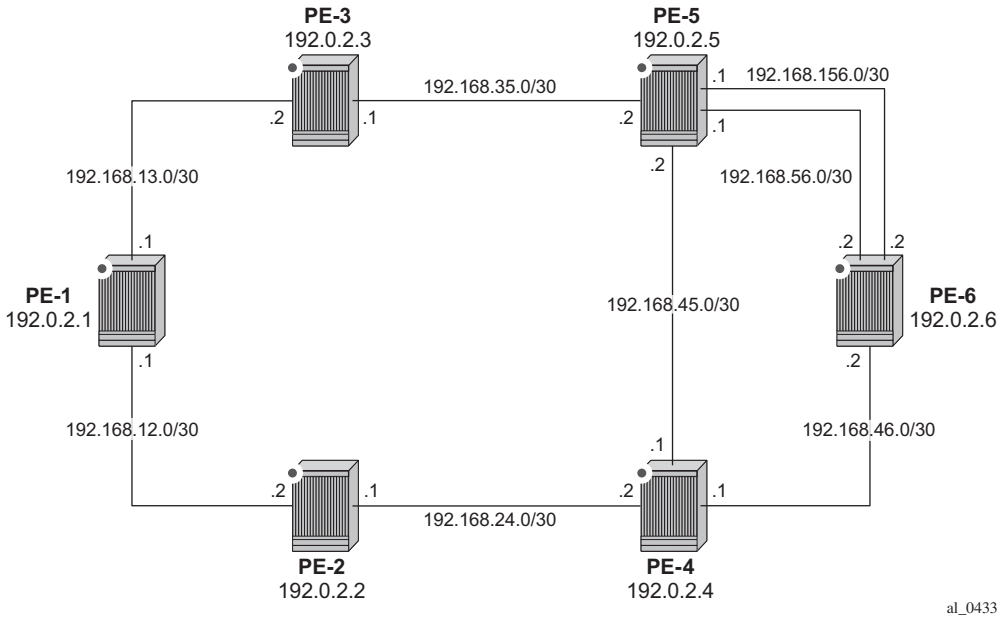
```
*A:PE-6# show service id 3 base | match "Service Access" post-lines 10
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:1/1/4:3                               q-tag        1518    1518    Up   Up
sdp:17406:4294967294 SB(192.0.2.1)        BgpVpws      0       1548    Up   Up
==============================================================================
*A:PE-6#
```

Automatic Creation of RSVP Single-Hop LSPs

As previously discussed, the purpose of a single-hop LSP mesh is to allow for ECMP load-balancing of traffic using LDP-over-RSVP. ECMP load-balancing could be implemented using LDP over a partial or full mesh of RSVP-TE LSPs, but the use of single-hop LSPs additionally allows for load-balancing across a number of parallel RSVP LSPs between nodes. To illustrate ECMP load-balancing over multiple parallel RSVP LSPs the example topology of Figure 102 is modified to include a parallel link between PE-6 and PE-5 as shown in Figure 104. In addition, all routers are enabled for ECMP=2.

```
configure router ecmp 2
```

Figure 104 **Test Topology for Single-Hop LDP-over-RSVP with ECMP**



Unlike the automatically created RSVP-TE LSP mesh previously discussed, the automatically created single-hop RSVP-TE LSPs have no requirement for a prefix-list to be referenced containing the prefixes of the remote nodes that form part of the mesh. In the case of automatically created single-hop LSPs the TE database keeps track of each TE link which comes up to a directly connected IGP neighbor. The system then establishes a single-hop LSP with a destination address matching the router-id of the neighbor and with a strict hop consisting of the address of the interface used by the TE link.

The first requirement is to create an **lsp-template** containing the common parameters used to establish each single-hop LSP. For a single-hop LSP mesh the **lsp-template** must be suffixed with the creation-time attribute **one-hop-p2p**. Upon creation of the template, **cspf** is automatically enabled (and cannot be disabled), and the **hop-limit** is set to a value of **2**. The hop-limit defines the number of nodes the LSP may traverse, and since these are single-hop LSPs to adjacent neighbors a limit of 2 is sufficient. The template must also reference a **default-path** before it can be placed in the no shutdown state. The following example references a path named “loose” that has no strict or loose hops defined. When the RSVP PATH message is actually generated to create the one-hop LSP, it contains one strict-hop to the interface address of the neighbor; and as destination the system address of the adjacent node.

The next requirement is to trigger the creation of single-hop LSPs, and this is achieved using the **auto-lsp lsp-template** command. In this example, the lsp-template “Single-Hop” is referenced, and the command is completed with the keyword **one-hop** to indicate the creation of single-hop LSPs. Unlike an RSVP-TE mesh there is no requirement to reference a route-policy. In the example, the auto-lsp with lsp-template “Full-Mesh” is removed.

```
configure router mpls no auto-lsp lsp-template "Full-Mesh"

configure
    router
        mpls
            path "loose"
                no shutdown
            exit
            lsp-template "Single-Hop" one-hop-p2p
                default-path "loose"
                cspf
                hop-limit 2
                no shutdown
            exit
            auto-lsp lsp-template "Single-Hop" one-hop
            no shutdown
        exit
    exit
exit
```

Once the **auto-lsp lsp-template** command is entered, the system starts the process of establishing the single-hop LSPs. A check is made of the TE database for every TE link to a directly connected IGP neighbor, and a single-hop LSP is established across each TE link. The following output is taken from PE-6 and shows the automatically created single-hop LSPs. The LSP names are automatically constructed as TemplateName-DestIPv4Address-TunnelId. The LSP name signaled in the session attribute object concatenates the LSP name with the path name (for example Single-Hop-192.0.2.4-61499::loose). Recall from Figure 104 that PE-6 has a single TE-enabled link to PE-4, and two TE-enabled links to PE-5, therefore with ECMP=2, there is one LSP to PE-4 (192.0.2.4) and two LSPs to PE-5 (192.0.2.5). However, if ECMP=1, only one single-hop LSP would be signaled to PE-5.

```
*A:PE-6# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                            To              Tun     Fastfail  Adm  Opr
                                                    Id      Config
-------------------------------------------------------------------------------
Single-Hop-192.0.2.4-61448          192.0.2.4       61448   No        Up   Up
Single-Hop-192.0.2.5-61449          192.0.2.5       61449   No        Up   Up
Single-Hop-192.0.2.5-61450          192.0.2.5       61450   No        Up   Up
-------------------------------------------------------------------------------
LSPs : 3
===============================================================================
*A:PE-6#
```

As the purpose of single-hop LSPs is to enable ECMP load-balancing using LDP-over-RSVP, there is a requirement to configure T-LDP sessions between RSVP LSP endpoints. This can be implemented using static peer provisioning, or it can be done using automatic creation of T-LDP sessions, both of which have been previously described and they are therefore not repeated. In this example, the automatic creation of T-LDP sessions approach is used, and T-LDP sessions are created to adjacent neighbors that are capable of tunneling inside RSVP.

```
*A:PE-6# show router ldp session

===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.4:0         Targeted  Established   17        18        0d 00:00:45
192.0.2.5:0         Targeted  Established   18        19        0d 00:00:51
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================

===============================================================================
LDP IPv6 Sessions
===============================================================================
Peer LDP Id
 Adj Type            State          Msg Sent       Msg Recv       Up Time
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-6#
```

To validate the ECMP load-balancing capability, PE-5 is configured to advertise prefix 172.16.5.0/24 to PE-6. In turn, PE-6 is configured for **ibgp-multipath** to enable load-balancing over IGP links to the BGP next hop address, **next-hop-resolution shortcut-tunnel resolution-filter ldp** to enable tunneling of traffic destined toward the BGP next hop in MPLS, and **ecmp 2**.

```
configure
    router
        bgp
            ibgp-multipath
            next-hop-resolution
                shortcut-tunnel
                    family ipv4
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                exit
            exit
            no shutdown
```

The prefix 172.16.5.0/24 advertised by PE-5 is learned at PE-6 and installed in the RIB/FIB with PE-5's system address (192.0.2.5) as next hop.

```
*A:PE-6# show router bgp routes 172.16.5.0/24
===============================================================================
 BGP Router ID:192.0.2.6        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  172.16.5.0/24                                     100         None
      192.0.2.5                                         None        -
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-6#
```

Checking the FIB for the next hop address 192.0.2.5, it can be verified that both links are installed as next hop addresses, meaning that ECMP load-balancing is active.

```
*A:PE-6# show router fib 1 192.0.2.5/32

===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                          Protocol
  NextHop
-------------------------------------------------------------------------------
192.0.2.5/32                                            ISIS
  192.168.56.1 (int-PE-6-PE-5)
  192.168.156.1 (int-PE-6-PE-5-2nd)
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-6#
```

Conclusion

Automatic creation of RSVP-TE LSPs provides a good solution for reducing the amount of provisioning activity required when configuring RSVP LSPs. However, there are some constraints with regard to the way that services are deployed on top of those LSPs. SDPs cannot explicitly reference automatically created RSVP LSPs, which means that automatically created SDPs need to be used for Layer 2 services. In turn, automatically created SDPs can only use LDP or BGP as a transport tunnel (not RSVP), therefore in order to use the automatically created RSVP mesh, LDP over RSVP must be used. These caveats need to be fully understood before considering deployment of automatically created RSVP-TE LSPs.

Class-Based Forwarding

This chapter provides information about class-based forwarding

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS routers and was initially written for release 13.0.R7. The CLI in this edition corresponds to release 14.0.R1.

Before SR OS 13.0.R6, class-based forwarding (CBF) was only applicable to resource reservation protocol (RSVP) label switched paths (LSPs) in multi-LSP service distribution points (SDPs) used by services. In release 13.0.R6, CBF of label distribution protocol (LDP) prefix packets over interior gateway protocol (IGP) shortcuts was introduced. Both implementations are described in this chapter.

Overview

In large networks, services are typically required from any PE to any other PE, and can traverse multiple domains. Within a service, different traffic classes can coexist, each with different requirements for latency and jitter.

With CBF, packets with different forwarding classes (FCs) can be forwarded on different LSPs. When equal-cost multipath (ECMP) routing without CBF is used, packets are distributed over the whole set of LSPs, without any distinction for the FC. CBF is based on FC, not on the traffic being in-profile or out-of-profile, and is a local decision, which makes it easier for interoperability.

This feature may be useful when certain links have less bandwidth and should only be used for high priority traffic. A service provider might decide to send real-time traffic over a shorter path with less bandwidth, while the bulk of the traffic takes a longer path with more bandwidth.

Configuration

The initial implementation for CBF was based on RSVP LSPs. This is described first, followed by CBF of LDP prefix packets over IGP shortcuts.

CBF over RSVP LSPs

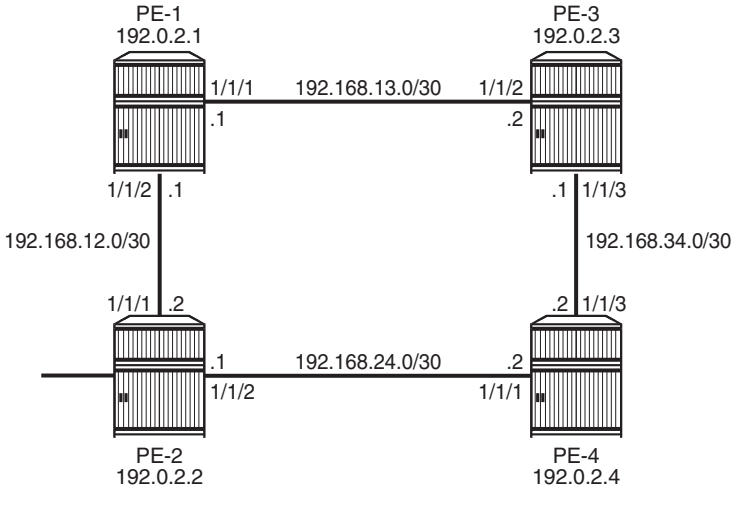
CBF over RSVP LSPs allows a service packet to be forwarded over a specific RSVP LSP, part of an SDP, based on its service ingress determined FC, typically controlled by a default or operator-defined **sap-ingress** policy. A default LSP is configured for all traffic that does not match the FCs that are explicitly configured in the SDP. CBF over RSVP LSPs is also used to forward a received packet that has been classified at the ingress service access point (SAP) into an FC, if the LSP that supports its FC is not available.

Note: CBF can also be enabled on static LSPs in the same way, but that scenario is not common.

A multicast LSP is configured for broadcast, unknown, and multicast (BUM) traffic. When no multicast LSP is defined, BUM traffic uses the default LSP. Because there are multiple LSPs per SDP, CBF over RSVP LSPs increases the number of RSVP sessions.

The test topology is shown in Figure 105. This topology will be extended in a subsequent use case.

Figure 105 **Test Topology for CBF on RSVP LSPs**



Initial Configuration

All nodes have the following initial configuration:

- Cards, media dependent adapters (MDAs), ports
- Router interfaces
- IGP open shortest path first (OSPF) or intermediate system to intermediate system (IS-IS)
- Multiprotocol label switching (MPLS) enabled on all router interfaces
- RSVP enabled
- RSVP LSPs

As an example, the configuration on PE-1 is shown. The configuration for PE-2 is similar.

```
*A:PE-1# configure router
        interface "int-PE-1-P-3"
            address 192.168.13.1/30
            port 1/1/1
        exit
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.1/32
        exit
```

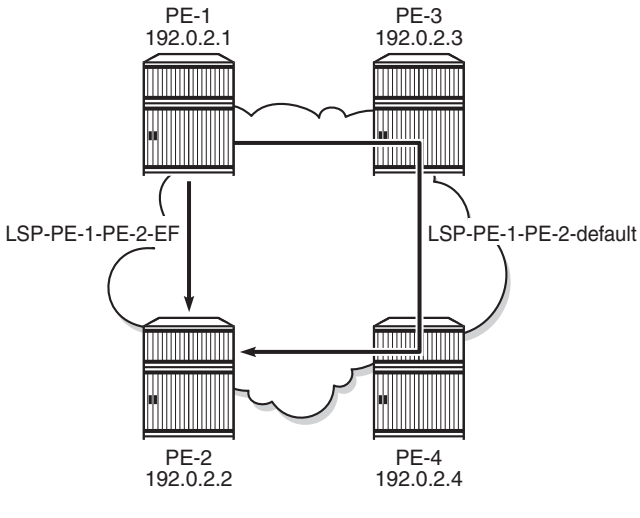
```
*A:PE-1# configure router
        ospf
            traffic-engineering
            area 0.0.0.0
                interface "system"
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-1-P-3"
                    interface-type point-to-point
                exit
            exit
        exit
*A:PE-1# configure router
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-P-3"
            exit
            no shutdown
        exit
*A:PE-1# configure router rsvp no shutdown
```

The RSVP LSPs use paths with strict hops:

```
*A:PE-1# configure router
        mpls
            path "direct-PE-1-PE-2"
                hop 10 192.168.12.2 strict
                no shutdown
            exit
            path "indirect-PE-1-PE-2"
                hop 10 192.168.13.2 strict
                hop 20 192.168.34.2 strict
                hop 30 192.168.24.1 strict
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2-EF"
                to 192.0.2.2
                primary "direct-PE-1-PE-2"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2-default"
                to 192.0.2.2
                primary "indirect-PE-1-PE-2"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The configured RSVP LSPs are shown in Figure 106.

Figure 106 **LSPs with Direct and Indirect Path toward PE-2**



Configure SDPs

In the initial configuration, the traffic is by default sent on LSP "LSP-PE-1-PE-2-default", except when the FC is expedited forwarding (EF). The traffic with FC EF is sent over LSP "LSP-PE-1-PE-2-EF". Both LSPs are assigned to SDP 122 on PE-1:

```
*A:PE-1# configure service
        sdp 122 mpls create
            description "SDP-PE-1-PE-2"
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2-EF"
            lsp "LSP-PE-1-PE-2-default"
            path-mtu 1514
            class-forwarding default-lsp "LSP-PE-1-PE-2-default"
                fc "ef" lsp "LSP-PE-1-PE-2-EF"
                no shutdown
            exit
            no shutdown
        exit
```

Note: A change in the CBF configuration may result in a change of forwarding behavior.

The configuration of SDP 212 from PE-2 to PE-1 is similar.

For the SDPs to get operational up, targeted-LDP (T-LDP) must be enabled on the PE nodes:

```
*A:PE-1# configure router ldp no shutdown
```

The state of the SDP can be verified as follows:

```
*A:PE-1# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr          Del     LSP   Sig
-------------------------------------------------------------------------------
122    1514    1514    192.0.2.2        Up   Up           MPLS    R     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
===============================================================================
*A:PE-1#
```

Some Considerations

- When CBF is enabled, a default LSP must be configured. An error is raised when class-forwarding is enabled without a default LSP:

```
*A:PE-1>config>service>sdp# class-forwarding
MINOR: CLI Default LSP must be specified.
```

- Only one LSP can be configured as the default LSP in the SDP context. Configuring another LSP as the default LSP overwrites the originally configured LSP.
- Only one LSP can be assigned to an FC in the SDP context. Configuring another LSP for an FC overwrites the originally configured LSP.
- An LSP can be assigned to multiple FCs in the SDP context.
- The default LSP can also be assigned to one or more FCs in the SDP context.

```
*A:PE-1>config>service# info
----------------------------------------------
        sdp 122 mpls create
            description "RSVP-SDP-PE-1-PE-2"
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2-EF"
            lsp "LSP-PE-1-PE-2-default"
            path-mtu 1514
            keep-alive
                shutdown
            exit
            class-forwarding default-lsp "LSP-PE-1-PE-2-default"
                fc "af" lsp "LSP-PE-1-PE-2-default"
                fc "be" lsp "LSP-PE-1-PE-2-default"
```

```
            fc "ef" lsp "LSP-PE-1-PE-2-EF"
            multicast-lsp "LSP-PE-1-PE-2-default"
            no shutdown
        exit
        no shutdown
    exit
```

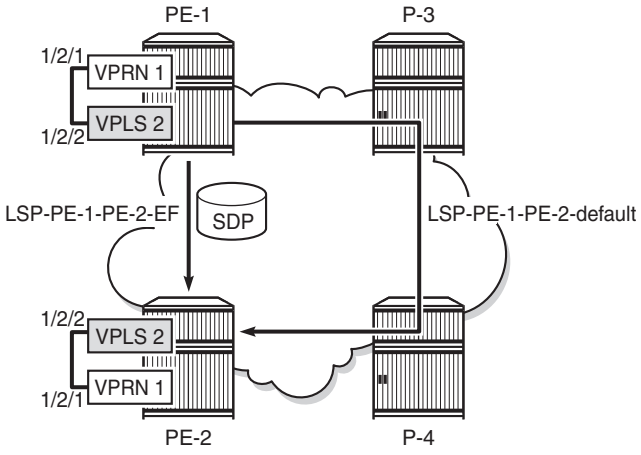
- The SDP goes down when the default LSP goes down.
- When CBF is configured on the LSP/LSP template as well as on the SDP, the configuration on the LSP/LSP template is ignored. Configuring CBF in the LSP/LSP template context is required in another use case: CBF of LDP Prefix Packets over IGP Shortcuts. This is described later in this chapter.

Configure Services for Traffic Verification

To verify that EF traffic is sent over the direct link while traffic with a different FC is sent over the longer path, two services are configured on PE-1 and PE-2.

To generate traffic, ping messages will be sent in virtual private routed network 1 (VPRN 1). The outgoing traffic is sent to virtual private LAN service 2 (VPLS 2) where the FC can be modified to EF if needed. VPLS 2 has the spoke SDP with the different LSPs and CBF. This is shown in Figure 107.

Figure 107 **CBF on RSVP LSPs - Services**




```
*A:PE-1# configure service
        vprn 1 customer 1 create
            route-distinguisher 64496:11
```

```
            vrf-target target:64496:1
            interface "loopback1" create
                address 172.31.1.1/32
                loopback
            exit
            interface "int-PE-1-PE-2-VPRN1" create
                address 192.168.112.1/30
                sap 1/2/1 create
                exit
            exit
            static-route-entry 172.31.2.1/32
                next-hop 192.168.112.2
                    no shutdown
                exit
            exit
            no shutdown
        exit
*A:PE-1# configure service
        vpls 2 customer 1 create
            description "VPLS to modify FC for traffic from VPRN 1"
            sap 1/2/2 create
            exit
            spoke-sdp 122:2 create
            exit
            no shutdown
        exit
```

The configuration of the services on PE-2 is similar.

The operational state of the spoke SDP can be verified as follows:

```
*A:PE-1# show service sdp-using
===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                         State
-------------------------------------------------------------------------------
2          122:2              Spok   192.0.2.2            Up    262139  262139
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

Verify CBF on LSP for Specific FC

To verify that EF traffic is sent out on "LSP-PE-1-PE-2-EF", via port 1/1/2 on PE-1, the configuration of VPLS 2 needs to be modified. The default FC is set to EF.

The SAP ingress policy to set the FC to EF is applied in VPLS 2 on PE-1 as follows:

```
*A:PE-1# configure qos sap-ingress 2 create
```

```
            default-fc ef
            exit
*A:PE-1# configure service
        vpls 2
            sap 1/2/2 create
                ingress qos 2
            exit
        exit
```

The configuration is identical for PE-2.

First, the port statistics are cleared.

```
*A:PE-1# clear port 1/1/[1..2] statistics
```

One thousand ping messages will be sent from VPRN 1 on PE-1 to the loopback address in VPRN 1 on PE-2:

```
*A:PE-1# ping router 1 172.31.2.1 rapid count 1000
```

The port statistics are verified. The FC is equal to EF, so the LSP "LSP-PE-1-PE-2-EF" is used and the traffic is sent via port 1/1/2.

```
*A:PE-1# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                        21            1740              21            1708
===============================================================================
*A:PE-1# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                      1033          126954            1033          126898
===============================================================================
```

Note: The traffic on the unused port (1/1/1) is not strictly equal to zero and the traffic on the used port (1/1/2) is not strictly equal to 1000, because of other traffic, such as OSPF messages.

Verify CBF on Default LSP

The SAP ingress policy to define the FC is removed from SAP 1/2/2 of VPLS 2. Traffic that enters a SAP where no SAP ingress policy with FC is defined always gets FC best effort (BE). Traffic with FC BE will be sent on the default LSP "LSP-PE-1-PE-2-default" on port 1/1/1 on PE-1, not on the direct link to PE-2.

```
*A:PE-1# configure service vpls 2 sap 1/2/2 ingress no qos 2
*A:PE-1# clear port 1/1/[1..2] statistics
*A:PE-1# ping router 1 172.31.2.1 rapid count 1000
```

The port statistics are verified. The FC is not equal to EF, so the default LSP is used and the traffic is sent via port 1/1/1.

```
*A:PE-1# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets          Octets
-------------------------------------------------------------------------------
1/1/1                      1015          125284            1015          125284
===============================================================================
*A:PE-1# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets          Octets
-------------------------------------------------------------------------------
1/1/2                        12            1068              12            1068
===============================================================================
```

Because no LSP has been configured to transport packets of classes other than EF, the default LSP will be used for all traffic classes different from EF. A similar outcome occurs when the FC of the ping messages is set to assured forwarding (AF), or any other FC different from EF.

The SAP ingress policy to set the FC to AF is applied in VPLS 2 on PE-1 as follows:

```
*A:PE-1# configure qos sap-ingress 2 create
            default-fc af
            exit
*A:PE-1# configure service
        vpls 2
            sap 1/2/2 create
                ingress qos 2
            exit
        exit
```

The configuration is identical for PE-2.

The port statistics are cleared and ping messages are sent. The result shows that the same port (that is, 1/1/1) is used for the generated traffic:

```
*A:PE-1# clear port 1/1/[1..2] statistics
*A:PE-1# ping router 1 172.31.2.1 rapid count 1000
*A:PE-1# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress           Egress
Id                      Packets          Octets         Packets          Octets
-------------------------------------------------------------------------------
1/1/1                      1005          124342            1007          124750
===============================================================================
*A:PE-1# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress           Egress
Id                      Packets          Octets         Packets          Octets
-------------------------------------------------------------------------------
1/1/2                        10             846              10             864
===============================================================================
*A:PE-1#
```

As indicated, the default LSP is used for all traffic with an FC that is not mapped to a specific LSP, which includes the case where a mapping FC-to-LSP has been defined but the LSP is not available. In the current example, all traffic with FC EF will be sent on LSP "LSP-PE-1-PE-2-EF", unless that LSP is unavailable. The default LSP will carry traffic with FC EF in that latter case.

The SAP ingress policy to assign the FC EF is applied in VPLS 2 on PE-1 as follows:

```
*A:PE-1# configure qos sap-ingress 2 create
            default-fc ef
            exit
*A:PE-1# configure service
        vpls 2
            sap 1/2/2 create
                ingress qos 2
            exit
        exit
```

The configuration is identical for PE-2.

To make the LSP "LSP-PE-1-PE-2-EF" unavailable, it is put in the shutdown state on PE-1. LSP "LSP-PE-2-PE-1-EF" on PE-2 is also put in the shutdown state.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2-EF" shutdown
*A:PE-2# configure router mpls lsp "LSP-PE-2-PE-1-EF" shutdown
```

The port statistics are cleared and ping messages are sent. The packets are classified as FC EF in VPLS 2:

```
*A:PE-1# clear port 1/1/[1..2] statistics
*A:PE-1# ping router 1 172.31.2.1 rapid count 1000
```

Because LSP "LSP-PE-1-PE-2-EF" (which uses port 1/1/2) is not operational, the traffic is sent on the default LSP on port 1/1/1 instead:

```
*A:PE-1# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                      1010          124858            1009         124620
===============================================================================
*A:PE-1# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                        14            1046              14           1046
===============================================================================
```

The LSP "LSP-PE-1-PE-2-EF" is re-enabled and the default LSP is now disabled. In this case, the SDP goes down. Ping messages will time out. The SAP ingress quality of service (QoS) policy in VPLS 2 is removed and the FC of the ping messages will no longer be changed to EF.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2-EF" no shutdown
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2-default" shutdown
*A:PE-1# configure service vpls 2 sap 1/2/2 ingress no qos 2
```

The SDP is operational down when the default LSP is down, as can be verified:

```
*A:PE-1# show service sdp-using
===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                         State
-------------------------------------------------------------------------------
2          122:2              Spok   192.0.2.2            Down  262139  262139
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The default LSP is re-enabled.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2-default" no shutdown
```

Define Multicast LSP for BUM Traffic

The multicast LSP specifies the LSP in the SDP to use to forward BUM traffic. The LSP name must exist and must have been associated with this SDP. In this example, the default LSP is configured as the multicast LSP.

```
*A:PE-1# configure service sdp 122 class-forwarding multicast-lsp "LSP-PE-1-PE-2-
default"
*A:PE-2# configure service sdp 212 class-forwarding multicast-lsp "LSP-PE-2-PE-1-
default"
```

The SAP ingress policy to set the default FC to EF is enabled on SAP 1/2/2 in VPLS 2 in PE-1 and PE-2.

```
*A:PE-1# configure service vpls 2 sap 1/2/2 ingress qos 2
```

Ping messages are classified as FC EF in VPLS 2. This traffic is sent over "LSP-PE-1-PE-2-EF". However, if the traffic is unknown, it will be sent over the multicast LSP, which is "LSP-PE-1-PE-2-default". To get unknown traffic, the forwarding database (FDB) is cleared and MAC learning is disabled in VPLS 2 on PE-1.

```
*A:PE-1# clear service id 2 fdb all
*A:PE-1# configure service vpls 2 disable-learning
```

Clear port statistics and launch one thousand rapid ping messages in VPRN 1 on PE-1.

```
*A:PE-1# clear port 1/1/[1..2] statistics
*A:PE-1# ping router 1 172.31.2.1 rapid count 1000
```

Verify the port statistics. BUM traffic is sent on the multicast LSP "LSP-PE-1-PE-2-default" and the egress port on PE-1 is 1/1/1:

```
*A:PE-1# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                      1010          124684            1011          124850
===============================================================================
*A:PE-1# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                        14            1200              15            1310
===============================================================================
```

Re-enable MAC learning in VPLS 2.

```
*A:PE-1# configure service vpls 2 no disable-learning
```

Verify CBF in SDP

The CBF related information for SDP 122 in VPLS 2 on PE-1can be verified as follows:

```
*A:PE-1# show service id 2 sdp 122 detail
---snip---
-------------------------------------------------------------------------------
RSVP/Static LSPs
-------------------------------------------------------------------------------
Associated LSP List :
Lsp Name           : LSP-PE-1-PE-2-EF
Admin State        : Up                        Oper State        : Up
Time Since Last Tr*: 00h26m41s

Lsp Name           : LSP-PE-1-PE-2-default
Admin State        : Up                        Oper State        : Up
Time Since Last Tr*: 00h21m35s

-------------------------------------------------------------------------------
Class-based forwarding :
-------------------------------------------------------------------------------
Class forwarding   : Enabled                   EnforceDSTELspFc  : Disabled
Default LSP        : LSP-PE-1-PE-2-default     Multicast LSP     : LSP-PE-1-PE-*
===============================================================================
FC Mapping Table
===============================================================================
FC Name             LSP Name
-------------------------------------------------------------------------------
ef                  LSP-PE-1-PE-2-EF
===============================================================================
---snip---
```

The detailed information for the SDP contains all the LSPs configured in the SDP. CBF is enabled with default LSP "LSP-PE-1-PE-2-default". The multicast LSP is "LSP-PE-PE-2-default", but the name is truncated. The FC mapping table lists all FCs that have an LSP assigned. Each FC can only have one LSP assigned. Different FCs can have the same LSP assigned. In this example, only the FC EF has an LSP assigned.

Traffic statistics can be displayed per service per SDP. These statistics are not per LSP, so no distinction can be made between the different FCs.

```
*A:PE-1# show service id 2 sdp 122 detail
---snip---
Statistics            :
```

```
I. Fwd. Pkts.      : 9003                  I. Dro. Pkts.      : 0
I. Fwd. Octs.      : 882180                I. Dro. Octs.      : 0
E. Fwd. Pkts.      : 9001                  E. Fwd. Octets     : 882060
---snip---
```

Note: In this configuration example, the service using the SDP with CBF is a VPLS. The SDP can be used in a similar way by a VPRN.

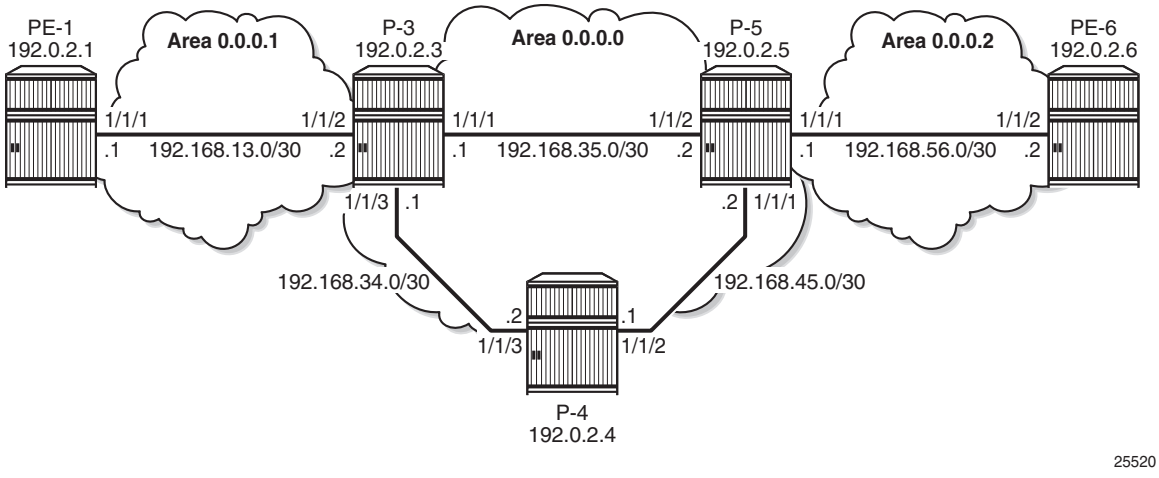
Note: An epipe service using CBF must have ingress shared-queuing enabled on access/ SAP to work correctly.

CBF of LDP Prefix Packets over IGP Shortcuts

For this implementation, a more complex test topology is needed. Different OSPF areas are used.

In this example, traffic will be sent from VPRN 3 in PE-1 in OSPF area 0.0.0.1 to VPRN 3 in PE-6 in OSPF area 0.0.0.2. Most nodes used in the previous use case are reused, but some reconfiguration is required (router interfaces, OSPF areas, LSPs). For simplicity, PE-2 is not used anymore. The test topology is shown in Figure 108.

Figure 108 **Test Topology for CBF of LDP Prefix Packets over IGP Shortcuts**



Initial Configuration

All nodes have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces
- IGP (here: OSPF) - PE-1 is now in OSPF area 0.0.0.1 instead of 0.0.0.0 in the preceding use case. P-3 and P-5 have interfaces in different areas:

```
*A:P-3# configure router
        ospf
            traffic-engineering
            area 0.0.0.0
                interface "system"
                exit
                interface "int-P-3-P-4"
                    interface-type point-to-point
                exit
                interface "int-P-3-P-5"
                    interface-type point-to-point
                exit
            exit
            area 0.0.0.1
                interface "int-P-3-PE-1"
                    interface-type point-to-point
                exit
            exit
            no shutdown
        exit
```

- MPLS enabled on all router interfaces
- RSVP enabled
- RSVP LSPs:
  - Between PE-1 and P-3 in area 0.0.0.1 and between PE-6 and P-5 in area 0.0.0.2.
  - Between P-3 and P-5 in area 0.0.0.0, there are two LSPs in each direction: one LSP uses the direct strict path and the other LSP uses an indirect strict path via P-4. On P-3, the LSP "LSP-P-3-P-4-P-5" with indirect path will be used as the default LSP, while the LSP "LSP-P-3-P-5" with the direct path will be used for traffic with FC EF. For simplicity, there are no LSPs configured to or from P-4.

```
*A:P-3# configure router
        mpls
            path "path-P-3-PE-1"
                hop 10 192.168.13.1 strict
                no shutdown
            exit
            path "path-P-3-P-5"
                hop 10 192.168.35.2 strict
                no shutdown
```

```
            exit
            path "path-P-3-P-4-P-5"
                hop 10 192.168.34.2 strict
                hop 20 192.168.45.2 strict
                no shutdown
            exit
            lsp "LSP-P-3-PE-1"
                to 192.0.2.1
                primary "path-P-3-PE-1"
                exit
                no shutdown
            exit
            lsp "LSP-P-3-P-5"
                to 192.0.2.5
                primary "path-P-3-P-5"
                exit
                no shutdown
            exit
            lsp "LSP-P-3-P-4-P-5"
                to 192.0.2.5
                primary "path-P-3-P-4-P-5"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

Configure IGP Shortcuts

IGP shortcut or forwarding adjacency must be enabled in one or more IGP instances. In the example, IGP shortcuts (RSVP shortcuts) are configured on all nodes:

- IGP shortcut:

```
*A:PE-1# configure router ospf rsvp-shortcut
```

- Forwarding adjacency (not configured in the example):

```
*A:PE-1# configure router ospf advertise-tunnel-link
```

By default, all LSPs are eligible for IGP shortcut. However, it is possible to exclude a specific RSVP LSP from being used in IGP shortcut as follows (this is not required in this example):

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-P-3" no igp-shortcut
```

For more information on IGP Shortcuts, see chapter “IGP Shortcuts”.

Configure ECMP

ECMP must be enabled in the global routing instance. Nokia recommends that ECMP is set to a value that is at least equal to the number of FCs used by the packets forwarded using the LDP over RSVP tunnel.

In this case, two traffic flows are distinguished: traffic with FC EF takes one LSP while all other traffic takes the default LSP. Here, ECMP should be at least equal to 2. For simplicity, ECMP equal to 2 is only configured on P-3 and P-5:

```
*A:P-3# configure router ecmp 2
```



Note: The selection of ECMP LSPs is done regardless of class-forwarding assignments, if any. The total number of LSPs with same cost, the number of those which have class-forwarding assignments, and the max-ecmp-routes parameter value influence the final set of LSPs, the consistency (from a CBF perspective) of which will be verified. These three factors must be carefully configured so as to ensure that the final set is consistent from a CBF perspective.

Enable LDP over RSVP

LDP over RSVP must be enabled between all PE routers and all P routers within the same area and between all P routers in the area 0:

```
*A:PE-1# configure router
        ldp
            targeted-session
                peer 192.0.2.3
                    tunneling
                    exit
                exit
            exit

*A:P-3# configure router
        ldp
            targeted-session
                peer 192.0.2.1
                    tunneling
                    exit
                exit
                peer 192.0.2.5
                    tunneling
                    exit
                exit
            exit
```

Note: The LSP names configured in the tunneling context (if any) are not directly used by LDP when the rsvp-shortcut option is enabled. With IGP shortcuts, the set of tunnel next-hops is always provided by IGP in the routing table manager (RTM). The class-based forwarding rules will not apply to these named LSPs unless they are populated by IGP in the RTM as next-hops for a prefix.

Note: The option prefer-tunnel-in-tunnel must be disabled (which is the default) for CBF to apply to LDP prefixes which are the endpoints of tunnels.

There is no need to enable LDP over RSVP on the LSPs. It is enabled by default, as can be verified with the following command:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2-EF" detail | match LdpOverRsvp
LdpOverRsvp : Enabled                          VprnAutoBind   : Enabled
```

For more information on LDP over RSVP, see chapter “LDP over RSVP Using OSPF as IGP”.

Configure CBF of LDP Prefix Packets over IGP Shortcuts

Enable CBF of LDP prefix packets over IGP shortcuts with the following command in the LDP context on P-3 and P-5:

```
*A:P-3# configure router ldp class-forwarding
```

In this example, the PEs only have one path to the P routers. There is no need to enable class-forwarding.

When CBF is enabled, LDP prefixes resolved to a set of ECMP tunnel next-hops will have their packets forwarded to the LSP configured to carry the forwarding class that the packet was classified to at the ingress SAP, access interface, or network interface.

In release 13.0.R6, this command is applicable for LSRs forwarding LDP forwarding equivalence class (FEC) prefix packets over a set of MPLS LSPs using IGP shortcuts. It is not supported for LER forwarding.

Note: The command configure router ldp class-forwarding applies to the following contexts:

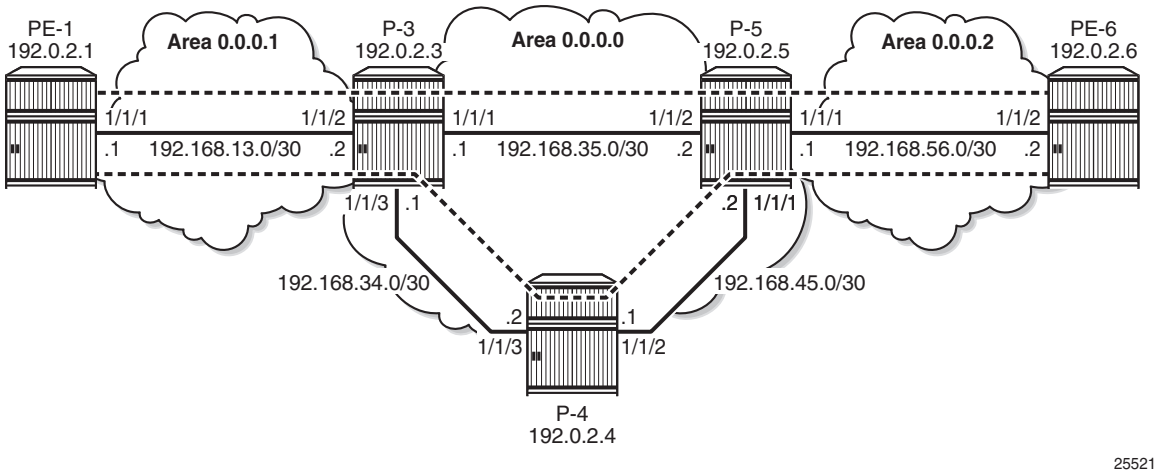
1. LER forwarding of packets of VPRN and L2 services that use auto-binding to LDP when the LDP FEC is resolved to a set of MPLS LSPs using IGP shortcuts. This is not supported in 13.0.R6.
2. LER forwarding of shortcut packets over LDP FEC that is resolved to a set of MPLS LSPs using IGP shortcuts. This is not supported in 13.0.R6.
3. LSR forwarding LDP FEC prefix packets over a set of MPLS LSPs using IGP shortcuts. This is supported in 13.0.R6.

This command does not apply to the following contexts:

- LER forwarding of VPRN and L2 services over a user-provisioned SDP of type LDP when the LDP FEC is resolved to a set of MPLS LSPs using IGP shortcuts.

In this use case, the direct link from P-3 to P-5 should only be used for traffic with FC EF. The default LSP will take the longer path from P-3 via P-4 to P-5. The two different paths are shown in Figure 109.

Figure 109 **Different Paths for Different FCs**



The LSPs can be configured as follows:

```
*A:P-3# configure router
        mpls
           lsp "LSP-P-3-P-5"
                class-forwarding
                    fc ef
                exit
            exit
            lsp "LSP-P-3-P-4-P-5"
                class-forwarding
```

```
                default-lsp
            exit
        exit
```

All traffic with an FC that is not mapped to a specific LSP will be mapped to the default LSP. The default LSP will also be used for traffic with FC EF, in case the dedicated LSP for that traffic is no longer available. It is possible to configure more than one default LSP, but only one will be used at a time.

The same LSP can be assigned to one or more FCs and be the default LSP at the same time.

```
*A:P-3# configure router
        mpls
            lsp "LSP-P-3-P-4-P-5"
            class-forwarding
                fc be
                default-lsp
            exit
        exit
```

The consistency of the configuration among the tunnel next-hops of an LDP FEC can only be verified by LDP at the time the FEC is resolved to IGP shortcuts. An example of an inconsistent configuration would be when class-forwarding is enabled while all tunnel next-hops for an LDP FEC have neither **fc** nor **default-lsp** assigned to them. LDP will then revert to ECMP routing for that FEC.

In this example, only P-3 and P-5 have ECMP equal to 2 and class-forwarding enabled.

Multiple LSPs can have the same FC assigned. Only one of these LSPs will be used to forward packets of this FC. That LSP is the one with the lowest tunnel ID.

Multiple LSPs can have the default-lsp configuration assigned, but only one of those will be the default LSP carrying all the traffic that should get the default treatment. That LSP is the one with the lowest tunnel ID.

If at least one LSP (amongst the ECMP set of LSPs) has an FC assigned, but no LSP has the default-lsp configuration, a single LSP will be automatically designated by LDP, even if all eight FCs have been mapped. A default LSP is needed to carry packets of an FC for which no explicit mapping to an LSP exists. The LSP with the lowest tunnel ID will be selected.

Note: If none of the LSPs has an FC or default LSP configuration, the set is inconsistent and no CBF happens.

When the active LDP bindings are displayed, FECs resolved with CBF can be recognized by the (C) after the prefix:

```
*A:P-3# show router ldp bindings active prefixes prefix 192.0.2.6/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                              Push              --      262136
192.0.2.5                                 LspId 2

192.0.2.6/32                              Push              --      262136
192.0.2.5                                 LspId 3

192.0.2.6/32(C)                           Swap            262133    262136
192.0.2.5                                 LspId 2

192.0.2.6/32(C)                           Swap            262133    262136
192.0.2.5                                 LspId 3

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:P-3#
```

Only the entries that correspond to swap operations get the (C) added. The entries that correspond to push operations do not get the (C) because CBF is not supported for LERs in 13.0.R6.

Traffic that is sent from PE-1 to PE-6 will be forwarded on P-3 based on the FC, while traffic that originates on P-3 will not be subjected to CBF.

LSP 2 uses the direct path from P-3 to P-5 while LSP 3 uses the indirect path from P-3 to P-5 via P-4. This can be verified in the tunnel table:

```
*A:P-3# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      rsvp      MPLS  1         7        192.168.13.1    16777215
192.0.2.1/32      ldp       MPLS  65537     9        192.0.2.1       65535
```

```
192.0.2.5/32      rsvp      MPLS  2         7        192.168.35.2   16777215
192.0.2.5/32      rsvp      MPLS  3         7        192.168.34.2   16777215
192.0.2.5/32      ldp       MPLS  65538     9        192.0.2.5      65535
192.0.2.6/32      ldp       MPLS  65539     9        192.0.2.5      131070
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:P-3#
```

The LSP that uses the direct path to P-5 has the lowest tunnel ID. When both LSPs are configured as default LSP, the LSP with the lowest tunnel ID will effectively be used as the default LSP. In the current configuration, only the LSP using the indirect path is configured as the default LSP.

Verify CBF

Traffic will be sent from a VPRN on PE-1 to a VPRN on PE-6 and back.

Configure VPRN

On PE-1 and PE-6, a VPRN service needs to be configured. BGP will be used to exchange VPN routes.

Import and export policies are configured as follows:

```
*A:PE-1# configure router
        policy-options
            begin
            community "VPN3" members "target:64496:3"
            policy-statement "VPN3-export"
                entry 10
                    from
                        protocol direct
                    exit
                    to
                        protocol bgp-vpn
                    exit
                    action accept
                        community add "VPN3"
                    exit
                exit
            exit
            policy-statement "VPN3-import"
                entry 10
                    from
                        protocol bgp-vpn
                        community "VPN3"
                    exit
                    action accept
```

```
                exit
            exit
        exit
        commit
    exit
```

The VPRN is configured using auto-bind-tunnel with resolution-filter ldp, as follows:

```
*A:PE-1# configure service
    vprn 3 customer 1 create
        vrf-import "VPN3-import"
        vrf-export "VPN3-export"
        route-distinguisher 64496:31
        auto-bind-tunnel
            resolution-filter
                ldp
            exit
            resolution filter
        exit
        interface "loopback3" create
            address 172.31.1.3/32
            loopback
        exit
        no shutdown
    exit
```

For BGP, P-5 is used as route reflector. The address family is VPN-IPv4.

```
*A:PE-1# configure router
    autonomous-system 64496
    bgp
        family vpn-ipv4
        group "internal"
            peer-as 64496
            neighbor 192.0.2.5
            exit
        exit
        no shutdown
    exit

*A:P-5# configure router
    bgp
    autonomous-system 64496
        family vpn-ipv4
        cluster 1.1.1.1
        group "internal"
            peer-as 64496
            neighbor 192.0.2.1
            exit
            neighbor 192.0.2.3
            exit
            neighbor 192.0.2.4
            exit
            neighbor 192.0.2.6
            exit
        exit
        no shutdown
```

```
        exit
```

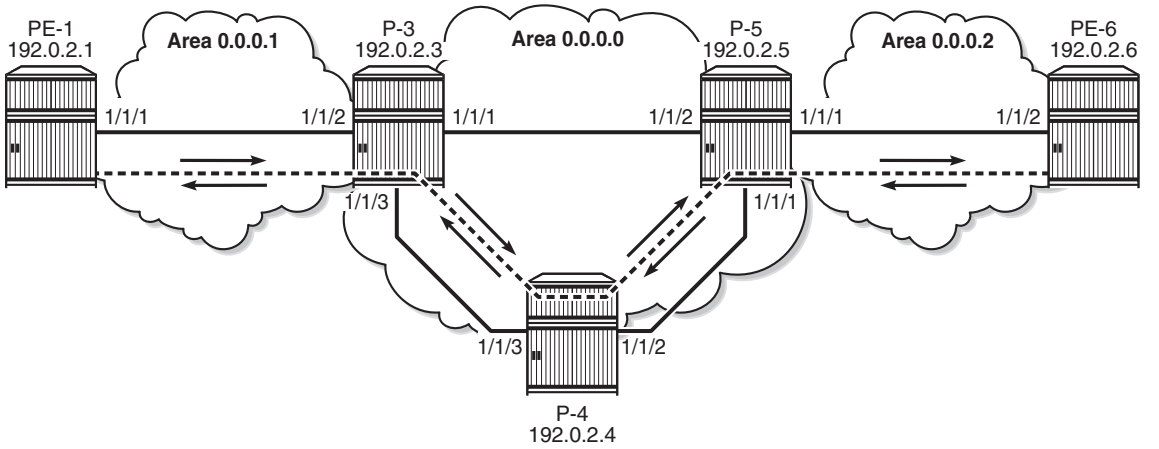
The routes are learned, as can be verified with the following command:

```
*A:PE-1# show router 3 route-table
===============================================================================
Route Table (Service: 3)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.31.1.3/32                                   Local   Local     00h21m08s   0
       loopback3                                                      0
172.31.6.3/32                                   Remote  BGP VPN   00h10m47s   170
       192.0.2.6 (tunneled)                                         0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

Verify Traffic Flow - Default LSP

When ping messages are sent from the VPRN on PE-1 to the VPRN on PE-6, the traffic will be forwarded according to the FC at the LSRs. The FC is not manipulated yet. For P-3, the traffic will be sent on the default LSP to P-5 via P-4. On node P-3, the ingress port for traffic destined to PE-6 is 1/1/2 and the egress port to P-4 is 1/1/3. The ping replies will follow the reverse path. The traffic flow is shown in Figure 110.

Figure 110 **Traffic on Default LSP**




```
*A:P-3# clear port 1/1/[1..3] statistics
*A:PE-1# ping router 3 172.31.6.3 rapid count 1000
*A:P-3# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                        25            2548              24            2322
===============================================================================
*A:P-3# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                     1029          116738            1030          116844
===============================================================================
*A:P-3# show port 1/1/3 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/3                     1037          118074            1040          118648
===============================================================================
*A:P-3#
```

With the same commands, the traffic flow on the other nodes can be verified. P-4 has traffic on port 1/1/3 (to P-3) and on port 1/1/2 (to P-5). P-5 has traffic on port 1/1/1 (to P-4) and port 1/1/3 (to PE-6).

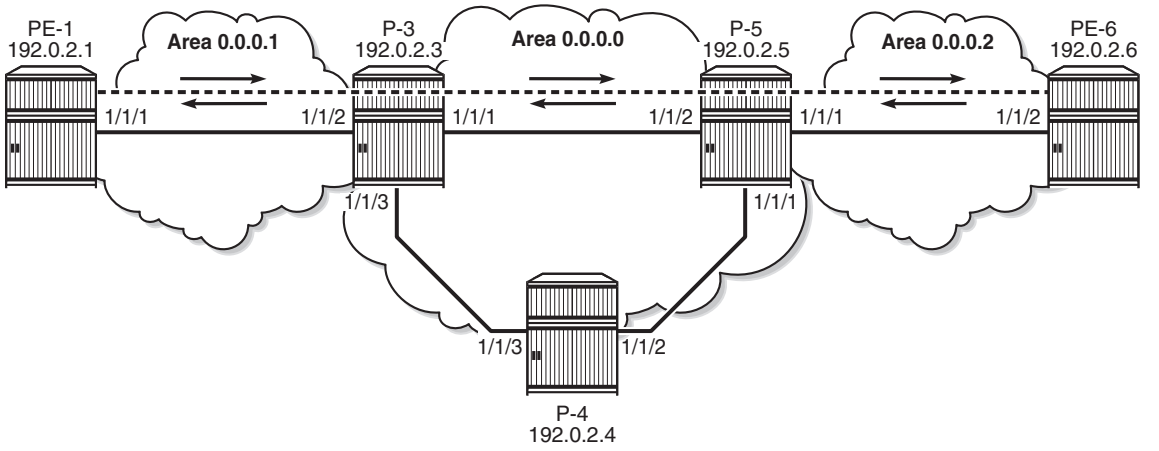
Verify Traffic Flow - LSP for FC EF

On nodes P-3 and P-5, the FC is set to EF at the ingress of the interface to the PE router.

```
*A:P-3# configure qos
        network 2 create
            ingress
                default-action fc ef profile in
            exit
        exit
*A:P-3# configure router interface "int-P-3-PE-1" qos 2
```

When the FC is modified to EF, the traffic takes the direct path between P-3 and P-5. On node P-3, traffic will be sent to and received from P-5 on port 1/1/1, as shown in Figure 111.

Figure 111 **Traffic with FC EF on Direct Path**




```
*A:P-3# clear port 1/1/[1..3] statistics
*A:PE-1# ping router 3 172.31.6.3 rapid count 1000
*A:P-3# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                      1015          115276            1015          115276
===============================================================================
*A:P-3# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
```

```
-------------------------------------------------------------------------------
1/1/2                      1024          116136            1024          116004
===============================================================================
*A:P-3# show port 1/1/3 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/3                        22            1834              23            2062
===============================================================================
*A:P-3#
```

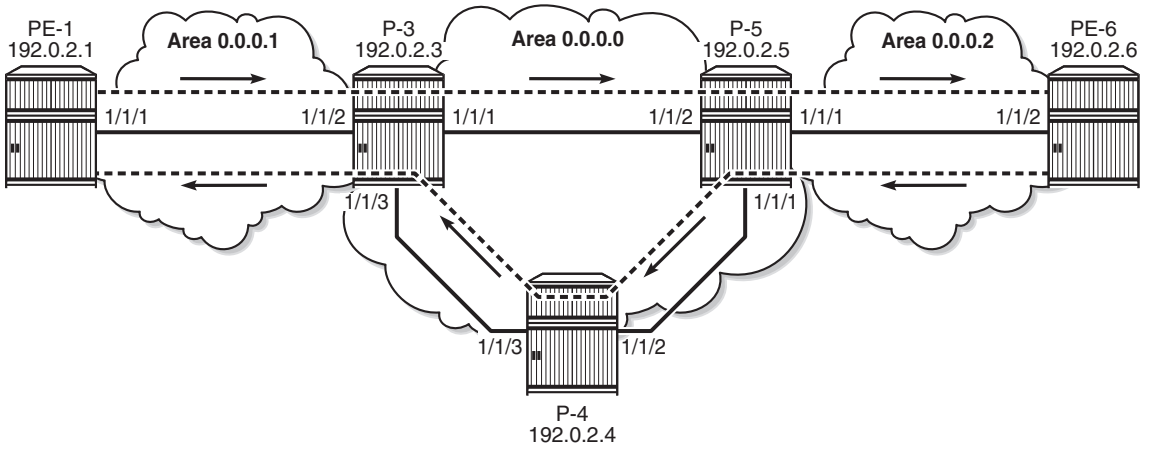
The original configuration is restored by removing the QoS network policy from the interface to the PE as follows:

```
*A:P-3# configure router interface "int-P-3-PE-1" no qos
```

Verify Traffic Flow - Multiple LSPs Configured as Default LSP on P-3

The FC is not manipulated anymore. On P-3, LSP "LSP-P-3-P-4-P-5" was configured as default LSP. Additionally, also LSP "LSP-P-3-P-5" is configured as a default LSP. So "LSP-P-3-P-5" becomes the used default LSP, because the tunnel ID (equal to 2) is lower than the tunnel ID of the LSP using the long path from P-3 to P-5 via P-4 (equal to 3). When this is configured on P-3, without changing anything to the configuration at P-5, the behavior is asymmetric. Incoming traffic on port 1/1/2 is forwarded to P-5 via port 1/1/1, while the ping replies still take the long path using ingress port 1/1/3 on P-3 The traffic flow is shown in Figure 112.

Figure 112 **Traffic on Default LSP with Lowest ID on P-3**




```
*A:P-3# configure router
        mpls
            lsp "LSP-P-3-P-5"
                class-forwarding
                    default-lsp
                exit
*A:P-3# clear port 1/1/[1..3] statistics
*A:PE-1# ping router 3 172.31.6.3 rapid count 1000
*A:P-3# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                        22            1904            1028          116819
===============================================================================
*A:P-3# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                       1029          116813            1025          116289
===============================================================================
*A:P-3# show port 1/1/3 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/3                       1031          117375              32            3520
===============================================================================
*A:P-3#
*A:P-3# show router tunnel-table
```

```
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      rsvp      MPLS  1         7        192.168.13.1    16777215
192.0.2.1/32      ldp       MPLS  65550     9        192.0.2.1       65535
192.0.2.5/32      rsvp      MPLS  2         7        192.168.35.2    16777215
192.0.2.5/32      rsvp      MPLS  3         7        192.168.34.2    16777215
192.0.2.5/32      ldp       MPLS  65553     9        192.0.2.5       65535
192.0.2.6/32      ldp       MPLS  65554     9        192.0.2.5       131070
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:P-3#
```

The original configuration is restored. The LSP with the direct path will only be used for traffic with FC EF ("LSP-P-3-P-5" remained an EF LSP even when it became a default LSP):

```
*A:P-3# configure router mpls lsp "LSP-P-3-P-5" class-forwarding no default-lsp
```

Verify Traffic Flow - No Default LSP Configured

When the class-forwarding configuration for the different LSPs always has one or more FCs and no default LSP, the system will select a default LSP itself. This can be tested by removing the default LSP configuration from all LSPs on P-3 and P-5 and adding FC AF instead. FC AF does not correspond to the FC of the ping messages; therefore, a default LSP will be required.

```
*A:P-3# configure router mpls lsp "LSP-P-3-P-4-P-5" class-forwarding no default-lsp
*A:P-3# configure router mpls lsp "LSP-P-3-P-4-P-5" class-forwarding fc af
*A:P-5# configure router mpls lsp "LSP-P-5-P-4-P-3" class-forwarding no default-lsp
*A:P-5# configure router mpls lsp "LSP-P-5-P-4-P-3" class-forwarding fc af
```

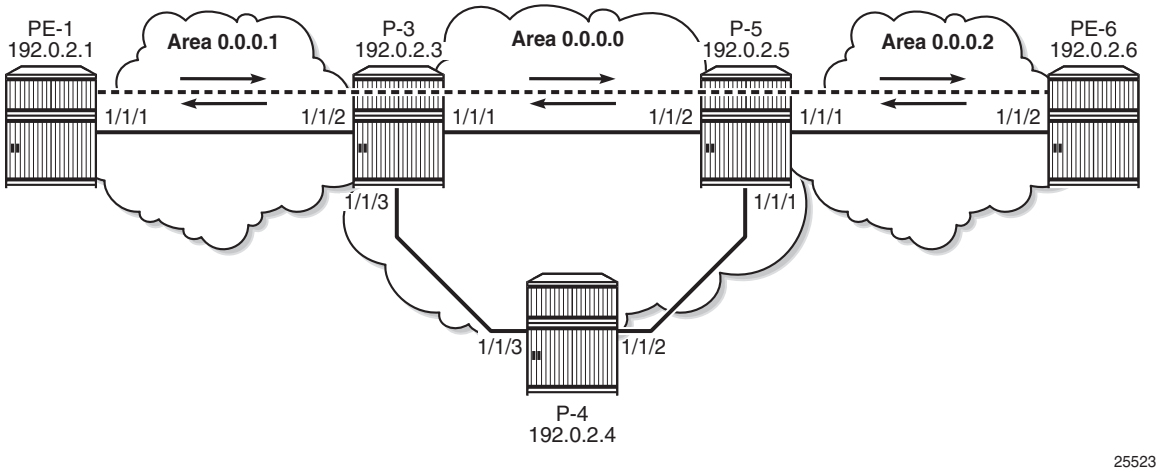
The actual CBF configuration does not include any default LSPs anymore:

```
*A:P-3>config>router>mpls# info
----------------------------------------------
---snip---
            lsp "LSP-P-3-P-5"
                to 192.0.2.5
                class-forwarding
                    fc ef
                exit
                primary "path-P-3-P-5"
                exit
                no shutdown
            exit
            lsp "LSP-P-3-P-4-P-5"
                to 192.0.2.5
```

```
            class-forwarding
                fc af
            exit
            primary "path-P-3-P-4-P-5"
            exit
            no shutdown
        exit
```

The system selects the LSP with the lowest tunnel ID as the default LSP, in this case the LSP using the direct path between P-3 and P-5, as shown in Figure 113.

Figure 113 **Traffic on System-Selected Default LSPs with Lowest Tunnel ID**



The tunnel ID can be verified in the tunnel table. The tunnel table for P-3 has already been shown in the preceding section.

On P-5, the LSP to P-3 with the tunnel ID 2 has next hop 192.168.35.1, as can be verified in the tunnel table. This corresponds to LSP "LSP-P-5-P-3". The LSP with tunnel ID 3 and next hop 192.168.45.1 corresponds to LSP "LSP-P-5-P-4-P-3".

```
*A:P-5# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65552     9        192.0.2.3       131070
192.0.2.3/32      rsvp      MPLS  2         7        192.168.35.1    16777215
192.0.2.3/32      rsvp      MPLS  3         7        192.168.45.1    16777215
192.0.2.3/32      ldp       MPLS  65551     9        192.0.2.3       65535
192.0.2.6/32      rsvp      MPLS  1         7        192.168.56.2    16777215
192.0.2.6/32      ldp       MPLS  65553     9        192.0.2.6       65535
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
```

```
===============================================================================
*A:P-5#
```

On P-3, incoming traffic from PE-1 will be forwarded on port 1/1/1 toward P-5.

```
*A:P-3# clear port 1/1/[1..3] statistics
*A:PE-1# ping router 3 172.31.6.3 rapid count 1000

*A:P-3# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                      1024          116120            1022         115840
===============================================================================
*A:P-3# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                      1026          116221            1028         116325
===============================================================================
*A:P-3# show port 1/1/3 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/3                        26            2142              25           2174
===============================================================================
*A:P-3#
```

No CBF in Case of Inconsistent Configuration

The FC LSP and default LSPs are removed from P-3 and P-5. Class-forwarding remains enabled, but the configuration is inconsistent and the forwarding behavior reverts to ECMP routing.

```
*A:P-3# configure router mpls lsp "LSP-P-3-P-4-P-5" class-forwarding no default-lsp
*A:P-3# configure router mpls lsp "LSP-P-3-P-4-P-5" class-forwarding no fc af
*A:P-3# configure router mpls lsp "LSP-P-3-P-5" class-forwarding no fc ef
*A:P-5# configure router mpls lsp "LSP-P-5-P-4-P-3" class-forwarding no default-lsp
*A:P-5# configure router mpls lsp "LSP-P-5-P-4-P-3" class-forwarding no fc af
*A:P-5# configure router mpls lsp "LSP-P-5-P-3" class-forwarding no fc ef
*A:P-3>config>router>mpls# info
----------------------------------------------
---snip---
            lsp "LSP-P-3-P-5"
                to 192.0.2.5
```

```
            class-forwarding
            exit
            primary "path-P-3-P-5"
            exit
            no shutdown
        exit
        lsp "LSP-P-3-P-4-P-5"
            to 192.0.2.5
            class-forwarding
            exit
            primary "path-P-3-P-4-P-5"
            exit
            no shutdown
        exit
        no shutdown
```

ECMP equals 2, so the traffic flows will be distributed over the two LSPs between P-3 and P-5, as can be verified in the routing tables at P-3 and P-5:

```
*A:P-3# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  OSPF      03h05m18s  10
       192.0.2.1 (tunneled:RSVP:1)                                  65535
192.0.2.3/32                                    Local   Local     04h56m20s  0
       system                                                       0
192.0.2.4/32                                    Remote  OSPF      04h40m38s  10
       192.168.34.2                                                 10
192.0.2.5/32                                    Remote  OSPF      00h00m06s  10
       192.0.2.5 (tunneled:RSVP:2)                                  65535
192.0.2.5/32                                    Remote  OSPF      00h00m06s  10
       192.0.2.5 (tunneled:RSVP:3)                                  65535
192.0.2.6/32                                    Remote  OSPF      00h00m06s  10
       192.0.2.5 (tunneled:RSVP:2)                                  131070
192.0.2.6/32                                    Remote  OSPF      00h00m06s  10
       192.0.2.5 (tunneled:RSVP:3)                                  131070
---snip---

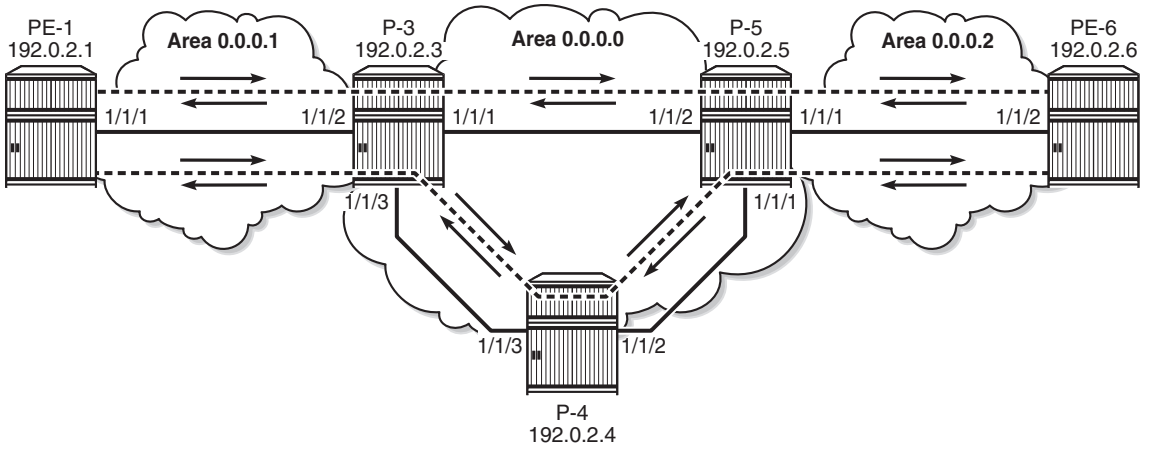
*A:P-5# show router route-table 192.0.2.1/32
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  OSPF      00h01m06s  10
       192.0.2.3 (tunneled:RSVP:2)                                  131070
192.0.2.1/32                                    Remote  OSPF      00h01m06s  10
       192.0.2.3 (tunneled:RSVP:3)                                  131070
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
```

```
       S = Sticky ECMP requested
===============================================================================
*A:P-5#
```

The ECMP hashing algorithm will hash on the source and destination IP address. As long as they are the same for each packet, the traffic will be sent on the same link. Therefore, different traffic flows need to be generated in parallel from PE-1 to PE-6. It is possible to create additional loopback interfaces in VPRN 3, but for this test, the system IP addresses of PE-1 and PE-6 can also be used. The ping messages are not sent after one another, but parallel in two different sessions with PE-1.

Figure 114 shows that different traffic flows are sprayed over the different LSPs. The returning traffic need not take the same path.

Figure 114 **Inconsistent CBF Configuration. Revert to ECMP Forwarding**




```
*A:P-3# clear port 1/1/[1..3] statistics
*A:PE-1# ping router 3 172.31.6.3 rapid count 2000
*A:PE-1# ping 192.0.2.6 rapid count 2000
*A:P-3# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                      2052          216350              54           4745
===============================================================================
*A:P-3# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                      4054          445019            4055          444877
```

```
===============================================================================
*A:P-3# show port 1/1/3 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/3                      2048          232178            4042         443622
===============================================================================
*A:P-3#

*A:P-3# clear port 1/1/[1..3] statistics
*A:PE-1# ping router 3 172.31.6.3 rapid count 2000
*A:PE-1# ping 192.0.2.6 rapid count 2000
*A:P-3# show port 1/1/1 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/1                      2052          216350              54           4745
===============================================================================
*A:P-3# show port 1/1/2 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/2                      4054          445019            4055         444877
===============================================================================
*A:P-3# show port 1/1/3 statistics
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                    Ingress         Ingress         Egress          Egress
Id                      Packets          Octets         Packets         Octets
-------------------------------------------------------------------------------
1/1/3                      2048          232178            4042         443622
===============================================================================
*A:P-3#
```

In this case, all traffic from PE-1 to PE-6 for both traffic flows takes the LSP "LSP-P-3-P-4-P-5". The returning traffic is distributed over the LSPs "LSP-P-5-P-3" and "LSP-P-5-P-4-P-3". ECMP works, but there are not enough traffic flows to show it more clearly.

Conclusion

Within a service, traffic with different FCs can be forwarded using different paths, depending on the requirements for latency and jitter. This is also the case for services traversing multiple domains.

CBF is a local decision and need not be enabled on all intermediate routers. This allows for better interoperability.

DiffServ Traffic Engineering

This chapter provides information about DiffServ traffic engineering.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R2.

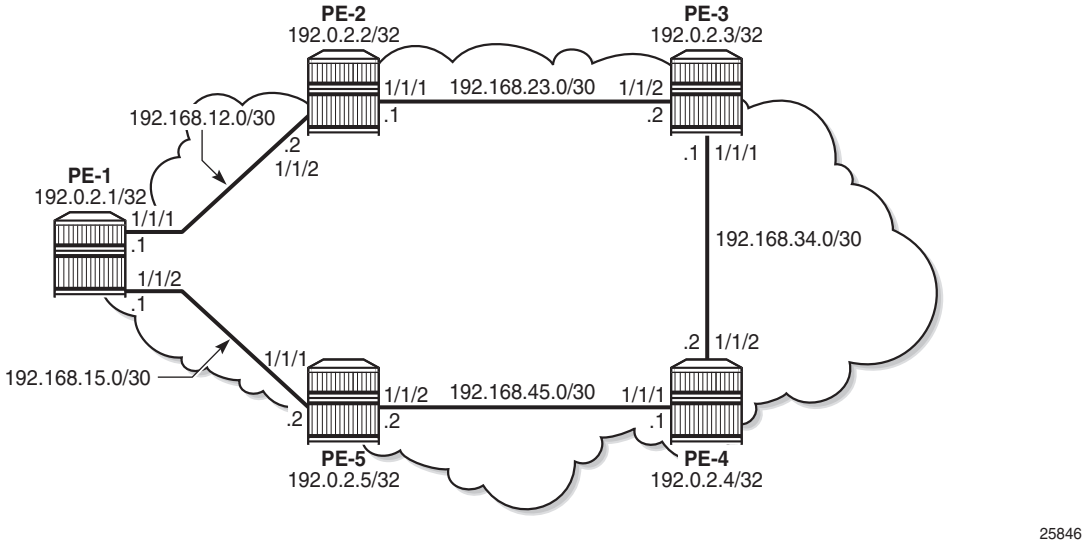
Overview

Differentiated services (DiffServ) is a mechanism to classify and manage network traffic to provide quality of service (QoS). DiffServ traffic engineering (DiffServ TE) reserves bandwidth for label switched paths (LSPs) on resource reservation protocol (RSVP) interfaces on a per TE class basis.

Example Topology

Figure 115 shows the example topology that contains five 7750 SRs in a ring topology.

Figure 115 **Example Topology**



Definitions

The following definitions are used in this chapter:

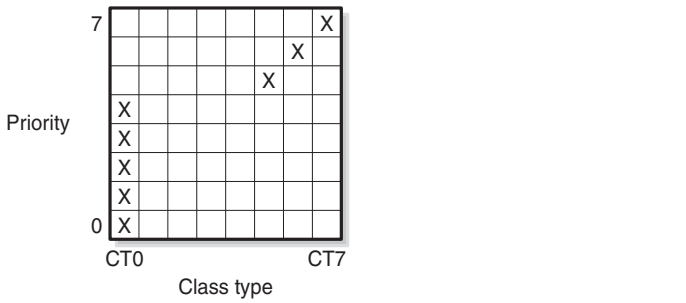
- Forwarding classes (FCs) classify micro-flows into macro-flows. FCs can be mapped to class types (CTs).
- A CT is macro-flow crossing a link governed by a specific bandwidth constraint (BC). The BC is defined on a per-link and per-CT basis. A CT can be considered as a network-wide FC, advertised by the IGP (OSPF opaque link state advertisement (LSA), IS-IS TE type length value (TLV)).
  - IGP-TE can reserve bandwidth per CT on a link (BC).
  - RSVP-TE can reserve bandwidth per LSP path, based on TE class.
- A TE class is a combination of a CT and a preemption priority.

There are eight FCs that can be mapped to CTs. The CTs range from CT0 (lowest) to CT7 (highest) and each gets a percentage of the bandwidth of the link. Each CT has eight different priority levels that are used for preemption. Even though there are 64 different potential combinations of CT and priority, only eight different combinations can be defined for TE classes. All CTs and priorities must be manually configured.

The system allows up to eight TE classes to be configured. The more TE classes are defined, the more RSVP LSPs need to be configured for each service. TE classes are consistently configured on all TE-aware label switching routers (LSRs) throughout the network and advertised through the IGP.

The following shows a DiffServ TE configuration where each CT can reserve up to 10% of the maximum reservable bandwidth meaning that 20% of the bandwidth is not allocated to any CT. MPLS needs to be shut down when DiffServ TE is configured. Figure 116 shows the mapping of the TE classes.

Figure 116 **Mapping of TE Classes**



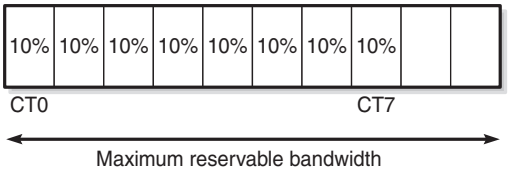

```
*A:PE-1# configure router
        mpls
            shutdown
        exit
        rsvp
            preemption-timer 5
            diffserv-te mam
                class-type-
bw ct0 10 ct1 10 ct2 10 ct3 10 ct4 10 ct5 10 ct6 10 ct7 10
                te-class 0 class-type 0 priority 0
                te-class 1 class-type 0 priority 1
                te-class 2 class-type 0 priority 2
                te-class 3 class-type 0 priority 3
                te-class 4 class-type 0 priority 4
                te-class 5 class-type 5 priority 5
                te-class 6 class-type 6 priority 6
                te-class 7 class-type 7 priority 7
                fc af class-type 0
                fc be class-type 0
                fc ef class-type 5
                fc h1 class-type 6
                fc h2 class-type 0
                fc l1 class-type 0
                fc l2 class-type 0
                fc nc class-type 7
            exit
        exit
        mpls
            no shutdown
```

```
        exit
```

This configuration is applied on all TE-aware LSRs for consistency.

Figure 117 shows the bandwidth reservation for the CTs.

Figure 117 **Bandwidth Reservation for the CTs**



Eight TE classes are defined, each with a different priority, using four CTs (0, 5, 6 and 7). There is no need to assign any bandwidth to the unused CTs. The configuration in this example is not a recommendation. In this example, the BC for each CT is 10% of the maximum reservable bandwidth.

On each node, the RSVP status shows whether DiffServ TE is enabled and how it is configured, as follows:

```
*A:PE-1# show router rsvp status
===============================================================================
RSVP Status
===============================================================================
Admin Status       : Up                  Oper Status        : Up
Keep Multiplier    : 3                   Refresh Time       : 30 sec
Message Pacing     : Disabled            Pacing Period      : 100 msec
Max Packet Burst   : 650 msgs            Refresh Bypass     : Disabled
Rapid Retransmit   : 5 hmsec             Rapid Retry Limit  : 3
Graceful Shutdown  : Disabled            SoftPreemptionTimer: 5 sec
GR Max Recovery    : 300 sec             GR Max Restart     : 120 sec
Implicit Null Label: Disabled            Node-id in RRO     : Exclude
P2P Merge Point Ab*: Disabled            P2MP Merge Point A*: Disabled
DiffServTE AdmModel: Mam                 Entropy Label      : Disabled
Percent Link Bw CT0: 10                  Percent Link Bw CT4: 10
Percent Link Bw CT1: 10                  Percent Link Bw CT5: 10
Percent Link Bw CT2: 10                  Percent Link Bw CT6: 10
Percent Link Bw CT3: 10                  Percent Link Bw CT7: 10
TE0 -> Class Type  : 0                   Priority           : 0
TE1 -> Class Type  : 0                   Priority           : 1
TE2 -> Class Type  : 0                   Priority           : 2
TE3 -> Class Type  : 0                   Priority           : 3
TE4 -> Class Type  : 0                   Priority           : 4
TE5 -> Class Type  : 5                   Priority           : 5
TE6 -> Class Type  : 6                   Priority           : 6
TE7 -> Class Type  : 7                   Priority           : 7
FCName             : af                  Class Type         : 0
FCName             : be                  Class Type         : 0
FCName             : ef                  Class Type         : 5
```

```
FCName             : h1                 Class Type         : 6
FCName             : h2                 Class Type         : 0
FCName             : l1                 Class Type         : 0
FCName             : l2                 Class Type         : 0
FCName             : nc                 Class Type         : 7
IgpThresholdUpdate : Disabled
Up Thresholds(%)   : 0 15 30 45 60 75 80 85 90 95 96 97 98 99 100
Down Thresholds(%) : 100 99 98 97 96 95 90 85 80 75 60 45 30 15 0
Update Timer       : N/A
Update on CAC Fail : Disabled
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The OSPF LSAs will contain BC information, as shown in the following output taken from PE-1:

```
*A:PE-1# show router ospf opaque-database adv-router 192.0.2.3 detail
===============================================================================
Rtr Base OSPFv2 Instance 0 Opaque Link State Database (type: All) (detail)
===============================================================================
-------------------------------------------------------------------------------
Opaque LSA
-------------------------------------------------------------------------------
Area Id          : 0.0.0.0              Adv Router Id    : 192.0.2.3
Link State Id    : 1.0.0.1              LSA Type         : Area Opaque
Sequence No      : 0x80000006           Checksum         : 0x922c
Age              : 1993                 Length           : 28
Options          :  E
Advertisement    : Traffic Engineering
     ROUTER-ID TLV  (0001) Len   4 : 192.0.2.3
-------------------------------------------------------------------------------
Opaque LSA
-------------------------------------------------------------------------------
Area Id          : 0.0.0.0              Adv Router Id    : 192.0.2.3
Link State Id    : 1.0.0.3              LSA Type         : Area Opaque
Sequence No      : 0x80000001           Checksum         : 0xf162
Age              : 281                  Length           : 164
Options          :  E
Advertisement    : Traffic Engineering
     LINK INFO TLV  (0002) Len 140 :
       Sub-TLV: 1     Len: 1     LINK_TYPE    : 1
       Sub-TLV: 2     Len: 4     LINK_ID      : 192.0.2.2
       Sub-TLV: 3     Len: 4     LOC_IP_ADDR  : 192.168.23.2
       Sub-TLV: 4     Len: 4     REM_IP_ADDR  : 192.168.23.1
       Sub-TLV: 5     Len: 4     TE_METRIC    : 10
       Sub-TLV: 6     Len: 4     MAX_BDWTH    : 10000000 Kbps
       Sub-TLV: 7     Len: 4     RSRVBL_BDWTH : 10000000 Kbps
       Sub-TLV: 8     Len: 32    UNRSRVD_CLS0 :
         P0: 1000000 Kbps P1: 1000000 Kbps P2: 1000000 Kbps P3: 1000000 Kbps
         P4: 1000000 Kbps P5: 1000000 Kbps P6: 1000000 Kbps P7: 1000000 Kbps
       Sub-TLV: 9     Len: 4     ADMIN_GROUP  : 0 None
       Sub-TLV: 17    Len: 36    TELK_BW_CONST:
         BW Model : MAM
         BC0: 1000000 Kbps BC1: 1000000 Kbps BC2: 1000000 Kbps BC3: 1000000 Kbps
         BC4: 1000000 Kbps BC5: 1000000 Kbps BC6: 1000000 Kbps BC7: 1000000 Kbps
-------------------------------------------------------------------------------
---snip---
```

In the preceding output, only the output for the interface between PE-2 and PE-3 is shown; the output is similar for the other interfaces. On each interface between nodes, there will be eight BCs for the eight CTs: from BC0 to BC7. In this example, each of the BCs has the same constraint of 1 Gb/s, which corresponds to 10% of the 10 Gb/s interfaces. As long as no LSP is configured with a CT and a bandwidth, no bandwidth will be reserved. The BCs for an interface, such as the interface between PE-1 and PE-2, can be shown as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
Interface         : int-PE-1-PE-2
Port ID           : 1/1/1
Admin State       : Up                  Oper State        : Up
Active Sessions   : 0                   Active Resvs      : 0
Total Sessions    : 0
Subscription      : 100 %               Port Speed        : 10000 Mbps
Total BW          : 10000 Mbps          Aggregate         : Dsabl
Hello Interval    : 3000 ms             Hello Timeouts    : 0
Key Type Auth     : Disabled
Keychain Auth     : Disabled
Auth Rx Seq Num   : n/a                 Auth Key Id       : n/a
Auth Tx Seq Num   : n/a                 Auth Win Size     : n/a
Refresh Reduc.    : Disabled            Reliable Deli.    : Disabled
Bfd Enabled       : No                  Graceful Shut.    : Disabled
ImplicitNullLabel : Disabled*           GR helper         : Disabled

Percent Link Bandwidth for Class Types*
Link Bw CT0       : 10                  Link Bw CT4       : 10
Link Bw CT1       : 10                  Link Bw CT5       : 10
Link Bw CT2       : 10                  Link Bw CT6       : 10
Link Bw CT3       : 10                  Link Bw CT7       : 10

Bandwidth Constraints for Class Types (Kbps)
BC0               : 1000000             BC4               : 1000000
BC1               : 1000000             BC5               : 1000000
BC2               : 1000000             BC6               : 1000000
BC3               : 1000000             BC7               : 1000000

Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE1->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE2->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE3->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE4->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE5->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE6->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE7->  Resv. Bw   : 0                   Unresv. Bw        : 1000000

IGP Update
Up Thresholds(%)   : 0 15 30 45 60 75 80 85 90 95 96 97 98 99 100  *
Down Thresholds(%) : 100 99 98 97 96 95 90 85 80 75 60 45 30 15 0  *
IGP Update Pending : No
```

```
Next Update        : N/A
No Neighbors.
* indicates inherited values
===============================================================================
*A:PE-1#
```

In this example, all BCs for the CTs are equal to 1 Gb/s, which is 10% of the maximum reservable bandwidth of 10 Gb/s. Currently no bandwidth is reserved for any of the TE classes. The unreserved bandwidth equals 1 Gb/s for TE0 through TE7.

The maximum bandwidth that can be allocated depends on the bandwidth of the link and the subscription percentage. When the subscription percentage is doubled to 200%, the BCs will be doubled too, as follows:

```
*A:PE-1# configure router rsvp interface "int-PE-1-PE-2" subscription 200
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0       : 10                  Link Bw CT4       : 10
Link Bw CT1       : 10                  Link Bw CT5       : 10
Link Bw CT2       : 10                  Link Bw CT6       : 10
Link Bw CT3       : 10                  Link Bw CT7       : 10

Bandwidth Constraints for Class Types (Kbps)
BC0               : 2000000             BC4               : 2000000
BC1               : 2000000             BC5               : 2000000
BC2               : 2000000             BC6               : 2000000
BC3               : 2000000             BC7               : 2000000

Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE1->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE2->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE3->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE4->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE5->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE6->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
TE7->  Resv. Bw   : 0                   Unresv. Bw        : 2000000
---snip---
```

The subscription percentage is restored to its default value of 100% as follows:

```
*A:PE-1# configure router rsvp interface "int-PE-1-PE-2" no subscription
```

Bandwidth Constraint Models

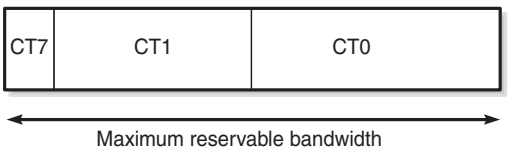
Two models are available for the bandwidth calculation that is required during the LSP setup: the maximum allocation model (MAM) and the Russian doll model (RDM). Table 13 shows a comparison between the two models.

Table 13 **Comparison Bandwidth Constraint Models**

Maximum allocation model (MAM)	Russian doll model (RDM)
Fixed BC per CT. No bandwidth sharing between CTs.	Maps one BC to one or more CTs. Lower CTs are allowed to reserve from the unused bandwidth of the pools defined for higher CTs.
Achieves isolation between CTs and guaranteed bandwidth to CTs without the need for preemption.	No isolation between CTs. Requires preemption to guarantee bandwidth to CTs other than the premium.
Bandwidth may be wasted.	Efficient use of bandwidth.
Easy to manage.	More complex.

Figure 118 shows the reserved bandwidth for the different class types according to the MAM model. In this example, there is only bandwidth reserved for CT0, CT1, and CT2.

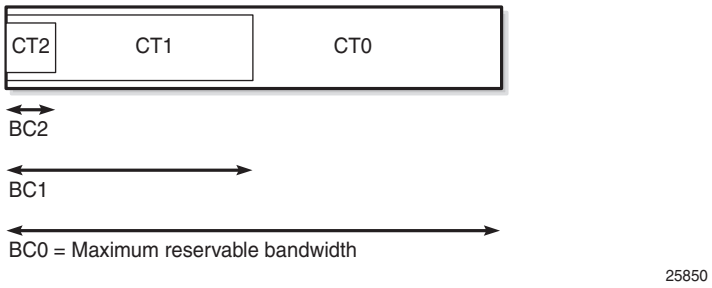
Figure 118 **Bandwidth Reservation in Maximum Allocation Model for Three CTs**



Bandwidth that is reserved for a specific CT cannot be used by any other CT. Therefore, bandwidth may be wasted.

The Russian doll model is more flexible: when CT1 has some spare bandwidth that might be used by CT0, this is allowed. Depending on the configured setup priority and hold priority, this may be reversed when CT1 requires the bandwidth. The bandwidth reservation in the Russian doll model is shown in Figure 119.

Figure 119 **Bandwidth Reservation in Russian Doll Model for Three CTs**



Backup Class Types

The main CT is defined at LSP level or primary path level. The main CT is used at the first attempt for the initial establishment and re-signal make-before-break (MBB) of the LSP primary path. Re-signaling of the LSP path can be triggered manually or timer-based. Subsequent retries use the backup CT, which is configured on the primary path level. Secondary paths are always signaled with the main CT. There is no verification whether the backup CT is lower than the main CT. This applies to CSPF and non-CSPF LSPs. An example of an LSP with main CT1 and backup CT0 is as follows:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3-withBackupCT"
                to 192.0.2.3
                cspf
                class-type 1
                primary "dyn"
                    bandwidth 50
                    priority 4 4
                    backup-class-type 0
                exit
                no shutdown
            exit
```

Possible triggers for using the backup CT are:

- Local interface failure or control plane failure (hello timeout)

- Received Resv message with the local-protection-in-use flag set (global revertive trigger)
- Received Patherr message with fast reroute (FRR) protection active notification (global revertive trigger)
- Received Patherr message with error code 34 (Reroute) and value 1 (soft preemption trigger)
- Received Patherr message with Preemption pending flag set (soft preemption trigger)
- Received ResvTear message

When the reservable bandwidth for a CT (including the bandwidth for the inner dolls in case of RDM) is insufficient, this does not trigger the backup CT to be used. If possible, an alternate path will be used for the LSP requiring this bandwidth.

Priorities

Two different priorities are linked to an LSP in a range from 0 to 7, where 0 is the highest priority and 7 the lowest. These values are important when preemption occurs, as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" primary "dyn" priority
  - no priority
  - priority <setup-priority> <hold-priority>
 <setup-priority>     : [0..7]
 <hold-priority>      : [0..7]
```

The following shows an LSP with both priorities equal to 4:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
                cspf
                primary "dyn"
                    bandwidth 50
                    priority 4 4
                exit
                no shutdown
            exit
```

- The first priority in the configuration is the setup priority. When an LSP is signaled and there is not enough bandwidth available on the Ethernet label edge router (eLER) or LSR, the LSP can preempt an established LSP with a hold priority lower than this setup priority. For a setup priority of 4, existing LSPs with a hold priority of 5, 6, or 7 can be preempted in case of insufficient bandwidth.

- The second priority in the configuration is the hold priority. When this LSP is established and a new LSP needs to be established, and there is insufficient bandwidth, this LSP can only be preempted by an LSP with a higher setup priority than this hold priority. For a hold priority of 4, the LSP can be preempted by any LSP with a setup priority of 0, 1, 2, or 3.

The default values are a setup priority of 7 and a hold priority of 0. A low setup priority of 7 means the LSP cannot preempt any LSP. A high hold priority of 0 implies that the LSP cannot be preempted by any other LSP.

The setup priority needs to be lower than or equal to the hold priority to avoid preemption loops. Nokia recommends that the setup priority and the hold priority are set to equal values.

Bandwidth, CT information, and priorities are shown as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 61442
From              : 192.0.2.1             To                   : 192.0.2.3
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262142
---snip---
Neg MTU           : 1564                  Oper MTU             : 1564
Bandwidth         : 50 Mbps               Oper Bandwidth       : 50 Mbps
Hop Limit         : 255                   Oper HopLimit        : 255
Record Route      : Record                Oper Record Route    : Record
Record Label      : Record                Oper Record Label    : Record
Setup Priority    : 4                     Oper Setup Priority  : 4
Hold Priority     : 4                     Oper Hold Priority   : 4
Class Type        : 0                     Oper CT              : 0
Backup CT         : None
MainCT Retry      : n/a
    Rem           :
MainCT Retry      : 0
    Limit         :
---snip---
```

When the LSP is being established, the path message contains the setup and hold priorities, and the required bandwidth, as follows:

```
*A:PE-1# debug router rsvp packet path detail
1 2016/04/13 13:17:10.56 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.3
          TTL:255, Checksum:0x9f8d, Flags:0x0
Session    - EndPt:192.0.2.3, TunnId:4, ExtTunnId:192.0.2.1
SessAttr   - Name:LSP-PE-1-PE-3::dyn
             SetupPri:4, HoldPri:4, Flags:0x46
RSVPHop    - Ctype:1, Addr:192.168.12.1, LIH:2
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.1, LspId:61442
SendTSpec  - Ctype:QOS, CDR:50.000 Mbps, PBS:50.000 Mbps, PDR:infinity
             MPU:20, MTU:1564
LabelReq   - IfType:General, L3ProtID:2048
RRO        - IpAddr:192.168.12.1, Flags:0x0
ERO        - IPv4Prefix 192.168.12.2/32, Strict
             IPv4Prefix 192.168.23.2/32, Strict
"
```

As soon as the LSP is established, the bandwidth is reserved on the interface int-PE-1-PE-2 on PE-1 in TE class 4 (configured as a combination of CT0 and priority 4), as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
Interface         : int-PE-1-PE-2
Port ID           : 1/1/1
Admin State       : Up                  Oper State        : Up
Active Sessions   : 1                   Active Resvs      : 1
Total Sessions    : 1
Subscription      : 100 %               Port Speed        : 10000 Mbps
Total BW          : 10000 Mbps          Aggregate         : Dsabl
---snip---

Percent Link Bandwidth for Class Types*
Link Bw CT0       : 10                  Link Bw CT4       : 10
Link Bw CT1       : 10                  Link Bw CT5       : 10
Link Bw CT2       : 10                  Link Bw CT6       : 10
Link Bw CT3       : 10                  Link Bw CT7       : 10

Bandwidth Constraints for Class Types (Kbps)
BC0               : 1000000             BC4               : 1000000
BC1               : 1000000             BC5               : 1000000
BC2               : 1000000             BC6               : 1000000
BC3               : 1000000             BC7               : 1000000

Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE1->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
```

```
TE2->  Resv. Bw   : 0                  Unresv. Bw        : 1000000
TE3->  Resv. Bw   : 0                  Unresv. Bw        : 1000000
TE4->  Resv. Bw   : 50000              Unresv. Bw        : 950000
TE5->  Resv. Bw   : 0                  Unresv. Bw        : 1000000
TE6->  Resv. Bw   : 0                  Unresv. Bw        : 1000000
TE7->  Resv. Bw   : 0                  Unresv. Bw        : 1000000
---snip---
```

Configuration

The setup consists of five 7750 SRs in a ring topology, as shown in Figure 115.

Initial Configuration

The nodes have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces. For PE-1:

```
*A:PE-1# configure router
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/1
        exit
        interface "int-PE-1-PE-5"
            address 192.168.15.1/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.1/32
        exit
```

- IGP: OSPF (alternatively, IS-IS could have been used) with TE enabled. For PE-1:

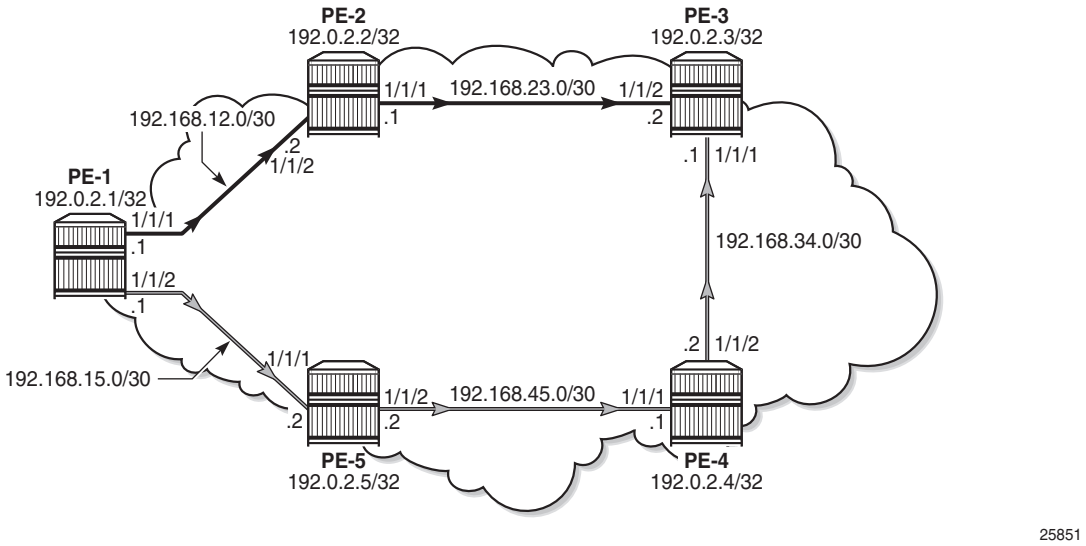
```
*A:PE-1# configure router
       ospf
            traffic-engineering
            area 0.0.0.0
                interface "system"
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-1-PE-5"
                    interface-type point-to-point
                exit
            exit
        exit
```

- MPLS and RSVP enabled on all interfaces. For PE-1:

```
*A:PE-1# configure router
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-5"
            exit
            no shutdown
        exit
        rsvp
            no shutdown
        exit
```

LSPs will be established from PE-1 to PE-3 with the short path via PE-2 as preferred. If insufficient bandwidth is available on the short path via PE-2, the longer path via PE-5 and PE-4 is taken, as shown in Figure 120.

Figure 120 **Paths from PE-1 to PE-3**



Initially, the default BC model, which is MAM, is enabled. Different LSPs will be created with different class type and priority. When an LSP is established, the bandwidth reservation is verified on the interfaces of PE-1. The same LSPs will later be established for the second BC model, RDM, where the bandwidth reservation is more efficient. For simplicity, FRR is not enabled on the LSPs.

Maximum Allocation Model

Enable DiffServ MAM

DiffServ TE can only be configured when MPLS is shutdown. This is something to consider during migration. An error is raised when MPLS is not shutdown, as follows:

```
*A:PE-1# configure router rsvp diffserv-te mam
MINOR: RSVP #1005 Invalid operation for RSVP instance -
 Diffserv can be enabled only when MPLS is shutdown
*A:PE-1#
```

The DiffServ TE configuration must be consistent on all the nodes in the setup. In this example, DiffServ TE is configured as follows:

```
*A:PE-1# configure router
        mpls
            shutdown
        exit
        rsvp
            diffserv-te mam
                class-type-bw ct0 50 ct1 40 ct2 0 ct3 0 ct4 0 ct5 0 ct6 0 ct7 10
                te-class 0 class-type 0 priority 7
                te-class 1 class-type 0 priority 4
                te-class 2 class-type 1 priority 7
                te-class 3 class-type 1 priority 4
                te-class 4 class-type 2 priority 7
                te-class 5 class-type 2 priority 2
                fc af class-type 1
                fc be class-type 0
                fc nc class-type 2
            exit
        exit
        mpls
            no shutdown
        exit
```

The bandwidth percentage for each CT must be configured. For unused CTs, the bandwidth percentage must be set to 0. An error is raised if unused CTs are missing , as follows:

```
*A:PE-1# configure router rsvp diffserv-te class-type-bw ct0 50 ct1 50
                                                                 ^
Error: Missing parameter
```

The sum of bandwidth percentages can be lower than, but must not exceed 100%, as follows:

```
*A:PE-1# configure router rsvp diffserv-te class-type-
bw ct0 50 ct1 50 ct2 0 ct3 0 ct4 0 ct5 0 ct6 0 ct7 10
```

```
MINOR: RSVP #1005 Invalid operation for RSVP instance -
 Total CT percent (110) exceeds 100
```

Fewer than eight classes can be configured, as in this example.

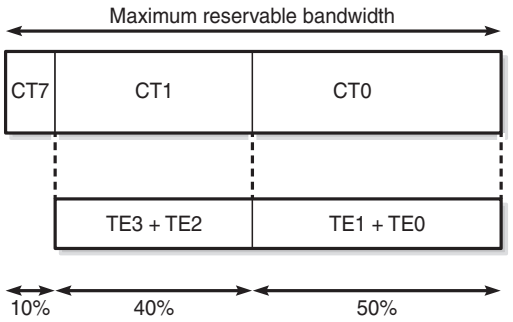
In the example, only three CTs are used by the TE classes: CT0, CT1, and CT2. However, 10% of the maximum reservable bandwidth is allocated to CT7. Because the MAM model does not allow bandwidth allocated to a CT to be used by other CTs, only 90% of the bandwidth can be reserved: 50% to be divided between TE0 and TE1, and 40% to be divided between TE2 and TE3. TE4 and TE5 do not have any bandwidth allocated. The bandwidth allocated to CT7 is completely wasted. This is just an example, not a recommendation.

The same settings will be repeated in the RDM model, where the bandwidth will not be wasted. The following bandwidth information can be seen on any interface:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0       : 50                  Link Bw CT4       : 0
Link Bw CT1       : 40                  Link Bw CT5       : 0
Link Bw CT2       : 0                   Link Bw CT6       : 0
Link Bw CT3       : 0                   Link Bw CT7       : 10
Bandwidth Constraints for Class Types (Kbps)
BC0               : 5000000             BC4               : 0
BC1               : 4000000             BC5               : 0
BC2               : 0                   BC6               : 0
BC3               : 0                   BC7               : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                   Unresv. Bw        : 5000000
TE1->  Resv. Bw   : 0                   Unresv. Bw        : 5000000
TE2->  Resv. Bw   : 0                   Unresv. Bw        : 4000000
TE3->  Resv. Bw   : 0                   Unresv. Bw        : 4000000
TE4->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE5->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                   Unresv. Bw        : 0
---snip---
```

Figure 121 shows the bandwidth allocation for the CTs and TE classes.

Figure 121 **MAM Bandwidth Allocation**



Establishing LSPs

TE class 5 corresponds to CT2 and priority 2. No bandwidth can be reserved for an RSVP LSP with CT2, setup priority 2, and hold priority 2. This can be verified by configuring an empty path and an LSP, as follows:

```
*A:PE-1# configure router
        mpls
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3-TE5"
                to 192.0.2.3
                cspf
                class-type 2
                primary "dyn"
                    bandwidth 1000
                    priority 2 2
                exit
                no shutdown
            exit
```

CSPF must be enabled. The class type is by default CT0, but can be changed to CT2 by configuration. The class type can be configured in the LSP context, as shown here, or in the primary path context. The setup priority and hold priority are configured in the primary path context.

The LSP cannot be established, because no bandwidth is allocated to TE class 5, as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE5" path detail | match "Failure Code"
Failure Code     : noCspfRouteToDestination
```

No bandwidth is reserved on the RSVP interfaces, as follows:

```
*A:PE-1# show router rsvp interface
===============================================================================
RSVP Interfaces
===============================================================================
Interface                       Total    Active    Total BW  Resv BW   Adm Opr
                                Sessions Sessions  (Mbps)    (Mbps)
-------------------------------------------------------------------------------
system                          -        -         -         -         Up  Up
int-PE-1-PE-2                   0        0         10000     0         Up  Up
int-PE-1-PE-5                   0        0         10000     0         Up  Up
-------------------------------------------------------------------------------
Interfaces : 3
===============================================================================
*A:PE-1#
```

Bandwidth can be reserved for an LSP with CT1 and priorities 4, as for the following LSP:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3-TE3"
                to 192.0.2.3
                cspf
                primary "dyn"
                    bandwidth 2000
                    priority 4 4
                    class-type 1
                exit
                no shutdown
            exit
```

In the example, the CT is configured in the primary path context while the CT in the previous example was configured in the LSP context.

The path is set up via PE-2 to PE-3, as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3-TE3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3-TE3 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE3
Path LSP ID      : 28160
From             : 192.0.2.1            To                  : 192.0.2.3
Admin State      : Up                   Oper State          : Up
Path Name        : dyn                  Path Type           : Primary
Path Admin       : Up                   Path Oper           : Up
```

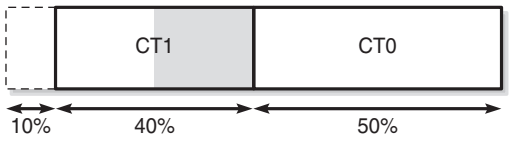
```
Out Interface    : 1/1/1                Out Label            : 262143
---snip---
Actual Hops      :
    192.168.12.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262143
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262143
---snip---
```

Bandwidth is reserved in TE class 3, as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
---snip---
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                    Unresv. Bw        : 5000000
TE1->  Resv. Bw   : 0                    Unresv. Bw        : 5000000
TE2->  Resv. Bw   : 0                    Unresv. Bw        : 2000000
TE3->  Resv. Bw   : 2000000              Unresv. Bw        : 2000000
TE4->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE5->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                    Unresv. Bw        : 0
---snip---
```

Figure 122 shows the bandwidth reservation for CT1 on interface int-PE-1-PE-2 on PE-1 and on interface int-PE-2-PE-3 on PE-2.

Figure 122 **Reserved and Unreserved Bandwidth**



An additional LSP is configured with CT1 and priority 4 and with CT0 as backup CT. The backup CT will not be used when the amount of unreserved bandwidth for CT1 is insufficient, as in the following case where int-PE-1-PE-2 and int-PE-1-PE-5 have insufficient unreserved bandwidth for CT1:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3-TE3-backupTE1"
                to 192.0.2.3
                cspf
                primary "dyn"
                    bandwidth 5000
                    priority 4 4
                    class-type 1
                    backup-class-type 0
```

```
            exit
            no shutdown
        exit
```

The LSP will not come up, as follows:

```
*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                            To              Tun     Fastfail  Adm  Opr
                                                    Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-PE-3-TE5                   192.0.2.3       3       No        Up   Dwn
LSP-PE-1-PE-3-TE3                   192.0.2.3       4       No        Up   Up
LSP-PE-1-PE-3-TE3-backupTE1         192.0.2.3       5       No        Up   Dwn
-------------------------------------------------------------------------------
LSPs : 3
```

The bandwidth requirement is lowered, as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" primary "dyn"
bandwidth 2500
```

Interface int-PE-1-PE-2 does not have sufficient bandwidth for CT1, but the longer path via PE-5 and PE-4 has sufficient unreserved bandwidth for CT1. The LSP will be operationally up, as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3-TE3-backupTE1 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3-TE3-backupTE1 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3-TE3-backupTE1
Path LSP ID       : 23552
From              : 192.0.2.1             To                   : 192.0.2.3
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/2                 Out Label            : 262143
---snip---
Setup Priority    : 4                     Oper Setup Priority  : 4
Hold Priority     : 4                     Oper Hold Priority   : 4
Class Type        : 1                     Oper CT              : 1
Backup CT         : 0
---snip---
Actual Hops       :
```

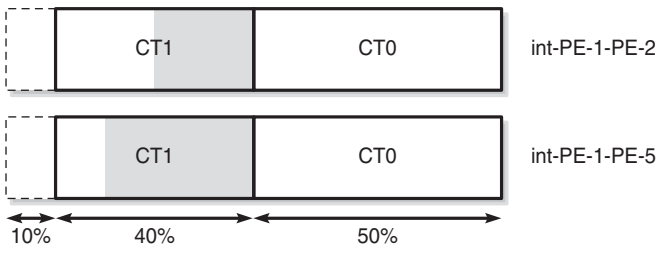
```
    192.168.15.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.15.2 (192.0.2.5)                    Record Label        : 262143
 -> 192.168.45.1 (192.0.2.4)                    Record Label        : 262143
 -> 192.168.34.1 (192.0.2.3)                    Record Label        : 262142
---snip---
```

The bandwidth reservation on RSVP interface int-PE-1-PE-2 remains unchanged, because the bandwidth is reserved on int-PE-1-PE-5, as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-5" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-5
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-5
-------------------------------------------------------------------------------
Interface          : int-PE-1-PE-5
Port ID            : 1/1/2
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0        : 50                    Link Bw CT4        : 0
Link Bw CT1        : 40                    Link Bw CT5        : 0
Link Bw CT2        : 0                     Link Bw CT6        : 0
Link Bw CT3        : 0                     Link Bw CT7        : 10
Bandwidth Constraints for Class Types (Kbps)
BC0                : 5000000               BC4                : 0
BC1                : 4000000               BC5                : 0
BC2                : 0                     BC6                : 0
BC3                : 0                     BC7                : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw    : 0                     Unresv. Bw         : 5000000
TE1->  Resv. Bw    : 0                     Unresv. Bw         : 5000000
TE2->  Resv. Bw    : 0                     Unresv. Bw         : 1500000
TE3->  Resv. Bw    : 2500000               Unresv. Bw         : 1500000
TE4->  Resv. Bw    : 0                     Unresv. Bw         : 0
TE5->  Resv. Bw    : 0                     Unresv. Bw         : 0
TE6->  Resv. Bw    : 0                     Unresv. Bw         : 0
TE7->  Resv. Bw    : 0                     Unresv. Bw         : 0
---snip---
```

Figure 123 shows the reserved and unreserved bandwidth on the RSVP interfaces on PE-1.

Figure 123 **Reserved and Unreserved Bandwidth on PE-1**



Trigger Backup Class-Type

This mechanism is described for MAM, but it is also supported in RDM.

Port 1/1/2 is shut down. The long path via PE-5 and PE-4 can no longer be used. However, the LSP has a backup CT (CT0), which is triggered by the port being down, as follows:

```
*A:PE-1# configure port 1/1/2 shutdown
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3-TE3-backupTE1 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3-TE3-backupTE1 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3-TE3-backupTE1
Path LSP ID       : 23554
From              : 192.0.2.1             To                   : 192.0.2.3
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262142
---snip---
Setup Priority    : 4                     Oper Setup Priority  : 4
Hold Priority     : 4                     Oper Hold Priority   : 4
Class Type        : 1                     Oper CT              : 0
Backup CT         : 0
---snip---
Actual Hops       :
    192.168.12.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262142
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262142
```

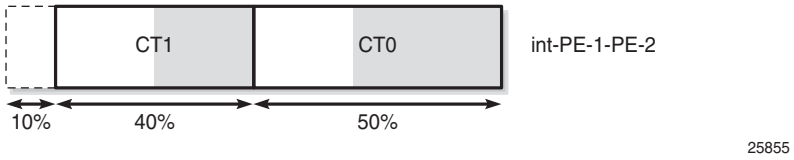
The bandwidth for this LSP is reserved in TE class 1, as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
Interface         : int-PE-1-PE-2
Port ID           : 1/1/1
Admin State       : Up                    Oper State        : Up
Active Sessions   : 2                     Active Resvs      : 2
Total Sessions    : 2
---snip---
```

```
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                  Unresv. Bw        : 2500000
TE1->  Resv. Bw   : 2500000            Unresv. Bw        : 2500000
TE2->  Resv. Bw   : 0                  Unresv. Bw        : 2000000
TE3->  Resv. Bw   : 2000000            Unresv. Bw        : 2000000
TE4->  Resv. Bw   : 0                  Unresv. Bw        : 0
TE5->  Resv. Bw   : 0                  Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                  Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                  Unresv. Bw        : 0
---snip---
```

Figure 124 shows the bandwidth reservation on interface int-PE-1-PE-2. The bandwidth reservation on interface int-PE-2-PE-3 on PE-2 is identical.

Figure 124 **Bandwidth Reservation**



The preceding examples illustrate that bandwidth can be wasted in the MAM model. The bandwidth allocated to CT7 cannot be used because there is no TE class configured with CT7. The bandwidth cannot be shared between CTs. The next section explains how the same LSPs will be used in the RDM model. They will be established one-by-one and, therefore, they are shut down, as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE5" shutdown
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE3" shutdown
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" shutdown
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE1" shutdown

*A:PE-1# configure port 1/1/2 no shutdown
```

Russian Doll Model

Enable DiffServ RDM

The DiffServ TE configuration needs to be consistent on all the nodes in the network, as follows:

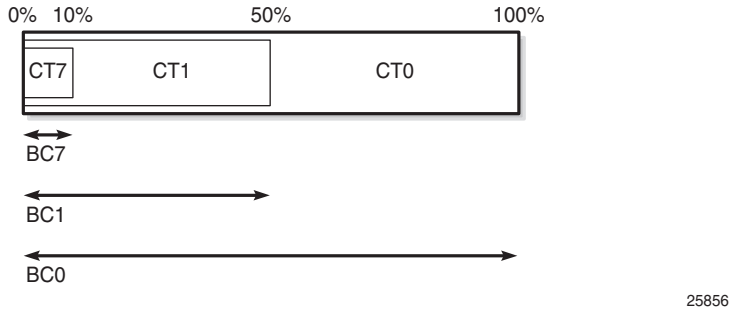
```
*A:PE-1# configure router
        mpls
            shutdown
```

```
exit
rsvp
    diffserv-te rdm
        class-type-bw ct0 50 ct1 40 ct2 0 ct3 0 ct4 0 ct5 0 ct6 0 ct7 10
        te-class 0 class-type 0 priority 7
        te-class 1 class-type 0 priority 4
        te-class 2 class-type 1 priority 7
        te-class 3 class-type 1 priority 4
        te-class 4 class-type 2 priority 7
        te-class 5 class-type 2 priority 2
        fc af class-type 1
        fc be class-type 0
        fc nc class-type 2
    exit
exit
mpls
    no shutdown
exit
```

In this example, three FCs are mapped to CTs. FC BE corresponds to CT0, FC AF to CT1, and FC NC to CT2. CT0 can be mapped to TE class 0 for priority 7, and to TE class 1 for priority 4. The mapping is similar for CT1 (with priorities 7 and 4), and CT2 (with priorities 7 or 2).

The RDM model allows the outer dolls (lower CT) to use the unused bandwidth allocated to the inner dolls (higher CT), as shown in Figure 125.

Figure 125 **Russian Doll Model for Three Class Types**



The calculation of the BCs takes into account the BCs of the inner dolls, as shown in the OSPF LSAs in the opaque database, as follows:

```
*A:PE-1# show router ospf opaque-database adv-router 192.0.2.1 detail
===============================================================================
Rtr Base OSPFv2 Instance 0 Opaque Link State Database (type: All) (detail)
===============================================================================
-------------------------------------------------------------------------------
Opaque LSA
-------------------------------------------------------------------------------
Area Id          : 0.0.0.0              Adv Router Id    : 192.0.2.1
Link State Id    : 1.0.0.1              LSA Type         : Area Opaque
```

```
Sequence No      : 0x80000003           Checksum         : 0x9035
Age              : 945                  Length           : 28
Options          :  E
Advertisement    : Traffic Engineering
     ROUTER-ID TLV  (0001) Len   4 : 192.0.2.1
-------------------------------------------------------------------------------
Opaque LSA
-------------------------------------------------------------------------------
Area Id          : 0.0.0.0              Adv Router Id    : 192.0.2.1
Link State Id    : 1.0.0.3              LSA Type         : Area Opaque
Sequence No      : 0x8000000d           Checksum         : 0x25f4
Age              : 15                   Length           : 164
Options          :  E
Advertisement    : Traffic Engineering
     LINK INFO TLV  (0002) Len 140 :
       Sub-TLV: 1     Len: 1     LINK_TYPE    : 1
       Sub-TLV: 2     Len: 4     LINK_ID      : 192.0.2.2
       Sub-TLV: 3     Len: 4     LOC_IP_ADDR  : 192.168.12.1
       Sub-TLV: 4     Len: 4     REM_IP_ADDR  : 192.168.12.2
       Sub-TLV: 5     Len: 4     TE_METRIC    : 10
       Sub-TLV: 6     Len: 4     MAX_BDWTH    : 10000000 Kbps
       Sub-TLV: 7     Len: 4     RSRVBL_BDWTH : 10000000 Kbps
       Sub-TLV: 8     Len: 32    UNRSRVD_CLS0 :
         P0: 10000000 Kbps P1: 10000000 Kbps P2: 5000000 Kbps P3: 5000000 Kbps
         P4: 1000000 Kbps P5: 1000000 Kbps P6:       0 Kbps P7:       0 Kbps
       Sub-TLV: 9     Len: 4     ADMIN_GROUP  : 0 None
       Sub-TLV: 17    Len: 36    TELK_BW_CONST:
         BW Model : RDM
         BC0: 10000000 Kbps BC1: 5000000 Kbps BC2: 1000000 Kbps BC3: 1000000 Kbps
         BC4: 1000000 Kbps BC5: 1000000 Kbps BC6: 1000000 Kbps BC7: 1000000 Kbps
-------------------------------------------------------------------------------
---snip---
```

Six TE classes are defined:

- TE0 and TE1 are defined for CT0. They can reserve all the available bandwidth, if it is not required by the other TE classes (100% = 50% for CT0 + 40% for CT1 + 10% for CT7)
- TE2 and TE3 are defined for CT1. They can reserve 50% of the bandwidth (50% = 40% for CT1 + 10% for CT7)
- TE4 and TE5 are defined for CT2. They can reserve 10% of the bandwidth, even though the configured bandwidth percentage for CT2 is 0. The 10% allocated to higher class CT7 can be used.

Bandwidth is more efficiently used in RDM than in MAM.

The BCs and bandwidth per TE class type show that bandwidth can be shared with the outer dolls, as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
```

```
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
Interface         : int-PE-1-PE-2
Port ID           : 1/1/1
Admin State       : Up                  Oper State        : Up
Active Sessions   : 0                   Active Resvs      : 0
Total Sessions    : 0
Subscription      : 100 %               Port Speed        : 10000 Mbps
Total BW          : 10000 Mbps          Aggregate         : Dsabl
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0       : 50                  Link Bw CT4       : 0
Link Bw CT1       : 40                  Link Bw CT5       : 0
Link Bw CT2       : 0                   Link Bw CT6       : 0
Link Bw CT3       : 0                   Link Bw CT7       : 10
Bandwidth Constraints for Class Types (Kbps)
BC0               : 10000000            BC4               : 1000000
BC1               : 5000000             BC5               : 1000000
BC2               : 1000000             BC6               : 1000000
BC3               : 1000000             BC7               : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                   Unresv. Bw        : 10000000
TE1->  Resv. Bw   : 0                   Unresv. Bw        : 10000000
TE2->  Resv. Bw   : 0                   Unresv. Bw        : 5000000
TE3->  Resv. Bw   : 0                   Unresv. Bw        : 5000000
TE4->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE5->  Resv. Bw   : 0                   Unresv. Bw        : 1000000
TE6->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                   Unresv. Bw        : 0
---snip---
```

Establishing LSPs

LSP-PE-1-PE-3-TE5 could not be established in the MAM model, because there was no bandwidth assigned to TE5 (CT2). However, in the RDM model, TE5 can use the bandwidth of the inner doll CT7 and the LSP will be operationally up, as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE5" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3-TE5 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3-TE5 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3-TE5
Path LSP ID       : 50176
From              : 192.0.2.1           To                  : 192.0.2.3
Admin State       : Up                  Oper State          : Up
```

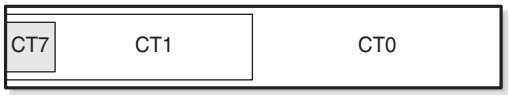
```
Path Name         : dyn                    Path Type            : Primary
Path Admin        : Up                     Path Oper            : Up
Out Interface     : 1/1/1                  Out Label            : 262143
---snip---
Setup Priority    : 2                      Oper Setup Priority  : 2
Hold Priority     : 2                      Oper Hold Priority   : 2
Class Type        : 2                      Oper CT              : 2
Backup CT         : None
---snip---
Actual Hops       :
    192.168.12.1 (192.0.2.1)                    Record Label         : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label         : 262143
 -> 192.168.23.2 (192.0.2.3)                    Record Label         : 262143
---snip---
```

The bandwidth reservation on interface int-PE-1-PE-2 is as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0       : 50                     Link Bw CT4       : 0
Link Bw CT1       : 40                     Link Bw CT5       : 0
Link Bw CT2       : 0                      Link Bw CT6       : 0
Link Bw CT3       : 0                      Link Bw CT7       : 10
Bandwidth Constraints for Class Types (Kbps)
BC0               : 10000000               BC4               : 1000000
BC1               : 5000000                BC5               : 1000000
BC2               : 1000000                BC6               : 1000000
BC3               : 1000000                BC7               : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                      Unresv. Bw        : 9000000
TE1->  Resv. Bw   : 0                      Unresv. Bw        : 9000000
TE2->  Resv. Bw   : 0                      Unresv. Bw        : 4000000
TE3->  Resv. Bw   : 0                      Unresv. Bw        : 4000000
TE4->  Resv. Bw   : 0                      Unresv. Bw        : 0
TE5->  Resv. Bw   : 1000000                Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                      Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                      Unresv. Bw        : 0
---snip---
```

This LSP uses all the available bandwidth for CT7. Because TE5 is defined with the best priority (2) of all TE classes, this LSP will not be preempted when a new LSP is enabled. Therefore, this bandwidth is subtracted from the amount of unreserved bandwidth. The remaining unreserved bandwidth is for CT0 and CT1 only. LSPs with other CTs cannot be established on this interface. Figure 126 shows the reserved bandwidth on interface int-PE-1-PE-2 for this LSP.

Figure 126 **Reserved Bandwidth for LSP with CT2 (One Session)**



Another LSP is established: LSP-PE-1-PE-3-TE3, with CT1 and priority 4, as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE3" no shutdown
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3-TE3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3-TE3 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE3
Path LSP ID      : 28162
From             : 192.0.2.1            To                  : 192.0.2.3
Admin State      : Up                   Oper State          : Up
Path Name        : dyn                  Path Type           : Primary
Path Admin       : Up                   Path Oper           : Up
Out Interface    : 1/1/1                Out Label           : 262142
---snip---
Setup Priority   : 4                    Oper Setup Priority  : 4
Hold Priority    : 4                    Oper Hold Priority   : 4
Class Type       : 1                    Oper CT              : 1
Backup CT        : None
---snip---
Actual Hops      :
    192.168.12.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262142
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262142
---snip---
```

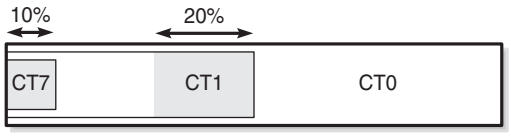
The bandwidth reservation on RSVP interface int-PE-1-PE-2 is as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
===============================================================================
RSVP Interface (Detailed) : int-PE-1-PE-2
===============================================================================
-------------------------------------------------------------------------------
Interface : int-PE-1-PE-2
-------------------------------------------------------------------------------
Interface         : int-PE-1-PE-2
Port ID           : 1/1/1
Admin State       : Up                  Oper State        : Up
```

```
Active Sessions    : 2                    Active Resvs       : 2
Total Sessions     : 2
---snip---
Bandwidth Constraints for Class Types (Kbps)
BC0                : 10000000             BC4                : 1000000
BC1                : 5000000              BC5                : 1000000
BC2                : 1000000              BC6                : 1000000
BC3                : 1000000              BC7                : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw    : 0                    Unresv. Bw         : 7000000
TE1->  Resv. Bw    : 0                    Unresv. Bw         : 7000000
TE2->  Resv. Bw    : 0                    Unresv. Bw         : 2000000
TE3->  Resv. Bw    : 2000000              Unresv. Bw         : 2000000
TE4->  Resv. Bw    : 0                    Unresv. Bw         : 0
TE5->  Resv. Bw    : 1000000              Unresv. Bw         : 0
TE6->  Resv. Bw    : 0                    Unresv. Bw         : 0
TE7->  Resv. Bw    : 0                    Unresv. Bw         : 0
---snip---
```

Figure 127 shows the bandwidth reservation for the two active sessions.

Figure 127 **Bandwidth Reservation for LSP with CT2 and LSP with CT1 (2 Sessions)**



Another LSP is established for CT1, requesting more bandwidth than the short path via PE-2 has available. Therefore, the longer path via PE-5 and PE-4 is set up for this LSP, as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" no shutdown

*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3-TE3-backupTE1 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3-TE3-backupTE1 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3-TE3-backupTE1
Path LSP ID       : 23556
From              : 192.0.2.1             To                  : 192.0.2.3
Admin State       : Up                    Oper State          : Up
```

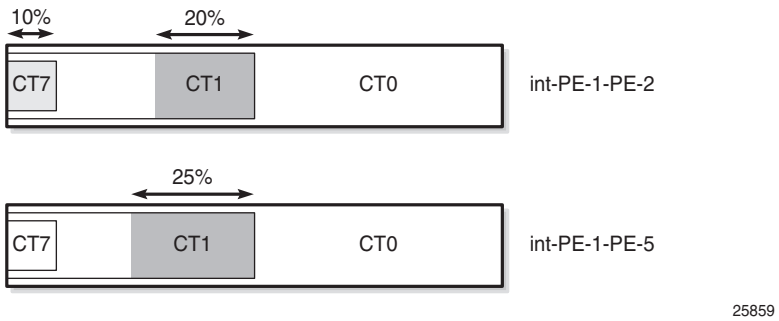
```
Path Name         : dyn                    Path Type            : Primary
Path Admin        : Up                     Path Oper            : Up
Out Interface     : 1/1/2                  Out Label            : 262143
---snip---
Setup Priority    : 4                      Oper Setup Priority  : 4
Hold Priority     : 4                      Oper Hold Priority   : 4
Class Type        : 1                      Oper CT              : 1
Backup CT         : 0
---snip---
Actual Hops       :
    192.168.15.1 (192.0.2.1)                    Record Label         : N/A
 -> 192.168.15.2 (192.0.2.5)                    Record Label         : 262143
 -> 192.168.45.1 (192.0.2.4)                    Record Label         : 262143
 -> 192.168.34.1 (192.0.2.3)                    Record Label         : 262141
---snip---
```

The bandwidth for this LSP is reserved on interface int-PE-1-PE-5, because the amount of unreserved bandwidth for TE3 is insufficient and inner dolls cannot use bandwidth assigned to outer dolls. Inner dolls are of higher priority than outer dolls, as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-5" detail
---snip---
Bandwidth Constraints for Class Types (Kbps)
BC0               : 10000000             BC4               : 1000000
BC1               : 5000000              BC5               : 1000000
BC2               : 1000000              BC6               : 1000000
BC3               : 1000000              BC7               : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                    Unresv. Bw        : 7500000
TE1->  Resv. Bw   : 0                    Unresv. Bw        : 7500000
TE2->  Resv. Bw   : 0                    Unresv. Bw        : 2500000
TE3->  Resv. Bw   : 2500000              Unresv. Bw        : 2500000
TE4->  Resv. Bw   : 0                    Unresv. Bw        : 1000000
TE5->  Resv. Bw   : 0                    Unresv. Bw        : 1000000
TE6->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                    Unresv. Bw        : 0
---snip---
```

Figure 128 shows the reserved bandwidth on both interfaces of PE-1.

Figure 128 **Reserved Bandwidth on Both Interfaces of PE-1 (3 Sessions)**



The following LSP is configured on PE-1:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3-TE1"
                to 192.0.2.3
                cspf
                primary "dyn"
                    bandwidth 3000
                    priority 4 4
                exit
                no shutdown
            exit
```

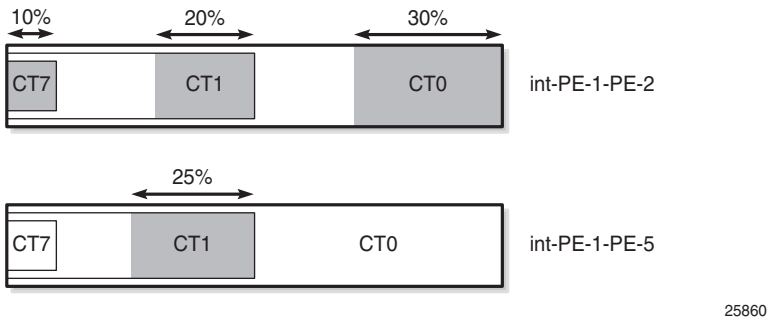
The class type is by default 0. CT0 and priority 4 corresponds to TE1. There is sufficient bandwidth available on the short path via PE-2. The bandwidth reservation on RSVP interface int-PE-1-PE-2 is as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
---snip---
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                  Unresv. Bw        : 4000000
TE1->  Resv. Bw   : 3000000            Unresv. Bw        : 4000000
TE2->  Resv. Bw   : 0                  Unresv. Bw        : 2000000
TE3->  Resv. Bw   : 2000000            Unresv. Bw        : 2000000
TE4->  Resv. Bw   : 0                  Unresv. Bw        : 0
TE5->  Resv. Bw   : 1000000            Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                  Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                  Unresv. Bw        : 0
---snip---
```

None of the established LSPs can be preempted. Therefore, the sum of the reserved and unreserved bandwidth does not exceed the total bandwidth.

Figure 129 shows the bandwidth reservation on both interfaces.

Figure 129 **Reserved Bandwidth on Both Interfaces on PE-1 (4 Sessions)**



The following LSP with CT0 and priority 7 is configured on PE-1:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3-TE0"
                to 192.0.2.3
                cspf
                primary "dyn"
                    bandwidth 4000
                    priority 7 7
                exit
                no shutdown
            exit
```

The bandwidth is reserved in TE class 0. There is sufficient bandwidth on the short path to PE-3. The bandwidth is now reserved for 100%, as follows:

```
*A:PE-1# show router rsvp interface
===============================================================================
RSVP Interfaces
===============================================================================
Interface                       Total    Active    Total BW  Resv BW   Adm Opr
                                Sessions Sessions  (Mbps)    (Mbps)
-------------------------------------------------------------------------------
system                          -        -         -         -         Up  Up
int-PE-1-PE-2                   4        4         10000     10000     Up  Up
int-PE-1-PE-5                   1        1         10000     2500      Up  Up
-------------------------------------------------------------------------------
Interfaces : 3
```

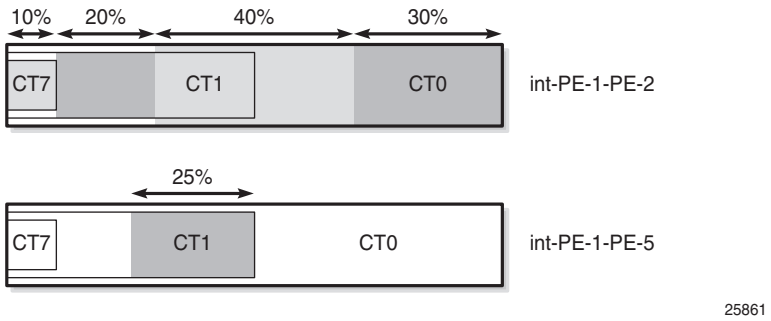
The bandwidth reservation on int-PE-1-PE-2 is as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0       : 50                  Link Bw CT4       : 0
Link Bw CT1       : 40                  Link Bw CT5       : 0
Link Bw CT2       : 0                   Link Bw CT6       : 0
Link Bw CT3       : 0                   Link Bw CT7       : 10
Bandwidth Constraints for Class Types (Kbps)
BC0               : 10000000            BC4               : 1000000
BC1               : 5000000             BC5               : 1000000
BC2               : 1000000             BC6               : 1000000
BC3               : 1000000             BC7               : 1000000
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 4000000             Unresv. Bw        : 0
TE1->  Resv. Bw   : 3000000             Unresv. Bw        : 4000000
TE2->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE3->  Resv. Bw   : 2000000             Unresv. Bw        : 2000000
TE4->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE5->  Resv. Bw   : 1000000             Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                   Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                   Unresv. Bw        : 0
---snip---
```

Even though the sum of the reserved bandwidth equals the maximum reservable bandwidth on the link, there is still unreserved bandwidth for specific TE classes. When an additional LSP is established requiring bandwidth in TE3 or TE1 (which have setup priority 4), it can preempt another LSP with a lower hold priority. LSPs requiring bandwidth in TE class TE2 have a setup priority 7 and cannot preempt any other LSP. The setup priority in TE1 and TE3 is 4, which is higher than the hold priority in TE2 and TE0 (7 is the lowest priority). There are no LSPs in TE2, so the only LSPs to preempt have bandwidth reserved in TE0.

Figure 130 shows the bandwidth reservation on the interfaces of PE-1.

Figure 130 **Reserved Bandwidth on Both Interfaces of PE-1 (5 Sessions)**



Preemption

The following LSP is configured with CT1, setup priority 4, and hold priority 4, which corresponds to TE class 3:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3-TE3-2nd"
                to 192.0.2.3
                cspf
                class-type 1
                primary "dyn"
                    bandwidth 750
                    priority 4 4
                exit
                no shutdown
            exit
```

Because the setup priority 4 exceeds the hold priority 7 of LSP-PE-1-PE-3-TE0, this LSP will preempt the existing one. The following output shows that the next hop for LSP-PE-1-PE-3-TE0 is 192.168.15.2 (PE-5), while the next hop for LSP-PE-1-PE-3-TE3-2nd is 192.168.12.2 (PE-2):

```
*A:PE-1# show router mpls lsp path
===============================================================================
MPLS LSP  Path
===============================================================================
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE5    To                  : 192.0.2.3
Adm State        : Up                   Oper State          : Up
-------------------------------------------------------------------------------
Path Name                         Next Hop        Type       Out I/F   Adm  Opr
-------------------------------------------------------------------------------
dyn                               192.168.12.2    Primary    1/1/1     Up   Up
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE3    To                  : 192.0.2.3
Adm State        : Up                   Oper State          : Up
-------------------------------------------------------------------------------
Path Name                         Next Hop        Type       Out I/F   Adm  Opr
-------------------------------------------------------------------------------
dyn                               192.168.12.2    Primary    1/1/1     Up   Up
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE3-   To                  : 192.0.2.3
                   backupTE1
Adm State        : Up                   Oper State          : Up
-------------------------------------------------------------------------------
Path Name                         Next Hop        Type       Out I/F   Adm  Opr
-------------------------------------------------------------------------------
dyn                               192.168.15.2    Primary    1/1/2     Up   Up
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE1    To                  : 192.0.2.3
Adm State        : Up                   Oper State          : Up
-------------------------------------------------------------------------------
Path Name                         Next Hop        Type       Out I/F   Adm  Opr
-------------------------------------------------------------------------------
dyn                               192.168.12.2    Primary    1/1/1     Up   Up
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE0    To                  : 192.0.2.3
Adm State        : Up                   Oper State          : Up
-------------------------------------------------------------------------------
Path Name                         Next Hop        Type       Out I/F   Adm  Opr
-------------------------------------------------------------------------------
dyn                               192.168.15.2    Primary    1/1/2     Up   Up
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3-TE3-   To                  : 192.0.2.3
                   2nd
Adm State        : Up                   Oper State          : Up
-------------------------------------------------------------------------------
Path Name                         Next Hop        Type       Out I/F   Adm  Opr
-------------------------------------------------------------------------------
dyn                               192.168.12.2    Primary    1/1/1     Up   Up
===============================================================================
*A:PE-1#
```

The bandwidth reservation on RSVP interface int-PE-1-PE-2 is as follows:

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-2" detail
---snip---
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 0                  Unresv. Bw        : 3250000
TE1->  Resv. Bw   : 3000000            Unresv. Bw        : 3250000
```

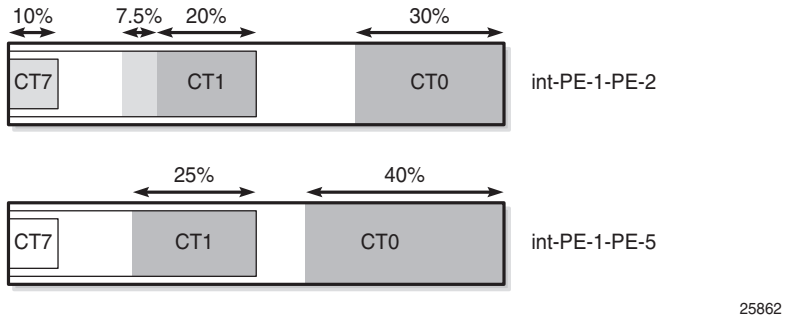
```
TE2->  Resv. Bw   : 0                    Unresv. Bw        : 1250000
TE3->  Resv. Bw   : 2750000              Unresv. Bw        : 1250000
TE4->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE5->  Resv. Bw   : 1000000              Unresv. Bw        : 0
TE6->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                    Unresv. Bw        : 0
---snip---
```

The bandwidth reservation on RSVP interface int-PE-1-PE-5 is as follows

```
*A:PE-1# show router rsvp interface "int-PE-1-PE-5" detail
---snip---
Bandwidth for TE Class Types (Kbps)
TE0->  Resv. Bw   : 4000000              Unresv. Bw        : 3500000
TE1->  Resv. Bw   : 0                    Unresv. Bw        : 7500000
TE2->  Resv. Bw   : 0                    Unresv. Bw        : 2500000
TE3->  Resv. Bw   : 2500000              Unresv. Bw        : 2500000
TE4->  Resv. Bw   : 0                    Unresv. Bw        : 1000000
TE5->  Resv. Bw   : 0                    Unresv. Bw        : 1000000
TE6->  Resv. Bw   : 0                    Unresv. Bw        : 0
TE7->  Resv. Bw   : 0                    Unresv. Bw        : 0
---snip---
```

Figure 131 shows the bandwidth reservation on both interfaces on PE-1 for the six sessions.

Figure 131 **Reserved Bandwidth on Both Interfaces on PE-1 (6 Sessions)**



Preemption can also be within the same CT. An LSP with CT0 and priority 4 (TE1) could have preempted the LSP with CT0 and priority 7 (TE0) equally well.

Bandwidth Availability Check

A tools command can be launched to verify the available bandwidth toward a node for a specific class type (by default CT0) and priority (by default setup priority 7 and hold priority 0). The options for this command are as follows:

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3
  - cspf to <ip-addr> [from <ip-addr>] [bandwidth <bandwidth>] [include-
bitmap <bitmap>] [exclude-bitmap <bitmap>] [hop-limit <limit>] [exclude-
address <excl-addr> [<excl-addr>...(upto 8 max)]] [use-te-metric] [strict-
srlg] [srlg-group <grp-id>...(upto 8 max)] [exclude-node <excl-node-id> [<excl-node-
id>..(upto 8 max)]] [skip-interface <interface-name>] [ds-class-type <class-
type>] [cspf-reqtype <req-type>] [least-fill-min-thd <thd>] [setup-
priority <val>] [hold-priority <val>]
 <ip-addr>            : a.b.c.d
 <bandwidth>          : [1..100000] in Mbps
 <bitmap>             : [0..4294967295] - accepted in decimal, hex(0x) or binary(0b)
 <limit>              : [2..255]
 <excl-addr>          : a.b.c.d (outbound interface)
 <use-te-metric>      : keyword
 <strict-srlg>        : keyword
 <grp-id>             : [0..4294967295]
 <excl-node-id>       : [a.b.c.d] (outbound interface)
 <interface-name>     : [max 32 chars]
 <class-type>         : [0..7]
 <req-type>           : all|random|least-fill : keywords
 <thd>                : [1..100]
 <priority>           : [0..7]
```

The following verifies whether an LSP can be set up from PE-1 to PE-3 requesting 100 Mb/s with CT0 (default) and both priorities equal to 4:

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 bandwidth 100 setup-
priority 4 hold-priority 4
Req CSPF for all ECMP paths
    from: this node to: 192.0.2.3 w/
(DiffServ = RDM) class: 0 , setup Priority 4, Hold Priority 4 TE Class: 1
CSPF Path
To        : 192.0.2.3
Path 1    : (cost 20)
    Src:   192.0.2.1   (= Rtr)
    Egr:   192.168.12.1     -> Ingr:   192.168.12.2    Rtr:   192.0.2.2   (met 10)
    Egr:   192.168.23.1     -> Ingr:   192.168.23.2    Rtr:   192.0.2.3   (met 10)
    Dst:   192.0.2.3   (= Rtr)
*A:PE-1#
```

The short path via PE-2 has sufficient bandwidth for an LSP with these TE requirements (TE class 1 with CT0 and both priorities 4). This is different for TE class 5 (CT2 and priorities 2), where the bandwidth is completely reserved. The following shows that the longer path via PE-5 and PE-4 must be taken:

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 bandwidth 100 ds-class-
type 2 setup-priority 2 hold-priority 2
Req CSPF for all ECMP paths
    from: this node to: 192.0.2.3 w/
(DiffServ = RDM) class: 2 , setup Priority 2, Hold Priority 2 TE Class: 5
CSPF Path
To        : 192.0.2.3
Path 1    : (cost 30)
    Src:   192.0.2.1   (= Rtr)
    Egr:   192.168.15.1     -> Ingr:   192.168.15.2    Rtr:   192.0.2.5   (met 10)
    Egr:   192.168.45.2     -> Ingr:   192.168.45.1    Rtr:   192.0.2.4   (met 10)
```

```
    Egr:   192.168.34.2     -> Ingr:   192.168.34.1    Rtr:   192.0.2.3   (met 10)
    Dst:   192.0.2.3   (= Rtr)
*A:PE-1#
```

This tools command can only be launched when a TE class is defined with the requested CT and priority. An error is raised when the request cannot be fulfilled, as follows:

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 setup-priority 5
MINOR: CLI No Te class mapped to Class Type 0 , Setup Priority 5.
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 setup-priority 4
MINOR: CLI No Te class mapped to Class Type 0 , Hold Priority 0.
```

Conclusion

DiffServ TE enforces different BCs for different classes of traffic. DiffServ TE controls overbooking and supports preemption. Two BC models are described in this chapter: MAM and RDM.

Entropy Label

This chapter provides information about the entropy label (EL).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4.

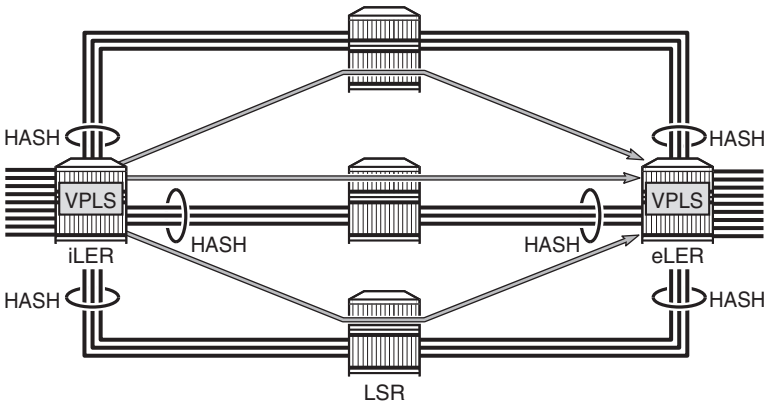
RFC 6391 hash label or flow-aware transport label is supported in SR OS release 8.0.R1, and later. RFC 6790 entropy labels (ELs) are supported on RSVP and BGP tunnels in release 14.0.R1, and later, and supported on LDP tunnels in release 14.0.R4, and later. Chassis-mode D is required.

Overview

Entropy is the degree of disorder or uncertainty in a system. SR OS supports both the MPLS EL and the hash label, but they are mutually exclusive. These labels allow label switched routers (LSRs) to load-balance labeled packets in a granular way without the need to inspect the IPv4 or IPv6 header. The hash label is applicable to services such as Epipe VLL, VPLS, IES (spoke-SDP), and VPRN services. The main advantage of the EL compared to the hash label is that the EL can be applied to a wider range of services, such as EVPN, BGP VPWS, Fpipe VLL, Ipipe VLL, H-VPLS, and BGP RFC 3107 tunnels.

Figure 132 shows that different flows from an ingress label edge router (iLER) are load-balanced across different paths in the MPLS network toward the egress label edge router (eLER).

Figure 132 **Load-Balancing of Flows Based on Hash Label or Entropy Label**

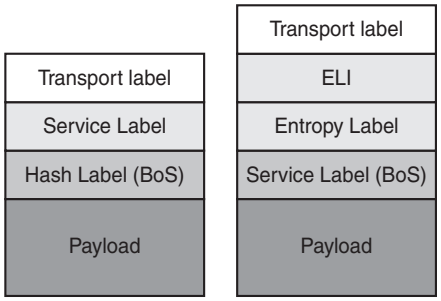


The EL is inserted below the innermost tunnel label (TL) in the label stack, along with the entropy label indicator (ELI), which is a special purpose MPLS label with a value of 7 to indicate that the next label in the stack is an EL. As with the hash label, the value of the EL is calculated based on a hash of the packet payload header (IP and Layer 4). The EL is inserted as deep as possible in the stack to ensure preservation for as far as possible through the network. When the EL and ELI are present, load-balancing on the transit LSRs can be configured to only take into account the EL label for LAG and ECMP. Load-balancing on the LSR can be configured to take into account the IP header also, but that is not required when the EL and ELI are present. The eLER removes the EL and ELI before forwarding the packet to its final destination.

The EL requires two additional labels in the label stack, which might result in an unsupported label stack depth in an intermediate (possibly third-party) LSR. SR OS allows control of the EL and ELI insertion and accounting of extra labels in the tunnel table. The supported label stack depth is 12 in SR OS release 14.0, including transport, service, hash, and OAM labels.

Figure 133 shows a comparison between a label stack with a hash label and a label stack with an EL and ELI.

Figure 133 **Label Stack with Hash Label versus Label Stack with EL and ELI**



Hash labels or ELs are inserted by the iLER, based on the incoming packet header. The iLER creates a unique label per conversation, based on the incoming packet header (IP and Layer 4). The hash label is a single label that is inserted at the bottom of the stack, below the service label. The EL is inserted together with the ELI below the innermost tunnel label, but above the service label. The entropy and hash label value is outside of the range for MPLS label values, because the most significant bit equals 1 for entropy and hash labels and 0 for MPLS labels.

The EL is used as part of the LSR hashing algorithm for spraying packets over multi-port LAGs, ECMP, or BGP tunnels with multiple downstream interfaces. The packet ordering is preserved because one label is used per conversation. Hashing based on a label stack containing an EL per service will become more granular than when based on a standard label stack alone.

Figure 134 shows how the eLER signals its EL capability to the iLER.

Figure 134 **Downstream LERs Signal EL Capability to ILER**



The entropy label capability (ELC) is signaled by the eLER and indicates the ability to receive and process the ELs. This can be advertised for LDP and RSVP. However, ELC signaling is not supported for BGP tunnels, because no agreed standard exists in the IETF. RFC 6790 introduced the ELC BGP attribute that can be signaled by the eLER to indicate that it supports EL. According to RFC 6790, LSRs incapable of processing ELs must remove the ELC BGP attribute, but this requirement could not be guaranteed; therefore, the ELC BGP attribute has been deprecated in RFC 7447.

As a workaround, at the iLER, an override of the ELC for a BGP tunnel can be manually configured. After this ELC has been overridden, the BGP sender assumes that the receiver is capable of receiving and processing the ELs, regardless of the signaled ELC. The iLER inserts an EL on a tunnel for which the ELC is confirmed by the downstream peer or when the ELC is overridden by configuration.

ELC signaling can be enabled for LDP on the LERs, as follows:

```
*A:PE-2# configure router ldp entropy-label-capability
```

ELC signaling for RSVP can be enabled as follows:

```
*A:PE-2# configure router rsvp entropy-label-capability
```

As previously described, the lack of an IETF standard for BGP tunnels means that ELC is ignored by the receiving LER, and no EL is inserted. Therefore, an override is required that assumes that the far-end LER has ELC, and allows insertion of the EL. The override is enabled as follows:

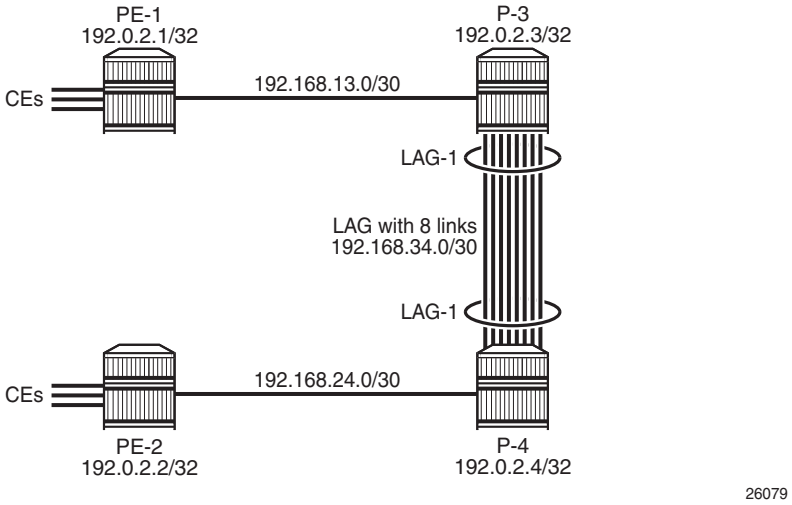
```
*A:PE-1# configure router bgp override-tunnel-elc
```

When the ELC is overridden for BGP, the iLER assumes that the receiver can handle the EL.

Configuration

In this section, an EVPN-VPLS will be configured on PE-1 and PE-2 to illustrate LAG hashing based on EL. Figure 135 shows the topology used for this example. Load-balancing of the traffic will be performed in the P-routers that are connected by a LAG with eight links.

Figure 135 **Example Topology**



The initial configuration of the PE/P-routers includes the following:

- Cards, MDAs, ports
- LAG with eight network links between P-3 and P-4
- Router interfaces
- IGP (IS-IS or OSPF)
- LDP enabled on all interfaces
- MPLS enabled on all interfaces. RSVP enabled.
- iBGP configured with P-3 as route reflector (RR)

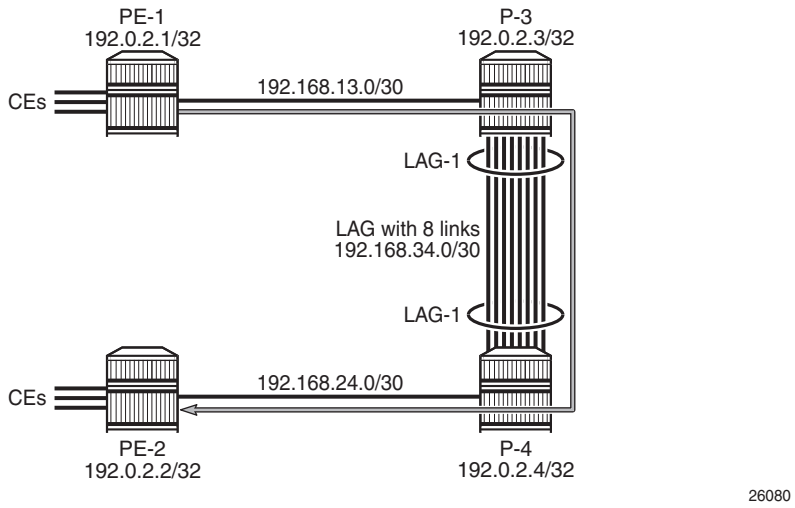
For the configuration of EVPN-MPLS, the BGP configuration needs to include the address family EVPN, as follows. See chapter EVPN for MPLS Tunnels for more information.

```
configure
    router
        bgp
            min-route-advertisement 1
            rapid-update evpn
            group "iBGP"
                family evpn
                peer-as 64500
                neighbor 192.0.2.3
                exit
            exit
        exit
```

EVPN Service Using RSVP Tunnel with EL

Figure 136 shows LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via core routers P-3 and P-4.

Figure 136 **RSVP LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via P-3 and P-4**



The LSP "LSP-PE-1-PE-2" is configured on PE-1, as follows:

```
configure
    router
        mpls
            path "path-PE-1-PE-2"
                hop 10 192.168.13.2 strict
                hop 20 192.168.34.2 strict
                hop 30 192.168.24.1 strict
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "path-PE-1-PE-2"
                exit
                no shutdown
            exit
        exit
    exit
exit
```

The configuration for LSP "LSP-PE-2-PE-1" on PE-2 is similar.

ELC is disabled by default, and needs to be enabled for RSVP on the eLERs, as follows:

```
*A:PE-2# configure router rsvp entropy-label-capability
```

The EL can be disabled (force-disable) or enabled in the MPLS context on the iLER, as follows:

```
*A:PE-1# configure router mpls entropy-label
  - entropy-label rsvp-te <rsvp-te>

 <rsvp-te>            : <force-disable | enable>
```

The EL can also be disabled (force-disable) or enabled per LSP, but there is a third option to inherit the EL settings from the MPLS context, as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" entropy-label
  - entropy-label {force-disable | inherit | enable}

 <force-disable | i*> : force-disable|inherit|enable
```

By default, the EL on the LSP inherits the EL settings in the MPLS context, as follows:

```
*A:PE-1# configure router mpls
*A:PE-1>config>router>mpls#  info detail
----------------------------------------------
            ---snip---
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                ---snip---
                entropy-label inherit
                ---snip---
```

When LSP templates are used, EL can be configured within the LSP template context for single-hop and mesh point-to-point LSPs, as follows:

```
*A:PE-1# configure router mpls lsp-template "LSPtemplate1" one-hop-p2p entropy-
label
  - entropy-label {force-disable | inherit | enable}

 <force-disable | i*> : force-disable|inherit|enable


*A:PE-1# configure router mpls lsp-template "LSPtemplate2" mesh-p2p entropy-label
  - entropy-label {force-disable | inherit | enable}

 <force-disable | i*> : force-disable|inherit|enable
```

When the EL settings are modified, for example, from inherit to enabled, the changes only take effect after the LSP has been cleared or MPLS has been bounced, using shutdown/no shutdown. The following message is raised when the configuration is modified for an LSP in no shutdown state:

```
*A:PE-1>config>router>mpls>lsp# entropy-label enable
INFO: MPLS #1029 Entropy Label change is not operational - LSP must be bounced
```

The following command shows that EL is enabled in MPLS:

```
*A:PE-1# show router mpls status | match Label
Entropy Label RSVP-TE     : Enabled
*A:PE-1#
```

The following command shows that EL is enabled in RSVP:

```
*A:PE-1# show router rsvp status | match Label
Implicit Null Label: Disabled           Node-id in RRO     : Exclude
DiffServTE AdmModel: Basic              Entropy Label      : Enabled
*A:PE-1#
```

The following command shows that EL is enabled and operational in LSP "LSP-PE-1-PE-2":

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" detail | match Label
Entropy Label   : Enabled                   Oper Entropy Label   : Enabled
*A:PE-1#
```

In this case, the EL was configured to be enabled on the LSP, instead of the default behavior, which would have been displayed as "Inherited", as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" detail | match Label
Entropy Label   : Inherited                   Oper Entropy Label   : Enabled
*A:PE-1#
```

The following command shows that the LSP "LSP-PE-1-PE-2" has the flag "entropy-label-capable" in the RSVP tunnel table:

```
*A:PE-1# show router tunnel-table protocol rsvp detail

===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination      : 192.0.2.2/32
NextHop          : 192.168.13.2
Tunnel Flags     : exclude-for-lfa entropy-label-capable
Age              : 00h14m59s
CBF Classes      : (Not Specified)
Owner            : rsvp                 Encap            : MPLS
Tunnel ID        : 1                    Preference       : 7
Tunnel Label     : 262137               Tunnel Metric    : 16777215
Tunnel MTU       : 1496                 Max Label Stack  : 1
LSP ID           : 49666                Bypass Label     : 0
LSP Bandwidth    : 0                    LSP Weight       : 0
-------------------------------------------------------------------------------
Number of tunnel-table entries          : 1
Number of tunnel-table entries with LFA : 0
===============================================================================
*A:PE-1#
```

EL is supported on many Layer 2 and Layer 3 services. In this example, VPLS 1 is configured on PE-1 with BGP EVPN-MPLS, as follows:

```
configure
    service
        vpls 1 customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls
                    entropy-label
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/2:1 create
            exit
            no shutdown
```

Auto-bind-tunnel is resolved to the RSVP LSP "LSP-PE-1-PE-2". A similar configuration is applied on PE-2.

The following command shows that the EL is enabled for BGP-EVPN, as follows:

```
*A:PE-1# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled           Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
VXLAN Admin Status : Disabled          Creation Origin    : manual
MAC Dup Detn Moves : 5                 MAC Dup Detn Window: 3
MAC Dup Detn Retry : 9                 Number of Dup MACs : 0
IP Route Advert    : Disabled

EVI                : 1
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled
-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses            Time Detected
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN MPLS Information
===============================================================================

Admin Status       : Enabled
Force Vlan Fwding  : Disabled          Control Word       : Disabled
Split Horizon Group: (Not Specified)
```

```
Ingress Rep BUM Lbl: Disabled           Max Ecmp Routes    : 0
Ingress Ucast Lbl  : 262138             Ingress Mcast Lbl  : 262138
Entropy Label      : Enabled
RestProtSrcMacAct  : none
===============================================================================
===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Resolution         : filter
Filter Tunnel Types: rsvp
===============================================================================
*A:PE-1#
```

The iLER PE-1 will add the EL and ELI to the label stack. Traffic load-balancing will be performed in P-3 where the traffic will be sprayed over all eight links of the LAG. By default, the load-balancing settings on P-3 are as follows:

```
*A:P-3>config>system>load-balancing# info detail
----------------------------------------------
            no l4-load-balancing
            no lsr-load-balancing
            no system-ip-load-balancing
            no mc-enh-load-balancing
            no service-id-lag-hashing
----------------------------------------------
```

When the EL and ELI are included in the stack, only the EL label is used for the load-balancing. It is not required to inspect the IP header or any other header. The options for LSR load-balancing are the following:

```
*A:P-3# configure system load-balancing lsr-load-balancing
  - lsr-load-balancing <hashing-algorithm>
  - no lsr-load-balancing

 <hashing-algorithm>  : lbl-only | lbl-ip | ip-only | eth-encap-ip
```

The option **lbl-only** is preferred when the EL is present, and is configured as follows:

```
*A:P-3# configure system load-balancing lsr-load-balancing lbl-only
```

For the reverse path, PE-2 is the iLER, and P-4 will do the spraying of the packets over all links of the LAG. Therefore, LSR load-balancing with lbl-only is also configured on P-4.

EVPN Service Using LDP Tunnel with EL

ELC is disabled by default for LDP, as follows:

```
*A:PE-1# show router ldp status
```

```
===============================================================================
LDP Status for IPv4 LSR ID 192.0.2.1
              IPv6 LSR ID ::
===============================================================================
Created at         : 09/29/16 13:39:26
Last Change        : 09/29/16 13:39:26
Admin State        : Up
IPv4 Oper State    : Up                    IPv6 Oper State      : Down
IPv4 Up Time       : 0d 00:01:21           IPv6 Down Time       : 0d 00:01:21
IPv4 Oper Down Rea*: n/a                   IPv6 Oper Down Reason: systemIpDown
IPv4 Oper Down Eve*: 0                     IPv6 Oper Down Events: 0
Tunn Down Damp Time: 3 sec
Label Withdraw Del*: 0 sec                 Implicit Null Label  : Disabled
Short. TTL Local   : Enabled               Short. TTL Transit   : Enabled
Import Policies    : None                  Export Policies      : None
Tunl Exp Policies  : None
FRR                : Disabled              Mcast Upstream FRR   : Disabled
MP MBB Time        : 3
Aggregate Prefix   : False                 Agg Prefix Policies  : None
Class Forwarding   : Disabled
Legacy LSR Interop : False
Entropy Label Capa*: False
-------------------------------------------------------------------------------
Capabilities
-------------------------------------------------------------------------------
Dynamic            : Enabled               P2MP                 : Enabled
IPv4 Prefix Fec    : Enabled               IPv6 Prefix Fec      : Enabled
Service Fec128     : Enabled               Service Fec129       : Enabled
MP MBB             : Enabled               Overload             : Enabled
Unrecognized Notif*: Enabled
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The command to enable ELC is as follows:

```
*A:PE-1# configure router ldp entropy-label-capability
```

In the list of LDP active bindings, the egress labels that are pushed by iLER PE-1 on traffic toward an eLER capable of handling an EL get the indication "e", as follows:

```
*A:PE-1# show router ldp bindings active prefixes ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
```

```
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                              Pop             262143      --
  --                                        --

192.0.2.2/32                              Push              --      262140e
192.168.13.2                              1/1/1

192.0.2.2/32                              Swap            262132    262140
192.168.13.2                              1/1/1

192.0.2.3/32                              Push              --      262143
192.168.13.2                              1/1/1

192.0.2.4/32                              Push              --      262141
192.168.13.2                              1/1/1

192.0.2.4/32                              Swap            262133    262141
192.168.13.2                              1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 6
===============================================================================
*A:PE-1#
```

Because the iLER adds the EL, only labels which are pushed get the "e"-indication. Labels which are swapped or popped do not get this label.

The details of the tunnel table for LDP show that the LDP tunnel toward PE-2 has tunnel flag entropy-label-capable, as follows:

```
*A:PE-1# show router tunnel-table protocol ldp detail

===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination      : 192.0.2.2/32
NextHop          : 192.168.13.2
Tunnel Flags     : entropy-label-capable
Age              : 00h00m30s
CBF Classes      : (Not Specified)
Owner            : ldp                 Encap            : MPLS
Tunnel ID        : 65539               Preference       : 9
Tunnel Label     : 262140              Tunnel Metric    : 30
Tunnel MTU       : 1496                Max Label Stack  : 1
-------------------------------------------------------------------------------
---snip---
```

The following EVPN-VPLS uses an LDP transport tunnel and has EL enabled:

```
configure
    service
        vpls 2 customer 1 create
            bgp
            exit
```

```
            bgp-evpn
                evi 2
                mpls
                    entropy-label
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/2:16 create
            exit
            no shutdown
```

The configuration on PE-1 and PE-2 is similar.

EL is enabled in this service, as follows:

```
*A:PE-2# show service id 2 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled            Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
VXLAN Admin Status : Disabled           Creation Origin    : manual
MAC Dup Detn Moves : 5                  MAC Dup Detn Window: 3
MAC Dup Detn Retry : 9                  Number of Dup MACs : 0
IP Route Advert    : Disabled

EVI                : 2
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled
-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN MPLS Information
===============================================================================

Admin Status       : Enabled
Force Vlan Fwding  : Disabled           Control Word       : Disabled
Split Horizon Group: (Not Specified)
Ingress Rep BUM Lbl: Disabled           Max Ecmp Routes    : 0
Ingress Ucast Lbl  : 262130             Ingress Mcast Lbl  : 262130
Entropy Label      : Enabled
RestProtSrcMacAct  : none
===============================================================================
===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Resolution         : filter
Filter Tunnel Types: ldp
```

```
===============================================================================
*A:PE-2#
```

LSR Load-Sharing Based on EL

The LSR load-sharing based on the EL results in a granular load-sharing, as shown in the following test results. In each direction, 9000 flows were generated; each flow with four unique variables: source IP address, destination IP address, source UDP port, and destination UDP port. For this test, an EVPN-VPLS with RSVP tunnel was used. Each flow gets a unique EL and the P-nodes distribute the load over all eight links in the LAG, based on the hashing algorithm "lbl-only", as follows:

```
A:P-4# monitor lag 1 rate
===============================================================================
Monitor statistics for LAG ID 1
===============================================================================
Port-id   Input       Input       Output      Output      Input      Output
          Bytes       Packets     Bytes       Packets     Errors     Errors
-------------------------------------------------------------------------------
At time t = 15 sec (Mode: Rate)
-------------------------------------------------------------------------------
2/1/12    93943884    61572       94542017    61964       0          0
  % Util  76.14       --          76.62       --          --         --
2/1/13    94156935    61711       94535383    61959       0          0
  % Util  76.31       --          76.61       --          --         --
2/1/14    94042970    61637       94005352    61612       0          0
  % Util  76.22       --          76.19       --          --         --
2/1/15    94075016    61657       93999248    61608       0          0
  % Util  76.24       --          76.18       --          --         --
2/1/16    94162530    61715       93992635    61603       0          0
  % Util  76.31       --          76.17       --          --         --
2/1/17    94030277    61628       93986023    61599       0          0
  % Util  76.21       --          76.17       --          --         --
2/1/18    93993653    61604       94502319    61938       0          0
  % Util  76.18       --          76.59       --          --         --
2/1/19    93765261    61455       94495707    61933       0          0
  % Util  75.99       --          76.58       --          --         --
-------------------------------------------------------------------------------
Totals    752170526   492979      754058684   494216      0          0
  % Util  76.20       --          76.39       --          --         --
```

When one of the links is unavailable, the traffic is redistributed over the remaining seven links, as follows:

```
A:P-4# monitor lag 1 rate
===============================================================================
Monitor statistics for LAG ID 1
===============================================================================
Port-id   Input       Input       Output      Output      Input      Output
          Bytes       Packets     Bytes       Packets     Errors     Errors
-------------------------------------------------------------------------------
At time t = 18 sec (Mode: Rate)
```

```
-------------------------------------------------------------------------------
2/1/12    107659513   70560       108164110   70891       0          0
  % Util  87.25       --          87.66       --          --         --
2/1/13    107683907   70576       108245984   70944       0          0
  % Util  87.27       --          87.73       --          --         --
2/1/14    107717479   70598       107593365   70516       0          0
  % Util  87.30       --          87.20       --          --         --
2/1/15    107213899   70267       107511469   70462       0          0
  % Util  86.89       --          87.13       --          --         --
2/1/16    107426013   70407       107674751   70570       0          0
  % Util  87.06       --          87.26       --          --         --
2/1/17    107387355   70381       107757155   70624       0          0
  % Util  87.03       --          87.33       --          --         --
2/1/18    107086224   70184       108326862   70997       0          0
  % Util  86.79       --          87.79       --          --         --
-------------------------------------------------------------------------------
Totals    752174390   492973      755273696   495004      0          0
  % Util  87.08       --          87.44       --          --         --
```

The flows are evenly spread in each direction.

Conclusion

The EL is a standard-compliant way to maintain packet ordering within a conversation while load-balancing across multiple ECMP paths or links in a LAG. The EL is supported for both Layer 2 and Layer 3 services and is, therefore, a more general solution than the MPLS hash label.

IGP Shortcuts

This chapter provides information about IGP shortcuts.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS when the feature is not related to BGP.There are no other pre-requisites for this configuration. This chapter was originally written for release 12.0.R3, but the CLI in this edition corresponds to release 14.0.R1.

Overview

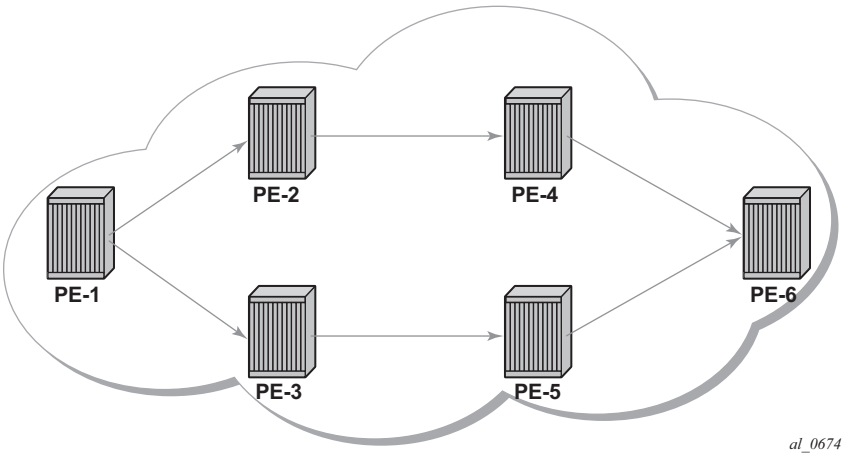
Interior gateway protocols (IGPs) are routing protocols that operate inside an autonomous system (AS). An AS is a network domain that is managed under a single administration. Because the scope of operation of an IGP is usually within an AS, IGPs are also called intra-AS protocols. The purpose of an IGP is to provide reachability information to destination nodes that are inside the domain. IGPs can be one or more of a variety of protocols, including routing protocols such as routing information protocol (RIP) version 1 or 2, open shortest path first (OSPF), and intermediate system to intermediate system (IS-IS).

IGPs such as OSPF and IS-IS are link-state protocols that use a shortest path first (SPF) algorithm to compute the shortest path tree to all nodes in a network. The results of such computations indicate the destination node, next hop address, and output interface, where the output interface is a physical interface. Optionally, multiprotocol label switching (MPLS) label switched paths (LSPs) can be included in the SPF algorithm on the node performing the calculations, as LSPs behave as

logical interfaces directly connected to remote nodes in the network. Because the SPF algorithm treats the LSPs in the same way as a physical interface (being a potential output interface), the computation results could be to select a destination node together with an output LSP, using the LSP as a shortcut through the network to the destination node.

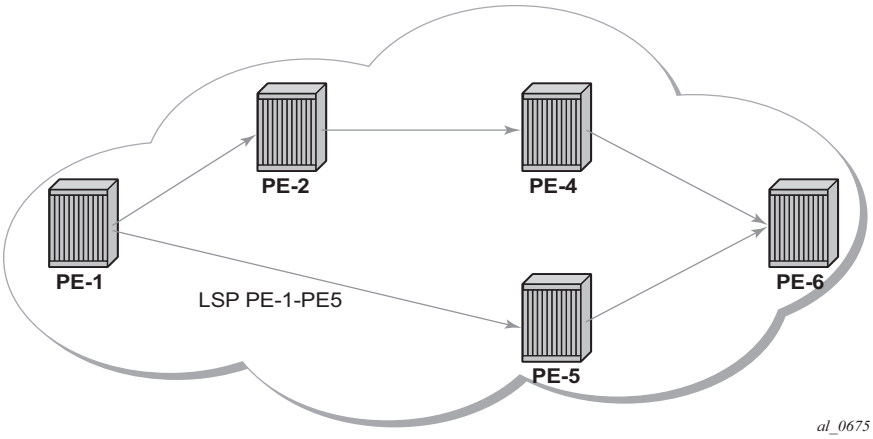
Figure 137 shows a normal SPF tree sourced by PE-1 (Provider Edge-1).

Figure 137 **Normal SPF Tree Sourced by PE-1**



If there is an LSP that connects PE-1 to PE-5, and IGP shortcuts are configured on PE-1, the SPF tree will be as shown in Figure 138.

Figure 138 **SPF Tree Sourced by PE-1 Using LSP Shortcuts**

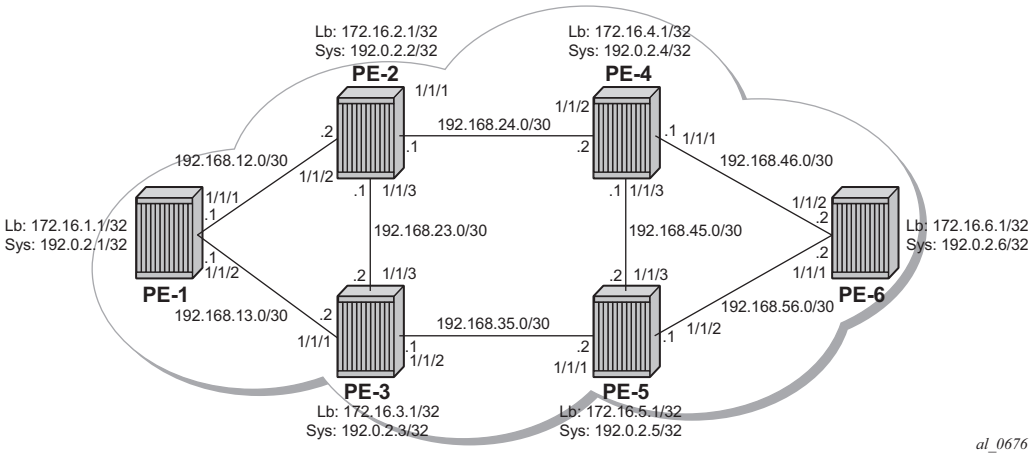


IGP shortcuts are enabled on a per router basis; SPF computations are independent and irrelevant to other routers, so there is no need to enable shortcuts on every single router.

The example topology used in this example is displayed in Figure 139. The setup consists of six 7750 service routers. There is a single AS and a single IGP area. The following configuration tasks should be completed first:

- IS-IS or OSPF on all interfaces within the AS (configuration has been done using IS-IS but using OSPF shows exactly the same behavior).
- Label distribution protocol (LDP) and resource reservation protocol (RSVP) on all interfaces within the AS.

Figure 139 **Example Topology**



In all figures, **Lb** stands for Loopback and **Sys** stands for the system IP addresses.

Configuration

The first step is to configure the IGP (IS-IS) on all nodes, where IS-IS redistributes route reachability to all routers. To facilitate IS-IS configuration, all routers are L2-L1 capable within the same IS-IS area-id so there is only a single topology area in the network (all routers share the same topology). Traffic engineering (TE) is enabled on the IGP as it is a requirement for RSVP. The metric is using the default values: since no reference bandwidth command is used, the default metric of 10 is applicable on all interfaces. The configuration for PE-2 is displayed below.

```
*A:PE-2# configure router
```

```
        interface "int-PE-2-PE-1"
            address 192.168.12.2/30
            port 1/1/2
        exit
        interface "int-PE-2-PE-3"
            address 192.168.23.1/30
            port 1/1/3
        exit
        interface "int-PE-2-PE-4"
            address 192.168.24.1/30
            port 1/1/1
        exit
        interface "system"
            address 192.0.2.2/32
        exit

*A:PE-2# configure router
        isis
            area-id 49.0001
            traffic-engineering
            interface "system"
                passive
            exit
            interface "int-PE-2-PE-1"
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-4"
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-3"
                interface-type point-to-point
            exit
```

The configuration for the other nodes is similar. The IP addresses can be derived from Figure 139.

The global route table (GRT) for PE-2 is as follows:

```
*A:PE-2# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                       Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                      Remote  ISIS      00h00m46s  15
       192.168.12.1                                                     10
192.0.2.2/32                                      Local   Local     00h02m00s  0
       system                                                           0
192.0.2.3/32                                      Remote  ISIS      00h00m38s  15
       192.168.23.2                                                     10
192.0.2.4/32                                      Remote  ISIS      00h00m23s  15
       192.168.24.2                                                     10
192.0.2.5/32                                      Remote  ISIS      00h00m18s  15
       192.168.23.2                                                     20
192.0.2.6/32                                      Remote  ISIS      00h00m08s  15
       192.168.24.2                                                     20
192.168.12.0/30                                   Local   Local     00h02m00s  0
```

```
       int-PE-2-PE-1                                                    0
192.168.13.0/30                          Remote  ISIS      00h00m46s  15
       192.168.12.1                                                     20
192.168.23.0/30                          Local   Local     00h02m00s  0
       int-PE-2-PE-3                                                    0
192.168.24.0/30                          Local   Local     00h02m00s  0
       int-PE-2-PE-4                                                    0
192.168.35.0/30                          Remote  ISIS      00h00m38s  15
       192.168.23.2                                                     20
192.168.45.0/30                          Remote  ISIS      00h00m23s  15
       192.168.24.2                                                     20
192.168.46.0/30                          Remote  ISIS      00h00m23s  15
       192.168.24.2                                                     20
192.168.56.0/30                          Remote  ISIS      00h00m17s  15
       192.168.23.2                                                     30
-------------------------------------------------------------------------------
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

LDP and RSVP Shortcuts

Interface label distribution protocol (iLDP) is enabled on all interfaces (except system interfaces, which is not allowed) in all routers. The configuration on all nodes is similar and the IP addresses are derived from Figure 139. The configuration of PE-4 is as follows:

```
*A:PE-4# configure router ldp
            interface-parameters
                interface "int-PE-4-PE-2"
                exit
                interface "int-PE-4-PE-5"
                exit
                interface "int-PE-4-PE-6"
                exit
            exit
```

With iLDP enabled, PE-4 establishes iLDP sessions with its directly connected neighbors, as follows:

```
*A:PE-4# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   1205      1204      0d 00:53:02
192.0.2.5:0         Link      Established   1198      1197      0d 00:52:55
192.0.2.6:0         Link      Established   181       183       0d 00:07:43
```

```
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 3
===============================================================================
===============================================================================
LDP IPv6 Sessions
===============================================================================
Peer LDP Id
 Adj Type            State          Msg Sent       Msg Recv       Up Time
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-4#
```

The following tunnel table shows that there is a label switched path (LSP) to every other router. The reason is that the LDP label distribution mode is downstream unsolicited (DU) by default, originating label bindings for system addresses only (which are used by iLDP as transport address by default). The command also shows the preference of the LSPs (where the preference is 9 for LDP) and the metric of the LSPs (the metric is inherited from the IGP, each hop counts as a metric of 10), as follows. The metric to destinations PE-1 and PE-3 is 20 because there are two hops in between (PE-4 is two hops away from PE-1 and PE-3).

```
*A:PE-4# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65538     9        192.168.24.1    20
192.0.2.2/32      ldp       MPLS  65537     9        192.168.24.1    10
192.0.2.3/32      ldp       MPLS  65539     9        192.168.24.1    20
192.0.2.5/32      ldp       MPLS  65540     9        192.168.45.2    10
192.0.2.6/32      ldp       MPLS  65545     9        192.168.46.2    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-4#
```

In order to configure RSVP shortcuts, RSVP must be enabled on all interfaces where traffic engineering is required. For simplicity, RSVP is configured on all interfaces of the network, including system interfaces. When RSVP is in **no shutdown**, it is automatically configured on the interfaces where MPLS is configured. The configuration for PE-6 is as follows.

```
*A:PE-6# configure router mpls no shutdown
*A:PE-6# configure router rsvp no shutdown

*A:PE-6# configure router
        mpls
            interface "int-PE-6-PE-4"
            exit
            interface "int-PE-6-PE-5"
            exit
```

The configuration of the other nodes is similar. The IP addresses can be derived from Figure 139. Because there are no RSVP LSPs configured yet, the tunnel-table has no RSVP LSPs and only contains LDP LSPs.

LDP Static Route (IP Tunneled in LDP Tunnel)

Using LDP LSP shortcuts for static route resolution enables forwarding of IPv4 packets over LDP LSPs instead of using a regular IP next hop. In other words, the traffic to the resolved static routes is forwarded using MPLS LDP LSP rather than plain IP.

The configuration defines a static route pointing to the destination PE (remote loopback, which is an indirect next hop in the example), and explicitly indicates that it should use LDP rather than IGP. Taking PE-1 and PE-6 as an example, two loopback interfaces are configured (172.16.X.1/32), where X = PE number, and a static route is defined according to the explanation above. The following shows the configuration on PE-1.

```
*A:PE-1# configure router
        interface "loopback"
            address 172.16.1.1/32
            loopback
        exit
        static-route-entry 172.16.6.1/32
            indirect 192.0.2.6
                tunnel-next-hop
                    resolution-filter
                        ldp
                    exit
                    disallow-igp
                    resolution filter
                exit
                no shutdown
            exit
        exit
```

Looking at the GRT or forwarding information base (FIB), there are two new entries corresponding to the two configured loopback interfaces. One entry is has the protocol set to local (the local loopback on the PE), and the other entry has the protocol set to static, where the next hop is reached using an LDP LSP.

```
*A:PE-1# show router fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                          Protocol
    NextHop
-------------------------------------------------------------------------------
172.16.1.1/32                                           LOCAL
```

```
    172.16.1.1 (loopback)
172.16.6.1/32                                                 STATIC
    192.0.2.6 (Transport:LDP)
---snip---
```

The next output shows that a ping sourced by the loopback interface on PE-1 is able to reach the loopback interface on PE-6, and traceroute demonstrates that the traffic is following the LDP LSP. The ping and traceroute traffic cannot follow the IGP path because the static route command states that the IGP is disallowed when no LDP LSP towards PE-6 is available (also, the loopback interfaces are not enabled on IS-IS).

```
*A:PE-1# ping 172.16.6.1 source 172.16.1.1
PING 172.16.6.1 56 data bytes
64 bytes from 172.16.6.1: icmp_seq=1 ttl=64 time=2.03ms.
64 bytes from 172.16.6.1: icmp_seq=2 ttl=64 time=2.16ms.
64 bytes from 172.16.6.1: icmp_seq=3 ttl=64 time=2.01ms.
64 bytes from 172.16.6.1: icmp_seq=4 ttl=64 time=2.78ms.
64 bytes from 172.16.6.1: icmp_seq=5 ttl=64 time=3.18ms.

---- 172.16.6.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.01ms, avg = 2.43ms, max = 3.18ms, stddev = 0.466ms

*A:PE-1# traceroute 172.16.6.1 source 172.16.1.1
traceroute to 172.16.6.1 from 172.16.1.1, 30 hops max, 40 byte packets
  1  0.0.0.0  * * *
  2  0.0.0.0  * * *
  3  172.16.6.1 (172.16.6.1)    1.69 ms  3.24 ms  2.46 ms
*A:PE-1#
```

With the traceroute command, there are three hops from PE-1 to PE-6. There is no information regarding IP for the first two hops because the traffic is encapsulated in an MPLS LSP. The reason why the hops are displayed even when there is an MPLS LSP tunnel is because by default, the SR router propagates (copies) the time to live (TTL) from the IP header in the MPLS header. This is known as uniform mode.

However, a service provider might not want to show how many MPLS hops (nodes) there are in their network if a **traceroute** command is executed from outside their network. To prevent internal hops being shown, **no propagate** commands are needed in the LDP configuration, as shown below. This is known as pipe mode.

```
*A:PE-1# configure router ldp
            no shortcut-local-ttl-propagate
            no shortcut-transit-ttl-propagate
            exit
```

Once TTL propagation is disabled, the hops are not displayed any longer when running the **traceroute** command.

```
*A:PE-1# traceroute 172.16.6.1 source 172.16.1.1
traceroute to 172.16.6.1 from 172.16.1.1, 30 hops max, 40 byte packets
```

```
  1  172.16.6.1 (172.16.6.1)    2.36 ms  2.24 ms  2.25 ms
*A:PE-1#
```

For more information about uniform mode and pipe mode, see chapter Tunneling of ICMP Reply Packets over MPLS LSPs.

RSVP Static Route (IP Tunneled in RSVP Tunnel)

Using RSVP LSP shortcuts for static route resolution enables forwarding of IPv4 packets over RSVP LSPs instead of using a regular IP next hop. In other words, the traffic to the resolved static routes is forwarded using an MPLS RSVP LSP rather than plain IP.

The configuration defines a static route pointing to a destination PE (remote loopback, which is an indirect next hop in the example), and explicitly indicates that it should use RSVP rather than IGP. Taking PE-6 and PE-1 as an example, two loopback interfaces are configured (172.16.X.1/32), where X = PE number, and a static route is defined according to the explanation above. The following shows the configuration on PE-6.

```
*A:PE-6# configure router
        interface "loopback"
            address 172.16.6.1/32
            loopback
            no shutdown
        exit
        static-route-entry 172.16.1.1/32
            indirect 192.0.2.1
                tunnel-next-hop
                    resolution-filter
                        rsvp-te
                        exit
                    exit
                    disallow-igp
                    resolution filter
                exit
                no shutdown
            exit
        exit
```

Also, an RSVP LSP needs to be configured with the system interface of PE-1 as the destination:

```
*A:PE-6# configure router mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-6-PE-1"
                to 192.0.2.1
                primary "loose"
```

```
            exit
            no shutdown
        exit
```

In the LSP tunnel table, an RSVP LSP is created:

```
*A:PE-6# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      rsvp      MPLS  1         7        192.168.46.1    30
192.0.2.1/32      ldp       MPLS  65553     9        192.168.46.1    30
192.0.2.2/32      ldp       MPLS  65552     9        192.168.46.1    20
192.0.2.3/32      ldp       MPLS  65540     9        192.168.56.1    20
192.0.2.4/32      ldp       MPLS  65551     9        192.168.46.1    10
192.0.2.5/32      ldp       MPLS  65541     9        192.168.56.1    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-6#
```

The default RSVP preference is 7 (preferred over that of LDP, which is 9) and the metric reflects that this LSP spans 3 hops (for a dynamic LSP not using constrained shortest path first (CSPF), the metric is inherited from IGP). See RSVP Shortcut for IGP Route Resolution for more details about the metric applied in LSPs.

The RSVP LSP is used to resolve the indirect next hop (PE-1 system address) in the static route (the LSP used is identified with the tunnel ID, in this case 1), hence the GRT is modified with the following entry:

```
*A:PE-6# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.1/32                                   Remote  Static    00h00m21s  5
       192.0.2.1 (tunneled:RSVP:1)                                     1

---snip---
-------------------------------------------------------------------------------
No. of Routes: 16
```

As in the LDP shortcut with static route example, between PE-6 and PE-1, TTL propagation is disabled.

```
*A:PE-6# configure router mpls
            no shortcut-local-ttl-propagate
            no shortcut-transit-ttl-propagate
```

The output is the following when running a traceroute:

```
*A:PE-6# traceroute 172.16.1.1 source 172.16.6.1
traceroute to 172.16.1.1 from 172.16.6.1, 30 hops max, 40 byte packets
  1  172.16.1.1 (172.16.1.1)    2.22 ms  2.04 ms  2.18 ms
*A:PE-6#
```

The two static routes that have been defined to use the LDP and RSVP shortcuts follow the static routes default values and have a preference of 5 and a metric of 1.

LDP Shortcut for IGP Route Resolution

Using LDP shortcuts for IGP route resolution enables forwarding of packets to IGP learned routes over an LDP LSP. The default is to disable the LDP shortcut across all interfaces in the node.

When LDP shortcuts are enabled, LDP populates the route table manager (RTM) with next hop entries corresponding to all prefixes for which it activated an LDP forwarding equivalence class (FEC). For a given prefix, two route entries are populated in RTM. One corresponds to the LDP shortcut next hop and has an owner of LDP. The other one is the regular IP next hop. The LDP shortcut next hop always takes preference over the regular IP next hop for forwarding user packets and specific control packets over a given outgoing interface to the route next hop.

Once LDP has activated a FEC for a given prefix and programmed the RTM, it also programs the ingress tunnel table in the line card with the LDP tunnel information.

When an IPv4 packet is received on an ingress network interface, a subscriber Internet enhanced service (IES) interface, or a regular IES interface, the lookup of the packet by the ingress line card results in the packet being sent labeled with the label stack corresponding to the next hop label forwarding entry (NHLFE) of the LDP LSP when the preferred RTM entry corresponds to an LDP shortcut. If the preferred RTM entry corresponds to an IP next hop, the IPv4 packet is forwarded unlabeled. The activation of the FEC by LDP is done by performing an exact match with an IGP route prefix in RTM but it can also be done by performing a longest prefix match with an IGP route in RTM if the aggregate-prefix-match option is enabled globally in LDP.

Handling of Control Packets

All control plane packets will not see the LDP shortcut route entry in RTM with the exception of the following control packets which will be forwarded over an LDP shortcut when enabled:

- A locally generated or in transit ICMP ping and UDP traceroute of an IGP route. The transit message appears as a user packet to the ingress LER node.
- A locally generated response to a received ICMP ping or UDP traceroute message.

All other control plane packets that require an RTM lookup and have knowledge of which destination is reachable over the LDP shortcut will continue to be forwarded over the IP next hop route in RTM.

Handling of Multicast Packets

LDP shortcuts apply to unicast FEC types and are used for forwarding IP unicast packets in the data path. IP multicast packets forwarded over an multicast label distribution protocol (mLDP) point-to-multipoint (P2MP) LSP make use of a multicast FEC and thus cannot make use of the LDP unicast shortcut.

ECMP Considerations

When equal cost multi-path (ECMP) is enabled and multiple equal cost next hops exist for the IGP route, the ingress line card will spray the packets for this route based on the hashing routine supported for IPv4 packets. When the preferred RTM entry corresponds to an LDP shortcut route, spraying is performed across the multiple next hops for the LDP FEC. The FEC next hops can either be direct link LDP neighbors or T-LDP (targeted LDP) neighbors reachable over RSVP LSPs in the case of LDP-over-RSVP but not both. This is as per ECMP for LDP in the existing implementation. When the preferred RTM entry corresponds to a regular IP route, spraying will be performed across regular IP next hops for the prefix. Spraying across regular IP next hops and LDP shortcut next hops concurrently is not supported.

Configuring IGP LDP shortcuts is straightforward, and only applies to the node where there is interest to provision the LDP shortcut. In this example, only PE-1 is provisioned with LDP shortcuts, as follows:

```
*A:PE-1#configure router ldp-shortcut
```

Now, all tunnel LSPs that resolve an IGP next hop will replace the IP next hops, as shown in the following output:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
```

```
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                 Local   Local     01h31m15s  0
       system                                                        0
192.0.2.2/32                                 Remote  LDP       00h04m15s  9
       192.168.12.2 (tunneled)                                       10
192.0.2.3/32                                 Remote  LDP       00h04m15s  9
       192.168.13.2 (tunneled)                                       10
192.0.2.4/32                                 Remote  LDP       00h04m15s  9
       192.168.12.2 (tunneled)                                       20
192.0.2.5/32                                 Remote  LDP       00h04m15s  9
       192.168.13.2 (tunneled)                                       20
192.0.2.6/32                                 Remote  LDP       00h04m15s  9
       192.168.12.2 (tunneled)                                       30
192.168.12.0/30                              Local   Local     01h31m15s  0
       int-PE-1-PE-2                                                 0
192.168.13.0/30                              Local   Local     01h31m15s  0
       int-PE-1-PE-3                                                 0
192.168.23.0/30                              Remote  ISIS      01h29m45s  15
       192.168.12.2                                                  20
192.168.24.0/30                              Remote  ISIS      01h29m45s  15
       192.168.12.2                                                  20
192.168.35.0/30                              Remote  ISIS      01h29m37s  15
       192.168.13.2                                                  20
192.168.45.0/30                              Remote  ISIS      01h29m22s  15
       192.168.12.2                                                  30
192.168.46.0/30                              Remote  ISIS      01h29m22s  15
       192.168.12.2                                                  30
192.168.56.0/30                              Remote  ISIS      01h29m16s  15
       192.168.13.2                                                  30
-------------------------------------------------------------------------------
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
*A:PE-1# show router fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                 Protocol
    NextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                                   LOCAL
    192.0.2.1 (system)
192.0.2.2/32                                                   LDP
    192.0.2.2 (Transport:LDP)
192.0.2.3/32                                                   LDP
    192.0.2.3 (Transport:LDP)
192.0.2.4/32                                                   LDP
    192.0.2.4 (Transport:LDP)
192.0.2.5/32                                                   LDP
    192.0.2.5 (Transport:LDP)
192.0.2.6/32                                                   LDP
    192.0.2.6 (Transport:LDP)
192.168.12.0/30                                                LOCAL
    192.168.12.0 (int-PE-1-PE-2)
```

```
192.168.13.0/30                                                 LOCAL
    192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.35.0/30                                                 ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.168.45.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.46.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.56.0/30                                                 ISIS
    192.168.13.2 (int-PE-1-PE-3)
-------------------------------------------------------------------------------
Total Entries : 14
-------------------------------------------------------------------------------
```

Applying LDP IGP shortcuts only on PE-1 implies that IP traffic from PE-1 to any of the system addresses of the rest of nodes will use the LDP shortcut, however, the traffic replied from any PE back to PE-1 will be native IP since IGP shortcuts have not been provisioned in the other nodes.

RSVP Shortcut for IGP Route Resolution

Using RSVP LSP shortcuts when resolving IGP routes enables forwarding of packets to IGP learned routes over an RSVP LSP. The use of RSVP shortcut for resolving IGP routes is enabled at the IS-IS (or OSPF) routing protocol level or at the LSP level, and instructs IS-IS and OSPF to include RSVP LSPs originating on this node and terminating on the system address (router-id) of a remote node and considers them as direct links. RSVP LSPs with a destination address corresponding to an interface address or any other loopback interface address of a remote node are automatically not considered by IS-IS or OSPF.

By default, **rsvp-shortcut** is disabled in all IGP instances.

RSVP LSPs are included in the IGP SPF computation with the following characteristics:

- RSVP LSP is modeled as a point-to-point link IP interface and its metric is used in the computation of the shortest path of IGP routes
- Next hop and interface include the NHLFE of the shortcut LSP when the IGP path cost using the RSVP LSP is the best.
- Shortcuts are not used when the destination RSVP LSP is in a different IGP area. In addition, IGP adjacencies across an RSVP LSP are not supported.

The next output shows the configuration commands:

```
configure router isis rsvp-shortcut
```

The configuration can be done at the IGP level or per LSP level. When rsvp-shortcut is enabled at the IGP instance level, all RSVP LSPs originating on this node are eligible by default. The user can, however, exclude a specific RSVP LSP from being used as a shortcut for resolving IGP routes by entering the command

```
*A:PE-6# configure router mpls lsp "LSP-PE-6-PE-1" no igp-shortcut
```

As RSVP shortcuts can coexist with LDP shortcuts or IP next hops, SPF computation and path selection follows the procedures in RFC 3906:

- SPF picks the RSVP shortcut next hop if there is an RSVP LSP directly to that address regardless of the path cost compared to the IGP next hop.
- SPF picks the RSVP shortcut next hop or the IGP next hop based on path lowest cost if there is an IGP path to the prefix that does not go via the tail-end of the LSP.
- If the IGP next hop is picked, then it can be an LDP shortcut next hop or a regular IP next hop. The LDP shortcut next hop always has preference over the regular IP next hop.

Handling of Control Packets

All control plane packets requiring an RTM lookup and whose destination is reachable over the RSVP shortcut are forwarded over the shortcut. This is because RTM keeps a single route entry for each prefix except if there is ECMP over different outgoing interfaces. Interface bound control packets are not impacted by the RSVP shortcut since RSVP LSPs with a destination address different than the router-id are not included by IGP in its SPF calculation.

Important note: RSVP shortcuts for IGP shortcut resolution should only be used with CSPF LSPs and/or with fully explicit path non-CSPF LSP. RSVP hop-by-hop Path messages will try to use the shortcut and consequently LSPs without CSPF enabled, or that use a loose/empty hop path, will not come up. However, LSPs with CSPF enabled or using a strict hop path will come up. This is because in the former case the RTM lookup to get the next hop results in using the shortcut and so the path messages are sent directly to the destination of the LSP, where they are dropped. With CSPF enabled, the next hop (and the entire path) is provided by CSPF and the path messages are sent unlabeled to the directly connected neighbor which corresponds to the next hop of the destination of the LSP. Similar processing occurs if a strict hop path is used, as is the case in the following example.

Handling of Multicast Packets

IP multicast packets cannot be forwarded over an RSVP shortcut, they can only be forwarded over an RSVP P2MP LSP. However, RSVP shortcut routes appear in RTM and are seen by all applications when they are the best route. When the reverse path forwarding (RPF) check for the source of the multicast packet matches an RSVP shortcut route, the check will pass if both the RSVP shortcut and the multicast-import options are enabled in the IGP, as shown below, as the RTM is populated with next hops only and not with tunnels (RPFs will fail for source prefixes resolved to a tunnel NH).

```
*A:PE-6# configure router isis multicast-import
  - multicast-import [both]
  - multicast-import [ipv4]
  - multicast-import [ipv6]
  - no multicast-import [both]
  - no multicast-import [ipv4]
  - no multicast-import [ipv6]

 <ipv4>               : keyword
 <ipv6>               : keyword
 <both>               : keyword

*A:PE-6#
```

The unicast RTM can still use the tunnel next hop for the same prefix. SPF keeps track of both the direct first hop and the tunneled first hop of a node which is added to the Dijkstra tree.

ECMP Considerations

When ECMP is enabled and multiple equal cost paths exist for the route over a set of tunnel next hops based on the hashing routine supported for IPv4 packets, there are two possibilities:

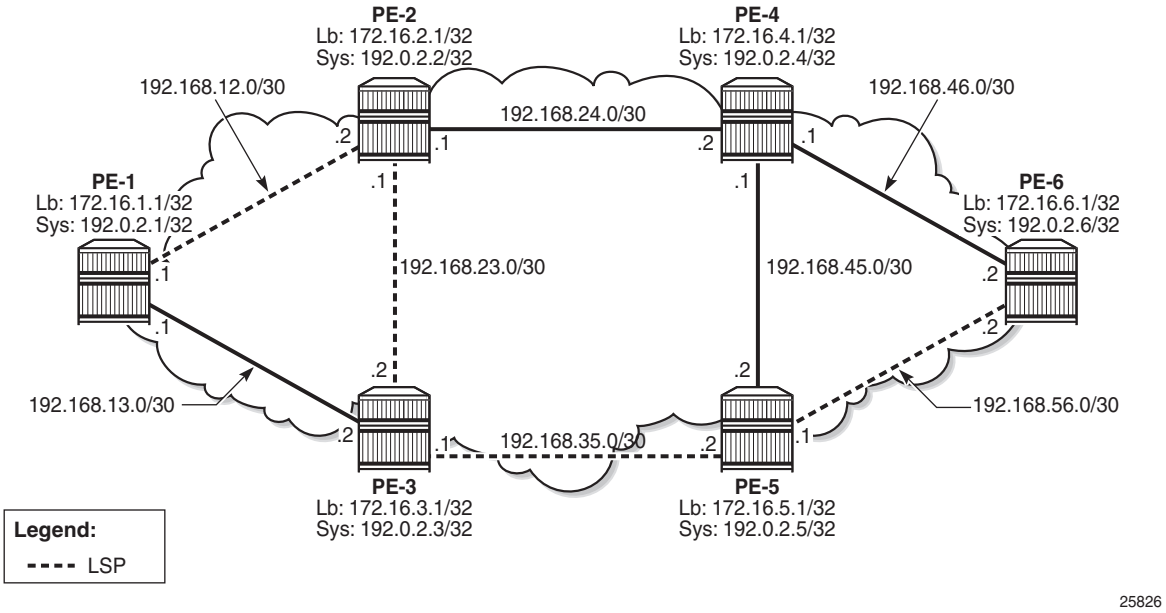
- Destination is tunnel endpoint: the system selects the tunnel with lowest tunnel ID (IP next hop is never used).
- Destination is different from the tunnel endpoint: it selects tunnel endpoints when the LSP metric is not greater than the IGP cost and it prefers tunnel endpoint over IP next hop.

ECMP is not performed across the IP and tunnel next hops simultaneously.

RSVP Shortcuts Configuration

Configuring RSVP LSP shortcuts is straightforward, and only applies to the node where there is interest to provision the RSVP shortcut. Two LSPs, from PE-6 to PE-1 and from PE-1 to PE-6, with strict hops, are provisioned according to Figure 140.

Figure 140 **LSPs Between PE-1 and PE-6**



The configuration on PE-1 and PE-6 is similar (replacing the IP addresses), so only the configuration for PE-6 is shown:

```
configure router isis rsvp-shortcut

configure router mpls
    path "path-to-PE-1"
        hop 10 192.0.2.5 strict
        hop 20 192.0.2.3 strict
        hop 30 192.0.2.2 strict
        hop 40 192.0.2.1 strict
        no shutdown
    exit
    lsp "LSP-PE-6-PE-1-strict"
        to 192.0.2.1
        primary "path-to-PE-1"
        exit
        no shutdown
    exit
```

The GRT output shows the change in the next hop, using an RSVP shortcut:

```
*A:PE-6# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m34s  15
       192.0.2.1 (tunneled:RSVP:2)                                  16777215
192.0.2.2/32                                    Remote  ISIS      00h00m34s  15
       192.168.46.1                                                 20
192.0.2.3/32                                    Remote  ISIS      00h00m34s  15
       192.168.56.1                                                 20
192.0.2.4/32                                    Remote  ISIS      00h00m34s  15
       192.168.46.1                                                 10
192.0.2.5/32                                    Remote  ISIS      00h00m34s  15
       192.168.56.1                                                 10
192.0.2.6/32                                    Local   Local     01h45m01s  0
       system                                                       0
192.168.12.0/30                                 Remote  ISIS      00h00m34s  15
       192.168.46.1                                                 30
192.168.13.0/30                                 Remote  ISIS      00h00m34s  15
       192.168.56.1                                                 30
192.168.23.0/30                                 Remote  ISIS      00h00m34s  15
       192.168.46.1                                                 30
192.168.24.0/30                                 Remote  ISIS      00h00m34s  15
       192.168.46.1                                                 20
192.168.35.0/30                                 Remote  ISIS      00h00m34s  15
       192.168.56.1                                                 20
192.168.45.0/30                                 Remote  ISIS      00h00m34s  15
       192.168.46.1                                                 20
192.168.46.0/30                                 Local   Local     01h45m02s  0
       int-PE-6-PE-4                                                0
192.168.56.0/30                                 Local   Local     01h45m01s  0
       int-PE-6-PE-5                                                0
-------------------------------------------------------------------------------
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-6#
```

The RSVP LSP in the output has a metric of 16777215, the LSP administrative metric matches the maximum value allowed for an IS-IS link using the wide-metric (24-bit value with a range of [0 — 16777215]). The following metric rules apply:

- A dynamic strict path non-CSPF LSP has the maximum metric (16777215).
- A dynamic CSPF LSP has a metric equal to the cumulative IGP cost.
  - If the user enabled the use of the TE metric on this LSP (configure router mpls lsp cspf use-te-metric), then the metric for the LSP is the maximum (16777215).

  - If the user enabled the use of the TE metric on this LSP (configure router mpls lsp cspf use-te-metric) and provisioned a specific metric on the lsp (configure router mpls lsp metric <metric>:<0..16777215>), then the metric for the LSP is the one provisioned. When configuring the metric of an LSP, the parameter “use-te-metric” is not required.
- A static LSP has a maximum metric (16777215).
- Manual and dynamic bypass LSPs have the maximum metric (16777215).

The RSVP shortcuts section detailed the importance of the LSP metric when using CSPF LSPs or when importing RSVP tunnel links into the IGP. The LSP metric can be inherited from the IGP, or can be manually modified by configuring a specific LSP metric or relative-metric offset. Because IP and LDP FECs resolve to RSVP LSPs when the metric is equal or lower compared to the regular routing metric, configuring a specific static LSP metric (lower than the IGP metric) or relative-metric offset is strongly recommended when using RSVP shortcuts, so that the GRT and LDP FEC resolution will always prefer the RSVP LSP shortcuts when the CSPF path computation is not using the shortest path.

For the example above, the first rule applies.

Advertising RSVP LSP Tunnel Links in the IGP: Forwarding Adjacency Feature

If configured, an RSVP LSP can also be advertised into the IGP similar to regular links so that other routers in the network can include that RSVP LSP into their SPF computations. The forwarding adjacency feature can be enabled independently from the RSVP shortcut feature in CLI. If both are configured for a given IGP instance, the forwarding adjacency takes precedence. An RSVP LSP must exist in the reverse direction in order for the advertised link to pass the bi-directional link check and be usable by other routers in the network. However, this is not required for the node which originates the LSP. The LSP is advertised as an unnumbered point-to-point link and the link state protocol data unit (LSP) and link state advertisement (LSA) have no traffic engineering opaque sub-TLVs as per RFC 3906.

Reusing the RSVP IGP shortcuts set up previously (PE-1 and PE-6 RSVP IGP shortcut example according to Figure 140), the outcome is a route linked with an RSVP LSP as next hop, as seen below:

```
*A:PE-6# show router route-table 192.0.2.1/32
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
```

```
192.0.2.1/32                                    Remote  ISIS      00h02m37s  15
       192.0.2.1 (tunneled:RSVP:2)                                    16777215
-------------------------------------------------------------------------------
No. of Routes: 1
```

The route tunneled through RSVP has a metric of 16777215, so it is not used by PE-6 GRT to reach any other routes since the metric is very high. After enabling the forwarding adjacency feature (tunnel links) to use shortcuts in the configuration, PE-1 and PE-6 have a direct connection through the RSVP LSP (as a virtual link). This configuration command must be executed in both routers.

```
configure router isis advertise-tunnel-link
```

Once the shortcut is advertised by IS-IS, the route will disappear from the RTM as the metric of the shortcut is greater than the IGP cost.

```
*A:PE-6# show router route-table 192.0.2.1/32
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m04s  15
       192.168.46.1                                                30
-------------------------------------------------------------------------------
No. of Routes: 1
```

If the LSP is reconfigured to use a metric equal to or smaller than the IGP cost, the router PE-6 will use the RSVP shortcut again. In the example, the LSP is reconfigured with a metric of 30:

```
*A:PE-6# configure router mpls lsp "LSP-PE-6-PE-1-strict" metric 30
```

Now the shortcut shows up as the preferred next hop to reach PE-1 from PE-6.

```
*A:PE-6# show router route-table 192.0.2.1/32
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m06s  15
       192.0.2.1 (tunneled:RSVP:2)                                 30
-------------------------------------------------------------------------------
No. of Routes: 1
```

As explained earlier, this could be combined together with ECMP, so if ECMP is configured to 2, the system shows the two equal cost paths.

```
*A:PE-6# configure router ecmp 2
```

```
*A:PE-6# show router route-table 192.0.2.1/32
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m05s  15
       192.0.2.1 (tunneled:RSVP:2)                                    30
192.0.2.1/32                                    Remote  ISIS      00h00m05s  15
       192.168.46.1                                                   30
-------------------------------------------------------------------------------
No. of Routes: 2
```

Checking GRT on PE-4, it displays the route to reach PE-1 (192.0.2.1/32) with a metric of 20 via PE-2 as next hop. Although now PE-6 is announcing the RSVP LSP-PE-6-PE-1 to the other routers, the LSP shortcut is not used by PE-4 because the metric to reach PE-6 (10) plus the metric of the LSP shortcut from PE-6 to PE-1 (metric 30) is greater than 20.

```
*A:PE-4# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      01h57m14s  15
       192.168.24.1                                                   20
192.0.2.2/32                                    Remote  ISIS      01h57m14s  15
       192.168.24.1                                                   10
192.0.2.3/32                                    Remote  ISIS      01h57m14s  15
       192.168.24.1                                                   20
192.0.2.4/32                                    Local   Local     01h58m26s  0
       system                                                         0
192.0.2.5/32                                    Remote  ISIS      01h57m08s  15
       192.168.45.2                                                   10
192.0.2.6/32                                    Remote  ISIS      01h57m01s  15
       192.168.46.2                                                   10
192.168.12.0/30                                 Remote  ISIS      01h57m14s  15
       192.168.24.1                                                   20
192.168.13.0/30                                 Remote  ISIS      01h57m14s  15
       192.168.24.1                                                   30
192.168.23.0/30                                 Remote  ISIS      01h57m14s  15
       192.168.24.1                                                   20
192.168.24.0/30                                 Local   Local     01h58m26s  0
       int-PE-4-PE-2                                                  0
192.168.35.0/30                                 Remote  ISIS      01h57m08s  15
       192.168.45.2                                                   20
192.168.45.0/30                                 Local   Local     01h58m26s  0
       int-PE-4-PE-5                                                  0
192.168.46.0/30                                 Local   Local     01h58m26s  0
       int-PE-4-PE-6                                                  0
192.168.56.0/30                                 Remote  ISIS      01h57m08s  15
       192.168.45.2                                                   20
-------------------------------------------------------------------------------
```

```
No. of Routes: 14
```

If the metric of the LSP LSP-PE-6-PE-1 is modified to a value between 1 and 9, there is a better metric (less than 20) so that PE-4 will change the next hop via PE-6. First the metric of the LSP is modified to 9:

```
*A:PE-6# configure router mpls lsp "LSP-PE-6-PE-1-strict" metric 9
```

And checking PE-4´s GRT the next hop to reach PE-1 has changed, from next hop PE-2 to next hop PE-6 (hence, using the LSP shortcut), and the metric is 19 (10 to reach PE-6 plus metric 9 of the LSP PE-6-PE-1 shortcut):

```
*A:PE-4# show router route-table 192.0.2.1/32
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m07s  15
       192.168.46.2                                                 19
-------------------------------------------------------------------------------
No. of Routes: 1
```

Because the metric of the LSP shortcut was modified to a value of 9, when displaying the GRT of PE-6 it is noted that the next hops of several routes have changed and are also using the shortcut LSP PE-6-PE-1 because the metric is better than the regular IS-IS metric. It is important to emphasize that IGP shortcuts will not be used to resolve prefixes downstream of the LSP endpoint when the LSP metric is higher than the underlying IGP cumulative metric.

```
*A:PE-6# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h01m15s  15
       192.0.2.1 (tunneled:RSVP:2)                                  9
192.0.2.2/32                                    Remote  ISIS      00h01m15s  15
       192.0.2.1 (tunneled:RSVP:2)                                  19
192.0.2.3/32                                    Remote  ISIS      00h01m15s  15
       192.0.2.1 (tunneled:RSVP:2)                                  19
192.0.2.4/32                                    Remote  ISIS      00h11m25s  15
       192.168.46.1                                                 10
192.0.2.5/32                                    Remote  ISIS      00h11m25s  15
       192.168.56.1                                                 10
192.0.2.6/32                                    Local   Local     02h01m04s  0
       system                                                       0
192.168.12.0/30                                 Remote  ISIS      00h01m15s  15
       192.0.2.1 (tunneled:RSVP:2)                                  19
192.168.13.0/30                                 Remote  ISIS      00h01m15s  15
       192.0.2.1 (tunneled:RSVP:2)                                  19
```

```
192.168.23.0/30                                 Remote  ISIS      00h01m15s  15
       192.0.2.1 (tunneled:RSVP:2)                                     29
192.168.24.0/30                                 Remote  ISIS      00h11m25s  15
       192.168.46.1                                                    20
192.168.35.0/30                                 Remote  ISIS      00h11m25s  15
       192.168.56.1                                                    20
192.168.45.0/30                                 Remote  ISIS      00h10m34s  15
       192.168.46.1                                                    20
192.168.46.0/30                                 Local   Local     02h01m04s  0
       int-PE-6-PE-4                                                   0
192.168.56.0/30                                 Local   Local     02h01m04s  0
       int-PE-6-PE-5                                                   0
-------------------------------------------------------------------------------
No. of Routes: 14
```

There are also cases where an LDP FEC can resolve to an RSVP LSP, if the user enables the LDP-over-RSVP feature or IGP shortcut feature when prefer-tunnel-in-tunnel is enabled in LDP and the endpoint of the RSVP LSP matches the FEC prefix. For those cases, the metric to the prefix is the sum of the RSVP LSP metric + remaining IGP path cost.

Table 14 provides a summary of the outcome when configuring the forwarding adjacency, LDPoRSVP and RSVP shortcut options at both the IGP instance level and at the LSP level.

Table 14 **RSVP LSP Role As Outcome of LSP Level and IGP Level Configuration Options**

	IGP Instance Level Configurations					
LSP Level Configuration	advertise-tunnel-link enabled/ rsvp-shortcut enabled/ldp-over-rsvp enabled	advertise-tunnel-link enabled/ rsvp-shortcut enabled/ldp-over-rsvp disabled	advertise-tunnel-link enabled/ rsvp-shortcut disabled/ldp-over-rsvp disabled	advertise-tunnel-link disabled/ rsvp-shortcut disabled/ldp-over-rsvp disabled	advertise-tunnel-link disabled/ rsvp-shortcut enabled/ldp-over-rsvp enabled	advertise-tunnel-link disabled/ rsvp-shortcut disabled/ldp-over-rsvp enabled
igp-shortcut enabled/ldp-over-rsvp enabled	Forwarding Adjacency	Forwarding Adjacency	Forwarding Adjacency	None	IGP Shortcut	LDP-over-RSVP
igp-shortcut enabled/ldp-over-rsvp disabled	Forwarding Adjacency	Forwarding Adjacency	Forwarding Adjacency	None	IGP Shortcut	None
igp-shortcut disabled/ldp-over-rsvp enabled	None	None	None	None	None	LDP-over-RSVP

Table 14 **RSVP LSP Role As Outcome of LSP Level and IGP Level Configuration Options**

	IGP Instance Level Configurations					
igp-shortcut disabled/ldp-over-rsvp disabled	None	None	None	None	None	None

LSP Relative Metric

It is possible to use relative metrics for IGP shortcuts as per RFC 3906, *Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels*, with the following command:

```
*A:PE-6# configure router mpls lsp "LSP-PE-6-PE-1-loose" igp-shortcut relative-metric
  - igp-shortcut [lfa-protect | lfa-only][relative-metric [offset]]
  - no igp-shortcut

 <lfa-protect>        : keyword
 <lfa-only>           : keyword
 <relative-metric>    : keyword
 <offset>             : [-10..10]
```

When this feature is enabled, IGP applies the shortest IGP cost between the endpoints of the LSP, plus the value of a configured offset when computing the cost of the prefix that is resolved to the LSP.

The offset value is optional and can have a value between -10 and 10, and defaults to zero (0). An offset value of zero (0) is used when the relative-metric option is enabled without specifying the offset parameter value. The minimum net cost for the prefix is capped to the value of one (1) after applying the offset:

Prefix cost = max (1, IGP Cost + relative metric offset)

The **relative-metric** option is ignored when advertise-tunnel-link is enabled in IS-IS or OSPF. In that case, the IGP advertises the LSP as a P2P unnumbered link using the LSP operational metric.

The **relative-metric** option is mutually exclusive with the **lfa-protect** (loop-free alternat (LFA)) or the **lfa-only** options. An LSP with **relative-metric** option enabled cannot be included in the LFA SPF and vice-versa when the **rsvp-shortcut** option is enabled in the IGP (see chapter LDP/IP FRR LFA for IGP Shortcut Using IS-IS/OSPF for more information).

The offset can be used to enforce the preference of the shortcut path over the other paths for the prefix. Using an example, a new CSPF LSP with empty path and relative metric of -10 is created between PE-6 and PE-1. While the operational or absolute metric is 30 (IGP cost and populated in the Tunnel Table Manager, TTM), the metric that the RTM shows is 20 after applying the offset:

```
*A:PE-6# configure router mpls
            lsp "LSP-PE-6-PE-1-loose"
                to 192.0.2.1
                cspf
                igp-shortcut relative-metric -10
                primary "loose"
                exit
                no shutdown
            exit
*A:PE-6# show router tunnel-table 192.0.2.1
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      rsvp      MPLS  2         7        192.168.56.1    16777215
192.0.2.1/32      rsvp      MPLS  3         7        192.168.56.1    30
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-6#
*A:PE-6# show router route-table 192.0.2.1
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m07s  15
       192.0.2.1 (tunneled:RSVP:3)                                    20
-------------------------------------------------------------------------------
No. of Routes: 1
```

LDP/IP FRR LFA for IGP Shortcut Using IS-IS/OSPF

MPLS LDP/IP FRR LFA for IGP shortcuts allows for the use of RSVP-LSP-based IGP shortcuts as a Loop-Free Alternate (LFA) backups, this way expanding the coverage of the IP Fast-Reroute (FRR) and the LDP FRR capabilities for IS-IS and OSPF prefixes. For a detailed description about IP and LDP FRR, see chapter MPLS LDP FRR using ISIS as IGP.

When an RSVP LSP is used as a shortcut by IS-IS or OSPF, it is included by the SPF as a P2P link and it can also be optionally advertised into the rest of the network by the IGP.

Two LSP-level configuration options are provided:

- The lfa-protect option includes the RSVP LSP in both the main SPF and the LFA SPFs. If the prefix primary next hop (NH) is tunneled, no LFA NH is computed. The protection in this case is provided by RSVP FRR. If the prefix primary NH is direct, then an LFA NH is computed. A direct LFA NH is preferred over a tunneled LFA NH. Within each LFA NH type, node-protection is preferred over link-protection. The configuration command is:

```
configure router mpls lsp <lsp-name> igp-shortcut lfa-protect
```

- The lfa-only option includes the LSP in the LFA SPFs only so that the introduction of IGP shortcuts does not impact the main SPF decision. The prefix primary NH is always direct and the prefix LFA NH is computed. A direct LFA NH is preferred over a tunneled LFA NH. Within each LFA NH type, node-protection is preferred over link-protection. The configuration command is:

```
configure router mpls lsp <lsp-name> igp-shortcut lfa-only
```

LDP/IP FRR is a local decision, so it can be enabled per node and there are no interoperability issues with other nodes. In the topology, PE-2 is provisioned with IS-IS LFA (OSPF configuration for the rest of this section is similar):

```
A:PE-2# configure router isis loopfree-alternate
```

The second item to configure is whether LDP or IP FRR is provisioned. To configure ip-fast-reroute, the command is:

```
A:PE-2# configure router ip-fast-reroute
```

IP FRR feature for IS-IS/OSPF prefixes is supported on 7950 XRS, on 7750 SR-7/12/12e in chassis mode D, on the 7450 ESS-6/6v/7/12 in chassis mode D with or without mixed-mode, and 7750 SR-c4/c12.

To configure LDP FRR (no chassis dependency), this command is used:

```
A:PE-2# configure router ldp fast-reroute
```

Although not shown, it is recommended to enable IGP-LDP synchronization per interface to avoid possible traffic blackholes.

Once LFA is enabled in all routers of the topology, looking at PE-2 (the configuration is done only on this node), the following command shows the LFA coverage where 4/5 nodes and 7/10 IPv4 prefixes are protected (IPv6 is not configured). The output shows L1 and L2 because this node is provisioned as an L1-L2 IS-IS router. PE-2, PE-3, PE-4 and PE-5 share the same results, whereas only PE-1 and PE-6 have a 100% of coverage.

```
*A:PE-2# show router isis lfa-coverage
===============================================================================
Router Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      4/5(80%)       7/10(70%)           0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      4/5(80%)       7/10(70%)           0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-2#

*A:PE-1# show router isis lfa-coverage
===============================================================================
Router Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      5/5(100%)      11/11(100%)         0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      5/5(100%)      11/11(100%)         0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-1#
```

Taking a deeper look into the IS-IS LFA on PE-2, it can be seen that the node which is not protected is PE-4 (system address 192.0.2.4, since it is the one missing):

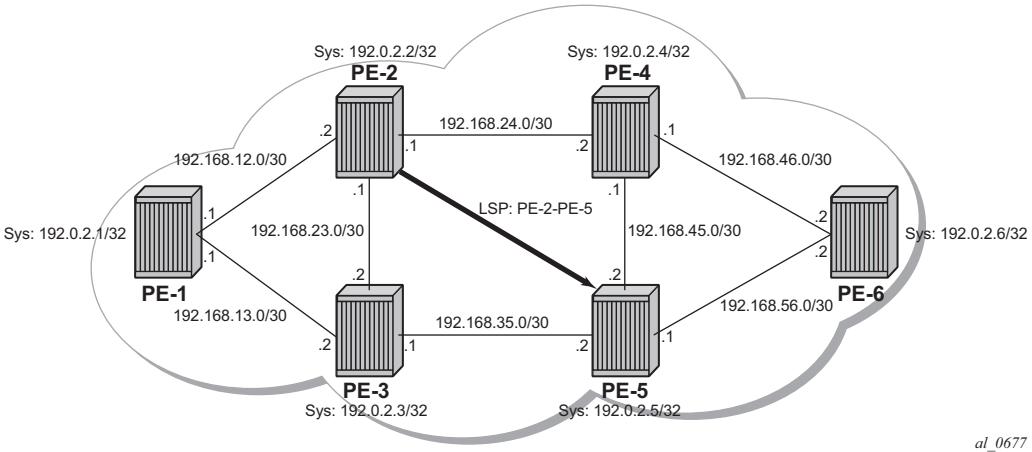
```
*A:PE-2# show router route-table alternative | match LFA pre-lines 2
192.0.2.1/32                                  Remote  ISIS      02h22m46s  15
       192.168.12.1                                                 10
       192.168.23.2 (LFA)                                           20
192.0.2.3/32                                  Remote  ISIS      02h22m38s  15
       192.168.23.2                                                 10
       192.168.12.1 (LFA)                                           20
192.0.2.5/32                                  Remote  ISIS      02h22m18s  15
       192.168.23.2                                                 20
       192.168.24.2 (LFA)                                           20
192.0.2.6/32                                  Remote  ISIS      02h22m08s  15
       192.168.24.2                                                 20
       192.168.23.2 (LFA)                                           30
192.168.13.0/30                               Remote  ISIS      02h22m46s  15
       192.168.12.1                                                 20
       192.168.23.2 (LFA)                                           30
192.168.35.0/30                               Remote  ISIS      02h22m38s  15
       192.168.23.2                                                 20
       192.168.12.1 (LFA)                                           30
192.168.56.0/30                               Remote  ISIS      02h22m17s  15
       192.168.23.2                                                 30
```

```
       192.168.24.2 (LFA)                                          30
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
*A:PE-2#
```

LFA is improved by taking advantage of RSVP shortcuts when it is properly provisioned. The reason why PE-4 cannot be protected with an LFA path is because the direct NH is using the direct link between PE-2 and PE-4 (the shortest IGP) and the intended LFA path through PE-3 is not valid (when LFA tries to find an alternate path via PE-3, the IGP cost from PE-3 to PE-4 is the same going via PE-5 as the path back via PE-2, invalidating that LFA calculation because there is a loop). This is normal because PE-2, PE-3, PE-4 and PE-5 are forming a ring. LFA coverage is increased by adding a link between PE-2 and PE-5, which can be done using a physical link or a virtual link with an RSVP shortcut. From the two possible options (lfa-only and lfa-protect), a new LSP “LSP-PE-2-PE-5” is configured with igp-shortcut lfa-only.

```
*A:PE-2# configure router mpls
            path "path-to-PE-5"
                hop 10 192.0.2.3 strict
                hop 20 192.0.2.5 strict
                no shutdown
            exit
            lsp "LSP-PE-2-PE-5"
                to 192.0.2.5
                igp-shortcut lfa-only
                primary "path-to-PE-5"
                exit
                no shutdown
            exit
```

Figure 141 **RSVP Shortcuts LFA Use Case Example**



Now the coverage is 100% as shown by the output:

```
*A:PE-2# show router isis lfa-coverage
===============================================================================
Router Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      5/5(100%)      10/10(100%)         0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      5/5(100%)      10/10(100%)         0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-2#
```

The GRT details the prefix information after the new LFA calculation using the lfa-only option (the shortcut is used by LFA SPF). The metric from PE-2 to PE-4 is the maximum plus the IGP cost (16777215 + 10) and the shortcut is also used to protect the rest of the previously unprotected prefixes:

```
*A:PE-2# show router route-table alternative | match LFA pre-lines 2
192.0.2.1/32                                    Remote  ISIS      02h27m59s  15
       192.168.12.1                                                     10
       192.168.23.2 (LFA)                                               20
192.0.2.3/32                                    Remote  ISIS      02h27m51s  15
       192.168.23.2                                                     10
       192.168.12.1 (LFA)                                               20
192.0.2.4/32                                    Remote  ISIS      02h27m36s  15
       192.168.24.2                                                     10
       192.0.2.5 (LFA) (tunneled:RSVP:1)                                16777225
192.0.2.5/32                                    Remote  ISIS      02h27m31s  15
       192.168.23.2                                                     20
       192.168.24.2 (LFA)                                               20
192.0.2.6/32                                    Remote  ISIS      02h27m21s  15
       192.168.24.2                                                     20
       192.168.23.2 (LFA)                                               30
192.168.13.0/30                                 Remote  ISIS      02h27m59s  15
       192.168.12.1                                                     20
       192.168.23.2 (LFA)                                               30
192.168.35.0/30                                 Remote  ISIS      02h27m51s  15
       192.168.23.2                                                     20
       192.168.12.1 (LFA)                                               30
192.168.45.0/30                                 Remote  ISIS      02h27m36s  15
       192.168.24.2                                                     20
       192.0.2.5 (LFA) (tunneled:RSVP:1)                                16777235
192.168.46.0/30                                 Remote  ISIS      02h27m36s  15
       192.168.24.2                                                     20
       192.0.2.5 (LFA) (tunneled:RSVP:1)                                16777235
192.168.56.0/30                                 Remote  ISIS      02h27m30s  15
       192.168.23.2                                                     30
       192.168.24.2 (LFA)                                               30
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
```

```
       LFA = Loop-Free Alternate nexthop
*A:PE-2#
```

The tunnel table shows the RSVP LSP used as a shortcut and its operational metric.

```
*A:PE-2# show router tunnel-table 192.0.2.5
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.5/32      rsvp      MPLS  1         7        192.168.23.2    16777215
192.0.2.5/32      ldp       MPLS  65540     9        192.168.23.2    20
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-2#
```

If the LSP “LSP-PE-2-PE-5” is provisioned with lfa-protect instead of lfa-only, the result is that the LSP "LSP-PE-2-PE-5“ is used by normal SPF to define the primary NH and it is not used by LFA SPF anymore.

```
A:PE-2# configure router mpls lsp "LSP-PE-2-PE-5" igp-shortcut lfa-protect
```

The coverage when lfa-protect is used also shows a 100% for nodes. The 112% coverage for prefixes shown below is not correct and should be 100% since the IGP shortcut used as a primary next hop to reach PE-5 from PE-2 is providing protection for prefixes reachable via PE-5. The display issue will be fixed in a later release and will be referenced by [201872] in the release notes.

```
*A:PE-2# show router isis lfa-coverage
===============================================================================
Router Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      4/4(100%)      9/8(112%)           0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      4/4(100%)      9/8(112%)           0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-2#
```

In this case, the GRT looks as follows, the main difference being that now PE-5 (192.0.2.5) has a direct shortcut from PE-2:

```
*A:PE-2# show router route-table alternative
```

```
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
      Alt-NextHop                                                 Alt-
                                                                  Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      02h32m17s  15
       192.168.12.1                                                 10
       192.168.23.2 (LFA)                                           20
192.0.2.2/32                                    Local   Local     02h33m31s  0
       system                                                       0
192.0.2.3/32                                    Remote  ISIS      02h32m09s  15
       192.168.23.2                                                 10
       192.168.12.1 (LFA)                                           20
192.0.2.4/32                                    Remote  ISIS      02h31m54s  15
       192.168.24.2                                                 10
       192.0.2.5 (LFA) (tunneled:RSVP:1)                            16777225
192.0.2.5/32                                    Remote  ISIS      00h01m56s  15
       192.0.2.5 (tunneled:RSVP:1)                                  16777215
192.0.2.6/32                                    Remote  ISIS      02h31m39s  15
       192.168.24.2                                                 20
       192.168.23.2 (LFA)                                           30
192.168.12.0/30                                 Local   Local     02h33m31s  0
       int-PE-2-PE-1                                                0
192.168.13.0/30                                 Remote  ISIS      02h32m17s  15
       192.168.12.1                                                 20
       192.168.23.2 (LFA)                                           30
192.168.23.0/30                                 Local   Local     02h33m31s  0
       int-PE-2-PE-3                                                0
192.168.24.0/30                                 Local   Local     02h33m31s  0
       int-PE-2-PE-4                                                0
192.168.35.0/30                                 Remote  ISIS      02h32m09s  15
       192.168.23.2                                                 20
       192.168.12.1 (LFA)                                           30
192.168.45.0/30                                 Remote  ISIS      02h31m54s  15
       192.168.24.2                                                 20
       192.0.2.5 (LFA) (tunneled:RSVP:1)                            16777235
192.168.46.0/30                                 Remote  ISIS      02h31m54s  15
       192.168.24.2                                                 20
       192.0.2.5 (LFA) (tunneled:RSVP:1)                            16777235
192.168.56.0/30                                 Remote  ISIS      00h01m56s  15
       192.168.24.2                                                 30
       192.168.23.2 (LFA)                                           40
-------------------------------------------------------------------------------
No. of Routes: 14
```

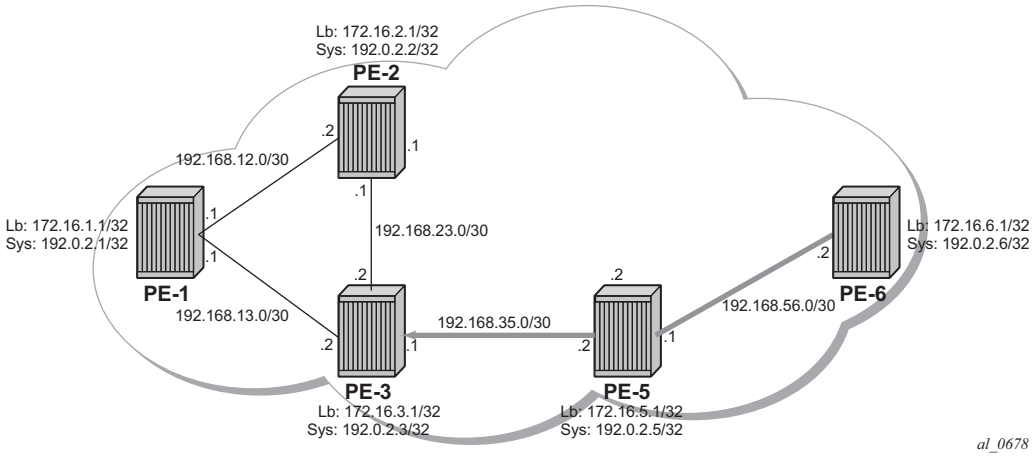
Rules Determining the Installation of Shortcuts into RTM

Although it was already mentioned in the RSVP-TE LSP shortcut for IGP route resolution section, the rules determining how shortcuts are installed into RTM are (sorted by higher priority):

- RSVP shortcut.
- LDP shortcut.
- IGP route with regular IP next hop.
- The implementation is compliant with RFC3906.

To check the rules, the network configuration is iLDP in all interfaces with LDP shortcuts enabled, there is also an RSVP LSP from PE-6 to PE-3 available but RSVP shortcuts are disabled. The topology is shown in Figure 142.

Figure 142 **Network Topology to Verify Installation of Shortcuts into RTM**



The following RSVP LSP is needed between PE-6 and PE-3.

```
configure router ldp-shortcut

configure router isis no rsvp-shortcut

configure router mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-6-PE-3"
                to 192.0.2.3
                cspf
                primary "loose"
                exit
                no shutdown
            exit
```

Displaying relevant info in PE-6, the routes are:

```
*A:PE-6# show router route-table
===============================================================================
Route Table (Router: Base)
```

```
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  LDP       00h00m03s  9
       192.168.56.1 (tunneled)                                         30
192.0.2.2/32                                    Remote  LDP       00h00m03s  9
       192.168.56.1 (tunneled)                                         30
192.0.2.3/32                                    Remote  LDP       00h00m03s  9
       192.168.56.1 (tunneled)                                         20
192.0.2.5/32                                    Remote  LDP       00h20m48s  9
       192.168.56.1 (tunneled)                                         10
192.0.2.6/32                                    Local   Local     02h57m46s  0
       system                                                          0
192.168.12.0/30                                 Remote  ISIS      00h00m03s  15
       192.168.56.1                                                    40
192.168.13.0/30                                 Remote  ISIS      00h00m03s  15
       192.168.56.1                                                    30
192.168.23.0/30                                 Remote  ISIS      00h00m03s  15
       192.168.56.1                                                    30
192.168.35.0/30                                 Remote  ISIS      00h00m03s  15
       192.168.56.1                                                    20
192.168.56.0/30                                 Local   Local     02h57m46s  0
       int-PE-6-PE-5                                                   0
-------------------------------------------------------------------------------
No. of Routes: 10
```

The tunnel table shows the LSPs available for the shortcuts, and therefore these are used in the GRT for LDP (but not for RSVP):

```
*A:PE-6# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65564     9        192.168.56.1    30
192.0.2.2/32      ldp       MPLS  65565     9        192.168.56.1    30
192.0.2.3/32      rsvp      MPLS  4         7        192.168.56.1    20
192.0.2.3/32      ldp       MPLS  65566     9        192.168.56.1    20
192.0.2.5/32      ldp       MPLS  65541     9        192.168.56.1    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-6#
```

So far, LDP shortcuts are preferred over the IGP next hops for the system addresses (router-id). After enabling RSVP shortcuts in the IS-IS context, the changes in the GRT are:

```
*A:PE-6# configure router isis rsvp-shortcut

*A:PE-6# show router route-table next-hop-type tunneled
===============================================================================
Route Table (Router: Base)
```

```
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h00m08s  15
       192.0.2.3 (tunneled:RSVP:4)                                     30
192.0.2.2/32                                    Remote  ISIS      00h00m08s  15
       192.0.2.3 (tunneled:RSVP:4)                                     30
192.0.2.3/32                                    Remote  ISIS      00h00m08s  15
       192.0.2.3 (tunneled:RSVP:4)                                     20
192.0.2.5/32                                    Remote  LDP       00h23m42s  9
       192.168.56.1 (tunneled)                                         10
192.168.12.0/30                                 Remote  ISIS      00h00m08s  15
       192.0.2.3 (tunneled:RSVP:4)                                     40
192.168.13.0/30                                 Remote  ISIS      00h00m08s  15
       192.0.2.3 (tunneled:RSVP:4)                                     30
192.168.23.0/30                                 Remote  ISIS      00h00m08s  15
       192.0.2.3 (tunneled:RSVP:4)                                     30
-------------------------------------------------------------------------------
No. of Routes: 7
```

The GRT shows that PE-6 is using an LDP shortcut to reach PE-5, but PE-6 is using the RSVP shortcut to reach not only PE-3's system address, but also PE-1 and PE-2 routes (including all interfaces) which were behind the RSVP LSP shortcut.

In summary, the behavior is:

1. When resolving a prefix, SPF picks the RSVP shortcut next hop if there is an RSVP LSP directly to that address regardless of the IGP path cost compared to the IGP next hop. When multiple RSVP LSPs to that address exist and all have the same lowest metric, if ECMP is enabled on the system, the LSP with the lowest tunnel ID is chosen. In this example, if LSP "LSP-PE-6-PE-3" is provisioned with a metric of 100 (IGP metric is 20), the GRT shows that the PE-3 system address is reachable via the LSP.

```
*A:PE-6# show router route-table 192.0.2.3
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.3/32                                    Remote  ISIS      00h00m12s  15
       192.0.2.3 (tunneled:RSVP:4)                                     100
-------------------------------------------------------------------------------
No. of Routes: 1
```

2. SPF also picks the RSVP LSP shortcut if both the LSP path and the IGP path to the prefix are via the tail-end of the LSP. This is regardless of the path cost compared to the IGP next hop. When paths over multiple RSVP shortcuts have the same lowest cost, if ECMP is enabled on the system, the LSP with the lowest tunnel ID is chosen. In this example, 192.168.13.0 and 192.168.23.0 are using the shortcut but 192.168.12.0 is not.

```
*A:PE-6# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
---snip---

192.168.12.0/30                                 Remote  ISIS      00h00m46s  15
       192.168.56.1                                                  40
192.168.13.0/30                                 Remote  ISIS      00h00m46s  15
       192.0.2.3 (tunneled:RSVP:4)                                   110
192.168.23.0/30                                 Remote  ISIS      00h00m46s  15
       192.0.2.3 (tunneled:RSVP:4)                                   110

---snip---
-------------------------------------------------------------------------------
No. of Routes: 10
```

LDP/RSVP LSP Shortcut for BGP NH Resolution

Using LDP/RSVP LSP shortcuts for resolving BGP next hops allows IPv4 packet forwarding to routes resolved via a BGP next hop using an LDP/RSVP LSP instead of using a regular IP next hop. In the network topology of Figure 139, both PE-3 and PE-6 have a single peer configured, initially without any shortcuts enabled under the BGP context. Also, one static route is configured in PE-3 and PE-6 and that is redistributed into BGP. The relevant configuration on PE-3 is the following:

```
*A:PE-3# configure router
        interface "static-route"
            address 172.16.33.1/30
            port 1/1/4:33
        exit
        autonomous-system 65536

        static-route-entry 10.10.10.0/24
            next hop 172.16.33.2
                no shutdown
            exit
        exit

        policy-options
            begin
            policy-statement "static-routes"
                description "export static-routes for I-BGP"
                entry 10
                    from
                        protocol static
                    exit
                    to
                        protocol bgp
                    exit
```

```
                action accept
                    next-hop-self
                exit
            exit
        exit
        commit
    exit

    bgp
        export "static-routes"
        group "ibgp"
            peer-as 65536
            neighbor 192.0.2.6
            exit
        exit
    exit
```

Checking the static route received on PE-6 via BGP, the next hop is the PE-3 system address:

```
*A:PE-6# show router bgp routes 10.10.10.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1
Local Pref.    : 100                    Interface Name : int-PE-6-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  Incomplete
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h00m10s
```

```
Modified Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1
---snip---
-------------------------------------------------------------------------------
Routes : 1
```

Three of the BGP peering configuration possibilities are LDP, RSVP, or BGP. The other resolution filter options are related to segment routing and are beyond the scope of this chapter. In case both LDP and RSVP are included in the filter, RSVP is preferred. Disabling the IGP is also allowed (meaning that unless there is a shortcut, the BGP peering will not fall back to IGP):

```
*A:PE-6# configure router bgp next-hop-resolution shortcut-tunnel family ipv4
resolution
  - resolution {any|filter|disabled}

*A:PE-6# configure router bgp next-hop-resolution shortcut-tunnel family ipv4
resolution-filter
  - resolution-filter

 [no] bgp             - Use BGP tunnelling for next hop resolution
 [no] ldp             - Use LDP tunnelling for next hop resolution
 [no] rsvp            - Use RSVP tunnelling for next hop resolution
 [no] sr-isis         - Use sr-isis for next hop resolution
 [no] sr-ospf         - Use sr-ospf for next hop resolution
 [no] sr-te           - Use sr-te for next hop resolution

*A:PE-6#
```

When enabling LDP shortcuts on PE-6, the output changes showing the detail of the received BGP route indicating that the next hop is resolved using LDP:

```
*A:PE-6# configure router bgp
            next-hop-resolution
                shortcut-tunnel
                    family ipv4
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                exit
            exit

*A:PE-6# show router bgp routes 10.10.10.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
```

```
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1 (LDP)
Local Pref.    : 100                    Interface Name : int-PE-6-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  Incomplete
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h20m53s

Modified Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1 (LDP)
---snip---
-------------------------------------------------------------------------------
Routes : 1
```

The GRT output command also shows that the route is reachable using LDP (indicated as tunneled):

```
*A:PE-6# show router route-table next-hop-type tunneled
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.10.10.0/24                                   Remote  BGP       00h20m30s  170
       192.0.2.3 (tunneled)                                            0
-------------------------------------------------------------------------------
No. of Routes: 1
```

It can be seen that the previously created LSP LSP-PE-6-PE-3 is up and running:

```
*A:PE-6# show router mpls lsp "LSP-PE-6-PE-3" path detail

===============================================================================
MPLS LSP LSP-PE-6-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-6-PE-3 Path loose
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-6-PE-3
Path LSP ID       : 59904
From              : 192.0.2.6             To                   : 192.0.2.3
Admin State       : Up                    Oper State           : Up
Path Name         : loose                 Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262139
Path Up Time      : 0d 00:30:08           Path Down Time       : 0d 00:00:00
Retry Limit       : 0                     Retry Timer          : 30 sec
Retry Attempt     : 0                     Next Retry In        : 0 sec
BFD Template      : None                  BFD Ping Interval    : 60
BFD Enable        : False

Adspec            : Disabled              Oper Adspec          : Disabled
CSPF              : Enabled               Oper CSPF            : Enabled
Least Fill        : Disabled              Oper LeastFill       : Disabled
FRR               : Disabled              Oper FRR             : Disabled
Propogate Adm Grp: Disabled               Oper Prop Adm Grp    : Disabled
Inter-area        : False

Neg MTU           : 1564                  Oper MTU             : 1564
Bandwidth         : No Reservation        Oper Bandwidth       : 0 Mbps
Hop Limit         : 255                   Oper HopLimit        : 255
Record Route      : Record                Oper Record Route    : Record
Record Label      : Record                Oper Record Label    : Record
Setup Priority    : 7                     Oper Setup Priority  : 7
Hold Priority     : 0                     Oper Hold Priority   : 0
Class Type        : 0                     Oper CT              : 0
Backup CT         : None
MainCT Retry      : n/a
    Rem           :
MainCT Retry      : 0
    Limit         :
Include Groups    :                       Oper Include Groups  :
None                                            None
Exclude Groups    :                       Oper Exclude Groups  :
None                                            None

Adaptive          : Enabled               Oper Metric          : 100
Preference        : n/a
Path Trans        : 1                     CSPF Queries         : 1
Failure Code      : noError
```

```
Failure Node     : n/a
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.168.56.2 (192.0.2.6)                    Record Label        : N/A
 -> 192.168.56.1 (192.0.2.5)                    Record Label        : 262139
 -> 192.168.35.1 (192.0.2.3)                    Record Label        : 262140
Computed Hops    :
    192.168.56.2(S)
 -> 192.168.56.1(S)
 -> 192.168.35.1(S)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 20
===============================================================================
*A:PE-6#
```

After adding **resolution-filter rsvp** to the shortcut-tunnel configuration in the bgp context, the output shows that the BGP peer is reachable using an RSVP LSP (switched from LDP to RSVP since RSVP is preferred):

```
*A:PE-6# configure router bgp
            next-hop-resolution
                shortcut-tunnel
                    family ipv4
                        resolution-filter
                            ldp
                            rsvp
                        exit
                        resolution filter
                    exit
                exit
            exit

*A:PE-6# show router bgp routes 10.10.10.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1 (RSVP LSP: 4)
Local Pref.    : 100                    Interface Name : int-PE-6-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
```

```
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.3
Fwd Class      : None                    Priority       : None
Flags          : Used  Valid  Best  Incomplete
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h37m12s

Modified Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1 (RSVP LSP: 4)

---snip---
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
```

The GRT output command also shows that the route is reachable using RSVP (indicated as tunneled:RSVP:4):

```
*A:PE-6# show router route-table next-hop-type tunneled
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.10.10.0/24                                   Remote  BGP       00h16m09s  170
       192.0.2.3 (tunneled:RSVP:4)                                     0
-------------------------------------------------------------------------------
No. of Routes: 1
```

If the RSVP LSP is **shutdown**, the system reverts back to the LDP LSP:

```
*A:PE-6# configure router mpls lsp "LSP-PE-6-PE-3" shutdown

*A:PE-6# show router bgp routes 10.10.10.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
```

```
Original Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1 (LDP)
Local Pref.    : 100                    Interface Name : int-PE-6-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  Incomplete
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h49m57s

Modified Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.56.1 (LDP)

---snip---

-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
```

When the shortcut tunnel with resolution-filter rsvp is enabled at the BGP level, all RSVP LSPs originating on this node are eligible to be used by default as long as the destination address of the LSP corresponds to that of the BGP next hop for that prefix. It is also possible to exclude a specific RSVP LSP from BGP next hop resolution, similar to the exclusion of a specific RSVP LSP being used as a shortcut for resolving IGP routes. In this example, if the RSVP LSP LSP-PE-6-PE-3 is excluded to be eligible for BGP next hop resolution, it reverts back to LDP.

```
*A:PE-6# configure router mpls lsp "LSP-PE-6-PE-3"
               no bgp-shortcut
               no shutdown
               exit

*A:PE-6# show router route-table 10.10.10.0
===============================================================================
Route Table (Router: Base)
```

```
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.10.10.0/24                                   Remote  BGP       00h04m56s  170
       192.0.2.3 (tunneled)                                          0
-------------------------------------------------------------------------------
No. of Routes: 1
```

If the configuration is using **disallow-igp**, and neither LDP nor RSVP LSPs are available, the remote route received via BGP is removed from the GRT although the BGP peer session remains up. A field in the detailed show BGP route output indicates that the next hop is “Unresolved”:

```
*A:PE-6# configure router bgp
            next-hop-resolution
                shortcut-tunnel
                    family ipv4
                        resolution-filter
                            ldp
                            rsvp
                        exit
                        disallow-igp
                        resolution filter
                    exit
                exit
            exit

*A:PE-6# configure router ldp shutdown

*A:PE-6# show router bgp routes 10.10.10.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : Unresolved
Local Pref.    : 100                    Interface Name : int-PE-6-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
```

```
Fwd Class      : None                   Priority       : None
Flags          : Invalid  Incomplete  Nexthop-Unresolved
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 03h20m08s

Modified Attributes

Network        : 10.10.10.0/24
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : Unresolved
Local Pref.    : 100                    Interface Name : int-PE-6-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Invalid  Incomplete  Nexthop-Unresolved
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 03h20m08s
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
```

Because the route is unresolved, it does not appear in the GRT:

```
*A:PE-6# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
20.20.20.0/24                                   Remote  Static    03h26m25s  5
       172.16.66.2                                                  1
172.16.6.1/32                                   Local   Local     00h00m11s  0
       loopback                                                     0
172.16.66.0/30                                  Local   Local     03h26m25s  0
       static-route                                                 0
192.0.2.1/32                                    Remote  ISIS      03h27m22s  15
       192.168.46.1                                                 30
192.0.2.2/32                                    Remote  ISIS      03h27m23s  15
```

```
       192.168.46.1                                                  20
192.0.2.3/32                                 Remote  ISIS      03h27m23s  15
       192.168.56.1                                                  20
192.0.2.4/32                                 Remote  ISIS      03h27m23s  15
       192.168.46.1                                                  10
192.0.2.5/32                                 Remote  ISIS      03h27m23s  15
       192.168.56.1                                                  10
192.0.2.6/32                                 Local   Local     01d00h03m  0
       system                                                        0
192.168.12.0/30                              Remote  ISIS      03h27m23s  15
       192.168.46.1                                                  30
192.168.13.0/30                              Remote  ISIS      03h27m23s  15
       192.168.56.1                                                  30
192.168.23.0/30                              Remote  ISIS      03h27m22s  15
       192.168.46.1                                                  30
192.168.24.0/30                              Remote  ISIS      03h27m23s  15
       192.168.46.1                                                  20
192.168.35.0/30                              Remote  ISIS      03h27m23s  15
       192.168.56.1                                                  20
192.168.45.0/30                              Remote  ISIS      03h27m22s  15
       192.168.46.1                                                  20
192.168.46.0/30                              Local   Local     03h27m24s  0
       int-PE-6-PE-4                                                 0
192.168.56.0/30                              Local   Local     01d00h03m  0
       int-PE-6-PE-5                                                 0
-------------------------------------------------------------------------------
No. of Routes: 17
```

MPLS/GRE Shortcut for BGP NH Resolution within a VRF

Using RSVP/LDP or GRE shortcuts for resolving BGP next hops within a Virtual Private Routed Network (VPRN), also known as auto-bind-tunnel, allows a VPRN service to automatically resolve the BGP next hop for VPRN routes to an MPLS LSP or a GRE tunnel. Three possible mechanisms to provide transport tunnels for forwarding traffic between PE routers within an RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, network exist:

- RSVP-TE protocol to create tunnel LSPs between PE routers.
- LDP protocol to create tunnel LSPs between PE routers.
- GRE tunnels between PE routers.

These transport tunneling mechanisms provide the flexibility to use dynamically created LSPs where the service tunnels are automatically bound (the “auto-bind-tunnel” feature), and the ability to provide certain VPN services with their own transport tunnels by explicitly binding SDPs if desired. All services using the auto-bind-tunnel feature use the same set of LSPs, which does not allow for alternate tunneling mechanisms (like GRE) or the ability to craft sets of LSPs with bandwidth reservations for specific customers, as is available with explicit SDPs for the service.

The auto-bind-tunnel configuration is as follows:

```
*A:PE-2# configure service vprn 1 auto-bind-tunnel resolution
  - resolution {disabled|any|filter}

 <disabled|any|filt*> : disabled|any|filter


*A:PE-2# configure service vprn 1 auto-bind-tunnel resolution-filter
  - resolution-filter

 [no] gre             - Enable/disable setting GRE type for auto-bind-tunnel
 [no] ldp             - Enable/disable setting LDP type for auto-bind-tunnel
 [no] rsvp            - Enable/disable setting RSVP-TE type for auto-bind-tunnel
 [no] sr-isis         - Enable/disable setting SR-ISIS type for auto-bind-tunnel
 [no] sr-ospf         - Enable/disable setting SR-OSPF type for auto-bind-tunnel
 [no] sr-te           - Enable/disable setting SR-TE type for auto-bind-tunnel
```

Parameter descriptions:

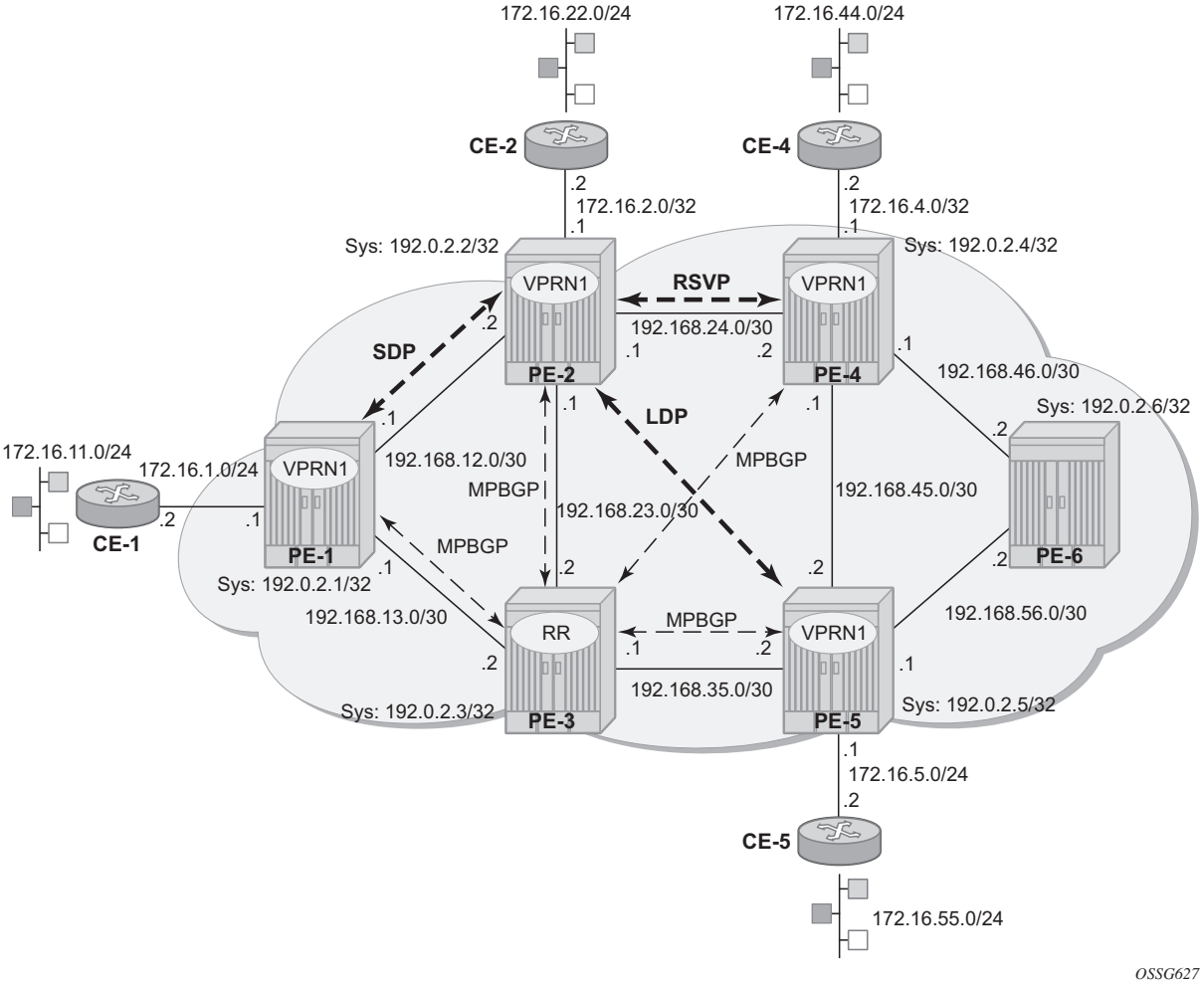
- **ldp** — Specifies LDP based LSPs should be used to resolve the BGP next hop for VPRN routes in an associated VPRN instance.
- **gre** — Specifies GRE based tunnels to be used to resolve the BGP next hop for VPRN routes in an associated VPRN instance. GRE is out of the scope regarding shortcuts, refer to SR OS documentation for further details.
- **rsvp** — Specifies RSVP-TE LSPs should be used to resolve the BGP next hop for VPRN routes in an associated VPRN instance.
- **ldp rsvp** — Chooses an existing RSVP-TE LSP if available, otherwise use LDP.
- **gre ldp rsvp** — When there are multiple tunnels to the BGP next hop address, the tunnel with the lowest tunnel-table preference value is selected (first RVSP, then LDP, then GRE).
- **sr-isis**— For segment routing with IS-IS (beyond the scope of this chapter)
- **sr-ospf** — For segment routing with OSPF (beyond the scope of this chapter)
- **sr-te** — For segment routing with TE (beyond the scope of this chapter)

In all cases, if an explicit spoke-sdp is specified in the VPRN, it is always preferred over automatically selected tunnels (even if the SDP is down, the route becomes inactive; there is no fallback to the automatic selection).

The network is configured according to the topology shown in Figure 6. Four PEs (PE-1, PE-2, PE-4 and PE-5) are connected forming a meshed IP-VPN (named VPRN 1), using a route reflector on PE-3 for MP-BGP peering. All PEs have LDP tunnels enabled so at a minimum all can establish LDP shortcut tunnels to the others. In order to have not only LDP but also RSVP-TE LSPs and static SDPs (using an RSVP LSP) in the network, a mix of tunneling methods is configured. For the sake of simplicity, the configuration of PE-2 only is given, providing the details about the

shortcuts created by auto-bind-tunnel. PE-2 has a static SDP (RSVP based) with PE-1, an RSVP LSP with PE-4, and an LDP LSP with PE-5. Every PE has a CE connected, so each PE has an interface connected to the CE as well as a static route to a CE LAN (although redistribution routing policies are needed, they are not shown for simplicity).

Figure 143 **Shortcuts Within a VRF Topology Network**



On PE-2, the following provides the configuration of VPRN1:

```
*A:PE-2# configure service
        sdp 1 mpls create
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
        vprn 1 customer 1 create
            vrf-import "VPN1-import"
```

```
            vrf-export "VPN1-export"
            route-distinguisher 65536:1
            auto-bind-tunnel
                resolution-filter
                    gre
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            interface "to-CE-2" create
                address 172.16.2.1/24
                sap 1/1/4:1 create
                exit
            exit
            static-route-entry 172.16.22.0/24
                next-hop 172.16.2.2
                    no shutdown
                exit
            exit
            spoke-sdp 1 create
            exit
            no shutdown
        exit
```

As previously mentioned, regarding IP-VPN meshed connectivity, the configuration shows that there is a static SDP 1 (pointing to PE-1), and the rest of the configuration is just **auto-bind-tunnel.** On PE-2, the connectivity towards the other PEs in the network can be verified by checking VPRN 1:

```
*A:PE-2# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                   Remote  BGP VPN   00h12m26s  170
       192.0.2.1 (tunneled)                                         0
172.16.2.0/24                                   Local   Local     00h20m27s  0
       to-CE-2                                                      0
172.16.4.0/24                                   Remote  BGP VPN   00h00m27s  170
       192.0.2.4 (tunneled:RSVP:3)                                  0
172.16.5.0/24                                   Remote  BGP VPN   00h12m26s  170
       192.0.2.5 (tunneled)                                         0
172.16.11.0/24                                  Remote  BGP VPN   00h12m26s  170
       192.0.2.1 (tunneled)                                         0
172.16.22.0/24                                  Remote  Static    00h20m27s  5
       172.16.2.2                                                   1
172.16.44.0/24                                  Remote  BGP VPN   00h00m27s  170
       192.0.2.4 (tunneled:RSVP:3)                                  0
172.16.55.0/24                                  Remote  BGP VPN   00h12m26s  170
       192.0.2.5 (tunneled)                                         0
-------------------------------------------------------------------------------
No. of Routes: 8
```

As can be seen, there are eight routes since every PE has two routes (one direct PE-CE interface and one static route), so six routes are received from other PEs via MP-BGP. The VPRN 1 routing table can be understood by looking at the tunnel table (active LSPs for remote system-ids):

```
*A:PE-2# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)

===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.1/32      sdp       MPLS  1         5        192.0.2.1      0
192.0.2.1/32      rsvp      MPLS  2         7        192.168.12.1   10
192.0.2.1/32      ldp       MPLS  65537     9        192.168.12.1   10
192.0.2.3/32      ldp       MPLS  65538     9        192.168.23.2   10
192.0.2.4/32      rsvp      MPLS  3         7        192.168.24.2   10
192.0.2.4/32      rsvp      MPLS  4         7        192.168.24.2   16777215
192.0.2.4/32      ldp       MPLS  65545     9        192.168.24.2   10
192.0.2.5/32      ldp       MPLS  65542     9        192.168.23.2   20
192.0.2.6/32      ldp       MPLS  65546     9        192.168.24.2   20
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route

===============================================================================

*A:PE-2#
```

The tunnel-table shows one entry per LSP per remote PE. The following tunnel selection rules apply:

- SDP has the lowest (best) preference, followed by RSVP and then by LDP.
- If the preference is the same, the lowest metric is selected (ECMP is possible with LDP).

PE-2 has three possibilities to reach PE-1 (192.0.2.1): an SDP Tunnel ID 1 with preference 5, an RSVP Tunnel ID 1 with preference 7, and an LDP LSP with preference 9. Because SDP Tunnel ID 1 has the lowest preference, it is the chosen option. PE-2 has three possibilities to reach PE-4 (192.0.2.4): an RSVP Tunnel ID 3 with preference 7 and metric 10, an RSVP Tunnel ID 4 with preference 7 and metric 16777215, and an LDP LSP with preference 9; hence RSVP Tunnel ID 3 is selected. PE-2 only has one option to reach PE-5 and PE-6 (192.0.2.5 and 192.0.2.6) using an LDP LSP.

As the VPRN 1 output does not show the details of the tunneling, displaying the FIB on router VPRN 1 provides more detailed information:

```
*A:PE-2# show router 1 fib 1
===============================================================================
FIB Display
```

```
===============================================================================
Prefix [Flags]                                                  Protocol
    NextHop
-------------------------------------------------------------------------------
172.16.1.0/24                                                   BGP_VPN
    192.0.2.1 (VPRN Label:262137 Transport:SDP:1)
172.16.2.0/24                                                   LOCAL
    172.16.2.0 (to-CE-2)
172.16.4.0/24                                                   BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:RSVP LSP:3)
172.16.5.0/24                                                   BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:LDP)
172.16.11.0/24                                                  BGP_VPN
    192.0.2.1 (VPRN Label:262137 Transport:SDP:1)
172.16.22.0/24                                                  STATIC
    172.16.2.2 (to-CE-2)
172.16.44.0/24                                                  BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:RSVP LSP:3)
172.16.55.0/24                                                  BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:LDP)
-------------------------------------------------------------------------------
Total Entries : 8
-------------------------------------------------------------------------------
```

The FIB shows the chosen transport tunnel, specifying SDP ID, RSVP Tunnel ID, and LDP, as well as service label information linked to the routes.

As static SDP tunnels are preferred over dynamic tunnels (RSVP or LDP auto-bind-tunnel), when the static SDP 1 is shut down with or the LSP goes down (there is no fallback to dynamic tunneling), the associated routes are removed:

```
*A:PE-2# configure service sdp 1 shutdown

*A:PE-2# show router 1 fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
    NextHop
-------------------------------------------------------------------------------
172.16.2.0/24                                                   LOCAL
    172.16.2.0 (to-CE-2)
172.16.4.0/24                                                   BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:RSVP LSP:3)
172.16.5.0/24                                                   BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:LDP)
172.16.22.0/24                                                  STATIC
    172.16.2.2 (to-CE-2)
172.16.44.0/24                                                  BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:RSVP LSP:3)
172.16.55.0/24                                                  BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:LDP)
-------------------------------------------------------------------------------
Total Entries : 6
```

To avoid this fallback issue, the configuration is modified and the manual spoke-sdps are removed from the configuration of PE-1 and PE-2, the rest of the configuration remains the same. Now the connectivity between PE-1 and PE-2 is using an RSVP LSP, as shown in the PE-1 following output (RSVP LSP which was used by SDP 1 has disappeared):

```
*A:PE-1# configure service vprn 1 no spoke-sdp 1
*A:PE-2# configure service vprn 1 no spoke-sdp 1

*A:PE-1# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                   Local   Local     00h33m21s  0
       to-CE-1                                                        0
172.16.2.0/24                                   Remote  BGP VPN   00h00m24s  170
       192.0.2.2 (tunneled:RSVP:2)                                    0
172.16.4.0/24                                   Remote  BGP VPN   00h13m29s  170
       192.0.2.4 (tunneled)                                           0
172.16.5.0/24                                   Remote  BGP VPN   00h25m09s  170
       192.0.2.5 (tunneled)                                           0
172.16.11.0/24                                  Remote  Static    00h33m21s  5
       172.16.1.2                                                     1
172.16.22.0/24                                  Remote  BGP VPN   00h00m24s  170
       192.0.2.2 (tunneled:RSVP:2)                                    0
172.16.44.0/24                                  Remote  BGP VPN   00h13m29s  170
       192.0.2.4 (tunneled)                                           0
172.16.55.0/24                                  Remote  BGP VPN   00h25m09s  170
       192.0.2.5 (tunneled)                                           0
-------------------------------------------------------------------------------
No. of Routes: 8
```

If RSVP is disabled, the connectivity falls back to LDP as the output shows:

```
*A:PE-1# configure router mpls shutdown

*A:PE-1# show router 1 fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
    NextHop
-------------------------------------------------------------------------------
172.16.1.0/24                                                   LOCAL
    172.16.1.0 (to-CE-1)
172.16.2.0/24                                                   BGP_VPN
    192.0.2.2 (VPRN Label:262136 Transport:LDP)
172.16.4.0/24                                                   BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:LDP)
172.16.5.0/24                                                   BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:LDP)
172.16.11.0/24                                                  STATIC
    172.16.1.2 (to-CE-1)
172.16.22.0/24                                                  BGP_VPN
```

```
    192.0.2.2 (VPRN Label:262136 Transport:LDP)
172.16.44.0/24                                                  BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:LDP)
172.16.55.0/24                                                  BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:LDP)
-------------------------------------------------------------------------------
Total Entries : 8
-------------------------------------------------------------------------------
```

If LDP is disabled, the connectivity falls back to GRE as the output shows:

```
*A:PE-1# configure router ldp shutdown

*A:PE-1# show router 1 fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
    NextHop
-------------------------------------------------------------------------------
172.16.1.0/24                                                   LOCAL
    172.16.1.0 (to-CE-1)
172.16.2.0/24                                                   BGP_VPN
    192.0.2.2 (VPRN Label:262136 Transport:GRE)
172.16.4.0/24                                                   BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:GRE)
172.16.5.0/24                                                   BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:GRE)
172.16.11.0/24                                                  STATIC
    172.16.1.2 (to-CE-1)
172.16.22.0/24                                                  BGP_VPN
    192.0.2.2 (VPRN Label:262136 Transport:GRE)
172.16.44.0/24                                                  BGP_VPN
    192.0.2.4 (VPRN Label:262136 Transport:GRE)
172.16.55.0/24                                                  BGP_VPN
    192.0.2.5 (VPRN Label:262136 Transport:GRE)
-------------------------------------------------------------------------------
Total Entries : 8
-------------------------------------------------------------------------------
```

Conclusion

IGP shortcuts provide a variety of shortcuts in IP, MPLS and IP-VPN scenarios to customers who want to use new options for building routing topologies. Because IGP shortcuts are enabled on a per router basis, SPF computations are independent and irrelevant to other routers, so there is no need to enable shortcuts globally. This network example shows the configuration of IGP shortcuts together with the associated show outputs which can be used for verification and troubleshooting.

Inter-Area TE Point-to-Point LSPs

This chapter describes inter-area TE point-to-point LSP configurations.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

Inter-Area Traffic Engineering (TE) point-to-point (P2P) label switched paths (LSPs) are supported on SR OS. This feature is supported on all IOM/IMM types. Initially, this chapter was written for release 11.0.R4, but the CLI in this edition corresponds to release 14.0.R1.

Summary

Multiprotocol label switching with traffic engineering (MPLS TE) is implemented on a wide scale in current ISP networks to steer traffic across their backbones to facilitate efficient use of available bandwidth between the routers and to guarantee fast convergence in case a link or node fails.

Regular TE LSPs in MPLS networks are confined to only a single interior gateway protocol (IGP) area/level. This is because the head-end has information in the TE database of only the local area (OSPF) or level (IS-IS). As the name implies, inter-area TE LSPs can cross the area/level borders of the IGP.

Inter-Area TE LSP Based On Explicit Route Expansion

Inter-area TE LSP using explicit route object (ERO) expansion enables the head-end to calculate the ERO path within its own area/level and keep the remaining ABRs of other areas/levels as loose hops in the ERO path. On receiving a PATH message with a loose hop ERO and based on local configuration, each area border router (ABR) does a partial constrained shortest path first (CSPF) calculation to the next ABR or a full CSPF calculation to reach the final destination.

Automatic selection of ABRs is supported so that the head-end node can work with an empty primary path. When the **to** field of an LSP definition is in an area/level different from the head-end node, CSPF will automatically compute the segment to the exit ABR router which advertised the prefix and which currently is the best path for resolving the prefix in route table manager (RTM).

ABR Protection

Link and node protection within the respective areas are supported through the TE capabilities of the IGP and resource reservation protocol (RSVP) in each area. To support ABR node protection, a bypass is required from the point of local repair (PLR; node prior to ABR) to the merge point (MP; next-hop node to ABR).

Two methods are possible: dynamic ABR protection and static ABR protection. Static ABR protection uses manual bypass tunnels (MBTs), statically configured by the operator between the PLR and the MP. For dynamic ABR protection, node ID propagation and signaling of an exclude route object (XRO) in RSVP PATH messages must both be supported.

Since the record route object (RRO) Node ID sub-object description in RFC 4561 (*Definition of a Record Route Object (RRO) Node-Id Sub-Object*) is not clear about the format of the included node-address (S), interface-address (I) and label (L), the system supports multiple formats: IL, SL, ISL, SIL, SLI, ILSL and SLIL. The system uses the SLIL (node-address, label, interface-address, label) format to include the node-ID itself.

The exclude route object (XRO) inclusion (RFC 4874, *Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering*) in bypass RSVP PATH messages is required to exclude the protected ABR from the bypass path. The XRO object contains the ABRs system IP address.

Overview

The example setup in this chapter contains ten nodes in three areas. Figure 144 shows the physical topology of the setup.

Figure 144 **Inter-Area TE LSP Setup**

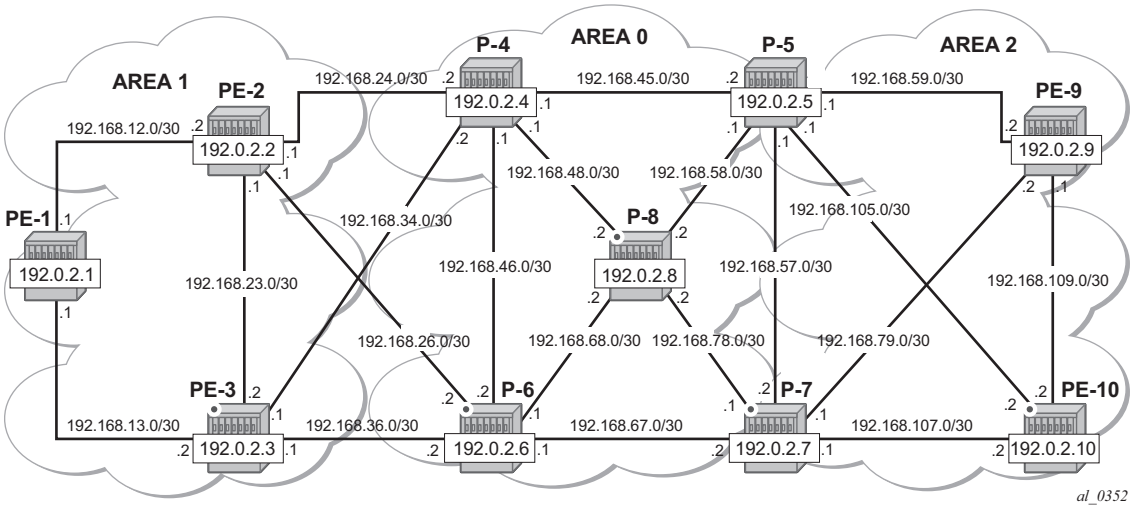
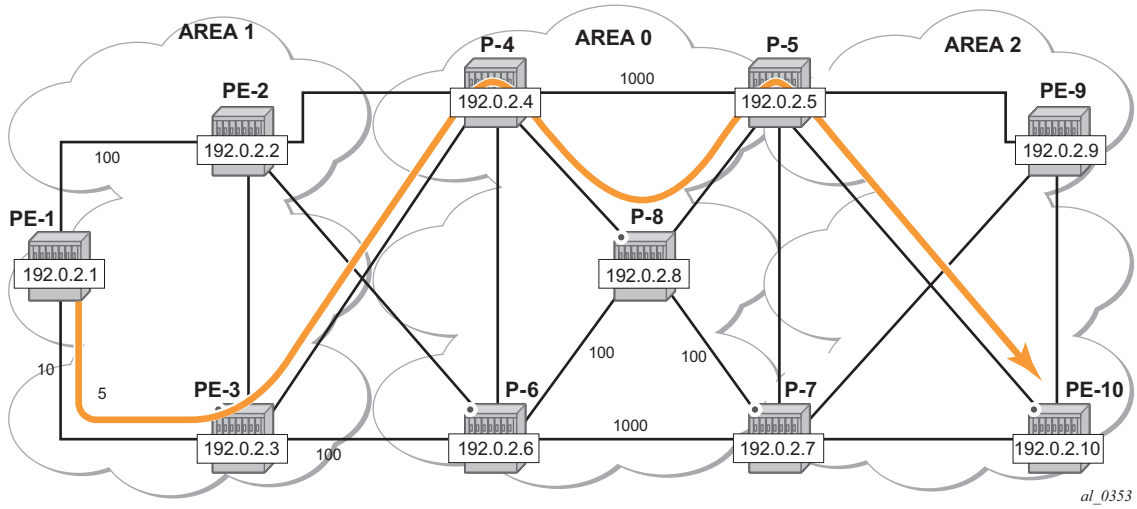


Figure 145 shows the LSP path intended to be set up through the network. An empty MPLS path is used. At the head-end node PE-1, the destination address PE-10 is learned via ABR node P-4 and ABR node P-5.

Figure 145 **Inter-Area TE LSP Path**



Configuration

The assumption is made that following base configuration has been implemented on the nodes:

- Cards, MDAs and ports configured
- Interfaces configured
- IGP areas configured and converged
- Traffic Engineering configured for the IGP
- MPLS and RSVP configured on all links in the network

OSPF or IS-IS can be configured as the IGP; OSPF is used in this chapter.

The following output shows the opaque database of PE-1:

```
A:PE-1# show router ospf opaque-database
===============================================================================
OSPF Opaque Link State Database (Type : All)
===============================================================================
Type  Id              Link State Id     Adv Rtr Id      Age  Sequence   Cksum
-------------------------------------------------------------------------------
Area  0.0.0.1         1.0.0.1           192.0.2.1       1503 0x80000002 0x9234
Area  0.0.0.1         1.0.0.3           192.0.2.1       1470 0x80000001 0x9b45
Area  0.0.0.1         1.0.0.4           192.0.2.1       1465 0x80000001 0xe9f2
Area  0.0.0.1         1.0.0.1           192.0.2.2       1498 0x80000002 0x962e
Area  0.0.0.1         1.0.0.3           192.0.2.2       1470 0x80000001 0xdce8
Area  0.0.0.1         1.0.0.4           192.0.2.2       1470 0x80000001 0x833b
Area  0.0.0.1         1.0.0.5           192.0.2.2       1471 0x80000001 0x637b
Area  0.0.0.1         1.0.0.6           192.0.2.2       1471 0x80000001 0x665f
Area  0.0.0.1         1.0.0.1           192.0.2.3       1504 0x80000002 0x9a28
Area  0.0.0.1         1.0.0.3           192.0.2.3       1472 0x80000001 0x6d43
Area  0.0.0.1         1.0.0.4           192.0.2.3       1481 0x80000001 0x1495
Area  0.0.0.1         1.0.0.5           192.0.2.3       1470 0x80000001 0x9f3c
Area  0.0.0.1         1.0.0.6           192.0.2.3       1472 0x80000001 0x4283
Area  0.0.0.1         1.0.0.1           192.0.2.4       1482 0x80000002 0x9e22
Area  0.0.0.1         1.0.0.6           192.0.2.4       1471 0x80000001 0x7e44
Area  0.0.0.1         1.0.0.7           192.0.2.4       1473 0x80000001 0x218b
Area  0.0.0.1         1.0.0.1           192.0.2.6       1584 0x80000002 0xa616
Area  0.0.0.1         1.0.0.6           192.0.2.6       1467 0x80000001 0xf6c5
Area  0.0.0.1         1.0.0.7           192.0.2.6       1482 0x80000001 0x990d
-------------------------------------------------------------------------------
No. of Opaque LSAs: 19
===============================================================================
A:PE-1#
```

The information is only about routers that are part of area 0.0.0.1. PE-1 cannot calculate an end-to-end CSPF path to node PE-10 since this would require TE topology information from area 0.0.0.0 and area 0.0.0.2.

Each node announces its router ID and each attached link that is part of that area, hence the 19 opaque LSAs in area 0.0.0.1. The system interfaces of P-4 and P-6 are configured in backbone area 0.0.0.0, not in area 0.0.0.1.

In Figure 145, the LSP passes through node PE-3 and node P-8. In order to prefer a dynamic path from PE-1 to P-4 via PE-3 rather than via PE-2, it is necessary to configure on PE-1 a lower IGP metric on the interface to PE-3 (the default metric is derived from the interface speed; in this case the metric is 10 by default).

```
*A:PE-1# configure router ospf area 1 interface "int-PE-1-PE-3" metric 5
```

Similarly, in the core, the IGP metric between P-4 and P-5, and between P-6 and P-7 is increased to force the LSP to pass through the core P-8 node.

```
*A:P-4# configure router ospf area 0 interface "int-P-4-P-5" metric 1000

*A:P-6# configure router ospf area 0 interface "int-P-6-P-7" metric 1000
```

Other metrics have also been manipulated as indicated on Figure 145.

MPLS Path Configuration

Since automatic ABR selection is performed, an empty MPLS path is sufficient on the head-end node PE-1. Using an empty MPLS path will ease the provisioning process and brings consistency since this empty MPLS path can be used for both intra and inter-area/level type LSPs.

```
*A:PE-1# configure router mpls path "dyn" no shutdown
```

MPLS LSP Configuration

Configure an LSP on PE-1 to PE-10 and include the previously created MPLS path as primary path. Enable CSPF and fast reroute (FRR) facility on the LSP.

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-10"
                to 192.0.2.10
                cspf
                fast-reroute facility
                exit
                primary "dyn"
                exit
                no shutdown
            exit
```

At this stage the LSP is in an operational Down state with a failure code of badNode at failure node 192.168.34.2 (ABR P-4).

In order to get around the intra-area CSPF confinement, enable the ERO-expansion feature on all possible ABR nodes.

```
*A:P-4# configure router mpls cspf-on-loose-hop
*A:P-6# configure router mpls cspf-on-loose-hop

*A:P-7# configure router mpls cspf-on-loose-hop
*A:P-5# configure router mpls cspf-on-loose-hop
```

Cspf-on-loose-hop is only required if FRR or TE parameters are configured on the LSP. If any of these parameters is configured on the LSP while one of the ABRs along the path is not configured with cspf-on-loose-hop, the LSP will stay operationally down with failure code: badNode and an indication of the interface address of the failure node.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail

---snip---

From        : 192.0.2.1                          To          : 192.0.2.10
Adm State   : Up                                 Oper State  : Down

---snip---

Failure Code     : badNode
Failure Node     : 192.168.34.2

---snip---
```

The LSP path can also contain other strict and/or loose hops. However, cspf-on-loose-hop must be configured inthe MPLS context whenever loose hops are configured in the MPLS path. This command enables ERO expansion and is only required for inter-area LSPs on all possible ABR nodes and all nodes not belonging to the area where the iLER is located, which have a loose hop reference in the MPLS path. However, for simplicity and consistency, it can be configured on all nodes without having a negative effect.

The following trace shows the ERO calculation on the head-end to the first ABR.

```
*A:PE-1# debug router rsvp packet path detail

4 2016/03/15 10:25:02.22 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:255, Checksum:0x3b17, Flags:0x0
Session    - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr   - Name:LSP-PE-1-PE-10::dyn
             SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop    - Ctype:1, Addr:192.168.13.1, LIH:3
TimeValue  - RefreshPeriod:30
```

```
SendTempl  - Sender:192.0.2.1, LspId:57862
SendTSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1564
LabelReq   - IfType:General, L3ProtID:2048
RRO        - IpAddr:192.168.13.1, Flags:0x0
ERO        - IPv4Prefix 192.168.13.2/32, Strict
             IPv4Prefix 192.168.34.2/32, Strict
             IPv4Prefix 192.0.2.10/32, Loose
FRRObj     - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
             ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

On the ABR P-4, the ERO is expanded to include the nodes of area 0.0.0.0 of which P-4 is also part. The RRO contains all the hops the PATH message has passed so far.

```
*A:P-4# debug router rsvp packet path detail

4 2016/03/15 10:24:40.42 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:253, Checksum:0xc992, Flags:0x0
Session    - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr   - Name:LSP-PE-1-PE-10::dyn
             SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop    - Ctype:1, Addr:192.168.48.1, LIH:4
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.1, LspId:57862
SendTSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1564
LabelReq   - IfType:General, L3ProtID:2048
RRO        - IpAddr:192.168.48.1, Flags:0x0
             IpAddr:192.168.34.1, Flags:0x0
             IpAddr:192.168.13.1, Flags:0x0
ERO        - IPv4Prefix 192.168.48.2/32, Strict
             IPv4Prefix 192.168.58.1/32, Strict
             IPv4Prefix 192.0.2.10/32, Loose
FRRObj     - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
             ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

Finally, the P-5 ABR will expand the ERO to the final destination PE-10:

```
*A:P-5# debug router rsvp packet path detail

8 2016/03/15 10:25:03.52 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:251, Checksum:0x71f1, Flags:0x0
Session    - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr   - Name:LSP-PE-1-PE-10::dyn
             SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop    - Ctype:1, Addr:192.168.105.1, LIH:5
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.1, LspId:57862
SendTSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1564
```

```
LabelReq   - IfType:General, L3ProtID:2048
RRO        - IpAddr:192.168.105.1, Flags:0x0
             IpAddr:192.168.58.2, Flags:0x0
             IpAddr:192.168.48.1, Flags:0x0
             IpAddr:192.168.34.1, Flags:0x0
             IpAddr:192.168.13.1, Flags:0x0
ERO        - IPv4Prefix 192.168.105.2/32, Strict
FRRObj     - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
             ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

The MPLS LSP is now operational up and the LSP path can be shown in detail on the head-end, PE-1:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-10 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-10 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-10
Path LSP ID       : 57862
From              : 192.0.2.1             To                  : 192.0.2.10
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up
Out Interface     : 1/1/2                 Out Label           : 262143
Path Up Time      : 0d 00:01:47           Path Down Time      : 0d 00:00:00
Retry Limit       : 0                     Retry Timer         : 30 sec
Retry Attempt     : 0                     Next Retry In       : 0 sec
BFD Template      : None                  BFD Ping Interval   : 60
BFD Enable        : False

Adspec            : Disabled              Oper Adspec         : Disabled
CSPF              : Enabled               Oper CSPF           : Enabled
Least Fill        : Disabled              Oper LeastFill      : Disabled
FRR               : Enabled               Oper FRR            : Enabled
FRR NodeProtect   : Enabled               Oper FRR NP         : Enabled
FR Hop Limit      : 16                    Oper FRHopLimit     : 16
FR Prop Admin Gr* : Disabled              Oper FRPropAdmGrp   : Disabled
Propogate Adm Grp : Disabled              Oper Prop Adm Grp   : Disabled
Inter-area        : True

Neg MTU           : 1560                  Oper MTU            : 1560
Bandwidth         : No Reservation        Oper Bandwidth      : 0 Mbps
Hop Limit         : 255                   Oper HopLimit       : 255
Record Route      : Record                Oper Record Route   : Record
Record Label      : Record                Oper Record Label   : Record
Setup Priority    : 7                     Oper Setup Priority : 7
Hold Priority     : 0                     Oper Hold Priority  : 0
```

```
Class Type       : 0                    Oper CT             : 0
Backup CT        : None
MainCT Retry     : n/a
    Rem          :
MainCT Retry     : 0
    Limit        :
Include Groups   :                      Oper Include Groups  :
None                                          None
Exclude Groups   :                      Oper Exclude Groups  :
None                                          None

Adaptive         : Enabled              Oper Metric          : 15
Preference       : n/a
Path Trans       : 1                    CSPF Queries         : 4
Failure Code     : noError
Failure Node     : n/a
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.168.13.1 (192.0.2.1) @ n                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @                  Record Label        : 262143
 -> 192.168.34.2 (192.0.2.4) @ n                Record Label        : 262143
 -> 192.168.48.2  @                             Record Label        : 262143
 -> 192.168.58.1  @                             Record Label        : 262143
 -> 192.168.105.2                               Record Label        : 262143

Computed Hops    :
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.34.2(SA)
 -> 192.0.2.10(L)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 15
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

ABR Node Protection

At this stage, the LSP is configured with facility FRR protection; link and node protection will be offered within each area, as was seen in the preceding output where there is no node protection for ABR P-4 because there is no bypass tunnel originating in PE-3 to offer node protection for the next hop. Likewise, there is no bypass tunnel originating in P-8 to offer node protection for ABR P-5. The bypass tunnels originating in PE-3 and P-8 only offer link protection. Dynamic ABR node protection requires the setup of a bypass tunnel from the PLR (node just upstream of the ABR) to the MP (node just downstream of the ABR). Two things are required for this:

- Firstly, the PLR node (part of area x) needs to know the system IP address of the MP node (part of area y) to set up the bypass. For this reason, the node-ID of the MP is included in the RESV message so that the PLR can link the manual bypass tunnel to the primary path to protect the ABR.
- Secondly, the other ABR node receiving the RSVP bypass PATH message for the protected ABR needs to do an ERO expansion toward the MP node. For this reason, the XRO object is included in the RSVP bypass PATH message, containing the node-ID of the protected ABR. As an example, a bypass PATH message is shown below on node PE-3.

The XRO object includes the system IP address of the protected ABR node P-4 and the ERO object has MP node P-8 as loose destination:

```
*A:PE-3# debug router rsvp packet path detail

1 2016/03/15 10:26:58.62 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.3, To:192.0.2.8
          TTL:17, Checksum:0xfddd, Flags:0x0
Session    - EndPt:192.0.2.8, TunnId:61442, ExtTunnId:192.0.2.3
SessAttr   - Name:bypass-node192.0.2.4-61442
             SetupPri:7, HoldPri:0, Flags:0x2
RSVPHop    - Ctype:1, Addr:192.168.36.1, LIH:3
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.3, LspId:4
SendTSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1564
LabelReq   - IfType:General, L3ProtID:2048
RRO        - IpAddr:192.168.36.1, Flags:0x0
ERO        - IPv4Prefix 192.168.36.2/32, Strict
             IPv4Prefix 192.0.2.8/32, Loose
XRO        - IPv4Prefix: 192.0.2.4/32, Attribute: Node, LBit: Exclude
AdSpec     - General BreakBit:0, NumISHops:0, PathBwEstimate:0
                     MinPathLatency:4294967295, CompPathMTU:1564
             Controlled BreakBit:0
"
```

Node-ID Inclusion in the RESV Message

P-8 will be the MP for the bypass of ABR P-4 and PE-10 will be the MP for the bypass of ABR P-5. So P-8 and PE-10 need to include their node-ID in the RESV message, inside the record route object (RRO).

```
*A:P-8# configure router rsvp node-id-in-rro include
*A:PE-10# configure router rsvp node-id-in-rro include
```

The default is **node-id-in-rro exclude**. As an example, the RESV message received on PLR node PE-3 is shown below. The RRO contains the MP node P-8 information in SLIL format:

```
*A:PE-3# debug router rsvp packet resv detail

7 2016/03/15 10:27:49.62 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESV Msg
Send RESV From:192.168.13.2, To:192.168.13.1
          TTL:255, Checksum:0xfcb3, Flags:0x0
Session    - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
RSVPHop    - Ctype:1, Addr:192.168.13.2, LIH:3
TimeValue  - RefreshPeriod:30
Style      - SE
FlowSpec   - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1560, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:192.0.2.1, LspId:57874, Label:262142
RRO        - ---snip---
             SystemIp:192.0.2.8, Flags:0x29
             Label:262141, Flags:0x1
             InterfaceIp:192.168.48.2, Flags:0x9
             Label:262141, Flags:0x1
---snip---
"
```

Bypass Configuration For ABR Protection

Since dynamic ABR protection is supported and used in this example, no explicit MBTs are configured to protect the ABRs. Each PLR first checks if an MBT tunnel exists between the PLR and the MP matching the constraints and protecting the ABR. If no MBT is available, the PLR will signal a bypass tunnel in a dynamic way toward the MP node.

Figure 146 shows the two dynamic ABR node protections that are signaled for this LSP.

Figure 146 **ABR Protection**

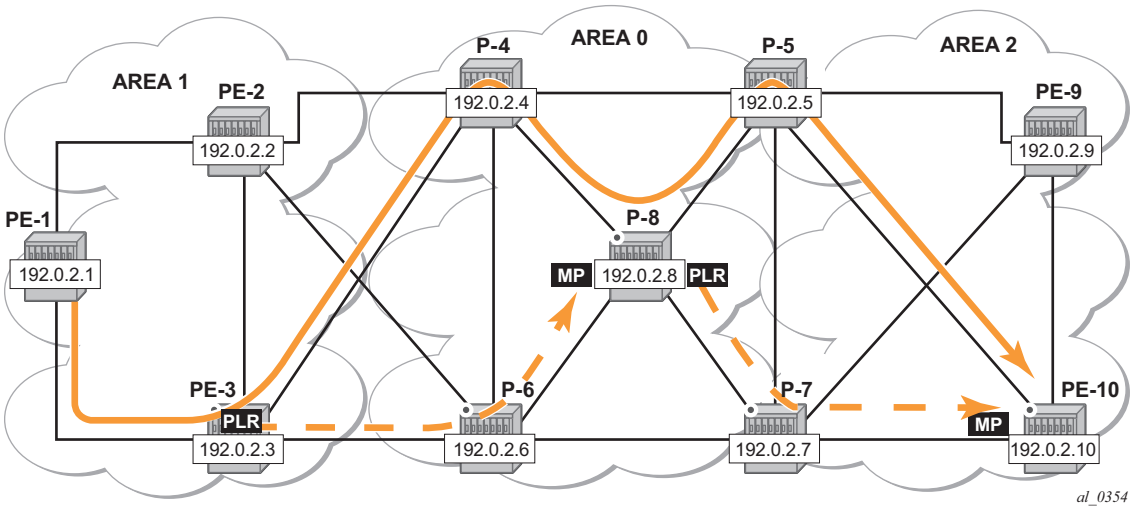
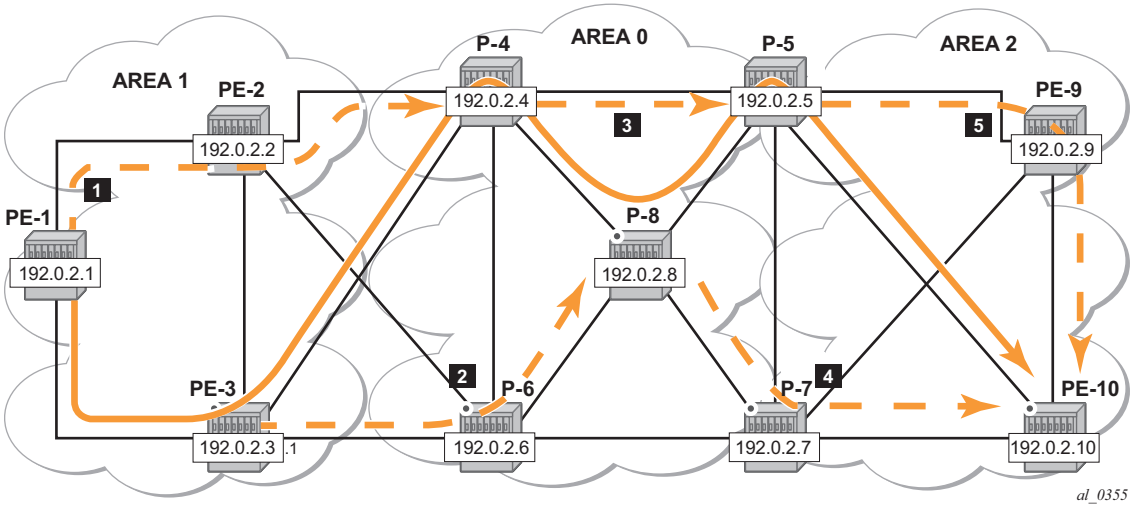


Figure 147 shows the complete picture of all the FRR protections and indicates each node/link protection in the setup.

Figure 147 **Protection of All Nodes/Links Along the LSP Path**



This can be seen in the detailed show output of the LSP path:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail

===============================================================================
MPLS LSP LSP-PE-1-PE-10 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available                # - Detour In Use
```

```
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-10 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-10
Path LSP ID      : 57862
From             : 192.0.2.1            To                  : 192.0.2.10
Admin State      : Up                   Oper State          : Up
Path Name        : dyn                  Path Type           : Primary
Path Admin       : Up                   Path Oper           : Up
---snip---

Inter-area       : True
---snip---

Actual Hops      :
    192.168.13.1 (192.0.2.1) @ n                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @ n                Record Label        : 262143
 -> 192.168.34.2 (192.0.2.4) @ n                Record Label        : 262143
 -> 192.0.2.8 (192.0.2.8) @ n                   Record Label        : 262143
 -> 192.168.48.2  @ n                           Record Label        : 262143
 -> 192.0.2.5 (192.0.2.5) @                     Record Label        : 262143
 -> 192.168.58.1  @                             Record Label        : 262143
 -> 192.0.2.10 (192.0.2.10)                     Record Label        : 262143
 -> 192.168.105.2                               Record Label        : 262143
---snip---
```

- The first bypass originates in PE-1 and protects node PE-3.
- The second bypass originates in PE-3 and protects node P-4.
- The third bypass originates in P-4 and protects node P-8.
- The fourth bypass originates in P-8 and protects node P-5. There are two entries for P-8: hop 192.0.2.8 and hop 192.168.48.2.
- The fifth bypass originates in P-5 and protects the link between P-5 and PE-10. There are two entries for P-5: hop 192.0.2.5 and hop 192.168.58.1.

There are two entries for P-8, P-5 and PE-10 in the ‘Actual Hops’ section in the previous output: one for the interface IP address and one for the system IP address. This is a consequence of configuring **node-id-in-rro include** on P-8, P-5 and PE-10.

The **node-id-in-rro include** command is not mandatory for this example on ABR node P-5, but to be able to cover cases where a new LSP is established in the network and P-5 acts as an MP node while the corresponding PLR node for that new LSP is in another area. This RSVP command can be executed on all possible MP nodes in the network.

The details of the bypass tunnel can be shown with the following command:

```
*A:PE-3# show router mpls bypass-tunnel protected-lsp detail
```

```
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-node192.0.2.4-61442
-------------------------------------------------------------------------------
To             : 192.0.2.8           State               : Up
Out I/F        : 1/1/2               Out Label           : 262143
Up Time        : 0d 00:01:17         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 100
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : 192.0.2.4           Inter-Area          : True
Computed Hops  :
    192.168.36.1(S)                  Egress Admin Groups : None
 -> 192.168.36.2(SA)                 Egress Admin Groups : None
 -> 192.0.2.8(L)                     Egress Admin Groups : None
Actual Hops    :
    192.168.36.1 (192.0.2.3)         Record Label        : N/A
 -> 192.168.36.2 (192.0.2.6)         Record Label        : 262143
 -> 192.0.2.8 (192.0.2.8)            Record Label        : 262142
 -> 192.168.68.2                     Record Label        : 262142
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-1-PE-10::dyn
From           : 192.0.2.1           To                  : 192.0.2.10
Avoid Node/Hop : 192.0.2.4           Downstream Label    : 262143
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-3#
```

The LSP could be protected with one or more additional secondary paths, pre-signaled or not, but this is outside the scope of this example.

When a link or node failure occurs along the LSP path, FRR protection kicks in and end-to-end path re-optimization is executed: a PATHERR message is forwarded to the head-end. Upon receiving the PATHERR message the head-end calculates a new path.

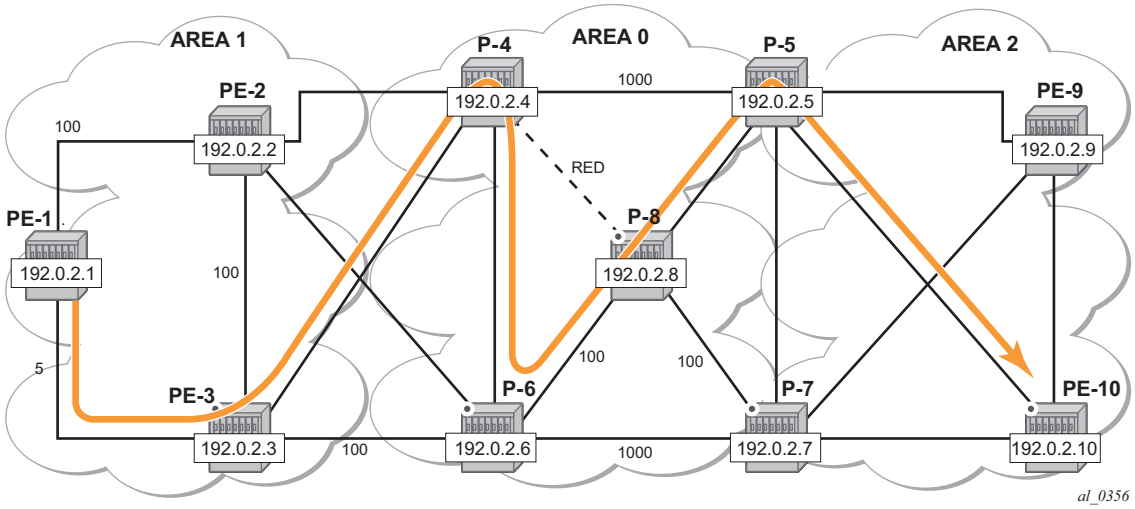
Admin Groups

The use of administrative groups is described in chapter RSVP Point-to-Point LSPs.

To support admin-groups for inter-area LSPs, the ingress node PE-1 must propagate the admin-groups within the session attribute object (SA) of the PATH message so that the ABRs along the path receive the admin group restrictions they have to take into account when further expanding the ERO in the PATH message.

In Figure 147 the LSP path avoids the link between P-4 and P-8. This is implemented by assigning admin group “red” to the link between P-4 and P-8 and then configuring the LSP to exclude the admin group “red”.

Figure 148 **Admin Group Example**



Admin Group Configuration

On P-4, configure admin group “red” and assign a group value. In this example group value 11 is used, but this can be any value between 0 and 31. Assign admin group “red” to the link to P-8.

This admin-group configuration is required on P-4 and on iLER PE-1. However, it is good practice to configure the admin group on all the nodes.

```
*A:Px# configure router if-attribute admin-group red value 11

*A:P-4# configure router
        mpls
            interface "int-P-4-P-8"
                admin-group "red"
            exit
```

On PE-1, change the LSP configuration as follows:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-10"
                exclude "red"
                propagate-admin-group
            exit
```

It is possible to have the same admin group constraint applied to the FRR bypass tunnels in the PLRs, but that is not the case here. The bypass tunnels ignore any admin group constraint. The **propagate-admin-group** command is required to include the admin group properties in the SA object of the PATH message. The admin-group value is mapped to a 32-bitmap. In this example, value 11 means that the 12th bit is set, which means in binary 100000000000 or hex 0x800.

```
*A:PE-1# debug router rsvp packet path detail

1 2016/03/15 10:29:25.22 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:255, Checksum:0x3301, Flags:0x0
Session    - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr   - Name:LSP-PE-1-PE-10::dyn
             SetupPri:7, HoldPri:0, Flags:0x17
             Ctype:RA, ExcAny:0x800, IncAny:0x0, IncAll:0x0
RSVPHop    - Ctype:1, Addr:192.168.13.1, LIH:3
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.1, LspId:57866
SendTSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1564
LabelReq   - IfType:General, L3ProtID:2048
RRO        - IpAddr:192.168.13.1, Flags:0x0
ERO        - IPv4Prefix 192.168.13.2/32, Strict
             IPv4Prefix 192.168.34.2/32, Strict
             IPv4Prefix 192.0.2.10/32, Loose
FRRObj     - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
             ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

The following two sets of output show that when P-4 expands the ERO it now excludes the link to node P-8 for the path calculation and the path is set up through P-6, P-8 and P-5.

```
*A:P-4# debug router rsvp packet path detail

8 2016/03/15 10:30:01.43 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:253, Checksum:0xa1ba, Flags:0x0
Session    - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr   - Name:LSP-PE-1-PE-10::dyn
             SetupPri:7, HoldPri:0, Flags:0x17
             Ctype:RA, ExcAny:0x800, IncAny:0x0, IncAll:0x0
RSVPHop    - Ctype:1, Addr:192.168.46.1, LIH:3
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.1, LspId:57866
```

```
SendTSpec   - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
              MPU:20, MTU:1564
LabelReq    - IfType:General, L3ProtID:2048
RRO         - IpAddr:192.168.46.1, Flags:0x0
              IpAddr:192.168.34.1, Flags:0x0
              IpAddr:192.168.13.1, Flags:0x0
ERO         - IPv4Prefix 192.168.46.2/32, Strict
              IPv4Prefix 192.168.68.2/32, Strict
              IPv4Prefix 192.168.58.1/32, Strict
              IPv4Prefix 192.0.2.10/32, Loose
FRRObj      - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
              ExcAny:0x0, IncAny:0x0, IncAll:0x0
"

*A:PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail

---snip---
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-10 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-10
Path LSP ID       : 57866
From              : 192.0.2.1             To                  : 192.0.2.10
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up

---snip---

FR Prop Admin Gr*: Disabled               Oper FRPropAdmGrp   : Disabled
Propogate Adm Grp: Enabled                Oper Prop Adm Grp   : Enabled
Inter-area        : True
---snip---
Include Groups    :                       Oper Include Groups  :
None                                            None
Exclude Groups    :                       Oper Exclude Groups  :
red                                             red
---snip---
Actual Hops       :
    192.168.13.1 (192.0.2.1) @ n                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @ n                Record Label        : 262141
 -> 192.168.34.2 (192.0.2.4) @ n                Record Label        : 262140
 -> 192.168.46.2  @ n                           Record Label        : 262142
 -> 192.0.2.8 (192.0.2.8) @ n                   Record Label        : 262140
 -> 192.168.68.2  @ n                           Record Label        : 262140
 -> 192.0.2.5 (192.0.2.5) @                     Record Label        : 262140
 -> 192.168.58.1  @                             Record Label        : 262140
 -> 192.0.2.10 (192.0.2.10)                     Record Label        : 262139
 -> 192.168.105.2                               Record Label        : 262139
---snip---
```

Shared Risk Link Groups (SRLG)

Shared risk link groups are explained in chapter Shared Risk Link Groups for RSVP-Based LSP.

SRLGs are also supported in the context of inter-area TE LSPs. SRLGs refer to situations where links in a network share a common fiber (or a common physical attribute). If one link fails, other links in the group may fail as well. Links in the group have a shared risk.

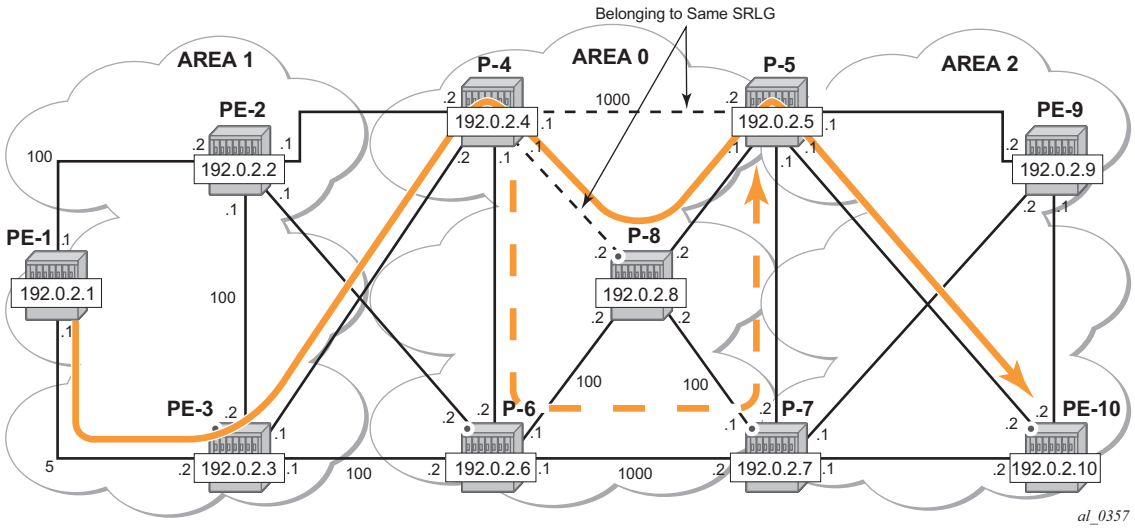
The MPLS TE SRLG feature enhances backup tunnel path selection so that a backup tunnel avoids using links that are in the same SRLG.

Consider the setup in Figure 149, where an inter-area LSP is set up from PE-1 to PE-10 and the path goes through P-8 because of a lower IGP metric. To protect against a node failure of P-8, P-4 (PLR) would normally set up an FRR backup directly to P-5 (MP), because of the lower IGP metric (P-4 to P-5:1000) compared to the IGP traffic via P-6 (P-4 to P-6 to P-7 to P-5:1020).

However, imagine that in this setup the link between P-4 and P-5 link and the link between P-4 and P-8 are part of the same transmission bundle. In this case a cut of that fiber bundle will bring down both the primary and the backup path.

This can be avoided by configuring these two links in the same SRLG group and enabling srlg-frr strict on P-4. In that case the backup will be set up via P-6 as indicated by the dotted line in Figure 149

Figure 149 **Share Risk Link Groups**



SRLG Configuration

On P-4 configure an SRLG group and add the link to P-5 and the link to P-8 to this SRLG group and enable srlg-frr strict.

The srlg-group configuration is required on all nodes that use srlg groups and on the ABR used by the inter-area TE LSP. In this example, it is configured on all nodes.

Note: Enabling or disabling SRLG for FRR is system-wide and requires the MPLS routing instance to be manually set to shutdown and then to no shutdown to activate the change. This may cause service outage. Nokia recommends that the operator incorporates the SRLG into the initial network design and implementation to minimize the traffic loss. In this case, it is sufficient to disable and re-enable RSVP on PE-1.

```
*A:Px# configure router if-attribute srlg-group bundle-red value 1


*A:P-4# configure router
        mpls
            srlg-frr strict
            interface "int-P-4-P-5"
                srlg-group "bundle-red"
            exit
            interface "int-P-4-P-8"
                srlg-group "bundle-red"
            exit


*A:PE-1# configure router rsvp shutdown
*A:PE-1# configure router rsvp no shutdown
```

LSP Configuration

Remove the admin group restriction from the LSP.

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-10"
                no exclude "red"
                no propagate-admin-group
            exit
```

Now check the LSP path on PE-1 and verify that FRR protection is in place.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-10"  path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-10 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
```

```
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-10 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-10
Path LSP ID      : 57876
From             : 192.0.2.1            To                  : 192.0.2.10
Admin State      : Up                   Oper State          : Up
Path Name        : dyn                  Path Type           : Primary
Path Admin       : Up                   Path Oper           : Up
Out Interface    : 1/1/2                Out Label           : 262143
Path Up Time     : 0d 00:01:38          Path Down Time      : 0d 00:00:00
Retry Limit      : 0                    Retry Timer         : 30 sec
Retry Attempt    : 0                    Next Retry In       : 0 sec
BFD Template     : None                 BFD Ping Interval   : 60
BFD Enable       : False

Adspec           : Disabled             Oper Adspec         : Disabled
CSPF             : Enabled              Oper CSPF           : Enabled
Least Fill       : Disabled             Oper LeastFill      : Disabled
FRR              : Enabled              Oper FRR            : Enabled
FRR NodeProtect  : Enabled              Oper FRR NP         : Enabled
FR Hop Limit     : 16                   Oper FRHopLimit     : 16
FR Prop Admin Gr*: Disabled             Oper FRPropAdmGrp   : Disabled
Propogate Adm Grp: Disabled             Oper Prop Adm Grp   : Disabled
Inter-area       : True

---snip---
Actual Hops      :
    192.168.13.1 (192.0.2.1) @ n                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @ n                Record Label        : 262143
 -> 192.168.34.2 (192.0.2.4) @ n                Record Label        : 262143
 -> 192.0.2.8 (192.0.2.8) @ n                   Record Label        : 262143
 -> 192.168.48.2  @ n                           Record Label        : 262143
 -> 192.0.2.5 (192.0.2.5) @                     Record Label        : 262143
 -> 192.168.58.1  @                             Record Label        : 262143
 -> 192.0.2.10 (192.0.2.10)                     Record Label        : 262143
 -> 192.168.105.2                               Record Label        : 262143
Computed Hops    :
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.34.2(SA)
 -> 192.0.2.10(L)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric         : 15
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

On P-4 check the SRLG configuration and verify that the backup is set up via P-6 rather than via P-5.

```
*A:P-4# show router if-attribute srlg-group

===============================================================================
Interface Srlg Groups
===============================================================================
Group Name                      Group Value      Penalty Weight
```

```
-------------------------------------------------------------------------
bundle-red                        1               0
-------------------------------------------------------------------------
No. of Groups: 1
=========================================================================
*A:P-4#

*A:P-4# show router mpls interface

===============================================================================
MPLS Interfaces
===============================================================================
Interface                         Port-id           Adm    Opr    TE-metric
-------------------------------------------------------------------------------
system                            system            Up     Up     None
  Admin Groups                    None
  SRLG Groups                     None
int-P-4-P-5                       1/1/1             Up     Up     None
  Admin Groups                    None
  SRLG Groups                     bundle-red
int-P-4-P-6                       1/1/3             Up     Up     None
  Admin Groups                    None
  SRLG Groups                     None
int-P-4-P-8                       1/2/1             Up     Up     None
  Admin Groups                    red
  SRLG Groups                     bundle-red
int-P-4-PE-2                      1/1/2             Up     Up     None
  Admin Groups                    None
  SRLG Groups                     None
int-P-4-PE-3                      1/1/4             Up     Up     None
  Admin Groups                    None
  SRLG Groups                     None
-------------------------------------------------------------------------------
Interfaces : 6
===============================================================================
*A:P-4#

*A:P-4# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-node192.0.2.8-61444
-------------------------------------------------------------------------------
To             : 192.168.57.1        State               : Up
Out I/F        : 1/1/3               Out Label           : 262140
Up Time        : 0d 00:01:17         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 2
Type           : Dynamic             Bypass Path Cost    : 1020
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.46.1(S)                  Egress Admin Groups : None
 -> 192.168.46.2(S)                  Egress Admin Groups : None
 -> 192.168.67.2(S)                  Egress Admin Groups : None
 -> 192.168.57.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.46.1 (192.0.2.4)         Record Label        : N/A
```

```
 -> 192.168.46.2 (192.0.2.6)         Record Label        : 262140
 -> 192.168.67.2 (192.0.2.7)         Record Label        : 262141
 -> 192.168.57.1 (192.0.2.5)         Record Label        : 262138
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-1-PE-10::dyn
From           : 192.0.2.1           To                  : 192.0.2.10
Avoid Node/Hop : 192.0.2.8           Downstream Label    : 262141
Bandwidth      : 0 Kbps

LSP Name       : LSP-PE-1-PE-10::dyn
From           : 192.0.2.1           To                  : 192.0.2.10
Avoid Node/Hop : 192.0.2.8           Downstream Label    : 262143
Bandwidth      : 0 Kbps

===============================================================================
*A:P-4#
```

Conclusion

Inter-area TE P2P LSPs can be set up based on ERO expansion. With this feature, the head-end does a partial CSPF calculation to its local ABR. On receiving a PATH message with a loose hop ERO, this ABR does a partial or full CSPF calculation to the next ABR to reach the final destination.

FRR protection within the area is available. FRR node protection of the ABR is possible through an MBT on the PLR (node just upstream of the ABR) to the MP (node just downstream of the ABR) or through a dynamically signaled bypass tunnel on the PLR. Dynamic ABR node protection requires that the node ID of the MP node is propagated in the RESV message and that an XRO object is included in the bypass PATH message which makes it possible for the ABR to calculate a path to an MP node.

TE features like BW, path prioritization, path pre-emption, graceful shutdown are supported, as well as propagation of the session attribute with affinity along the LSP path (admin groups) and SRLG.

LDP FEC to BGP Label Route Stitching

This chapter provides information about LDP FEC to BGP label route stitching.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS routers and was initially written for release 13.0.R7. The CLI in the current edition corresponds to release 14.0.R1. Label distribution protocol (LDP) forwarding equivalence class (FEC) to border gateway protocol (BGP) label route stitching was first implemented in SR OS release 8.0.

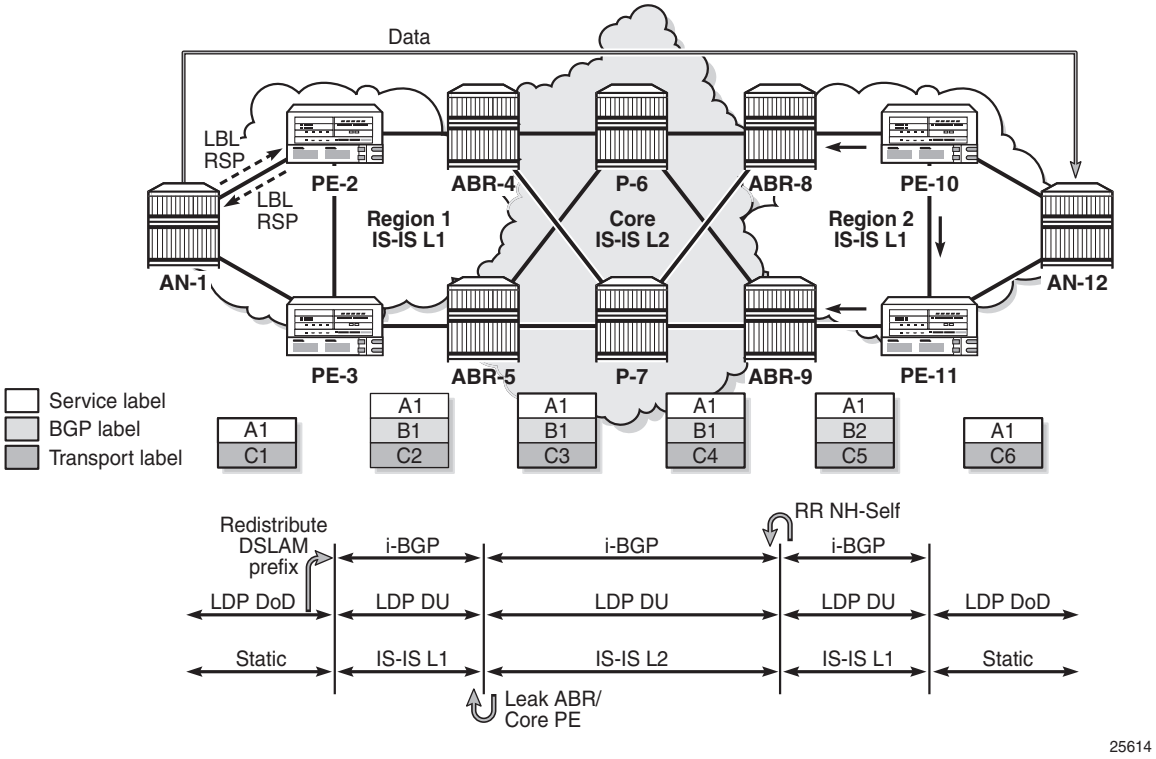
Overview

Stitching of an LDP FEC to a BGP labeled route allows LDP-capable PE devices, such as digital subscriber line access multiplexers (DSLAMs), to offer services to LDP-capable PE devices in other areas or domains without the need to support BGP labeled routes. This feature is used in a large network to provide services across multiple areas or autonomous systems.

When BGP is used to distribute a particular route, it can at the same time be used to distribute a multiprotocol label switching (MPLS) label that is mapped to that route. The label mapping information for a particular route is appended to the same BGP update message that is used to distribute the route. This is described in RFC 3701, *Carrying Label Information in BGPv4*.

Figure 150 shows a network with a core area and regional areas. The components of the network are defined in the paragraphs that follow. For simplification, the control plane is displayed from right to left and the data plane from left to right.

Figure 150 **LDP FEC to BGP Label Route Stitching**



The access nodes (ANs) are DSLAMs that support LDP. In seamless MPLS networks, LDP downstream-on-demand (DoD) label advertisement can be used between the ANs and their next-hop PEs. Usually, MPLS routers implement LDP downstream unsolicited (DU) label distribution, advertising MPLS labels for all routes in their routing information base (RIB). The ANs do not need to have LDP bindings for all prefixes in the network. The ANs will request the LDP labels they need. LDP DoD improves scalability in large networks.

BGP route reflectors (RRs) can be used to improve scalability. The RR can be any node; it does not need to be an area border router (ABR) as in Figure 150. If the RR is not in the forwarding path, it does not need to be capable of forwarding MPLS packets.

There are different areas for IS-IS: routers in the core network have level 2 (L2) capability, while the routers in the regional areas have level 1 (L1) capability and the ABRs have both. In each ABR, an IS-IS export policy is defined to leak the routes from the core to the regional networks.

Passing L1 routes (regional) into L2 (core) is inherent to IS-IS and cannot be controlled through policy. Passing L2 routes to L1 can be controlled through policy.

Only nodes within a regional area, and the ABR nodes in the same area, exchange LDP FECs. PE routers in a regional area learn the reachability of PE routers in other regional areas by way of RFC 3107 BGP labeled routes redistributed by the remote nodes.

The label stack contains three labels for packets sent in an Epipe service between the access nodes:

- The DSLAMs push a service label to the packets sent in the Epipe service. The service label remains unchanged end-to-end between the DSLAMs. The service label is popped by the remote DSLAM and is the inner label of the label stack.
- The BGP label is the middle label of the label stack and should be regarded as a transport label. The transport label stack has increased to two labels: BGP and LDP transport label. BGP labeled routes are not supported on the DSLAMs. The BGP label is pushed by the PE nearest to the local DSLAM and is swapped at the BGP next hop, which can be a BGP peer configured with next-hop-self or the PE that is the remote endpoint of the BGP tunnel. The BGP label is popped by the PE at the end of the BGP tunnel.
- The DSLAMs push an LDP transport label to the packets sent to the remote DSLAM. At the PE nearest to the local DSLAM, the LDP transport label is stitched to the BGP label. At the same time, that same PE pushes the LDP transport label to reach the BGP next hop. The LDP transport label is swapped in every label switching router (LSR) and popped by the PE nearest to the remote DSLAM. That PE also pops the BGP label, which is stitched to the LDP transport label that is pushed to the packets sent to the remote DSLAM. This LDP label is the top label of the label stack.

When PE-2 is an ingress label edge router (iLER) sending a service packet to the remote PE, PE-2 inserts the BGP route label to reach the remote PE and an LDP label to reach the next-hop router. In Figure 150, this is the remote ABR because it has set next-hop-self (NH-Self).

The access node AN-1, which is a DSLAM, can behave as a PE router for Epipe services. It will need to establish a pseudowire (PW) to a PE in a different regional area via LSR PE-2. In this case, PE-2 performs the following actions:

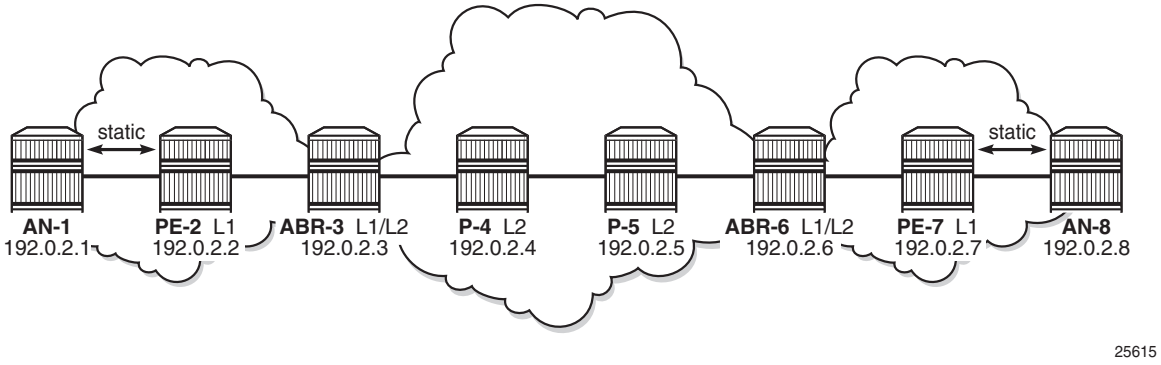
- Translates the LDP FEC it learned from AN-1 into a BGP labeled route and redistributes it using iBGP within its area. This is in addition to redistributing the FEC to its LDP neighbors in the same area.
- Translates the BGP labeled routes it learns through iBGP into an LDP FEC and redistributes it to its LDP neighbors in the same area. AN-1 requests the LDP FEC of the remote DSLAM (AN-12) using LDP DoD.

- When a data packet is received from AN-1 with destination AN-12, PE-2 swaps the LDP label into a BGP label and pushes the LDP label to reach the BGP next hop. When a data packet with destination AN-1 is received on PE-2 from the local ABR (ABR-4), the top transport label (LDP) is removed and the BGP label is swapped for the LDP label corresponding to AN-1.

Configuration

Figure 151 shows the example topology that is used in this chapter. An Epipe will be established between the access nodes AN-1 and AN-8. PE-2 and PE-7 will stitch the LDP FECs to BGP label routes. In the regional areas, IS-IS L1 capability is used while in the core area IS-IS L2 is used. The ABR nodes support both IS-IS L1 and L2 and export routes from L2 to L1. Static routes are configured between the access nodes and the next-hop PEs.

Figure 151 **Example Topology**



Initial Configuration

Note: In the example topology, all nodes are 7750 SRs, while the ANs should be access devices, such as DSLAMs. The limitation of this approach is that the ANs (SRs) in this setup can only request a label for the directly connected PE and not for their remote peer AN; however, DSLAMs do not have this limitation. Consequently, the Epipe service in this configuration will be operationally down because the transport tunnel is down.

All nodes have the following initial configuration:

- Cards, media dependent adapters (MDAs), ports

- Router interfaces



Note: The IP addresses for the link between node A and node B are in subnet 192.168.AB.0/0. The node with the lowest ID has IP address 192.168.AB.1/30 and the node with the highest ID has IP address 192.168.AB.2/30.

```
*A:PE-2# configure router
        interface "int-PE-2-ABR-3"
            address 192.168.23.1/30
            port 1/1/1
        exit
        interface "int-PE-2-AN-1"
            address 192.168.12.2/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.2/32
        exit
```

- Static routes are configured between AN-1 and PE-2 and between PE-7 and AN-8:

```
*A:AN-1# configure router
        static-route-entry 0.0.0.0/0
            next-hop 192.168.12.2
                no shutdown
            exit
        exit

*A:PE-2# configure router
 static-route-entry 192.0.2.1/32
            next-hop 192.168.12.1
                no shutdown
            exit
        exit
```

- IS-IS (alternatively, OSPF could have been used)
  - PE-2 and PE-7 have L1 capability.

```
*A:PE-2# configure router
        isis
            level-capability level-1
            area-id 49.0001
            interface "system"
            exit
            interface "int-PE-2-ABR-3"
                interface-type point-to-point
            exit
        exit
```

  - P-4 and P-5 have L2 capability.

– ABR-3 and ABR-6 have L1 capability on the interfaces toward the PE routers in the regional areas and L2 capability on the interfaces toward the P routers in the core area. A policy is applied to export the system IP addresses from L2 to L1:

```
*A:ABR-3# configure router
        policy-options
            begin
            prefix-list "systemIP"
                prefix 192.0.2.0/24 longer
            exit
            policy-statement "export-L2-to-L1"
                entry 10
                    from
                        protocol isis
                        prefix-list "systemIP"
                        level 2
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
        isis
            area-id 49.0001
            export "export-L2-to-L1"
            interface "system"
            exit
            interface "int-ABR-3-PE-2"
                level-capability level-1
                interface-type point-to-point
            exit
            interface "int-ABR-3-P-4"
                level-capability level-2
                interface-type point-to-point
            exit
        exit
```

• LDP

– Link LDP is enabled on all router interfaces on all nodes, including the ANs.

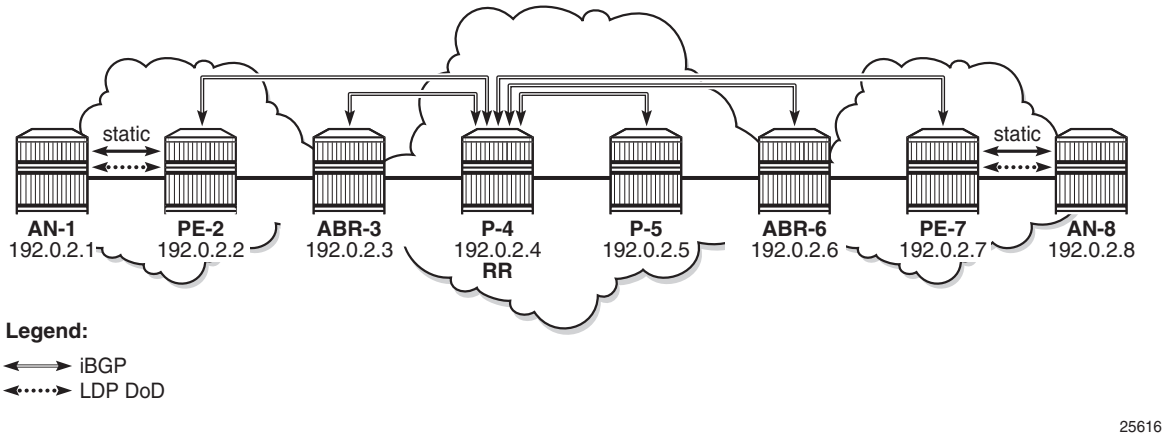
– On PE-2 and PE-7, DoD is enabled in the session parameters for the peering sessions with the ANs:

```
*A:PE-2# configure router
        ldp
            session-parameters
                peer 192.0.2.1
                    dod-label-distribution
                exit
            exit
            interface-parameters
                interface "int-PE-2-AN-1
                exit
                interface "int-PE-2-ABR-3"
                exit
            exit
```

Configure BGP

BGP is configured on all nodes except the ANs. Figure 152 shows that P-4 is the RR.

Figure 152 **BGP Enabled with P-4 as RR**



The initial BGP configuration on PE-2 is the following:

```
*A:PE-2# configure router
        autonomous-system 64496
        bgp
            group "internal"
                peer-as 64496
                neighbor 192.0.2.4
                exit
            exit
            no shutdown
        exit
```

The configuration is identical for ABR-3, P-5, ABR-6, and PE-7. The initial BGP configuration on the RR P-4 is:

```
*A:P-4# configure router
        autonomous-system 64496
        bgp
            cluster 1.1.1.1
            group "internal"
                peer-as 64496
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.6
                exit
                neighbor 192.0.2.7
```

```
            exit
        exit
        no shutdown
    exit
```

The BGP sessions can be shown with the following command:

```
*A:P-4# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 5
Total BGP Paths         : 13          Total Path Memory           : 2392
---snip---
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
                64496       3    0 00h00m18s 0/0/0 (IPv4)
                            3    0
192.0.2.3
                64496       3    0 00h00m18s 0/0/0 (IPv4)
                            3    0
192.0.2.5
                64496       3    0 00h00m18s 0/0/0 (IPv4)
                            3    0
192.0.2.6
                64496       3    0 00h00m18s 0/0/0 (IPv4)
                            3    0
192.0.2.7
                64496       3    0 00h00m09s 0/0/0 (IPv4)
                            3    0
-------------------------------------------------------------------------------
*A:P-4#
```

Export Policies for BGP and LDP

LDP FEC to BGP label route stitching is established by configuring separate tunnel table route export policies in both protocols. At the local next-hop PE, the LDP FEC of the local AN must be translated into a BGP label and at the remote PE, the BGP label must be translated into an LDP FEC.

An export policy for the export from LDP to BGP must be defined on the PE nodes.

```
*A:PE-2# configure router
```

```
        policy-options
            begin
            prefix-list "local-AN"
                prefix 192.0.2.1/32 exact
            exit
            prefix-list "remote-AN"
                prefix 192.0.2.8/32 exact
            exit
            policy-statement "export-BGP"
                entry 10
                    from
                        protocol ldp
                        prefix-list "local-AN"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

On PE-7, the policy statement is identical, but the prefix lists are different.

This export policy must be applied in the BGP context: either in the general settings or per group or per neighbor.

```
*A:PE-2# configure router
        bgp
            group "internal"
                export "export-BGP"
            exit
        exit
```

In a similar way, BGP labels must be exported to LDP on the PE routers. The export policy is configured as follows, with the prefix list already defined earlier:

```
*A:PE-2# configure router
        policy-options
            begin
            prefix-list "remote-AN"
                prefix 192.0.2.8/32 exact
            exit
            policy-statement "export-LDP"
                entry 10
                    from
                        protocol bgp
                        prefix-list "remote-AN"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

This export policy is applied in the LDP context, as follows:

```
*A:PE-2# configure router
        ldp
            export-tunnel-table "export-LDP"
        exit
```

Advertise Labels in BGP Updates

BGP should evaluate the activated /32 LDP prefixes in the export policy. This needs to be configured on the endpoints of the BGP tunnel and on the RR:

```
*A:PE-2# configure router bgp group "internal" neighbor 192.0.2.4 advertise-
label ipv4 include-ldp-prefix
*A:PE-7# configure router bgp group "internal" neighbor 192.0.2.4 advertise-
label ipv4 include-ldp-prefix
*A:P-4# configure router
        bgp
            group "internal"
                neighbor 192.0.2.2
                    advertise-label ipv4 include-ldp-prefix
                exit
                neighbor 192.0.2.7
                    advertise-label ipv4 include-ldp-prefix
                exit
            exit
```

Configuring advertise-label with the include-ldp-prefix argument implies that all activated /32 LDP FEC prefixes will be sent to the remote BGP peer as an RFC 3107 formatted label.

Configuring advertise-label without the include-ldp-prefix argument implies that only core IPv4 routes learned from the route table manager (RTM) are advertised as RFC 3107 BGP labeled routes to this neighbor. No stitching of LDP FEC to the BGP labeled route will be performed for this neighbor, even if the same prefix was learned from LDP.

The BGP open messages contain address family AFI=1 and SAFI=1 between the RR and peers without the advertise-label argument. This address family is used for IPv4 unicast. See Cap_Code MP-BGP. Bytes 0x0 0x1 (AFI=1) 0x0 0x1 (SAFI=1).

```
*A:ABR-3# show debug
debug
    router "Base"
        bgp
            open
            update
        exit
    exit
exit
2 2016/03/30 08:02:25.73 UTC MINOR: DEBUG #2001 Base BGP
"BGP: OPEN
Peer 1: 192.0.2.4 - Received BGP OPEN: Version 4
```

```
   AS Num 64496: Holdtime 90: BGP_ID 192.0.2.4: Opt Length 16
   Opt Para: Type CAPABILITY: Length = 14: Data:
     Cap_Code MP-BGP: Length 4
       Bytes: 0x0 0x1 0x0 0x1
     Cap_Code ROUTE-REFRESH: Length 0
     Cap_Code 4-OCTET-ASN: Length 4
       Bytes: 0x0 0x0 0xfb 0xf0
"
```

Between peers that advertise the labels, AFI=1 and SAFI=4. This address family is used for labeled IPv4 unicast. The following BGP open message is seen on PE-2:

```
8 2016/03/30 08:06:24.34 UTC MINOR: DEBUG #2001 Base BGP
"BGP: OPEN
Peer 1: 192.0.2.4 - Received BGP OPEN: Version 4
   AS Num 64496: Holdtime 90: BGP_ID 192.0.2.4: Opt Length 16
   Opt Para: Type CAPABILITY: Length = 14: Data:
     Cap_Code MP-BGP: Length 4
       Bytes: 0x0 0x1 0x0 0x4
     Cap_Code ROUTE-REFRESH: Length 0
     Cap_Code 4-OCTET-ASN: Length 4
       Bytes: 0x0 0x0 0xfb 0xf0
"
```

The BGP update messages also show a difference. As expected, there is no label information between the RR and ABR-3 because the option to advertise labels is not enabled.

```
4 2016/03/30 08:07:22.58 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 42
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 3 Len: 4 Nexthop: 192.0.2.4
    Flag: 0x80 Type: 4 Len: 4 MED: 1
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.7
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    NLRI: Length = 5
        192.0.2.8/32
"
```

The BGP update messages between peers that advertise labels contain additional information for the same prefix 192.0.2.8/32. The address family is IPV4-Labeled and the label is 262133. This message is seen on PE-2:

```
9 2016/03/30 08:06:56.26 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 63
    Flag: 0x90 Type: 14 Len: 17 Multiprotocol Reachable NLRI:
```

```
        Address Family IPV4-Labeled
        NextHop len 4 NextHop 192.0.2.7
        192.0.2.8/32 Label 262136
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 3 Len: 4 Nexthop: 192.0.2.7
    Flag: 0x80 Type: 4 Len: 4 MED: 1
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.7
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
"
```

After applying the export policy from BGP to LDP and enabling the option in BGP to advertise labels for the LDP FEC prefixes, LDP will look for BGP route entries in the tunnel table. If a /32 BGP labeled route matches a prefix entry in the export policy, LDP originates an LDP FEC for this prefix, stitches it to the BGP labeled route, and redistributes the LDP FEC to its BGP neighbors. This can be shown on PE-7, as follows.

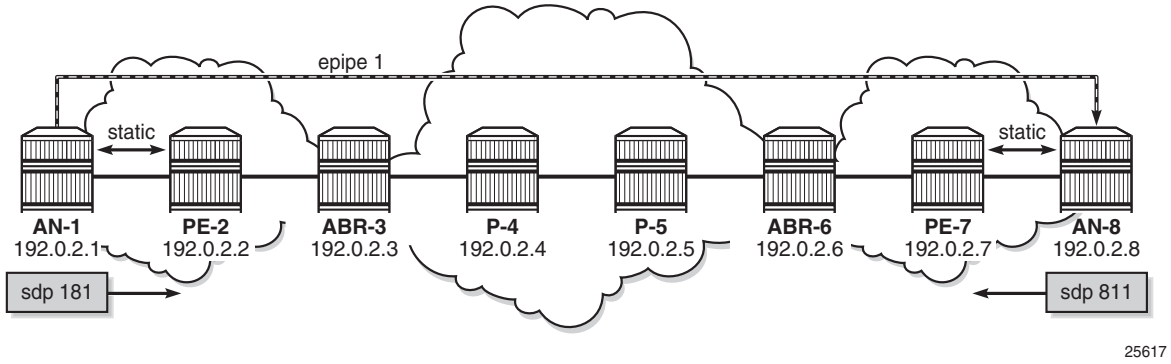
```
*A:PE-7# show router bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.8                     262143         262136         InternalLdp
-------------------------------------------------------------------------------
Total Labels allocated:   1
===============================================================================
*A:PE-7#
```

The label received from AN-8 is 262143. The label origin is InternalLdp. This LDP label is stitched to BGP label 262136 that will be advertised by PE-7 to its BGP peers, such as PE-2. Traffic sent from AN-1 toward AN-8 will be forwarded from PE-2 to its BGP NH PE-7 using BGP label 262136. In PE-7, the BGP label is stitched to LDP label 262143 that will be used to forward the packet to AN-8.

Configure SDP and Epipe

An end-to-end Epipe service is established between AN-1 and AN-8, as shown in Figure 153.

Figure 153 **End-to-End Epipe Service**



Note: : In this setup, ANs are simulated by 7750 SRs. Due to this limitation, the SDP used by the Epipe service will not become operational. 7750 SR only supports single-hop DoD, which implies that AN-1 can only request a label for the LSR ID of the directly connected router, PE-2, not of remote nodes, such as AN-8. Similarly, AN-8 cannot request a label for AN-1. Therefore, it is not possible to have an LDP LSP between the ANs and the SDP will be down because there is no transport tunnel.

The SDP is configured on AN-1, as follows:

```
*A:AN-1# configure service
        sdp 181 mpls create
            far-end 192.0.2.8
            ldp
            no shutdown
        exit
```

Epipe 1 is configured on AN-1, as follows:

```
*A:AN-1# configure service
        epipe 1 customer 1 create
            sap 1/2/1:1 create
            exit
            spoke-sdp 181:1 create
            exit
            no shutdown
        exit
```

The configuration of the SDP and Epipe on AN-8 is similar.

The SDP is down because there is no transport tunnel, which can be shown as follows:

```
*A:AN-1# show service sdp detail
===============================================================================
Services: Service Destination Points Details
```

```
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 181  -192.0.2.8
-------------------------------------------------------------------------------
Description           : (Not Specified)
SDP Id                : 181                    SDP Source         : manual
Admin Path MTU        : 0                      Oper Path MTU      : 0
Delivery              : MPLS
Far End               : 192.0.2.8
Tunnel Far End        : 192.0.2.8              LSP Types          : LDP

Admin State           : Up                     Oper State         : Down
Signaling             : TLDP                   Metric             : 0
Acct. Pol             : None                   Collect Stats      : Disabled
Last Status Change    : 03/30/2016 08:35:28    Adv. MTU Over.     : No
Last Mgmt Change      : 03/30/2016 08:09:06    VLAN VC Etype      : 0x8100
Bw BookingFactor      : 100                    PBB Etype          : 0x88e7
Oper Max BW(Kbps)     : 0                      Avail BW(Kbps)     : 0
Net-Domain            : default                Egr Interfaces     : n/a
Flags                 : TranspTunnDown
---snip---
```

A targeted LDP session is established between AN-1 and AN-8, which can be shown as follows:

```
*A:AN-1# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   34389     34348     1d 01:29:46
192.0.2.8:0         Targeted  Established   1967      1969      0d 02:55:36
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
---snip---
```

LDP FEC Resolution at PE-2 for Traffic from AN-8 to AN-1

The following steps occur at PE-2 for the LDP FEC resolution for traffic from AN-1 toward AN-8. The situation is similar for PE-7.

Step 1. After receiving an LDP label binding message for LDP FEC for the system address of AN-1 (192.0.2.1/32), PE-2 installs this prefix in the label forwarding information base (LFIB). PE-2 programs a push and a swap next hop label forwarding entry (NHLFE) in the egress data path to forward packets to prefix 192.0.2.1/32.



Note: PE-2 installs this LDP FEC in the LFIB only if there is an exact match of the prefix 192.0.2.1/32 in the routing table or a longest match of the prefix in the routing table, in case aggregate-prefix-match is configured on PE-2. The advertising LDP neighbor (AN-1) must be the next hop to reach the FEC prefix.

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262143
192.168.12.1                            1/1/2

192.0.2.1/32                            Swap            262142    262143
192.168.12.1                            1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-2#
```

Step 2. PE-2 programs a tunnel entry for prefix 192.0.2.1/32 in the tunnel table.

```
*A:PE-2# show router tunnel-table 192.0.2.1/32
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65537     9        192.168.12.1   1
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-2#
```

Step 3. PE-2 advertises a new FEC label binding for prefix 192.0.2.1/32 toward all its LDP neighbors. The result can be shown on ABR-3, as follows:

```
*A:ABR-3# show router ldp bindings prefixes prefix 192.0.2.1/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
```

```
            (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                      IngLbl                    EgrLbl
Peer                                        EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                 --                       262142
192.0.2.2:0                                  --
  --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 1
===============================================================================
*A:ABR-3#
```

Step 4. When BGP learns the LDP FEC via the tunnel table and the FEC prefix exists in the BGP route policy, PE-2 originates a BGP labeled route toward all its neighbors that have the advertise label for LDP FEC prefixes enabled. The following output shows the BGP labeled route to RR P-4 for prefix 192.0.2.1/32.

```
*A:PE-2# show router bgp routes ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.0.2.2
Path Id        : None
To             : 192.0.2.4
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 1
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
```

```
IPv4 Label     : 262136
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-2
```

BGP Labeled Route Resolution at PE-2 for Traffic from AN-1 to AN-8

The following steps occur at PE-2 for the BGP labeled route resolution for traffic from AN-1 toward AN-8. The situation is similar for PE-7.

Step 1. When there is an LDP LSP to the BGP neighbor advertising the route (PE-7) and PE-2 has received a BGP labeled route via iBGP for AN-8, PE-2 installs the prefix 192.0.2.8/32 in BGP. The LDP tunnel toward PE-7 is shown, then the BGP route toward AN-8, as advertised by PE-7.

```
*A:PE-2# show router tunnel-table 192.0.2.7
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.7/32      ldp       MPLS  65542     9        192.168.23.2    50
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-2#

*A:PE-2# show router bgp routes 192.0.2.8/32
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.0.2.8/32                                         100         1
      192.0.2.7                                            None        262136
```

```
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

The BGP label for traffic toward AN-8 is 262136. This is the middle label in the label stack. The next hop is PE-7.

Step 2. PE-2 programs a swap NHLFE in the egress data path to forward packets to 192.0.2.8/32, as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.8/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.8/32(B)                           Swap            262135    262136
192.0.2.7                                 LspId 65542

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 1
===============================================================================
*A:PE-2#
```

The (B) indicates that 192.0.2.8/32 is a BGP next hop. The ingress label is the LDP transport label from AN-1 for prefix 192.0.2.8/32. The LSP ID 65542 corresponds to the LDP LSP toward egress next-hop PE-7, as shown earlier in the tunnel table. The BGP egress label for traffic toward AN-8 is 262136.

Step 3. PE-2 programs a tunnel table entry for 192.0.2.8/32.

```
*A:PE-2# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65537     9        192.168.12.1    1
192.0.2.3/32      ldp       MPLS  65538     9        192.168.23.2    10
192.0.2.4/32      ldp       MPLS  65539     9        192.168.23.2    20
192.0.2.5/32      ldp       MPLS  65540     9        192.168.23.2    30
192.0.2.6/32      ldp       MPLS  65541     9        192.168.23.2    40
192.0.2.7/32      ldp       MPLS  65542     9        192.168.23.2    50
```

```
192.0.2.8/32      bgp       MPLS  262145    12       192.0.2.7      1000
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-2#
```

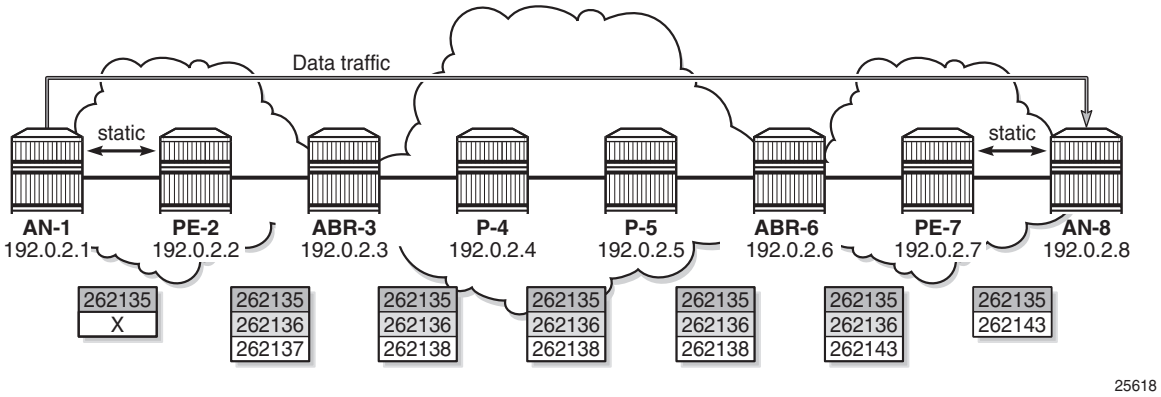
This is the only BGP tunnel in the tunnel table; all tunnels toward the other nodes are LDP tunnels. LDP routes have preference over BGP labeled routes, but there is no LDP route toward 192.0.2.8/32. Therefore, the BGP tunnel will be used for traffic destined to AN-8.

Step 4. PE-2 advertises a new FEC label binding for prefix 192.0.2.8/32 toward AN-1. This is only done after AN-1 requests a label for prefix 192.0.2.8/32, because LDP DoD is enabled. This is possible if the ANs are DSLAMs, but not in this setup with SRs.

Data Plane Overview for PE-2

Figure 154 shows the label stacks that are used for traffic from AN-1 to AN-8.

Figure 154 **Label Stacks for Traffic from AN-1 to AN-8**



Note: The LDP transport label that is pushed by AN-1 is not known because of the single-hop LDP DoD implementation in 7750 SR. AN-1 cannot request the LDP label for AN-8. Therefore, the LDP transport label is represented by "X".

The service label added for Epipe 1 on AN-1 for egress traffic to AN-8 is 262135. Ingress traffic on AN-1 has service label 262135. This can be shown as follows:

```
*A:AN-1# show service id 1 labels
===============================================================================
```

```
Martini Service Labels
==============================================================================
Svc Id     Sdp Binding        Type  I.Lbl                E.Lbl
------------------------------------------------------------------------------
1          181:1              Spok  262135               262135
------------------------------------------------------------------------------
Number of Bound SDPs : 1
------------------------------------------------------------------------------
==============================================================================
*A:AN-1#
```

This service label remains unchanged end-to-end.

As shown earlier, the (middle) BGP label for traffic with destination AN-8 is 262136, as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.8/32
==============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
==============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
==============================================================================
LDP IPv4 Prefix Bindings (Active)
==============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
------------------------------------------------------------------------------
192.0.2.8/32(B)                         Swap            262135    262136
192.0.2.7                               LspId 65542

------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 1
==============================================================================
*A:PE-2#
```

The next hop is PE-7, which is the PE nearest to AN-8. The BGP label will not be swapped between PE-2 and PE-7 because there is no intermediate node that has set next-hop-self. An intermediate node with next-hop-self would become the next hop instead of PE-7. The BGP label is only added or removed by the next-hop PE.

On PE-2, when a service packet with destination AN-8 is received, the ingress LDP transport label X is swapped into BGP label 262136. To reach PE-7, which is the BGP next hop for traffic toward AN-8, another LDP transport label 262137 is pushed to the packet, as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.7/32
==============================================================================
```

```
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.7/32                              Push              --      262137
192.168.23.2                              1/1/1

192.0.2.7/32                              Swap            262137    262137
192.168.23.2                              1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-2#
```

The next hop is ABR-3, where the ingress label 262137 is swapped to egress label 262138, as follows:

```
*A:ABR-3# show router ldp bindings active prefixes prefix 192.0.2.7/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.7/32                              Push              --      262138
192.168.34.2                              1/1/1

192.0.2.7/32                              Swap            262137    262138
192.168.34.2                              1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:ABR-3#
```

In the subsequent LSRs, the transport label is swapped, as follows:

On P-4:

```
*A:P-4# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---

192.0.2.7/32                                Swap            262138    262138
192.168.45.2                                1/1/1

-------------------------------------------------------------------------------
```

On P-5:

```
*A:P-5# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---

192.0.2.7/32                                Swap            262138    262138
192.168.56.2                                1/1/1

-------------------------------------------------------------------------------
```

On ABR-6:

```
*A:ABR-6# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---
192.0.2.7/32                                Swap            262138    262143
192.168.67.2                                1/1/1

-------------------------------------------------------------------------------
```

On PE-7, the LDP label 262143 is popped, as follows:

```
*A:PE-7# show router ldp bindings active prefixes prefix 192.0.2.7/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.7)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                      Op              IngLbl    EgrLbl
EgrNextHop                                  EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.7/32                                Pop             262143      --
  --                                          --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 1
```

```
===============================================================================
*A:PE-7#
```

The BGP label is also popped and mapped onto LDP label 262143 that will be pushed by PE-7 on packets toward AN-8.

```
*A:PE-7# show router ldp bindings active prefixes prefix 192.0.2.8/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.7)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.8/32                            Push              --      262143
192.168.78.2                            1/1/1

192.0.2.8/32                            Swap            262137    262143
192.168.78.2                            1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-7#
```

OAM

The following operations, administration, and maintenance (OAM) commands can be launched to validate an LDP FEC stitched to a BGP IPv4 labeled route and vice versa.

```
*A:PE-2# oam lsp-ping bgp-label prefix 192.0.2.8/32
LSP-PING 192.0.2.8/32: 80 bytes MPLS payload
Seq=1, send from intf int-PE-2-ABR-3, reply from 192.0.2.8
       udp-data-len=32 ttl=255 rtt=6.27ms rc=4 (NoFECMapping)
---- LSP 192.0.2.8/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 6.27ms, avg = 6.27ms, max = 6.27ms, stddev = 0.000ms
*A:PE-2#
```

The detailed output includes the LDP bindings for the destination prefix.

```
*A:PE-2# oam lsp-ping bgp-label prefix 192.0.2.8/32 detail
LSP-PING 192.0.2.8/32: 80 bytes MPLS payload
Seq=1, send from intf int-PE-2-ABR-3, reply from 192.0.2.8
       udp-data-len=32 ttl=255 rtt=4.59ms rc=4 (NoFECMapping)
---- LSP 192.0.2.8/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 4.59ms, avg = 4.59ms, max = 4.59ms, stddev = 0.000ms
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                      IngLbl                    EgrLbl
Peer                                        EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.8/32                                262135U                   262136
192.0.2.1:0                                 LspId 65542
192.0.2.7

192.0.2.8/32                                262135U                   262136
192.0.2.3:0                                 LspId 65542
192.0.2.7

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 2
===============================================================================
*A:PE-2#
```

In a similar way, LSP trace can validate LDP FEC to BGP label route stitching:

```
*A:PE-2# oam lsp-trace bgp-label prefix 192.0.2.8/32
lsp-trace to 192.0.2.8/32: 0 hops min, 0 hops max, 104 byte packets
1  192.0.2.3  rtt=0.696ms rc=8(DSRtrMatchLabel)
2  192.0.2.4  rtt=3.08ms rc=8(DSRtrMatchLabel)
3  192.0.2.5  rtt=3.33ms rc=8(DSRtrMatchLabel)
4  192.0.2.6  rtt=4.78ms rc=8(DSRtrMatchLabel)
5  192.0.2.7  rtt=5.76ms rc=8(DSRtrMatchLabel) rsc=1
6  192.0.2.8  rtt=6.38ms rc=4(NoFECMapping) rsc=1
*A:PE-2#
```

The detailed output includes the BGP label to LDP label mapping information at the PE:

```
*A:PE-2# oam lsp-trace bgp-label prefix 192.0.2.8/32 detail
lsp-trace to 192.0.2.8/32: 0 hops min, 0 hops max, 104 byte packets
1  192.0.2.3  rtt=1.52ms rc=8(DSRtrMatchLabel)
2  192.0.2.4  rtt=0.955ms rc=8(DSRtrMatchLabel)
3  192.0.2.5  rtt=1.52ms rc=8(DSRtrMatchLabel)
4  192.0.2.6  rtt=3.55ms rc=8(DSRtrMatchLabel)
```

```
5  192.0.2.7  rtt=4.47ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.78.2 ifaddr=192.168.78.2 iftype=ipv4Numbered MRU=1560
           label[1]=262143 protocol=3(LDP)
6  192.0.2.8  rtt=5.37ms rc=4(NoFECMapping) rsc=1
*A:PE-2#
```

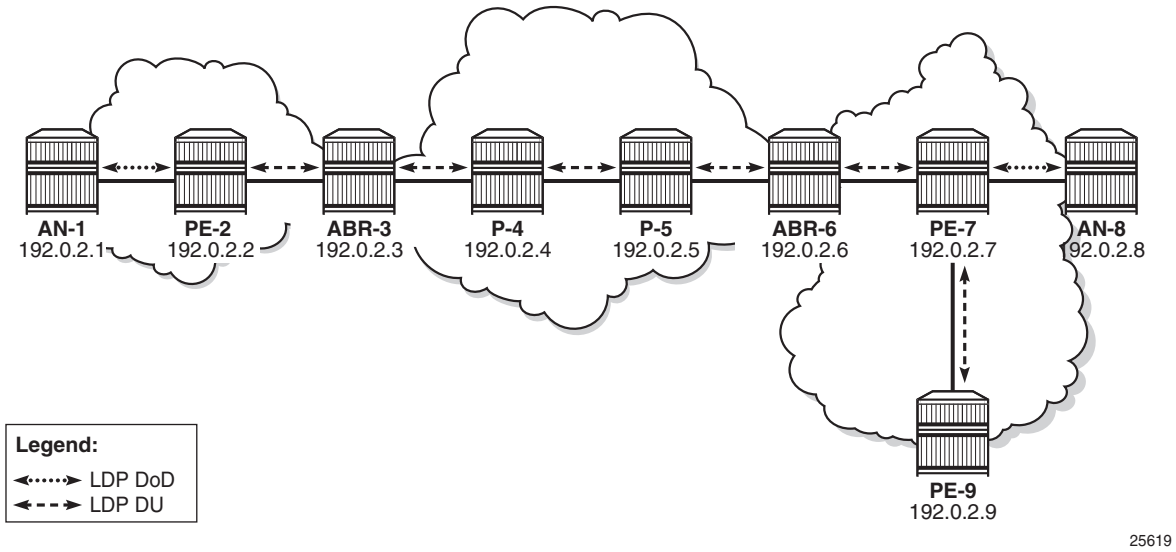
Block BGP Label Bindings to LDP DU Peers

On a PE, labeled BGP prefixes are exported to LDP to allow LDP DoD peers to request these labels. LDP DU peers will also get all labeled BGP prefixes if not explicitly blocked by an LDP export policy, based on prefix lists. This can result in a high administrative and operational effort in large networks.

Blocking BGP label bindings to LDP DU peers is less labor-intensive because per-peer export policies are re-evaluated on NH type change (such as from BGP to LDP or to "unresolved state"), not only on a configuration change.

Figure 155 shows the extended topology used for this configuration. The additional PE router, PE-9, does not need to know the BGP labeled prefixes. LDP DU is used between PE-7 and PE-9.

Figure 155 **Block BGP Label Bindings to LDP DU Peer PE-9**



Blocking BGP label bindings to LDP DU peers can be achieved in two ways:

1. LDP export policy based on prefix list.
2. LDP export policy based on BGP NH type change. No prefix list is required.

To compare the two, both are described.

LDP Export Policy Based on Prefix List

Before applying the policy to block BGP label bindings from PE-7 to PE-9, the LDP bindings on PE-9 for prefix 192.0.2.1 are the following:

```
*A:PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.9)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                    IngLbl                    EgrLbl
Peer                                      EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                --                      262135
192.0.2.7:0                                 --
  --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 1
===============================================================================
*A:PE-9#
```

The following policy created on PE-7 is based on a prefix list that only contains the system address of the remote AN: 192.0.2.1.

```
*A:PE-7# configure router
        policy-options
            begin
            prefix-list "remote-AN"
                prefix 192.0.2.1/32 exact
            exit
            policy-statement "block-BGP-bindings-remote-AN"
                entry 10
                    from
                        prefix-list "remote-AN"
                    exit
                    action reject
                exit
            exit
            commit
        exit
```

The policy is applied on PE-7 in the LDP session-parameters context for peer 192.0.2.9.

```
*A:PE-7# configure router
        ldp
            session-parameters
                peer 192.0.2.9
                    export-prefixes "block-BGP-bindings-remote-AN"
                exit
            exit
        exit
```

After the policy is applied, there are no LDP bindings for prefix 192.0.2.1 on PE-9:

```
*A:PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.9)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                      IngLbl                    EgrLbl
Peer                                        EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-9#
```

The original situation is restored by removing the export prefixes in the LDP session-parameters context on PE-7.

```
*A:PE-7# configure router ldp session-parameters peer 192.0.2.9 no export-prefixes
*A:PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.9)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                      IngLbl                    EgrLbl
Peer                                        EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                  --                      262135
192.0.2.7:0                                   --
  --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 1
===============================================================================
*A:PE-9#
```

LDP Export Policy Based on BGP NH Type Change

The "from protocol bgp" argument will have a different meaning in the context of per-peer and targeted export policies. For those types of policies, policies are re-evaluated on NH type change; for example, from BGP to LDP or from LDP to "unresolved state". This requires less configuration because no prefix list needs to be specified. The following policy is configured on PE-7.

```
*A:PE-7# configure router
        policy-options
            begin
            policy-statement "block-BGP-to-LDP-DU"
                entry 10
                    from
                        protocol bgp
                    exit
                    action reject
                exit
            exit
            commit
        exit
```

The policy is applied in the LDP session-parameter context for peer 192.0.2.9.

```
*A:PE-7# configure router
        ldp
            session-parameters
                peer 192.0.2.9
                    export-prefixes "block-BGP-to-LDP-DU"
                exit
            exit
```

PE-7 will not send BGP label mapping information for prefix 192.0.2.1/32 to PE-9, or for any other prefix of a remote AN. In this example, AN-1 with prefix 192.0.2.1/32 is the only remote AN for PE-7.

```
*A:PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.9)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                  IngLbl                    EgrLbl
Peer                                    EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-9#
```

Conclusion

LDP FEC to BGP label route stitching allows LDP-capable PE devices, such as DSLAMs, to offer services to LDP-capable PE devices in other areas or domains without the need to support BGP labeled routes. This feature can be used in a seamless MPLS environment.

LDP over RSVP Using OSPF as IGP

This chapter provides information about label distribution protocol (LDP) over resource reservation protocol with traffic engineering (RSVP-TE), also called LDPoRSVP, that uses RSVP label switched paths (LSPs) as a transport vehicle to carry the packets using LDP LSPs.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Additional Topics
- Conclusion

Applicability

This chapter is applicable to SR OS. Initially, it was written for and tested on 7.0.R5, but the CLI in this edition corresponds to release 14.0.R1. There are no prerequisites.

Overview

Introduction

Only user packets are tunneled over RSVP LSPs; targeted LDP (T-LDP) control messages are still sent unlabeled using the interior gateway protocol (IGP) shortest path. Since LDP does not have traffic engineering (TE), it can now benefit from the RSVP-TE features. LDP FRR is loopfree alternate (LFA), but with LDPoRSVP, it can use RSVP FRR detour or bypass tunnels.

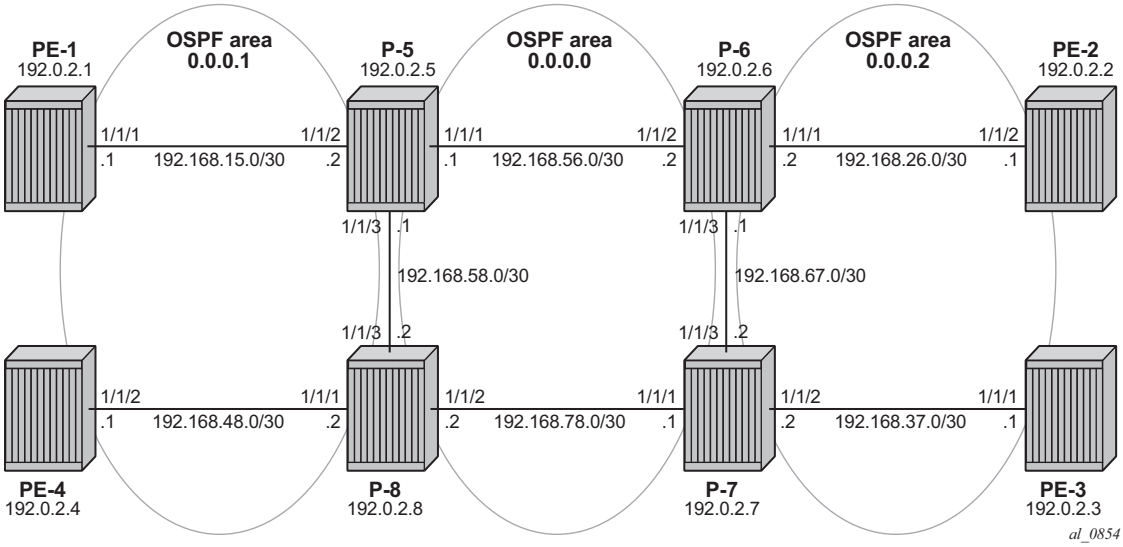
The main advantage of LDPoRSVP is seen in large networks. A full mesh of intra-area RSVP LSPs between PE nodes (which in some cases is not scalable) is not needed anymore. While a label edge router (LER) may not have that many tunnels, any transit node will potentially have thousands of LSPs, and if each transit node also has to deal with detour tunnels or bypass tunnels, this number can make the label switching router (LSR) overly burdened.

LDPoRSVP can be configured in an intra-area domain and an inter-area domain. Any router in an area can be a stitching point for LDP over RSVP. LDPoRSVP introduces a new tunnel type, tunnel-in-tunnel (in addition to the existing LDP tunnel type and RSVP tunnel type). If multiple tunnel types match the destination PE forwarding equivalence class (FEC) lookup, LDP will prefer an LDP tunnel over an LDPoRSVP tunnel by default.

First, it is important to understand how LDP FEC resolution is working (with LDPoRSVP enabled). A more detailed explanation can be found later on in this chapter. The ingress LER receives an LDP label message including a FEC with prefix **P** and label **L** from a peer by a T-LDP session. LDP tries to resolve prefix **P** by performing a lookup in the routing table manager (RTM). The result of this is a next hop (NH) to the destination PE, either an intra-area PE (intra-area context) or an area border router (ABR) (inter-area context). When the NH matches the targeted LDP peer, LDP performs a second lookup for that NH in the tunnel table manager (TTM) which returns a user configured RSVP LSP with the best metric. If there are more than one configured RSVP LSPs with the best metric, LDP selects the first available RSVP LSP. If all user configured RSVP LSPs are down, no more action is taken. If the user did not configure any RSVP LSPs under the T-LDP context, the lookup in TTM will return the first available RSVP LSP which terminates on the ABR (inter-area) or intra-area PE with the lowest metric.

If the lookup in TTM results in no RSVP LSP, the system can fall back to link-level LDP (iLDP). In that way, it is possible that the NH is reachable using iLDP. Accordingly, the egress label will be installed on the ingress LER.

Figure 156 **Initial Topology**



OSPF area 0.0.0.1 and OSPF area 0.0.0.2 should be seen as two metro areas, connected to each other via a core area, represented by OSPF backbone area (area 0.0.0.0). Therefore, P-5, P-6, P-7 and P-8 are all acting as area border routers (ABRs). LDPoRSVP principles will be shown for intra-area PE communication (PE-1 <=> PE-4) and inter-area communication (PE-1 <=> PE-2).

Configuration

Step 1. Configuring the IP/MPLS network.

The system addresses and IP interface addresses are configured according to the Initial Topology. An interior gateway protocol (IGP) is needed to distribute routing information on all routers. In this case, the IGP is open shortest path first (OSPF) using the backbone area 0.0.0.0 in the core and normal areas (area 0.0.0.1 and area 0.0.0.2) in the two metro regions, connected toward the backbone area via ABRs. A configuration example is shown for PE-1 and P-5. A similar configuration can be derived for the other P and PE nodes.

```
A:PE-1# configure router ospf
            traffic-engineering
            area 0.0.0.1
                interface "system"
                exit
                interface "int-PE-1-P-5"
                    interface-type point-to-point
```

```
            exit
        exit
A:P-5# configure router ospf
        traffic-engineering
        area 0.0.0.0
            interface "system"
            exit
            interface "int-P-5-P-6"
                interface-type point-to-point
            exit
            interface "int-P-5-P-8"
                interface-type point-to-point
            exit
        exit
        area 0.0.0.1
            interface "int-P-5-PE-1"
                interface-type point-to-point
            exit
        exit
```

Because fast reroute will be enabled on the RSVP LSPs in the core area, traffic engineering is needed on the IGP. By doing this, OSPF will generate opaque link state advertisements (LSAs) which are collected in a traffic engineering database (TED), separate from the traditional OSPF topology database. OSPF interfaces are set up as type point-to-point to improve convergence, no designated router/backup designated router (DR/BDR) election process is performed. Convergence is beyond the scope of this document.

On all nodes originating and terminating a T-LDP session, an explicit **ldp-over-rsvp** parameter must be configured to enable this OSPF instance for LDPoRSVP. In the example, this becomes.

```
A:PE-[1..4]# configure router ospf ldp-over-rsvp
A:P-[5..8]# configure router ospf ldp-over-rsvp
```

To verify that OSPF neighbors are up (state:full), the **show router ospf neighbor** command is executed. To check if IP interface addresses/ subnets are known on all PEs, **show router route-table** or **show router fib** *IOM-card-slot* will display the content of the forwarding information base (FIB).

```
*A:PE-1# show router ospf neighbor

===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
 Area-Id
-------------------------------------------------------------------------------
int-PE-1-P-5                     192.0.2.5       Full       1    0       33
 0.0.0.1
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
```

```
*A:PE-1#

*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     01h10m49s  0
       system                                                         0
192.0.2.2/32                                    Remote  OSPF      00h01m01s  10
       192.168.15.2                                                   30
192.0.2.3/32                                    Remote  OSPF      00h00m48s  10
       192.168.15.2                                                   40
192.0.2.4/32                                    Remote  OSPF      00h00m33s  10
       192.168.15.2                                                   30
192.0.2.5/32                                    Remote  OSPF      00h01m19s  10
       192.168.15.2                                                   10
192.0.2.6/32                                    Remote  OSPF      00h01m13s  10
       192.168.15.2                                                   20
192.0.2.7/32                                    Remote  OSPF      00h00m48s  10
       192.168.15.2                                                   30
192.0.2.8/32                                    Remote  OSPF      00h00m33s  10
       192.168.15.2                                                   20
192.168.15.0/30                                 Local   Local     01h10m37s  0
       int-PE-1-P-5                                                   0
192.168.26.0/30                                 Remote  OSPF      00h01m07s  10
       192.168.15.2                                                   30
192.168.37.0/30                                 Remote  OSPF      00h00m48s  10
       192.168.15.2                                                   40
192.168.48.0/30                                 Remote  OSPF      00h00m33s  10
       192.168.15.2                                                   30
192.168.56.0/30                                 Remote  OSPF      00h01m19s  10
       192.168.15.2                                                   20
192.168.58.0/30                                 Remote  OSPF      00h01m19s  10
       192.168.15.2                                                   20
192.168.67.0/30                                 Remote  OSPF      00h01m07s  10
       192.168.15.2                                                   30
192.168.78.0/30                                 Remote  OSPF      00h00m39s  10
       192.168.15.2                                                   30
-------------------------------------------------------------------------------
No. of Routes: 16
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#


*A:PE-1# show router fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
    NextHop
-------------------------------------------------------------------------------
```

```
192.0.2.1/32                                                    LOCAL
    192.0.2.1 (system)
192.0.2.2/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.0.2.3/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.0.2.4/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.0.2.5/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.0.2.6/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.0.2.7/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.0.2.8/32                                                    OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.15.0/30                                                 LOCAL
    192.168.15.0 (int-PE-1-P-5)
192.168.26.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.37.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.48.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.56.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.58.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.67.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
192.168.78.0/30                                                 OSPF
    192.168.15.2 (int-PE-1-P-5)
-------------------------------------------------------------------------------
Total Entries : 16
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The next step in the process of setting up the IP/MPLS network, is enabling the IP interfaces in the MPLS and RSVP context on all involved nodes (PE and P nodes). By default, the system interface is put automatically within the MPLS/RSVP context. When an interface is put in the MPLS context, the system also copies it into the RSVP context. Explicit enabling of MPLS and RSVP context is done by the **no shutdown** command. The following output displays the MPLS/RSVP configuration for PE-1.

```
A:PE-1# configure router rsvp no shutdown

A:PE-1# configure router mpls
            interface "system"
            exit
            interface "int-PE-1-P-5"
            exit
            no shutdown
```

Step 2. Configure the RSVP LSPs. In both metro areas, RSVP LSPs are set up from all PEs toward the ABRs, no intra-area PE-PE RSVP LSPs are needed. In the core/backbone, a full RSVP LSP mesh is required. To simplify the RSVP LSP configuration, no fast reroute is enabled on the RSVP LSPs in the metro areas, only in the backbone area. All RSVP paths are configured as **strict paths.** As an example, the configuration for PE-1 and P-5 is as follows:

```
*A:PE-1# configure router mpls
            path "path-PE-1-P-5"
                hop 1 192.168.15.2 strict
                no shutdown
            exit
            path "path-PE-1-P-5-P-8"
                hop 10 192.168.15.2 strict
                hop 20 192.168.58.2 strict
                no shutdown
            exit
            lsp "LSP-PE-1-P-5"
                to 192.0.2.5
                primary "path-PE-1-P-5"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-P-8"
                to 192.0.2.8
                primary "path-PE-1-P-5-P-8"
                exit
                no shutdown
            exit

*A:P-5# configure router mpls
            path "path-P-5-P-6"
                hop 10 192.168.56.2 strict
                no shutdown
            exit
            path "path-P-5-P-8"
                hop 10 192.168.58.2 strict
                no shutdown
            exit
            path "path-P-5-P-6-P-7"
                hop 10 192.168.56.2 strict
                hop 20 192.168.67.2 strict
                no shutdown
            exit
            path "path-P-5-PE-1"
                hop 10 192.168.15.1 strict
                no shutdown
            exit
            path "path-P-5-P-8-PE-4"
                hop 10 192.168.58.2 strict
                hop 20 192.168.48.1 strict
                no shutdown
            exit
            lsp "LSP-P-5-PE-1"
                to 192.0.2.1
                primary "path-P-5-PE-1"
```

```
            exit
            no shutdown
        exit
        lsp "LSP-P-5-PE-4"
            to 192.0.2.4
            primary "path-P-5-P-8-PE-4"
            exit
            no shutdown
        exit
        lsp "LSP-P-5-P-6"
            to 192.0.2.6
            cspf
            fast-reroute facility
            exit
            primary "path-P-5-P-6"
            exit
            no shutdown
        exit
        lsp "LSP-P-5-P-7"
            to 192.0.2.7
            cspf
            fast-reroute facility
            exit
            primary "path-P-5-P-6-P-7"
            exit
            no shutdown
        exit
        lsp "LSP-P-5-P-8"
            to 192.0.2.8
            cspf
            fast-reroute facility
            exit
            primary "path-P-5-P-8"
            exit
            no shutdown
        exit
```

To display the state of RSVP LSPs, several show commands can be used. A command to show the TTM is **show router tunnel-table** with parameter **rsvp** to reference to the RSVP LSP signaling protocol. By default, an RSVP LSP has preference **7**.

```
*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                          To              Tun     Fastfail  Adm  Opr
                                                  Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-P-5                      192.0.2.5       1       No        Up   Up
LSP-PE-1-P-8                      192.0.2.8       2       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 2
===============================================================================
*A:PE-1#
*A:PE-1# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
```

```
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.5/32      rsvp      MPLS  1         7        192.168.15.2   16777215
192.0.2.8/32      rsvp      MPLS  2         7        192.168.15.2   16777215
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#

On ABR P-5:

*A:P-5# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                              To              Tun     Fastfail  Adm  Opr
                                                      Id      Config
-------------------------------------------------------------------------------
LSP-P-5-PE-1                          192.0.2.1       1       No        Up   Up
LSP-P-5-PE-4                          192.0.2.4       2       No        Up   Up
LSP-P-5-P-6                           192.0.2.6       3       Yes       Up   Up
LSP-P-5-P-7                           192.0.2.7       4       Yes       Up   Up
LSP-P-5-P-8                           192.0.2.8       5       Yes       Up   Up
-------------------------------------------------------------------------------
LSPs : 5
===============================================================================
*A:P-5#
*A:P-5# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.1/32      rsvp      MPLS  1         7        192.168.15.1   16777215
192.0.2.4/32      rsvp      MPLS  2         7        192.168.58.2   16777215
192.0.2.6/32      rsvp      MPLS  3         7        192.168.56.2   10
192.0.2.7/32      rsvp      MPLS  4         7        192.168.56.2   20
192.0.2.8/32      rsvp      MPLS  5         7        192.168.58.2   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:P-5#
```

By default, the metric for strict LSPs configured without constrained shortest path first (CSPF) (RSVP LSPs in metro areas) is infinite (value = 16777215). The LSP metric for CSPF LSPs (RSVP LSPs in the core area) follows the IGP cost. LSP metrics can be explicitly set on the LSP level, see also in the Additional Topics section.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-P-5" metric
  - metric <metric>
  - no metric

 <metric>            : <0..16777215>
```

Whenever an RSVP LSP comes up it is by default, eligible for LDPoRSVP, meaning that RSVP will signal to the relevant IGP (OSPF in this case) that the LSP should be included in the IGP shortest path first (SPF) run. The destination of the LSP (192.0.2.5) will be considered as a potential endpoint in the FEC resolution. With the **info detail** command, all default settings of a context are shown.

```
A:PE-1# configure router mpls lsp "LSP-PE-1-P-5"
A:PE-1>config>router>mpls>lsp# info detail
----------------------------------------------
                to 192.0.2.5

---snip---
                ldp-over-rsvp include

---snip---

*A:PE-1# show router mpls lsp "LSP-PE-1-P-5" detail
===============================================================================
MPLS LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-P-5
LSP Type    : RegularLsp                        LSP Tunnel ID  : 1
LSP Index   : 1                                 TTM Tunnel Id  : 1
From        : 192.0.2.1                         To             : 192.0.2.5
Adm State   : Up                                Oper State     : Up

---snip---

LdpOverRsvp : Enabled

---snip---

Primary(a)  : path-PE-1-P-5                     Up Time        : 0d 00:04:55
Bandwidth   : 0 Mbps
===============================================================================
*A:PE-1#
```

To make an RSVP LSP ineligible for LDPoRSVP, use the **exclude** command.

```
A:PE-1# configure router mpls lsp <LSP-name> ldp-over-rsvp exclude
```

Step 3. Create T-LDP sessions according to RSVP LSPs. It is a must that when configuring an RSVP LSP eligible for LDPoRSVP, also a T-LDP session is initiated. This must be done on all PE and P nodes.

```
*A:PE-1# configure router ldp
            targeted-session
                peer 192.0.2.5
                exit
                peer 192.0.2.8
                exit
```

```
            exit
*A:PE-1#

*A:PE-1# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.5:0         Targeted  Established   15        16        0d 00:01:04
192.0.2.8:0         Targeted  Established   5         7         0d 00:00:19
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================

===============================================================================
LDP IPv6 Sessions
===============================================================================
Peer LDP Id
 Adj Type            State          Msg Sent       Msg Recv       Up Time
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
```

Step 4. Enable LDPoRSVP. This is done using the **tunneling** keyword inside the T-LDP session context. Configuration is needed on all PE and ABR nodes.

```
*A:PE-1# configure router ldp
            targeted-session
                peer 192.0.2.5
                    tunneling
                    exit
                exit
                peer 192.0.2.8
                    tunneling
                    exit
                exit
            exit
```

As a result of the **tunneling** command, the LDPoRSVP process of FEC resolving is initiated. As already stated in the introduction, FEC resolution is a three-step process. First run an SPF calculation to the destination, then select an endpoint close to that destination followed by a tunnel to that endpoint. The next two steps go more into detail on this FEC resolution process. Step 5 will handle inter-area FEC resolving and Step 6 will handle intra-area FEC resolving.

Step 5. Inter-area FEC resolving (ingress LER is PE-1, egress LER is PE-2)

i. Verification endpoint nodes and associated RSVP tunnels.

The first thing to do in the inter-area FEC resolving process is for PE-1 to perform an SPF calculation toward PE-2 with the purpose to search for an eligible endpoint, as close as possible to PE-2. An endpoint is eligible when a T-LDP session exists between PE-1 and the endpoint node, tunneling is configured on the endpoint node, PE-1 received a label for the destination FEC from the endpoint and an RSVP LSP exists between PE-1 and endpoint node that can be used for LDPoRSVP.

Endpoint node in OSPF area 1 can be either P-5 or P-8 (only those nodes have a T-LDP session toward PE-1). With **show router ldp bindings active prefixes prefix 192.0.2.2/32**, it can be concluded that P-5 will be the endpoint node (EgrNextHop).

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262124
192.0.2.5                               LspId 1

192.0.2.2/32                            Swap            262137    262124
192.0.2.5                               LspId 1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#

*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                          To              Tun     Fastfail  Adm  Opr
                                                  Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-P-5                      192.0.2.5       1       No        Up   Up
LSP-PE-1-P-8                      192.0.2.8       2       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 2
===============================================================================
*A:PE-1#

*A:PE-1# show router tunnel-table
```

```
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------

---snip---

192.0.2.5/32      rsvp      MPLS  1         7        192.168.15.2   16777215

---snip---
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

Endpoint node in OSPF area 0 can be either P-6, P-7 or P-8 (only those nodes have a T-LDP session toward P-5). With **show router ldp bindings active prefixes prefix 192.0.2.2/32**, it can be concluded that P-6 will be the endpoint node.

```
*A:P-5# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262126
192.0.2.6                               LspId 3

192.0.2.2/32                            Swap            262124    262126
192.0.2.6                               LspId 3

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-5#
```

```
*A:P-5# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                          To              Tun     Fastfail  Adm  Opr
```

```
                                                    Id      Config
-------------------------------------------------------------------------------
LSP-P-5-PE-1                        192.0.2.1       1       No        Up   Up
LSP-P-5-PE-4                        192.0.2.4       2       No        Up   Up
LSP-P-5-P-6                         192.0.2.6       3       Yes       Up   Up
LSP-P-5-P-7                         192.0.2.7       4       Yes       Up   Up
LSP-P-5-P-8                         192.0.2.8       5       Yes       Up   Up
-------------------------------------------------------------------------------
LSPs : 5
===============================================================================
*A:P-5#

*A:P-5# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------

---snip---

192.0.2.6/32      rsvp      MPLS  3         7        192.168.56.2   10

---snip---

-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:P-5#
```

On node P-6, the same commands can be repeated for the final destination node PE-2. Also there, an RSVP LSP toward PE-2 will be used as transport tunnel for user packets.

```
*A:P-6# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.6)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262141
192.0.2.2                               LspId 4

192.0.2.2/32                            Swap            262126    262141
192.0.2.2                               LspId 4
```

```
-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-6#

*A:P-6# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                              To               Tun     Fastfail  Adm  Opr
                                                       Id      Config
-------------------------------------------------------------------------------
LSP-P-6-P-5                           192.0.2.5        1       Yes       Up   Up
LSP-P-6-P-7                           192.0.2.7        2       Yes       Up   Up
LSP-P-6-P-8                           192.0.2.8        3       Yes       Up   Up
LSP-P-6-PE-2                          192.0.2.2        4       No        Up   Up
LSP-P-6-PE-3                          192.0.2.3        5       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 5
===============================================================================
A:P-6#

*A:P-6# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------

---snip---

192.0.2.2/32      rsvp      MPLS  4         7        192.168.26.1    16777215

---snip---

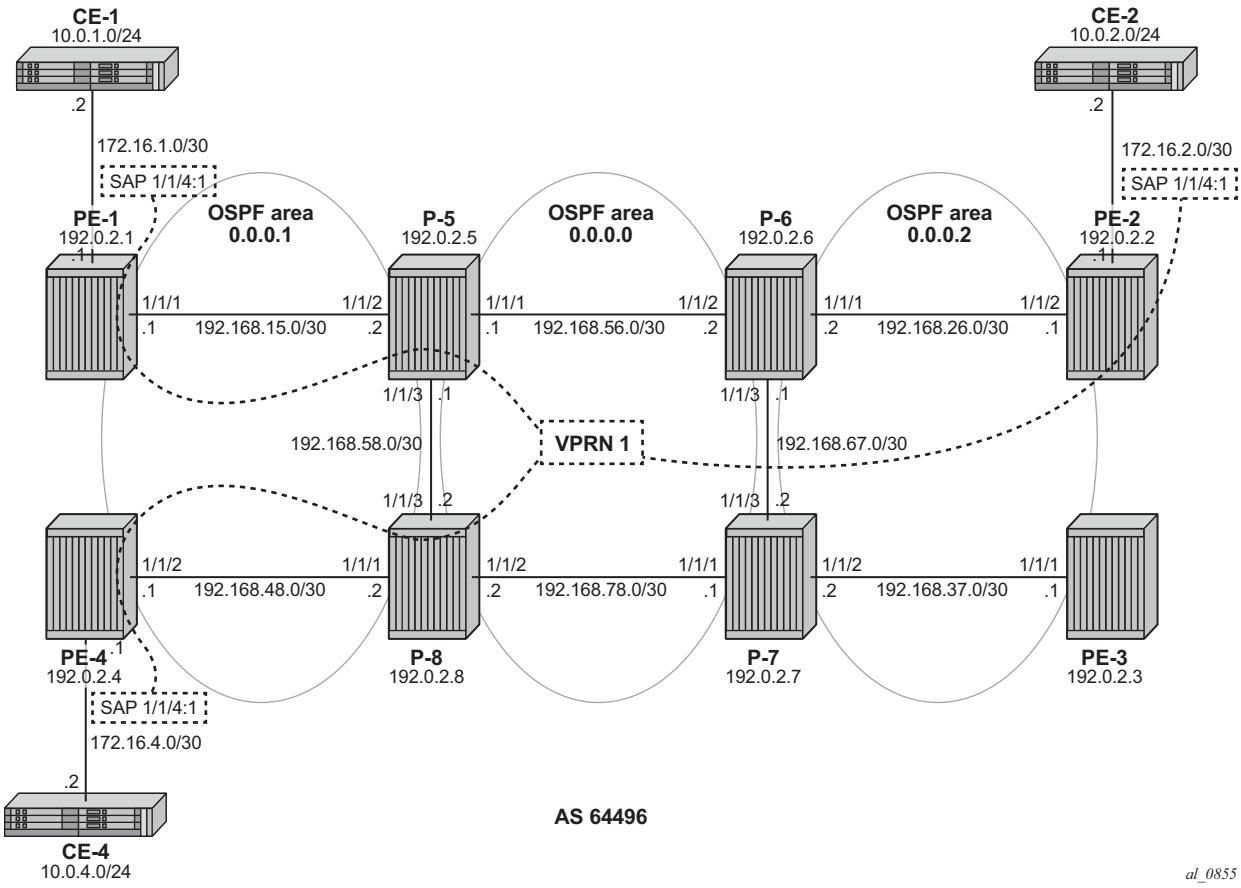
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:P-6#
```

Nodes P-5 and P-6 behave as stitching nodes to stitch RSVP LSPs. P-5 will stitch LSP-PE-1-P-5 and LSP-P-5-P-6 together while P-6 node will stitch LSP-P-5-P-6 and LSP-P-6-PE-2 together.

When the endpoints are defined, one corresponding RSVP LSP to those endpoints will be chosen (when ECMP=1). Selection criteria are as follows. When RSVP LSPs are configured under the T-LDP **tunneling** command (maximum 4), the one with the lowest LSP metric will be selected. When no RSVP LSPs are configured under the T-LDP **tunneling** command, LDP checks the TTM for all available RSVP LSPs. The RSVP LSP with the lowest metric and operational state up will be selected.

ii. Traffic verification using a virtual private routed network (VPRN) service.

Figure 157 **VPRN 1 with LDPoRSVP and No Intra-Area PE Connectivity**



VPRN service 1 is set up between three PE nodes (PE-1, PE-2, and PE-4) using the **auto-bind-tunnel resolution-filter ldp resolution filter** command. See also Figure 157 for the exact addressing scheme.

```
*A:PE-1# configure service vprn 1 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-1-CE-1" create
                address 172.16.1.1/30
                sap 1/1/4:1 create
                exit
```

```
            exit
            static-route-entry 10.0.1.0/24
                next-hop 172.16.1.2
                    no shutdown
                exit
            exit
            no shutdown
            info
            exit all
```

In order to distribute VPRN information (vpn-ipv4 routes and VPRN service labels) across the service provider network, multi-protocol border gateway protocol (MP-BGP) is needed. MP-BGP is configured on PE-1, PE-2 and PE-4 with P-5 (192.0.2.5) being the route reflector (RR). In this way, no full BGP mesh between the three PE-nodes is needed, only a BGP peering toward RR.

```
*A:PE-1# configure router
        autonomous-system 64496
        bgp
            group "internal"
                family ipv4 vpn-ipv4
                peer-as 64496
                neighbor 192.0.2.5
                exit
            exit
            no shutdown

*A:P-5# configure router
        autonomous-system 64496
        bgp
            group "internal"
                family ipv4 vpn-ipv4
                peer-as 64496
                cluster 1.1.1.1
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.4
                exit
            exit
            no shutdown
```

If user traffic is monitored between PE-1 (ingress LER) and PE-2 (egress LER) three labels should be seen. The outer label is the transport label distributed using the RSVP protocol, the inner label is the service label distributed using MP-BGP. LDPoRSVP will add an extra MPLS label between transport and service label (distributed using LDP). This middle label is used to tell the endpoint nodes (P-5 and P-6 acting as ABR) what to do. The transport label stack contains two labels: an RSVP label and an LDP label.

Translated into show commands for traffic on ingress port 1/1/2 on ABR P-5 (PE-1<=>P-5 link):

RSVP transport label 262143 is added as the top label on each user packet

```
*A:PE-1# show router rsvp session lsp-name "LSP-PE-1-P-5::path-PE-1-P-5" detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-PE-1-P-5::path-PE-1-P-5
-------------------------------------------------------------------------------
From            : 192.0.2.1             To             : 192.0.2.5
Tunnel ID       : 1                     LSP ID         : 6144
Style           : SE                    State          : Up
Session Type    : Originate
In Interface    : n/a                   Out Interface  : 1/1/1
In Label        : n/a                   Out Label      : 262143
Previous Hop    : n/a                   Next Hop       : 192.168.15.2
---snip---
```

LDP label 262124 is added as the middle label on each user packet

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262124
192.0.2.5                               LspId 1

192.0.2.2/32                            Swap            262137    262124
192.0.2.5                               LspId 1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

Service label 262133 is added as the inner MP-BGP label on each user packet



Note: This label will not change at endpoint nodes (P-5 and P-6). Ingress LER (PE-1) will push the service label to the user packet while the egress LER (PE-2) will pop the service label.

```
*A:PE-1# show router bgp neighbor 192.0.2.5 received-routes vpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     64496:1:10.0.1.0/24                                  100         None
      192.0.2.1                                            None        262133
      No As-Path
u*>i  64496:1:10.0.2.0/24                                  100         None
      192.0.2.2                                            None        262133
      No As-Path
u*>i  64496:1:10.0.4.0/24                                  100         None
      192.0.2.4                                            None        262133
      No As-Path
i     64496:1:172.16.1.0/30                                100         None
      192.0.2.1                                            None        262133
      No As-Path
u*>i  64496:1:172.16.2.0/30                                100         None
      192.0.2.2                                            None        262133
      No As-Path
u*>i  64496:1:172.16.4.0/30                                100         None
      192.0.2.4                                            None        262133
      No As-Path
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

Translated into show commands for traffic on ingress port 1/1/2 on ABR P-6 (P-5<=>P-6 link):

RSVP transport label 262140 is added as the top label on each user packet.

```
*A:P-5# show router rsvp session lsp-name "LSP-P-5-P-6::path-P-5-P-6" detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-P-5-P-6::path-P-5-P-6
-------------------------------------------------------------------------------
```

```
From            : 192.0.2.5             To              : 192.0.2.6
Tunnel ID       : 3                     LSP ID          : 56832
Style           : SE                    State           : Up
Session Type    : Originate
In Interface    : n/a                   Out Interface   : 1/1/1
In Label        : n/a                   Out Label       : 262140
Previous Hop    : n/a                   Next Hop        : 192.168.56.2
---snip---
```

LDP label 262126 is added as the middle label on each user packet.

```
*A:P-5# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262126
192.0.2.6                               LspId 3

192.0.2.2/32                            Swap            262124    262126
192.0.2.6                               LspId 3

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-5#
```

Service label 262133 is added as the inner MP-BGP label on each user packet.

Translated into show commands for traffic on ingress port 1/1/2 on node PE-2 (P-6<=>PE-2 link).

RSVP transport label 262143 is added as the outer label on each user packet.

```
*A:P-6# show router rsvp session lsp-name "LSP-P-6-PE-2::path-P-6-PE-2" detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-P-6-PE-2::path-P-6-PE-2
-------------------------------------------------------------------------------
```

```
From            : 192.0.2.6             To             : 192.0.2.2
Tunnel ID       : 4                     LSP ID         : 48128
Style           : SE                    State          : Up
Session Type    : Originate
In Interface    : n/a                   Out Interface  : 1/1/1
In Label        : n/a                   Out Label      : 262143
Previous Hop    : n/a                   Next Hop       : 192.168.26.1*
---snip---
```

iii. LDP label 262141 is added as the middle label on each user packet.

```
*A:P-6# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.6)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262141
192.0.2.2                               LspId 4

192.0.2.2/32                            Swap            262126    262141
192.0.2.2                               LspId 4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-6#
```

Service label 262133 is added as the inner MP-BGP label on each user packet.

Step 6. Intra-area FEC resolving (ingress LER is PE-1, egress LER is PE-4).

i. Verification endpoint node and associated RSVP tunnel.

The first thing to do in the intra-area FEC resolving process is for PE-1 to perform an SPF calculation toward PE4 to search for an eligible endpoint, as close as possible to PE-4. An endpoint is eligible when a T-LDP session exists between PE-1 and the endpoint node, tunneling is configured on the endpoint node, PE-1 received a label for the destination FEC from the endpoint and an RSVP LSP exists between PE-1 and endpoint node that can be used for LDPoRSVP.

First endpoint node in OSPF area 1 can be either P-5 or P-8 (only those nodes have a T-LDP session toward PE-1). With **show router ldp bindings active prefixes prefix 192.0.2.4/32** it can be concluded that P-5 will be the endpoint node.

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                            Push              --      262125
192.0.2.5                               LspId 1

192.0.2.4/32                            Swap            262139    262125
192.0.2.5                               LspId 1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#

*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                            To               Tun     Fastfail  Adm  Opr
                                                     Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-P-5                        192.0.2.5        1       No        Up   Up
LSP-PE-1-P-8                        192.0.2.8        2       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 2
===============================================================================
*A:PE-1#

*A:PE-1# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------

192.0.2.5/32      rsvp      MPLS  1         7        192.168.15.2   16777215

---snip---
```

```
===============================================================================
*A:PE-1#
```

On node P-5, the same commands can be repeated for the final destination node (PE-4). Also there, an RSVP LSP toward PE-4 will be used as transport tunnel for user packets.

```
*A:P-5# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                            Push              --      262141
192.0.2.4                               LspId 2

192.0.2.4/32                            Swap            262125    262141
192.0.2.4                               LspId 2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-5#

*A:P-5# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                            To               Tun     Fastfail  Adm  Opr
                                                     Id      Config
-------------------------------------------------------------------------------
LSP-P-5-PE-1                        192.0.2.1        1       No        Up   Up
LSP-P-5-PE-4                        192.0.2.4        2       No        Up   Up
LSP-P-5-P-6                         192.0.2.6        3       Yes       Up   Up
LSP-P-5-P-7                         192.0.2.7        4       Yes       Up   Up
LSP-P-5-P-8                         192.0.2.8        5       Yes       Up   Up
-------------------------------------------------------------------------------
LSPs : 5
===============================================================================
*A:P-5#

*A:P-5# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
```

```
-------------------------------------------------------------------------------

---snip---

192.0.2.4/32      rsvp      MPLS  2         7        192.168.58.2   16777215

---snip---

===============================================================================
*A:P-5#
```

P-5 node acts as a stitching node to stitch RSVP LSPs. P-5 will stitch LSP-PE-1-P-5 and LSP-P-5-PE-4 together.

When the endpoint node (P-5) is defined, the corresponding RSVP LSP to this endpoint will be chosen. Selection criteria are as follows (when ECMP=1). When RSVP LSPs are configured under the T-LDP **tunneling** command (maximum 4), the one with the lowest LSP metric will be selected. When no RSVP LSPs are configured under the T-LDP **tunneling** command, LDP checks TTM for all available RSVP LSPs. The RSVP LSP with the lowest metric and operational state **up** will be selected.

ii. Traffic verification using a VPRN service (see VPRN 1 with LDPoRSVP and No Intra-Area PE Connectivity).

If user traffic between PE-1 (ingress LER) and PE-4 (egress LER) is monitored, three labels are seen. The outer label is the transport label (distributed using RSVP protocol), the inner label is the service label (distributed using MP-BGP). LDPoRSVP will add an extra MPLS transport label between outer and inner label (distributed using LDP). This middle label is used to tell the endpoint node (P-5) what to do.

Translated into show commands for traffic on ingress port 1/1/2 on P-5 node (PE-1<=>P-5 link):

Transport label 262143 is added as the top RSVP label on each user packet.

```
*A:PE-1# show router rsvp session lsp-name "LSP-PE-1-P-5::path-PE-1-P-5" detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-PE-1-P-5::path-PE-1-P-5
-------------------------------------------------------------------------------
From            : 192.0.2.1             To              : 192.0.2.5
Tunnel ID       : 1                     LSP ID          : 6144
Style           : SE                    State           : Up
Session Type    : Originate
In Interface    : n/a                   Out Interface   : 1/1/1
In Label        : n/a                   Out Label       : 262143
Previous Hop    : n/a                   Next Hop        : 192.168.15.2
---snip---
```

LDPoRSVP label 262125 is added as the middle LDP label on each user packet.

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                            Push              --      262125
192.0.2.5                               LspId 1

192.0.2.4/32                            Swap            262139    262125
192.0.2.5                               LspId 1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

Service label 262133 is added as the inner MP-BGP label on each user packet.

Note: This label will not change at endpoint node (P-5). Ingress LER (PE-1) will push the service label to the user packet while the egress LER (PE-4) will pop the service label.

```
*A:PE-1# show router bgp neighbor 192.0.2.5 received-routes vpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
```

```
      As-Path
-------------------------------------------------------------------------------
i     64496:1:10.0.1.0/24                                     100         None
      192.0.2.1                                               None        262133
      No As-Path
u*>i  64496:1:10.0.2.0/24                                     100         None
      192.0.2.2                                               None        262133
      No As-Path
u*>i  64496:1:10.0.4.0/24                                     100         None
      192.0.2.4                                               None        262133
      No As-Path
i     64496:1:172.16.1.0/30                                   100         None
      192.0.2.1                                               None        262133
      No As-Path
u*>i  64496:1:172.16.2.0/30                                   100         None
      192.0.2.2                                               None        262133
      No As-Path
u*>i  64496:1:172.16.4.0/30                                   100         None
      192.0.2.4                                               None        262133
      No As-Path
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

Translated into show commands for traffic on ingress port 1/1/2 on node PE-4 (PE-4<=>P-8 link):

P-5 pushes RSVP transport label 262140 as the top label on each user packet. This RSVP transport label is swapped by P-8 to label 262143.

```
*A:P-5# show router mpls lsp "LSP-P-5-PE-4" path detail
===============================================================================
MPLS LSP LSP-P-5-PE-4 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-P-5-PE-4 Path path-P-5-P-8-PE-4
-------------------------------------------------------------------------------
LSP Name          : LSP-P-5-PE-4
Path LSP ID       : 18944
From              : 192.0.2.5             To                  : 192.0.2.4
Admin State       : Up                    Oper State          : Up
Path Name         : path-P-5-P-8-PE-4     Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up
Out Interface     : 1/1/3                 Out Label           : 262140

---snip---

Explicit Hops     :
    192.168.58.2(S)     -> 192.168.48.1(S)
Actual Hops       :
```

```
    192.168.58.1 (192.0.2.5)                    Record Label        : N/A
 -> 192.168.58.2 (192.0.2.8)                    Record Label        : 262140
 -> 192.168.48.1                                Record Label        : 262143
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 0
===============================================================================
*A:P-5#
```



Note: show router rsvp session lsp-name LSP-P-5-PE-4::path-P-5-P-8-PE-4 detail cannot be used since it only shows the outgoing RSVP label toward node P-8. On node P-8, RSVP transport label 262140 will be swapped into RSVP transport label 262143 for the link P-8 <=> PE-4.

LDP label 262141 is added as the middle label on each user packet.

```
*A:P-5# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                            Push              --      262141
192.0.2.4                               LspId 2

192.0.2.4/32                            Swap            262125    262141
192.0.2.4                               LspId 2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-5#
```

Service label 262133 is added as the inner MP-BGP label on each user packet.

Additional Topics

prefer-tunnel-in-tunnel

If the next-hop router advertised the same FEC over link-level LDP (iLDP), LDP will prefer the iLDP tunnel by default unless the user explicitly changed the default preference using the **prefer-tunnel-in-tunnel** command. In this case an LDPoRSVP tunnel will have precedence.

Until now, no RSVP LSPs are configured inside the **ldp targeted-session peer tunneling** context. Therefore, two additional strict non-CSPF RSVP LSPs are added between ingress LER PE-1 and egress LER P-5. Both LSPs will have an explicit metric setting and will be applied inside the **ldp tunneling** context. On the Layer 3 interface between PE-1 and P-5, iLDP is enabled.

```
A:PE-1# configure router ldp
            interface-parameters
                interface "int-PE-1-P-5"
                exit
            exit


A:P-5# configure router ldp
            interface-parameters
                interface "int-P-5-PE-1"
                exit
            exit


*A:PE-1# configure router mpls
            lsp "LSP-PE-1-P-5-metric100"
                to 192.0.2.5
                metric 100
                primary "path-PE-1-P-5"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-P-5-metric200"
                to 192.0.2.5
                metric 200
                primary "path-PE-1-P-5"
                exit
                no shutdown
            exit


*A:PE-1# configure router ldp
            targeted-session
                peer 192.0.2.5
                    tunneling
                        lsp "LSP-PE-1-P-5-metric100"
                        lsp "LSP-PE-1-P-5-metric200"
```

```
                exit
            exit
        exit
```

TTM on node PE-1 will look like this:

```
*A:PE-1# show router tunnel-table
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------

---snip---

192.0.2.5/32      rsvp      MPLS  1         7        192.168.15.2    16777215
192.0.2.5/32      rsvp      MPLS  3         7        192.168.15.2    100
192.0.2.5/32      rsvp      MPLS  4         7        192.168.15.2    200
192.0.2.5/32      ldp       MPLS  65537     9        192.168.15.2    10

---snip---

===============================================================================
```

Four LSPs are set up toward node P-5, three RSVP LSPs and one LDP LSP. Tunnel ID 1 is a reference to LSP-PE-1-P-5. Tunnel ID 3 is a reference to LSP-PE-1-P-5-metric100. Tunnel ID 4 is a reference to LSP-PE-1-P-5-metric200 and owner LDP is a reference to iLDP.

Taken into account the FEC resolution rules, iLDP prevails (no LDPoRSVP tunnel will be used).

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                            Push              --      262127
192.168.15.2                            1/1/1

192.0.2.5/32                            Swap            262140    262127
192.168.15.2                            1/1/1
```

```
-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

This behavior can be changed by setting the **prefer-tunnel-in-tunnel** command in the LDP context. Now, the LDPoRSVP tunnel with the best (= lowest) metric is taken.

```
*A:PE-1# configure router ldp prefer-tunnel-in-tunnel

*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                              Push              --      262127
192.0.2.5                                 LspId 3

192.0.2.5/32                              Swap            262140    262127
192.0.2.5                                 LspId 3

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#

*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                          To              Tun     Fastfail  Adm  Opr
                                                  Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-P-5                      192.0.2.5       1       No        Up   Up
LSP-PE-1-P-8                      192.0.2.8       2       No        Up   Up
LSP-PE-1-P-5-metric100            192.0.2.5       3       No        Up   Up
LSP-PE-1-P-5-metric200            192.0.2.5       4       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 4
===============================================================================
*A:PE-1#
```

If the LSP-PE-1-P-5-metric 100 is shut down, then the LSP-PE-1-P-5-metric 200 will become active.

```
*A:PE-1# configure router mpls
            lsp "LSP-PE-1-P-5-metric100"
            shutdown
            exit

*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                            Push              --      262127
192.0.2.5                               LspId 4

192.0.2.5/32                            Swap            262140    262127
192.0.2.5                               LspId 4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#*

*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                        To              Tun     Fastfail  Adm  Opr
                                                Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-P-5                    192.0.2.5       1       No        Up   Up
LSP-PE-1-P-8                    192.0.2.8       2       No        Up   Up
LSP-PE-1-P-5-metric100          192.0.2.5       3       No        Dwn  Dwn
LSP-PE-1-P-5-metric200          192.0.2.5       4       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 4
===============================================================================
*A:PE-1#
```

If LSP-PE-1-P-5-metric 200 is shut down, iLDP resumes.

```
*A:PE-1# configure router mpls
            lsp "LSP-PE-1-P-5-metric200"
            shutdown
```

```
            exit

*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                              Push              --      262127
192.168.15.2                              1/1/1

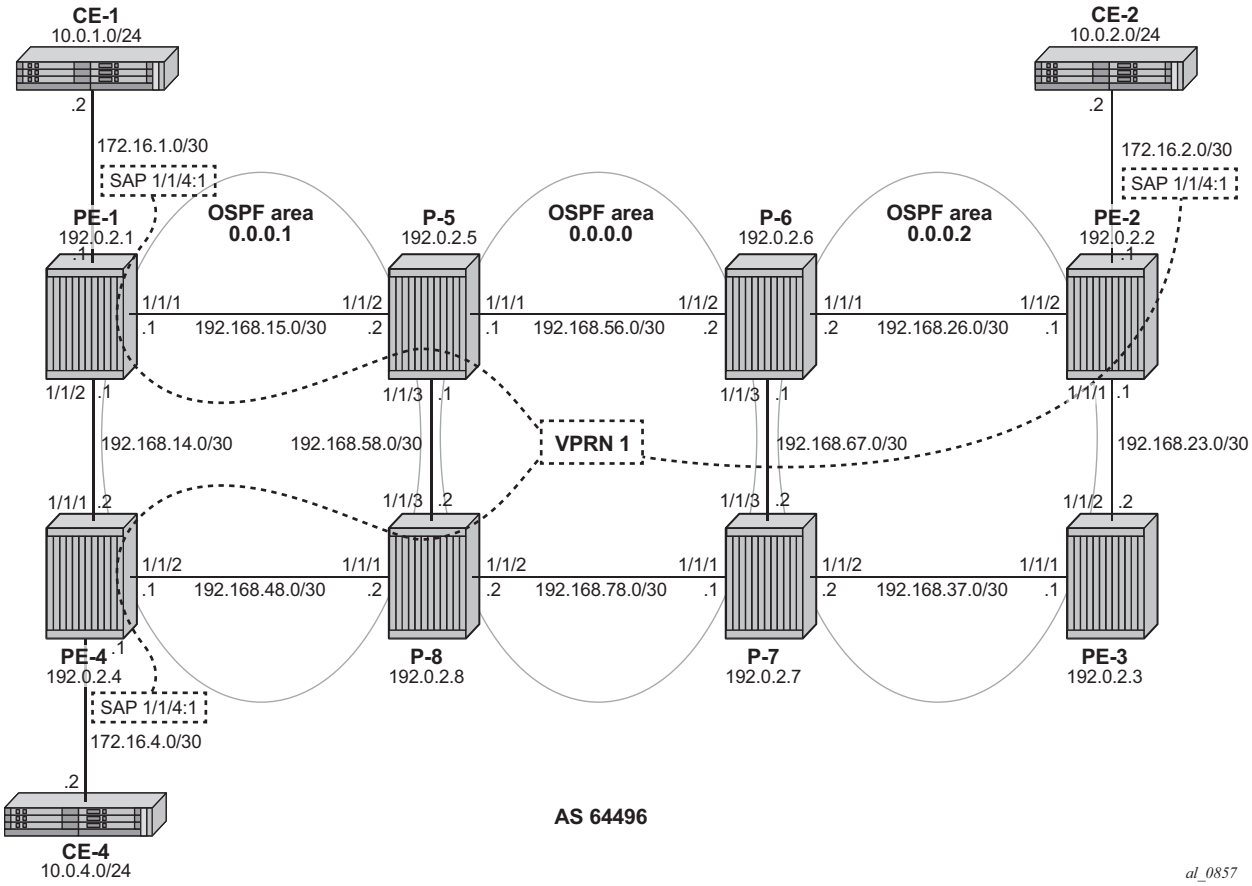
192.0.2.5/32                              Swap            262140    262127
192.168.15.2                              1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

Intra-PE Connectivity Will Change LDPoRSVP Behavior

Refer to Figure 158. In the two metro areas, both of the intra PEs are physically connected with each other. Compared with the previous figures, PE-1 is directly connected to PE-4 and PE-2 is directly connected to PE-3 (up to the OSPF level).

Figure 158 **VPRN 1 with LDPoRSVP and Intra-Area PE Connectivity**



The SPF path calculation on PE-1 toward destination (PE-4) will not point to node P-5 anymore (as was seen before) but will now point directly to PE-4 (shortest, lowest IGP metric). As a conclusion, it can be said that when possible intra-area endpoint nodes are not part of the calculated SPF path, LDPoRSVP will be not be preferred anymore. For this situation, it is advisable to configure iLDP on the intra-PE interfaces to have a fallback mechanism.

This is configured on PE-1 and PE-4 as follows:

```
*A:PE-1# configure router
        interface "int-PE-1-PE-4"
            address 192.168.14.1/30
            port 1/1/2
        exit

*A:PE-4# configure router
        interface "int-PE-4-PE-1"
            address 192.168.14.2/30
            port 1/1/1
```

```
        exit

*A:PE-1# configure router ospf
            area 0.0.0.1
                interface "int-PE-1-PE-4"
                    interface-type point-to-point
                exit
            exit

*A:PE-4# configure router ospf
            area 0.0.0.1
                interface "int-PE-4-PE-1"
                    interface-type point-to-point
                exit
            exit
```

From the moment iLDP is configured, an LDP LSP is set up. Intra-area PE traffic will flow over this LDP LSP.

```
*A:PE-1# configure router ldp
            interface-parameters
                interface "int-PE-1-PE-4"
                exit
            exit

*A:PE-4# configure router ldp
            interface-parameters
                interface "int-PE-4-PE-1"
                exit
            exit

A:PE-1# show router tunnel-table 192.0.2.4/32
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.4/32      ldp       MPLS  65544     9        192.168.14.2   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
A:PE-1#
```

If user traffic is monitored, between PE-1 (ingress LER) and PE-4 (egress LER) only two labels are seen. The outer label is the transport label (distributed using LDP), the inner label is the service label (distributed using MP-BGP). No LDPoRSVP label is present anymore. Translated into show commands for traffic on ingress port 1/1/1 on PE-4 node (PE-1<=>PE-4 link):

LDP transport label 262141 is added as the outer label on each user packet.

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
```

```
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                            Push              --      262141
192.168.14.2                            1/1/2

192.0.2.4/32                            Swap            262139    262141
192.168.14.2                            1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

Service label 262133 is added as the inner MP-BGP label on each user packet.

```
*A:PE-1# show router bgp neighbor 192.0.2.5 received-routes vpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     64496:1:10.0.1.0/24                               100         None
      192.0.2.1                                         None        262133
      No As-Path
u*>i  64496:1:10.0.2.0/24                               100         None
      192.0.2.2                                         None        262133
      No As-Path
u*>i  64496:1:10.0.4.0/24                               100         None
      192.0.2.4                                         None        262133
      No As-Path
i     64496:1:172.16.1.0/30                             100         None
      192.0.2.1                                         None        262133
      No As-Path
u*>i  64496:1:172.16.2.0/30                             100         None
      192.0.2.2                                         None        262133
```

```
      No As-Path
u*>i  64496:1:172.16.4.0/30                                  100         None
      192.0.2.4                                              None        262133
      No As-Path
-------------------------------------------------------------------------------
Routes : 6
===============================================================================
*A:PE-1#
```

Conclusion

LDPoRSVP allows tunneling of user packets toward an LDP far-end destination inside an RSVP LSP (with the benefits of RSVP LSPs, fast-reroute (FRR) and traffic engineering (TE)). The main application of this feature is for deployment of MPLS based services, for example, VPRN, virtual leased line (VLL), and virtual private LAN service (VPLS) services, in large networks where a full mesh of LSPs reaches the limits of scalability.

LDP Point-to-Point LSPs

This chapter provides information about label distribution protocol (LDP) point-to-point label switched paths (LSPs)

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was originally written for release 7.0.R5. The output in this edition corresponds to release 14.0.R1. There are no pre-requisites or conditions on the hardware for this configuration.

Overview

Due to the connectionless nature of the network layer protocol IP, packets travel through the network on a hop-by-hop basis with routing decisions made at each node. As a result, hyperaggregation of data on certain links may occur and it may impact the provider's ability to provide guaranteed service levels across the network end-to-end. To address these shortcomings, multiprotocol label switching (MPLS) was developed. The technology provides the capability to establish connection oriented paths, called label switched paths (LSPs), over a connectionless (IP) network. The LSP offers a mechanism to engineer network traffic independently from the underlying network routing protocol (mostly IP) to improve the network resiliency and recovery options and to permit delivery of new services that are not readily supported by conventional IP routing techniques (Layer 2 IP virtual private networks (VPNs)). These benefits are essential for today's communication network explaining the wide deployment base of the MPLS technology.

RFC 3031, *Multiprotocol Label Switching Architecture*, specifies the MPLS architecture while this document describes the configuration and troubleshooting of point-to-point LSPs on SR OS.

Packet Forwarding

As a packet of a connectionless network layer protocol travels from one router to the next, each router in the network makes an independent forwarding decision by performing the following basic tasks: first analyzing the packet's header, then referencing the local routing table to find the longest match based on the destination address in the IP header, and finally sending out the packet on the selected interface. In other terms, the first function partitions the entire set of possible packets into a set of forwarding equivalence classes (FECs). All packets associated to a particular FEC will be forwarded along the same logical path to the same destination. The second function maps each FEC to a next hop destination router. Each router along the packet's path performs these actions.

On the other hand, in MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network. In turn, the FEC is mapped to an LSP, which is established prior to packet forwarding. An MPLS label, representing the FEC to which the packet is assigned, is attached to the packet (push operation) and once labeled, the packet is forwarded to the next hop router along that LSP path. At subsequent hops, no analysis of the packet's network layer header is needed. Instead, the label is used as an index into a table which specifies the next hop and a new label. The old label is replaced with the new label (swap operation), and the packet is forwarded to its next hop. At the MPLS network egress, the label is removed from the packet (pop operation). If this router is the final destination (based on the remaining packet), the packet is handed to the receiving application (such as a virtual private LAN service (VPLS) domain). If this router is not the final destination of the packet, the packet will be sent into a new MPLS tunnel or forwarded by conventional IP forwarding towards the Layer 3 destination

Terminology

Figure 159 **Generic MPLS Network, MPLS Label Operations**

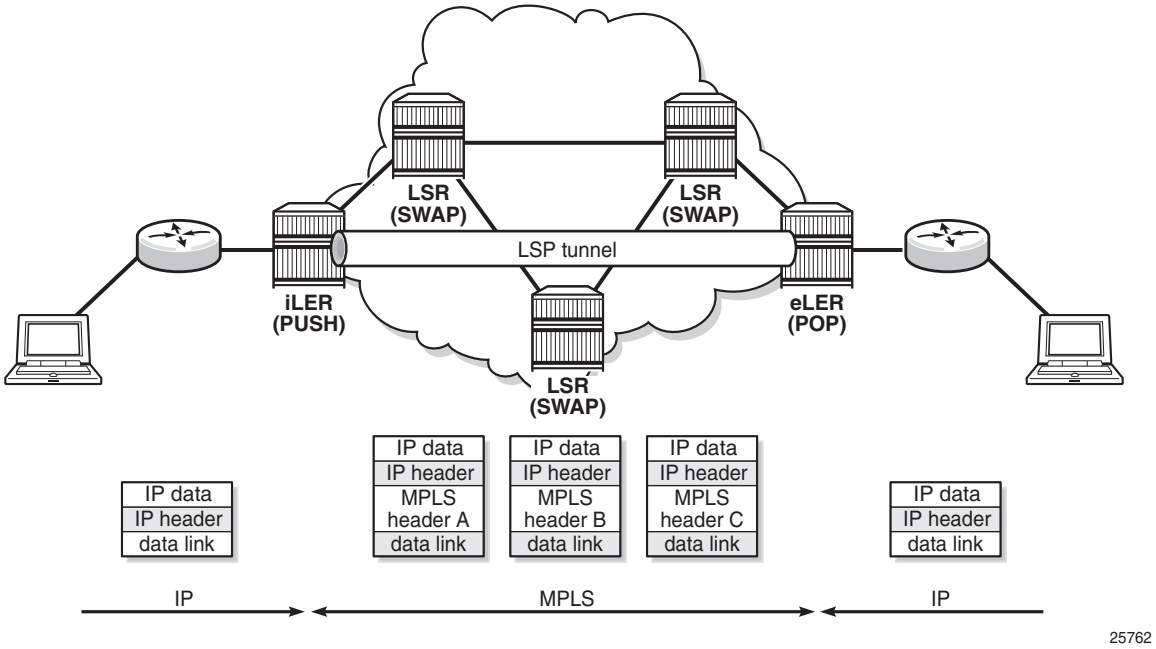


Figure 159 depicts a general network topology clarifying the MPLS-related terms. A label edge router (LER) is a device at the edge of an MPLS network, with at least one interface outside the MPLS domain. A router is usually defined as an LER based on its position relative to a particular LSP. The MPLS router at the head-end of an LSP is called the ingress label edge router (iLER). The MPLS router at the tail-end of an LSP is called the egress label edge router (eLER). The iLER receives unlabeled packets from outside the MPLS domain, then applies MPLS labels to the packets, and forwards the labeled packets into the MPLS domain. The eLER receives labeled packets from the MPLS domain, then removes the labels, and forwards unlabeled packets outside the MPLS domain. The eLER can signal an implicit-null label (numeric value 3). This informs the previous hop to send MPLS packets without an outer label and so is known as penultimate hop popping (PHP).

A label switching router (LSR) is a device internal to an MPLS network, with all interfaces inside the MPLS domain. These devices switch labeled packets inside the MPLS domain. In the core of the network, LSR ignore the packet's network layer (IP) header and simply forward the packet using the MPLS label swapping mechanism.

LSP Establishment

Prior to packet forwarding, the LSP must be established. In order to do so, labels need to be distributed for the path. Labels are usually distributed by a downstream router in the upstream direction (relative to the data flow). There are a number of ways used for label distribution: static, LDP, and RSVP. For static P2P LSPs, see chapter Static Point-to-Point LSPs; for RSVP-TE P2P LSPs, see chapter RSVP Point-to-Point LSPs.

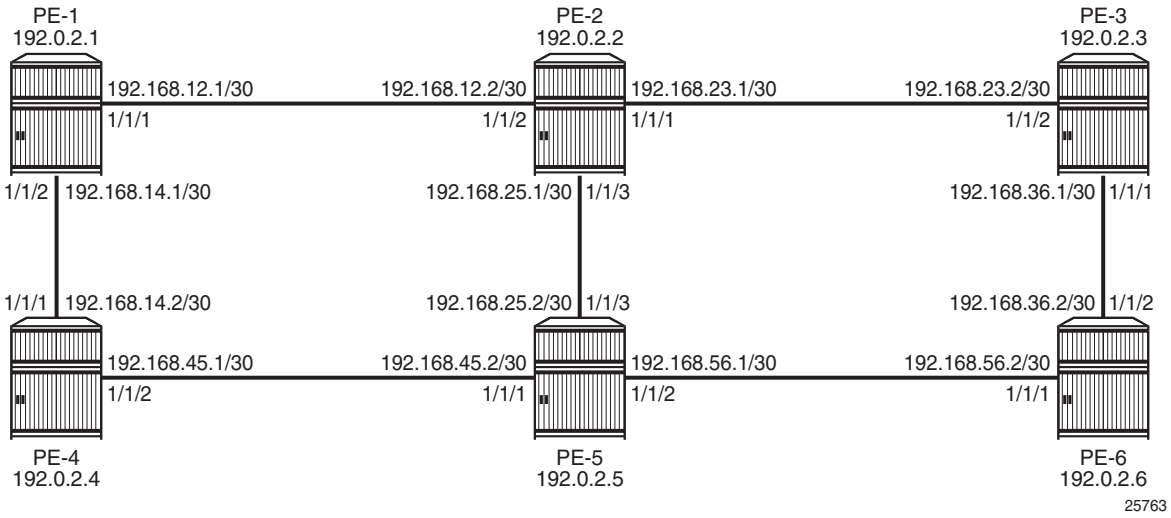
LDP (RFC 5036, *LDP Specification*) can be considered as an extension to the network interior gateway protocol (IGP). As routers become aware of new destination networks, they advertise labels in the upstream direction that will allow upstream routers to reach the destination.

Fast reroute (FRR) allows for establishing backup paths before a failure takes place. This way traffic can flow almost continuously, without waiting for routing protocol convergence; see chapter MPLS LDP FRR using ISIS as IGP.

Example Topology

The example topology is displayed in Figure 160. The setup consists of six 7750 SR nodes located in a single autonomous system.

Figure 160 **MPLS Example Topology**



Configuration

As a general pre-requisite for the configuration of MPLS LSPs, a correctly working IGP is required. Open shortest path first (OSPF) or intermediate system to intermediate system (IS-IS) can be used as IGP.

LDP is a simple label distribution protocol with basic MPLS functionality (no traffic engineering). Fast Reroute is supported; see chapter MPLS LDP FRR using ISIS as IGP. LDP relies on the underlying routing information provided by an IGP in order to forward labeled packets. Each LDP configured LSR will originate a label for its system address and a label for each FEC for which it has a next-hop that is external to the MPLS domain, without the explicit need to manually configure the LSPs. When deviations from this default behavior are desired, import and export policies can be applied.

The configuration is as simple as enabling the LDP protocol instance and adding all network interfaces, for each node. As an example the configuration on node PE-1 is displayed below; similar configurations apply on the other nodes.

```
*A:PE-1# configure router
        ldp
            interface-parameters
                interface "int-PE-1-PE-2"
                exit
                interface "int-PE-1-PE-4"
                exit
```

The **show router ldp discovery** and **show router ldp session** commands can be used to verify the LDP hello adjacencies and sessions. The adjacency type (AdjType) needs to be **Link** while the state should be **Established**.

```
*A:PE-1# show router ldp discovery
===============================================================================
LDP IPv4 Hello Adjacencies
===============================================================================
Interface Name                   Local Addr                          State
AdjType                          Peer Addr
-------------------------------------------------------------------------------
int-PE-1-PE-2                    192.0.2.1:0                         Estab
link                             192.0.2.2:0

int-PE-1-PE-4                    192.0.2.1:0                         Estab
link                             192.0.2.4:0

-------------------------------------------------------------------------------
No. of IPv4 Hello Adjacencies: 2
===============================================================================
===============================================================================
LDP IPv6 Hello Adjacencies
===============================================================================
Interface Name                   Local Addr                          State
```

```
AdjType                          Peer Addr
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
*A:PE-1# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   35        37        0d 00:01:19
192.0.2.4:0         Link      Established   29        31        0d 00:00:58
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
===============================================================================
LDP IPv6 Sessions
===============================================================================
Peer LDP Id
 Adj Type            State          Msg Sent       Msg Recv       Up Time
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
```

The **show router ldp bindings** command displays the contents of the LIB (Label Information Base) and contains all labels locally generated (IngLbl) and those received from any LDP neighbors (EgrLbl), whether they are in use or not. The following output is for IPv4 prefixes:

```
*A:PE-1# show router ldp bindings ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, WP - Label Withdraw Pending, C - Cpipe Service
        BU - Alternate For Fast Re-Route, TLV - (Type, Length: Value)
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                  IngLbl                    EgrLbl
Peer                                    EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.1/32                            262143U                     --
192.0.2.2:0                               --
  --

192.0.2.1/32                            262143U                     --
192.0.2.4:0                               --
  --
```

```
192.0.2.2/32                                --                      262143
192.0.2.2:0                               1/1/1
192.168.12.2

192.0.2.2/32                              262142U                   262141
192.0.2.4:0                                 --
  --

192.0.2.3/32                              262141N                   262141
192.0.2.2:0                               1/1/1
192.168.12.2

192.0.2.3/32                              262141U                   262140
192.0.2.4:0                                 --
  --

192.0.2.4/32                              262140U                   262140
192.0.2.2:0                                 --
  --

192.0.2.4/32                                --                      262143
192.0.2.4:0                               1/1/2
192.168.14.2

192.0.2.5/32                              262139N                   262139
192.0.2.2:0                               1/1/1
192.168.12.2

192.0.2.5/32                              262139U                   262139
192.0.2.4:0                                 --
  --

192.0.2.6/32                              262138N                   262138
192.0.2.2:0                               1/1/1
192.168.12.2

192.0.2.6/32                              262138U                   262138
192.0.2.4:0                                 --
  --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 12
===============================================================================
---snip---
```

The **show router ldp bindings active** command displays the content of the Label Forwarding Information Base (LFIB) and contains all active labels and the associated label actions used for label switching packets. The active LDP bindings for IPv4 prefixes are the following:

```
*A:PE-1# show router ldp bindings active ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
```

```
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Pop             262143      --
  --                                      --

192.0.2.2/32                            Push              --      262143
192.168.12.2                            1/1/1

192.0.2.2/32                            Swap            262142    262143
192.168.12.2                            1/1/1

192.0.2.3/32                            Push              --      262141
192.168.12.2                            1/1/1

192.0.2.3/32                            Swap            262141    262141
192.168.12.2                            1/1/1

192.0.2.4/32                            Push              --      262143
192.168.14.2                            1/1/2

192.0.2.4/32                            Swap            262140    262143
192.168.14.2                            1/1/2

192.0.2.5/32                            Push              --      262139
192.168.12.2                            1/1/1

192.0.2.5/32                            Swap            262139    262139
192.168.12.2                            1/1/1

192.0.2.6/32                            Push              --      262138
192.168.12.2                            1/1/1

192.0.2.6/32                            Swap            262138    262138
192.168.12.2                            1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 11
===============================================================================
---snip---
```

In the tunnel table, there are LDP LSPs to all other nodes:

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
```

```
192.0.2.2/32      ldp       MPLS  65542     9        192.168.12.2   10
192.0.2.3/32      ldp       MPLS  65543     9        192.168.12.2   20
192.0.2.4/32      ldp       MPLS  65544     9        192.168.14.2   10
192.0.2.5/32      ldp       MPLS  65545     9        192.168.12.2   20
192.0.2.6/32      ldp       MPLS  65546     9        192.168.12.2   30
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

In order to signal PHP with LDP, implicit-null must be configured on the eLER.

```
A:PE-6# configure router ldp implicit-null-label
```

The implicit-null is signaled immediately, all related labels are withdrawn and re-advertised with label value of 3. The new label would show up on PE-5 as a swap from the ingress label to an egress label of 3, although label 3 is not pushed on to the frame.

Import and Export Policies

The default label handling behavior is to originate label bindings for the system address and to propagate all FECs received. If this is not the desired behavior, an import/export policy can be applied. An LDP import policy impacts inbound filtering; an LDP export policy impacts outbound filtering. An export policy may be configured to control the set of LDP label bindings advertised by the LER (sending to LDP peers). As such, export policies are used to include additional FECs rather than filtering FECs from those advertised. An import policy can be used to control for which FECs a router will generate labels (accepting from LDP peers). This functionality is not unique to LDP; it can be used for RSVP-TE, OSPF, and IS-IS as well as others.

The policy can be global or LDP peer FEC prefix filtering, both for import and export. LDP peer FEC prefix filtering uses a similar policy context as the LDP global policies and works in addition to these global policies.

```
*A:PE-1# tree flat detail | match import-pref
configure router ldp session-parameters peer import-prefixes <policy-name>
                     [<policy-name>...(upto 5 max)]
configure router ldp session-parameters peer no import-prefixes
configure router ldp targeted-session import-prefixes <policy-name>
                     [<policy-name>...(upto 5 max)]
configure router ldp targeted-session no import-prefixes
*A:PE-1#
*A:PE-1# tree flat detail | match export-pref
configure router ldp session-parameters peer export-prefixes <policy-name>
                     [<policy-name>...(upto 5 max)]
configure router ldp session-parameters peer no export-prefixes
```

```
configure router ldp targeted-session export-prefixes <policy-name>
                     [<policy-name>...(upto 5 max)]
configure router ldp targeted-session no export-prefixes
*A:PE-1#
```

By default no labels are generated for directly connected (local) interfaces. To change this behavior, an export policy is created and applied to the LDP instance. There is no configuration difference in defining an import and export policy.

A policy starts with the keyword begin and contains a list of entries (of which each has a number), and ends with the keyword commit. An entry typically contains matching criteria (however, it is not required in cases where everything matches) and a corresponding action. Entries without an action are considered incomplete and are rendered inactive. When processing the policy, the router executes the specified action on the first matching statement; it does not process any further matches. For this reason, entries must be sequenced correctly from most to least specific.

The configuration of the LDP export policy for local interfaces is as follows:

```
*A:PE-1# configure router
        policy-options
            begin
            policy-statement "LDP-export"
                entry 10
                    from
                        protocol direct
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

There are 11 active LDP bindings before applying the export policy, as shown earlier.

The LDP export or import policy is applied to the LDP instance on the router, with the **export** or **import** keyword.

```
*A:PE-1# configure router ldp export "LDP-export"
```

When the export policy is applied, the active LDP binding table contains additional entries: the local interfaces of PE-x.

```
*A:PE-1# show router ldp bindings active ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
```

```
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Pop             262143      --
  --                                      --

192.0.2.2/32                            Push              --      262143
192.168.12.2                            1/1/1

192.0.2.2/32                            Swap            262142    262143
192.168.12.2                            1/1/1

192.0.2.3/32                            Push              --      262141
192.168.12.2                            1/1/1

192.0.2.3/32                            Swap            262141    262141
192.168.12.2                            1/1/1

192.0.2.4/32                            Push              --      262143
192.168.14.2                            1/1/2

192.0.2.4/32                            Swap            262140    262143
192.168.14.2                            1/1/2

192.0.2.5/32                            Push              --      262139
192.168.12.2                            1/1/1

192.0.2.5/32                            Swap            262139    262139
192.168.12.2                            1/1/1

192.0.2.6/32                            Push              --      262138
192.168.12.2                            1/1/1

192.0.2.6/32                            Swap            262138    262138
192.168.12.2                            1/1/1

192.168.12.0/30                         Pop             262136      --
  --                                      --

192.168.14.0/30                         Pop             262133      --
  --                                      --

192.168.23.0/30                         Swap            262135    262137
192.168.12.2                            1/1/1

192.168.25.0/30                         Swap            262134    262134
192.168.12.2                            1/1/1

192.168.36.0/30                         Swap            262137    262136
192.168.12.2                            1/1/1

192.168.45.0/30                         Swap            262131    262132
```

```
192.168.14.2                                1/1/2

192.168.56.0/30                             Swap            262132    262132
192.168.12.2                                1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 18
===============================================================================
---snip---
```

OAM

The following operations, administration, and maintenance operations can be launched on an LDP LSP:

- oam lsp-ping
- oam lsp-trace

The following options are supported for LSP ping:

```
*A:PE-1# oam lsp-ping
  - lsp-ping <lsp-name> [path <path-name>]
  - lsp-ping bgp-label prefix <ip-prefix/prefix-length> [path-destination <ip-
address> [interface
    <if-name> | next-hop <ip-address>]]
  - lsp-ping prefix <ip-prefix/prefix-length> [path-destination <ip-
address> [interface <if-name> |
    next-hop <ip-address>]]
  - lsp-ping static <lsp-name> [assoc-channel <ipv4|non-ip|none>] [dest-global-
id <global-id>
    dest-node-id <node-id>] [force] [path-type <active|working|protect>]
  - options common to all lsp-ping cases:  [detail] [fc <fc-
name> [profile <in|out>]] [interval
    <interval>] [send-count <send-count>] [size <octets>] [src-ip-address <ip-
address>] [timeout
    <timeout>] [ttl <label-ttl>]
 <octets>             : [1..9198] - Default: 1
 <label-ttl>          : [1..255] - Default: 255
 <timeout>            : [1..10] seconds - Default: 5
 <interval>           : [1..10] seconds - Default: 1
 <in|out>             : keywords - Default: out
 <send-count>         : [1..100] packets - Default: 1
 <path-name>          : [32 chars max]
 <detail>             : keyword - display detailed information
 <if-name>            : [32 chars max]
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <static>             : keyword - use the Target FEC Stack sub-type 'Static LSP'
 <global-id>          : [0..4294967295] - Default: 0
 <node-id>            : a.b.c.d|<1..4294967295> - Default: 0
```

```
 <ipv4|non-ip|none>   : keywords - Default: non-ip
 <active|working|pr*> : keywords - Default: active
 <bgp-label>          : keyword - use the Target FEC Stack sub-
type 'BGP labeled IPv4 prefix'
 <force>              : keyword -
 start the test regardless of the local operational state of the LSP
                        path
 <lsp-name>           : [64 chars max]
 <ip-prefix/prefix-*> : <ipv4-prefix>/32 | <ipv6-prefix>/128
                        ipv4-prefix    - a.b.c.d
                        ipv6-prefix    - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <fc-name>            : be|l2|af|l1|h2|ef|h1|nc - Default: be
*A:PE-1#
```

As an example, an LSP ping is sent from PE-1 to PE-6:

```
*A:PE-1# oam lsp-ping prefix 192.0.2.6/32
LSP-PING 192.0.2.6/32: 80 bytes MPLS payload
Seq=1, send from intf int-PE-1-PE-2, reply from 192.0.2.6
       udp-data-len=32 ttl=255 rtt=2.56ms rc=3 (EgressRtr)
---- LSP 192.0.2.6/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 2.56ms, avg = 2.56ms, max = 2.56ms, stddev = 0.000ms
*A:PE-1#
```

An LSP trace is sent from PE-1 to PE-6:

```
*A:PE-1# oam lsp-trace prefix 192.0.2.6/32
lsp-trace to 192.0.2.6/32: 0 hops min, 0 hops max, 104 byte packets
1  192.0.2.2  rtt=0.636ms rc=8(DSRtrMatchLabel) rsc=1
2  192.0.2.3  rtt=2.04ms rc=8(DSRtrMatchLabel) rsc=1
3  192.0.2.6  rtt=2.91ms rc=3(EgressRtr) rsc=1
*A:PE-1#
```

The return code (rc) is 8 for the LSRs and 3 for the eLER.

The detailed output for this LSP trace includes the interface IP address, the interface type, maximum receive unit (MRU), label and protocol; as follows:

```
*A:PE-1# oam lsp-trace prefix 192.0.2.6/32 detail
lsp-trace to 192.0.2.6/32: 0 hops min, 0 hops max, 104 byte packets
1  192.0.2.2  rtt=0.590ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.23.2 ifaddr=192.168.23.2 iftype=ipv4Numbered MRU=1564
           label[1]=262138 protocol=3(LDP)
2  192.0.2.3  rtt=1.01ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.36.2 ifaddr=192.168.36.2 iftype=ipv4Numbered MRU=1564
           label[1]=262143 protocol=3(LDP)
3  192.0.2.6  rtt=2.06ms rc=3(EgressRtr) rsc=1
*A:PE-1#
```

LDP Statistics

LDP-related statistics can be collected in files. First a file needs to be configured.

```
*A:PE-1# configure log
        file-id 1
            location cf1:
            rollover 5 retention 1
        exit
```

The next step is to configure an accounting policy that will define which statistics should be recorded.

```
*A:PE-1# configure log accounting-policy 1 record
  - no record
  - record <record-name>
 <record-name>        : aa-admit-deny|aa-app-group|aa-application|aa-partition|aa-
performance|
                        aa-protocol|aa-subscriber-application|aa-subscriber-
protocol|
                        com-bined-ldp-lsp-egress|combined-mpls-lsp-egress|combined-
mpls-lsp-ingress|
                        combined-network-ing-egr-octets|combined-queue-group|
                        combined-sdp-ingress-egress|combined-service-ing-egr-octets|
                        combined-service-ingress|compact-service-ingress-
octets|complete-pm|
                        complete-sdp-ingress-egress|complete-service-ingress-egress|
                        complete-subscriber-ingress-egress|custom-record-aa-sub|
                        custom-record-service|custom-record-subscriber|network-
egress-octets|
                        network-egress-packets|network-ingress-octets|network-
ingress-packets|
                        queue-group-octets|queue-group-packets|saa|service-egress-
octets|
                        ser-vice-egress-packets|service-ingress-octets|service-
ingress-packets|video|
                        complete-ethernet-port|extended-service-ingress-egress|
                        complete-network-ing-egr
*A:PE-1#
```

As an example, the following accounting-policy is configured:

```
*A:PE-1# configure log
        accounting-policy 1
            record combined-ldp-lsp-egress
            to file 1
            no shutdown
        exit
```

The collection of statistics for prefix 192.0.2.6/32 is enabled on PE-1 in the LDP context, as follows:

```
*A:PE-1# configure router
        ldp
```

```
            egress-statistics
                fec-prefix 192.0.2.6/32
                    collect-stats
                    accounting-policy 1
                    no shutdown
                exit
            exit
```

The following FEC egress statistics can be displayed:

```
*A:PE-1# show router ldp fec-egress-stats
  - fec-egress-stats [<ip-prefix/mask>]
  - fec-egress-stats [active] [family]
 <ip-prefix/mask>     : <ipv4-prefix>/32 | <ipv6-prefix>/128
                        ipv4-prefix    - a.b.c.d
                        ipv6-prefix    - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <active>             : keyword
 <family>             : ipv4|ipv6
*A:PE-1#
```

The FEC egress stats for prefix 192.0.2.6/32 can be retrieved as follows:

```
*A:PE-1# show router ldp fec-egress-stats 192.0.2.6/32
===============================================================================
LDP IPv4 FEC Egress Statistics
===============================================================================
-------------------------------------------------------------------------------
FEC Prefix/Mask    : 192.0.2.6/32
-------------------------------------------------------------------------------
Collect Stats      : Enabled              Accounting Plcy.     : 1
Admin State        : Up
FC BE
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC L2
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC AF
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC L1
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC H2
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC EF
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC H1
InProf Pkts        : 0                    OutProf Pkts         : 0
InProf Octets      : 0                    OutProf Octets       : 0
FC NC
InProf Pkts        : 0                    OutProf Pkts         : 0
```

```
InProf Octets      : 0                    OutProf Octets       : 0
===============================================================================
LDP IPv4 FEC Egress Statistics: 1
===============================================================================
*A:PE-1#
```

Statistics can be cleared as follows:

```
*A:PE-1# clear router ldp fec-egress-statistics 192.0.2.6/32
```

Debug

LDP debugging can be configured per LDP interface or per LDP peer, as follows:

```
*A:PE-1# debug router ldp
  - ldp
  - no ldp
 [no] interface       + Enable/disable and configure debugging for an LDP interface
 [no] peer            + Enable/disable and configure debugging for an LDP peer
*A:PE-1#
```

A particular peer is specified by its IPv4 or IPv6 address. It is possible to configure debugging for specific LDP events: bindings or messages, as follows:

```
*A:PE-1# debug router ldp peer 192.0.2.2
  - no peer <ip-address>
  - peer <ip-address>
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 [no] event           + Configure debugging for specific LDP events
 [no] packet          + Enable/disable debugging for specific LDP packets
*A:PE-1# debug router ldp peer 192.0.2.2 event
  - event
  - no event
 [no] bindings        - Enable/disable debugging for LDP bindings
 [no] messages        - Enable/disable debugging for LDP messages
*A:PE-1#
```

It is also possible to configure debugging for specific packets, such as label packets:

```
*A:PE-1# debug router ldp peer 192.0.2.2 packet
  - no packet
  - packet
 [no] hello           - Enable/disable debugging for LDP Hello packets
 [no] init            - Enable/disable debugging for LDP Init packets
 [no] keepalive       - Enable/disable debugging for LDP Keepalive packets
 [no] label           - Enable/disable debugging for LDP Label packets
```

The following debugging is configured on PE-1:

```
*A:PE-1# debug router ldp peer 192.0.2.2 packet label detail
```

Some label mapping packets sent to peer 192.0.2.2 are the following, with the label mapping for prefixes 192.0.2.1/32 and 192.0.2.4/32:

```
2 2016/02/18 14:18:59.71 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 7) to 192.0.2.2:0
Protocol version = 1
Label 262132 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.4/32
"
1 2016/02/18 14:18:59.71 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 6) to 192.0.2.2:0
Protocol version = 1
Label 262143 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.1/32
"
```

Conclusion

MPLS provides the capability to establish connection-oriented paths over a connectionless network.. LDP point-to-point LSPs are dynamically signaled and FRR is supported. This can greatly improve network resiliency. In this chapter, the configuration of several LDP point-to-point LSP features is given together with the associated show output which can be used to verify and troubleshoot.

LDP-SR Stitching for IPv4 Prefixes (IS-IS)

This chapter provides information about LDP-SR Stitching for IPv4 Prefixes (IS-IS).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R5.

Overview

Segment Routing (SR) allows for the construction of source-routed Label Switched Paths (LSPs) where the series of hops to be taken through the network are indicated by one or more Segment Identifiers (SIDs) assigned at the ingress PE. In the case of an MPLS data plane, these SIDs are MPLS labels learned through extensions to the OSPF/IS-IS control plane. SR provides benefits to the MPLS data plane, such as high scalability (due to lack of soft-state), traffic engineering capability, and topology-independent fast reroute.

When SR is configured in an IP/MPLS network that runs the Label Distribution Protocol (LDP), it is possible that SR and LDP will coexist, in which case preference for LDP or SR is a local matter at the LSP head end. It is equally possible that not all devices will have the capability to support SR, in which case some kind of interworking between SR and LDP is necessary to create an end-to-end LSP. Fast reroute coverage can also benefit from this SR-LDP interworking function, where SR is used to increase Loop Free Alternate (LFA) coverage using Remote or Directed LFA.

This chapter describes the configuration requirements for the interworking of LDP and SR to form a single end-to-end LSP when using IS-IS as an IGP. The chapter shows how this interworking function can be used to extend fast reroute coverage for LDP-based LSPs.

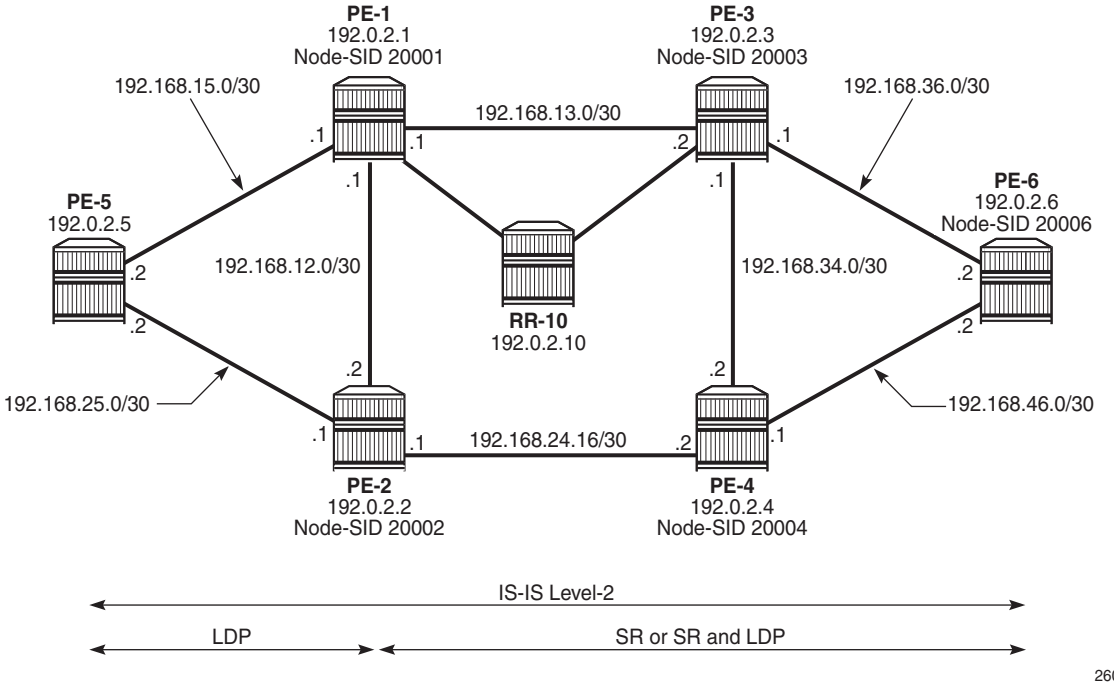
Configuration

Example Topology

The topology shown in Figure 161 provides an example of SR-LDP interworking. All routers within the topology form part of Autonomous System 64496 and are IBGP clients of RR-10 for the VPN-IPv4 address family. All routers in the topology belong to the same IS-IS Level-2 area, and all link metrics are set to 100. RR-10 does not participate in any MPLS data plane, and signals the IS-IS overload bit to avoid being used for transit traffic.

PE-5 is a router that does not support SR and, therefore, runs only LDP to its connected peers PE-1 and PE-2. PE-1, PE-2, PE-3, PE-4, and PE-6 are capable of running both SR and LDP, but are initially configured to only run SR with the associated node-SIDs shown in Figure 161. When explicitly described, LDP will be enabled in conjunction with SR on these routers to show the difference between the two approaches, and to show how SR can be used as a fast reroute backup for SR primary LSPs.

Figure 161 **Example Topology**



The LDP configuration at PE-1 toward PE-5 is shown in the following output. The configuration at PE-2 is similar with the only exception being IP addressing.

```
configure
    router
        ldp
            interface-parameters
                interface "int-PE-1-PE-5" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
            targeted-session
            exit
            no shutdown
        exit
```

PE-1, PE-2, PE-3, PE-4, and PE-6 run SR. The following output provides an example of the relevant SR configuration parameters at PE-1, with similar configurations on the remaining SR routers. For an explanation of these parameters, see chapter Segment Routing with IS-IS Control Plane.

```
configure
    router Base
        mpls-labels
            sr-labels start 20000 end 20099
        exit
        isis 0
            advertise-router-capability as
            interface "system"
                level-capability level-2
                ipv4-node-sid label 20001
                passive
                no shutdown
            exit
            segment-routing
                prefix-sid-range start-label 20000 max-index 99
                no shutdown
            exit
            no shutdown
        exit
```

SR Mapping Server

An SR Mapping Server (SR-MS) is an integral part of SR-LDP interoperability and has the responsibility for advertising prefixes to SID/label mappings on behalf of routers that do not support SR. When using IS-IS, a SID/label-binding TLV (TLV 149) containing a prefix-SID sub-TLV is used to advertise one or more SID index/labels and one or more prefixes. In the example topology, PE-4 is selected as the SR-MS and will advertise a prefix SID for the non-SR-capable router, PE-5.

The following output provides an example of the configuration required to implement SR-MS functionality. Under the **segment-routing** node, a **mapping-server** context is created that allows for origination of a SID/label-binding TLV and prefix-SID sub-TLV. The syntax begins with **sid-map node-sid** and is followed by an index. In SR OS, prefix-SIDs are always advertised with an index value (as opposed to an absolute label value), and the formula {start-label + SID index} is used to derive the label value. In this example, **index 5** is used and, therefore, the derived label value is {20000+5} 20005 for the PE-5 prefix 192.0.2.5/32.

An optional **range** argument allows for advertisement of a contiguous range of prefixes and associated SIDs using the configured index/prefix as the beginning of the range. Non-contiguous ranges require multiple entries and are advertised as separate SID/label-binding TLVs. An additional optional **set-flags s** argument can also be used to set the S-flag, which controls the flooding scope. When set, the flooding scope is the entire IS-IS domain. When not set, the flooding scope is the IS-IS level into which the TLV was advertised.

```
configure
    router
        isis 0
            segment-routing
                prefix-sid-range start-label 20000 max-index 99
         mapping-server
                    sid-map node-sid index 5 prefix 192.0.2.5/32
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The relevant part of the IS-IS LSP generated by PE-4, showing the SID/label-binding TLV, is shown in the following output:

```
*A:PE-4# show router isis database PE-4.00-00 detail
===============================================================================
Rtr Base ISIS Instance 0 Database (detail)
===============================================================================
---snip---

TLVs :
```

```
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols     : IPv4
  IS-Hostname   : PE-4
  Router ID   :
    Router ID   : 192.0.2.4
  Router Cap : 192.0.2.4, D:0, S:0
    TE Node Cap : B E M
    SR Cap: IPv4 MPLS-IPv6
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
  SID Label Binding:
    Prefix: 192.0.2.5/32 Range:1 Weight:0 bFlgs:v4 SID:5 Algo:0 pFlgs:N
---snip---
```

At other routers within the SR domain, the presence of the advertised prefix can be validated as shown in the following output taken at PE-1. The SRMS field is set to Y for prefix 192.0.2.5/32, indicating that the prefix was advertised by an SR-MS. (In the case of IS-IS, the prefix-SID is a sub-TLV of the SID/label-binding TLV and the "N" (node-SID) flag is set; therefore, it can be recognized as being advertised by a mapping server.) The Y is followed by an "(S)" flag, indicating that the SRMS prefix-SID is selected to be programmed. This indication is provided in case there are multiple advertisements for the same prefix and/or node-SID from different SR mapping servers that result in some kind of conflict or inconsistency. If there are multiple mapping servers advertising the same prefix-SID, the advertising router with the lowest system/router ID is preferred.

```
*A:PE-1# show router isis prefix-sids
===============================================================================
Rtr Base ISIS Instance 0 Prefix/SID Table
===============================================================================
Prefix                          SID        Lvl/Typ    SRMS   AdvRtr
                                                       MT     Flags
-------------------------------------------------------------------------------
192.0.2.1/32                    1          2/Int.     N      PE-1
                                                         0        NnP
192.0.2.2/32                    2          2/Int.     N      PE-2
                                                         0        NnP
192.0.2.3/32                    3          2/Int.     N      PE-3
                                                         0        NnP
192.0.2.4/32                    4          2/Int.     N      PE-4
                                                         0        NnP
192.0.2.5/32                    5          2/Int.     Y(S)   PE-4
                                                         0        NnP
192.0.2.6/32                    6          2/Int.     N      PE-6
                                                         0        NnP
-------------------------------------------------------------------------------
No. of Prefix/SIDs: 8 (8 unique)
-------------------------------------------------------------------------------
SRMS : Y/N  = prefix SID advertised by SR Mapping Server (Y) or not (N)
       S    = SRMS prefix SID is selected to be programmed
Flags: R    = Re-advertisement
       N    = Node-SID
       nP   = no penultimate hop POP
```

```
       E    = Explicit-Null
       V    = Prefix-SID carries a value
       L    = value/index has local significance
===============================================================================
```

SR-LDP Interworking

Interworking SR and LDP essentially consists of stitching an LDP FEC and an SR node-SID route for the same prefix. In the example topology, PE-1 and PE-2 will act as the SR-LDP interworking nodes.

In the LDP-to-SR data plane direction, LDP uses an **export-tunnel-table** command under the **ldp** context to reference a policy that defines which prefixes should be redistributed from the IS-IS/SR domain into LDP. When applied, the LDP process monitors the tunnel-table until it locates a /32 SR tunnel of type sr-isis that matches a prefix defined in the export policy. LDP then programs an LDP Incoming Label Map (ILM) entry and stitches it to the SR node-SID tunnel endpoint. The LDP process also originates a FEC for the prefix and advertises that FEC to its peers.

The following output provides an example of the route policy and application of the policy at PE-1. In this policy, PE-1 advertises LDP FECs to PE-5 for PE-3 (192.0.2.3), PE-4 (192.0.2.4), and PE-6 (192.0.2.6), provided that PE-1 has a /32 SR tunnel of type sr-isis in the tunnel-table for those same prefixes. PE-1 also programs an LDP ILM entry for each prefix and stitches it to the appropriate SR tunnel. An identical configuration exists at PE-2.

```
configure
    router
        policy-options
            begin
            prefix-list "sr-domain"
                prefix 192.0.2.3/32 exact
                prefix 192.0.2.4/32 exact
                prefix 192.0.2.6/32 exact
            exit
            policy-statement "SR-to-LDP"
                entry 10
                    from
                        protocol isis
                        prefix-list "sr-domain"
                    exit
                    to
                        protocol ldp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

```
        ldp
            export-tunnel-table "SR-to-LDP"
            interface-parameters
                interface "int-PE-1-PE-5" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

In the SR-to-LDP data plane direction, the **export-tunnel-table ldp** command within the **segment-routing** context is the only required configuration. Unlike the LDP-to-SR data plane direction, where policy is used to control which prefixes are stitched, in the SR-to-LDP direction, no policy is explicitly referenced because the SR-MS provides a network-wide policy for the prefixes that SR needs to stitch to a corresponding LDP FEC. With the **export-tunnel-table ldp** command applied, whenever a /32 LDP tunnel destination matches a prefix for which a prefix-SID sub-TLV was received from a mapping server, the SR ILM is stitched to the corresponding LDP tunnel endpoint.

The following output shows the configuration applied at PE-1 to implement SR-to-LDP data plane interworking:

```
configure
    router
        isis 0
            segment-routing
                export-tunnel-table ldp
                no shutdown
            exit
            no shutdown
        exit
```

With the required configuration in the SR-LDP interworking routers (PE-1 and PE-2), it is possible to validate the correct ILM entries. In the LDP-to-SR data plane direction, the following output shows the active LDP bindings at PE-1. Each of the entries for PE-3 (192.0.2.3), PE-4 (192.0.2.4), and PE-6 (192.0.2.6) have an "(I)" flag to indicate that the prefix has an SR-ISIS next-hop. Each entry also has an ingress label and an egress label. The ingress label represents the LDP FEC advertised for the corresponding prefix (in this case, advertised only to PE-5). The egress label represents the SR node-SID for the same prefix. Therefore, a mapping exists between LDP FEC and SR node-SID.

```
*A:PE-1# show router ldp bindings active prefixes ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
```

```
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Pop             262143      --
  --                                      --

192.0.2.3/32(I)                         Swap            262135    20003
192.168.13.2                            1/1/1:100

192.0.2.4/32(I)                         Swap            262136    20004
192.168.12.2                            1/1/3:100

192.0.2.5/32                            Push              --      262143
192.168.15.2                            1/1/2:100

192.0.2.6/32(I)                         Swap            262134    20006
192.168.13.2                            1/1/1:100

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 5
===============================================================================
```

In the SR-to-LDP data plane direction, the following output, taken at PE-1, shows a dump of the SR database for next-hops resolved to LDP. There is a single entry with index 5 (label value 20005) advertised by the SR-MS for the PE-5 prefix 192.0.2.5. The final line of the entry shows that an LDP FEC is the SID next-hop for SR-LDP stitching. The tunnel LSP ID is 65537. The tunnel-table verifies that this is an LDP tunnel to PE-5 (192.0.2.5).

```
*A:PE-1# tools dump router isis sr-database nh-type ldp detail
===============================================================================
Rtr Base ISIS Instance 0 SR Database
===============================================================================
-------------------------------------------------------------------------------
SID 5
-------------------------------------------------------------------------------
Label              : 20005                 Adv System Id       : 1920.0000.2005
Prefix             : 192.0.2.5
Route Level        : 2                     MT Id               : 0
Rtm Preference     : 18                    Ttm Preference      : 0
Metric             : 0                     Last Action         : AddTnl
Num Ip NextHop     : 1                     Num SR-Tnl NextHop  : 1
Mtu                : 0
Mtu Prim           : 0                     Mtu Backup          : 0
Exclude from LFA   : 0                     Remote LFA          : 0
Duplicate Pending  : 0                     Tunnel Active State : Reported/Ack
SR Error           : SR_ERR_OK
LDP Next-Hop IP    : 192.0.2.5
```

```
Tunnel LSP Id      : 65537                 Tunnel Type         : 2
-------------------------------------------------------------------------------
No. of Entries: 1
-------------------------------------------------------------------------------
LDP = LDP FEC is the SID NH for SR-LDP stitching
===============================================================================
*A:PE-1#
```

To verify that the data plane is intact from end-to-end, a VPRN service is configured at the non-SR-capable PE-5 and the SR-capable PE-6, each with a locally configured subnet that is used to test IP connectivity. The configuration of the VPRN at PE-5 is shown in the following output. The **auto-bind-tunnel** configuration uses a resolution filter allowing only **ldp** to be used to resolve BGP next-hops for VPN-IPv4 routes. Usually, this could be configured for **resolution any**, but this configuration shows that LDP is being used. The local IP address at PE-5 is 172.31.5.1/24.

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "Local-Subnet" create
                address 172.31.5.1/24
                sap 1/1/1:1 create
                exit
            exit
            no shutdown
        exit
```

The configuration of the VPRN at PE-6 is shown in the following output. Again, the **auto-bind-tunnel** configuration uses a resolution filter, but this time it is configured for **sr-isis**. It could be set to **resolution any**, so that the tunnel-table preference would resolve an LSP with the lowest preference/metric, but the resolution filter configuration again shows that SR is being used. The **auto-bind-tunnel** context allows the transport mechanism to be a local decision at service level. The local IP address at PE-6 is 172.31.6.1/24.



Note: An alternative approach would be to configure the auto-bind-tunnel context for **resolution any**, then modify the tunnel-table preference for SR using the **tunnel-table-pref** command in the **segment-routing** context.

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
```

```
            auto-bind-tunnel
                resolution-filter
                    sr-isis
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "Local-Subnet" create
                address 172.31.6.1/24
                sap 1/1/2:1 create
                exit
            exit
            no shutdown
        exit
```

A VPRN ping between 172.31.5.1 at PE-5 and 172.31.6.1 at PE-6 verifies that the data plane is intact:

```
*A:PE-5# ping router 1 172.31.6.1 source 172.31.5.1
PING 172.31.6.1 56 data bytes
64 bytes from 172.31.6.1: icmp_seq=1 ttl=64 time=7.05ms.
64 bytes from 172.31.6.1: icmp_seq=2 ttl=64 time=6.93ms.
64 bytes from 172.31.6.1: icmp_seq=3 ttl=64 time=6.88ms.
64 bytes from 172.31.6.1: icmp_seq=4 ttl=64 time=6.54ms.
64 bytes from 172.31.6.1: icmp_seq=5 ttl=64 time=9.98ms.
---- 172.31.6.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 6.54ms, avg = 7.47ms, max = 9.98ms, stddev = 1.26ms
```

SR and LDP Coexistence

The previous example demonstrates the use of SR-LDP interworking when the SR domain runs only SR. A more common scenario is that SR will coexist with LDP, because LDP is already deployed and the SR deployment will be added. In this sub-section, PE-1, PE-2, PE-3, PE-4, and PE-6 are configured to run LDP in conjunction with SR. PE-5 remains the same as the previous sub-section in that it runs only LDP to its connected peers PE-1 and PE-2.

In the SR-to-LDP data plane direction, there is no notable change when LDP coexists in the SR domain. Whenever a /32 LDP tunnel destination matches a prefix for which a prefix-SID sub-TLV was received from a mapping server, the SR ILM is stitched to the corresponding LDP tunnel endpoint.

In the LDP-to-SR data plane direction, there is a significant change. If only SR is running within the SR domain, the LDP process monitors the tunnel-table and when a /32 SR tunnel of type sr-isis is found that matches a prefix in the (**export-tunnel-table**) export policy, LDP programs an LDP ILM and stitches it to the SR node-SID tunnel endpoint. However, if an LDP FEC exists for the same /32 prefix, SR OS will resolve the LDP ILM entry to the LDP FEC. This is because LDP attempts to resolve the prefix in the route table first before looking in the tunnel-table and, therefore, prefers the LDP tunnel to the SR tunnel.

The following output is taken at PE-1 when LDP and SR coexist in the SR domain. The previous version of this output (when LDP was not running in the SR domain) showed the prefixes for PE-3, PE-4, and PE-6 as known via an SR-ISIS next-hop, and the egress labels as node-SIDs. When LDP is active in conjunction with SR, the egress labels resolve to an LDP FEC.

```
*A:PE-1# show router ldp bindings active prefixes ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Pop             262143      --
  --                                      --

192.0.2.2/32                            Push              --      262142
192.168.12.2                            1/1/3:100

192.0.2.2/32                            Swap            262139    262142
192.168.12.2                            1/1/3:100

192.0.2.3/32                            Push              --      262139
192.168.13.2                            1/1/1:100

192.0.2.3/32                            Swap            262135    262139
192.168.13.2                            1/1/1:100

192.0.2.4/32                            Push              --      262136
192.168.12.2                            1/1/3:100

192.0.2.4/32                            Swap            262136    262136
192.168.12.2                            1/1/3:100

192.0.2.5/32                            Push              --      262143
```

```
192.168.15.2                                1/1/2:100

192.0.2.5/32                                Swap            262138    262143
192.168.15.2                                1/1/2:100

192.0.2.6/32                                Push              --      262134
192.168.13.2                                1/1/1:100

192.0.2.6/32                                Swap            262134    262134
192.168.13.2                                1/1/1:100

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 11
===============================================================================
```

That the LDP-to-SR data path resolves to LDP FECs rather than SR tunnels may result in an asymmetric data path. Taking the previously used VPRN service between PE-5 and PE-6 as an example:

- Traffic from PE-6 to PE-5 will use SR between PE-6 and one of the SR-LDP interworking gateways at PE-1 or PE-2, after which it will use LDP.
- Traffic from PE-5 to PE-6 will use LDP between ingress and egress. The interworking function between SR and LDP has no effect.

Both directions still use an MPLS data plane. However, the MPLS control plane differs in each direction.

LDP Fast Reroute using SR Tunnels

With the ability to interwork LDP and SR, primary LSPs signaled using LDP can select a remote LFA SR tunnel as backup. This provides the potential to increase fast reroute coverage. As with any other backup or fast reroute mechanism, the SR backup tunnel can be installed in the forwarding database before any failure, but can only be activated when the failure of the primary path has been detected.

The ability to detect a failure quickly forms a significant part of the overall recovergence time and may require the use of failure detection mechanisms, such as Bidirectional Forwarding Detection (BFD), the 802.3ah Ethernet in the First Mile (EFM), or just loss of signal (LoS). These mechanisms are beyond the scope of this chapter.

To use SR as a backup for LDP, the **fast-reroute backup-sr-tunnel** command must be configured in the **ldp** context. The **export-tunnel-table** command previously described should also be present, and should reference a policy including all of the prefixes for which backup is required. There is no requirement for an SR-MS when using SR tunnels for LDP backup, nor is there a requirement to enable SR-to-LDP interworking using the **export-tunnel-table ldp** command within the **segment-routing** context.

The following output shows the configuration applied at PE-6. When this configuration is applied, if the LFA SPF does not find an adjacent IP next-hop prefix for an LDP FEC, but can compute a remote LFA tunnel next-hop, LDP programs the LDP FEC using an LDP Next-Hop Label Forwarding Entry (NHLFE), and a backup next-hop using an LDP NHLFE pointing to the SR tunnel endpoint. The LDP packet is not tunneled over the SR tunnel, but rather the LDP label is stitched to the segment-routing label stack. This behavior is similar to the LDP-SR interworking function previously described within this chapter, but is modified such that the stitching of an LDP ILM entry to an SR tunnel only takes place if no adjacent LFA next-hop could be found for the prefix.

```
configure
    router Base
        isis
            loopfree-alternate remote-lfa
        exit
        ldp
            export-tunnel-table "SR-to-LDP"
            fast-reroute backup-sr-tunnel
        exit
        policy-options
            begin
            prefix-list "sr-domain"
                prefix 192.0.2.0/24 longer
            exit
            policy-statement "SR-to-LDP"
                entry 10
                    from
                        protocol isis
                        prefix-list "sr-domain"
                    exit
                    to
                        protocol ldp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

With the preceding configuration in place at PE-6, it is possible to verify whether a backup exists for a specific prefix, using the command shown in the following output. In this example, the backup is displayed for the PE-6 adjacent neighbor PE-3 (192.0.2.3). There are two LSPs for the prefix 192.0.2.3/32; one is known via LDP and one is known via SR-ISIS, indicated in the protocol column. The entries are defined as follows:

- The first line of the LDP entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3).
- The second line of the LDP entry is the backup indicated by a "(B)" flag, with a next-hop of 192.168.46.1 using interface 1/1/1:100 (via PE-4). This backup is a basic LFA, which is possible to compute due to the example topology, or more explicitly the triangular mesh between PE-6, PE-4, and PE-3. Due to this topology, if the link between PE6 and PE3 fails, PE-6 can forward packets destined for PE-3 toward PE-4. PE-4 will then forward them directly toward PE-3, not return them to PE-6 (which would create a transient micro-loop until the next SPF is run).
- The first line of the SR-ISIS entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3).
- The second line of the SR-ISIS entry is the backup LSP indicated by the "(B)" flag, with a next-hop of 192.168.46.1 using interface 1/1/1:100 (via PE-4). Both the primary and backup LSPs use the label 20003, representing the PE-3 node-SID. As with the LDP backup entry, the SR-ISIS backup is a basic LFA.

```
*A:PE-6# show router fp-tunnel-table 1 192.0.2.3/32
===============================================================================
Tunnel Table Display
Legend:
B - FRR Backup
===============================================================================
Destination                                             Protocol   Tunnel-ID
    Lbl                 NextHop           Intf/Tunnel
-------------------------------------------------------------------------------
192.0.2.3/32                                            LDP         -
  262139                192.168.36.1      1/1/2:100
  262138                192.168.46.1(B)   1/1/1:100
192.0.2.3/32                                            SR-ISIS-0   -
  20003                 192.168.36.1      1/1/2:100
  20003                 192.168.46.1(B)   1/1/1:100
-------------------------------------------------------------------------------
Total Entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

To show the benefits that SR provides in increasing fast reroute coverage, the link between PE-4 and PE-3 is removed from the example topology, creating a ring topology. With this link removed, it is no longer possible for PE-6 to compute a basic LFA to PE-3 for the link between PE-6 and PE-3. If that link failed and PE-6 forwarded packets destined for PE-3 toward PE-4, PE-4 would return them to PE-6 until the next SPF was complete. Therefore, a backup tunnel is needed to a place in the network that will not loop packets back; essentially a remote LFA.

The following output at PE-6 shows the primary and backup LSPs for PE-3 (192.0.2.3) with the modified topology. Again there are two LSPs: one known through via LDP and one known via SR-ISIS. The entries are defined as follows:

- The first line of the LDP entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3).
- The second line of the LDP entry is the backup indicated by a "(B)" flag with a next-hop of 192.0.2.3 (PE-3), which uses an SR tunnel. The label of "3" (implicit-null) indicates that the LDP label is not tunneled through the SR tunnel, but rather popped before the primary LDP LSP is stitched to the backup SR LSP.
- The first line of the SR-ISIS entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3). This LSP assigns a single label of value 20003, representing the node-SID of PE-3.
- The second line of the SR-ISIS entry is the backup indicated by a "(B)" flag with a next-hop of 192.168.46.1 using interface 1/1/1:100 (via PE-4). There are two labels assigned to this backup tunnel. The upper label has a value of 20002, which represents the node-SID of PE-2. This is the remote LFA "PQ-node". The second label has a value of 20003, which represents the node-SID of the destination, PE-3.

  When this backup tunnel is operational, PE-6 encapsulates traffic destined for PE-3 to a point in the network where it will not be looped back toward the source. In the example topology, that node is PE-2. When traffic arrives at PE-2, it pops the top label (20002) and forwards traffic for PE-3 (with label 20003) on the shortest path toward the destination.

```
*A:PE-6# show router fp-tunnel-table 1 192.0.2.3/32
===============================================================================
Tunnel Table Display
Legend:
B - FRR Backup
===============================================================================
Destination                                             Protocol   Tunnel-ID
    Lbl                 NextHop           Intf/Tunnel
-------------------------------------------------------------------------------
192.0.2.3/32                                            LDP         -
  262139                192.168.36.1      1/1/2:100
  3                     192.0.2.3(B)      SR
192.0.2.3/32                                            SR-ISIS-0   -
  20003                 192.168.36.1      1/1/2:100
  20003/20002           192.168.46.1(B)   1/1/1:100
-------------------------------------------------------------------------------
```

```
Total Entries : 2
------------------------------------------------------------------------------
==============================================================================
```

Conclusion

The SR control plane can (and likely will) coexist with other MPLS control plane clients, such as RSVP, LDP, or BGP. It is possible that these control plane clients will operate independently. However, where a mix of SR-capable and non-SR-capable routers exist within the same domain, SR-LDP interworking is necessary to form an end-to-end LSP. This chapter shows how that is possible using one or more SR mapping servers and one or more interworking routers.

SR-LDP interworking also provides an opportunity to increase fast reroute coverage in LDP-based networks. Before the introduction of SR-LDP interworking, a remote LFA could only be constructed using LDP-over-RSVP, which required the RSVP LSP to be manually configured and placed. When SR-LDP interworking is used, primary LDP LSPs can use a backup tunnel to a remote LFA signaled using SR. This requires no manual configuration, which provides the potential to greatly increase fast reroute coverage with minimal effort.

MPLS LDP FRR using ISIS as IGP

This chapter describes multiprotocol label switching (MPLS) label distribution protocol (LDP) fast reroute (FRR) using intermediate system to intermediate system (IS-IS) as the interior gateway protocol (IGP).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

MPLS label distribution protocol fast reroute (LDP FRR) is supported on all 7x50 platforms. This feature is supported on all input/output modules (IOMs), integrated media modules (IMMs), media dependent adapter (MDA), and compact media adapter (CMA) types that support network interfaces from 9.0.R4 and higher. This chapter was originally written for and tested on release 9.0.R6, but the CLI in this edition corresponds to release 14.0.R1. There are no prerequisites for this configuration.

Overview

LDP FRR improves convergence in case of a single link or single node failure in the network. Convergence times will be in the order of tens of milliseconds. This is important to some application services, such as voice over IP (VoIP), which are sensitive to traffic loss when running over the MPLS network.

Without FRR, link and/or node failures inside an MPLS LDP network result in traffic loss in the order of hundreds of milliseconds. The reason for that is that LDP depends on the convergence of the underlying IGP (IS-IS sending link state PDUs (LSPs) in this case). After IGP convergence, LDP itself needs to compute new primary next hop label forwarding entries (NHLFEs) for all affected forwarding equivalence classes (FECs). Finally, the different label forwarding information bases (LFIBs) are updated.

When FRR is configured on a node, the node pre-computes primary NHLFEs for all FECs and in addition it will pre-compute backup NHLFEs for all FECs. The backup NHLFE corresponds to the label received for the same FEC from a Loop-Free Alternate (LFA) next-hop (see also RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*). Both primary NHLFEs and backup NHLFEs are programmed in the IOM/IMM which makes it possible to converge very quickly.

The SROS software has implemented Inequality 1 (link criterion) and Inequality 3 (node criterion) of RFC 5286. Similar to the shortest path tree (SPT) computation that is part of standard link-state routing functionality, also the LFA next-hop computation is based on the IGP metric.

The underlying LFA formulas appear in the following format:

Inequality 1:

- SP(backup NHR, D) < {SP(backup NHR, S) + SP(S, D)}

Inequality 3:

- SP(backup NHR, D) < {SP(backup NHR, PN) + SP(PN, D)}

In these inequalities ‘SP’ is ‘shortest IGP metric path’, ‘NHR’ is ‘next hop router’, ‘D’ is ‘destination’, ‘S’ is ‘source node or upstream node doing the actual LFA next-hop computation’, and ‘PN’ is ‘protected node’. The inequality 3 rule is stricter than the inequality 1 rule. See Additional Topics for a practical example on these inequalities.

Configuration

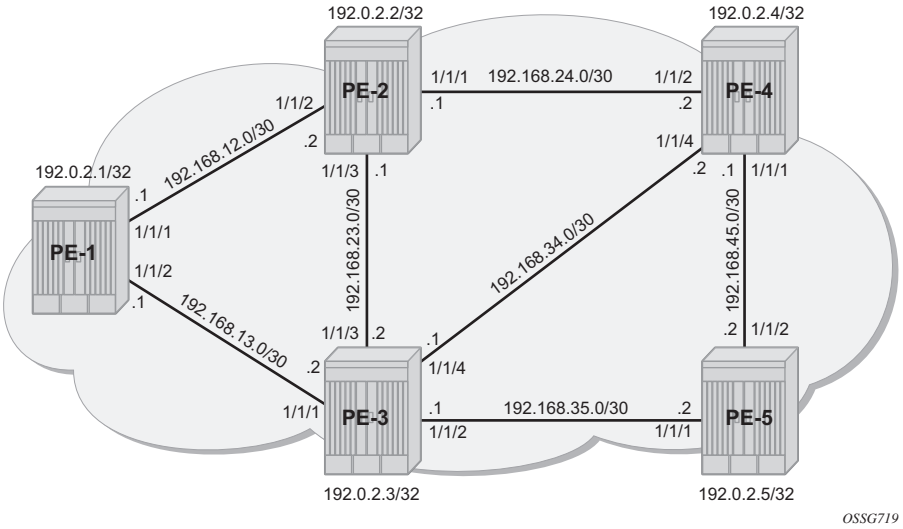
This section provides information to configure:

- Configure the IP/MPLS network
- Enable LDP FRR and Verification
- Enable Synchronization Timer
- Data Path Verification

Additional topics include:

- Metric Change
- IS-IS Overload Bit

Figure 162 **Initial Topology**



Configure the IP/MPLS network

The system addresses and IP interface addresses are configured according to Figure 162. An interior gateway protocol (IGP) is needed to distribute routing information on all PEs. In our case, the IGP is IS-IS where each PE is acting as a level 2 router. A configuration example is shown for PE-1. Similar configurations can be derived for the other PEs.

```
*A:PE-1# configure router
        isis
            level-capability level-2
            level 2
                wide-metrics-only
            exit
            interface "system"
            exit
            interface "int-PE-1-PE-2"
                interface-type point-to-point
            exit
            interface "int-PE-1-PE-3"
                interface-type point-to-point
            exit
```

IS-IS interfaces are set up as type point-to-point to improve convergence since no designated router/backup designated router (DR/BDR) election process is done. To verify that IS-IS adjacencies are up, **show router isis adjacency** is performed. To check if IP interface addresses/subnets are known on all PEs, **show router route-table** or **show router fib** *slot-number* will display the content of the forwarding information base (FIB).

```
*A:PE-1# show router isis adjacency
===============================================================================
Rtr Base ISIS Instance 0 Adjacency
===============================================================================
System ID               Usage State Hold Interface                     MT-ID
-------------------------------------------------------------------------------
PE-2                    L2    Up    26   int-PE-1-PE-2                 0
PE-3                    L2    Up    20   int-PE-1-PE-3                 0
-------------------------------------------------------------------------------
Adjacencies : 2
===============================================================================
*A:PE-1#

*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     07d23h34m  0
       system                                                       0
192.0.2.2/32                                    Remote  ISIS      07d23h34m  18
       192.168.12.2                                                 10
192.0.2.3/32                                    Remote  ISIS      07d23h17m  18
       192.168.13.2                                                 10
192.0.2.4/32                                    Remote  ISIS      07d22h58m  18
       192.168.12.2                                                 20
192.0.2.5/32                                    Remote  ISIS      07d23h17m  18
       192.168.13.2                                                 20
192.168.12.0/30                                 Local   Local     07d23h34m  0
       int-PE-1-PE-2                                                0
192.168.13.0/30                                 Local   Local     07d23h34m  0
       int-PE-1-PE-3                                                0
192.168.23.0/30                                 Remote  ISIS      07d23h16m  18
       192.168.12.2                                                 20
192.168.24.0/30                                 Remote  ISIS      07d23h34m  18
       192.168.12.2                                                 20
192.168.34.0/30                                 Remote  ISIS      07d23h17m  18
       192.168.13.2                                                 20
192.168.35.0/30                                 Remote  ISIS      07d23h17m  18
       192.168.13.2                                                 20
192.168.45.0/30                                 Remote  ISIS      03h59m10s  18
       192.168.12.2                                                 30
-------------------------------------------------------------------------------
No. of Routes: 12
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
```

```
===============================================================================
*A:PE-1#

*A:PE-1# show router fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
    NextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                                    LOCAL
    192.0.2.1 (system)
192.0.2.2/32                                                    ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.3/32                                                    ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.0.2.4/32                                                    ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.5/32                                                    ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.168.12.0/30                                                 LOCAL
    192.168.12.0 (int-PE-1-PE-2)
192.168.13.0/30                                                 LOCAL
    192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.34.0/30                                                 ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.168.35.0/30                                                 ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.168.45.0/30                                                 ISIS
    192.168.12.2 (int-PE-1-PE-2)
-------------------------------------------------------------------------------
Total Entries : 12
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

Initially, the default IS-IS Level 2 metric is applied on all interfaces (value 10).

```
*A:PE-1# show router isis status | match "L2 Default Metric"
L2 Default Metric    : 10
*A:PE-1#
```

The next step in the process of setting up the IP/MPLS network is setting up interface-LDP sessions on all interfaces. If the keyword dual-stack and ipv4 no shutdown isn't included in the command, as is the case for interface "int-PE-1PE-2", it will be added automatically.

```
*A:PE-1# configure router
        ldp
            interface-parameters
                interface "int-PE-1-PE-2"
                exit
```

```
                interface "int-PE-1-PE-3" dual-stack
                    ipv4
                        no shutdown
                    exit
                exit
            exit
            targeted-session
            exit
            no shutdown
        exit all
```

There is now a full mesh of LDP label switched paths (LSPs) set up between all system interfaces of the PEs, and the tunnel table on PE-1 will look like this:

```
*A:PE-1# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.2/32      ldp       MPLS  65537     9        192.168.12.2   10
192.0.2.3/32      ldp       MPLS  65538     9        192.168.13.2   10
192.0.2.4/32      ldp       MPLS  65539     9        192.168.12.2   20
192.0.2.5/32      ldp       MPLS  65540     9        192.168.13.2   20
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

The LDP LSP metric follows the IGP cost. Optionally, LSP metrics can be applied but that is out of the scope for this chapter.

Enable LDP FRR and Verification

Because LDP FRR is using LFA next-hop pre-computation by the IGP (as described in RFC 5286), the IGP CLI configuration is as follows:

```
*A:PE-1# configure router isis loopfree-alternate
*A:PE-1# show router isis status | match Loopfree
Loopfree-Alternate   : Enabled
*A:PE-1#
```

After enabling LFA inside the IGP context, FRR needs to be enabled within the LDP context:

```
*A:PE-1# configure router ldp fast-reroute
*A:PE-1# show router ldp status | match FRR
FRR                : Enabled              Mcast Upstream FRR   : Disabled
*A:PE-1#
```

This chapter describes FRR for unicast LDP. For multicast upstream FRR, see chapter Multicast Label Distribution Protocol. After these two CLI commands, the software pre-computes for each LDP FEC in the network both a primary and a backup NHLFE and uploads it to the IOM/IMM. The primary NHLFE corresponds to the label of the FEC received from the primary next-hop as per standard LDP resolution of the FEC prefix in the Routing Table Manager (RTM). The backup NHLFE corresponds to the label received for the same FEC from an LFA next-hop.

For point-to-point interfaces, when multiple LFA next-hops are found for a given primary next-hop, the following selection criteria are used:

- It will pick the node-protect type in favor of the link-protect type.
- If there is more than one LFA next-hop within the selected type, then it will pick one based on the lowest cost.
- If more than one LFA next-hop with the same cost, SPF will select the first one. This is not a deterministic selection and will vary following each SPF calculation.

Several show commands are possible to display LFA information:

The **show router isis statistics** command displays the number of LFA runs on a specific node.

```
*A:PE-1# show router isis statistics
===============================================================================
Rtr Base ISIS Instance 0 Statistics
===============================================================================
---snip---
LFA Statistics
LFA Runs          : 20
 Last scheduled   : 03/10/2016 08:58:54
Partial LFA Runs  : 1
 Last scheduled   : 03/10/2016 08:53:11

RLFA Statistics
RLFA Runs         : 0
---snip---
```

Remote LFA (RLFA) is related to segment routing and is described in chapter Segment Routing with IS-IS Control Plane.

The **show router isis lfa-coverage** command performs a mathematical calculation between the number of nodes and IPv4/IPv6 routes in the network versus present LFA next-hop protections. In our network (see Figure 162), all IS-IS links have a default level 2 metric of 10. This results in all four nodes and all IS-IS routes learned by PE-1 being 100% LFA protected (link or node). See the following output:

```
*A:PE-1# show router isis lfa-coverage
===============================================================================
Rtr Base ISIS Instance 0 LFA Coverage
===============================================================================
```

```
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      0/0(0%)        9/9(100%)           0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      4/4(100%)      9/9(100%)           0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-1#
```

The **show router isis topology lfa detail** command shows the LFA protection type (link or node).

```
*A:PE-1# show router isis topology lfa detail
===============================================================================
Rtr Base ISIS Instance 0 Topology Table
===============================================================================
-------------------------------------------------------------------------------
IS-IS IP paths (MT-ID 0), Level 2
-------------------------------------------------------------------------------
Node      : PE-2.00                          Metric      : 10
Interface : int-PE-1-PE-2                    SNPA        : none
Nexthop   : PE-2

LFA intf  : int-PE-1-PE-3                    LFA Metric  : 20
LFA nh    : PE-3                             LFA type    : linkProtection

Node      : PE-3.00                          Metric      : 10
Interface : int-PE-1-PE-3                    SNPA        : none
Nexthop   : PE-3

LFA intf  : int-PE-1-PE-2                    LFA Metric  : 20
LFA nh    : PE-2                             LFA type    : linkProtection

Node      : PE-4.00                          Metric      : 20
Interface : int-PE-1-PE-2                    SNPA        : none
Nexthop   : PE-2

LFA intf  : int-PE-1-PE-3                    LFA Metric  : 20
LFA nh    : PE-3                             LFA type    : nodeProtection

Node      : PE-5.00                          Metric      : 20
Interface : int-PE-1-PE-3                    SNPA        : none
Nexthop   : PE-3

LFA intf  : int-PE-1-PE-2                    LFA Metric  : 30
LFA nh    : PE-2                             LFA type    : linkProtection
===============================================================================
*A:PE-1#
```

The **show router route-table** command adds an 'L' flag as reference that the associated prefix is having also an LFA next-hop available. For detailed interface address information used by the LFA calculation, use the **show router route-table alternative** or **show router isis alternative** command. The output on PE-1 for PE-4 will look like this:

```
*A:PE-1# show router route-table 192.0.2.4
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.4/32 [L]                                Remote  ISIS      07d23h01m  18
       192.168.12.2                                                    20
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#

*A:PE-1# show router route-table alternative 192.0.2.4
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
      Alt-NextHop                                                   Alt-
                                                                    Metric
-------------------------------------------------------------------------------
192.0.2.4/32                                    Remote  ISIS      07d23h02m  18
       192.168.12.2                                                    20
       192.168.13.2 (LFA)                                              20
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#

*A:PE-1# show router isis routes 192.0.2.4 alternative

===============================================================================
Rtr Base ISIS Instance 0 Route Table (alternative)
===============================================================================
Prefix[Flags]                    Metric     Lvl/Typ     Ver.  SysID/Hostname
  NextHop                                                MT     AdminTag/SID[F]
Alt-Nexthop                                              Alt-   Alt-Type
                                                         Metric
-------------------------------------------------------------------------------
192.0.2.4/32                     20         2/Int.      7     PE-2
   192.168.12.2                                            0       0
   192.168.13.2(L)                                        20      NP
```

```
-------------------------------------------------------------------------------
No. of Routes: 1
Flags        : L = Loop-Free Alternate nexthop
Legend       : LP = linkProtection, NP = nodeProtection
-------------------------------------------------------------------------------
SID[F]       : R  = Re-advertisement
               N  = Node-SID
               nP = no penultimate hop POP
               E  = Explicit-Null
               V  = Prefix-SID carries a value
               L  = value/index has local significance
===============================================================================
*A:PE-1#
```

On PE-1, PE-4 (192.0.2.4/32) has a primary SPF next-hop pointing toward PE-2 (192.168.12.2) and an LFA next-hop pointing toward PE-3 (192.168.13.2).

Using the Inequality 3 formula on PE-1 for prefix 192.0.2.4/32, this becomes:

Inequality 3:

- [SP(backup NHR, D) < {SP(backup NHR, PN) + SP(PN, D)}] or
  [SP (PE-3, PE-4) < {SP (PE-3, PE-2) + SP(PE-2, PE-4)}] or
  [10 < {10 + 10}]

This means that Inequality 3 is met. The calculated LFA next-hop for prefix 192.0.2.4/32 on PE-1 is protecting node PE-2, see Figure 162 for a graphical representation.

The **show router ldp bindings** command displays the label information base (LIB). A BU flag is present in case the associated label is used as backup NHLFE for the given prefix. As an example, a display on PE-1 for prefix PE-4 is as follows.

This is only possible because the 7x50 LDP implementation is using liberal retention mode which means that every label mapping received by a peer is retained regardless of whether the LSR is the next-hop for the advertised mapping.

```
*A:PE-1# show router ldp bindings prefixes prefix 192.0.2.4/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                  IngLbl                    EgrLbl
Peer                                    EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.4/32                            262140N                   262140
```

```
192.0.2.2:0                                 1/1/1
192.168.12.2

192.0.2.4/32                                262140U                 262140BU
192.0.2.3:0                                 1/1/2
192.168.13.2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 2
===============================================================================
*A:PE-1#
```

The **show router ldp bindings active** command displays the label forwarding information base (LFIB). Also the BU flag is present and in addition a reference to the label action itself: **pop** for eLER, **push** for iLER and **swap** for LSR.

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                      Op              IngLbl    EgrLbl
EgrNextHop                                  EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                                Push              --      262140
192.168.12.2                                1/1/1

192.0.2.4/32                                Push              --      262140BU
192.168.13.2                                1/1/2

192.0.2.4/32                                Swap            262140    262140
192.168.12.2                                1/1/1

192.0.2.4/32                                Swap            262140    262140BU
192.168.13.2                                1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-1#
```

Enable Synchronization Timer

Within an MPLS network using LDP it is common practice to enable a synchronization timer between LDP and the IGP. Also when LDP FRR is enabled, a situation can occur in which a synchronization timer between IGP and LDP will help: the revert scenario. When the interface for the previous primary next hop is restored, IGP may re-converge before LDP completed the FEC exchange with its neighbor over that interface. This may cause LDP to remove the LFA next-hop from the FEC and blackhole traffic.

In order to avoid these situations, it is recommended to first enable IGP-LDP synchronization on the LDP interface. The time is expressed in seconds and can have a value between 1 and 1800 seconds. The LDP synchronization timer is enabled as follows:

```
*A:PE-1# configure router interface "int-PE-1-PE-2" ldp-sync-timer 10
*A:PE-1# configure router interface "int-PE-1-PE-3" ldp-sync-timer 10
```

The configuration on the other nodes is similar.

When this timer is enabled, it means that when an interface is restored again, the IGP will advertise this link in the network with an infinite metric. The **ldp-sync-timer** is started, LDP adjacencies are brought up together with a label exchange. After the **ldp-sync-timer** expires, the normal metric is advertised in the network again.

Data Path Verification

Data path verification is performed using a Layer 2/VLL service. Traffic generator ports are connected toward PE-1 and PE-5, and an Epipe service is created using an MPLS LDP based service distribution path (SDP) on both PE-1 and PE-5.

```
*A:PE-1# configure service
        sdp 15 mpls create
            far-end 192.0.2.5
            ldp
            no shutdown

*A:PE-1# configure service
        epipe 1 customer 1 create
            service-mtu 1450
            sap 1/1/3:1 create
            exit
            spoke-sdp 15:1 create
            exit
            no shutdown
```

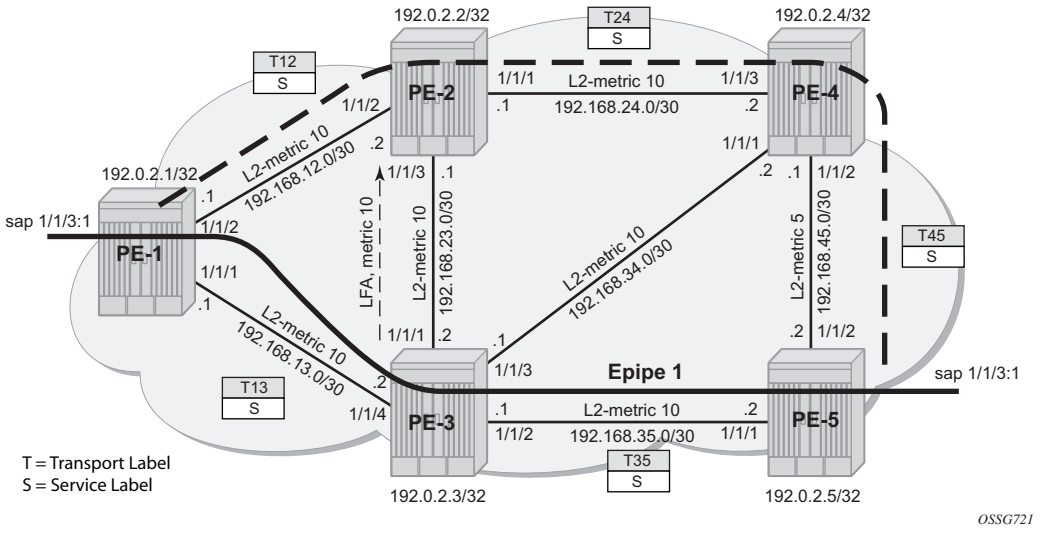
A similar configuration is configured on PE-5.

The IS-IS Level 2 metric value on the interface between PE-4 and PE-5 is decreased to 5, see Figure 163.

```
*A:PE-4# configure router isis interface int-PE-4-PE-5 level 2 metric 5

*A:PE-5# configure router isis interface int-PE-5-PE-4 level 2 metric 5
```

Figure 163 **Data Verification, Direction PE-1 => PE-5 Using VLL Service**



In this setup, there is a loopfree alternate for prefix PE-5 from PE-1 that is protecting the node PE-3:

```
*A:PE-1# show router isis topology lfa detail
===============================================================================
Rtr Base ISIS Instance 0 Topology Table
===============================================================================
-------------------------------------------------------------------------------
IS-IS IP paths (MT-ID 0),   Level 2
-------------------------------------------------------------------------------
---snip---
Node      : PE-5.00                          Metric      : 20
Interface : int-PE-1-PE-3                    SNPA        : none
Nexthop   : PE-3

LFA intf  : int-PE-1-PE-2                    LFA Metric  : 25
LFA nh    : PE-2                             LFA type    : nodeProtection
===============================================================================
*A:PE-1#

*A:PE-1# show router isis routes alternative 192.0.2.5
===============================================================================
Rtr Base ISIS Instance 0 Route Table (alternative)
```

```
===============================================================================
Prefix[Flags]                       Metric     Lvl/Typ     Ver.  SysID/Hostname
  NextHop                                                  MT     AdminTag/SID[F]
Alt-Nexthop                                                Alt-   Alt-Type
                                                          Metric
-------------------------------------------------------------------------------
192.0.2.5/32                        20         2/Int.      10    PE-3
   192.168.13.2                                              0       0
   192.168.12.2(L)                                          25      NP
-------------------------------------------------------------------------------
No. of Routes: 1
Flags        : L = Loop-Free Alternate nexthop
Legend       : LP = linkProtection, NP = nodeProtection
-------------------------------------------------------------------------------
SID[F]       : R  = Re-advertisement
               N  = Node-SID
               nP = no penultimate hop POP
               E  = Explicit-Null
               V  = Prefix-SID carries a value
               L  = value/index has local significance
===============================================================================
*A:PE-1#
```

In normal conditions, MPLS traffic from PE-1 toward PE-5 over Epipe 1 will have two MPLS labels: 1) outer (transport) label given by LDP protocol, swapped on each intermediate LSR and 2) inner (service) label given by T-LDP, the same end-to-end. See the following show commands.

The T-LDP service label is S (262138):

```
*A:PE-1# show router ldp bindings services service-id 1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, WP - Label Withdraw Pending, C - Cpipe Service
        BU - Alternate For Fast Re-Route, TLV - (Type, Length: Value)
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Service FEC 128 Bindings
===============================================================================
Type                                        VCId       SDPId     IngLbl  LMTU
Peer                                        SvcId                EgrLbl  RMTU
-------------------------------------------------------------------------------
E-Eth                                       1          15        262137U 1436
192.0.2.5:0                                 1                    262138S 1436
-------------------------------------------------------------------------------
No. of VC Labels: 1
===============================================================================
===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                          AGII        IngLbl    LMTU
```

```
TAII                                                Type        EgrLbl    RMTU
Peer                                                SvcId       SDPId
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
```

The transport LDP label between PE-1 and PE-3 for prefix 192.0.2.5/32 is T13 (262139):

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                      Op              IngLbl    EgrLbl
EgrNextHop                                  EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                                Push              --      262139
192.168.13.2                                1/1/2

192.0.2.5/32                                Push              --      262139BU
192.168.12.2                                1/1/1

192.0.2.5/32                                Swap            262139    262139
192.168.13.2                                1/1/2

192.0.2.5/32                                Swap            262139    262139BU
192.168.12.2                                1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-1#
```

The transport LDP label between PE-3 and PE-5 for prefix 192.0.2.5/32 is T35 (262143):

```
*A:PE-3# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
```

```
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                            Push              --      262143
192.168.35.2                            1/1/2

192.0.2.5/32                            Push              --      262139BU
192.168.34.2                            1/1/4

192.0.2.5/32                            Swap            262139    262143
192.168.35.2                            1/1/2

192.0.2.5/32                            Swap            262139    262139BU
192.168.34.2                            1/1/4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
*A:PE-3#
```

When PE-3 reboots, PE-1 performs an immediate swap to LFA next-hop for prefix 192.0.2.5/32 bypassing PE-3. The service label remains the same, only the transport labels can change on the network ports PE-1 <=> PE-2, PE-2 <=> PE-4 and PE-4 <=> PE-5. Refer to the following show commands.



Note: The LDP FRR MPLS label stack will never contain more than two labels. This is different from RSVP-TE FRR facility mode which uses a three-label MPLS stack.

The T-LDP service label is S (262138):

```
*A:PE-1# show router ldp bindings services service-id 1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, WP - Label Withdraw Pending, C - Cpipe Service
        BU - Alternate For Fast Re-Route, TLV - (Type, Length: Value)
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Service FEC 128 Bindings
===============================================================================
Type                                      VCId       SDPId     IngLbl  LMTU
Peer                                      SvcId                EgrLbl  RMTU
-------------------------------------------------------------------------------
```

```
E-Eth                                          1          15        262137U 1436
192.0.2.5:0                                    1                    262138S 1436

-------------------------------------------------------------------------------
No. of VC Labels: 1
===============================================================================

===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                            AGII        IngLbl    LMTU
TAII                                            Type        EgrLbl    RMTU
Peer                                            SvcId       SDPId
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-1#
```

The transport LDP label value between PE-1 and PE-2 for prefix 192.0.2.5/32 is the same label (previously tagged as BU) as before the node failure event: T12 (262139):

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                      Op              IngLbl    EgrLbl
EgrNextHop                                  EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                                Push              --      262139
192.168.12.2                                1/1/1

192.0.2.5/32                                Swap            262139    262139
192.168.12.2                                1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

The transport LDP label between PE-2 and PE-4 for prefix 192.0.2.5/32 is T24 (262139):

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.5/32
```

```
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                            Push              --      262139
192.168.24.2                            1/1/1

192.0.2.5/32                            Swap            262139    262139
192.168.24.2                            1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-2#
```

The transport LDP label between PE-4 and PE-5 for prefix 192.0.2.5/32 is T45 (262143):

```
*A:PE-4# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                            Push              --      262143
192.168.45.2                            1/1/1

192.0.2.5/32                            Swap            262139    262143
192.168.45.2                            1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
```

```
*A:PE-4#
```

Additional Topics

Metric Change

Restore all the level 2 metrics back to their default value (10) before applying further changes:

```
*A:PE-4# configure router isis interface "int-PE-4-PE-5" level 2 no metric
*A:PE-5# configure router isis interface "int-PE-5-PE-4" level 2 no metric
```

Suppose that the IS-IS level 2 metric between PE-2 and PE-3 changes to 30, then 100% LFA coverage is no longer possible. The IS-IS level 2 metric is modified as follows:

```
*A:PE-3# configure router isis interface "int-PE-3-PE-2" level 2 metric 30
*A:PE-2# configure router isis interface "int-PE-2-PE-3" level 2 metric 30
```

On PE-1, Inequality 3 formula will find LFA next-hop coverages for prefix PE-4 and PE-5. Inequality formula 1 will find LFA next-hop coverages for prefix PE-4, PE-5, and the subnet between PE-4 and PE-5.

Both inequality formulas are visualized in Figure 165 and Figure 164 for prefix 192.0.2.5/32 (= PE-5) on PE-1 which serves as the source node for LFA next-hop computation.

Figure 164 **LFA Computation, Inequality 1 for Prefix PE-5 (D) on PE-1 (S)**

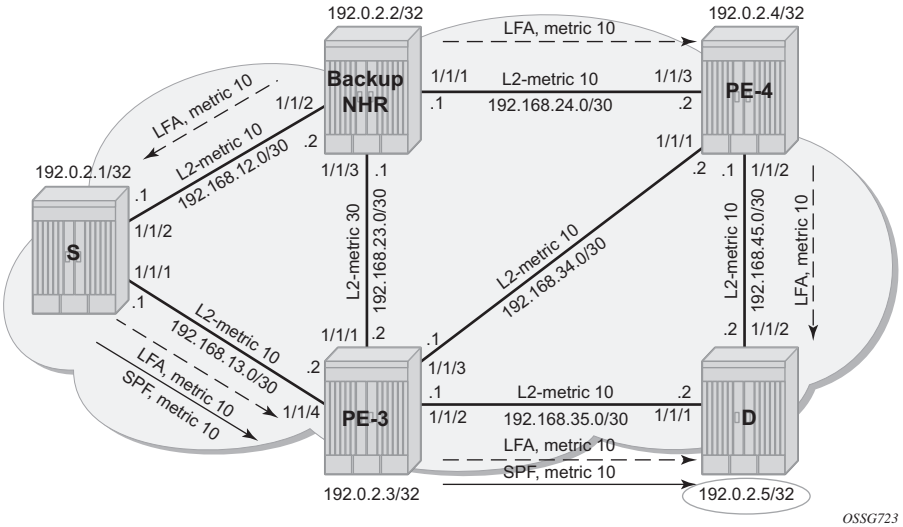
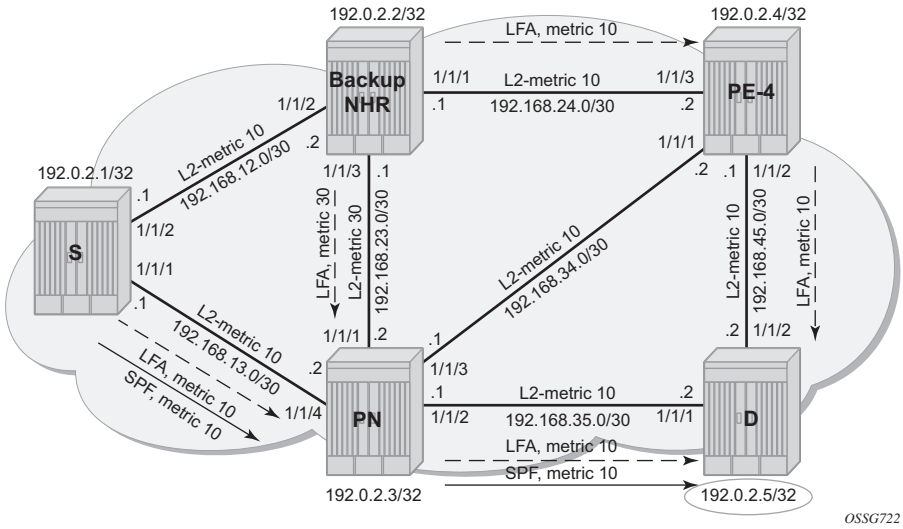


Figure 165 **LFA Computation, Inequality 3 for Prefix PE-5 (D) on PE-1 (S)**



Inequality 3 formula:

- [SP(backup NHR, D) < {SP(backup NHR, PN) + SP(PN, D)}]

For a node LFA next-hop calculation of prefix 192.0.2.5/32 (D) on PE-1, this means that the shortest path from backup next-hop router PE-2 toward destination PE-5 must be smaller than the sum of the shortest path from backup next-hop router PE-2 toward protected node PE-3 with the shortest path from protected node PE-3 to destination PE-5

The shortest path from backup next-hop router PE-2 toward destination PE-5 is going via PE-4, using IS-IS level 2 metric 10 for interface “int-PE-2-PE-4” and IS-IS level 2 metric 10 for interface “int-PE-4-PE-5”. The shortest path from backup next-hop router (PE-2) toward protected node (PE-3) uses IS-IS level 2 metric 30 for interface “int-PE-2-PE-3”. The shortest path from protected node (PE-3) to destination (PE-5) uses IS-IS level 2 metric 10 for interface “int-PE-3-PE-5”. The calculation is as follows:

```
Prefix 192.0.2.5/32: SP (PE-2, PE-5) < SP (PE-2, PE-3) + SP (PE-3, PE-5)
                           10 + 10  < 30 + 10 => OK
```

Inequality 1 formula:

- SP(backup NHR, D) < {SP(backup NHR, S) + SP(S, D)}

For a link LFA next-hop calculation of prefix 192.0.2.5/32 (D) on PE-1, this means that the shortest path from backup next-hop router PE-2 toward destination PE-5 must be smaller than the sum of the shortest path from backup next-hop router PE-2 toward source PE-1 with the shortest path from source PE-1 to destination PE-5.

The shortest path from backup next-hop router PE-2toward destination PE-5 is going over PE-4, using IS-IS level 2 metric 10 for interface “int-PE-2-PE-4” and IS-IS level 2 metric 10 for interface “int-PE-4-PE-5”. The shortest path from backup next-hop router PE-2 toward source PE-1 uses IS-IS level 2 metric 10 for interface “int-PE-2-PE-1”. The shortest path from source PE-1 to destination PE-5 follows the normal SPF calculation, going over PE-3, using IS-IS level 2 metric 10 for interface “int-PE-1-PE-3”, and IS-IS level 2 metric 10 for interface “int-PE-3-PE-5”.

The calculation is as follows:

```
Prefix 192.0.2.5/32 :  SP(PE-2,PE-5) < SP(PE-2,PE-1) + SP(PE-1,PE-5)
                          10 + 10    <    10 + (10 + 10)            => OK
```

For completeness, all the other Inequality 1 calculations on PE-1 are given:

```
Prefix 192.0.2.2/32 :  SP(PE-3,PE-2) < SP(PE-3,PE-1) + SP(PE-1,PE-2)
                             30     <      10 + 10                    => NOK
Prefix 192.0.2.3/32 :  SP(PE-2,PE-3) < SP(PE-2,PE-1) + SP(PE-1,PE-3)
                             30     <      10 + 10                    => NOK
Prefix 192.0.2.4/32 :  SP(PE-3,PE-4) < SP(PE-3,PE-1) + SP(PE-1,PE-4)
                             10     <      10 + (10 + 10)             => OK
Prefix 192.168.23.0/30 :  SP(PE-3,D) < SP(PE-3,PE-1) + SP(PE-1,D)
                             30     <      10 + (10 + 10)             => NOK
```

```
Prefix 192.168.24.0/30 :  SP(PE-3,D) < SP(PE-3,PE-1) + SP(PE-1,D)
                           30 + 10   <      10 + (10 + 10)               => NOK
Prefix 192.168.34.0/30 :  SP(PE-2,D) < SP(PE-2,PE-1) + SP(PE-1,D)
                           30 + 10   <      10 + (10 + 10)               => NOK
Prefix 192.168.35.0/30 :  SP(PE-2,D) < SP(PE-2,PE-1) + SP(PE-1,D)
                           30 + 10   <      10 + (10 + 10)               => NOK
Prefix 192.168.45.0/30 :  SP(PE-3,D) < SP(PE-3,PE-1) + SP(PE-1,D)
                           10 + 10   <      10 + (10 + 10 + 10)          => OK
```

Considering all inequality 1 calculations, only three of these are valid (OK).

In SR OS, the following summary command exists for LFA coverage on the router:

```
*A:PE-1# show router isis lfa-coverage
===============================================================================
LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      0/0(0%)        3/9(33%)            0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      2/4(50%)       3/9(33%)            0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-1#
```

Restore all the IS-IS level 2 metrics back to the default value as follows:

```
*A:PE-2# configure router isis interface "int-PE-2-PE-3" level 2 no metric
*A:PE-3# configure router isis interface "int-PE-3-PE-2" level 2 no metric
```

IS-IS Overload Bit

As stated in RFC 3137, O*SPF Stub Router Advertisement*, sometimes it is useful and desirable for a router not to be a transit node. For those cases, it is also desirable not to have that router used as transit node during the LFA next-hop computation. Within IS-IS protocol this is achieved by setting the overload bit. When other routers detect that this bit is set, they will only use this router for packets destined to the overloaded router's directly connected networks and IP prefixes.

As an example, setting the IS-IS overload condition for a specific time on PE-2 provides following result on PE-1:

```
*A:PE-2# configure router isis overload
  - no overload
  - overload [timeout <seconds>] [max-metric]
```

```
 <seconds>            : [60..1800]

*A:PE-2# configure router isis overload timeout 60

*A:PE-1# show router isis lfa-coverage
===============================================================================
Rtr Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      0/0(0%)        3/9(33%)            0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Unicast     L2      1/4(25%)       3/9(33%)            0/0(0%)
IPV6 Unicast     L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)        0/0(0%)             0/0(0%)
===============================================================================
*A:PE-1#

*A:PE-1# show router route-table alternative
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
      Alt-NextHop                                                  Alt-
                                                                   Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     07d23h49m   0
       system                                                          0
192.0.2.2/32                                    Remote  ISIS      07d23h48m   18
       192.168.12.2                                                    10
       192.168.13.2 (LFA)                                              20
192.0.2.3/32                                    Remote  ISIS      00h01m59s   18
       192.168.13.2                                                    10
192.0.2.4/32                                    Remote  ISIS      00h00m40s   18
       192.168.13.2                                                    20
192.0.2.5/32                                    Remote  ISIS      00h01m59s   18
       192.168.13.2                                                    20
192.168.12.0/30                                 Local   Local     07d23h48m   0
       int-PE-1-PE-2                                                   0
192.168.13.0/30                                 Local   Local     07d23h48m   0
       int-PE-1-PE-3                                                   0
192.168.23.0/30                                 Remote  ISIS      00h01m11s   18
       192.168.12.2                                                    20
       192.168.13.2 (LFA)                                              30
192.168.24.0/30                                 Remote  ISIS      07d23h48m   18
       192.168.12.2                                                    20
       192.168.13.2 (LFA)                                              30
192.168.34.0/30                                 Remote  ISIS      00h01m59s   18
       192.168.13.2                                                    20
192.168.35.0/30                                 Remote  ISIS      00h01m59s   18
       192.168.13.2                                                    20
192.168.45.0/30                                 Remote  ISIS      00h00m40s   18
       192.168.13.2                                                    25
-------------------------------------------------------------------------------
No. of Routes: 12
```

```
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#

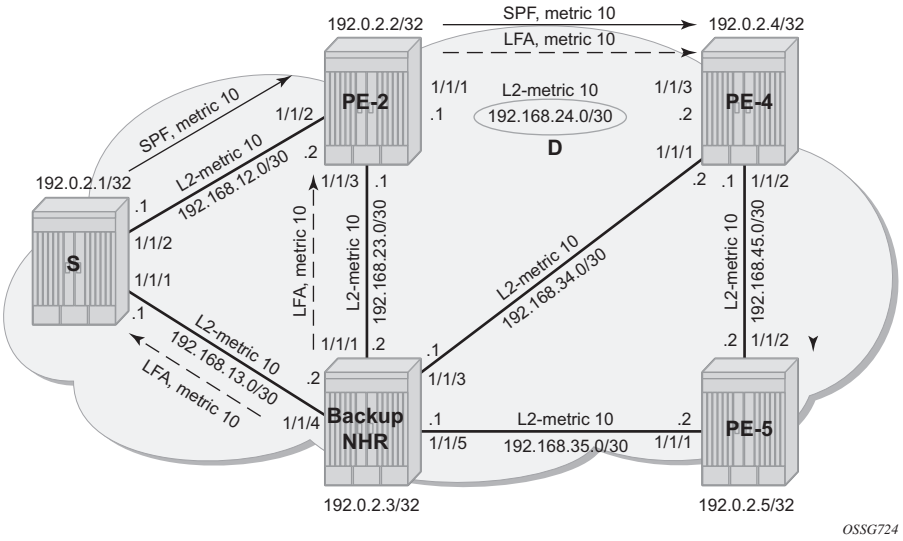
*A:PE-1# show router isis routes alternative

===============================================================================
Rtr Base ISIS Instance 0 Route Table (alternative)
===============================================================================
Prefix[Flags]                     Metric     Lvl/Typ     Ver.  SysID/Hostname
  NextHop                                                MT     AdminTag/SID[F]
Alt-Nexthop                                              Alt-   Alt-Type
                                                        Metric
-------------------------------------------------------------------------------
192.0.2.1/32                      0          2/Int.      4     PE-1
   0.0.0.0                                                 0       0
192.0.2.2/32                      10         2/Int.      6     PE-2
   192.168.12.2                                            0       0
   192.168.13.2(L)                                        20      LP
192.0.2.3/32                      10         2/Int.      24    PE-3
   192.168.13.2                                            0       0
192.0.2.4/32                      20         2/Int.      29    PE-3
   192.168.13.2                                            0       0
192.0.2.5/32                      20         2/Int.      25    PE-3
   192.168.13.2                                            0       0
192.168.12.0/30                   10         2/Int.      4     PE-1
   0.0.0.0                                                 0       0
192.168.13.0/30                   10         2/Int.      24    PE-1
   0.0.0.0                                                 0       0
192.168.23.0/30                   20         2/Int.      27    PE-2
   192.168.12.2                                            0       0
   192.168.13.2(L)                                        30      LP
192.168.24.0/30                   20         2/Int.      6     PE-2
   192.168.12.2                                            0       0
   192.168.13.2(L)                                        30      LP
192.168.34.0/30                   20         2/Int.      24    PE-3
   192.168.13.2                                            0       0
192.168.35.0/30                   20         2/Int.      24    PE-3
   192.168.13.2                                            0       0
192.168.45.0/30                   30         2/Int.      29    PE-3
   192.168.13.2                                            0       0
-------------------------------------------------------------------------------
No. of Routes: 12
Flags        : L = Loop-Free Alternate nexthop
Legend       : LP = linkProtection, NP = nodeProtection
-------------------------------------------------------------------------------
SID[F]       : R  = Re-advertisement
               N  = Node-SID
               nP = no penultimate hop POP
               E  = Explicit-Null
               V  = Prefix-SID carries a value
               L  = value/index has local significance
===============================================================================
*A:PE-1#
```

On PE-1, only three Inequality 1 calculations are possible, as seen in the previous show commands. The Inequality 1 calculation on PE-1 for destination 192.168.24.0/30 is as follows:

```
[SP(backup NHR,D) < {SP(backup NHR,S) + SP(S,D)}]
      SP(PE-3,D) <  SP(PE-3,PE-1) + SP(PE-1,D)
        10 + 10  <       10 + (10 + 10)                 => OK
```

Figure 166 **IS-IS Overload on PE-2, Inequality 1 for 192.168.24.0/30 (D) on PE-1 (S)**



Conclusion

In production MPLS networks where FRR needs to be deployed, a trade off must be made between RSVP-TE FRR versus LDP FRR. The two main advantages of using LDP FRR compared to RSVP FRR are the simple configuration and the fact that LFA next-hop calculation is a local decision, which means there are no interoperability issues when working in a multi-vendor environment. The main disadvantage of using LDP FRR is that LFA next-hop calculation has to deal with the source-route paradigm (inequality formulas exclude a path going over the original source router).

MPLS Transport Profile

This chapter provides information about multiprotocol label switching transport profile (MPLS-TP).

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was originally written for release 12.0.R2. The CLI in the current edition corresponds to 14.0.R1. Multiprotocol label switching transport profile (MPLS-TP) requires a minimum of Flex Path 2 (FP2) or higher based hardware. A control processing module (CPM) 3 or higher is required for the highest bi-directional forwarding detection (BFD) scale using 10ms control packet timers.

The reader should be familiar with the configuration of IP/MPLS and virtual leased line (VLL) services in SR OS.

MPLS-TP was first introduced in SR OS release 11.0.R4 and further enhancements were added in subsequent releases.

Summary

MPLS-TP is intended to allow MPLS to be operated in a manner similar to existing transport technologies, with static configuration of transport paths (particularly with no requirement for a dynamic control plane), proactive in-band and on-demand operations and maintenance (OAM), and protection mechanisms that do not rely on a control plane (for example, resource reservation protocol with traffic engineering

(RSVP-TE)) to operate. The 7x50 can operate both as a label edge router (LER) and label switching router (LSR) for MPLS-TP LSPs, and as a terminating provider edge (T-PE) and switching provider edge (S-PE) for pseudowires (PWs) with MPLS-TP OAM. The 7x50 can therefore act as a node within an MPLS-TP network, or as a gateway between MPLS-TP and IP/MPLS domains.

Overview

MPLS can provide a network layer with packet transport services. In some operational environments it is desirable that the operation and maintenance of such an MPLS based packet transport network follows the operational models typically used in traditional optical transport networks (for example with SONET, SDH) while providing additional OAM, survivability and other maintenance functions targeted at that environment.

MPLS-TP defines a profile of MPLS targeted at transport applications. This profile defines specific MPLS characteristics and extensions required to meet the transport requirements, while retaining compliance with the standard IETF MPLS architecture and label-switching paradigm. The basic architecture and requirements for MPLS-TP are described by the IETF in RFC 5654, RFC 5921 and RFC 5960, in order to meet two objectives:

- To enable MPLS to be deployed in a transport network and operated in a manner similar to existing transport technologies.
- To enable MPLS to support packet transport services with a similar degree of predictability to that found in existing transport networks.

In order to meet these objectives, MPLS-TP has a number of high-level characteristics:

- MPLS-TP, including resilience and protection, operates in the absence of an IP control plane and IP. MPLS-TP does not modify the MPLS forwarding architecture, which is based on existing pseudowire and LSP constructs. Point-to-point LSPs may be unidirectional or bi-directional. Bi-directional LSPs must be congruent (i.e. co-routed and follow the same path in each direction) and are the only supported type on the 7x50. MPLS-TP is only supported on static LSPs and pseudowires (PWs). Also, there is no LSP merging.
- LSP and pseudowire monitoring is achieved using in-band OAM and does not rely on control plane or IP routing functions to determine the health of a path, for example, LDP hello failures do not trigger protection.

The system supports MPLS-TP on LSPs and PWs with static labels. MPLS-TP is not supported on dynamically signaled LSPs and PWs, although switching a static MPLS-TP PW to a targeted LDP (T-LDP) signaled PW is supported. MPLS-TP is supported for Epipe, Apipe and Cpipe VLLs, and Epipe spoke SDP termination on IES, VPRN and VPLS. Static PWs may use SDPs in addition to static MPLS-TP LSPs or RSVP-TE LSPs.

The following MPLS-TP OAM and protection mechanisms defined by the IETF are supported:

- MPLS-TP generic associated channel for LSPs and PWs (RFC 5586)
- MPLS-TP identifiers (RFC 6370)
- Proactive continuity check (CC), connectivity verification (CV), and remote defect indicator (RDI) using bi-directional forwarding detection (BFD) for LSPs (RFC 6428)
- On-demand CV for LSPs and PWs using LSP ping and LSP trace (RFC 6426)
- 1-for-1 Linear protection for LSPs (RFC 6378)
- Static PW status signaling (RFC 6478)

The system can play the role of an LER and an LSR for static MPLS-TP LSPs, and a PE/T-PE and an S-PE for static MPLS-TP PWs. It can also act as an S-PE for MPLS-TP segments between an MPLS network that strictly follows the transport profile and an MPLS network that supports both MPLS-TP and dynamic IP/MPLS.

Configuration

This section details the configuration steps for a set of simple MPLS-TP examples.

The following example network is used (Figure 167). It consists of four nodes and two Epipe VLL services. One service is used to transport traffic across a network domain consisting of only static MPLS-TP LSPs (Epipe 10) from PE-1 to PE-2. The other Epipe (Epipe 20) is used to transport traffic from PE-1 in the MPLS-TP domain to a VPLS service on PE-4 in an IP/MPLS domain. A static MPLS-TP LSP exists between PE-1 and PE-2, while a dynamic RSVP-TE LSP exists between PE-2 and PE-4.

Figure 167 **MPLS-TP Example Network Showing LSPs**

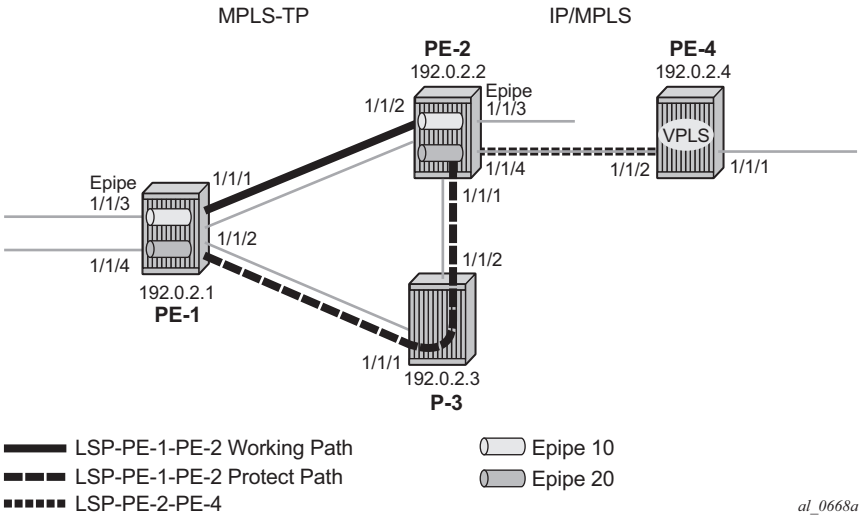


Figure 168 shows further details of the logical architecture of the services in the example network. The Epipe spoke-sdps use the static MPLS-TP transport LSP between PE-1 and PE-2, and the dynamically signaled RSVP-TE LSP between PE-2 and PE-4. The MPLS-TP LSP is protected using 1:1 linear protection, with a working path from PE-1 to PE-2, and a protect path from PE-1, through LSR P-3, to PE-2. The Ethernet PW for Epipe 10 connects an Ethernet SAP on port 1/1/3 on PE-1 to an Ethernet SAP on port 1/1/3 on PE-2. The PW for Epipe 20 connects an Ethernet SAP on port 1/1/4 on PE-1 to the VPLS on PE-4 and is switched between a static MPLS-TP segment and a dynamic targeted LDP (T-LDP) segment at PE-2. PE-2 thus acts as a gateway between the MPLS-TP domain and the IP/MPLS domain.

Figure 168 **MPLS-TP Example Network Showing Services Detail**

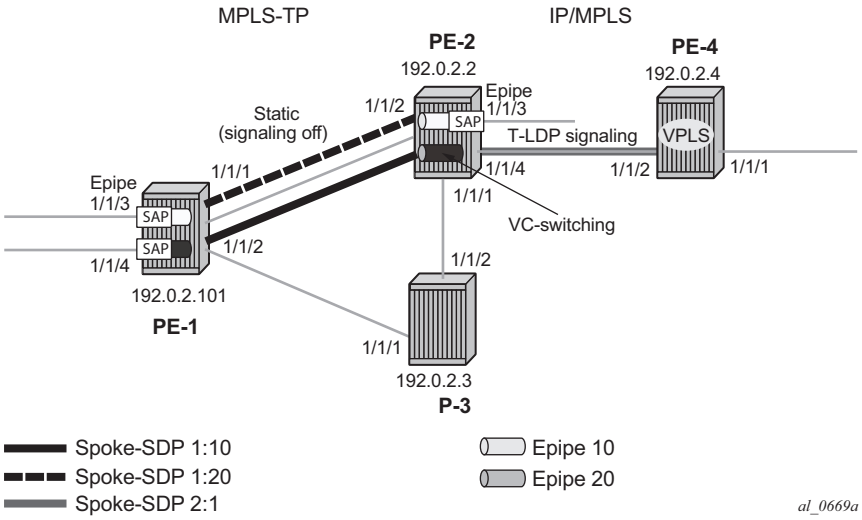
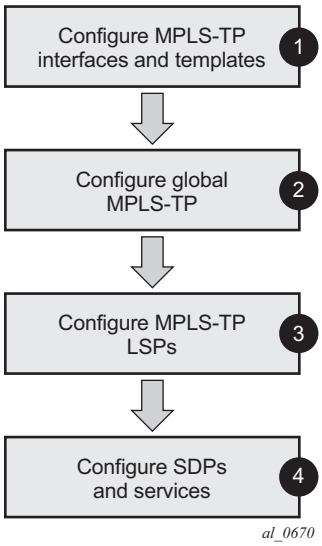


Figure 169 shows the configuration process to be followed when setting up MPLS-TP.

Figure 169 **MPLS-TP Configuration Steps**



Configure MPLS-TP Interfaces and Templates

MPLS-TP LSPs can use either numbered or unnumbered network IP interfaces, or unnumbered network interfaces that have been configured to operate without relying on IP routing. This non-IP interface type does not have an IP address associated with it and may be configured to have either a unicast, broadcast or multicast MAC address. The intent of using a broadcast or multicast MAC address is to enable a standard set of MAC addresses to be configured for a network without requiring any changes to the configuration of neighboring router interfaces each time an interface to which a router is connected is changed. If a broadcast or multicast MAC address is used, then the operator should take care that only a point-to-point link is connected to the Ethernet port used by the interface. Otherwise, MPLS-TP packets may be replicated to each remote port to which the link is connected.

The non-IP network interface type is known as an unnumbered-mpls-tp interface. Only MPLS-TP can use this interface type. That is, other IP protocols are blocked from using it. Furthermore, ARP is not used for next hop resolution. This example uses unnumbered-mpls-tp interfaces.

Unnumbered MPLS-TP interfaces are configured on each network-facing interface for the nodes in the MPLS-TP domain, as shown below. This is done using the **unnumbered-mpls-tp** keyword at create time. In addition, the **static-arp unnumbered** command is used to set the next-hop unicast, broadcast or multicast MAC address of the interface. The system interface should also be configured. Numbered IP Network interfaces, bound to port 1/1/4 of PE-2 and port 1/1/2 of PE-4 are used for the IP/MPLS portion of the network in Figure 167.

```
*A:PE-1# configure router
        interface "int-PE-1-P-3" unnumbered-mpls-tp
            port 1/1/2
            static-arp unnumbered 01:00:5e:90:00:00
            no shutdown
        exit
        interface "int-PE-1-PE-2" unnumbered-mpls-tp
            port 1/1/1
            static-arp unnumbered 01:00:5e:90:00:00
            no shutdown
        exit
        interface "system"
            address 192.0.2.1/32
        exit
        autonomous-system 64511


*A:PE-2# configure router
        interface "int-PE-2-P-3" unnumbered-mpls-tp
            port 1/1/1
            static-arp unnumbered 01:00:5e:90:00:00
            no shutdown
        exit
```

```
        interface "int-PE-2-PE-1" unnumbered-mpls-tp
            port 1/1/2
            static-arp unnumbered 01:00:5e:90:00:00
            no shutdown
        exit
        interface "int-PE-2-PE-4"
            address 192.168.24.1/30
            port 1/1/4
        exit
        interface "system"
            address 192.0.2.2/32
        exit
        autonomous-system 65535
        static-route-entry 192.0.2.4/32
            next-hop 192.168.24.2
                no shutdown
            exit
        exit

*A:P-3# configure router
        interface "int-P-3-PE-1" unnumbered-mpls-tp
            port 1/1/1
            static-arp unnumbered 01:00:5e:90:00:00
            no shutdown
        exit
        interface "int-P-3-PE-2" unnumbered-mpls-tp
            port 1/1/2
            static-arp unnumbered 01:00:5e:90:00:00
            no shutdown
        exit
        interface "system"
            address 192.0.2.3/32
        exit
        autonomous-system 65535

*A:PE-4# configure router
        interface "int-PE-4-PE-2"
            address 192.168.24.2/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.4/32
        exit
        autonomous-system 65535
        static-route-entry 192.0.2.2/32
            next-hop 192.168.24.1
                no shutdown
            exit
        exit
```

Next, MPLS should be configured on each of the interfaces to be used by MPLS-TP. As an example, only PE-1 configuration is shown although a similar configuration is provisioned on PE-2 and P-3.

```
*A:PE-1# configure router
        mpls
            mpls-tp
            exit
```

```
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-P-3"
            exit
            no shutdown
        exit
```

PE-4 is an IP/MPLS only node so there is no MPLS TP configuration

```
*A:PE-4# configure router
        mpls
            interface "int-PE-4-PE-2"
            exit
            no shutdown
        exit
```

The **mpls** context must be in the **no shutdown** state to enable MPLS-TP.

Static labels are used by MPLS-TP LSPs and PWs. By default, SR OS splits the full range in a static and a dynamic range,and these ranges are as follows:

```
*A:PE-1# show router mpls-labels label-range

===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505856      505856
    Seg-Route   0           0           -           0           505856
===============================================================================
*A:PE-1#
```

This can be modified by configuration. To reserve 2000 labels starting from label 32, the following command is launched:

```
*A:PE-1# configure router mpls-labels static-label-range 2000
```

As a result, the range from label 32 to 2031 is reserved for static labels.

```
*A:PE-1# show router mpls-labels label-range

===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          2031        -           2000        2000
Dynamic         2032        524287      0           522256      522256
    Seg-Route   0           0           -           0           522256
===============================================================================
*A:PE-1#
```

In this case, there is no need to modify the range for static labels. The labels that will be chosen are in the default range.

Next, one or more bidirectional forwarding detection (BFD) templates are configured on the LERs. These templates are used to define BFD state machine parameters used for BFD continuity check (CC) on an LSP, including the transmit and receive timer intervals (in milliseconds). CPM network processor BFD is required if timer intervals as short as 10ms are to be used, but depending on the platform, 100ms BFD may use CPU based BFD (as shown in the example here).

```
A:PE-1>config>router>bfd#
      abort           - Discard the changes that have been made to bfd template
                        during a session
      begin           - Switch to edit mode for bfd template - use commit to save
                        or abort to discard the changes made in a session
 [no] bfd-template    + Configure a bfd template
      commit          - Save the changes made to bfd template during a session


*A:PE-1# configure router bfd bfd-template
  - bfd-template <[32 chars max]>
  - no bfd-template <[32 chars max]>

 [no] echo-receive    - Configure echo receive interval
 [no] multiplier      - Configure multiplier
 [no] receive-interv* - Configure receive interval
 [no] transmit-inter* - Configure transmit interval
 [no] type            - Configure the bfd session endpoint type

*A:PE-1#
```

A subset of these parameters is used by MPLS-TP BFD sessions, as follows:

- **transmit-interval** *transmit-interval* and the **receive-interval** *receive-interval* — These are the transmit and receive timers for BFD packets. For MPLS-TP, these are the timers used by BFD CC packets. Values are in milliseconds: 10ms to 100,000ms, with 1ms granularity. Default 10ms for CPM3 or higher, 1 sec for other hardware. The minimum interval that can be supported is hardware dependent. For MPLS-TP BFD Connectivity Verification (CV) packets, a transmit interval of 1 sec is used.
- **multiplier** *multiplier* — Integer 3 – 20. Default: 3. The configured parameter is used for MPLS-TP CC BFD sessions. It is ignored for MPLS-TP combined CC/CV BFD sessions, and the default of 3 is used.
- **type cpm-np** — This selects the CPM network processor as the local termination point for the BFD session. This is used by default for MPLS-TP. Type CPM-NP is needed to configure a transmit interval down to 10ms.

The following CLI illustrates the BFD template configuration at PE-1. Since default parameters are sufficient, only the bfd-template name is configured. BFD templates use a begin/commit model for configuration. Create or modify a template with the **begin** statement. Changes to an existing template or the creation of a new template is not effected until the **commit** statement is entered.

```
*A:PE-1# configure router
        bfd
            begin
            bfd-template "tp-bfd"
            exit
            commit
        exit
```

The following **info detail** command shows the values that are assigned by default.

```
*A:PE-1>config>router>bfd# info detail
----------------------------------------------
            bfd-template "tp-bfd"
                no type
                transmit-interval 100
                receive-interval 100
                multiplier 3
                echo-receive 100
            exit
----------------------------------------------
```

Configure Global MPLS-TP Parameters

MPLS-TP global parameters are configured under **config>router>mpls>mpls-tp**. These include the MPLS-TP identifiers for the node and the range of tunnel identifiers that should be reserved for MPLS-TP LSPs.

Node identifiers include the global ID and the node ID. The node ID may be defined as an unsigned integer or use dotted quad notation (a.b.c.d), but the node ID does not have to be a routable IP address.

The CLI tree for configuring the MPLS-TP identifiers for a node is as follows:

```
*A:PE-1# configure router
        mpls
            mpls-tp
  - mpls-tp
  - no mpls-tp

 [no] global-id       - Global id for MPLS-TP
 [no] node-id         - Node id for MPLS-TP local router
 [no] oam-template    + Configure a MPLS-TP OAM Template
 [no] protection-tem* + Configure a MPLS-TP Protection Template
 [no] shutdown        - Administratively enable/disable the MPLS-TP
```

```
 [no] tp-tunnel-id-r* - Configure MPLS-TP tunnel id range on the ingress router
 [no] transit-path    + Configure a MPLS-TP Transit Path
```

The default value for the global-id is 0. This is used if the global-id is not configured. If an operator expects that inter-domain LSPs will be configured, then it is recommended to set the global ID to the local autonomous system number (ASN) of the node, as configured under **config>router**, to ensure that the combination of global-id and node-id is globally unique. If two-byte ASNs are used, then the two most significant bytes of the global-id are padded with zeros.

The default value of the **node-id** is the system interface IPv4 address. The MPLS-TP context cannot be administratively enabled unless at least a system interface IPv4 address is configured because MPLS requires that this value be configured.

In order to change the values, **config>router>mpls>mpls-tp** must be in the shutdown state. This will bring down all of the MPLS-TP LSPs on the node. New values are propagated to the system when a **no shutdown** is performed.

The following CLI shows the MPLS-TP node identifier configuration for PE-1. A similar configuration is implemented in all routers in this chapter, except that the node-ids must be different (PE-2 is 10.0.0. 2 and P-3 is 10.0.0. 3). In this example the global-id for PE-2 and P-3 equals 65535.

```
*A:PE-1# configure router
        mpls
            mpls-tp
                global-id 64511
                node-id 10.0.0.1
            exit

*A:PE-2# configure router
        mpls
            mpls-tp
                global-id 65535
                node-id 10.0.0.2
            exit
```

Next, protection and OAM templates should be configured at the MPLS-TP LERs. A protection template defines the parameters for the linear protection state coordination mechanism. MPLS-TP Linear Protection is specified in RFC6378. It provides protection for an LSP using a working and a protect path. A protection state coordination (PSC) protocol is used by the LERs at each end of the protected LSP to coordinate whether the working or protect path is used for forwarding. BFD is run on both the working and protect paths.

The linear protection parameters include revertive or non-revertive behavior, the wait-to-restore timer, the rapid-psc-timer and the slow-psc-timer. The wait-to-restore timer (in seconds) defines the time to wait before reverting to the working path if, on restoration of connectivity, the revertive behavior is selected.

The following CLI tree is used to configure the protection template:

```
*A:PE-1>config>router>mpls>tp$ protection-template
  - no protection-template <[32 chars max]>
  - protection-template <[32 chars max]>


     rapid-psc-timer - Configure the rapid Protection Switch Coordination (PSC) timer
 [no] revertive       - Enable/Disable the template's revertive mode
     slow-psc-timer  - Configure the slow Protection Switch Coordination (PSC) timer
 [no] wait-to-restore - Configure the WTR timer for the template
```

See the CLI command descriptions in the MPLS User Guide for further details of these commands.

The OAM template defines generic proactive OAM parameters, such as BFD hold down and hold up timer values (which can be used to introduce some hysteresis if BFD bounces) and the BFD template to use.

The following CLI tree is used to configure the OAM template:

```
*A:PE-1# configure router
        mpls
            mpls-tp
                oam-template
  - no oam-template <template-name>
  - oam-template <template-name>

 <template-name>      : [32 chars max]

 [no] bfd-template - Configure the Bidirectional Forwarding Detection (BFD) template
 [no] hold-time-down  - Configure hold-down dampening timer
 [no] hold-time-up    - Configure the hold-up dampening timer
```

See the CLI command descriptions in the MPLS User Guide for further details of these commands.

MPLS-TP requires the reservation of a tunnel ID range, dedicated for the use of MPLS-TP LSPs. This range is reserved using the following CLI tree:

```
*A:PE-1# configure router
        mpls
            mpls-tp
                tp-tunnel-id-range
  - tp-tunnel-id-range <min> <max>
  - no tp-tunnel-id-range

 <min>                : [1..61440]
 <max>                : [1..61440]
```

The default parameter values are used as shown below, where PE-1 and PE-2 have the same configuration:

```
*A:PE-1# configure router mpls
            mpls-tp
                tp-tunnel-id-range 100 1000
                protection-template "tp-protect"
                exit
                oam-template "tp-oam"
                    bfd-template "tp-bfd"
                exit
                no shutdown
            exit
```

Configure MPLS-TP LSPs

Once the global MPLS-TP parameters have been configured, the system is ready to configure MPLS-TP LSPs. An MPLS-TP LSP is configured under the **config>router>mpls>lsp** context.

Because LSP labels are statically configured, both ends of the LSP must be explicitly configured. The LSP paths must also be explicitly configured in the LSR nodes. MPLS-TP LSPs must use the mpls-tp keyword including a source tunnel number at creation time.

The following commands are used to configure an MPLS-TP LSP at an LER:

```
configure
   router
      mpls
         lsp <lsp-name> mpls-tp <src-tunnel-num>]
            to node-id {<a.b.c.d> | <1.. .4,294,967,295>}
            dest-global-id <global-id>
            dest-tunnel-number <tunnel-num>
            [no] working-tp-path
                lsp-num <lsp-num>
                in-label <in-label>
                out-label <out-label> out-link <if-name> [next-hop <ipv4-address>]
                [no] mep
                   [no] oam-template <name>
                   [no] bfd-enable [cc | cc-cv]
                   [no] shutdown
                   exit
                [no] shutdown
                exit
             [no] protect-tp-path
                lsp-num <lsp-num>
                in-label <in-label>
                out-label <out-label> out-link <if-name> [next-hop <ipv4-address> ]
                [no] mep
                   [no] protection-template <name>
                   [no] oam-template <name>
                   [no] bfd-enable [cc | cc-cv]
                   [no] shutdown
                   exit
                [no] shutdown
```

```
            exit
```

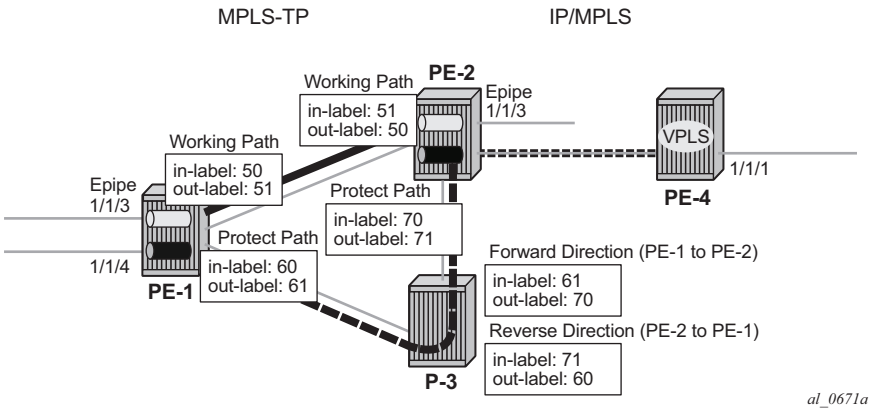
See the CLI command descriptions in the MPLS User Guide for further details of these commands.

A working path and a protect path for LSP LSP-PE-1-P-2 must be configured between PE-1 and PE-2. Each LSP is configured with the full set of MPLS-TP identifiers required to build the LSP ID. Each working path and protect path must have an incoming label, outgoing label and outgoing link configured.

Each working path and protect path also includes a maintenance entity group endpoint (MEP) configuration, under which the applicable OAM template is configured. BFD is also enabled under the MEP context for the path. In this example, BFD operating in CC mode is enabled on the working and protect paths. The protection template, containing parameters for linear protection, is only applied under the protect path context.

Figure 170 shows the LSP working and protect path label values configured at PE-1, PE-2 and P-3. At each node the outgoing label must match the incoming label on the next hop for a given direction. At the LERs (PE-1 and PE-2), the incoming and outgoing label values for each LSP path are configured together. At the LSR (P-3), the label values for the label mapping between ingress and egress for each direction of the path (that is, forward and reverse) are configured together.

Figure 170 **LSP Path Label Value Configurations**



The following shows the LER LSP configuration of PE-1 and PE-2.

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-2" mpls-tp 100
                to node-id 10.0.0.2
                dest-global-id 65535
                dest-tunnel-number 100
                working-tp-path
```

```
                        in-label 50
                        out-label 51 out-link "int-PE-1-PE-2"
                        mep
                            oam-template "tp-oam"
                            bfd-enable cc
                            no shutdown
                        exit
                        no shutdown
                    exit
                    protect-tp-path
                        in-label 60
                        out-label 61 out-link "int-PE-1-P-3"
                        mep
                            protection-template "tp-protect"
                            oam-template "tp-oam"
                            bfd-enable cc
                            no shutdown
                        exit
                        no shutdown
                    exit
                    no shutdown
                exit
                no shutdown
            exit

*A:PE-2# configure router
            mpls
               lsp "LSP-PE-1-PE-2" mpls-tp 100
                    to node-id 10.0.0.1
                    dest-global-id 64511
                    dest-tunnel-number 100
                    working-tp-path
                        in-label 51
                        out-label 50 out-link "int-PE-2-PE-1"
                        mep
                            oam-template "tp-oam"
                            bfd-enable cc
                            no shutdown
                        exit
                        no shutdown
                    exit
                    protect-tp-path
                        in-label 70
                        out-label 71 out-link "int-PE-2-P-3"
                        mep
                            protection-template "tp-protect"
                            oam-template "tp-oam"
                            bfd-enable cc
                            no shutdown
                        exit
                        no shutdown
                    exit
                    no shutdown
                exit
                no shutdown
            exit
```

Since this example requires a protect path to be switched via P-3, a transit path must be configured in P-3. The CLI tree for configuring MPLS-TP transit paths is as follows:

```
configure
   router
      mpls
         mpls-tp
            transit-path <path-name>
               [no] path-id {lsp-num <lsp-num>|working-path|protect-path

                  [src-global-id <global-id>]
                  src-node-id {<ipv4address> | <1.. .4,294,967,295>}
                  src-tunnel-num <tunnel-num>
                  [dest-global-id <global-id>]
                  dest-node-id {<ipv4address> | <1.. .4,294,967,295>}
                  [dest-tunnel-num <tunnel-num>]}
                  lsp-num <lsp-num>

               forward-path
                  in-label <in-label> out-label <out-label>
                  out-link <if-name> [next-hop <ipv4-next-hop>]
                  exit
               reverse-path
                  in-label <in-label> out-label <out-label>
                  [out-link <if-name> [next-hop <ipv4-next-hop>]]
                   exit
               [no] shutdown
```

See the CLI command descriptions in the MPLS User Guide for further details of these commands.

The CLI configuration for the forward and reverse directions of the transit path (that is, the protect path of the LSP) at P-3 is as follows:

```
*A:P-3# configure router
        mpls
            mpls-tp
                transit-path "LSP-PE-1-PE-2"
                    forward-path
                        in-label 61 out-label 70 out-link "int-P-3-PE-2"
                    exit
                    reverse-path
                        in-label 71 out-label 60 out-link "int-P-3-PE-1"
                    exit
                    path-id src-global-id 64511 src-node-id 10.0.0.1 src-tunnel-num
                            100 dest-global-id 65535 dest-node-id 10.0.0.2
                            dest-tunnel-num 100 lsp-num 2
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The example also requires an LSP across the IP/MPLS network to backhaul traffic from PE-2 at the edge of the MPLS-TP network to the VPLS service hosted in PE-4. An RSVP LSP is configured at PE-2 for this purpose, as follows:

```
*A:PE-2# configure router
        mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-2-PE-4"
                to 192.0.2.4
                primary "loose"
                exit
                no shutdown
            exit

*A:PE-2# configure router rsvp no shutdown
```

Create a T-LDP session toward PE-4. LDP over RSVP is preferred (prefer-tunnel-in-tunnel).

```
*A:PE-2# configure router
        ldp
            prefer-tunnel-in-tunnel
            interface-parameters
            exit
            targeted-session
                peer 192.0.2.4
                exit
            exit
        exit
```

A similar configuration is implemented in PE-4.

At this point in the configuration process, it is recommended to check the MPLS-TP LSP configuration and operation of BFD and linear protection.

First, check that the BFD sessions on both the working and protect paths are up:

```
*A:PE-1# show router bfd session

===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
wp::LSP-PE-1-PE-2             Up                   100       100       3
    65535::10.0.0.2           mplsTp               10566     10554     central
pp::LSP-PE-1-PE-2             Up                   100       100       3
    65535::10.0.0.2           mplsTp               7313      6938      central
-------------------------------------------------------------------------------
```

```
No. of BFD sessions: 2
===============================================================================
*A:PE-1#
```

Next, check the currently active path. This can be done using the oam lsp-trace command. The static option must be specified for MPLS-TP LSPs.

```
*A:PE-1# oam lsp-trace static "LSP-PE-1-PE-2"
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 100 byte packets
1  GlobalId 65535 NodeId 10.0.0.2
   rtt=0.564ms rc=3(EgressRtr)
*A:PE-1#
```

This shows that data packets currently follow the working path of the LSP (no transit node is shown).

In order to test the operation of linear protection, the port used by the working path can be shutdown, and the BFD session state checked again:

```
*A:PE-1# configure port 1/1/1 shutdown

*A:PE-1# show router bfd session

===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
wp::LSP-PE-1-PE-2             Down                 1000      100       3
    65535::10.0.0.2           mplsTp               12818     12806     central
pp::LSP-PE-1-PE-2             Up                   100       100       3
    65535::10.0.0.2           mplsTp               9565      9190      central
-------------------------------------------------------------------------------
No. of BFD sessions: 2
===============================================================================
*A:PE-1#
```

Execute LSP trace again to check that the LSP has failed over to use the protect path:

```
*A:PE-1# oam lsp-trace static  "LSP-PE-1-PE-2"
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 100 byte packets
1  GlobalId 65535 NodeId 10.0.0.3
   rtt=0.585ms rc=8(DSRtrMatchLabel)
2  GlobalId 65535 NodeId 10.0.0.2
   rtt=1.07ms rc=3(EgressRtr)
*A:PE-1#
```

This shows that packets are now forwarded via the protect path through P-3, which has node ID 10.0.0.3.

Finally bring the LSP back to the working path by bringing port 1/1/1 up, and either waiting for the LSP to revert to the working path or forcing it onto the working path and clearing the revert timer by executing a tools command as follows:

```
*A:PE-1# configure port 1/1/1 no shutdown

*A:PE-1# tools perform router mpls tp-tunnel force "LSP-PE-1-PE-2"

*A:PE-1# tools perform router mpls tp-tunnel clear "LSP-PE-1-PE-2"

*A:PE-1# oam lsp-trace static "LSP-PE-1-PE-2"
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 100 byte packets
1  GlobalId 65535 NodeId 10.0.0.2
   rtt=0.582ms rc=3(EgressRtr)
*A:PE-1#
```

Configure SDPs and Services

Services can be configured to use MPLS-TP LSPs once the LSP configuration is completed. SDPs and services are configured in a similar manner to those using static-labeled pseudowires without MPLS-TP.

Distributed services are configured to use MPLS-TP with the following steps:

- Configure an SDP with signaling off. With signaling off, the SDP far-end may then be configured as an MPLS-TP node-id or an IPv4 address. SDP keep-alive should be disabled.
- Configure the service, including the spoke-sdp using the SDP. To use MPLS-TP, the spoke-sdp must have statically assigned ingress and egress labels, the control-word must be enabled, and it must have an MPLS-TP identifier for the PW (the PW Path ID) configured. This is comprised of two parts, a Source Attachment Individual Identifier (SAII) and a Target Attachment Individual Identifier (TAII), both of which must be configured. Control channel status signaling may also be configured to support PW status signaling on the static MPLS-TP PW.

In this example, an SDP is configured to use the MPLS-TP LSP from PE-1 to PE-2, which will act as a transport for the static MPLS-TP PWs corresponding to Epipe 10 and Epipe 20. Another SDP is configured for the targeted LDP (T-LDP) PW segment corresponding to Epipe 20 between PE-2 and PE-4.

The following CLI shows the SDP between PE-1 and PE-2 and the SDP between PE-2 and PE-4:

```
*A:PE-1# configure service
        sdp 1 mpls create
            signaling off
```

```
            far-end node-id 10.0.0.2 global-id 65535
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit

*A:PE-2# configure service
        sdp 1 mpls create
            signaling off
            far-end node-id 10.0.0.1 global-id 64511
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
        sdp 2 mpls create
            far-end 192.0.2.4
            lsp "LSP-PE-2-PE-4"
            no shutdown
        exit

*A:PE-4# configure service
        sdp 2 mpls create
            far-end 192.0.2.2
            lsp "LSP-PE-4-PE-2"
            no shutdown
        exit
```

Next, configure the services that will use the MPLS-TP LSPs.

The service configuration CLI tree for an Epipe service using MPLS-TP is as follows:

```
— configure
      service
        epipe
          [no] spoke-sdp sdp-id[:vc-id]
              [no] hash-label
              [no] standby-signaling-slave
          [no] spoke-sdp sdp-id[:vc-id] [vc-type {ether|vlan}]
             [create] [vc-switching] [no-endpoint | {endpoint [icb]}]
            egress
              vc-label <out-label>
            ingress
              vc-label <in-label>
—           [no] control-word
            [no] pw-path-id
              agi <agi>
              saii-type2 <global-id:node-id:ac-id>
              taii-type2 <global-id:node-id:ac-id>
              exit
            control-channel-status
              [no] acknowledgment
              [no] refresh-timer <value>
              [no] request-timer <value> retry-timer <value> [timeout-multiplier <value>]
              [no] shutdown
```

See the CLI command descriptions in the user guides for further details of these commands.

The following CLI examples show the Epipe service configuration at PE-1, PE-2, and the VPLS spoke-sdp termination point at PE-4.

Epipe 10 belongs to customer 1, and Epipe 20 belongs to customer 2 in this example.

```
*A:PE-1# configure service
        epipe 10 customer 1 create
            sap 1/1/3 create
            exit
            spoke-sdp 1:10 create
                ingress
                    vc-label 150
                exit
                egress
                    vc-label 151
                exit
                control-word
                pw-path-id
                    saii-type2 64511:10.0.0.1:1
                    taii-type2 65535:10.0.0.2:1
                exit
                control-channel-status
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        epipe 20 customer 2 create
            sap 1/1/4 create
            exit
            spoke-sdp 1:20 create
                ingress
                    vc-label 200
                exit
                egress
                    vc-label 201
                exit
                control-word
                pw-path-id
                    saii-type2 64511:10.0.0.1:2
                    taii-type2 65535:10.0.0.2:2
                exit
                control-channel-status
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

At PE-2, Epipe 10 terminates on a SAP on port 1/1/3, while Epipe 20 is switched between a static MPLS-TP PW segment (spoke-sdp 1:20) and a T-LDP signaled PW segment (spoke-sdp 2:1) for backhaul to the remote PE-4 containing the VPLS service.

```
*A:PE-2# configure service
        epipe 10 customer 1 create
            sap 1/1/3 create
            exit
            spoke-sdp 1:10 create
                ingress
                    vc-label 151
                exit
                egress
                    vc-label 150
                exit
                control-word
                pw-path-id
                    saii-type2 65535:10.0.0.2:1
                    taii-type2 64511:10.0.0.1:1
                exit
                control-channel-status
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        epipe 20 customer 2 vc-switching create
            spoke-sdp 1:20 create
                ingress
                    vc-label 201
                exit
                egress
                    vc-label 200
                exit
                control-word
                pw-path-id
                    saii-type2 65535:10.0.0.2:2
                    taii-type2 64511:10.0.0.1:2
                exit
                control-channel-status
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 2:1 create
                control-word
                no shutdown
            exit
            no shutdown
        exit
```

At PE-4, the T-LDP signaled PW segment for Epipe 20 is terminated on a VPLS service:

```
*A:PE-4# configure service
```

```
        vpls 1 customer 2 create
            sap 1/1/1 create
            exit
            spoke-sdp 2:1 create
                control-word
                no shutdown
            exit
            no shutdown
        exit
```

Epipe 10 uses a static MPLS-TP PW from end to end, which can be tested using the virtual circuit connectivity verification vccv-ping command at PE-1, as follows:

```
*A:PE-1# oam vccv-ping static 1:10
VCCV-PING 1:10 84 bytes MPLS payload
Seq=1, send from intf int-PE-1-PE-2
       send from lsp LSP-PE-1-PE-2
       reply via Control Channel
       src id tlv received: GlobalId 65535 NodeId 10.0.0.2
       cv-data-len=44 rtt=0.597ms rc=3 (EgressRtr)

---- VCCV PING 1:10 Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 0.597ms, avg = 0.597ms, max = 0.597ms, stddev = 0.000ms
*A:PE-1#
```

The operation of control channel status signaling can also be tested for this Epipe, as follows:

Shutdown the port the SAP on PE-2 is using:

```
*A:PE-2# configure port 1/1/3 shutdown
```

The PW peer status bits for the spoke-sdp for Epipe 10, signaled using control channel status signaling, can be displayed at node PE-1 using the following command (some of the show command output has been removed for brevity). The peer PW status bits are shown in bold in the output below.

```
*A:PE-1# show service id 10 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 1:10  -(10.0.0.2:65535)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id            : 1:10                      Type              : Spoke
Spoke Descr     : (Not Specified)
VC Type           : Ether                     VC Tag            : n/a
Admin Path MTU    : 0                         Oper Path MTU     : 1556
Delivery          : MPLS
Far End           : 10.0.0.2:65535
Tunnel Far End    : n/a                       LSP Types         : MPLSTP
Hash Label        : Disabled                  Hash Lbl Sig Cap  : Disabled
```

```
Oper Hash Label    : Disabled
Entropy Label      : Disabled

Admin State        : Up                        Oper State        : Up
MinReqd SdpOperMTU : 1552
Acct. Pol          : None                      Collect Stats     : Disabled
Ingress Label      : 150                       Egress Label      : 151
Ingr Mac Fltr-Id   : n/a                       Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id    : n/a                       Egr IP Fltr-Id    : n/a
Ingr IPv6 Fltr-Id  : n/a                       Egr IPv6 Fltr-Id  : n/a
Admin ControlWord  : Preferred                 Oper ControlWord  : True
Admin BW(Kbps)     : 0                         Oper BW(Kbps)     : 0
BFD Template       : None
BFD-Enabled        : no                        BFD-Encap         : ipv4
Last Status Change : 03/22/2016 14:30:34       Signaling         : None
Last Mgmt Change   : 03/22/2016 14:30:34
Endpoint           : N/A                       Precedence        : 4
PW Status Sig      : Enabled
Force Vlan-Vc      : Disabled                  Force Qinq-Vc     : Disabled
Class Fwding State : Down
Flags              : None
Local Pw Bits      : None
Peer Pw Bits       : lacIngressFault lacEgressFault
Peer Fault Ip      : None
Peer Vccv CV Bits  : None
Peer Vccv CC Bits  : None
---snip---
```

Epipe 20 uses a static MPLS-TP PW from PE-1 to PE-2, identified by a static PW forwarding equivalence class (FEC), and a T-LDP segment with FEC128 from PE-2 to PE-4. Therefore the target FEC used for a vccv-ping command from PE-1 to PE-4 is different from the local FEC for the PW at PE-1. VCCV-trace provides a useful tool to test the resulting multi-segment PW (MS-PW). The same associated channel type must be used for both segments. This is the IPv4 channel.

```
*A:PE-1# oam vccv-trace static 1:20 assoc-channel ipv4 detail
VCCV-TRACE 1:20  with 116 bytes of MPLS payload
1  192.0.2.2 GlobalId 65535 NodeId 10.0.0.2
   rtt=0.599ms rc=8(DSRtrMatchLabel)
   Next segment: VcId=1 VcType=Ether Source=192.0.2.2 Remote=192.0.2.4
2  192.0.2.4  rtt=1.06ms rc=3(EgressRtr)

*A:PE-1#
```

The system supports the interworking of control channel status on a static MPLS-TP PW segment with T-LDP-signaled PW status on a T-LDP PW segment. This can be tested as follows.

Shut down the port the spoke SDP on PE-4 is using:

```
*A:PE-4# configure port 1/1/2 shutdown
```

The PW peer status bits for the spoke-sdp for Epipe 20 can then be displayed at node PE-1 using the following command (some of the show command output has been removed for brevity). The peer PW status bits are shown in bold in the output below.

```
*A:PE-1# show service id 20 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 1:20  -(10.0.0.2:65535)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id            : 1:20                      Type              : Spoke
Spoke Descr     : (Not Specified)
VC Type           : Ether                     VC Tag            : n/a
Admin Path MTU    : 0                         Oper Path MTU     : 1556
Delivery          : MPLS
Far End           : 10.0.0.2:65535
Tunnel Far End    : n/a                       LSP Types         : MPLSTP
Hash Label        : Disabled                  Hash Lbl Sig Cap  : Disabled
Oper Hash Label   : Disabled
Entropy Label     : Disabled

Admin State       : Up                        Oper State        : Up
MinReqd SdpOperMTU : 1552
Acct. Pol         : None                      Collect Stats     : Disabled
Ingress Label     : 200                       Egress Label      : 201
Ingr Mac Fltr-Id  : n/a                       Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id   : n/a                       Egr IP Fltr-Id    : n/a
Ingr IPv6 Fltr-Id : n/a                       Egr IPv6 Fltr-Id  : n/a
Admin ControlWord : Preferred                 Oper ControlWord  : True
Admin BW(Kbps)    : 0                         Oper BW(Kbps)     : 0
BFD Template      : None
BFD-Enabled       : no                        BFD-Encap         : ipv4
Last Status Change : 03/22/2016 14:30:34      Signaling         : None
Last Mgmt Change  : 03/22/2016 14:30:34
Endpoint          : N/A                       Precedence        : 4
PW Status Sig     : Enabled
Force Vlan-Vc     : Disabled                  Force Qinq-Vc     : Disabled
Class Fwding State : Down
Flags             : None
Local Pw Bits     : None
Peer Pw Bits      : psnIngressFault psnEgressFault
Peer Fault Ip     : None
Peer Vccv CV Bits : None
Peer Vccv CC Bits : None
---snip---
```

Conclusion

Release 11.0.R4 of SR OS introduced extensive MPLS Transport Profile (MPLS-TP) capabilities. MPLS-TP is intended to allow MPLS to be operated in a manner similar to existing transport technologies, with proactive in-band and on-demand operations and maintenance (OAM), and protection mechanisms that do not rely on a control plane to operate. The 7x50 can operate both as an LER and LSR for MPLS-TP LSPs, and as a T-PE and S-PE for PWs with MPLS-TP OAM. The 7x50 can therefore act as a node within an MPLS-TP network, or as a gateway between MPLS-TP and IP/MPLS domains.

This example has illustrated a simple configuration, demonstrating the role of the 7x50 as an LER and LSR for MPLS-TP LSPs, and how its already extensive multi-service capabilities can be extended over an MPLS-TP network and between MPLS-TP and IP/MPLS networks.

Multicast Label Distribution Protocol

This chapter provides information about multicast label distribution protocol.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for release 13.0.R6. The CLI in this edition corresponds to release 14.0.R1. Chassis mode C or later must be used.

In this chapter, the emphasis is on IPv4. However, multicast label distribution protocol (mLDP) is also supported on IPv6 interfaces.

Overview

Multicast label distribution protocol provides extensions to LDP for the setup of point-to-multipoint (P2MP) label switched paths (LSPs) and multipoint-to-multipoint (MP2MP) LSPs in MPLS networks.

The protocol is described in RFC 6388 - Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths.

Multicast LSPs can be applied for IP multicast or support for multicast in BGP/MPLS Layer 3 virtual private networks (L3 VPNs).

Compared to RSVP P2MP LSPs, mLDP P2MP LSPs are easier to configure and the setup direction is different. While the RSVP P2MP LSPs are set up from the root node toward the leaf nodes, mLDP P2MP LSPs are set up from the leaf nodes toward the root node.

P2MP Terminology

The following terminology will be used.

Table 15 **Terminology**

Node	Description
Ingress / Root	P2MP LSPs have just one ingress (root) node. The root node receives IP multicast traffic and maps the traffic to a P2MP LSP (push). The node might perform MPLS multicast replication.
Egress / Leaf	P2MP LSPs have multiple egress (leaf) nodes. A leaf node removes data packets from a P2MP LSP (pop) for further processing. The node might perform IP multicast replication.
Transit	A transit label switching router (LSR) can reach the root node via a directly connected upstream LSR. A transit LSR also has one or more directly connected downstream LSRs. The LSR swaps the MPLS label and might perform MPLS multicast replication.
Branch	A branch LSR is a transit LSR that has several directly connected LSRs. The LSR swaps the MPLS label and performs MPLS multicast replication.
Bud	A bud node is an egress node, but also a transit node. The node has directly connected receivers and also one or more directly connected downstream LSRs.

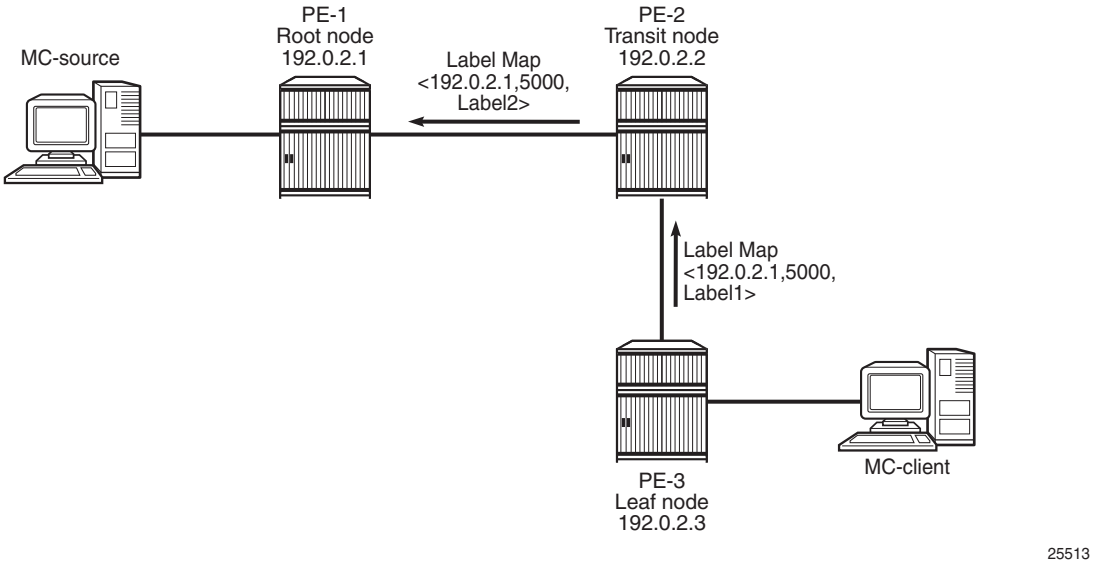
Setup of mLDP P2MP LSP

The setup of the P2MP LSP in the control plane is as follows.

1. The leaf node initiates a tree setup according to what is configured. Mandatory parameters are the IP address of the root and an opaque value. The leaf node sends an LDP label map message to its upstream hop toward the root node of the tree.
2. Each transit node receives the LDP label map message and sends another LDP label map message to its upstream hop toward the root node of the tree. Each label can be different.
3. The root node receives the LDP label map message.

The label map message contains the root node address, an opaque value, and a label. In the example in Figure 1, the root node address is 192.0.2.1 and the opaque value is 5000.

Figure 171 **Setup of mLDP P2MP LSP**



After the LDP label map messages are sent in the control plane, the nodes program pop, swap, or push entries for the corresponding labels in the data plane.

1. The leaf node programs a pop entry for the label sent upstream.
2. The transit node programs a swap entry for the label it sent upstream with the next-hop address and the label it received from the downstream node.
3. The root node programs a push entry and a next-hop address for the label it received from the downstream node.

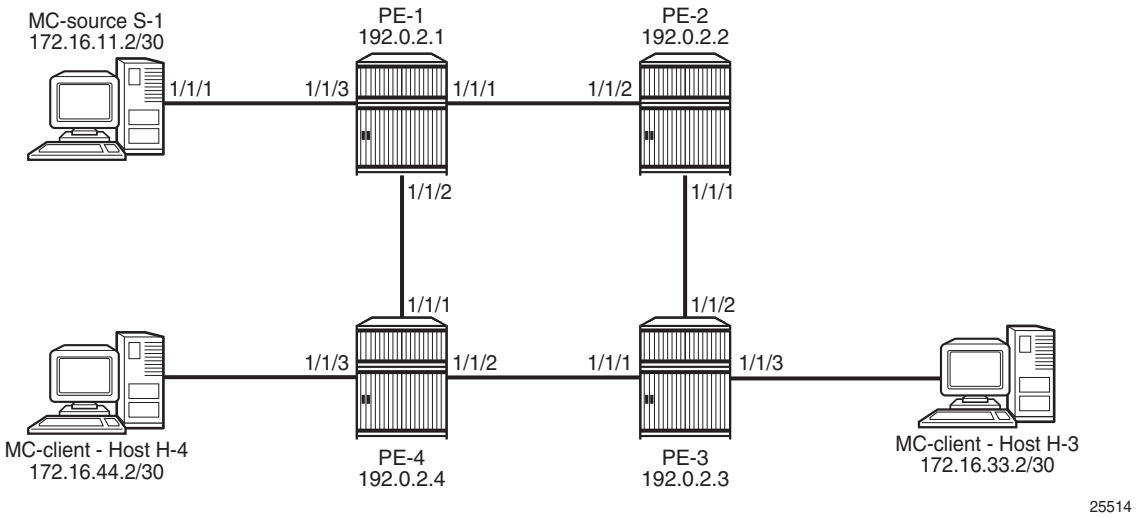
If multiple equal cost multi-path (ECMP) paths exist between two adjacent nodes, the upstream node of the multicast receiver must program all the entries in the forwarding plane. Only one entry must be active based on the ECMP hashing algorithm.

Configuration

The example setup shown in Figure 172 will be used. The multicast source S-1 is connected to root node PE-1. PE-2 or PE-4 will be the transit node for traffic destined to PE-3. There are two leaf nodes: PE-3 and PE-4. Multicast client H-3 is connected to PE-3, while multicast client H-4 is connected to PE-4.

Under normal circumstances, PE-2 is the transit node for traffic toward PE-3 and PE-4 is an egress node. If PE-4 is the transit node for traffic toward PE-3, while it also has a directly connected receiver, PE-4 is a bud node.

Figure 172 **Test Topology**



Configure LDP P2MP LSP

Initial Configuration

The PEs should have the following initial configuration:

- Chassis mode must be C or higher
- Cards, MDAs, ports
- Router interfaces
- IGP (OSPF or IS-IS)

As an example, the router interfaces and OSPF configuration on PE-1 are as follows:

```
*A:PE-1# configure router
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/1
        exit
        interface "int-PE-1-PE-4"
            address 192.168.14.1/30
            port 1/1/2
```

```
        exit
        interface "int-PE-1-S-1"
            address 172.16.11.1/30
            port 1/1/3
        exit
        interface "system"
            address 192.0.2.1/32
        exit
    exit

*A:PE-1# configure router
        ospf
            area 0.0.0.0
                interface "system"
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-1-PE-4"
                    interface-type point-to-point
                exit
                interface "int-PE-1-S-1"
                    interface-type point-to-point
                exit
            exit
        exit
    exit
```

Enabling mLDP

When LDP is enabled, mLDP is enabled by default

- For releases earlier than 13.0.R1, mLDP had to be enabled with the command:

```
configure router ldp interface-parameters interface <ip-int-name> multicast-traffic
enable
```

- For release 13.0.R1 or later, the command is as follows:

```
configure router ldp interface-parameters interface <ip-int-name> ipv4 fec-type-
capability p2mp-ipv4 enable
```

Enable LDP (including mLDP) on PE-1, as follows:

```
*A:PE-1# configure router
        ldp
            interface-parameters
                interface "int-PE-1-PE-2"
                exit
                interface "int-PE-1-PE-4"
                exit
            exit
```

Verify that mLDP is enabled (P2MP: Enabled), as follows:

```
*A:PE-1# show router ldp status
===============================================================================
LDP Status for IPv4 LSR ID 192.0.2.1
               IPv6 LSR ID ::
===============================================================================
Created at         : 03/25/16 14:02:42
Last Change        : 03/25/16 14:02:42
Admin State        : Up
IPv4 Oper State    : Up                    IPv6 Oper State      : Down
IPv4 Up Time       : 0d 00:01:47           IPv6 Down Time       : 0d 00:01:47
IPv4 Oper Down Rea*: n/a                   IPv6 Oper Down Reason: systemIpDown
IPv4 Oper Down Eve*: 0                     IPv6 Oper Down Events: 0
Tunn Down Damp Time: 3 sec
Label Withdraw Del*: 0 sec                 Implicit Null Label  : Disabled
Short. TTL Local   : Enabled               Short. TTL Transit   : Enabled
Import Policies    : None                  Export Policies      : None
Tunl Exp Policies  : None
FRR                : Disabled              Mcast Upstream FRR   : Disabled
MP MBB Time        : 3
Aggregate Prefix   : False                 Agg Prefix Policies  : None
Class Forwarding   : Disabled
Legacy LSR Interop : False
-------------------------------------------------------------------------------
Capabilities
-------------------------------------------------------------------------------
Dynamic            : Enabled               P2MP                 : Enabled
IPv4 Prefix Fec    : Enabled               IPv6 Prefix Fec      : Enabled
Service Fec128     : Enabled               Service Fec129       : Enabled
MP MBB             : Enabled               Overload             : Enabled
Unrecognized Notif*: Enabled
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

Verify that mLDP is enabled on the interface "int-PE-1-PE-2" (IPv4 P2MP Fec Cap), as follows:

```
*A:PE-1# show router ldp interface "int-PE-1-PE-2" detail
===============================================================================
LDP Interfaces
===============================================================================
===============================================================================
Interface "int-PE-1-PE-2"
===============================================================================
-------------------------------------------------------------------------------
BASE
-------------------------------------------------------------------------------
Admin State        : Up                    Oper State       : Up
BFD Status         : Disabled
-------------------------------------------------------------------------------
IPv4
-------------------------------------------------------------------------------
IPv4 Admin State   : Up                    IPv4 Oper State  : Up
Last Oper Chg      : 0d 00:01:19
Hold Time          : 15                    Hello Factor     : 3
Oper Hold Time     : 15
```

```
Keepalive Timeout  : 30                  Keepalive Factor : 3
Transport Addr     : System              Last Modified    : 03/25/16 14:02:42
Active Adjacencies : 1
Local LSR Type     : System
Local LSR          : None
IPv4 Pfx Fec Cap   : Enabled             IPv6 Pfx Fec Cap : Enabled
IPv4 P2MP Fec Cap  : Enabled             IPv6 P2MP Fec Cap: Enabled
===============================================================================
No. of Interfaces: 1
===============================================================================
*A:PE-1#
```

Disable mLDP on interface "int-PE-1-PE-2" and verify the LDP status again, as follows:

```
*A:PE-1# configure router
        ldp
            interface-parameters
                interface "int-PE-1-PE-2" dual-stack
                    ipv4
                        fec-type-capability
                            p2mp-ipv4 disable
                        exit all

*A:PE-1# show router ldp status
===============================================================================
LDP Status for IPv4 LSR ID 192.0.2.1
               IPv6 LSR ID ::
===============================================================================
Created at         : 03/25/16 14:02:42
Last Change        : 03/25/16 14:02:42
Admin State        : Up
IPv4 Oper State    : Up                   IPv6 Oper State      : Down
IPv4 Up Time       : 0d 00:01:47          IPv6 Down Time       : 0d 00:01:47
IPv4 Oper Down Rea*: n/a                  IPv6 Oper Down Reason: systemIpDown
IPv4 Oper Down Eve*: 0                    IPv6 Oper Down Events: 0
Tunn Down Damp Time: 3 sec
Label Withdraw Del*: 0 sec                Implicit Null Label  : Disabled
Short. TTL Local   : Enabled              Short. TTL Transit   : Enabled
Import Policies    : None                 Export Policies      : None
Tunl Exp Policies  : None
FRR                : Disabled             Mcast Upstream FRR   : Disabled
MP MBB Time        : 3
Aggregate Prefix   : False                Agg Prefix Policies  : None
Class Forwarding   : Disabled
Legacy LSR Interop : False
-------------------------------------------------------------------------------
Capabilities
-------------------------------------------------------------------------------
Dynamic            : Enabled              P2MP                 : Enabled
IPv4 Prefix Fec    : Enabled              IPv6 Prefix Fec      : Enabled
Service Fec128     : Enabled              Service Fec129       : Enabled
MP MBB             : Enabled              Overload             : Enabled
Unrecognized Notif*: Enabled
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

P2MP LDP is still enabled on the router, but it is disabled on interface "int-PE-1-PE-2", which can be verified as follows:

```
*A:PE-1# show router ldp interface "int-PE-1-PE-2" detail
===============================================================================
LDP Interfaces
===============================================================================
===============================================================================
Interface "int-PE-1-PE-2"
===============================================================================
-------------------------------------------------------------------------------
BASE
-------------------------------------------------------------------------------
Admin State        : Up                  Oper State       : Up
BFD Status         : Disabled
-------------------------------------------------------------------------------
IPv4
-------------------------------------------------------------------------------
IPv4 Admin State   : Up                  IPv4 Oper State  : Up
Last Oper Chg      : 0d 00:01:53
Hold Time          : 15                  Hello Factor     : 3
Oper Hold Time     : 15
Keepalive Timeout  : 30                  Keepalive Factor : 3
Transport Addr     : System              Last Modified    : 03/25/16 14:04:15
Active Adjacencies : 1
Local LSR Type     : System
Local LSR          : None
IPv4 Pfx Fec Cap   : Enabled             IPv6 Pfx Fec Cap : Enabled
IPv4 P2MP Fec Cap  : Disabled            IPv6 P2MP Fec Cap: Enabled
===============================================================================
No. of Interfaces: 1
===============================================================================
*A:PE-1#
```

P2MP multicast forwarding can be disabled per LDP interface. Disabling P2MP multicast forwarding will not prevent LDP from exchanging P2MP FEC elements on that interface in the control plane. In the data plane, the forwarding plane is not programmed with the next hop on the outgoing interface that is P2MP disabled.

Configure Tunnel Interface on the Root and Leaf Nodes

Multicast LDP can be applied in different scenarios. In the following example, a tunnel interface is created on the root and leaf nodes. Other examples are multicast virtual private network (MVPN) with mLDP and dynamic PIM-mLDP mapping. In several ACG chapters on MVPN, mLDP is chosen; for example, in *Multicast VPN: Use of Wildcard Selective PMSI*.

A tunnel interface needs to be created on the root node, as follows:

```
*A:PE-1# configure router tunnel-interface ldp-p2mp 5000 sender 192.0.2.1 root-node
```

In this example, the tunnel interface gets interface index 73728, as follows:

```
*A:PE-1# show router tunnel-interface
===============================================================================
 P2MP-RSVP P2MP-LDP Tunnel-Interfaces
===============================================================================
LSP/LDP           Type        SenderAddr          IfIndex             RootNode
-------------------------------------------------------------------------------
5000              ldp         192.0.2.1           73728               Yes
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
*A:PE-1#
```

A similar command will be launched on the leaf nodes, but without the keyword **root-node**, as follows:

```
*A:PE-3# configure router tunnel-interface ldp-p2mp 5000 sender 192.0.2.1
*A:PE-4# configure router tunnel-interface ldp-p2mp 5000 sender 192.0.2.1
*A:PE-3# show router tunnel-interface
===============================================================================
 P2MP-RSVP P2MP-LDP Tunnel-Interfaces
===============================================================================
LSP/LDP           Type        SenderAddr          IfIndex             RootNode
-------------------------------------------------------------------------------
5000              ldp         192.0.2.1           73728               No
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
*A:PE-3#
```

A P2MP LSP ping can be sent to verify the P2MP LSP. The options for P2MP LSP ping are as follows:

```
*A:PE-1# oam p2mp-lsp-ping
  - p2mp-lsp-ping {<lsp-name> [p2mp-instance <instance-name>
                  [s2l-dest-address <ipv4-address> [... up to 5]]]
                  [ttl <label-ttl>]}
  - p2mp-lsp-ping {ldp <p2mp-identifier> [sender-addr <ipv4-address>]
                  [leaf-addr <ipv4-address> [... up to 5]]}
  - p2mp-lsp-ping {ldp-ssm source <ip-address> group <ip-address>
                  [router <router-instance>|service-name <service-name>]
                  [sender-addr <ipv4-address>] [leaf-addr <ipv4-address>
                  [... up to 5]]}
  - options common to all p2mp-lsp-ping cases:  [fc <fc-name> [profile {in|out}]]
                     [size <octets>] [timeout <timeout>] [detail]
 <lsp-name>           : [64 chars max]
 <instance-name>      : [32 chars max]
 <ipv4-address>       : a.b.c.d
 <in|out>             : in|out - Default: out
 <fc-name>            : be|l2|af|l1|h2|ef|h1|nc - Default: be
 <octets>             : [1..9198] - Default: 1
 <label-ttl>          : [1..255] - Default: 255
 <timeout>            : [1..120] seconds - Default: 10
 <detail>             : keyword - displays detailed information
 <p2mp-identifier>    : [1..4294967295]
```

```
 <ldp-ssm>            : keyword
 <ip-address>         : ipv4-address    - a.b.c.d
                        ipv6-address    - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                          x:x:x:x:x:x:d.d.d.d
                                          x - [0..FFFF]H
                                          d - [0..255]D
 <router-instance>    : <router-name>|<service-id>
                        router-name     - "Base"|"management"  Default - Base
                        service-id      - [1..2147483647]
 <service-name>       : [64 chars max]
*A:PE-1#
```

Verify the P2MP LSP with the following ping command:

```
*A:PE-1# oam p2mp-lsp-ping ldp 5000
P2MP identifier 5000: 88 bytes MPLS payload
Total Leafs responded = 2
          round-trip min/avg/max  = 0.898 / 1.01 / 1.12 ms
Responses based on return code:
        EgressRtr(3)=2
*A:PE-1#
```

Both leaf nodes have sent a reply. The return code 3 indicates that the replying router is an egress for the forwarding equivalence class (FEC).

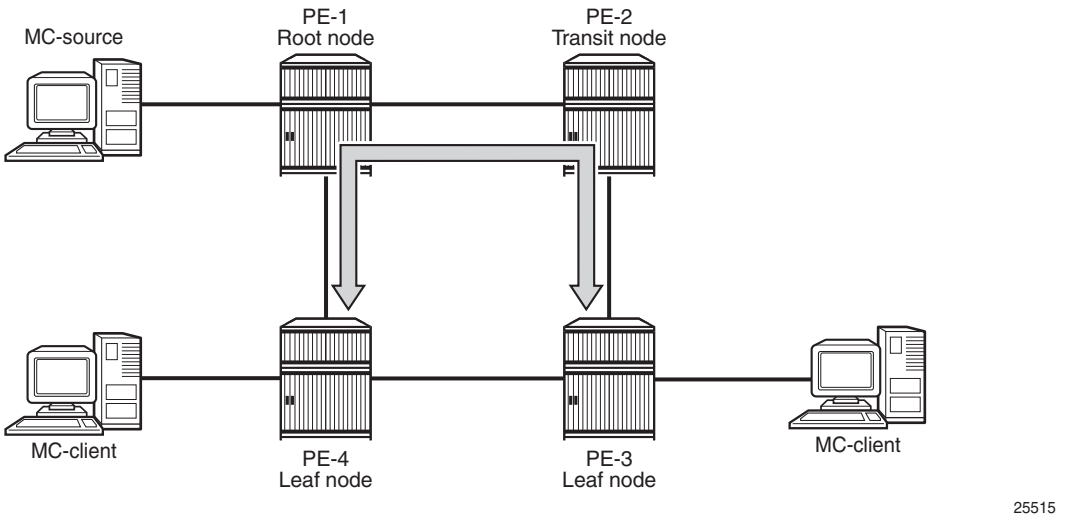
For a detailed output per leaf, use the following command:

```
*A:PE-1# oam p2mp-lsp-ping ldp 5000 detail
P2MP identifier 5000: 88 bytes MPLS payload
===============================================================================
Leaf Information
===============================================================================
From             RTT                  Return Code
-------------------------------------------------------------------------------
192.0.2.4        =0.889ms             EgressRtr(3)
192.0.2.3        =1.36ms              EgressRtr(3)
===============================================================================
Total Leafs responded = 2
          round-trip min/avg/max  = 0.889 / 1.12 / 1.36 ms
Responses based on return code:
        EgressRtr(3)=2
*A:PE-1#
```

Verify LDP P2MP Bindings

The example LDP P2MP LSP is shown in Figure 173. In this case, PE-4 is only an egress node and not a bud node.

Figure 173 **LDP P2MP LSP**



Verify the LDP P2MP bindings on the leaf node PE-4, as follows.

The leaf node programs a pop entry for the label sent upstream.

```
*A:PE-4# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
5000                                        73728
192.0.2.1                                   Pop             262139      --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
---snip---
*A:PE-4# show router ldp bindings p2mp detail
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
```

```
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 5000
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.1:0
Ing Lbl        : 262139U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
===============================================================================
---snip---
```

PE-4 is only an egress node and not a transit node. There is no next hop.

Verify the LDP P2MP bindings on the leaf node PE-3, as follows:

```
*A:PE-3# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
5000                                        73728
192.0.2.1                                   Pop             262139      --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
---snip---
```

Because PE-3 is an egress node, there is no next hop. The traffic toward PE-3 is sent via transit PE-2 and not via PE-4, as can be verified as follows:

```
*A:PE-3# show router ldp bindings p2mp detail
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
```

```
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 5000
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.2:0
Ing Lbl        : 262139U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
===============================================================================
```

PE-2 has programmed a swap entry for the label it sent to its upstream node PE-1 with the next-hop address and the label it received from the downstream node, as follows:

```
*A:PE-2# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
-------------------------------------------------------------------------------
5000                                     Unknw
192.0.2.1                                Swap            262139    262139
192.168.23.2                             1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
---snip---

*A:PE-2# show router ldp bindings p2mp detail
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 5000
```

```
Root-Addr      : 192.0.2.1
------------------------------------------------------------------------------
Peer           : 192.0.2.1:0
Ing Lbl        : 262139U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 5000
Root-Addr      : 192.0.2.1
------------------------------------------------------------------------------
Peer           : 192.0.2.3:0
Ing Lbl        :   --
Egr Lbl        : 262139
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.23.2
Egr. Flags     : None               Ing. Flags : None
Egr If Name    : int-PE-2-PE-3
Metric         : 1                  Mtu        : 1564
==============================================================================
No. of Generic IPv4 P2MP Bindings: 2
==============================================================================
---snip---
```

The egress next hop is PE-3.

On the root node PE-1, there is MPLS multicast replication. One traffic stream goes via transit node PE-2 toward leaf node PE-3 and the other traffic stream goes directly toward leaf node PE-4. There are two push entries with the corresponding next-hop address, as follows:

```
*A:PE-1# show router ldp bindings active p2mp
==============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
==============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
==============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
------------------------------------------------------------------------------
5000                                     73728
192.0.2.1                                Push              --      262139
192.168.12.2                             1/1/1

5000                                     73728
192.0.2.1                                Push              --      262139
192.168.14.2                             1/1/2

------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
```

```
===============================================================================
---snip---
*A:PE-1# show router ldp bindings p2mp detail
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 5000
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.2:0
Ing Lbl        :   --
Egr Lbl        : 262139
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.12.2
Egr. Flags     : None               Ing. Flags : None
Egr If Name    : int-PE-1-PE-2
Metric         : 1                  Mtu        : 1564
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 5000
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.4:0
Ing Lbl        :   --
Egr Lbl        : 262139
Egr Int/LspId  : 1/1/2
EgrNextHop     : 192.168.14.2
Egr. Flags     : None               Ing. Flags : None
Egr If Name    : int-PE-1-PE-4
Metric         : 1                  Mtu        : 1564
===============================================================================
No. of Generic IPv4 P2MP Bindings: 2
===============================================================================
---snip---
```

Tools Command

The following tools command can be launched on any of the nodes in the P2MP LSP.

For the ingress node PE-1, where one branch goes to transit node PE-2 (192.0.2.2) and another branch to leaf node PE-4 (192.0.2.4), the output is as follows:

```
*A:PE-1# tools dump router ldp fec p2mp-id 5000 root 192.0.2.1
P2MP: root: 192.0.2.1, T: 1, L: 4, TunnelId: 5000
  Create Time   : 03/25/16 14:05:29.747 (elapsed: 0d 00:01:44)
  Last Mod. Time: 03/25/16 14:05:39.017 (elapsed: 0d 00:01:35)
  FEC Flags     : Push Mttm
  TunlIfId      : 73728   (OperState : up)
```

```
  LSP ID        : 65540
  LSP ID Acct.  : 4
  isIngressMttm : Yes        HasLeaf     : No
  isIngrItermdte: No
  CanProgIngress: Yes
  InPhopFrr     : No
  isStitchedUpr : No
  RslvdPhop(p)  : 0.0.0.0:0 (seqNum 0)
  RslvdPhop(b)  : 0.0.0.0:0 (seqNum 0)
  pri Upstream  : None
  mbb Upstream  : None
  bkp Upstream  : None
  AdvInLabel(p) : 0
  AdvInLabel(b) : 0
  Num Resolved   Nhops  : 2
  Num MBB Req.   Nhops  : 0
  Num Programmed Nhops  : 2
    Programmed Nhop[01] : 192.0.2.2:0, OutLabel 262139
    Programmed Nhop[02] : 192.0.2.4:0, OutLabel 262139
  Metric        : 1          Mtu         : 1564
  Num of Peers : 2
  FEC Peer: 192.0.2.2:0
    Peer Flags: MPush  (0x800)
    ModTime   : 03/25/16 14:05:37.598 (elapsed.: 0d 00:01:36)
    ->Num Egress Labels:
      -> (Label: 262139    Status: UsePush)
      Flow Label Tx: no, Rx: no
    <-Num Ingress Labels:
      None
    <Resolved as Next Hop>
     Next Hop Info :
       metric: 1  mtu: 1564
       [01]: Next Hop: 192.168.12.2    Interface: 2
             owner   : 192.0.2.2:0  egress label: 262139
  FEC Peer: 192.0.2.4:0
    Peer Flags: MPush  (0x800)
    ModTime   : 03/25/16 14:05:39.019 (elapsed.: 0d 00:01:35)
    ->Num Egress Labels:
      -> (Label: 262139    Status: UsePush)
      Flow Label Tx: no, Rx: no
    <-Num Ingress Labels:
      None
    <Resolved as Next Hop>
     Next Hop Info :
       metric: 1  mtu: 1564
       [01]: Next Hop: 192.168.14.2    Interface: 3
             owner   : 192.0.2.4:0  egress label: 262139
*A:PE-1#
```

The labels that are pushed at PE-1 are 262139 for traffic to PE-2 and 262139 for traffic to PE-4.

On transit node PE-2, the incoming label 262139 is swapped to outgoing label 262139 toward PE-3, as follows:

```
*A:PE-2# tools dump router ldp fec p2mp-id 5000 root 192.0.2.1
P2MP: root: 192.0.2.1, T: 1, L: 4, TunnelId: 5000
```

```
  Create Time   : 03/25/16 14:05:27.209 (elapsed: 0d 00:01:36)
  Last Mod. Time: 03/25/16 14:05:27.209 (elapsed: 0d 00:01:36)
  FEC Flags     : Swap
  TunlIfId      : 0       (OperState : dn)
  LSP ID        : 0
  LSP ID Acct.  : 0
  isIngressMttm : No         HasLeaf     : No
  isIngrItermdte: No
  CanProgIngress: No
  InPhopFrr     : No
  isStitchedUpr : No
  RslvdPhop(p)  : 192.0.2.1:0 (seqNum 2)
  RslvdPhop(b)  : 0.0.0.0:0 (seqNum 0)
  pri Upstream  : 192.0.2.1:0, AdvLabel 262139
  mbb Upstream  : None
  bkp Upstream  : None
  AdvInLabel(p) : 262139
  AdvInLabel(b) : 0
  PrgInLabel(p) : 1
  Num Resolved   Nhops  : 1
  Num MBB Req.   Nhops  : 0
  Num Programmed Nhops  : 1
    Programmed Nhop[01] : 192.0.2.3:0, OutLabel 262139
  Metric        : 1          Mtu         : 1564
  Num of Peers : 2
  FEC Peer: 192.0.2.1:0
    Peer Flags: none  (0x0)
    ModTime   : 03/25/16 14:05:27.210 (elapsed.: 0d 00:01:36)
    ->Num Egress Labels:
      None
    <-Num Ingress Labels:
      <- (Label: 262139    Status: UseSwap)
      Rej Status: OK
      Flow Label Tx: no, Rx: no
      Flow Label Tx Sent: no, Rx Sent: no
    <Resolved as CUR Upstream>
  FEC Peer: 192.0.2.3:0
    Peer Flags: MSwap  (0x1000)
    ModTime   : 03/25/16 14:05:27.210 (elapsed.: 0d 00:01:36)
    ->Num Egress Labels:
      -> (Label: 262139    Status: UseSwap)
      Flow Label Tx: no, Rx: no
    <-Num Ingress Labels:
      None
    <Resolved as Next Hop>
     Next Hop Info :
       metric: 1  mtu: 1564
       [01]: Next Hop: 192.168.23.2    Interface: 3
             owner   : 192.0.2.3:0  egress label: 262139
*A:PE-2#
```

On leaf node PE-3, the incoming label from PE-2 (262139) is popped. There is no next hop.

```
*A:PE-3# tools dump router ldp fec p2mp-id 5000 root 192.0.2.1
P2MP: root: 192.0.2.1, T: 1, L: 4, TunnelId: 5000
  Create Time   : 03/25/16 14:05:30.157 (elapsed: 0d 00:01:36)
  Last Mod. Time: 03/25/16 14:05:30.157 (elapsed: 0d 00:01:36)
```

```
  FEC Flags     : Pop Mttm
  TunlIfId      : 73728   (OperState : up)
  LSP ID        : 0
  LSP ID Acct.  : 0
  isIngressMttm : No         HasLeaf     : Yes
  isIngrItermdte: No
  CanProgIngress: No
  InPhopFrr     : No
  isStitchedUpr : No
  RslvdPhop(p)  : 192.0.2.2:0 (seqNum 2)
  RslvdPhop(b)  : 0.0.0.0:0 (seqNum 0)
  pri Upstream  : 192.0.2.2:0, AdvLabel 262139
  mbb Upstream  : None
  bkp Upstream  : None
  AdvInLabel(p) : 262139
  AdvInLabel(b) : 0
  PrgInLabel(p) : 1
  Num Resolved   Nhops  : 1
  Num MBB Req.   Nhops  : 0
  Num Programmed Nhops  : 1
    Programmed Nhop[01] : 0.0.0.0:0, OutLabel 0 (Leaf)
  Metric        : 0          Mtu         : 0
  Num of Peers : 1
  FEC Peer: 192.0.2.2:0
    Peer Flags: none  (0x0)
    ModTime   : 03/25/16 14:05:30.158 (elapsed.: 0d 00:01:36)
    ->Num Egress Labels:
      None
    <-Num Ingress Labels:
      <- (Label: 262139    Status: UsePop)
      Rej Status: OK
      Flow Label Tx: no, Rx: no
      Flow Label Tx Sent: no, Rx Sent: no
    <Resolved as CUR Upstream>
*A:PE-3#
The output for leaf node PE-4 is similar.
```

Debug Commands

Debugging was enabled on the nodes when LDP was configured. To distinguish which messages are being logged for a debug command, the debug configuration is different for the nodes, as follows:

```
*A:PE-2# debug router ldp peer 192.0.2.1 event bindings
*A:PE-3# debug router ldp peer 192.0.2.2 packet label detail
*A:PE-4# debug router ldp peer 192.0.2.1 packet init detail
```

The following messages were logged. The first two messages correspond to the label mapping messages to establish LDP bindings. The first message is sent from transit node PE-2 to root node PE-1, while the second is sent by the leaf node PE-3 to the transit node PE-2.

```
*A:PE-2# debug router ldp peer 192.0.2.1 event bindings
```

```
---snip---
7 2016/03/25 14:05:27.21 UTC MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 262139 for P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId:
 5000
 to peer 192.0.2.1:0."

*A:PE-3# debug router ldp peer 192.0.2.2 packet label detail
---snip---
7 2016/03/25 14:05:30.15 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 30) to 192.0.2.2:0
Protocol version = 1
Label 262139 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"
```

The following message shows the negotiation of capabilities when LDP bindings are initialized.

```
*A:PE-4# debug router ldp peer 192.0.2.1 packet init detail
---snip---
1 2016/03/25 14:02:54.54 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Initialization packet (msgId 2) to 192.0.2.1:0
Protocol version = 1
Keepalive Timeout = 30  Label Advertisement = downStreamUnsolicited
Loop Detection = Off  PathVector Limit = 0  Max Pdu = 4096
P2MP Capability = yes
MP MBB Capability = yes
Overload Capability = yes
Dynamic Capability = yes
Unrecognized Notification Capability = yes
"
```

Configure Multicast LDP and Verify Traffic

Configure PIM and IGMP on the Root and Leaf Nodes

PIM needs to be enabled on the root node on the interface toward the multicast source S-1, as follows:

```
*A:PE-1# configure router pim interface "int-PE-1-S-1"
```

On the leaf nodes, PIM needs to be enabled (no shutdown), but no interfaces need to be assigned.

The IGMP configuration for root node PE-1 is needed to forward the incoming traffic for multicast group 232.1.1.1 from source 172.16.11.2 to the tunnel-interface. If IGMP is not configured, the incoming traffic on the interface toward the multicast source will be dropped, because no outgoing interface is defined.

```
*A:PE-1# configure router
        igmp
            tunnel-interface ldp-p2mp "5000" sender 192.0.2.1
                static
                    group 232.1.1.1
                        source 172.16.11.2
                    exit
                exit
            exit
        exit
```

The IGMP configuration for leaf node PE-3 is as follows:

```
*A:PE-3# configure router
        igmp
            interface "int-PE-3-H-3"
                static
                    group 232.1.1.1
                        source 172.16.11.2
                    exit
                exit
            exit
        exit
```

The incoming traffic from the tunnel interface will be forwarded to the outgoing interface toward the receiving multicast host H-3.

The IGMP configuration for leaf node PE-4 is similar.

At this point, the IGMP/PIM configuration on the root node is complete. This can be verified, as follows:

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
232.1.1.1                   (S,G)                        int-PE-1-S-1   1
   172.16.11.2
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
*A:PE-1# show router pim group detail
===============================================================================
PIM Source Group ipv4
```

```
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 172.16.11.2
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
MRIB Next Hop      : 172.16.11.2
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:02:39        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.11.2
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73728

Curr Fwding Rate   : 6319.3 kbps
Forwarded Packets  : 150360             Discarded Packets  : 0
Forwarded Octets   : 6916560            RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The incoming interface is the interface facing the multicast source S-1. The outgoing interface is a reference to the tunnel interface. The name for the outgoing interface (mpls-if-73728) contains the tunnel interface index 73728 as in previous CLI output. The multicast source S-1 is already sending traffic, but the receivers cannot receive it yet.

The configuration on the leaf nodes is still incomplete. A multicast policy needs to be configured and applied first.

Configure and Apply Multicast Policy on Leaf Nodes

The leaf nodes need to get multicast traffic off the LDP P2MP LSP. Therefore, a multicast policy needs to be created and applied. For leaf node PE-3, this is as follows:

```
*A:PE-3# configure mcast-management
        multicast-info-policy "p2mp-pol" create
            bundle "bundle1" create
                primary-tunnel-interface ldp-p2mp 5000 sender 192.0.2.1
                channel "232.1.1.1" create
                exit
```

```
            exit
         exit
*A:PE-3# configure router multicast-info-policy "p2mp-pol"
```

The configuration for leaf node PE-4 is identical.

Verify Multicast Traffic on Leaf Nodes

Verify the multicast traffic, as follows:

```
*A:PE-3# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
232.1.1.1                   (S,G)                        mpls-if-73728  1
   172.16.11.2
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-3#
```

The multicast source S-1 sends a multicast stream with group address 232.1.1.1. The multicast traffic is received by the leaf nodes, which can be verified as follows:

```
*A:PE-3# show router pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 172.16.11.2
RP Address         : 0
Advt Router        :
Flags              :                    Type               : (S,G)
MRIB Next Hop      :
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:02:36        Resolved By        : unresolved

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : mpls-if-73728
Outgoing Intf List : int-PE-3-H-3
```

```
Curr Fwding Rate   : 6235.2 kbps
Forwarded Packets  : 204013             Discarded Packets  : 0
Forwarded Octets   : 9384598            RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-3#
```

The incoming interface is from the tunnel interface, while the outgoing interface is toward the receiving multicast host H-3.

mLDP Fast Upstream Switchover

mLDP fast upstream switchover allows a downstream node of an mLDP FEC to perform a fast switchover and source the traffic from another upstream node. This switchover is necessary when IGP and LDP are converging due to a failure of the upstream LSR, which is the primary next hop of the root LSR for the P2MP FEC. There will be traffic duplication toward the node that has the upstream alternate backup (in this case to PE-3), but only one stream will be accepted. The multicast stream will be sent to the primary next hop as well as to the loopfree alternate backup. As long as there is no failure, the primary next hop accepts the traffic and forwards it. The backup rejects the traffic. When a failure occurs and the primary LDP session goes down, the backup will start accepting packets.

mLDP fast upstream switchover provides an upstream Fast Reroute (FRR) node-protection capability for the mLDP FEC packets. This multicast upstream FRR node protection is at the expense of traffic duplication from two different upstream nodes into the node that performs the fast upstream switchover. This feature is described in *draft-pdutta-mpls-mldp-up-redundancy*.

Multicast upstream FRR can be configured for mLDP, as follows:

```
*A:PE-1# configure router ospf loopfree-alternate

*A:PE-1# configure router ldp mcast-upstream-frr
```

This configuration can be repeated on some or all of the nodes. In this example, it is configured on all nodes. FRR for unicast can be configured in combination with this, but that is not required. FRR for unicast can be enabled as follows:

```
*A:PE-1# configure router ldp fast-reroute
*A:PE-1# configure router ip-fast-reroute
```

In this example, it is assumed that unicast IP and unicast LDP prefixes do not need to be protected. Therefore, unicast FRR remains disabled.

FRR can be verified as disabled for unicast (FRR) and enabled for multicast (Mcast Upstream FRR), as follows:

```
*A:PE-1# show router ldp status
===============================================================================
LDP Status for IPv4 LSR ID 192.0.2.1
               IPv6 LSR ID ::
===============================================================================
Created at         : 03/25/16 14:02:42
Last Change        : 03/25/16 14:10:40
Admin State        : Up
IPv4 Oper State    : Up                    IPv6 Oper State      : Down
IPv4 Up Time       : 0d 00:07:59           IPv6 Down Time       : 0d 00:07:59
IPv4 Oper Down Rea*: n/a                   IPv6 Oper Down Reason: systemIpDown
IPv4 Oper Down Eve*: 0                     IPv6 Oper Down Events: 0
Tunn Down Damp Time: 3 sec
Label Withdraw Del*: 0 sec                 Implicit Null Label  : Disabled
Short. TTL Local   : Enabled               Short. TTL Transit   : Enabled
Import Policies    : None                  Export Policies      : None
Tunl Exp Policies  : None
FRR                : Disabled              Mcast Upstream FRR   : Enabled
MP MBB Time        : 3
Aggregate Prefix   : False                 Agg Prefix Policies  : None
Class Forwarding   : Disabled
Legacy LSR Interop : False
-------------------------------------------------------------------------------
Capabilities
-------------------------------------------------------------------------------
Dynamic            : Enabled               P2MP                 : Enabled
IPv4 Prefix Fec    : Enabled               IPv6 Prefix Fec      : Enabled
Service Fec128     : Enabled               Service Fec129       : Enabled
MP MBB             : Enabled               Overload             : Enabled
Unrecognized Notif*: Enabled
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

Of the three nodes in the test topology that have upstream nodes, only PE-3 has an upstream alternate for FRR. PE-4 becomes a transit node for traffic destined to PE-3 (but PE-3 will drop it, until the primary LDP session fails). PE-3 sends a label mapping message to PE-4 for label 262138, as in the following trace message.

```
8 2016/03/25 14:10:34.07 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 63) to 192.0.2.4:0
Protocol version = 1
Label 262138 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"
```

PE-4 will have an additional LDP P2MP binding where the label is swapped, as follows:

```
*A:PE-4# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
5000                                        73728
192.0.2.1                                   Pop             262139      --
  --                                          --

5000                                        73728
192.0.2.1                                   Swap            262139    262138
192.168.34.1                                1/1/2

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
---snip---
```

PE-3 has an additional entry for the FRR backup that is available (BU - Alternate for Fast Re-Route). PE-3 will get duplicated traffic, but will reject all traffic from PE-4 and only accept traffic from PE-2 as long as there is no failover.

```
*A:PE-3# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
5000                                        73728
192.0.2.1                                   Pop             262139      --
  --                                          --

5000                                        73728
192.0.2.1                                   Pop             262138BU    --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
```

```
---snip---
```

Because LoopFree Alternate (LFA) and ECMP are mutually exclusive, LFA is only useful when ECMP is disabled. When both are enabled, ECMP will have preference.

mLDP fast upstream switchover relies on the fast detection of loss of the LDP session to the upstream peer to which the primary ingress label map (ILM) label had been advertised. As a result, Nokia recommends to perform the following:

1. Enable bidirectional forwarding detection (BFD) on all LDP interfaces to upstream LSR nodes. When BFD detects the loss of the last adjacency to the upstream LSR, BFD brings down the LDP session immediately. The backup ILM is activated.
2. If there is a concurrent T-LDP adjacency to the same LSR node, enable BFD on the T-LDP peer as well as on the interface.
3. Enable the ldp-sync-timer option on all interfaces to the upstream LSR nodes.

   If the LDP session for the primary ILM to the upstream LSR goes down for any other reason than a failure of the interface or of the upstream LSR, routing and LDP will go out of sync. The backup ILM will remain activated until the interior gateway protocol (IGP) seeks the next shortest path first (SPF). By enabling the ldp-sync-timer, this process is accelerated because the advertised link metric will get the maximum value as soon as the LDP session goes down. This triggers the IGP to calculate an SPF route.

The FRR configuration can be removed, as follows:

```
*A:PE-1# configure router ldp no mcast-upstream-frr
*A:PE-1# configure router ospf no loopfree-alternate
```

Multipoint Make-Before-Break (MP MBB)

Multipoint MBB is performed when the best path to the root changes, but the existing path can still be used, such as when a link comes up or when the routing metric changes. The goal of MBB is to establish a new P2MP LSP before the old P2MP is removed, in order to avoid traffic loss.

Leaf or transit nodes must allocate a new label and program the ILM with a duplicate set of existing next-hop label forwarding entries (NHLFEs) toward the upstream nodes. This may lead to traffic duplication for a short period of time.

Multipoint MBB is enabled by default, as follows:

```
*A:PE-3# show router ldp status
===============================================================================
```

```
LDP Status for IPv4 LSR ID 192.0.2.3
               IPv6 LSR ID ::
===============================================================================
Created at          : 03/25/16 14:02:47
Last Change         : 03/25/16 14:11:28
Admin State         : Up
IPv4 Oper State     : Up                    IPv6 Oper State      : Down
IPv4 Up Time        : 0d 00:08:42           IPv6 Down Time       : 0d 00:08:42
IPv4 Oper Down Rea*: n/a                    IPv6 Oper Down Reason: systemIpDown
IPv4 Oper Down Eve*: 0                      IPv6 Oper Down Events: 0
Tunn Down Damp Time: 3 sec
Label Withdraw Del*: 0 sec                  Implicit Null Label  : Disabled
Short. TTL Local    : Enabled               Short. TTL Transit   : Enabled
Import Policies     : None                  Export Policies      : None
Tunl Exp Policies   : None
FRR                 : Disabled              Mcast Upstream FRR   : Disabled
MP MBB Time         : 3
Aggregate Prefix    : False                 Agg Prefix Policies  : None
Class Forwarding    : Disabled
Legacy LSR Interop  : False
-------------------------------------------------------------------------------
Capabilities
-------------------------------------------------------------------------------
Dynamic             : Enabled               P2MP                 : Enabled
IPv4 Prefix Fec     : Enabled               IPv6 Prefix Fec      : Enabled
Service Fec128      : Enabled               Service Fec129       : Enabled
MP MBB              : Enabled               Overload             : Enabled
Unrecognized Notif*: Enabled
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

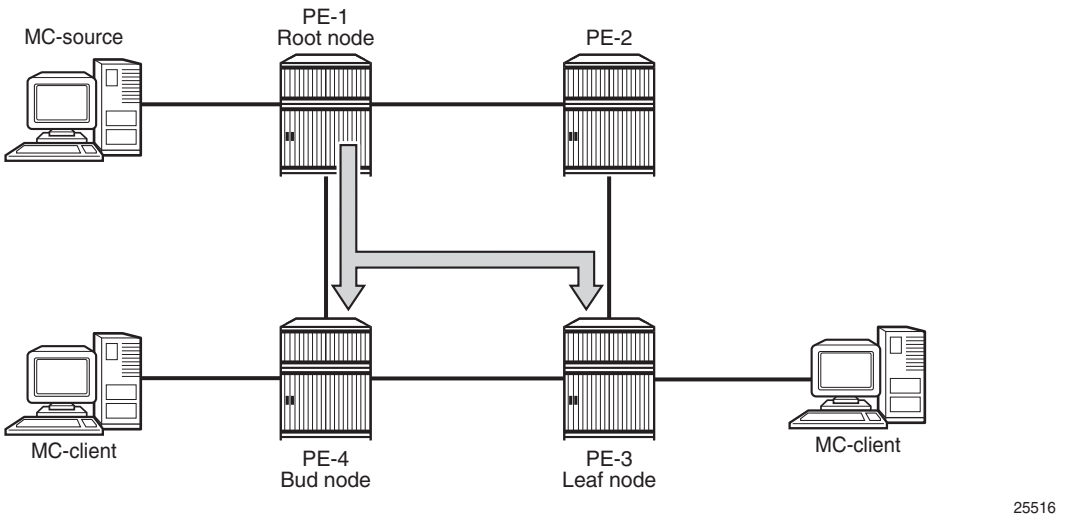
When the metric is increased on the interface (int-PE-3-PE-2) toward the active upstream node, PE-3 sends out an OSPF link status update. Traffic still arrives at PE-3 using the original P2MP LSP. The MBB P2MP LSP is set up.

PE-3 sends a label mapping message toward PE-4, including an MP status TLV carrying an MBB status code indicating that MBB procedures apply to the LSP. PE-4 sends an LDP notification toward PE-3, including an MP status TLV indicating that PE-4 has a state for the existing P2MP LSP.

PE-3 sends an LDP withdrawal message to PE-2. PE-2 replies with an LDP release message.

The multicast traffic arrives at PE-3 using the new LDP P2MP LSP. This way, PE-4 becomes a bud node, and PE-2 is not used for transit anymore; see Figure 174.

Figure 174 **New LDP P2MP LSP after Metric Change**



Originally, leaf node PE-3 preferred the route via PE-2 toward root node PE-1, as follows:

```
*A:PE-3# show router route-table 192.0.2.1
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                        Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                      Remote  OSPF      00h00m52s  10
       192.168.23.1                                                    20
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-3#
```

Consequently, the label map messages were originally sent to PE-2, not to PE-4. PE-2 is the transit node for traffic destined to PE-3, as follows:

```
*A:PE-2# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
```

```
===============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
-------------------------------------------------------------------------------
5000                                     Unknw
192.0.2.1                                Swap            262139    262139
192.168.23.2                             1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
---snip---
```

The metric is changed on the interface between PE-3 and PE-2, as follows:

```
*A:PE-3# configure router ospf area 0 interface "int-PE-3-PE-2" metric 1000
```

The preferred route from leaf node PE-3 to root node PE-1 is now via PE-4, as follows:

```
*A:PE-3# show router route-table 192.0.2.1
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                           Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                 Remote  OSPF      00h13m02s  10
       192.168.34.2                                                 20
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-3#
```

The leaf node PE-3 will prefer to set up a path from PE-4 rather than PE-2. PE-3 will send label mapping messages to PE-4. The old P2MP LSP will be used until the new P2MP LSP is set up. There will be no traffic interruption.

PE-3 will send a label withdrawal message to PE-2 and PE-2 will no longer be a transit node, as follows:

```
*A:PE-2# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
```

```
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
```

PE-4 is the transit node for traffic to PE-3, and also has a local multicast client H-4, so it is a bud node, as follows:

```
*A:PE-4# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
5000                                        73728
192.0.2.1                                   Pop             262139      --
  --                                          --

5000                                        73728
192.0.2.1                                   Swap            262139    262138
192.168.34.1                                1/1/2

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
---snip---
```

There is no traffic multiplication at the root node PE-1. All traffic goes to PE-4, as follows:

```
*A:PE-1# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
```

```
-------------------------------------------------------------------------------
5000                                        73728
192.0.2.1                                   Push              --      262139
192.168.14.2                                1/1/2

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
---snip---
```

The switchover to this new P2MP LSP occurred without traffic loss.

The following debugging was enabled before the metric change:

```
*A:PE-3# debug router ldp peer 192.0.2.2 packet label detail
*A:PE-3# debug router ldp peer 192.0.2.4 packet label detail
*A:PE-3# debug router ldp peer 192.0.2.2 packet init detail
*A:PE-3# debug router ldp peer 192.0.2.4 packet init detail
```

The first trace message shows that label 262138 is advertised to PE-4. MBB is requested, as follows:

```
11 2016/03/25 14:12:16.89 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 76) to 192.0.2.4:0
Protocol version = 1
Label 262138 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
MP Status MBB = REQ
"
```

The next message is a notification from PE-4 confirming that there is no fatal error and MBB can be applied, as follows:

```
12 2016/03/25 14:12:16.89 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Notification packet (msgId 76) from 192.0.2.4:0
Protocol version = 1
Status Code = MPStatus (0x00000040) Non-fatal
Causing message Id = 0
Causing message type = NULL
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
MP Status MBB = ACK
"
```

The following message is a label withdraw message for label 262139 sent to PE-2:

```
13 2016/03/25 14:12:16.89 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Withdraw packet (msgId 76) to 192.0.2.2:0
Protocol version = 1
Label 262139 withdrawn for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"
```

The last message is a label release message for label 262139 received from PE-2, as follows. This message is only sent after the new P2MP LSP is set up.

```
14 2016/03/25 14:12:17.01 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Release packet (msgId 75) from 192.0.2.2:0
Protocol version = 1
Label 262139 released for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"
```

Conclusion

Multicast LDP provides extensions to the LDP protocol for the setup of P2MP and MP2MP LSPs in MPLS networks. mLDP is simple to configure compared to RSVP. FRR and MBB are supported for mLDP.

RSVP Point-to-Point LSPs

This chapter provides information about point-to-point label switched paths (LSPs) established using resource reservation protocol (RSVP) with or without traffic engineering (TE).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was written for release 14.0.R1. There are no pre-requisites or conditions on the hardware for this configuration.

Overview

Due to the connectionless nature of the network layer protocol IP, packets travel through the network on a hop-by-hop basis with routing decisions made at each node. As a result, hyperaggregation of data on certain links may occur and it may impact the provider's ability to provide guaranteed service levels across the network end-to-end. To address these shortcomings, multiprotocol label switching (MPLS) was developed. The technology provides the capability to establish connection oriented paths, called label switched paths (LSPs), over a connectionless (IP) network. The LSP offers a mechanism to engineer network traffic independently from the underlying network routing protocol (mostly IP) to improve the network resiliency and recovery options and to permit delivery of new services that are not readily supported by conventional IP routing techniques, such as Layer 2 IP virtual private networks (VPNs). These benefits are essential for today's communication network explaining the wide deployment base of the MPLS technology.

RFC 3031, *Multiprotocol Label Switching Architecture*, specifies the MPLS architecture while this document describes the configuration and troubleshooting of RSVP point-to-point LSPs on SR OS. Besides RSVP P2P LSPs, there are also Static Point-to-Point LSPs, LDP Point-to-Point LSPs, and Segment Routing with IS-IS Control Plane.

Packet Forwarding

As a packet of a connectionless network layer protocol travels from one router to the next, each router in the network makes an independent forwarding decision by performing the following basic tasks: first analyzing the packet header, then referencing the local routing table to find the longest match based on the destination address in the IP header, and finally sending out the packet on the selected interface. In other words, the first function partitions the entire set of possible packets into a set of forwarding equivalence classes (FECs). All packets associated to a particular FEC will be forwarded along the same logical path to the same destination. The second function maps each FEC to a next hop destination router. Each router along the data path performs these actions.

On the other hand, in MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network. In turn, the FEC is mapped to an LSP, which is pre-signaled prior to any data flowing. An MPLS label, representing the FEC to which the packet is assigned, is attached to the packet (push operation) and once labeled, the packet is forwarded to the next hop router along that LSP path. At subsequent hops, there is no further analysis of the packet network layer header. Instead, the label is used as an index into a table which specifies the next hop and a new label. The old label is replaced with the new label (swap operation), and the packet is forwarded to its next hop. At the MPLS network egress, the label is removed from the packet (pop operation). If this router is the final destination (based on the remaining packet), the packet is handed to the receiving application (such as a virtual private LAN service (VPLS) domain). If this router is not the final destination of the packet, the packet will be sent into a new MPLS tunnel or forwarded by conventional IP forwarding towards the Layer 3 destination.

Terminology

Figure 175 **Generic MPLS Network, MPLS Label Operations**

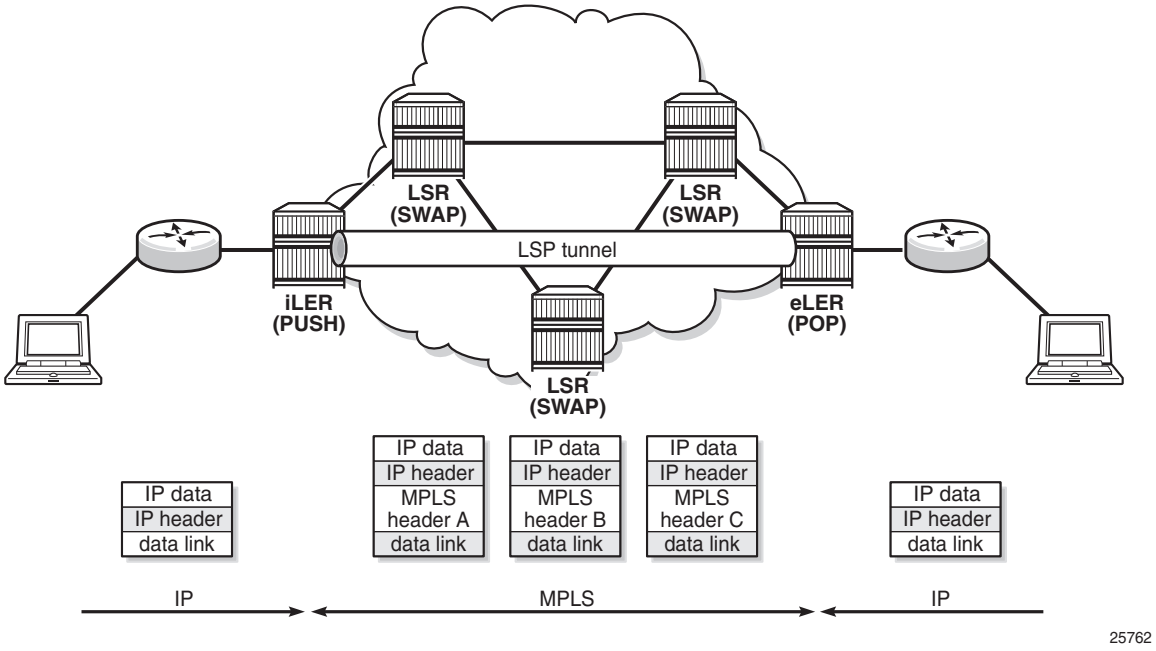


Figure 175 depicts a general network topology clarifying the MPLS-related terms. A label edge router (LER) is a device at the edge of an MPLS network, with at least one interface outside the MPLS domain. A router is usually defined as an LER based on its position relative to a particular LSP. The MPLS router at the head-end of an LSP is called the ingress label edge router (iLER). The MPLS router at the tail-end of an LSP is called the egress label edge router (eLER). The iLER receives unlabeled packets from outside the MPLS domain, then applies MPLS labels to the packets, and forwards the labeled packets into the MPLS domain. The eLER receives labeled packets from the MPLS domain, then removes the labels, and forwards unlabeled packets outside the MPLS domain. The eLER can signal an implicit-null label (numeric value 3). This informs the previous hop to send MPLS packets without an outer label and is known as penultimate hop popping (PHP).

A label switching router (LSR) is a device internal to an MPLS network, with all interfaces inside the MPLS domain. These devices switch labeled packets inside the MPLS domain. In the core of the network, LSR ignore the packet network layer (IP) header and simply forward the packet using the MPLS label swapping mechanism.

A single LSP is unidirectional. In common practice, because the bidirectional nature of most traffic flows is implied, the term LSP often is used to define the pair of LSPs that enable the bidirectional flow. For ease of terminology and discussion however, the LSP in this chapter is referred to as a single entity.

LSP Establishment

Prior to packet forwarding, the LSP must be established. In order to do so, labels need to be distributed for the path. Labels are usually distributed by a downstream router in the upstream direction (relative to the data flow). There are a number of ways used for label distribution: static, LDP, and RSVP. For static P2P LSPs, see chapter Static Point-to-Point LSPs; for LDP P2P LSPs, see chapter LDP Point-to-Point LSPs.

RSVP-TE (RFC 3209, *RSVP-TE: Extensions to RSVP for LSP Tunnels*) can be used to signal LSPs across the network. RSVP-TE is used for traffic engineering when the ingress router creates an LSP with specific constraints beyond the best route chosen by the IGP. RSVP-TE identifies the specific path desired for the LSP and may include resource requirements for the path.

The most important benefit of the label swapping mechanism RSVP-TE is its ability to map any type of user traffic to an LSP that has been specifically engineered to satisfy user traffic requirements. Customized LSPs may be created based on hop count, bandwidth requirements, administrative groups, or shared risk link groups (SRLGs). They can even be routed through a strict path with specific network links or nodes, as specified by the ingress node. This offers service providers precise control over the flow of traffic in their networks and results in a network that operates more efficiently and provides more predictable and scalable services. For information about SRLG, see chapter Shared Risk Link Groups for RSVP-Based LSP.

Fast reroute (FRR) allows to signal backup paths before a failure takes place. This allows traffic to flow almost continuously, without waiting for routing protocol convergence. There a two different methods for FRR for an RSVP-TE LSP: one-to-one and facility.

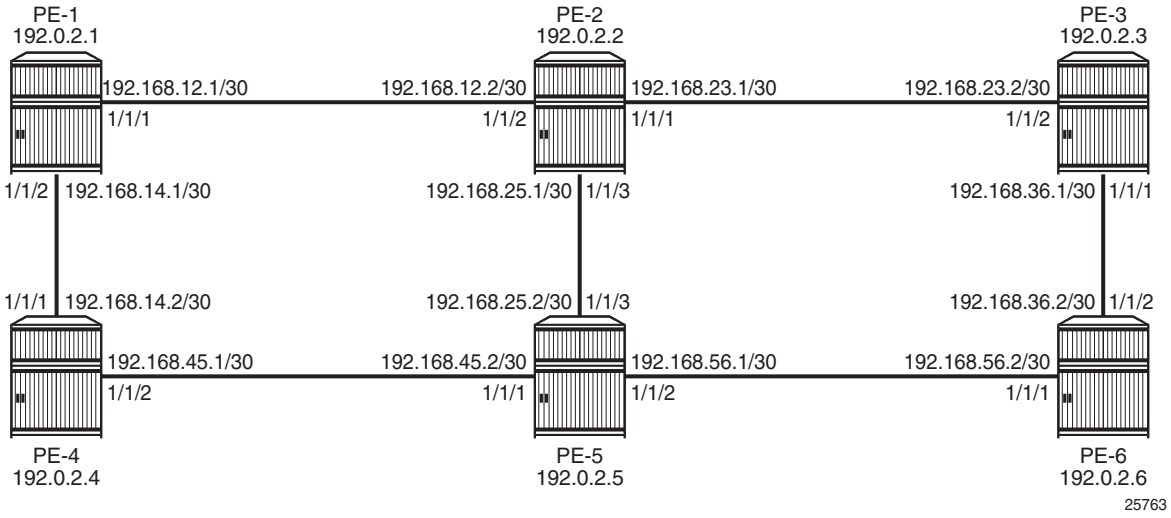
- FRR one-to-one defines detour tunnels toward the eLER for a particular LSP only. The advantage is that the detour tunnel is the best path to the eLER that avoids the node or link at the point of failure. The drawback is that when different LSPs would need the same detour, a dedicated RSVP-TE detour LSP needs to be signaled for each LSP.

- FRR facility defines local repair tunnels avoiding one particular node (the next hop in the data path) or one particular link (the next link in the data path), ignoring the eLER. These bypass tunnels originate in a point of local repair (PLR) and terminate in a merge point (MP) on the LSP. Bypass tunnels are shared between LSPs.

Example Topology

The example topology is displayed in Figure 176. The setup consists of six 7750 SR nodes located in a single autonomous system.

Figure 176 **MPLS Example Topology**



Configuration

In this chapter, RSVP LSPs are configured manually, but they can also be configured automatically using LSP templates; see chapter Automatic Creation of RSVP-TE LSPs.

As a general pre-requisite for the configuration of MPLS LSPs, a correctly working IGP is required. Open shortest path first (OSPF) or intermediate system to intermediate system (IS-IS) can be used as IGP.

RSVP-TE, an extension of the original RSVP protocol, has two major benefits adding to the basic MPLS functionality. The first benefit is traffic engineering, which allows the ingress router to create an LSP with specific constraints beyond the best route chosen by the IGP. The second benefit is improved network resiliency when a link or node fails in the network, using FRR and secondary paths. FRR is also supported for LDP, see chapter MPLS LDP FRR using ISIS as IGP.In this chapter, several RSVP-TE LSPs are configured:

- A simple LSP with a primary path that has strict hops and no specific TE constraints
- A simple LSP with a dynamic path without any configured hops is created. Initially, there are no constraints and the actual path is calculated based on the IGP best route.
- An LSP configured with constrained shortest path first (CSPF) that will use the TE metric, even though the IGP metric can also be used
- An LSP with fast reroute (FRR) one-to-one enabled
- An LSP with FRR facility enabled
- An LSP including an admin group "blue" and an LSP excluding admin group "red"
- An LSP with a hop limit configured

There is no configuration example with bandwidth constraints configured in this chapter. See chapter Automatic Bandwidth Adjustment in P2P LSPs for a configuration with bandwidth constraint with or without automatic adjustment.

Initially, no traffic engineering is enabled in the ISIS context, but it will be enabled when required. For RSVP LSPs, the MPLS instance needs to be enabled on each router and all network interfaces facing the MPLS domain. By default, the system interface is put automatically within the MPLS context. When adding interfaces to the MPLS instance, they are automatically added to the RSVP instance as well, but the instance itself is still in an administrative shutdown state. The next step is to enable the RSVP instance on all routers in the MPLS network. As a result, all interfaces facing the MPLS domain as well as the system interface are added to the MPLS and RSVP instance and both instances are in a no shutdown state. For PE-1, the following configuration is required:

```
*A:PE-1# configure router
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-4"
            exit
            no shutdown
        exit
*A:PE-1# configure router rsvp no shutdown
```

Strict or Loose Path

On the iLER, first the definition of a path is required. A path is a sequence of MPLS routers (hops) through which the LSP using that path has to pass. It is not uniquely bound to a particular LSP; it can be used by any LSP originating in that node. A hop in a path can be strict or loose: strict or loose meaning that the LSP must take either a direct path from the previous hop router to this router (strict) or can traverse through other routers (loose). The hops not explicitly defined in the loose path definition are created by calculating the IGP shortest path. A third possibility is an empty path implying not a single node is required to be present in the LSP path and the shortest path from the IGP is used to define the LSP path. Other techniques, such as the use of admin groups or shared risk link groups, can also be used to influence the decision which hops to include in the path. Three paths will be configured, respectively:

1. Only strict hops
2. Mixed strict and loose hops
3. Empty path

To find a valid path, the last hop in the path sequence needs to be the system IP or an interface address of the terminating router (eLER). The IP addresses in the hop command can be the system IP addresses or the interface addresses of the node. However, it is recommended to use the system IP addresses with keyword **loose** as this allows more flexibility when finding new paths in failover scenarios (because the upstream node could use any of multiple paths to the system address, whereas specifying the interface address would restrict the upstream node to a single entry-point). The recommendation when using the keyword strict in the hop command context, is to use the physical link addresses. However, the last hop in the path should be a system address to make it appear in the list on the 5620 SAM (service-aware manager).

```
*A:PE-1# configure router
        mpls
            path "path-PE-1-PE-6-strict"
                hop 10 192.168.12.2 strict
                hop 20 192.168.25.2 strict
                hop 30 192.168.56.2 strict
                no shutdown
            exit
            path "path-PE-1-PE-6-semiLoose"
                hop 10 192.0.2.5 loose
                hop 20 192.168.56.2 strict
                no shutdown
            exit
            path "dyn"
                no shutdown
            exit
```

The paths can be checked with the **show router mpls path** command.

```
*A:PE-1# show router mpls path
===============================================================================
MPLS Path:
===============================================================================
Path Name                       Adm  Hop Index   IP Address       Strict/Loose
-------------------------------------------------------------------------------
path-PE-1-PE-6-strict           Up   10          192.168.12.2     Strict
                                     20          192.168.25.2     Strict
                                     30          192.168.56.2     Strict

path-PE-1-PE-6-semiLoose        Up   10          192.0.2.5        Loose
                                     20          192.168.56.2     Strict

dyn                             Up   no hops     n/a              n/a

-------------------------------------------------------------------------------
Total Paths : 3
===============================================================================
*A:PE-1#
```

Simple RSVP LSP with Strict Primary Path

The configuration of a simple LSP using RSVP signaling contains at least on the iLER:

- System IP address of the terminating node (to)
- Path the LSP will take to the eLER (primary)
- Administratively enabled (no shutdown)

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-6"
                to 192.0.2.6
                primary "path-PE-1-PE-6-strict"
                exit
                secondary "dyn"
                exit
                no shutdown
            exit
```

All the hops in the strict path are already defined and there is no need to look up the IGP best route. The configuration of secondary paths is optional. In case the primary path fails, the secondary path can be signaled to take over the traffic. It can even be signaled as standby while the primary path is operational for a faster switchover when the keyword **standby** is added, which is not the case here. The secondary path has no hops defined. The hops will be calculated based on the IGP best route. The nodes through which the LSP will pass (LSRs and eLER) require no additional configuration: enabling MPLS and RSVP on their interfaces suffices.

An overview of all LSPs configured on a particular node is given by the **show router mpls lsp** command. More details about a particular LSP can be retrieved by adding the keyword **detail** to the previous command.

```
*A:PE-1# show router mpls lsp
===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                            To              Tun     Fastfail  Adm  Opr
                                                    Id      Config
-------------------------------------------------------------------------------
LSP-PE-1-PE-6                       192.0.2.6       1       No        Up   Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-1#
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-6" detail
===============================================================================
MPLS LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-PE-1-PE-6
LSP Type    : RegularLsp                        LSP Tunnel ID  : 1
LSP Index   : 1                                 TTM Tunnel Id  : 1
From        : 192.0.2.1                         To             : 192.0.2.6
Adm State   : Up                                Oper State     : Up
LSP Up Time : 0d 00:01:27                       LSP Down Time  : 0d 00:00:00
Transitions : 1                                 Path Changes   : 1
Retry Limit : 0                                 Retry Timer    : 30 sec
Signaling   : RSVP                              Resv. Style    : SE
Hop Limit   : 255                               Negotiated MTU : 1564
Adaptive    : Enabled                           ClassType      : 0
FastReroute : Disabled                          Oper FR        : Disabled
CSPF        : Disabled                          ADSPEC         : Disabled
Metric      : N/A
Load Bal Wt : N/A                               ClassForwarding: Disabled
Include Grps:                                   Exclude Grps   :
None                                            None
Least Fill  : Disabled
BFD Template: None                              BFD Ping Intvl : 60
BFD Enable  : False

Revert Timer: Disabled                          Next Revert In : N/A
EntropyLbl  : Inherited
Auto BW     : Disabled
LdpOverRsvp : Enabled
VprnAutoBind: Enabled
IGP Shortcut: Enabled                           BGP Shortcut   : Enabled
IGP LFA     : Disabled                          IGP Rel Metric : Disabled
BGPTransTun : Enabled
Oper Metric : 16777215
Prop Adm Grp: Disabled

Primary(a)  : path-PE-1-PE-6-strict             Up Time        : 0d 00:01:27
Bandwidth   : 0 Mbps
Secondary   : dyn                               Down Time      : 0d 00:01:27
```

```
Bandwidth   : 0 Mbps
===============================================================================
*A:PE-1#
```

In each hop (originating, transit and terminate), the RSVP sessions can be verified as follows:

```
*A:PE-1# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.6       1      24576 LSP-PE-1-PE-6::path-PE-1-PE* Up
-------------------------------------------------------------------------------
Sessions : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The detailed output of this command includes among others the session type (here: originate), the incoming and outgoing labels, the previous and next hop, and - for originating LSPs - also the list of hops):

```
*A:PE-1# show router rsvp session detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-PE-1-PE-6::path-PE-1-PE-6-strict
-------------------------------------------------------------------------------
From            : 192.0.2.1             To              : 192.0.2.6
Tunnel ID       : 1                     LSP ID          : 24576
Style           : SE                    State           : Up
Session Type    : Originate
In Interface    : n/a                   Out Interface   : 1/1/1
In Label        : n/a                   Out Label       : 262143
Previous Hop    : n/a                   Next Hop        : 192.168.12.2
Hops            :
    192.168.12.2(S)    -> 192.168.25.2(S)    -> 192.168.56.2(S)
SetupPriority   : 7                     Hold Priority   : 0
Class Type      : 0
SubGrpOrig ID   : 0                     SubGrpOrig Addr:
P2MP ID         : 0
FrrAvailType    : N/A
FrrSrlgStrict   : N/A                   SrlgDisjoint    : N/A

Path Recd       : 0                     Path Sent       : 1
Resv Recd       : 2                     Resv Sent       : 0
Summary msgs    :
SPath Recd      : 0                     SPath Sent      : 0
SResv Recd      : 0                     SResv Sent      : 0
LSP Attr Flags  : N/A
===============================================================================
*A:PE-1#
```

In the tunnel table, there is an RSVP LSP:

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.6/32      rsvp      MPLS  1         7        192.168.12.2   16777215
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

In order to signal PHP with RSVP, implicit-null must be configured on the eLER (RSVP must be shut down to perform this command).

```
*A:PE-6# configure router
        rsvp
            shutdown
            implicit-null-label
            no shutdown
            exit
```

The implicit-null is signaled after re-enabling RSVP and is shown on PE-5 as an egress label of 3. This label is not actually sent toward PE-6.

```
*A:PE-5# show router rsvp session detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-PE-1-PE-6::path-PE-1-PE-6-strict
-------------------------------------------------------------------------------
From            : 192.0.2.1             To              : 192.0.2.6
Tunnel ID       : 1                     LSP ID          : 41476
Style           : SE                    State           : Up
Session Type    : Transit
In Interface    : 1/1/3                 Out Interface   : 1/1/2
In Label        : 262143                Out Label       : 3
Previous Hop    : 192.168.25.1          Next Hop        : 192.168.56.2
---snip---
```

The use of implicit-null can also be enabled/disabled on a per interface basis (either RSVP, or the interface within RSVP, must be shut down to perform this change).

```
A:PE-6>config>router>rsvp# interface "int-PE-6-PE-5"
A:PE-6>config>router>rsvp>if# implicit-null-label
  - implicit-null-label {<enable|disable>}
  - no implicit-null-label
 <<enable|disable>>   : keyword
```

In the remainder of the chapter, LSPs with empty paths will be used. LSP "LSP-PE-1-PE-6" is shut down.

Simple RSVP LSP with Dynamic Path

In this section, an LSP is configured from PE-1 to PE-3 with a dynamic path that is empty. There is no secondary path. LSP "LSP-PE-1-PE-3" is configured as follows:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
                primary "dyn"
                exit
                no shutdown
            exit
```

Interfaces with a lower metric will be preferred over links with a high metric. The default IGP metric in this example is 10. The metric is lower for higher speed links, but can be configured manually; as follows:

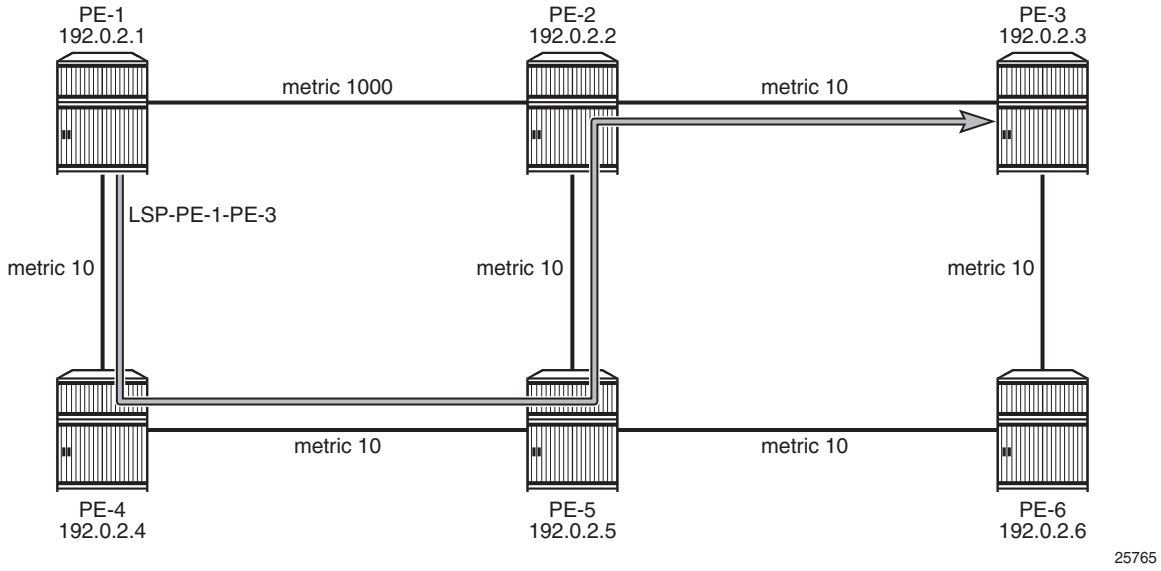
```
*A:PE-1# configure router isis interface "int-PE-1-PE-2" level 1 metric 1000
```

The link between PE-1 and PE-2 has a higher metric and will not be selected for forwarding traffic because the detour via PE-4 has a lower metric. The routing table shows that the route to prefix 192.0.2.3 has PE-4 as next hop instead of PE-2 and that the metric is 40:

```
*A:PE-1# show router route-table 192.0.2.3
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.3/32                                    Remote  ISIS      00h14m02s  15
       192.168.14.2                                                    40
-------------------------------------------------------------------------------
No. of Routes: 1
```

Figure 177 shows the path used by the LSP:

Figure 177 **LSP with Dynamic Path Takes IGP Best Route**



The actual hops can be verified in the following output. Since the path is dynamic, there are no explicit hops configured.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 46106
From              : 192.0.2.1             To                  : 192.0.2.3
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up
Out Interface     : 1/1/2                 Out Label           : 262143
---snip---
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.168.14.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)                    Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)                    Record Label        : 262143
 -> 192.168.25.1 (192.0.2.2)                    Record Label        : 262142
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262142
```

```
---snip---
```

The tunnel table shows the RSVP LSP with PE-4 as the next hop and a metric of 40:

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.3/32      rsvp      MPLS  2         7        192.168.14.2   40
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

Another path with metric 40 is as follows: PE-1, PE-4, PE-5, PE-6, PE-3. This can equally well be selected.

RSVP-TE LSP with Dynamic Path

Traffic engineering is enabled in the ISIS context on all nodes; as follows:

```
*A:PE-1# configure router isis traffic-engineering
```

The LSP can be configured with constrained shortest path first (CSPF); as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" cspf
```

For this LSP, it will not make any difference. By default, the IGP metrics are used and the LSP path takes the IGP shortest path.

Besides IGP metrics, also TE metrics can be configured; as follows:

```
*A:PE-2# configure router
        mpls
            interface "int-PE-2-PE-1"
                te-metric 10
            exit
            interface "int-PE-2-PE-3"
                te-metric 500
            exit
            interface "int-PE-2-PE-5"
                te-metric 10
            exit
```

In this example, all interfaces on all PEs get a TE metric of 10, except for the interfaces between PE-2 and PE-3, which get a TE metric of 500. Even with these TE metrics configured, the LSP path will not change, because the IGP metric is used by default, as can be verified as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
---snip---
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 18434
From              : 192.0.2.1             To                  : 192.0.2.3
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up
Out Interface     : 1/1/2                 Out Label           : 262142
---snip---
CSPF              : Enabled               Oper CSPF           : Enabled
---snip---
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.168.14.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)                    Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)                    Record Label        : 262143
 -> 192.168.25.1 (192.0.2.2)                    Record Label        : 262142
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262142
---snip---
Last Resignal     : n/a                   CSPF Metric         : 40
---snip---
```

The RSVP LSP in the tunnel table has next hop PE-4 and a metric of 40:

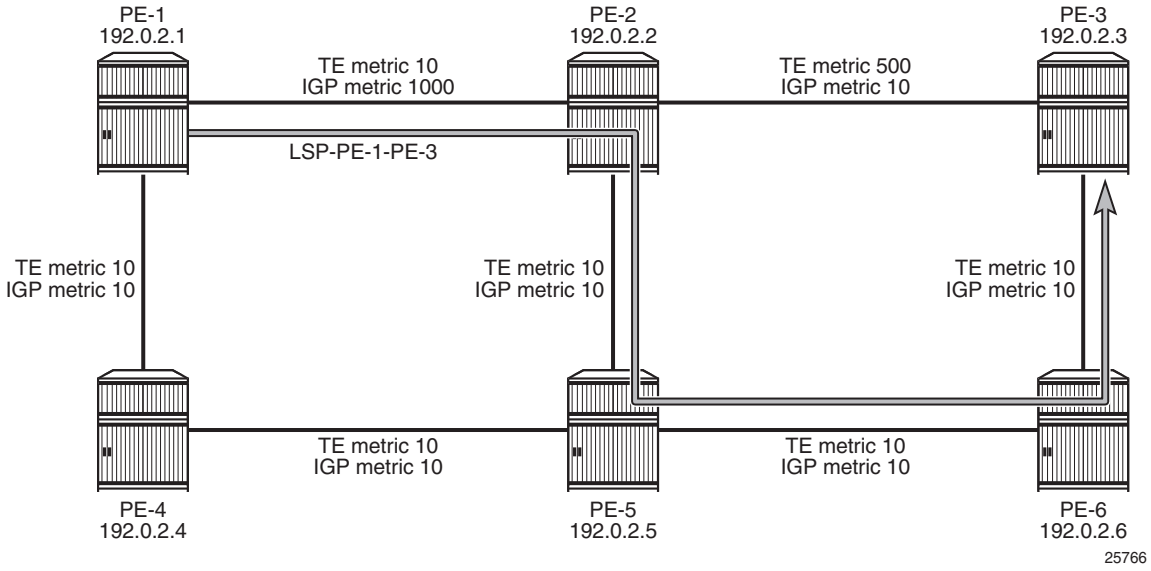
```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.3/32      rsvp      MPLS  2         7        192.168.14.2   40
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

To force the LSP to use the TE metric, the LSP is reconfigured as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" cspf use-te-metric
```

The LSP path is shown in Figure 178:

Figure 178 **RSVP-TE LSP with Dynamic Path Using TE Metric**



The LSP path goes from PE-1 to PE-2 and via PE-5 and PE-6 to PE-3, as can be seen in the following output. The CSPF metric is 40, which corresponds to the TE metric in this case:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
---snip---
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 18436
From              : 192.0.2.1             To                  : 192.0.2.3
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up
Out Interface     : 1/1/1                 Out Label           : 262143
---snip---
CSPF              : Enabled               Oper CSPF           : Enabled
---snip---
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.168.12.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262143
 -> 192.168.25.2 (192.0.2.5)                    Record Label        : 262143
 -> 192.168.56.2 (192.0.2.6)                    Record Label        : 262142
 -> 192.168.36.1 (192.0.2.3)                    Record Label        : 262143
Computed Hops     :
    192.168.12.1(S)
 -> 192.168.12.2(S)
 -> 192.168.25.2(S)
 -> 192.168.56.2(S)
```

```
 -> 192.168.36.1(S)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 40
Last MBB    :
 MBB Type        : ConfigChange         MBB State            : Success
 Ended At        : 03/02/2016 12:32:12  Old Metric           : 40
 Signaled BW     : 0 Mbps
===============================================================================
*A:PE-1#
```

The tunnel table shows the RSVP LSP with next hop PE-2 and a metric of 16777215 (infinity) because the IGP metric is not used:

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.3/32      rsvp      MPLS  2         7        192.168.12.2   16777215
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

The IGP metric values are restored to their default value as follows:

```
*A:PE-1# configure router isis interface "int-PE-1-PE-2" level 1 no metric
*A:PE-2# configure router isis interface "int-PE-2-PE-1" level 1 no metric
```

The TE metrics are configured with the same value for all interfaces; as follows:

```
*A:PE-2# configure router mpls interface "int-PE-2-PE-1" te-metric 10
*A:PE-3# configure router mpls interface "int-PE-3-PE-2" te-metric 10
```

Since all metrics have the same value, it does not matter whether CSPF uses the IGP or TE metric.

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" cspf
```

The primary path will go from PE-1 to PE-2 and then to PE-3 with a CSPF (IGP) metric of 20.

Fast Reroute for RSVP-TE LSPs

It is mandatory to have CSPF enabled for FRR.

Fast reroute can be configured on the RSVP LSP in two ways:

1. One-to-one: for each potential point of failure, the best detour tunnel to the eLER is signaled. This detour tunnel is signaled for this particular LSP only and cannot be shared among LSPs
2. Facility: local bypass tunnels are signaled from each point of local repair avoiding the next link or the next node. The bypass tunnels can be shared among LSPs.

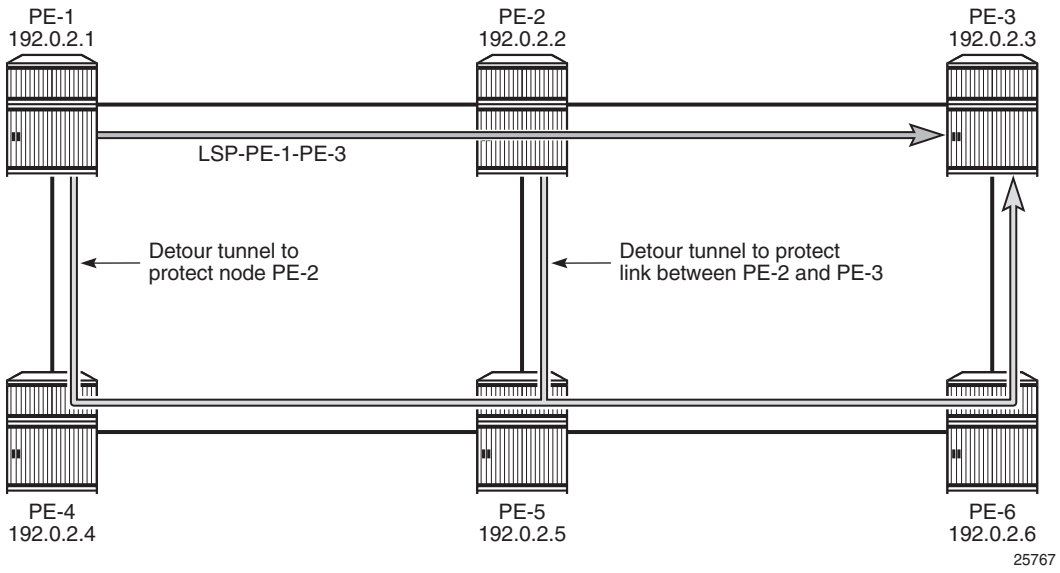
FRR One-to-One

The LSP "LSP-PE-1-PE-3" is configured with FRR one-to-one; as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" fast-reroute one-to-one
```

The preferred path from PE-1 to PE-3 is via PE-2. There will be two detour tunnels: one originating in PE-1 to protect node PE-2, and a second detour tunnel originating in PE-2 to protect the link between PE-2 and PE-3. Both detour tunnels use the same path from PE-5 to PE-3 and there is no need to signal this path twice. One detour tunnel terminates in PE-5, and the diverted traffic in this tunnel will be sent to PE-6 and PE-3 via the established detour tunnel. Depending on which detour tunnel is established first, the other detour tunnel terminates in PE-5. The preferred tunnel and the detour tunnels are shown in Figure 179:

Figure 179 **Fast Reroute One-to-One Detour Tunnels**



The protection can be seen in the list of actual hops in the path. In PE-1, a detour tunnel for node protection originates (indicated by @ n; see legend) and in PE-2 a detour tunnel for link protection (indicated by @):

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
---snip---
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 55810
From              : 192.0.2.1             To                   : 192.0.2.3
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262143
---snip---
CSPF              : Enabled               Oper CSPF            : Enabled
Least Fill        : Disabled              Oper LeastFill       : Disabled
FRR               : Enabled               Oper FRR             : Enabled
FRR NodeProtect   : Enabled               Oper FRR NP          : Enabled
FR Hop Limit      : 16                    Oper FRHopLimit      : 16
---snip---
Actual Hops       :
    192.168.12.1 (192.0.2.1) @ n                  Record Label         : N/A
 -> 192.168.12.2 (192.0.2.2) @                    Record Label         : 262143
 -> 192.168.23.2 (192.0.2.3)                      Record Label         : 262142
---snip---
Detour Status     : Standby               Detour Type          : Originate
Detour Avoid Nod*: 192.0.2.2              Detour Origin        : 192.0.2.1
Setup Priority    : 7                     Hold Priority        : 0
Class Type        : 0
Detour Active Ti*: n/a                    Detour Up Time       : 0d 00:00:42
In Interface      : n/a                   In Label             : n/a
Out Interface     : 1/1/2                 Out Label            : 262143
NextHop           : 192.168.14.2
Explicit Hops     :
    192.168.14.1(S)
 -> 192.168.14.2(S)
 -> 192.168.45.2(S)
 -> 192.168.56.2(S)
 -> 192.168.36.1(S)
===============================================================================
```

The output also contains information about the detour tunnel originating in PE-1 that protects node PE-2. Since the detour tunnel is dedicated for this LSP, that information can be included in the LSP information. The out label is the same as for the preferred path.

The RSVP detour sessions can be retrieved in the originating, transit, and terminating nodes. On originating node PE-1:

```
*A:PE-1# show router rsvp session detour
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.3       2      55812 LSP-PE-1-PE-3::dyn_detour    Up
-------------------------------------------------------------------------------
Sessions : 1
```

In the transit/terminating node PE-5:

```
*A:PE-5# show router rsvp session detour-transit
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.3       2      55810 LSP-PE-1-PE-3::dyn_detour    Up
-------------------------------------------------------------------------------
Sessions : 1
===============================================================================
*A:PE-5# show router rsvp session detour-terminate
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.3       2      55810 LSP-PE-1-PE-3::dyn_detour    Up
-------------------------------------------------------------------------------
Sessions : 1
```

More detailed information can be retrieved as follows:

```
*A:PE-5# show router rsvp session detail
===============================================================================
RSVP Sessions (Detailed)
===============================================================================
-------------------------------------------------------------------------------
LSP : LSP-PE-1-PE-3::dyn_detour
-------------------------------------------------------------------------------
From            : 192.0.2.1             To             : 192.0.2.3
Tunnel ID       : 2                     LSP ID         : 55810
Style           : SE                    State          : Up
Session Type    : Transit (Detour)
In Interface    : 1/1/1                 Out Interface  : 1/1/2
In Label        : 262142                Out Label      : 262142
Previous Hop    : 192.168.45.1          Next Hop       : 192.168.56.2
---snip---
-------------------------------------------------------------------------------
LSP : LSP-PE-1-PE-3::dyn_detour
```

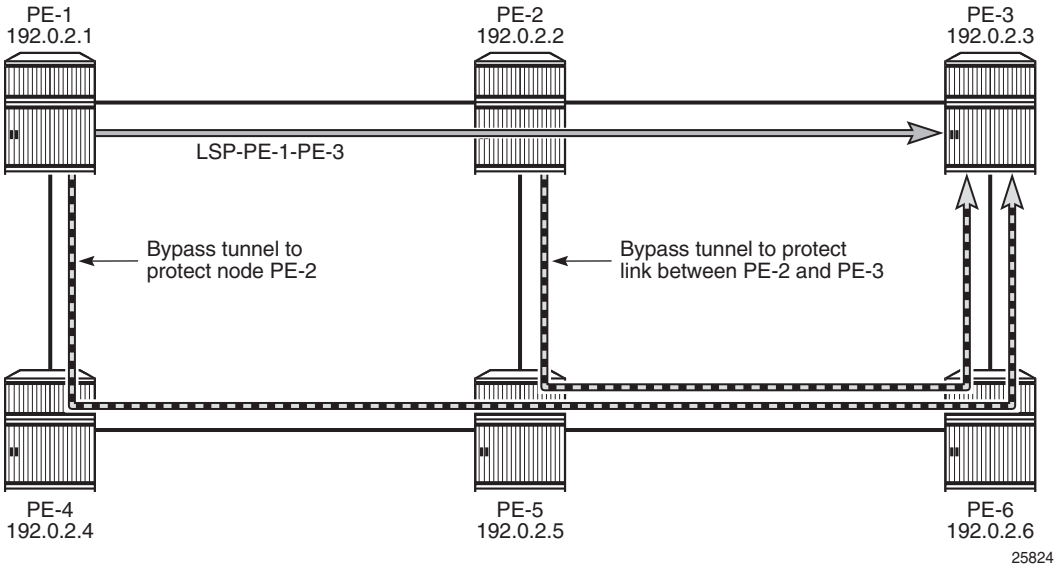
```
-------------------------------------------------------------------------------
From            : 192.0.2.1             To              : 192.0.2.3
Tunnel ID       : 2                     LSP ID          : 55810
Style           : SE                    State           : Up
Session Type    : Terminate (Detour)
In Interface    : 1/1/3                 Out Interface   : 1/1/2
In Label        : 262143                Out Label       : 262142
Previous Hop    : 192.168.25.1          Next Hop        : 192.168.56.2
---snip---
```

PE-5 is a transit node for the detour tunnel with previous hop PE-4 and a terminating node for the detour tunnel with previous hop PE-2. In both cases, the next hop is PE-6.

FRR Facility

The drawback of FRR one-to-one is that each LSP requires its own detour tunnels to be signaled. FRR facility does not have this issue, because it offers local repair for the next node or the next link that uses bypass tunnels that can be shared by LSPs. FRR facility bypass tunnels terminate in the merge point (MP), which is a hop in the primary path. FRR facility bypass tunnels for link protection terminate in the next hop in the primary path and FRR facility bypass tunnels for node protection terminate in the next hop of that next hop. FRR bypass tunnels are unaware of the final destination of the LSP and need not terminate in the final destination, but in this case they do, because the number of hops in the primary path is limited. Figure 180 shows the FRR facility bypass tunnels for LSP "LSP-PE-1-PE-3":

Figure 180 **Fast-Reroute Facility Bypass Tunnels**



Fast reroute facility is enabled on LSP "LSP-PE-1-PE-3" as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" fast-reroute facility
```

The LSP path detail output shows that there is a bypass tunnel available in PE-1 that offers node protection for the next node in the primary path: PE-2. In PE-2, there is a bypass tunnel offering link protection for the next link, which is the link between PE-2 and PE-3; as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
---snip---
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 55820
From              : 192.0.2.1           To                  : 192.0.2.3
Admin State       : Up                  Oper State          : Up
Path Name         : dyn                 Path Type           : Primary
Path Admin        : Up                  Path Oper           : Up
Out Interface     : 1/1/1               Out Label           : 262143
---snip---
CSPF              : Enabled             Oper CSPF           : Enabled
Least Fill        : Disabled            Oper LeastFill      : Disabled
FRR               : Enabled             Oper FRR            : Enabled
```

```
FRR NodeProtect  : Enabled              Oper FRR NP          : Enabled
---snip---
Actual Hops      :
    192.168.12.1 (192.0.2.1) @ n                Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2) @                  Record Label        : 262143
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262143
---snip---
```

In FRR facility mode, the bypass tunnels are shared. They are not included in the LSP information. The bypass tunnels can be retrieved as follows:

```
*A:PE-1# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-node192.0.2.2-61442
-------------------------------------------------------------------------------
To             : 192.168.36.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262143
Up Time        : 0d 00:19:28         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 40
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.14.1(S)                  Egress Admin Groups : None
 -> 192.168.14.2(S)                  Egress Admin Groups : None
 -> 192.168.45.2(S)                  Egress Admin Groups : None
 -> 192.168.56.2(S)                  Egress Admin Groups : None
 -> 192.168.36.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.14.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)         Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)         Record Label        : 262142
 -> 192.168.56.2 (192.0.2.6)         Record Label        : 262142
 -> 192.168.36.1 (192.0.2.3)         Record Label        : 262141
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-1-PE-3::dyn
From           : 192.0.2.1           To                  : 192.0.2.3
Avoid Node/Hop : 192.0.2.2           Downstream Label    : 262143
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-1#
```

This is the bypass tunnel that originates in PE-1 to protect (avoid) PE-2. In this example, there is only one LSP protected by this bypass tunnel, but the list of protected LSPs can be longer. The same command can be launched on PE-2, where a bypass tunnel originates that protects the link between PE-2 and PE-3.

The RSVP sessions can be displayed as follows:

```
*A:PE-3# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.3       2      55820 LSP-PE-1-PE-3::dyn            Up
192.0.2.2       192.168.36.1    61442  4     bypass-link192.168.23.2-614*  Up
192.0.2.1       192.168.36.1    61442  4     bypass-node192.0.2.2-61442    Up
-------------------------------------------------------------------------------
Sessions : 3
```

In PE-3, there is an RSVP session for the regular LSP and two bypass tunnels. In this case, the bypass tunnels all go to PE-3, which is the terminating node for the LSP, but that need not be the case. All bypass tunnels are signaled from the point of local repair to the merge point on the LSP path.

To force a FRR facility switchover to a bypass tunnel, a failure is simulated as follows:

```
*A:PE-2# configure port 1/1/1 shutdown
```

The detailed output for the LSP path on PE-1 shows that the tunnel is locally repaired.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
---snip---
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-3
---snip---
Failure Code     : tunnelLocallyRepaired
Failure Node     : 192.0.2.2
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.168.12.1 (192.0.2.1) @ n                Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2) @ #                Record Label        : 262141
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262141
---snip---
```

The failure code is tunnelLocallyRepaired and next to the actual hop 192.168.12.2 (PE-2), the symbol # indicates that the detour is in use.

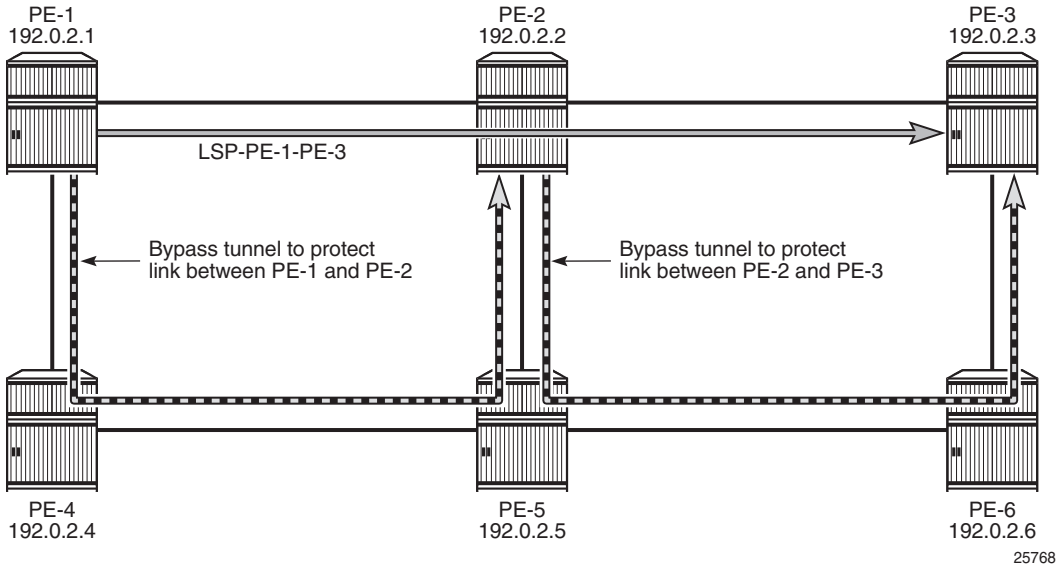
FRR Facility without Node Protection

Node protection is by default enabled, but can be disabled as follows:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-3"
                fast-reroute facility
                    no node-protect
                exit
```

As a result, there is only link protection. The bypass tunnels from PE-1 and PE-2 terminate in the next hop in the primary path, as shown in Figure 181:

Figure 181 **FRR Facility without Node Protection**



The LSP path detail output shows that there is no node protection (@):

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
---snip---
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3
Path LSP ID       : 46090
From              : 192.0.2.1             To                  : 192.0.2.3
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
```

```
Path Admin       : Up                    Path Oper            : Up
Out Interface    : 1/1/1                 Out Label            : 262140
---snip---
FRR              : Enabled               Oper FRR             : Enabled
FRR NodeProtect  : Disabled              Oper FRR NP          : Disabled
---snip---
Actual Hops      :
    192.168.12.1 (192.0.2.1) @                   Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2) @                   Record Label        : 262140
 -> 192.168.23.2 (192.0.2.3)                     Record Label        : 262139
---snip---
```

The bypass tunnel originating in PE-1 is now terminating in PE-2 instead of PE-3; as follows:

```
*A:PE-1# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.12.2-61443
-------------------------------------------------------------------------------
To             : 192.168.25.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262143
Up Time        : 0d 00:01:04         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 30
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.14.1(S)                  Egress Admin Groups : None
 -> 192.168.14.2(S)                  Egress Admin Groups : None
 -> 192.168.45.2(S)                  Egress Admin Groups : None
 -> 192.168.25.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.14.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)         Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)         Record Label        : 262140
 -> 192.168.25.1 (192.0.2.2)         Record Label        : 262139
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-1-PE-3::dyn
From           : 192.0.2.1           To                  : 192.0.2.3
Avoid Node/Hop : 192.168.12.2        Downstream Label    : 262140
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-1#
```

In the remainder of this chapter, this LSP is no longer used. Therefore, the LSP is shut down:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-3" shutdown
```

Administrative Groups for RSVP-TE LSPs

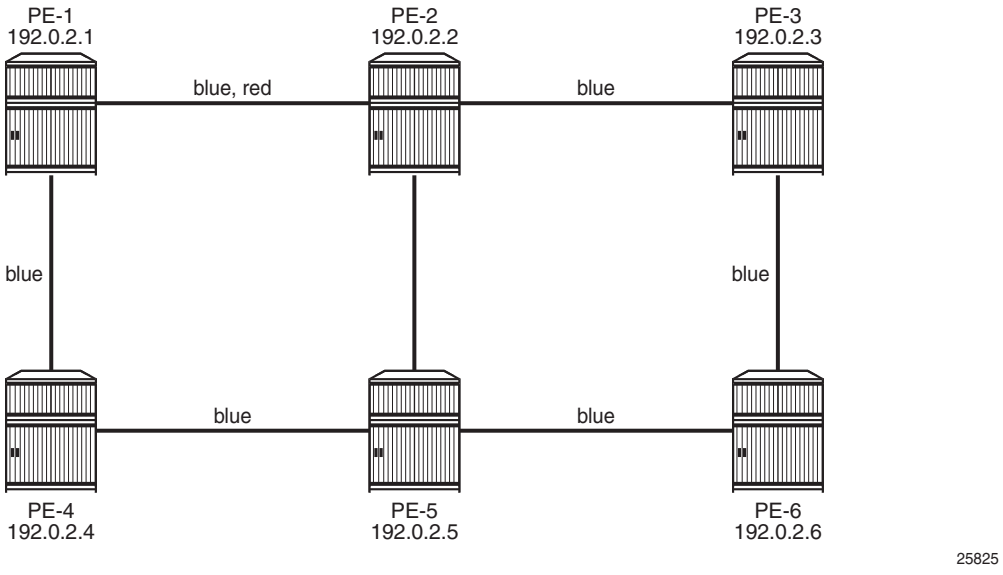
Administrative groups (link-coloring) can be used to calculate a path with the restriction to only include links of a particular admin group (color) or to exclude links of a particular admin group. Paths can be disjointed from each other, without the need for an explicit hops list.

Two admin groups are configured on all nodes; as follows:

```
*A:PE-1# configure router if-attribute admin-group "red" value 0
*A:PE-1# configure router if-attribute admin-group "blue" value 1
```

Admin group "blue" is assigned to all MPLS interfaces, except for the link between PE-2 and PE-5 while admin group "red" is only assigned to the link between PE-1 and PE-2; see Figure 182:

Figure 182 **Admin Groups 'Blue' and 'Red'**



The admin groups are assigned to the MPLS interfaces as follows:

```
*A:PE-1# configure router
        mpls
            interface "int-PE-1-PE-2"
                admin-group "blue"
                admin-group "red"
            exit
            interface "int-PE-1-PE-4"
                admin-group "blue"
            exit
```

The configuration on the other nodes is similar.

To ensure that FRR bypass tunnels will adhere to the same admin group constraints as defined in the LSP, the following is configured on all nodes. It is required on all points of local repair (PLRs):

```
*A:PE-1# configure router mpls admin-group-frr
```

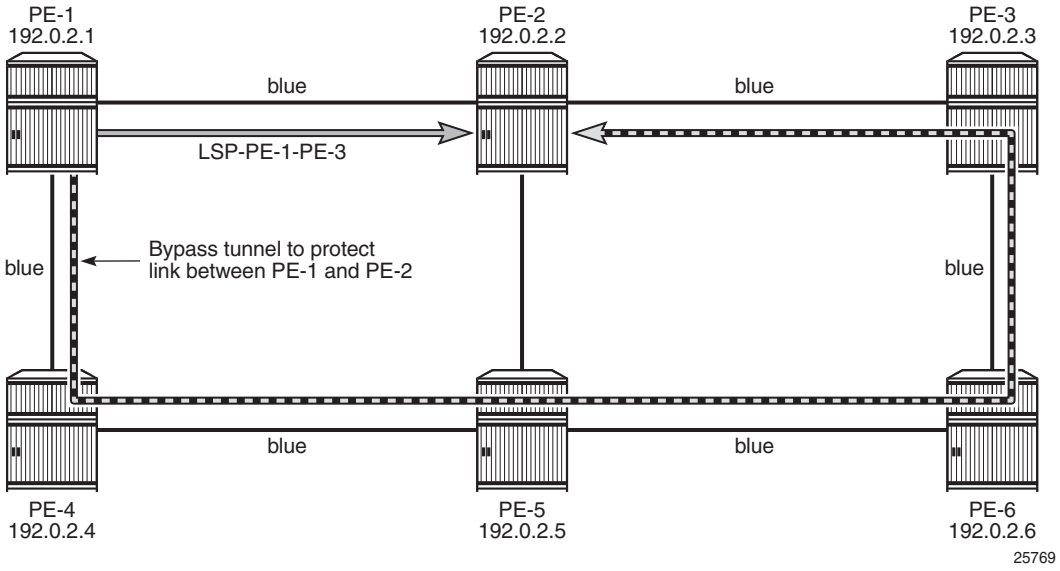
LSP Includes Admin Group 'blue'

LSP "LSP-PE-1-PE-2" is created on PE-1 with a dynamic primary path. FRR facility is enabled. The LSP includes admin group blue and both the primary path as the bypass tunnel must use links in admin group "blue" (propagate-admin-group). Admin-group-frr is enabled in the MPLS context, to ensure that the admin group restriction is respected for FRR.

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                cspf
                include "blue"
                propagate-admin-group
                fast-reroute facility
                    propagate-admin-group
                exit
                primary "dyn"
                exit
                no shutdown
            exit
```

The bypass tunnel cannot include the link between PE-2 and PE-5, since that link does not belong to admin group "blue". The LSP and its bypass tunnel are shown in Figure 183:

Figure 183 **LSP and Bypass within Admin Group 'Blue'**



The LSP path detailed information is as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
---snip---
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-2
Path LSP ID       : 25600
From              : 192.0.2.1             To                   : 192.0.2.2
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262143
---snip---
FR Prop Admin Gr*: Enabled                Oper FRPropAdmGrp    : Enabled
Propogate Adm Grp: Enabled                Oper Prop Adm Grp    : Enabled
---snip---
Include Groups    :                       Oper Include Groups  :
blue                                            blue
Exclude Groups    :                       Oper Exclude Groups  :
None                                            None
---snip---
Actual Hops       :
    192.168.12.1 (192.0.2.1) @                  Record Label         : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label         : 262143
---snip---
```

There is a bypass tunnel originating in PE-1 that offers protection for the link between PE-1 and PE-2. More information about this bypass tunnel can be retrieved as follows:

```
*A:PE-1# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.12.2-61446
-------------------------------------------------------------------------------
To             : 192.168.23.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262143
Up Time        : 0d 00:09:40         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 50
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.14.1(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.14.2(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.45.2(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.56.2(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.36.1(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.23.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.14.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)         Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)         Record Label        : 262143
 -> 192.168.56.2 (192.0.2.6)         Record Label        : 262143
 -> 192.168.36.1 (192.0.2.3)         Record Label        : 262143
 -> 192.168.23.1 (192.0.2.2)         Record Label        : 262142
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-1-PE-2::dyn
From           : 192.0.2.1           To                  : 192.0.2.2
Avoid Node/Hop : 192.168.12.2        Downstream Label    : 262143
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-1#
```

All egress links are in admin group blue on the originating and transit nodes.

LSP Excludes Admin Group 'red'

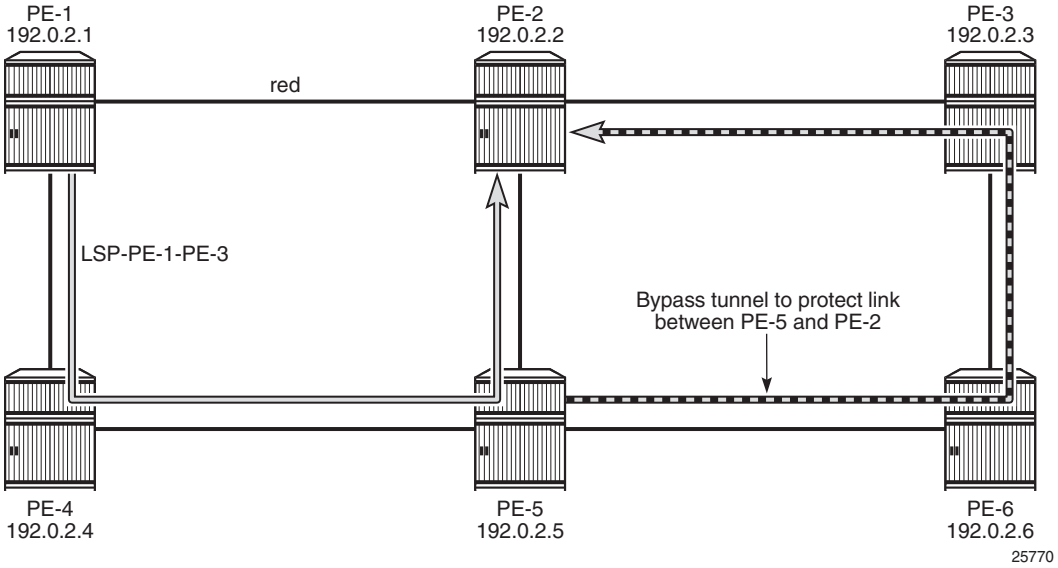
The LSP is reconfigured: instead of including admin group 'blue', it will exclude admin group 'red'. Nothing is changed to the configuration of FRR.

The MPLS configuration is modified as follows:

```
*A:PE-1# configure router
        mpls
            lsp "LSP-PE-1-PE-2"
                no include "blue"
                exclude "red"
            exit
```

The LSP cannot use the red link between PE-1 and PE-2. The path that avoids the red link, is from PE-1 via PE-4 and PE-5 to PE-2. On all PLRs, admin-group-frr is configured, which implies that the originating FRR bypass tunnels need to respect the admin-group constraint of the LSP. There can be no node protection for PE-4 or PE-5 without using the red link between PE-1 and PE-2. The only link that can be protected without using the red link between PE-1 and PE-2, is the link between PE-5 and PE-2. The LSP and the FRR bypass tunnel are shown in Figure 184:

Figure 184 **LSP and FRR Bypass Tunnel Excluding Admin Group 'Red'**



The LSP path can be verified as follows:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail
---snip---
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-2 Path dyn
```

```
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-1-PE-2
Path LSP ID      : 25610
From             : 192.0.2.1            To                  : 192.0.2.2
Admin State      : Up                   Oper State          : Up
---snip---
Include Groups   :                      Oper Include Groups  :
None                                           None
Exclude Groups   :                      Oper Exclude Groups  :
red                                            red
---snip---
Actual Hops      :
    192.168.14.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)                    Record Label        : 262142
 -> 192.168.45.2 (192.0.2.5) @                  Record Label        : 262142
 -> 192.168.25.1 (192.0.2.2)                    Record Label        : 262140
---snip---
```

There is only link protection for the link from PE-5 to PE-2. The bypass tunnel originates in PE-5 and has no links belonging to admin group 'red':

```
*A:PE-5# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.25.1-61460
-------------------------------------------------------------------------------
To             : 192.168.23.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262142
Up Time        : 0d 00:02:26         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 30
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.56.1(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.56.2(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.36.1(S)                  Egress Admin Groups :
                                     blue
 -> 192.168.23.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.56.1 (192.0.2.5)         Record Label        : N/A
 -> 192.168.56.2 (192.0.2.6)         Record Label        : 262142
 -> 192.168.36.1 (192.0.2.3)         Record Label        : 262142
 -> 192.168.23.1 (192.0.2.2)         Record Label        : 262139
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-1-PE-2::dyn
From           : 192.0.2.1           To                  : 192.0.2.2
Avoid Node/Hop : 192.168.25.1        Downstream Label    : 262140
Bandwidth      : 0 Kbps
```

```
===============================================================================
*A:PE-5#
```

This configuration is preserved for the following example.

Hop Limit for RSVP-TE LSPs

Another constraint to influence the path selection, is hop limit. This can be configured on the LSP, on a secondary path, or on FRR in case the path should not contain too many hops. In this example, it will be configured on the LSP and later also for FRR on that LSP. By default, the LSP hop limit is 255, but it can be configured as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" hop-limit 5
```

This hop limit of 5 is enough for the path via PE-4 and PE-5, but it will not be sufficient when the link between PE-2 and PE-5 is down:

```
*A:PE-5# configure port 1/1/3 shutdown
```

In this case, the only possible path that excludes the 'red' link between PE-1 and PE-2, has to go to PE-2 via PE-4, PE-5, PE-6, and PE-3. There are too many hops. The FRR bypass tunnel can do a local repair, but no new LSP path can be signaled, with failure code: noCspfRouteToDestination:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail
---snip---
Include Groups   :                      Oper Include Groups  :
None                                           None
Exclude Groups   :                      Oper Exclude Groups  :
red                                            red

Adaptive         : Enabled              Oper Metric          : 30
Preference       : n/a
Path Trans       : 5                    CSPF Queries         : 16
Failure Code     : tunnelLocallyRepaired
Failure Node     : 192.0.2.5
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.168.14.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)                    Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5) @ #                Record Label        : 262143
 -> 192.168.23.1 (192.0.2.2)                    Record Label        : 262143
---snip---
In Prog MBB :
 MBB Type        : GlobalRevert         Next Retry In        : 10 sec
 Started At      : 03/02/2016 13:47:18  Retry Attempt        : 12
 Failure Code    : noCspfRouteToDestina Failure Node         : 192.0.2.1
                   tion
```

```
 Signaled BW     : 0 Mbps
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

FRR tunnels also have a hop limit. The FRR hop limit is by default 16, but can be configured as follows:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" fast-reroute hop-limit 3
```

When the LSP is recalculated, it is impossible to establish the primary path with a hop limit of 5 and it is also impossible to establish a bypass tunnel protecting the link between PE-5 and PE-2 when the FRR hop limit is 3. The LSP will remain operationally down with failure code: noCspfRouteToDestination:

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-2
Path LSP ID       : 1036
From              : 192.0.2.1             To                   : 192.0.2.2
Admin State       : Up                    Oper State           : Down
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Down
Out Interface     : n/a                   Out Label            : n/a
Path Up Time      : 0d 00:00:00           Path Down Time       : 0d 00:00:52
---snip---
CSPF              : Enabled               Oper CSPF            : N/A
Least Fill        : Disabled              Oper LeastFill       : N/A
FRR               : Enabled               Oper FRR             : N/A
FRR NodeProtect   : Enabled               Oper FRR NP          : N/A
FR Hop Limit      : 3                     Oper FRHopLimit      : N/A
FR Prop Admin Gr* : Enabled               Oper FRPropAdmGrp    : N/A
Propogate Adm Grp : Enabled               Oper Prop Adm Grp    : N/A
Inter-area        : N/A

Neg MTU           : 0                     Oper MTU             : N/A
Bandwidth         : No Reservation        Oper Bandwidth       : N/A
Hop Limit         : 5                     Oper HopLimit        : N/A
---snip---
Include Groups    :                       Oper Include Groups  :
None                                            N/A
Exclude Groups    :                       Oper Exclude Groups  :
red                                             N/A
```

```
---snip---
Failure Code     : noCspfRouteToDestination
Failure Node     : 192.0.2.1
---snip---
```

For the remainder of the examples, FRR is disabled and the hop limit is restored to the default value, which is 255:

```
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" no fast-reroute
*A:PE-1# configure router mpls lsp "LSP-PE-1-PE-2" no hop-limit
```

The port is put in a no shutdown state:

```
*A:PE-5# configure port 1/1/3 no shutdown
```

Manual Resignal

Instead of waiting for the resignal timer to expire, one can manually trigger the resignal process.

The command to resignal the path "dyn" of LSP "LSP-PE-1-PE-2":

```
*A:PE-1# tools perform router mpls resignal lsp "LSP-PE-1-PE-2" path "dyn"
```

The command to resignal all RSVP LSPs originating at node PE-1:

*A:PE-1# tools perform router mpls resignal delay 0This command can only be launched after the resignal timer is configured.

```
*A:PE-1# configure router mpls resignal-timer
  - no resignal-timer
  - resignal-timer <minutes>
 <minutes>            : [30..10080]
```

The configuration timer is configured to 30 minutes as follows:

```
*A:PE-1# configure router mpls resignal-timer 30
```

Whenever an LSP is resignaled, the resignal timer is restarted.

LSP OAM

The LSP diagnostics are modeled after ICMP echo request/reply which provides a mechanism to detect data plane failures in MPLS LSPs. For a given FEC, LSP ping verifies whether the packet reaches the egress label edge router (LER).

```
*A:PE-1# oam lsp-ping "LSP-PE-1-PE-2"
LSP-PING LSP-PE-1-PE-2: 92 bytes MPLS payload
Seq=1, send from intf int-PE-1-PE-4, reply from 192.0.2.2
       udp-data-len=32 ttl=255 rtt=1.09ms rc=3 (EgressRtr)
---- LSP LSP-PE-1-PE-2 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 1.09ms, avg = 1.09ms, max = 1.09ms, stddev = 0.000ms
*A:PE-1#
```

In LSP traceroute mode, the packet is sent to the control plane of each transit label switched router (LSR) which performs various checks to see if it is actually a transit LSR for the path.

```
*A:PE-1# oam lsp-trace "LSP-PE-1-PE-2"
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 116 byte packets
1  192.0.2.4  rtt=0.679ms rc=8(DSRtrMatchLabel) rsc=1
2  192.0.2.5  rtt=1.10ms rc=8(DSRtrMatchLabel) rsc=1
3  192.0.2.2  rtt=1.88ms rc=3(EgressRtr) rsc=1
*A:PE-1#
*A:PE-1# oam lsp-trace "LSP-PE-1-PE-2" detail
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 116 byte packets
1  192.0.2.4  rtt=0.566ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.45.2 ifaddr=192.168.45.2 iftype=ipv4Numbered MRU=1564
           label[1]=262143 protocol=4(RSVP-TE)
2  192.0.2.5  rtt=2.00ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.25.1 ifaddr=192.168.25.1 iftype=ipv4Numbered MRU=1564
           label[1]=262143 protocol=4(RSVP-TE)
3  192.0.2.2  rtt=1.84ms rc=3(EgressRtr) rsc=1
*A:PE-1#
```

RSVP LSP Statistics

Statistics can be collected for RSVP LSPs. For each accounting record, a file ID is configured; as follows:

```
*A:PE-1# configure log
        file-id 2
            location cf1:
            rollover 5 retention 1
        exit
```

An accounting policy is configured for each record type; as follows:

```
*A:PE-1# configure log
```

```
        accounting-policy 2
            record combined-mpls-lsp-ingress
            to file 2
            no shutdown
        exit
```

The collection of statistics is enabled in the MPLS context as follows:

```
*A:PE-1# configure router
        mpls
            ingress-statistics
                lsp "LSP-PE-1-PE-2" sender 192.0.2.1
                    no shutdown
                    collect-stats
                    accounting-policy 2
                exit
            exit
```

To display the statistics, the following options are available for lsp-ingress-stats:

```
*A:PE-1# show router mpls lsp-ingress-stats
  - - lsp-ingress-stats [type <lsp-type>] [active] [template-
match <SessionNameString> [sender
    <ip-address>]]
  -  - lsp-ingress-stats lsp <lsp-name> sender <ip-address>
 <lsp-name>           : max 64 chars
 <ip-address>         : a.b.c.d
 <lsp-type>           : p2p|p2mp
 <active>             : match on all stats enabled lsp
 <template-match>     : match on p2p/p2mp stats template
 <SessionNameString>  : [Max 64 chars]
```

The following command retrieves the LSP ingress statistics for LSP "LPS-PE-1-PE-2" with sender 192.0.2.1:

```
*A:PE-1# show router mpls lsp-ingress-stats lsp "LSP-PE-1-PE-2" sender 192.0.2.1
===============================================================================
MPLS LSP Ingress Statistics
===============================================================================
-------------------------------------------------------------------------------
LSP Name      : LSP-PE-1-PE-2
Sender        : 192.0.2.1
-------------------------------------------------------------------------------
Collect Stats : Enabled                Accting Plcy. : 2
Adm State     : Up                     PSB Match     : False
FC BE
InProf Pkts   : 0                      OutProf Pkts  : 0
InProf Octets : 0                      OutProf Octets: 0
FC L2
InProf Pkts   : 0                      OutProf Pkts  : 0
InProf Octets : 0                      OutProf Octets: 0
FC AF
InProf Pkts   : 0                      OutProf Pkts  : 0
InProf Octets : 0                      OutProf Octets: 0
FC L1
InProf Pkts   : 0                      OutProf Pkts  : 0
```

```
InProf Octets : 0                       OutProf Octets: 0
FC H2
InProf Pkts   : 0                       OutProf Pkts  : 0
InProf Octets : 0                       OutProf Octets: 0
FC EF
InProf Pkts   : 0                       OutProf Pkts  : 0
InProf Octets : 0                       OutProf Octets: 0
FC H1
InProf Pkts   : 0                       OutProf Pkts  : 0
InProf Octets : 0                       OutProf Octets: 0
FC NC
InProf Pkts   : 0                       OutProf Pkts  : 0
InProf Octets : 0                       OutProf Octets: 0
===============================================================================
*A:PE-1#
```

Statistics can be cleared as follows:

```
*A:PE-1# clear router mpls lsp-ingress-stats 192.0.2.1 lsp "LSP-PE-1-PE-2"
```

Debug Tools

A wide range of debug tools are available which can be tuned to the specific information of importance for a certain troubleshooting task. In the **debug router mpls** context, the LSP object to trace or monitor can be selected by the following parameters:

- LSP name
- Source address of the LSP (the from parameter in the LSP definition)
- Termination point of the LSP (the to parameter in the LSP definition)
- Tunnel ID of the LSP
- LSP ID

```
A:PE-1# debug router rsvp
  - no rsvp
  - rsvp [lsp <lsp-name>] [sender <sender-address>] [endpoint <endpoint-address>]
         [tunnel-id <tunnel-id>] [lsp-id <lsp-id>] [interface <ip-int-name>]
 <lsp-name>           : [80 chars max]
 <sender-address>     : a.b.c.d
 <endpoint-address>   : a.b.c.d
 <tunnel-id>          : [0..4294967295]
 <lsp-id>             : [1..65535]
 <ip-int-name>        : [32 chars max]
 [no] event           + Enable/disable debugging for specific RSVP events
 [no] packet          + Enable/disable debugging for specific RSVP packets
A:PE-1# debug router mpls
  - mpls [lsp <lsp-name>] [sender <source-address>] [endpoint <endpoint-address>]
[tunnel-id <tunnel-id>] [lsp-id <lsp-id>]
  - no mpls
 <lsp-name>           : [80 chars max]
 <source-address>     : a.b.c.d
```

```
 <endpoint-address>   : a.b.c.d
 <tunnel-id>          : [0..4294967295]
 <lsp-id>             : [1..65535]
 [no] event           + Enable/disable debugging for specific MPLS events
```

In the **debug** command tree, the MPLS event type can be selected (tracing must be enabled):

```
*A:PE-1# debug router mpls lsp "LSP-PE-1-PE-2" event
  - event
  - no event
 [no] all             -  Enable/disable debugging for MPLS all
 [no] frr             -  Enable/disable debugging for MPLS frr
 [no] iom             -  Enable/disable debugging for MPLS iom
 [no] lsp-setup       -  Enable/disable debugging for MPLS lsp setup
 [no] mbb             -  Enable/disable debugging for MPLS mbb
 [no] misc            -  Enable/disable debugging for MPLS misc
 [no] pcc             -  Enable/disable debugging for MPLS PCC
 [no] xc              -  Enable/disable debugging for MPLS xc
```

As an example, the **all** keyword is entered, logging all MPLS events related to the selected LSP:

```
*A:PE-1# debug router mpls lsp "LSP-PE-1-PE-2" event all
```

The last step is to create a log container which will gather all MPLS debugging information according to the criteria set in the debug context. The **from debug-trace** parameter must be configured but there are several options where the different captured entries will be stored: console, a syslog server, SNMP, local file on the compact flash card, a temporary circular memory buffer, or the telnet/SSH session from which you are logged into the node.

The ID of the log container is a local number without any other significance.

```
A:PE-1# configure log log-id 2
A:PE-1>config>log>log-id# from debug-trace
A:PE-1>config>log>log-id# to
  - to console
  - to file <log-file-id>
  - to memory [<size>]
  - to session
  - to snmp [<size>]
  - to syslog <syslog-id>
 <console>            : keyword - specifies console as destination
 <syslog-id>          : [1..10]
 <snmp>               : keyword - specifies SNMP as destination
 <log-file-id>        : [1..99]
 <memory>             : keyword - specifies memory as destination
 <size>               : [50..3000]
 <session>            : keyword - specifies telnet session as destination
```

For this example, the temporary buffer (with adjustable size) is chosen and the log container is enabled (no shutdown).

```
A:PE-1>config>log>log-id# to memory
A:PE-1>config>log>log-id# no shutdown
```

All MPLS events related to the selected LSP are stored in the location (memory) specified. The content of this log container can be viewed through the **show log log-id 2** command. After disabling port 1/1/3 on PE-5, the LSP goes down and a new path needs to be calculated and signaled. FRR is disabled.

```
*A:PE-1# clear log 2
*A:PE-1# show log log-id 2
===============================================================================
Event Log 2
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=100   next event=34  (not wrapped)]
33 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP
Set operational state for LSP LSP-PE-1-PE-2 to Up, previous state is Down"
32 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: RTM
Add tunnel table entry for TunnelId 3  Dest 192.0.2.2
Owner RSVP  Pref 7  Metric 50  MTU 1560  LDPoRSVP Yes  VprnAutoBind Yes  PropTtl Yes
IgpShcut Yes BgpShcut No BgpTransTunn No ELICapable Yes BW 0
NHLFE[1]: outIf (3) outLbl (262143) nhAddr 192.168.14.2 nhIndex 81"
31 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LTN
Add PUSH for LspKey P2P: Session(192.0.2.2, 3, 192.0.2.1) Sender(192.0.2.1, 25626) P
HOP(0.0.0.0)
Primary NHLFE - OutLabel 262143 OutIfIndex 3 Next-hop 192.168.14.2"
30 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP
Set LspPath LSP-PE-1-PE-2::dyn(LspId 25626) as active path for LSP LSP-PE-1-PE-2"
29 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP
LSP LSP-PE-1-PE-2 has no active path. Set path LSP-PE-1-PE-
2::dyn(LspId 25626) as active"
28 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP
Primary path LSP-PE-1-PE-2::dyn(LspId 25626) is now operationally up for LSP LSP-PE-
1-PE-2"
27 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP Path
Set operational MTU for LspPath LSP-PE-1-PE-2::dyn(LspId 25626) to 1564"
26 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: Path : RTM
Process RTM New event for Pref 192.168.14.2/32
[1] NextHop 192.168.14.2 OutIfId 3"
25 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP Path
Label request sent for LspPath LSP-PE-1-PE-2::dyn(LspId 25626) -
 wait for retry interval to receive response"
24 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: Path
P2P: Session(192.0.2.2, 3, 192.0.2.1) Sender(192.0.2.1, 25626) PHOP(0.0.0.0)
ERO in label request to next-hop 192.168.14.2
ERO(1) -> IPv4Prefix 192.168.14.2/32, Strict
ERO(2) -> IPv4Prefix 192.168.45.2/32, Strict
```

```
ERO(3) -> IPv4Prefix 192.168.56.2/32, Strict
ERO(4) -> IPv4Prefix 192.168.36.1/32, Strict
ERO(5) -> IPv4Prefix 192.168.23.1/32, Strict"
23 2016/02/25 15:49:01.47 UTC MINOR: DEBUG #2001 Base MPLS
"MPLS: Path
P2P: Session(192.0.2.2, 3, 192.0.2.1) Sender(192.0.2.1, 25626) PHOP(0.0.0.0)
Using next hop 192.168.14.2 (OutIntf 3)"
---snip---
```

Conclusion

MPLS provides the capability to establish connection oriented paths over a connectionless network. The LSP offers a mechanism to engineer network traffic on constraint-based paths rather than the IGP shortest path. This can greatly improve network resiliency. In this chapter, the configuration of several RSVP LSP features is given together with the associated show output which can be used to verify and troubleshoot.

RSVP Signaled Point-to-Multipoint LSPs

This chapter provides information about RSVP signaled point-to-multipoint LSPs.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This feature is applicable to SR OS. This chapter was originally written for and tested on release7.0.R5, but the CLI in this edition corresponds to 14.0.R1. On all nodes involved with the LSP, at least chassis mode C is required. This means that modular systems should be equipped with IOM-2 line cards or higher.

Overview

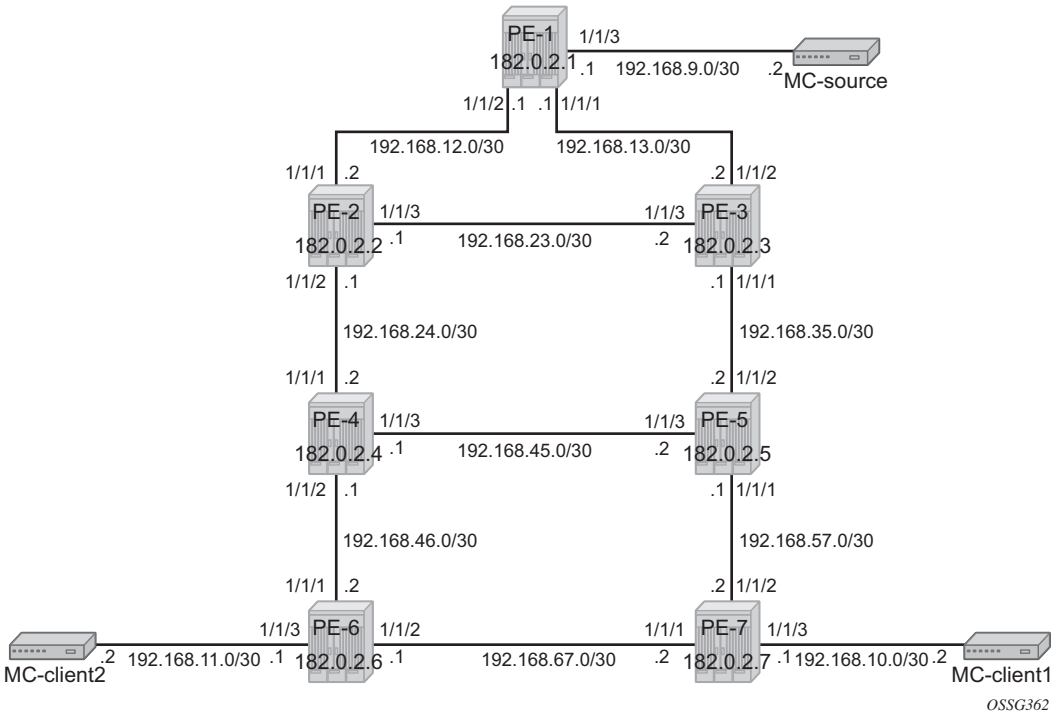
Point-to-MultiPoint (P2MP) MPLS label switched paths (LSP) allow the source of multicast traffic to forward packets to one or many multicast receivers over a network without requiring a multicast protocol, such as protocol independent multicast (PIM), to be configured in the network. A P2MP LSP tree is established in the control plane, and the path consists of a head-end node, one or many branch and bud nodes, and the leaf nodes. A bud node combines the roles of branch node and leaf node (for different source-to-leaf LSPs). Packets injected by the head-end node are replicated in the data plane at the branching nodes before they are delivered to the leaf nodes.

Similar to point-to-point (P2P) LSPs, also P2MP LSPs are unidirectional, originating on a head-end node (the ingress LER) and terminating on one or more leaf node(s) (the egress LER(s)). RSVP is used as signaling protocol. A P2MP LSP is modeled as a set of root-to-leaf sub LSPs (Source-to-Leaf: S2L). Each S2L is modeled as a point-to-point LSP in the control plane. This means that each S2L has its own PATH/RESV messages. This is called the de-aggregated method.

The forwarding of multicast packets to the LSP tree is based on static multicast routes initially, and will evolve to BGP based VPN routes in the future. Forwarding multicast packets is done over P2MP RSVP LSPs in the base router instance, but will evolve to VPRNs.

RSVP signaled P2MP LSPs can have fast reroute (FRR) enabled, the facility method (one-to-many) with link protection is supported.

Figure 185 **P2MP Network Topology**



Configuration

The following sections describe the tasks which must be performed to configure RSVP signaled point-to-multipoint LSPs.

Configuring the IP/MPLS Network

After configuring the cards and MDAs, the correct chassis-mode must be set (C or higher) on all MPLS nodes along the P2MP LSP.

```
A:PE-1# configure system chassis-mode

  - chassis-mode <chassis-mode> [force]
 <chassis-mode>       : a|b|c|d
                        Chassis Mode corresponds to scaling and feature sets associated
                        with a card.
                        mode 'a' corresponds to the iom-20g.
                        mode 'b' corresponds to the iom-20g-b.
                        mode 'c' corresponds to the iom2-20g.
                        mode 'd' corresponds to the iom3-xp.
 <force>             : keyword - forces an upgrade from a lesser scaling and feature
                        set to a greater one.
A:PE-1#

A:PE-1# configure system chassis-mode c

*A:PE-1# show chassis | match "chassis mode"
    Admin chassis mode            : c
    Oper chassis mode             : c
```

The system addresses and Layer 3 interface addresses are configured according to Figure 185. An interior gateway protocol (IGP) is needed to distribute routing information to all PEs. In this case, the IGP is OSPF using the backbone area 0.0.0.0. A configuration example is shown for PE-1. A similar configuration is needed on all PEs.

```
*A:PE-1# configure router
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/2
        exit
        interface "int-PE-1-PE-3"
            address 192.168.13.1/30
            port 1/1/1
        exit
        interface "system"
            address 192.0.2.1/32
        exit

*A:PE-1# configure router
        ospf
            traffic-engineering
            area 0.0.0.0
                interface "system"
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-1-PE-3"
                    interface-type point-to-point
                exit
            exit
        exit
```

Since fast reroute (FRR) is enabled for the P2MP LSP, traffic engineering (TE) is needed on the IGP. By doing this, OSPF will generate opaque LSAs which are collected in a traffic engineering database (TED), separate from the traditional OSPF topology database. OSPF interfaces are set up as type *'point-to-point'* to improve convergence, no designated router/backup designated router (DR/BDR) election process is done. However, convergence is out of the scope of this chapter.

To verify that OSPF neighbors are up (state **Full**), **show router ospf neighbor** is performed. To check if Layer 3 interface addresses/subnets are known on all PEs, **show router route-table** or **show router fib** *iom-card-slot* will display the content of the forwarding information base (FIB).

```
*A:PE-1# show router ospf neighbor
===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-PE-2                    192.0.2.2       Full       1    0       32
   0.0.0.0
int-PE-1-PE-3                    192.0.2.3       Full       1    0       31
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 2
===============================================================================
*A:PE-1#


*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix                                    Type    Proto    Age          Pref
       Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                   Local   Local    03h25m22s   0
       system                                                          0
192.0.2.2/32                                   Remote  OSPF     02h08m41s   10
       192.168.12.2                                                    10
192.0.2.3/32                                   Remote  OSPF     03h24m21s   10
       192.168.13.2                                                    10
192.0.2.4/32                                   Remote  OSPF     02h08m28s   10
       192.168.12.2                                                    20
192.0.2.5/32                                   Remote  OSPF     02h05m36s   10
       192.168.12.2                                                    30
192.0.2.6/32                                   Remote  OSPF     01h49m05s   10
       192.168.12.2                                                    30
192.0.2.7/32                                   Remote  OSPF     02h05m36s   10
       192.168.12.2                                                    40
192.168.12.0/30                                Local   Local    03h25m22s   0
       int-PE-1-PE-2                                                   0
192.168.13.0/30                                Local   Local    03h25m22s   0
       int-PE-1-PE-3                                                   0
```

```
192.168.23.0/30                                 Remote  OSPF     02h08m28s   10
       192.168.12.2                                                  20
192.168.24.0/30                                 Remote  OSPF     02h08m28s   10
       192.168.12.2                                                  20
192.168.35.0/30                                 Remote  OSPF     02h05m36s   10
       192.168.12.2                                                  40
192.168.45.0/30                                 Remote  OSPF     02h08m28s   10
       192.168.12.2                                                  30
192.168.46.0/30                                 Remote  OSPF     01h49m10s   10
       192.168.12.2                                                  30
192.168.57.0/30                                 Remote  OSPF     02h05m36s   10
       192.168.12.2                                                  40
192.168.67.0/30                                 Remote  OSPF     01h49m05s   10
       192.168.12.2                                                  40
-------------------------------------------------------------------------------
No. of Routes: 16
===============================================================================
*A:PE-1#

*A:PE-1# show router fib 1
===============================================================================
FIB Display
===============================================================================
Prefix                                                      Protocol
    NextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                                LOCAL
    192.0.2.1 (system)
192.0.2.2/32                                                OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.3/32                                                OSPF
    192.168.13.2 (int-PE-1-PE-3)
192.0.2.4/32                                                OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.5/32                                                OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.6/32                                                OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.7/32                                                OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.12.0/30                                             LOCAL
    192.168.12.0 (int-PE-1-PE-2)
192.168.13.0/30                                             LOCAL
    192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.35.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.45.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.46.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.57.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
192.168.67.0/30                                             OSPF
    192.168.12.2 (int-PE-1-PE-2)
-------------------------------------------------------------------------------
```

```
Total Entries : 16
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

On PE-1, the interface toward the multicast source is configured in an IES service. This could have been on a router interface instead.

```
*A:PE-1# configure service
        ies 1 customer 1 create
            interface "int-PE-1-MC-source" create
                address 192.168.9.1/30
                sap 1/1/3 create
                exit
            exit
            no shutdown
        exit
```

Similar IES services are configured on PE-7 and PE-6 for multicast client 1 and multicast client 2.

```
*A:PE-7# configure service
        ies 1 customer 1 create
            interface "int-PE-7-MC-client1" create
                address 192.168.10.1/30
                sap 1/1/3 create
                exit
            exit
            no shutdown
        exit

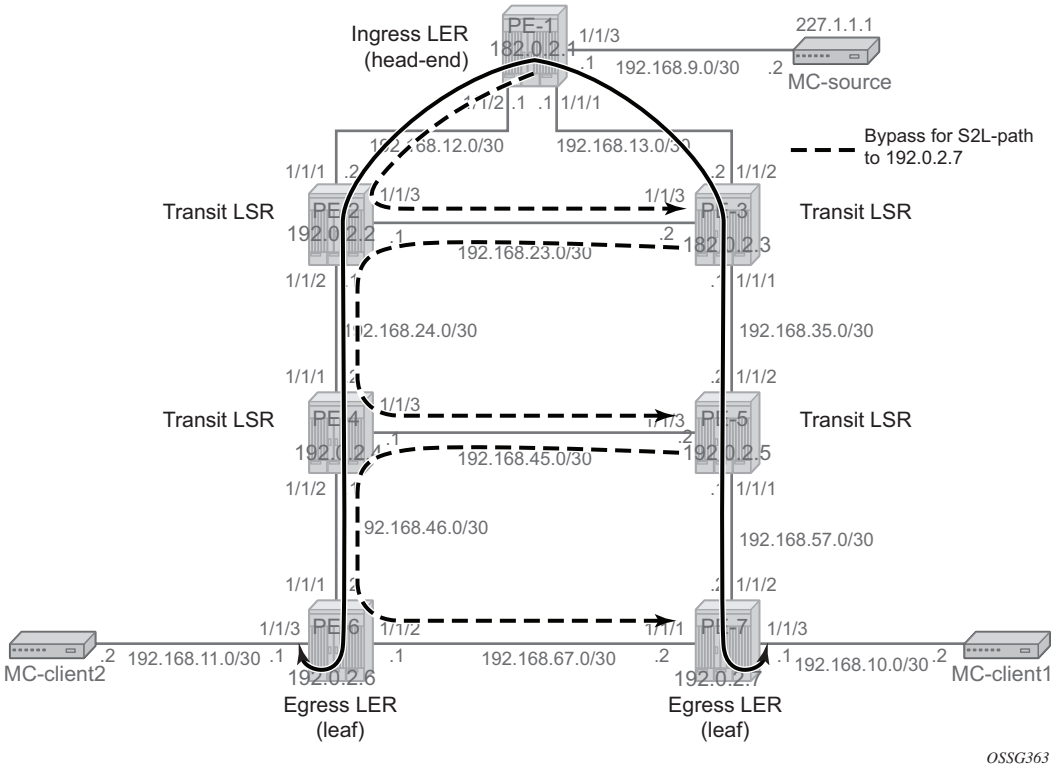
*A:PE-6# configure service
        ies 1 customer 1 create
            interface "int-PE-6-MC-client2" create
                address 192.168.11.1/30
                sap 1/1/3 create
                exit
            exit
            no shutdown
        exit
```

The next step in the process of setting up a P2MP LSP, is enabling the L3 interfaces in the MPLS and RSVP context on all involved PE nodes (from PE-1 to PE-7). By default, the system interface is put automatically within the MPLS/RSVP context. When an interface is put in the MPLS context, 7x50 copies it also in the RSVP context. Explicit enabling of MPLS and RSVP context is done by the **no shutdown** command. The MPLS/RSVP configuration for PE-1 is as follows:

```
*A:PE-1# configure router mpls no shutdown
*A:PE-1# configure router rsvp no shutdown
*A:PE-1#
```

```
*A:PE-1# configure router mpls
            interface "system"
            exit
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-3"
            exit
```

Figure 186 **P2MP LSP LSP-p2mp-1**



Configuring P2MP RSVP LSP

See Figure 186 for P2MP LSP LSP-p2mp-1 with facility backup.

A P2MP LSP (LSP-p2mp-1) will be set up from PE-1 acting as head-end node and PE-6 and PE-7 acting as leaf nodes. Because FRR is enabled, constrained shortest path first (CSPF) is enabled to do route calculations on the traffic engineering database (TED). FRR method **facility** is used without node protection, **facility** stands for one-to-many, meaning that one bypass tunnel can protect a set of primary LSPs with similar backup constraints. When a link failure occurs on one of the active S2L paths, the point of local repair (PLR) node will push an additional MPLS label on the incoming MPLS packet before sending it into the bypass tunnel downstream toward the merge point (MP) node.

In the first example, the IGP OSPF will do the path calculation to the two destinations PE-6 and PE-7. The intermediate hops of the LSP are dynamically assigned by OSPF best route selection, thus the S2L paths follow the IGP least cost path. Therefore, an MPLS path called “**loose”** is configured without specifying any hops.

```
*A:PE-1# configure router mpls
            path "loose"
                no shutdown
            exit
```

Creation of the P2MP LSP itself is done on the ingress LER or head-end node (PE-1 in the example) and can be seen in following CLI output. The name of the P2MP LSP is “**LSP-p2mp-1”**. A create time keyword **p2mp-lsp** is added in addition to the P2MP name to make a distinction in configuration between normal point-to-point LSPs and point-to-multipoint LSPs. A primary P2MP instance is initiated using the **primary-p2mp-instance** keyword accompanied with the P2MP instance name “p-LSP-p2mp-1”. Within this primary P2MP instance, the different S2Ls are defined using the **s2l-path** keyword. Be aware that the same MPLS path name can be used for different S2Ls as long as the destination is different (**to** command).

```
*A:PE-1# configure router mpls
            lsp "LSP-p2mp-1" p2mp-lsp
                cspf
                fast-reroute facility
                    no node-protect
                exit
                primary-p2mp-instance "p-LSP-p2mp-1"
                    s2l-path "loose" to 192.0.2.6
                    exit
                    s2l-path "loose" to 192.0.2.7
                    exit
                exit
                no shutdown
            exit
```

On the head-end LER node of the P2MP LSP, several show commands can be used. A first set of show commands is used to verify the administrative and operational state of the P2MP LSP and its different S2L paths (including FRR bypass information). In this example, ”LSP-p2mp-1” P2MP LSP has two active S2L paths: one toward leaf node PE-6 and one to leaf node PE-7.

```
*A:PE-1# show router mpls p2mp-lsp
===============================================================================
MPLS P2MP LSPs (Originating)
===============================================================================
LSP Name                                                Tun     Fastfail  Adm  Opr
                                                        Id      Config
-------------------------------------------------------------------------------
LSP-p2mp-1                                              1       Yes       Up   Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" detail
===============================================================================
MPLS P2MP LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-1
LSP Type    : P2mpLsp                       LSP Tunnel ID  : 1
LSP Index   : 1                             TTM Tunnel Id  : 1
From        : 192.0.2.1
Adm State   : Up                            Oper State     : Up
LSP Up Time : 0d 00:00:18                   LSP Down Time  : 0d 00:00:00
Transitions : 1                             Path Changes   : 1
Retry Limit : 0                             Retry Timer    : 30 sec
Signaling   : RSVP                          Resv. Style    : SE
Hop Limit   : 255                           Negotiated MTU : n/a
Adaptive    : Enabled                       ClassType      : 0
FastReroute : Enabled                       Oper FR        : Enabled
FR Method   : Facility                      FR Hop Limit   : 16
FR Node Pro*: Disabled                      FR Prop Adm Grp: Disabled
FR Object   : Enabled
CSPF        : Enabled                       ADSPEC         : Disabled
Metric      : Disabled                      Use TE metric  : Disabled
Load Bal Wt : N/A                           ClassForwarding: Disabled
Include Grps:                               Exclude Grps   :
None                                        None
Least Fill  : Disabled

Revert Timer: Disabled                      Next Revert In : N/A
Auto BW     : Disabled
LdpOverRsvp : Disabled
VprnAutoBind: Disabled
IGP Shortcut: Disabled                      BGP Shortcut   : Disabled
IGP LFA     : Disabled                      IGP Rel Metric : Disabled
BGPTransTun : Disabled
Oper Metric : Disabled
Prop Adm Grp: Disabled

P2MPInstance: p-LSP-p2mp-1                  P2MP-Inst-type : Primary
S2L Cfg Cou*: 2                             S2L Oper Count*: 2
```

```
S2l-Name    : loose                             To             : 192.0.2.6
S2l-Name    : loose                             To             : 192.0.2.7
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
*A:PE-1# show router mpls p2mp-info
===============================================================================
MPLS P2MP Cross Connect Information
===============================================================================
-------------------------------------------------------------------------------
S2L LSP-p2mp-1::loose
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 1
P2MP ID              : 0                     Lsp ID        : 31232
S2L Name             : LSP-p2mp-1::loose     To            : 192.0.2.6
Out Interface        : 1/1/2                 Out Label     : 262143
Num. of S2ls         : 1
-------------------------------------------------------------------------------
S2L LSP-p2mp-1::loose
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 1
P2MP ID              : 0                     Lsp ID        : 31232
S2L Name             : LSP-p2mp-1::loose     To            : 192.0.2.7
Out Interface        : 1/1/1                 Out Label     : 262143
Num. of S2ls         : 1
-------------------------------------------------------------------------------
P2MP Cross-connect instances : 2
===============================================================================
*A:PE-1#
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1"
===============================================================================
MPLS P2MP Instance (Originating)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-1
P2MP ID     : 0                                 LSP Tunnel ID  : 1
Adm State   : Up                                Oper State     : Up

P2MPInstance: p-LSP-p2mp-1                      P2MP-Inst-type : Primary
P2MP Inst Id: 1                                 P2MP Lsp Id    : 31232
Inst Admin  : Up                                Inst Oper      : Up
Inst Up Time: 0d 00:05:45                       Inst Dn Time   : 0d 00:00:00
Hop Limit   : 255                               Adaptive       : Enabled
Record Route: Record                            Record Label   : Record
Include Grps:                                   Exclude Grps   :
None                                            None
Bandwidth   : No Reservation                    Oper Bw        : 0 Mbps
S2l-Name    : loose                             To             : 192.0.2.6
S2l Admin   : Up                                S2l Oper       : Up
S2l-Name    : loose                             To             : 192.0.2.7
S2l Admin   : Up                                S2l Oper       : Up
-------------------------------------------------------------------------------
P2MP instances : 1
===============================================================================
*A:PE-1#
```

Note: As long as one S2L path is operationally up (show router mpls p2mp-lsp lsp-name p2mp-instance instance-name) , the Oper State of the P2MP LSP is Up.

FRR information can be displayed in detail for each S2L path. From this moment onward, the focus is on the S2L path toward PE-7. As can be seen in the show command, link protection is present for links PE-1 <=> PE-3, PE-3 <=> PE-5 and PE-5 <=> PE-7 ('@'-reference inside show command).

```
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l loose
to 192.0.2.7 detail
===============================================================================
MPLS LSP LSP-p2mp-1 S2L loose (Detail)
===============================================================================
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-1 S2L loose
-------------------------------------------------------------------------------
LSP Name     : LSP-p2mp-1                     S2l LSP ID  : 31232
P2MP ID      : 0                              S2l Grp Id  : 2
Adm State    : Up                             Oper State  : Up
S2l State:   : Active                                     :
S2L Name     : loose                          To          : 192.0.2.7
S2l Admin    : Up                             S2l Oper    : Up
OutInterface: 1/1/1                           Out Label   : 262143
S2L Up Time  : 0d 00:08:07                    S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                               NextRetryIn : 0 sec
S2L Trans    : 1                              CSPF Queries: 1
Failure Code: noError                         Failure Node: n/a
Inter-area   : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.13.1 (192.0.2.1) @                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @                Record Label        : 262143
 -> 192.168.35.2 (192.0.2.5) @                Record Label        : 262143
 -> 192.168.57.2 (192.0.2.7)                  Record Label        : 262143
ComputedHops:
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.35.2(S)
 -> 192.168.57.2(S)
LastResignal: n/a
===============================================================================
*A:PE-1#
```

More in detail, **show router mpls bypass-tunnel** can be used. **Actual Hops** provides the explicit hops of the bypass tunnel used to avoid link PE-1 <=> PE-3. On node PE-1 the MPLS path PE-1 <=> PE-2 <=> PE-3 is followed (see also Figure 186).

```
*A:PE-1# show router mpls bypass-tunnel detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.13.2-61441
-------------------------------------------------------------------------------
To             : 192.168.23.2        State               : Up
Out I/F        : 1/1/2               Out Label           : 262141
Up Time        : 0d 00:09:51         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : P2mp                Bypass Path Cost    : 20
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.12.1(S)                  Egress Admin Groups : None
 -> 192.168.12.2(S)                  Egress Admin Groups : None
 -> 192.168.23.2(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.12.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)         Record Label        : 262141
 -> 192.168.23.2 (192.0.2.3)         Record Label        : 262142

Protected LSPs -
LSP Name       : LSP-p2mp-1::loose
From           : 192.0.2.1           To                  : 192.0.2.7
Avoid Node/Hop : 192.168.13.2        Downstream Label    : 262143
Bandwidth      : 0 Kbps
-------------------------------------------------------------------------------
---snip---
```

On node PE-3, the MPLS path PE-3 <=> PE-2 <=> PE-4 <=> PE-5 is followed (see P2MP LSP LSP-p2mp-1) to avoid link PE-3 <=> PE-5.

```
*A:PE-3# show router mpls bypass-tunnel protected-lsp p2mp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.35.2-61441
-------------------------------------------------------------------------------
To             : 192.168.45.2        State               : Up
Out I/F        : 1/1/3               Out Label           : 262142
Up Time        : 0d 00:13:14         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : P2mp                Bypass Path Cost    : 30
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.23.2(S)                  Egress Admin Groups : None
```

```
 -> 192.168.23.1(S)                  Egress Admin Groups : None
 -> 192.168.24.2(S)                  Egress Admin Groups : None
 -> 192.168.45.2(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.23.2 (192.0.2.3)         Record Label        : N/A
 -> 192.168.23.1 (192.0.2.2)         Record Label        : 262142
 -> 192.168.24.2 (192.0.2.4)         Record Label        : 262141
 -> 192.168.45.2 (192.0.2.5)         Record Label        : 262142

Protected LSPs -
LSP Name       : LSP-p2mp-1::loose
From           : 192.0.2.1           To                  : 192.0.2.7
Avoid Node/Hop : 192.168.35.2        Downstream Label    : 262143
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-3#
```

A similar output can be seen on PE-5 node also. The MPLS path PE-5 <=> PE-4 <=> PE-6 <=> PE-7 is followed (see also Figure 186).

On the transit LSRs and egress LER/leaf node (see also Figure 186), the **show router mpls p2mp-info** command can be used. See the show command on node PE-3 for the S2L path to 192.0.2.7. Similar outputs are possible for nodes PE-5 and PE-7.

```
*A:PE-3# show router mpls p2mp-info
  - p2mp-info [type {originate|transit|terminate}] [s2l-endpoint <ip-address>]

 <originate|transit*> : keywords
 <ip-address>         : [a.b.c.d]

*A:PE-3# show router mpls p2mp-info
===============================================================================
MPLS P2MP Cross Connect Information
===============================================================================
-------------------------------------------------------------------------------
S2L LSP-p2mp-1::loose
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 1
P2MP ID              : 0                     Lsp ID        : 31232
S2L Name             : LSP-p2mp-1::loose     To            : 192.0.2.7
Out Interface        : 1/1/1                 Out Label     : 262143
Num. of S2ls         : 1
-------------------------------------------------------------------------------
P2MP Cross-connect instances : 1
===============================================================================
*A:PE-3#
```

Mapping Multicast Traffic

To map multicast traffic into the LSP tree from the head-end node until leaf node, PIM and internet group management protocol (IGMP) configurations are needed on the head-end node (PE-1) and leaf nodes (PE-6 and PE-7) of the P2MP RSVP LSP. The intermediate nodes (transit LSR or branch LSR) do not need any explicit configuration for that.

Head-end Node (Ingress LER)

PIM must be enabled on the interface toward the MC source and PIM must be enabled on the tunnel interface. A tunnel interface should be seen as an internal representation of a specific P2MP LSP. Creation is done within the PIM context using the **tunnel-interface rsvp-p2mp** command followed by the P2MP LSP name. This is configured as follows:

```
*A:PE-1# configure router pim
            interface "int-PE-1-MC-source"
            exit
            tunnel-interface rsvp-p2mp "LSP-p2mp-1"
```

For multicast packets received on an interface to pass through the data plane, a successful reverse path forwarding (RPF) check must be done on the source address otherwise the packet will be dropped.

Besides enabling PIM on the tunnel interface, also IGMP is enabled to do a static <S,G> or <*,G> join of a multicast group address (227.1.1.1 in the example) to the tunnel interface/P2MP LSP. Be aware that there is always a one-to-one mapping between <S,G> or <*,G> and a tunnel interface/P2MP LSP. In the example a < S,G > will be configured. A <*,G> join scenario is included in Additional Topics.

```
*A:PE-1# configure router igmp        …
            tunnel-interface rsvp-p2mp "LSP-p2mp-1"
                static
                    group 227.1.1.1
                        source 192.168.9.2
                    exit
                exit
            exit
            no shutdown
```

The **show router pim tunnel-interface** command shows you the admin state of the tunnel interface and an association to an internal local ifindex (**73728** in the example).

```
*A:PE-1# show router pim tunnel-interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                       Originator Address   Adm  Opr  Transport Type
-------------------------------------------------------------------------------
mpls-if-73728                   N/A                  Up   Up   Tx-IPMSI
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
*A:PE-1#
```

The **show router igmp group** command provides the configured <S,G> entry and outgoing interface (= tunnel interface), represented by mpls-if-73728.

```
*A:PE-1# show router igmp group 227.1.1.1
===============================================================================
IGMP Interface Groups
===============================================================================

(192.168.9.2,227.1.1.1)                                  UpTime: 0d 00:11:10
    Fwd List  : mpls-if-73728
-------------------------------------------------------------------------------
Entries : 1
===============================================================================
IGMP Host Groups
===============================================================================
No Matching Entries
===============================================================================
IGMP SAP Groups
===============================================================================
No Matching Entries
===============================================================================
*A:PE-1#
```

Users can verify if multicast traffic is using P2MP LSP at the head-end node using the **show router pim group** *group-address* **detail** command.

```
*A:PE-1# show router pim group 227.1.1.1 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 227.1.1.1
Source Address     : 192.168.9.2
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
MRIB Next Hop      : 192.168.9.2
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:02:42        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No
```

```
Rpf Neighbor       : 192.168.9.2
Incoming Intf      : int-PE-1-MC-source
Outgoing Intf List : mpls-if-73728

Curr Fwding Rate   : 5357.3 kbps
Forwarded Packets  : 124213             Discarded Packets  : 0
Forwarded Octets   : 5713798            RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

Leaf Node (Egress LER)

In the PIM context, the same tunnel interface must be created as the head-end node. An explicit reference to the head-end system address, using the **sender** *systemIP_head-end_node* parameter is needed.

```
*A:PE-7# configure router pim
           tunnel-interface rsvp-p2mp LSP-p2mp-1 sender 192.0.2.1
```

The **show router pim tunnel-interface** command provides the admin state of the tunnel interface and an association to an internal local ifindex (73728 in this example, by coincidence the same ifindex as the one on the head-end node/PE-1).

```
*A:PE-7# show router pim tunnel-interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                       Originator Address   Adm  Opr  Transport Type
-------------------------------------------------------------------------------
mpls-if-73728                   N/A                  Up   Up   Tx-IPMSI
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
*A:PE-7#
```

The main goal on the leaf node(s) is to get traffic off the P2MP LSP/tunnel interface. This is done using a multicast information policy (*multicast-info-policy*). Inside this MC policy, a range of multicast group addresses must be defined under a bundle context *(bundle1*) in order to see traffic *(channel*). Also inside the bundle context, the P2MP LSP is presented by the tunnel interface *(primary-tunnel-interface*). This is configured as follows:

```
*A:PE-7# configure mcast-management
```

```
multicast-info-policy "p2mp-pol" create
    bundle "bundle1" create
        primary-tunnel-interface rsvp-p2mp LSP-p2mp-1 sender 192.0.2.1
        channel "227.1.1.1" "227.1.1.1" create
        exit
    exit
    bundle "default" create
    exit
exit
```



Note: The **channel** command must be seen as a range command with a start-mc-group-address and an end-mc-group-address. In this example, only one MC group address, 227.1.1.1 is seen.

The configured multicast information policy must be applied to the base router instance.

```
*A:PE-7# configure router multicast-info-policy "p2mp-pol"
```

On the leaf node (PE-7/PE-6), MC clients are connected. IGMP is enabled on those MC clients with a static <S,G> join to redirect MC traffic downstream to the MC client. This is configured as follows:

```
*A:PE-7# configure router igmp
            interface "int-PE-7-MC-client1"
                static
                    group 227.1.1.1
                        source 192.168.9.2
                    exit
                exit
            exit
```

The **show router igmp group** provides the configured <S,G> entry and outgoing interface “int-PE-7-MC-client1”.

```
*A:PE-7# show router igmp group 227.1.1.1
===============================================================================
IGMP Interface Groups
===============================================================================

(192.168.9.2,227.1.1.1)                                 UpTime: 0d 00:09:02
    Fwd List  : int-PE-7-MC-client1
-------------------------------------------------------------------------------
Entries : 1
===============================================================================
IGMP Host Groups
===============================================================================
No Matching Entries
===============================================================================
IGMP SAP Groups
```

```
===============================================================================
No Matching Entries
===============================================================================
*A:PE-7#
```

Now, users can verify if multicast traffic is sent to the MC client using the **show router pim group** *group-address* **detail** command

```
*A:PE-7# show router pim group 227.1.1.1 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 227.1.1.1
Source Address     : 192.168.9.2
RP Address         : 0
Advt Router        :
Flags              :                    Type               : (S,G)
MRIB Next Hop      :
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:01:39        Resolved By        : unresolved

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:21
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : mpls-if-73728
Outgoing Intf List : int-PE-7-MC-client1

Curr Fwding Rate   : 4590.8 kbps
Forwarded Packets  : 226349             Discarded Packets  : 0
Forwarded Octets   : 10412054           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-7#
```

OAM Tools

P2P LSP operation and maintenance (OAM) commands (**oam lsp-ping** and **oam lsp-trace**) are extended for P2MP LSP. The user can instruct the head-end node to generate an P2MP LSP ping or a P2MP LSP trace by entering the command **oam p2mp-lsp-ping** or **oam p2mp-lsp-trace**. The P2MP OAM extensions are defined in **draft-ietf-mpls-p2mp-lsp-ping**.

For P2MP LSP ping, the echo request is sent on the active P2MP instance and replicated in the data path over all branches of the P2MP LSP instance. By default, all egress LER nodes which are leaves of the P2MP LSP instance will reply. Echo reply messages can be reduced by configuring the **s2l-dest-address** (a maximum of up to five egress nodes in a single run of the OAM command). Replies are sent by IP.

```
*A:PE-1# oam p2mp-lsp-ping
  - p2mp-lsp-ping {<lsp-name> [p2mp-instance <instance-name> [s2l-dest-address
                  <ipv4-  address> [... up to 5]]] [ttl <label-ttl>]}
  - p2mp-lsp-ping {ldp <p2mp-identifier> [sender-addr <ipv4-address>] [leaf-addr
                  <ipv4-address> [... up to 5]]}
  - p2mp-lsp-ping {ldp-ssm source <ip-address> group <ip-address> [router
                  <router-instance>|service-name <service-name>][sender-addr
                  <ipv4-address>] [leaf-addr <ipv4-address> [... up to 5]]}
  - options common to all p2mp-lsp-ping cases:  [fc <fc-name> [profile {in|out}]]
                  [size <octets>][timeout <timeout>] [detail]

 <lsp-name>           : [64 chars max]
 <instance-name>      : [32 chars max]
 <ipv4-address>       : a.b.c.d
 <in|out>             : in|out - Default: out
 <fc-name>            : be|l2|af|l1|h2|ef|h1|nc - Default: be
 <octets>             : [1..9198] - Default: 1
 <label-ttl>          : [1..255] - Default: 255
 <timeout>            : [1..120] seconds - Default: 10
 <detail>             : keyword - displays detailed information
 <p2mp-identifier>    : [1..4294967295]
 <ldp-ssm>            : keyword
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <router-instance>    : <router-name>|<service-id>
                        router-name    - "Base"|"management"  Default - Base
                        service-id     - [1..2147483647]
 <service-name>       : [64 chars max]

*A:PE-1#


*A:PE-1# oam p2mp-lsp-ping "LSP-p2mp-1" detail
P2MP LSP LSP-p2mp-1: 92 bytes MPLS payload

===============================================================================
S2L Information
===============================================================================
From             RTT                  Return Code
-------------------------------------------------------------------------------
192.0.2.6        =5.91ms              EgressRtr(3)
192.0.2.7        =6.07ms              EgressRtr(3)
===============================================================================

Total S2L configured/up/responded = 2/2/2,
          round-trip min/avg/max  = 5.91 / 5.99 / 6.07 ms
```

```
Responses based on return code:
        EgressRtr(3)=2

*A:PE-1#
```

Return codes are based on RFC4379. Value 3 means the replying router is an egress for the FEC at stack-depth.

P2MP LSP trace allows the user to trace the path of a single S2L path of a P2MP LSP from head-end node to leaf node. Using the downstream mapping TLV, each node along the S2L path can fill in the appropriate flags: B or E flag. The B-flag is set when the responding node is a branch LSR and the E-flag is set when the responding node is an egress LER.

```
*A:PE-1# oam p2mp-lsp-trace
  - p2mp-lsp-trace <lsp-name> p2mp-instance <instance-name> s2l-dest-address
    <ip-address> [fc <fc-name> [profile {in|out}]] [size <octets>] [max-fail
    <no-response-count>] [probe-count <probes-per-hop>] [min-ttl <min-label-ttl>]
    [max-ttl <max-label-ttl>] [timeout <timeout>]  [interval <interval>] [detail]

 <lsp-name>           : [64 chars max]
 <instance-name>      : [32 chars max]
 <ip-address>         : ipv4 address    a.b.c.d
 <fc-name>            : be|l2|af|l1|h2|ef|h1|nc - Default: be
 <in|out>             : in|out - Default: out
 <octets>             : [1..9198] - Default: 1
 <no-response-count>  : [1..10] - Default: 5
 <probes-per-hop>     : [1..10] - Default: 1
 <min-label-ttl>      : [1..255] - Default: 1
 <max-label-ttl>      : [1..255] - Default: 30
 <timeout>            : [1..60] seconds - Default: 3
 <detail>             : keyword - displays detailed information
 <interval>           : [1..10] seconds - Default: 1

*A:PE-1#

*A:PE-1# oam p2mp-lsp-trace "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l-dest-
address 192.0.2.7 detail
P2MP LSP LSP-p2mp-1: 132 bytes MPLS payload
P2MP Instance p-LSP-p2mp-1, S2L Egress 192.0.2.7

  1  192.0.2.3  rtt=2.10 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.35.2 ifaddr=192.168.35.2 iftype=ipv4Numbered MRU=1564
     label=262143 proto=4(RSVP-TE) B/E flags:0/0
  2  192.0.2.5  rtt=3.32 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.57.2 ifaddr=192.168.57.2 iftype=ipv4Numbered MRU=1564
     label=262143 proto=4(RSVP-TE) B/E flags:0/0
  3  192.0.2.7  rtt=4.55 ms rc=3(EgressRtr)

*A:PE-1#
```

Return codes are based on RFC4279. Value 8 means that the label is switched at stack-depth. This is the case for a transit LSR doing MPLS label swapping. No B or E flag is set.

Additional Topics

<*,G> IGMP join instead of <S,G> IGMP join

In the Head-end Node (Ingress LER) and Leaf Node (Egress LER) steps, a source specific IGMP join (<S,G> join) was used at the head-end node and leaf nodes. Another possibility is to use a source unknown or starg IGMP join (<*,G> join). When doing the latter, a rendezvous point (RP) must be defined in the PIM network. The RP allows multicast data flows between sources and receivers to meet at a predefined network location (in this example, the loopback address of node PE-1). It must be seen as an intermediate device to establish a multicast flow.

The RP can be defined in a dynamic way (BSR protocol) or a static way. In this example the static way is chosen meaning that on all involved PIM nodes, the RP address will be statically configured. The following configuration is needed on head-end and leaf nodes.

```
*A:PE-1/PE-6/PE-7# configure router pim
            rp
                static
                    address 192.0.2.1
                        group-prefix 227.1.1.1/32
                    exit
                exit
```

The **group-prefix** is a mandatory keyword. It references a group address or group address range for which this rendezvous point will be used.

```
*A:PE-1/PE-6/PE_7# show router pim rp
===============================================================================
PIM RP Set ipv4
===============================================================================
Group Address                                                  Hold Expiry
  RP Address                                 Type     Prio Time Time
-------------------------------------------------------------------------------
227.1.1.1/32
  192.0.2.1                                  Static   1    N/A  N/A
-------------------------------------------------------------------------------
Group Prefixes : 1
===============================================================================
*A:PE-1#
```

As previously mentioned, the configuration of the <*,G> IGMP join is done on the head-end node (PE-1) and leaf nodes (PE-6 and PE-7)

```
*A:PE-1# configure router igmp
            tunnel-interface rsvp-p2mp "LSP-p2mp-1"
                no shutdown
                static
```

```
                    group 227.1.1.1
                        starg
                    exit
                exit
            exit


*A:PE-6# configure router igmp
            interface "int-PE-6-MC-client2"
                static
                    group 227.1.1.1
                        starg
                    exit
                exit
            exit


*A:PE-7# configure router igmp
            interface "int-PE-7-MC-client1"
                static
                    group 227.1.1.1
                        starg
                    exit
                exit
            exit
```

The same previous **show** command can be used to verify the multicast traffic on head-end node and leaf nodes, **show router igmp group 227.1.1.1** and **show router pim group 227.1.1.1 detail**.

```
*A:PE-7# show router igmp group 227.1.1.1
===============================================================================
IGMP Interface Groups
===============================================================================

(*,227.1.1.1)                                                 UpTime: 0d 00:06:58
    Fwd List  : int-PE-7-MC-client1
-------------------------------------------------------------------------------
Entries : 1
===============================================================================
IGMP Host Groups
===============================================================================
No Matching Entries
===============================================================================
IGMP SAP Groups
===============================================================================
No Matching Entries
===============================================================================
*A:PE-7#
*A:PE-7# show router pim group 227.1.1.1 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 227.1.1.1
Source Address     : *
RP Address         : 192.0.2.1
Advt Router        :
Flags              :                    Type               : (*,G)
```

```
MRIB Next Hop      :
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:05:34        Resolved By        : unresolved

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:25
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor       :
Incoming Intf      : mpls-if-73728
Outgoing Intf List : int-PE-7-MC-client1

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 31                 Discarded Packets  : 0
Forwarded Octets   : 1426               RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 227.1.1.1
Source Address     : 192.168.9.2
RP Address         : 192.0.2.1
Advt Router        :
Flags              : spt                Type               : (S,G)
MRIB Next Hop      :
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:01:27
Up Time            : 0d 00:05:34        Resolved By        : unresolved

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:25
Up JP Rpt          : Not Pruned         Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : mpls-if-73728
Outgoing Intf List : int-PE-7-MC-client1

Curr Fwding Rate   : 3901.2 kbps
Forwarded Packets  : 3539800            Discarded Packets  : 0
Forwarded Octets   : 162830800          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-7#
```

Influence IGP Metric

Suppose that the IGP metric is increased on all links pointing to/from PE-2 and on the link between PE-5 and PE-7.

```
*A:PE-1# configure router ospf area 0 interface "int-PE-1-PE-2" metric 10000

*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-1" metric 10000
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-3" metric 10000
*A:PE-2# configure router ospf area 0 interface "int-PE-2-PE-4" metric 10000


*A:PE-3# configure router ospf area 0 interface "int-PE-3-PE-2" metric 10000

*A:PE-4# configure router ospf area 0 interface "int-PE-4-PE-2" metric 10000

*A:PE-5# configure router ospf area 0 interface "int-PE-5-PE-7" metric 10000

*A:PE-7# configure router ospf area 0 interface "int-PE-7-PE-5" metric 10000
```

The existing P2MP LSP *LSP-p2mp-1* will not take into account these new constraints. The two S2L paths (one *loose* toward PE-6 and another one *loose* toward PE-7) are calculated using the default OSPF metric. To trigger MPLS to re-compute the S2L paths, configure a p2mp-resignal-timer on the head-end node inside the global MPLS context. Each time this timer expires (in the example, every 60 minutes), MPLS will trigger CSPF to re-compute the whole set of S2L paths of all active P2MP instances. MPLS performs a global make-before-break (MBB) and moves each S2L sub-LSP in the instance into its new path using a new P2MP LSP ID if the global MBB is successful. **show router mpls status** gives an indication when the P2MP resignal timer will expire and which types of LSPs are set up on the node.

```
*A:PE-1# configure router mpls p2mp-resignal-timer 60

*A:PE-1# show router mpls status
===============================================================================
MPLS Status
===============================================================================
Admin Status              : Up          Oper Status               : Up
Oper Down Reason          : n/a
FRR Object                : Enabled     Resignal Timer            : Disabled
Hold Timer                : 1 seconds   Next Resignal             : N/A
Srlg Frr                  : Disabled    Srlg Frr Strict           : Disabled
Admin Group Frr           : Disabled
Dynamic Bypass            : Enabled     User Srlg Database        : Disabled
BypassResignalTimer       : Disabled    BypassNextResignal        : N/A
LeastFill Min Thd         : 5 percent   LeastFill Reopti Thd      : 10 percent
Local TTL Prop            : Enabled     Transit TTL Prop          : Enabled
AB Sample Multiplier      : 1           AB Adjust Multiplier      : 288
Exp Backoff Retry         : Disabled    CSPF On Loose Hop         : Disabled
Lsp Init RetryTimeout     : 30 seconds  MBB Pref Current Hops     : Disabled
Logger Event Bundling     : Disabled
RetryIgpOverload          : Disabled

P2mp Resignal Timer       : 60 minutes  P2mp Next Resignal        : 41 minutes
Sec FastRetryTimer        : Disabled    Static LSP FR Timer       : 30 seconds
P2P Max Bypass Association: 1000
P2PActPathFastRetry       : Disabled    P2MP S2L Fast Retry       : Disabled
In Maintenance Mode       : No
MplsTp                    : Disabled
```

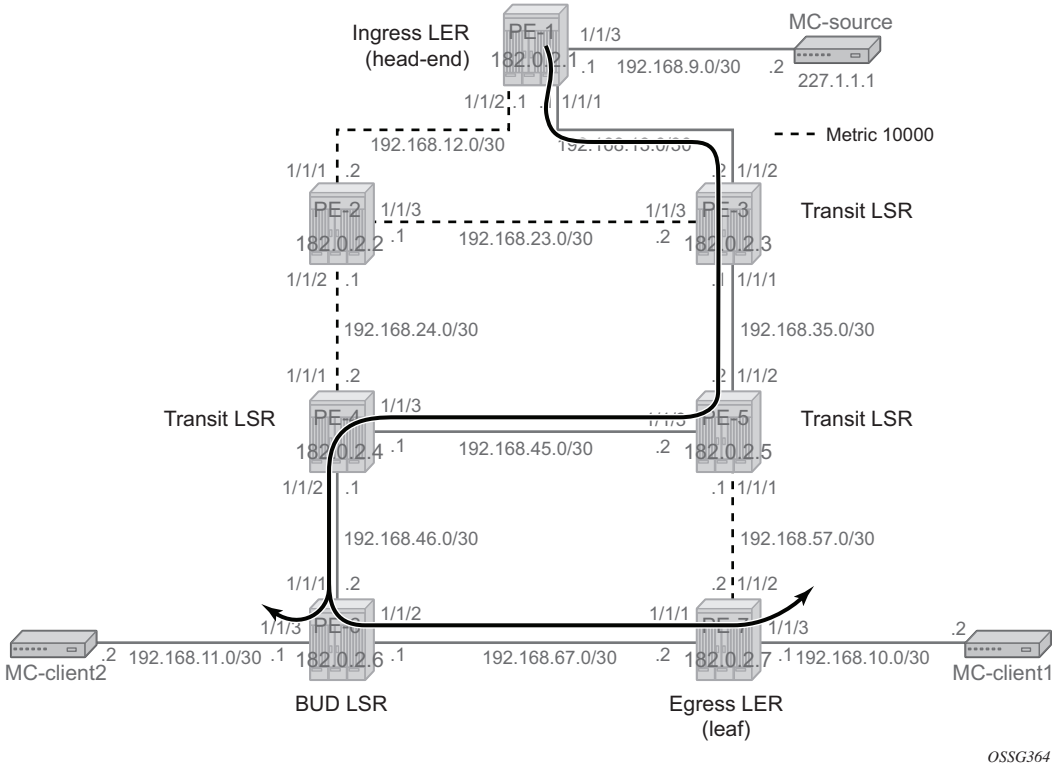
```
Pce-report                : None
Next Available Lsp Index  : 5
Entropy Label RSVP-TE     : Enabled
---snip---
===============================================================================
```

As an alternative the user can also perform a manual resignal of a P2MP instance on the head-end node using a tools command.

```
*A:PE-1# tools perform router mpls resignal p2mp-lsp "LSP-p2mp-1" p2mp-instance"p-LSP-
p2mp-1"

*A:PE-1# tools perform router mpls resignal p2mp-delay 0
```

Figure 187 **P2MP LSP p-to-mp-1 with Metric Change**



After an instantaneous tools resignal command which is executed with no delay (p2mp-delay 0), the S2L paths can be verified, also check Figure 187. Node PE-6 is now a bud LSR node (instead of egress LER before).

```
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l loose
to 192.0.2.7 detail
===============================================================================
MPLS LSP LSP-p2mp-1 S2L loose (Detail)
===============================================================================
```

```
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-1 S2L loose
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-1                      S2l LSP ID  : 53764
P2MP ID     : 0                               S2l Grp Id  : 1
Adm State   : Up                              Oper State  : Up
S2l State:  : Active                                      :
S2L Name    : loose                           To          : 192.0.2.7
S2l Admin   : Up                              S2l Oper    : Up
OutInterface: 1/1/1                           Out Label   : 262143
S2L Up Time : 0d 04:45:36                     S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                               NextRetryIn : 0 sec
S2L Trans   : 3                               CSPF Queries: 3
Failure Code: noError                         Failure Node: n/a
Inter-area  : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.13.1 (192.0.2.1) @                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @                Record Label        : 262143
 -> 192.168.35.2 (192.0.2.5) @                Record Label        : 262143
 -> 192.168.45.1 (192.0.2.4) @                Record Label        : 262143
 -> 192.168.46.2 (192.0.2.6) @                Record Label        : 262143
 -> 192.168.67.2 (192.0.2.7)                  Record Label        : 262143
ComputedHops:
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.35.2(S)
 -> 192.168.45.1(S)
 -> 192.168.46.2(S)
 -> 192.168.67.2(S)
LastResignal: n/a
===============================================================================
*A:PE-1#
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l loose
to 192.0.2.6 detail
===============================================================================
MPLS LSP LSP-p2mp-1 S2L loose (Detail)
===============================================================================
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-1 S2L loose
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-1                      S2l LSP ID  : 53764
P2MP ID     : 0                               S2l Grp Id  : 2
Adm State   : Up                              Oper State  : Up
```

```
S2l State:  : Active                                        :
S2L Name    : loose                                 To      : 192.0.2.6
S2l Admin   : Up                                    S2l Oper    : Up
OutInterface: 1/1/1                                 Out Label   : 262143
S2L Up Time : 0d 04:45:02                           S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                                     NextRetryIn : 0 sec
S2L Trans   : 3                                     CSPF Queries: 4
Failure Code: noError                               Failure Node: n/a
Inter-area  : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.13.1 (192.0.2.1) @                      Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @                      Record Label        : 262143
 -> 192.168.35.2 (192.0.2.5) @                      Record Label        : 262143
 -> 192.168.45.1 (192.0.2.4) @                      Record Label        : 262143
 -> 192.168.46.2 (192.0.2.6)                        Record Label        : 262143
ComputedHops:
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.35.2(S)
 -> 192.168.45.1(S)
 -> 192.168.46.2(S)
LastResignal: n/a
===============================================================================
*A:PE-1#
```

An **oam p2mp-lsp-trace** command toward PE-7 will now set the E flag on PE-6 since that PE acts also as an egress LER node.

```
*A:PE-1# oam p2mp-lsp-trace "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l-dest-
address 192.0.2.7 detail
P2MP LSP LSP-p2mp-1: 132 bytes MPLS payload
P2MP Instance p-LSP-p2mp-1, S2L Egress 192.0.2.7

  1  192.0.2.3  rtt=2.05 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.35.2 ifaddr=192.168.35.2 iftype=ipv4Numbered MRU=1564
     label=262143 proto=4(RSVP-TE) B/E flags:0/0
  2  192.0.2.5  rtt=3.25 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.45.1 ifaddr=192.168.45.1 iftype=ipv4Numbered MRU=1564
     label=262143 proto=4(RSVP-TE) B/E flags:0/0
  3  192.0.2.4  rtt=4.88 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.46.2 ifaddr=192.168.46.2 iftype=ipv4Numbered MRU=1564
     label=262143 proto=4(RSVP-TE) B/E flags:0/0
  4  192.0.2.6  rtt=7.79 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.67.2 ifaddr=192.168.67.2 iftype=ipv4Numbered MRU=1564
     label=262143 proto=4(RSVP-TE) B/E flags:0/1
  5  192.0.2.7  rtt=67.2 ms rc=3(EgressRtr)

*A:PE-1#
```

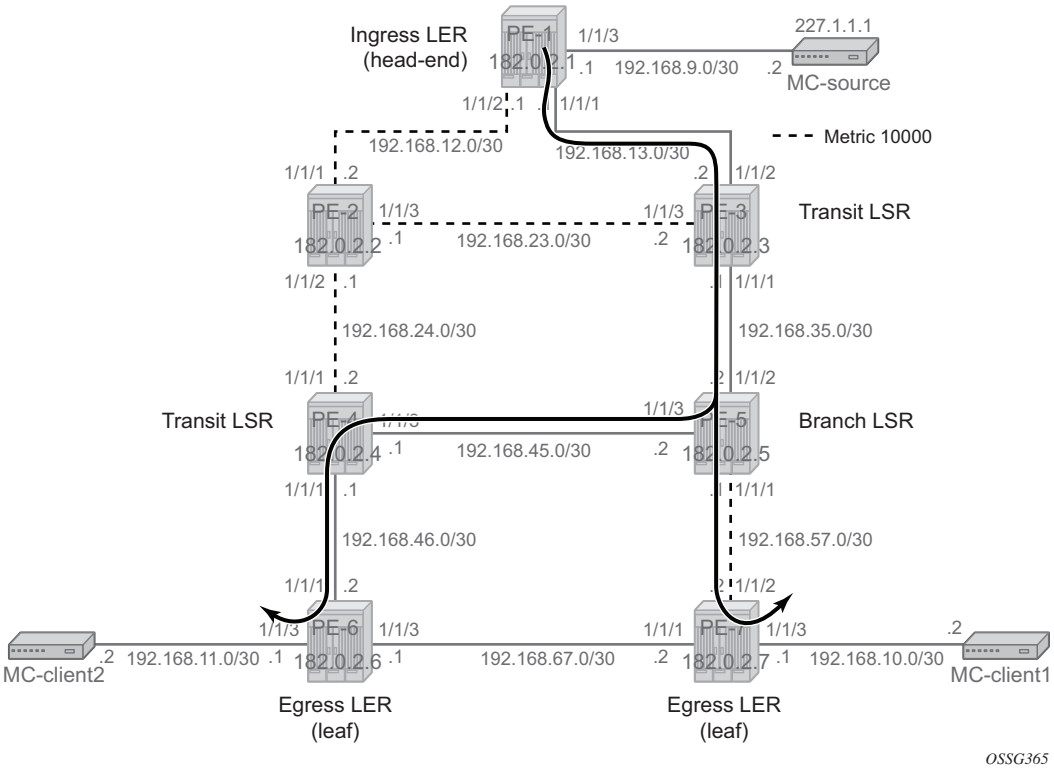
Inthenext step, the S2L path toward PE-7 is changed from **loose** to a **strict** direct MPLS path (*strict-to-PE-7)*. In that way, OSPF is not calculating anymore the shortest path to the leaf node.

```
*A:PE-1# configure router mpls
            path "path-strict-to-PE-7"
                hop 10 192.168.13.2 strict
                hop 20 192.168.35.2 strict
                hop 30 192.168.57.2 strict
                no shutdown
            exit
```

Before applying this new S2L path to the existing P2MP LSP (*LSP-p2mp-1*), the existing S2L path toward PE-7 must be removed.

```
*A:PE-1# configure router mpls lsp "LSP-p2mp-1" primary-p2mp-instance "p-LSP-p2mp-1"
                    s2l-path "loose" to 192.0.2.7 shutdown
                    no s2l-path "loose" to 192.0.2.7
                    s2l-path"path-strict-to-PE-7" to 192.0.2.7
                    exit
                exit
```

Figure 188 **P2MP LSP LSP-p2mp-1 with Strict S2L Path toward PE-7**



As a consequence of this, only the **S2l Grp Id** has changed while **S2L LSP ID** remains the same as before. Now, S2L paths can be verified according to Figure 188. PE-5 is now a branch LSR node (instead of a transit LSR before).

```
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l path-
strict-to-PE-7 to 192.0.2.7 detail
===============================================================================
MPLS LSP LSP-p2mp-1 S2L path-strict-to-PE-7 (Detail)
===============================================================================
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-1 S2L path-strict-to-PE-7
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-1                      S2l LSP ID  : 53764
P2MP ID     : 0                               S2l Grp Id  : 3
Adm State   : Up                              Oper State  : Up
S2l State:  : Active                                      :
S2L Name    : path-strict-to-PE-7             To          : 192.0.2.7
S2l Admin   : Up                              S2l Oper    : Up
OutInterface: 1/1/1                           Out Label   : 262143
S2L Up Time : 0d 00:09:25                     S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                               NextRetryIn : 0 sec
S2L Trans   : 1                               CSPF Queries: 1
Failure Code: noError                         Failure Node: n/a
Inter-area  : False
ExplicitHops:
    192.168.13.2(S)    -> 192.168.35.2(S)    -> 192.168.57.2(S)
Actual Hops :
    192.168.13.1 (192.0.2.1) @                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @                Record Label        : 262143
 -> 192.168.35.2 (192.0.2.5) @                Record Label        : 262143
 -> 192.168.57.2 (192.0.2.7)                  Record Label        : 262142
ComputedHops:
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.35.2(S)
 -> 192.168.57.2(S)
LastResignal: n/a
===============================================================================
*A:PE-1#
```

An **oam p2mp-lsp-trace** command toward PE-7 will now set the B flag on PE-5 since that became a branch LSR now.

```
*A:PE-1# oam p2mp-lsp-trace "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l-dest-
address  192.0.2.7 detail
P2MP LSP LSP-p2mp-1: 132 bytes MPLS payload
P2MP Instance p-LSP-p2mp-1, S2L Egress 192.0.2.7

  1  192.0.2.3  rtt=2.05 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.35.2 ifaddr=192.168.35.2 iftype=ipv4Numbered MRU=1564
label=262143 proto=4(RSVP-TE) B/E flags:0/0
  2  192.0.2.5  rtt=3.26 ms rc=8(DSRtrMatchLabel)
     DS 1: ipaddr=192.168.57.2 ifaddr=192.168.57.2 iftype=ipv4Numbered MRU=1564
label=262142 proto=4(RSVP-TE) B/E flags:1/0
  3  192.0.2.7  rtt=6.12 ms rc=3(EgressRtr)
```

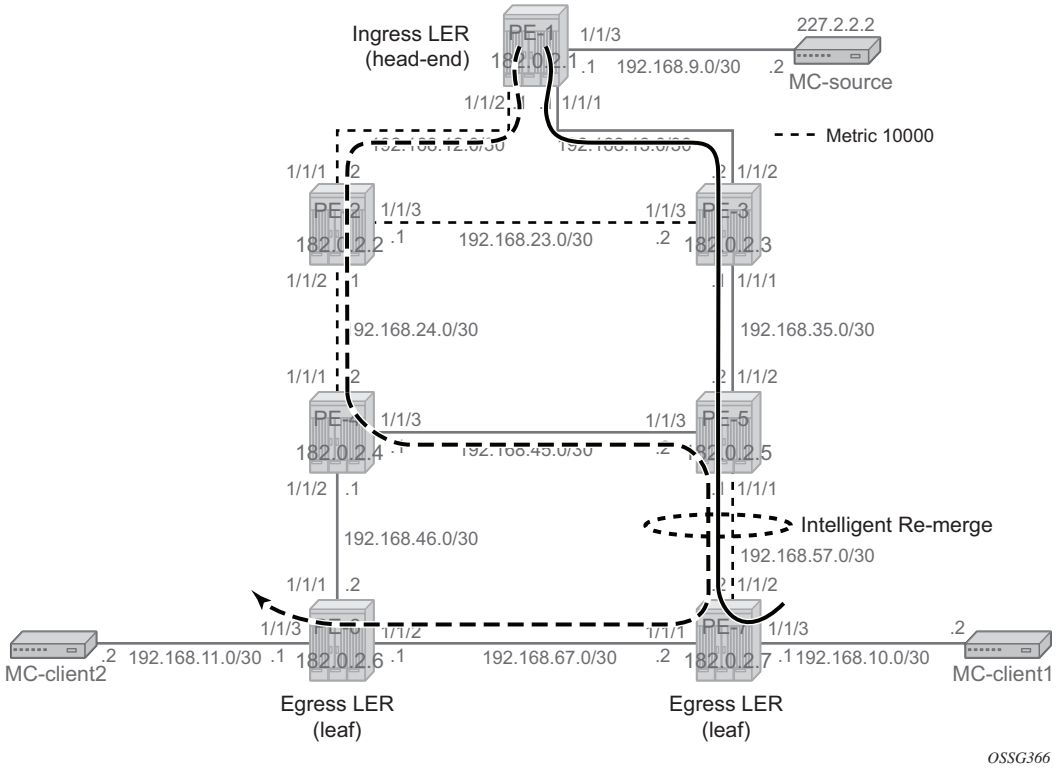
Intelligent Re-merge

Intelligent re-merge protects users from receiving duplicate multicast traffic during convergence. It also protects against duplicate traffic in case of badly designed S2L paths. Three cases are described for which intelligent re-merge is implemented.

Case 1

When the paths of two different S2Ls of the same P2MP LSP instance have ingress label maps (ILMs) on different ports but go out on the same next-hop label forwarding entry (NHLFE).

Figure 189 **Intelligent Remerge, Case 1**



On the head-end node (PE-1), a new P2MP LSP *('LSP-p2mp-2'*) will be created with two strict direct MPLS paths *('strict-to-PE-7'* and *'strict-to-PE-6'*). See Figure 190 for detailed address information. Intelligent re-merge is performed at node PE-5.

```
*A:PE-1# configure router mpls
            path "path-strict-to-PE-7"
                hop 10 192.168.13.2 strict
                hop 20 192.168.35.2 strict
                hop 30 192.168.57.2 strict
                no shutdown
```

```
            exit
            path "path-strict-to-PE-6"
                hop 10 192.168.12.2 strict
                hop 20 192.168.24.2 strict
                hop 30 192.168.45.2 strict
                hop 40 192.168.57.2 strict
                hop 50 192.168.67.1 strict
                no shutdown
            exit
            lsp "LSP-p2mp-2" p2mp-lsp
                primary-p2mp-instance "p-LSP-p2mp-2"
                    s2l-path "path-strict-to-PE-7" to 192.0.2.7
                    exit
                    s2l-path "path-strict-to-PE-6" to 192.0.2.6
                    exit
                exit
                no shutdown
            exit
            no shutdown

*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-2" p2mp-instance "p-LSP-p2mp-2"
===============================================================================
MPLS P2MP Instance (Originating)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-2
P2MP ID     : 0                                LSP Tunnel ID  : 2
Adm State   : Up                               Oper State     : Up

P2MPInstance: p-LSP-p2mp-2                     P2MP-Inst-type : Primary
P2MP Inst Id: 2                                P2MP Lsp Id    : 38912
Inst Admin  : Up                               Inst Oper      : Up
Inst Up Time: 0d 00:34:10                      Inst Dn Time   : 0d 00:00:00
Hop Limit   : 255                              Adaptive       : Enabled
Record Route: Record                           Record Label   : Record
Include Grps:                                  Exclude Grps   :
None                                           None
Bandwidth   : No Reservation                   Oper Bw        : 0 Mbps
S2l-Name    : path-strict-to-PE-7              To             : 192.0.2.7
S2l Admin   : Up                               S2l Oper       : Up
S2l-Name    : path-strict-to-PE-6              To             : 192.0.2.6
S2l Admin   : Up                               S2l Oper       : Up
-------------------------------------------------------------------------------
P2MP instances : 1
===============================================================================
*A:PE-1#
```

To verify that node PE-5 is not sending duplicate multicast traffic downstream toward PE-7 while it receives two incoming multicast streams, a new tunnel interface and a new static <S,G> IGMP join will be configured on head-end node (PE-1) and leaf nodes (PE-6 and PE-7). Also on the leaf nodes, an extension to the existing multicast information policy is needed. This is configured as follows:

```
*A:PE-1# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-2"
```

```
*A:PE-1# configure router igmp
            tunnel-interface rsvp-p2mp "LSP-p2mp-2"
                static
                    group 227.2.2.2
                        source 192.168.9.2
                    exit
                exit
            exit

*A:PE-6# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-2" sender 192.0.2.1

*A:PE-6# configure router igmp
            interface "int-PE-6-MC-client2"
                static
                    group 227.2.2.2
                        source 192.168.9.2
                    exit
                exit
            exit

*A:PE-6# configure mcast-management
        multicast-info-policy "p2mp-pol" create
            bundle "bundle2" create
                primary-tunnel-interface rsvp-p2mp "LSP-p2mp-2" sender 192.0.2.1
                    channel 227.2.2.2 create
                    exit
                exit
            exit

*A:PE-7# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-2" sender 192.0.2.1

*A:PE-7# configure router igmp
            interface "int-PE-7-MC-client1"
                static
                    group 227.2.2.2
                        source 192.168.9.2
                    exit
                exit
            exit

*A:PE-7# configure mcast-management
        multicast-info-policy "p2mp-pol" create
            bundle "bundle2" create
                primary-tunnel-interface rsvp-p2mp "LSP-p2mp-2" sender 192.0.2.1
                    channel 227.2.2.2 create
                    exit
                exit
            exit
```

For verification of incoming/outgoing multicast traffic at node PE-5, the **monitor** command is used.

```
*A:PE-5# monitor port 1/1/1 1/1/2 1/1/3 rate interval 3 repeat 100
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                    Input                Output
-------------------------------------------------------------------------------
```

```
---snip---
-------------------------------------------------------------------------------
At time t = 18 sec (Mode: Rate)
-------------------------------------------------------------------------------
Port 1/1/1
-------------------------------------------------------------------------------
Octets                                          43                  891751
Packets                                          1                   13114
Errors                                           0                       0
Utilization (% of port capacity)             ~0.00                    0.09

Port 1/1/2
-------------------------------------------------------------------------------
Octets                                      891751                      50
Packets                                      13114                       0
Errors                                           0                       0
Utilization (% of port capacity)              0.09                   ~0.00

Port 1/1/3
-------------------------------------------------------------------------------
Octets                                      891779                      50
Packets                                      13114                       0
Errors                                           0                       0
Utilization (% of port capacity)              0.09                   ~0.00

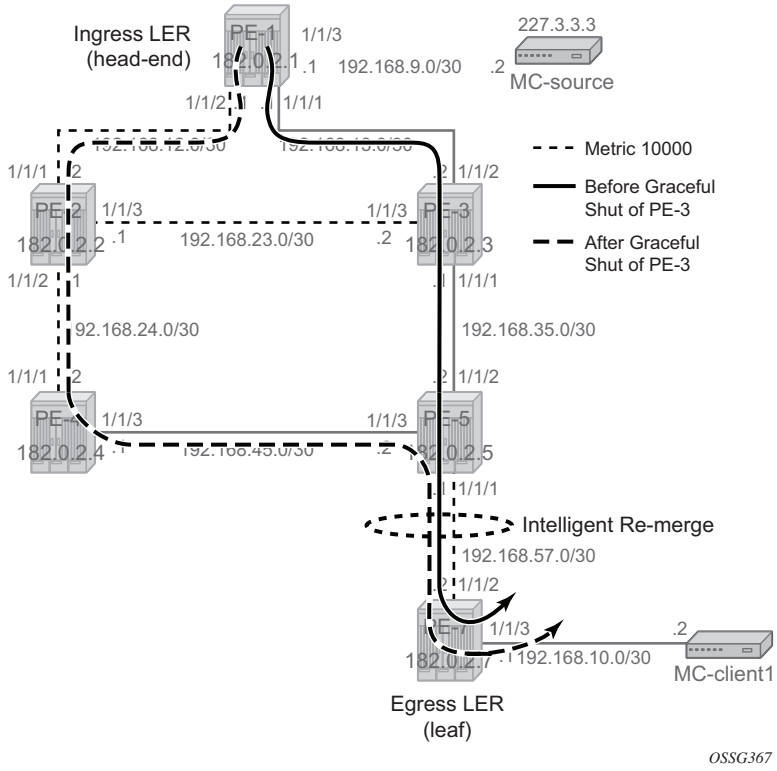
---snip---
```

Two incoming multicast streams are seen at PE-5 node (**port 1/1/2** and **port 1/1/3**) and only one outgoing multicast stream (**port 1/1/1**) is sent. No traffic duplication is seen.

Case 2

When two paths of the same S2L have ILMs on different incoming ports and go out on the same NHLFE. This is the case when we perform make-before-break (MBB) on an S2L path due to graceful shutdown or global revertive. This is only a temporary situation since the original path will be torn down.

Figure 190 **Intelligent Re-merge, Case 2**



For this test, only one MC-client will be looked at (the one connected to leaf node PE-7). On nodes PE-4 and PE-7, the port to PE-6 will be shut down to isolate PE-6. On the head-end node PE-1, a new P2MP LSP “LSP-p2mp-3” will be created with one loose MPLS path “loose” and keyword **cspf use-te-metric** to ensure that CSPF will use the TE metric instead of the IGP metric. In this case, the TE metric has the same value on all MPLS interfaces and the path has the hops PE-1, PE-3, PE-5, and PE-7. See Figure 190 for detailed address information. Also in this case, intelligent re-merge is performed at node PE-5.

```
*A:PE-1# configure router mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-p2mp-3" p2mp-lsp
                cspf use-te-metric
                primary-p2mp-instance "p-LSP-p2mp-3"
                    s2l-path "loose" to 192.0.2.7
                    exit
                exit
                no shutdown
            exit
            no shutdown

*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-3" p2mp-instance "p-LSP-p2mp-3" s2l loose
```

```
to 192.0.2.7 detail
===============================================================================
MPLS LSP LSP-p2mp-3 S2L loose (Detail)
===============================================================================
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-3 S2L loose
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-3                      S2l LSP ID  : 47104
P2MP ID     : 0                               S2l Grp Id  : 1
Adm State   : Up                              Oper State  : Up
S2l State:  : Active                                      :
S2L Name    : loose                           To          : 192.0.2.7
S2l Admin   : Up                              S2l Oper    : Up
OutInterface: 1/1/1                           Out Label   : 262141
S2L Up Time : 0d 00:04:23                     S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                               NextRetryIn : 0 sec
S2L Trans   : 1                               CSPF Queries: 0
Failure Code: noError                         Failure Node: n/a
Inter-area  : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.13.1 (192.0.2.1)                  Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3)                  Record Label        : 262141
 -> 192.168.35.2 (192.0.2.5)                  Record Label        : 262142
 -> 192.168.57.2 (192.0.2.7)                  Record Label        : 262142
LastResignal: n/a
===============================================================================
*A:PE-1#
```

In a normal situation, the P2MP LSP would follow the nodes PE-1, PE-3, PE-5, and PE-7. This can be verified with MC traffic. Therefore, a new tunnel interface and a new static <S,G> IGMP join will be configured on head-end node PE-1 and leaf node PE-7. On the leaf node, an extension to the existing multicast information policy is needed. This is configured as follows:

```
*A:PE-1# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-3"

*A:PE-1# configure router igmp
            tunnel-interface rsvp-p2mp "LSP-p2mp-3"
                static
                    group 227.3.3.3
                        source 192.168.9.2
                    exit
                exit
            exit


*A:PE-7# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-3" sender 192.0.2.1
```

```
*A:PE-7# configure router igmp
            interface "int-PE-7-MC-client1"
                static
                    group 227.3.3.3
                        source 192.168.9.2
                    exit
                exit
            exit

*A:PE-7# configure mcast-management
        multicast-info-policy "p2mp-pol" create
            bundle "bundle3" create
                primary-tunnel-interface rsvp-p2mp LSP-p2mp-3 sender 192.0.2.1
                channel 227.3.3.3 create
                exit
            exit
        exit
```

Monitor the traffic on PE-5. Under normal circumstances, the ingress port is 1/1/2 and the egress port is 1/1/1.

```
*A:PE-5# monitor port 1/1/1 1/1/2 1/1/3 rate interval 3 repeat 999
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                    Input                Output
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------

---snip---
-------------------------------------------------------------------------------
At time t = 843 sec (Mode: Rate)
-------------------------------------------------------------------------------
Port 1/1/1
-------------------------------------------------------------------------------
Octets                                              21              1446744
Packets                                              0                21276
Errors                                               0                    0
Utilization (% of port capacity)                 ~0.00                 0.14

Port 1/1/2
-------------------------------------------------------------------------------
Octets                                         1446836                  109
Packets                                          21276                    1
Errors                                               0                    0
Utilization (% of port capacity)                  0.14                ~0.00

Port 1/1/3
-------------------------------------------------------------------------------
Octets                                              21                  111
Packets                                              0                    1
Errors                                               0                    0
Utilization (% of port capacity)                 ~0.00                ~0.00
```

An RSVP graceful shutdown is performed on node PE-3, as follows:

```
*A:PE-3# configure router rsvp graceful-shutdown
```

Global revertive is triggered on head-end node PE-1. A new MPLS path will be calculated (see the dotted line in Figure 190). For a few seconds or even less than a second, the old path and new path are active (two incoming MC streams on node PE-5). Node PE-5 is doing intelligent re-merge, not sending duplicate multicast traffic downstream toward PE-7:

```
*A:PE-5# monitor port 1/1/1 1/1/2 1/1/3 rate interval 3 repeat 100
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                      Input                  Output
-------------------------------------------------------------------------------
---snip---

-------------------------------------------------------------------------------
At time t = 30 sec (Mode: Rate)
-------------------------------------------------------------------------------
Port 1/1/1
-------------------------------------------------------------------------------
Octets                                                  265                1228986
Packets                                                   2                  18070
Errors                                                    0                      0
Utilization (% of port capacity)                      ~0.00                   0.12

Port 1/1/2
-------------------------------------------------------------------------------
Octets                                                1229058                     91
Packets                                                 18071                      1
Errors                                                      0                      0
Utilization (% of port capacity)                         0.12                  ~0.00

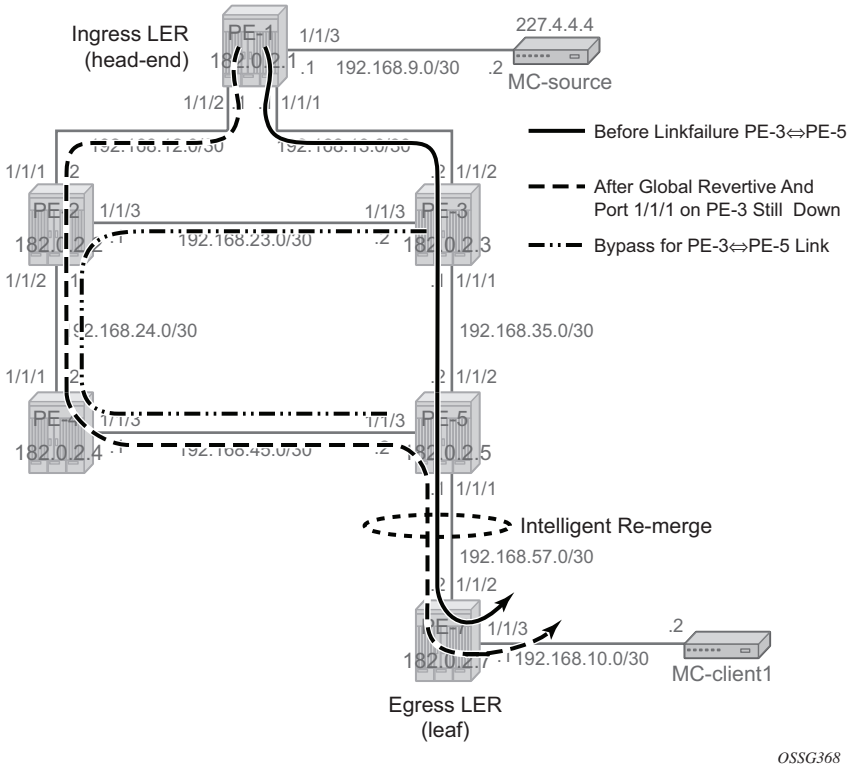
Port 1/1/3
-------------------------------------------------------------------------------
Octets                                                204539                    319
Packets                                                 3005                      2
Errors                                                      0                      0
Utilization (% of port capacity)                         0.02                  ~0.00
-------------------------------------------------------------------------------
```

The granularity of the monitoring command is 3 seconds. The graceful shutdown takes less than 3 seconds. However, we can clearly see that the number of outgoing packets on port 1/1/1 equals the number of incoming packets on port 1/1/2. A number of these incoming packets also arrived on port 1/1/3, but no duplicate packets were sent on the outgoing port. No traffic duplication is seen.

Case 3

When a bypass is active on the S2L path and the new global revertive path of the same S2L arrives on the same incoming interface as the original path (interface flapped) at the FRR merge point node, the node will do intelligent re-merge. The implementation recognizes this specific case and will signal a different label from the original S2L path coming on that same interface.

Figure 191 **Intelligent Re-merge, Case 3**



For this test, all the non-default OSPF metrics are removed from the interfaces. Only one MC-client will be looked at (the one connected to leaf node PE-7). On nodes PE-4 and PE-7, the port toward PE-6 will be shut down to isolate PE-6. On the head-end node PE-1, a new P2MP LSP “LSP-p2mp-4” will be created with one loose MPLS path “loose” and FRR enabled. See Figure 191 for detailed address information. Also in this case, intelligent re-merge is performed at node PE-5.

```
*A:PE-1# configure router mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-p2mp-4" p2mp-lsp
                cspf
                fast-reroute facility
                    no node-protect
                exit
                primary-p2mp-instance "p-LSP-p2mp-4"
```

```
                    s2l-path "loose" to 192.0.2.7
                    exit
                exit
                no shutdown
            exit
            no shutdown


*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-4" p2mp-instance "p-LSP-p2mp-4" s2l loose
to 192.0.2.7 detail
===============================================================================
MPLS LSP LSP-p2mp-4 S2L loose (Detail)
===============================================================================
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-4 S2L loose
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-4                      S2l LSP ID  : 46592
P2MP ID     : 0                               S2l Grp Id  : 1
Adm State   : Up                              Oper State  : Up
S2l State:  : Active                                      :
S2L Name    : loose                           To          : 192.0.2.7
S2l Admin   : Up                              S2l Oper    : Up
OutInterface: 1/1/1                           Out Label   : 262139
S2L Up Time : 0d 00:02:47                     S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                               NextRetryIn : 0 sec
S2L Trans   : 1                               CSPF Queries: 1
Failure Code: noError                         Failure Node: n/a
Inter-area  : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.13.1 (192.0.2.1) @                Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @                Record Label        : 262139
 -> 192.168.35.2 (192.0.2.5)                  Record Label        : 262141
 -> 192.168.57.2 (192.0.2.7)                  Record Label        : 262141
ComputedHops:
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.35.2(S)
 -> 192.168.57.2(S)
LastResignal: n/a
===============================================================================
*A:PE-1#
```

In the normal situation, the P2MP LSP follows the nodes PE-1, PE-3, PE-5 and PE-7. This can be verified with MC traffic. Therefore, a new tunnel interface and a new static <S,G> IGMP join will be configured on head-end node PE-1 and leaf node PE-7. On the leaf node, an extension to the existing multicast information policy is needed. This is configured as follows:

```
*A:PE-1# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-4"

*A:PE-1# configure router igmp
            tunnel-interface rsvp-p2mp LSP-p2mp-4
                static
                    group 227.4.4.4
                        source 192.168.9.2
                    exit
                exit
            exit


*A:PE-7# configure router pim tunnel-interface rsvp-p2mp "LSP-p2mp-4" sender 192.0.2.1

*A:PE-7# configure router igmp
            interface "int-PE-7-MC-client1"
                static
                    group 227.4.4.4
                        source 192.168.9.2
                    exit
                exit
            exit

*A:PE-7# configure mcast-management
        multicast-info-policy "p2mp-pol" create
            bundle "bundle4" create
                primary-tunnel-interface rsvp-p2mp LSP-p2mp-4 sender 192.0.2.1
                channel 227.4.4.4 create
                exit
            exit
        exit


*A:PE-5# monitor port 1/1/1 1/1/2 1/1/3 rate interval 3 repeat 999
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                    Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 1701 sec (Mode: Rate)
-------------------------------------------------------------------------------
Port 1/1/1
-------------------------------------------------------------------------------
Octets                                                21               1304494
Packets                                                0                 19183
Errors                                                 0                     0
Utilization (% of port capacity)                   ~0.00                  0.13

Port 1/1/2
-------------------------------------------------------------------------------
Octets                                           1304549                    87
Packets                                            19184                     1
Errors                                                 0                     0
Utilization (% of port capacity)                    0.13                 ~0.00

Port 1/1/3
-------------------------------------------------------------------------------
```

```
Octets                                              111                  116
Packets                                               1                    1
Errors                                                0                    0
Utilization (% of port capacity)                ~0.00                ~0.00
```

Now a link failure on the interface from PE-3 to PE-5 is emulated.

```
*A:PE-3# configure port 1/1/1 shutdown
```

As a consequence of this, traffic will be flowing over the bypass link (see Figure 191 and note the '#' symbol in the next **show** command, as well as the failure code).

```
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-4" p2mp-instance "p-LSP-p2mp-4" s2l loose
to 192.0.2.7 detail
===============================================================================
MPLS LSP LSP-p2mp-4 S2L loose (Detail)
===============================================================================
Legend :
    @ - Detour Available                    # - Detour In Use
    b - Bandwidth Protected                 n - Node Protected
    S - Strict                              L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-4 S2L loose
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-4                    S2l LSP ID  : 46594
P2MP ID     : 0                             S2l Grp Id  : 1
Adm State   : Up                            Oper State  : Up
S2l State:  : Active                                    :
S2L Name    : loose                         To          : 192.0.2.7
S2l Admin   : Up                            S2l Oper    : Up
OutInterface: 1/1/1                         Out Label   : 262140
S2L Up Time : 0d 00:44:42                   S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                             NextRetryIn : 0 sec
S2L Trans   : 3                             CSPF Queries: 3
Failure Code: tunnelLocallyRepaired         Failure Node: 192.0.2.3
Inter-area  : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.13.1 (192.0.2.1) @              Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3) @ #            Record Label        : 262140
 -> 192.168.35.2 (192.0.2.5)                Record Label        : 262138
 -> 192.168.57.2 (192.0.2.7)                Record Label        : 262140
ComputedHops:
    192.168.13.1(S)
 -> 192.168.13.2(S)
 -> 192.168.35.2(S)
 -> 192.168.57.2(S)
LastResignal: n/a
In Prog MBB :
 MBB Type   : GlobalRevert                  NextRetryIn : 15 sec
 Started At : 04/10/2015 09:14:45           RetryAttempt: 0
 FailureCode: noError                       Failure Node: n/a
```

```
===============================================================================
*A:PE-1#
```

In the meantime, PE-3 will trigger a global revertive action (sending PathErr message) toward the head-end node PE-1.

```
*A:PE-1# show router mpls p2mp-lsp "LSP-p2mp-4" p2mp-instance "p-LSP-p2mp-4" s2l loose
to 192.0.2.7 detail

===============================================================================
MPLS LSP LSP-p2mp-4 S2L loose (Detail)
===============================================================================
Legend :
    @ - Detour Available                          # - Detour In Use
    b - Bandwidth Protected                       n - Node Protected
    S - Strict                                    L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-p2mp-4 S2L loose
-------------------------------------------------------------------------------
LSP Name    : LSP-p2mp-4                        S2l LSP ID  : 46594
P2MP ID     : 0                                 S2l Grp Id  : 2
Adm State   : Up                                Oper State  : Up
S2l State:  : Active                                        :
S2L Name    : loose                             To          : 192.0.2.7
S2l Admin   : Up                                S2l Oper    : Up
OutInterface: 1/1/2                             Out Label   : 262142
S2L Up Time : 0d 00:45:19                       S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                                 NextRetryIn : 0 sec
S2L Trans   : 4                                 CSPF Queries: 4
Failure Code: noError                           Failure Node: n/a
Inter-area  : False
ExplicitHops:
    No Hops Specified
Actual Hops :
    192.168.12.1 (192.0.2.1) @                  Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262142
 -> 192.168.24.2 (192.0.2.4)                    Record Label        : 262142
 -> 192.168.45.2 (192.0.2.5)                    Record Label        : 262143
 -> 192.168.57.2 (192.0.2.7)                    Record Label        : 262140
ComputedHops:
    192.168.12.1(S)
 -> 192.168.12.2(S)
 -> 192.168.24.2(S)
 -> 192.168.45.2(S)
 -> 192.168.57.2(S)
LastResignal: n/a
Last MBB    :
 MBB Type   : GlobalRevert                      MBB State   : Success
 Ended At   : 04/10/2015 09:15:22
===============================================================================
*A:PE-1#
```

For a short time, PE-5 will receive two incoming MC streams (both arriving on port 1/1/3). One from ‘bypass’ path (PE-3 => PE-2 => PE-4 => PE-5) and one from new MPLS path (PE-1 => PE-2 => PE-4 => PE-5 => PE-7). Port 1/1/1 on PE-5 performs intelligent remerge, only one MC stream is sent downstream toward leaf node PE-7.

Conclusion

From a configuration point of view, a P2MP LSP is only configured on the head-end node of that P2MP LSP; no explicit configuration is needed on the transit LSRs, branch LSRs, bud LSRs, and egress LERs/leaf nodes.

Since the PIM protocol is only needed on the head-end node and the leaf nodes, we can work in a PIM-free core network. Although convergence is not covered in this chapter, failures in the core will be resolved by MPLS (in case of FRR, traffic loss for less than 50ms is expected). This is a major improvement compared to PIM convergence.

Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP

This chapter provides information about Seamless MPLS: Isolated IGP/LDP domains and Labeled BGP.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS routers and was initially written for release 13.0.R7. The CLI in this edition is based on release 14.0.R1.

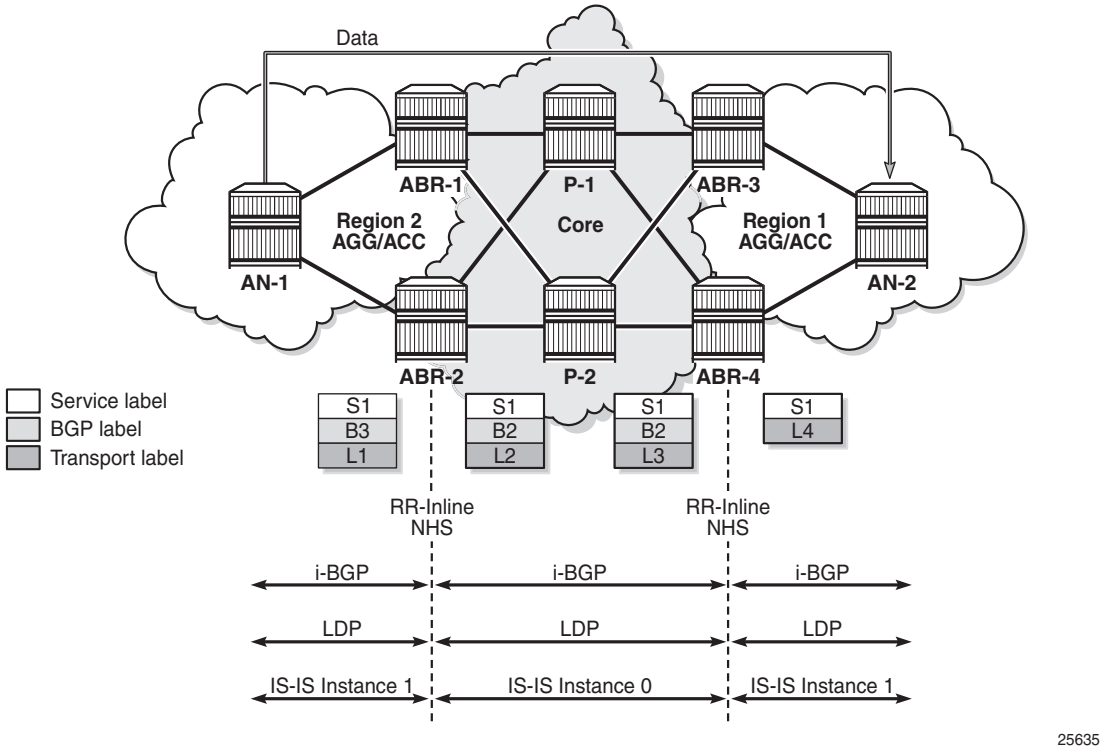
Overview

Seamless multiprotocol label switching (MPLS) is a network architecture that extends MPLS networks to integrate access and aggregation networks into a single MPLS domain, to solve the scaling problems in flat MPLS-based deployments. The Seamless MPLS transport concept described in this chapter partitions the core, aggregation, and access networks into isolated IGP/LDP domains. Seamless MPLS does not define any new protocols or technologies and is based on existing and well-known ones. Seamless MPLS provides end-to-end service-independent transport, separating the service and transport plane. Therefore, it removes the need for service-specific configurations in network transport nodes. Service provisioning is restricted only at the points of the network where it is required.

When BGP is used to distribute a route, it can also distribute an MPLS label that is mapped to that route. The label mapping information is appended to the BGP update message that is used to distribute the route. This is described in RFC 3701, *Carrying Label Information in BGPv4*.

Figure 192 shows a network with a core area and regional areas. Figure 1 also shows the control plane used in this Seamless MPLS implementation. For simplification, the control plane is displayed from right to left and the data plane from left to right. In this example, LDP will be used as the underlying transport inside each IGP domain. Alternatively, RSVP-TE could be used.

Figure 192 **Seamless MPLS - Network Topology, control and data plane**



In typical Seamless MPLS solutions, multiple ABRs are in place that result in some specific BGP configurations to send/receive multiple paths, such as the add-path feature. Due to this, ANs and ABRs will have several next hops for the same prefix, allowing the use of redundancy mechanisms such as BGP prefix independent convergence (PIC) edge, also known as BGP fast reroute (FRR). These mechanisms are beyond the scope of this chapter.

AN routers in a regional area learn the reachability of AN routers in other regional areas through BGP labeled routes redistributed by the local ABRs (RFC 3107).

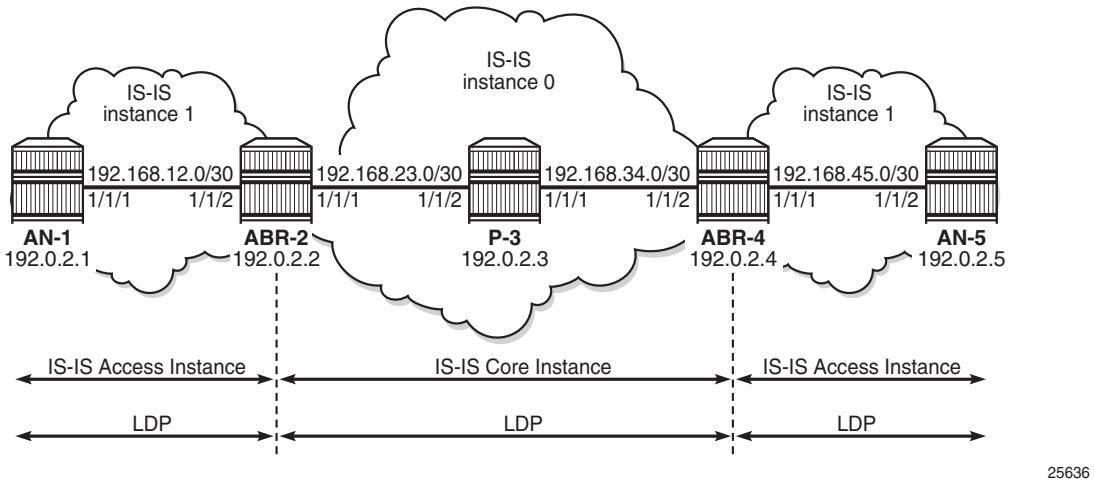
The label stack contains three labels for packets sent in a VPN service between the access nodes:

- The ANs push a service label to the packets sent in the VPN service. The service label remains unchanged end-to-end between ANs. The service label is popped by the remote AN and is the inner label of the label stack.
- The BGP label is the middle label of the label stack and should be regarded as a transport label. The transport label stack is increased to two labels: BGP and LDP transport labels. The BGP label is pushed by the iLER AN and is swapped at the BGP next hop, which can be one of the two local ABRs. Both ABRs are configured with next-hop-self. The BGP label is popped by the remote ABR.
- The iLER AN pushes an LDP transport label to the packets sent to the remote AN to reach the BGP next hop. At the local ABR, the LDP transport label is popped and a new LDP transport label is pushed to reach the BGP next hop (remote ABR). The LDP transport label is swapped in every label switching router (LSR) and popped by the ABR nearest to the remote AN. That ABR pops the BGP label and the LDP transport label, and pushes an LDP transport label to reach the remote eLER AN.

Configuration

Figure 193 shows the example topology that is used in this chapter. An Epipe and VPRN will be established between the access nodes AN-1 and AN-5. In the regional areas, and in the core area, IS-IS L2 capability is used.

Figure 193 **Seamless MPLS - IGP/LDP domains**



Initial Configuration

All nodes have the following initial configuration:

- Cards, media dependent adapters (MDAs), ports
  - Router interfaces:

```
*A:ABR-2# configure router
        interface "int-ABR-2-AN-1"
            address 192.168.12.2/30
            port 1/1/2
        exit
        interface "int-ABR-2-P-3"
            address 192.168.23.1/30
            port 1/1/1
        exit
        interface "system"
            address 192.0.2.2/32
        exit
```

- IS-IS (alternatively, OSPF could be used). Core area and regional areas run an isolated IS-IS instance. ABRs run two IS-IS instances: instance 0 belongs to the core and instance 1 belongs to the access network.
  - **Core instance**. All ABRs and Ps have level 2 (L2) capability, as follows:

```
*A:ABR-2# configure router
        isis 0
            level-capability level-2
            area-id 49.0001
            interface "system"
            exit
            interface "int-ABR-2-P-3"
                interface-type point-to-point
            exit
        exit
```

  - **Access instance**. All ABRs and ANs have L2 capability, as follows:

```
*A:ABR-2# configure router
        isis 1
            level-capability level-2
            interface "system"
            exit
            interface "int-ABR-2-AN-1"
                interface-type point-to-point
            exit
        exit
```

- LDP

  Link LDP is enabled on all router interfaces on all nodes, as follows:

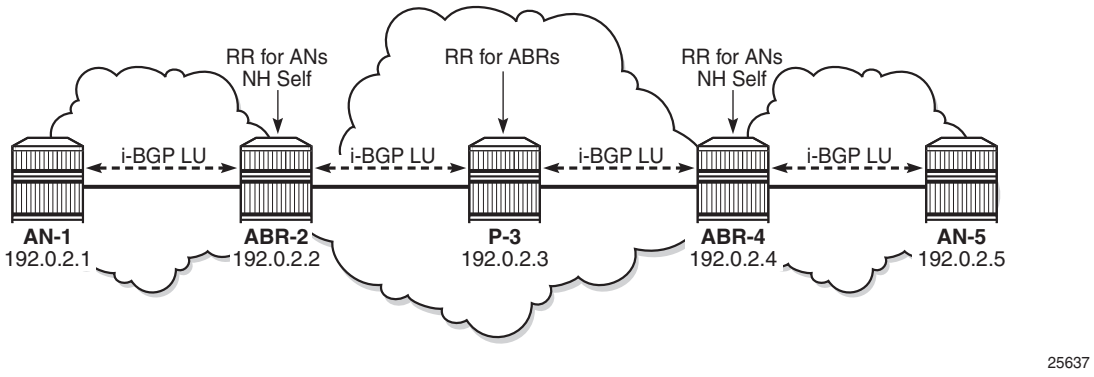
```
*A:ABR-2# configure router
        ldp
            interface-parameters
```

```
            interface "int-ABR-2-AN-1"
            exit
            interface "int-ABR-2-P-3"
            exit
        exit
    exit
```

Configure BGP

BGP is configured on all ABRs and all ANs. P-3 acts as a core route reflector (RR). To allow for separation of core/access IGP domains, the ABRs become RRs inline and implement next-hop-self on labeled IPv4 BGP prefixes. Figure 194 shows the exchange of iBGP labeled unicast (LU) routes.

Figure 194 **Seamless MPLS - BGP**



BGP configuration on ABRs

There are two groups. One group toward the core RR and another group toward the AN, as follows:

```
*A:ABR-2# configure router
        autonomous-system 64496
        bgp
            group "core"
                family ipv4 vpn-ipv4
                peer-as 64496
                advertise-inactive
                neighbor 192.0.2.3
                    description "coreRR_P-3"
                    next-hop-self
                    advertise-label ipv4
                exit
            exit
```

```
            no shutdown
        exit
```

Advertise-inactive must be enabled on the BGP group toward the core. The /32 system IP addresses, learned in labeled BGP, will also be learned in IS-IS. Because IS-IS has a lower preference compared to iBGP, the IS-IS routes will be installed in the routing table. BGP default behavior only advertises those prefixes that were elected by RTM and used. The VPNv4 address family is also included, along with labeled IPv4, to allow setting up L3 VPN services, as shown in next sections. The next-hop attribute of VPNv4 prefixes remains unchanged.

```
*A:ABR-2# configure router
        bgp
            group "ANs_Label_IPv4+VPNv4"
                family ipv4 vpn-ipv4
                cluster 2.2.2.2
                peer-as 64496
                neighbor 192.0.2.1
                    description "AN-1"
                    next-hop-self
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
```

BGP configuration on the core RR

```
*A:P-3# configure router
        autonomous-system 64496
        bgp
            cluster 3.3.3.3
            group "core"
                family ipv4 vpn-ipv4
                peer-as 64496
                advertise-inactive
                neighbor 192.0.2.2
                    description "ABR-2"
                    advertise-label ipv4
                exit
                neighbor 192.0.2.4
                    description "ABR-4"
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
```

BGP configuration on ANs toward ABRs

```
*A:AN-1# configure router
        autonomous-system 64496
        bgp
            group "ABRs_Label_IPv4+VPNv4"
                family ipv4 vpn-ipv4
                peer-as 64496
                neighbor 192.0.2.2
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
```

Configuring **advertise-label ipv4** implies that all IPv4 prefixes advertised will be sent to the remote BGP peer as an RFC 3107 formatted label. The next-hop-self command only applies to IPv4 prefixes.

The BGP sessions can be shown with the following command:

```
*A:P-3# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.3        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 2
Total BGP Paths         : 7           Total Path Memory           : 1288
---snip---
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
ABR-2
                64496      67    0 00h30m40s 0/0/0 (IPv4)
                           70    0           0/0/0 (VpnIPv4)
192.0.2.4
ABR-4
                64496      69    0 00h30m34s 0/0/0 (IPv4)
                           69    0           0/0/0 (VpnIPv4)
-------------------------------------------------------------------------------
*A:P-3#

*A:AN-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
```

```
Total BGP Paths         : 5           Total Path Memory           : 920
---snip---
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
                64496      39    0 00h16m36s 0/0/0 (IPv4)
                           41    0           0/0/0 (VpnIPv4)
-------------------------------------------------------------------------------
*A:AN-1#
```

Export Policies for BGP

A policy is required on the ANs to advertise the system IP address in labeled BGP toward the ABRs. The same policy is required on the ABRs to advertise their system IP address in labeled BGP toward the core and the AN.

Policy configuration on ANs and ABRs

```
configure
    router
        policy-options
            begin
            prefix-list "system"
                prefix 192.0.2.1/32 exact
            exit
            policy-statement "export-system"
                entry 10
                    from
                        protocol direct
                        prefix-list "system"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

This export policy must be applied in the BGP context on AN-1: either in the general settings or per group or per neighbor, as follows:

```
*A:AN-1# configure router
```

```
        bgp
            group "ABRs_Label_IPv4+VPNv4"
                export "export-system"
            exit
        exit
```

This export policy is applied in the group "core" on the ABRs, as follows:

```
*A:ABR-2# configure router
        bgp
            group "core"
                export "export-system"
            exit
        exit
```

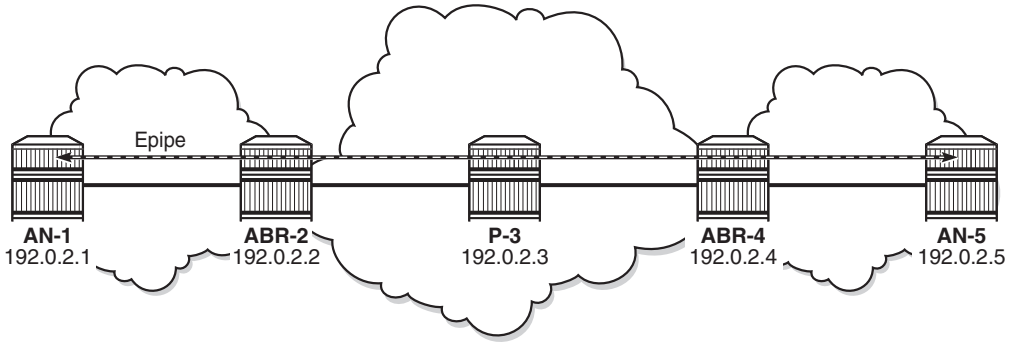
The prefix of the remote AN is added to the routing table in AN-1 and services can be configured in the ANs. No service configuration is required in the transit nodes.

```
*A:AN-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     00h06m58s   0
       system                                                          0
192.0.2.2/32                                    Remote  ISIS      00h06m12s   18
       192.168.12.2                                                    10
192.0.2.5/32                                    Remote  BGP       00h00m40s   170
       192.0.2.2 (tunneled)                                            0
192.168.12.0/30                                 Local   Local     00h06m58s   0
       int-AN-1-ABR-2                                                  0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:AN-1#
```

Configure SDP and Epipe

An end-to-end Epipe service is established between AN-1 and AN-5, as shown in Figure 195.

Figure 195 **End-to-End Epipe Service**



The SDP is configured on AN-1 and AN-5, as follows:

```
*A:AN-1# configure service
        sdp 15 mpls create
            far-end 192.0.2.5
            bgp-tunnel
            no shutdown
        exit

*A:AN-5# configure service
        sdp 51 mpls create
            far-end 192.0.2.1
            bgp-tunnel
            no shutdown
        exit
```

Epipe 1 is configured on AN-1 and AN-5, as follows:

```
*A:AN-1# configure service
        epipe 1 customer 1 create
            sap 1/1/3:1 create
            exit
            spoke-sdp 15:1 create
                no shutdown
            exit
            no shutdown
        exit

*A:AN-5# configure service
        epipe 1 customer 1 create
            sap 1/1/3:1 create
            exit
            spoke-sdp 51:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The state of the SDP and of the Epipe service can be verified on AN-1, as follows:

```
*A:AN-1# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
15     0       1552    192.0.2.5        Up   Up          MPLS    B     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
===============================================================================
*A:AN-1#

*A:AN-1# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 03/30/2016 13:21:11
Last Mgmt Change  : 03/30/2016 13:20:57
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                             Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/3:1                            q-tag        1518    1518    Up   Up
sdp:15:1 S(192.0.2.5)                  Spok         0       1552    Up   Up
===============================================================================
*A:AN-1#
```

The state of the SDP and of the Epipe service can be verified on AN-5, as follows:

```
*A:AN-5# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
51     0       1552    192.0.2.1        Up   Up          MPLS    B     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
```

```
===============================================================================
*A:AN-5#

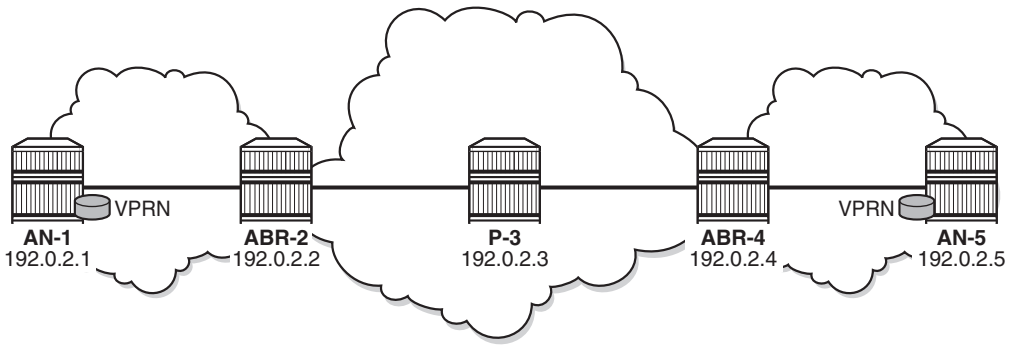
*A:AN-5# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 03/30/2016 13:21:20
Last Mgmt Change  : 03/30/2016 13:21:13
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm   Opr
-------------------------------------------------------------------------------
sap:1/1/3:1                             q-tag        1518    1518    Up    Up
sdp:51:1 S(192.0.2.1)                   Spok         0       1552    Up    Up
===============================================================================
*A:AN-5#
```

Configure VPRN

An L3 VPN service is established on AN-1 and AN-5, as shown in Figure 196.

Figure 196 **L3 VPN service**



The VPRN service is configured on AN-1 and AN-5, as follows. For simplicity, no CEs are attached to the ANs and only one loopback is created for verification.

```
*A:AN-1# configure service
        vprn 2 customer 1 create
            route-distinguisher 64496:2
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64496:2
            interface "loopback" create
                address 192.0.1.1/32
                loopback
            exit
            no shutdown
        exit

*A:AN-5#  configure service
        vprn 2 customer 1 create
            route-distinguisher 64496:2
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64496:2
            interface "loopback" create
                address 192.0.1.5/32
                loopback
            exit
            no shutdown
        exit
```

The routing table for VPRN 2 contains the local and the remote loopback addresses. On AN-1, this can be verified as follows:

```
*A:AN-1# show router 2 route-table
===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
192.0.1.1/32                                    Local   Local     00h04m28s  0
       loopback                                                       0
192.0.1.5/32                                    Remote  BGP VPN   00h03m05s  170
       192.0.2.5 (tunneled:BGP)                                       0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:AN-1#
```

On AN-5, this can be verified as follows:

```
*A:AN-5# show router 2 route-table
===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.1.1/32                                    Remote  BGP VPN   00h00m49s  170
       192.0.2.1 (tunneled:BGP)                                        0
192.0.1.5/32                                    Local   Local     00h01m38s  0
       loopback                                                        0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:AN-5#
```

Ping messages can be sent from the loopback address in VPRN 2 on AN-1 to the remote loopback address in VPRN 2 on AN-5, as follows:

```
*A:AN-1# ping router 2 192.0.1.5
PING 192.0.1.5 56 data bytes
64 bytes from 192.0.1.5: icmp_seq=1 ttl=64 time=2.22ms.
64 bytes from 192.0.1.5: icmp_seq=2 ttl=64 time=2.11ms.
64 bytes from 192.0.1.5: icmp_seq=3 ttl=64 time=3.00ms.
64 bytes from 192.0.1.5: icmp_seq=4 ttl=64 time=2.14ms.
64 bytes from 192.0.1.5: icmp_seq=5 ttl=64 time=2.88ms.
---- 192.0.1.5 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.11ms, avg = 2.47ms, max = 3.00ms, stddev = 0.387ms
*A:AN-1#
```

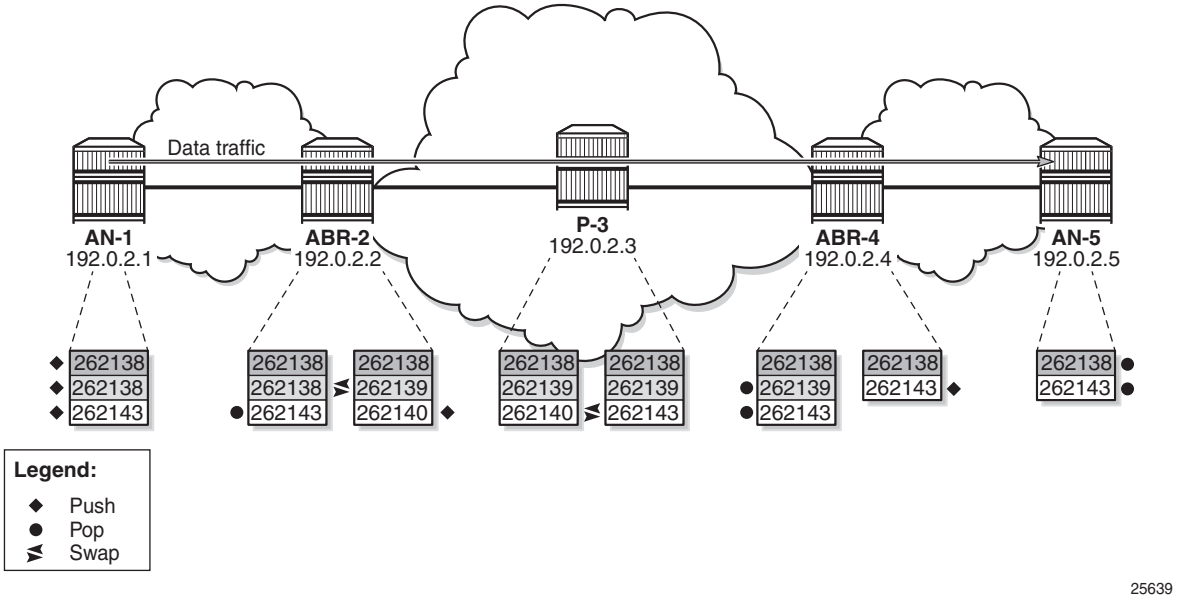
In a similar way, ping messages are sent from the loopback address in VPRN 2 on AN-5 to the loopback address in VPRN 2 on AN-1, as follows:

```
*A:AN-5# ping router 2 192.0.1.1
PING 192.0.1.1 56 data bytes
64 bytes from 192.0.1.1: icmp_seq=1 ttl=64 time=2.10ms.
64 bytes from 192.0.1.1: icmp_seq=2 ttl=64 time=1.94ms.
64 bytes from 192.0.1.1: icmp_seq=3 ttl=64 time=1.91ms.
64 bytes from 192.0.1.1: icmp_seq=4 ttl=64 time=2.11ms.
64 bytes from 192.0.1.1: icmp_seq=5 ttl=64 time=3.10ms.
---- 192.0.1.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.91ms, avg = 2.23ms, max = 3.10ms, stddev = 0.441ms
*A:AN-5#
```

Data Plane Overview

Figure 197 shows the label stacks used for traffic from AN-1 to AN-5. As an example, an Epipe service is used.

Figure 197 **Label Stacks for Traffic from AN-1 to AN-5**



1. The service label added for Epipe 1 on AN-1 for egress traffic to AN-5 is 262138. Ingress traffic on AN-1 has service label 262138. This can be shown as follows:

```
*A:AN-1# show service id 1 labels
===============================================================================
Martini Service Labels
===============================================================================
Svc Id     Sdp Binding        Type  I.Lbl                E.Lbl
-------------------------------------------------------------------------------
1          15:1               Spok  262138               262138
-------------------------------------------------------------------------------
Number of Bound SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:AN-1#
```

   This service label remains unchanged end-to-end.

   On AN-1, the (middle) BGP label for traffic with destination AN-5 is 262138, as follows:

```
*A:AN-1# show router bgp routes 192.0.2.5
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
```

```
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.0.2.5/32                                         100         None
      192.0.2.2                                            None        262138
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:AN-1#
```

The next hop is ABR-2. AN-1 pushes the LDP label 262143 to reach ABR-2, as follows:

```
*A:AN-1# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Push              --      262143
192.168.12.2                            1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 1
===============================================================================
*A:AN-1#
```

2. At ABR-2, the service label 262138 remains unchanged. The LDP label 262143 is popped, as follows:

```
*A:ABR-2# show router ldp bindings active prefixes prefix 192.0.2.2/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
```

```
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.2/32                            Pop             262143      --
  --                                      --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 1
===============================================================================
*A:ABR-2#
```

On ABR-2, the BGP next hop is ABR-4 for prefix 192.0.2.5, as follows:

```
*A:ABR-2# show router bgp routes 192.0.2.5
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.0.2.5/32                                      100         None
      192.0.2.4                                         None        262139
      No As-Path
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:ABR-2#
```

On ABR-2, the BGP label 262138 is swapped with 262139 for BGP next hop ABR-4, as follows:

```
*A:ABR-2# show router bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                      Received       Advertised     Label
                             Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.1                    262139         262139         Internal
192.0.2.4                    262139         262138         Internal
-------------------------------------------------------------------------------
Total Labels allocated:   2
```

```
===============================================================================
*A:ABR-2#
```

ABR-2 pushes a new LDP label (262140) to reach the BGP next hop (ABR-4), as follows:

```
*A:ABR-2# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                              Push              --      262140
192.168.23.2                              1/1/1

192.0.2.4/32                              Swap            262140    262140
192.168.23.2                              1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:ABR-2#
```

3. At LSR P-3, only an LDP label swap occurs. P-3 swaps LDP label 262140 with 262143, as follows:

```
*A:P-3# show router ldp bindings active prefixes prefix 192.0.2.4/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                              Push              --      262143
192.168.34.2                              1/1/1
```

```
192.0.2.4/32                                    Swap            262140    262143
192.168.34.2                                    1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:P-3#
```

4. At ABR-4, LDP label 262143 is popped and BGP label 262139 is popped because AN-5 is known by LDP, as follows:

```
*A:ABR-4# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      bgp       MPLS  262145    12       192.0.2.2       1000
192.0.2.2/32      ldp       MPLS  65538     9        192.168.34.1    20
192.0.2.3/32      ldp       MPLS  65537     9        192.168.34.1    10
192.0.2.5/32      ldp       MPLS  65539     9        192.168.45.2    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:ABR-4#
```

ABR-4 pushes a new LDP label 262143 to reach AN-5, as follows:

```
*A:ABR-4# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                          Op              IngLbl    EgrLbl
EgrNextHop                                      EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                                    Push              --      262143
192.168.45.2                                    1/1/1

192.0.2.5/32                                    Swap            262140    262143
192.168.45.2                                    1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:ABR-4#
```

5. Finally, at AN-5, the LDP label 262143 is popped, as follows:

```
*A:AN-5# show router ldp bindings active prefixes prefix 192.0.2.5/32
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.5/32                            Pop             262143      --
  --                                      --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 1
===============================================================================
*A:AN-5#
```

Then, the ingress service label 262138 is popped, as follows:

```
*A:AN-5# show service id 1 labels
===============================================================================
Martini Service Labels
===============================================================================
Svc Id     Sdp Binding        Type  I.Lbl                E.Lbl
-------------------------------------------------------------------------------
1          51:1               Spok  262138               262138
-------------------------------------------------------------------------------
Number of Bound SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:AN-5#
```

OAM

The following operations, administration, and maintenance (OAM) commands can be launched to validate reachability between regions using BGP IPv4 labeled routes.

```
*A:AN-1# oam lsp-ping bgp-label prefix 192.0.2.5/32
LSP-PING 192.0.2.5/32: 80 bytes MPLS payload
Seq=1, send from intf int-AN-1-ABR-2, reply from 192.0.2.5
       udp-data-len=32 ttl=255 rtt=2.11ms rc=3 (EgressRtr)
---- LSP 192.0.2.5/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 2.11ms, avg = 2.11ms, max = 2.11ms, stddev = 0.000ms
```

```
*A:AN-1#

*A:AN-5# oam lsp-ping bgp-label prefix 192.0.2.1/32
LSP-PING 192.0.2.1/32: 80 bytes MPLS payload
Seq=1, send from intf int-AN-5-ABR-4, reply from 192.0.2.1
       udp-data-len=32 ttl=255 rtt=2.21ms rc=3 (EgressRtr)
---- LSP 192.0.2.1/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 2.21ms, avg = 2.21ms, max = 2.21ms, stddev = 0.000ms
*A:AN-5#
```

In a similar way, LSP trace can validate the reachability of the remote AN, as follows:

```
*A:AN-1# oam lsp-trace bgp-label prefix 192.0.2.5/32
lsp-trace to 192.0.2.5/32: 0 hops min, 0 hops max, 104 byte packets
1  192.0.2.2  rtt=0.666ms rc=8(DSRtrMatchLabel) rsc=1
2  192.0.2.4  rtt=1.78ms rc=8(DSRtrMatchLabel)
3  192.0.2.5  rtt=2.09ms rc=3(EgressRtr) rsc=1
*A:AN-1#

*A:AN-5# oam lsp-trace bgp-label prefix 192.0.2.1/32
lsp-trace to 192.0.2.1/32: 0 hops min, 0 hops max, 104 byte packets
1  192.0.2.4  rtt=0.631ms rc=8(DSRtrMatchLabel) rsc=1
2  192.0.2.2  rtt=1.48ms rc=8(DSRtrMatchLabel)
3  192.0.2.1  rtt=2.06ms rc=3(EgressRtr) rsc=1
*A:AN-5#
```

Conclusion

Seamless MPLS helps to solve the scalability problems of large networks. Seamless MPLS partitions the core, aggregation, and access networks into isolated IGP/LDP domains, which helps to maintain IGP databases small and controlled. Label BGP allows the establishment of hierarchical LSPs for end-to-end service set up.

Segment Routing with IS-IS Control Plane

This chapter provides information about segment routing (SR) with intermediate system to intermediate system (IS-IS) control plane.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

Segment routing is supported in SR OS release 13.0, and later. This chapter was originally written for SR OS release 13.0.R3, but the CLI in this edition corresponds to release 14.0.R1.

Overview

Segment routing (SR) is a technology for IP/multiprotocol label switching (MPLS) networks that enables source routing. With source routing, operators can specify a forwarding path, from ingress to egress, that is independent of the shortest path determined by the interior gateway protocol (IGP).

The main benefit of segment routing compared to other source routing protocols (such as resource reservation protocol with traffic engineering (RSVP-TE)) is that, from a control plane perspective, no signaling protocol is required. Segment routing provides a path or tunnel, encoded as a sequential list of sub-paths or segments that are advertised within the segment routing domain, using extensions to well-known link state routing protocols, such as IS-IS or open shortest path first (OSPF).

Implementation

A segment routing tunnel can contain a single segment that represents the destination node, or it can contain a list of segments that the tunnel must traverse. The tunnel can be established over an IPv4/IPv6 MPLS or IPv6 data plane, encoded as a stack of MPLS labels or as a number of IPv6 addresses contained in an IPv6 extension header.

Network elements are modeled as segments. For each segment, IGP advertises an identifier referred to as a segment ID (SID).

The two segment types are:

- Prefix segment — Globally unique and allocated from a segment routing global block (SRGB), typically multi-hop and signaled by the IGP. It is the equal cost multi-path ECMP-aware shortest path IGP route to a related prefix. A typical example of a prefix segment is a node SID. Within the SR OS implementation, the node SID is either the system address or another interface address in the global routing table (GRT) of type loopback. Node SIDs are advertised in IS-IS using a prefix SID sub-TLV (type length value).
- Adjacency segment — Locally unique and allocated from the (local) dynamic label space, so that other routers in the SR domain can use the same label space. Adjacency segments are signaled by the IGP. Within the SR OS implementation, Adjacency SIDs are automatically assigned and advertised when the SR context within the IGP instance is set in no shutdown. Adjacency SIDs are advertised in IS-IS using an adjacency SID sub-TLV.

To make prefix segments globally unique within the segment routing domain, an indexing mechanism is required, because production networks consist of multiple vendors and multiple products. As a result, it is often difficult to agree on a common SRGB for the prefix SIDs.

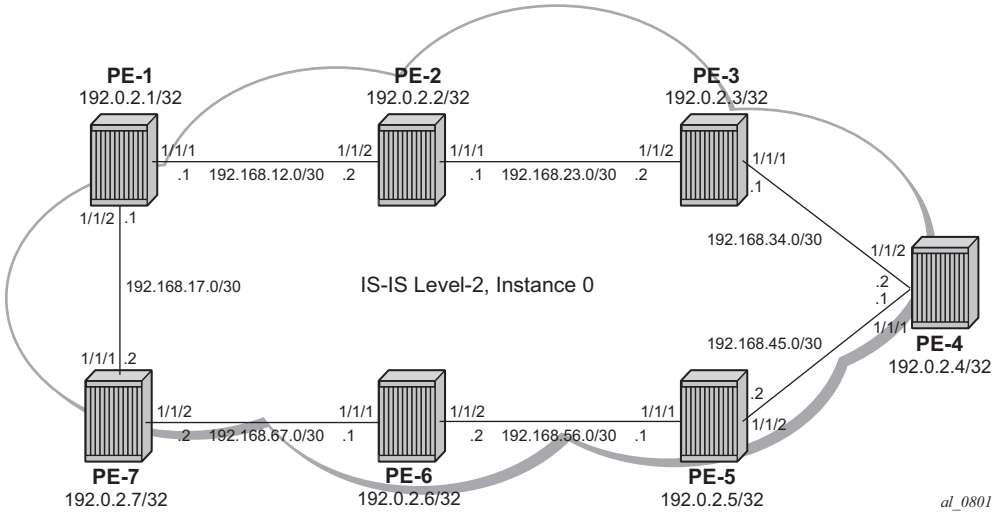
All routers within the SR domain are expected to configure and advertise the same Prefix SID index range for an IGP instance. The label value used by each router to represent a prefix can be local to that router by the use of an offset label, referred to as a start label:

Local label (for a prefix) = (local) start label + {Prefix SID index}

Within the SR OS implementation, prefix loop-free alternate (LFA) is supported for segment routing to improve the fast reroute (FRR) coverage. Remote LFA (RLFA) is also supported. With RLFA, segment routing shortest path tunnels are used as a virtual LFA or repair tunnel toward the PQ node.

The following example uses IS-IS as an IGP protocol, with an MPLS data plane and services enabled using LFA and RLFA.

Figure 198 **Network Topology**



Configuration

Step 1. Configure an IP/MPLS network according to Figure 198.

- The system and IP interface addresses are configured according to Figure 1.
- IS-IS level 2 is selected as the IGP to distribute routing information between all PEs. All IS-IS interfaces are of type point-to-point to avoid running the DR/BDR (Designated Router/Backup Designated Router) election process.

Step 2. Configure segment routing.

Before enabling segment routing on a router, define a dedicated SRGB. This SRGB is required on each individual router part of the SR domain and is used to allocate the Prefix SIDs.

By default, an SRGB is not instantiated and, when configured by the operator, it is taken from the system dynamic label range. Translated into a show command:

```
*A:PE-1# show router mpls-labels label-range
===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505856      505856
    Seg-Route   0           0           -           0           505856
===============================================================================
```

```
*A:PE-1#
```

For simplicity, the same SRGB is used in this example for all SR domain routers. Within the command, a start value and end value defines the size of the SRGB. Translated into configuration commands for an SRGB of 100 MPLS labels, this becomes:

```
*A:PE-1# configure router mpls-labels sr-labels start 20000 end 20099

*A:PE-1# show router mpls-labels label-range
===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505756      505856
    Seg-Route   20000       20099       -           0           100
===============================================================================
*A:PE-1#
```

This command is repeated for all other nodes. The allocated MPLS labels are only for the prefix SIDs. The adjacency SIDs, which are only locally unique, are taken from the dynamic range; in this example, between 18432 and 524287.

1. Enable router capability in the IGP instance.

   It is mandatory to enable the router-capability parameter inside the IS-IS instance, to advertise SR support among the IS-IS adjacencies. By configuring this command within the IGP instance, the SR capability sub-TLV is propagated and is used to indicate the index range and the start label. The SR algorithm sub-TLV is also used to advertise the algorithm used for path calculations. Only shortest path first (SPF) (value 0) is defined. This is configured as follows:

```
<all-nodes-within-SR-domain># configure router isis advertise-router-capability area
```

   The flooding parameter is a mandatory parameter in this CLI command. The keyword area or as indicates that the router capabilities label switched path (LSP) should be advertised throughout the same level or throughout the whole autonomous system (AS). In the preceding example, all routers belong to the same level, so the area argument is sufficient. When the SR context within the IGP instance is set in no shutdown, both IS-IS sub-TLVs are flooded.

2. Define the Prefix SID index range.

The SR OS implementation for SR provides two mutually exclusive modes of operation to define the Prefix SID index range: global mode and per-instance mode. Per-instance mode is useful in a seamless MPLS environment when multiple IGP instances are used. The main difference between the modes is the way that the start label and index range are calculated.

A comparison of the modes is shown in following table:

Table 16 **Mode Comparison**

Global	Per Instance
Applicable for all IGP instances on that node	Applicable for one dedicated IGP instance
Start label is first label of SRGB	Start label is configurable (but part of SRGB range); use of non-overlapping sub-ranges of SRGB
Prefix SID index range is “size” of SRGB	Prefix SID index-range is configurable
If SRGB needs to change, shutdown SR and delete prefix-SID-ranges in all IGP instances	If prefix SID index and/or label range needs to change, shut down SR in that specific IGP instance
SW checks whether any allocated SID index/label goes out of range. SW checks also for overlaps of the resulting net label value range across IGP instances.	

For simplicity, global mode is used for this example, as follows:

```
<all-nodes-within-SR-domain># configure router isis segment-routing prefix-sid-range
global
```

3. Assign a Prefix SID index or label to the prefix representing a node.

   To be able to set up SR shortest path tunnels to all routers of the SR domain, each router needs to be uniquely defined within the SR domain. Therefore, the system address or other loopback interface in the GRT will be assigned an ipv4-node-sid index or label value that is unique within the SR domain.

   Translated into configuration commands, this becomes:

```
*A:PE-1# configure router isis interface "system" ipv4-node-sid index 1
*A:PE-2# configure router isis interface "system" ipv4-node-sid index 2
*A:PE-3# configure router isis interface "system" ipv4-node-sid index 3
*A:PE-4# configure router isis interface "system" ipv4-node-sid index 4
*A:PE-5# configure router isis interface "system" ipv4-node-sid index 5
*A:PE-6# configure router isis interface "system" ipv4-node-sid index 6
*A:PE-7# configure router isis interface "system" ipv4-node-sid index 7
```

Because the SRGB is the same on all nodes, each node in the network can be reached using the same MPLS label. For example, the node SID for PE-5 on all nodes has a start label (first label of the SRGB (= 20000) + ipv4-node-sid index on node PE-5 (= 5)) of 20005.

When there is one consistent SRGB for the SR domain, the SR OS CLI allows the use of absolute MPLS label values instead of index values. For example, on PE-1, an operator can use an explicit MPLS label value:

```
*A:PE-1# configure router isis interface "system" ipv4-node-sid label 20001
```

Internally, this explicit value is translated into an index value (index-value 1) before advertising it toward its neighbors, taking into account the prefix SID index-range mode (global or per-instance) and the SRGB.

4. Enable SR context within the IGP instance, as follows:

```
<all-nodes-within-SR-domain># configure router isis segment-routing no shutdown
```

After enabling the SR context within an IGP instance, the SR Capability sub-TLV, and the SR Algorithm sub-TLV between all routers within the SR domain, are flooded. A show command is available to display the SR related router capability information. For example, on PE-1, this becomes:

```
*A:PE-1# show router isis capabilities level 2
===============================================================================
Rtr Base ISIS Instance 0 Capabilities
===============================================================================
Displaying Level 2 capabilities
-------------------------------------------------------------------------------
LSP ID    : PE-1.00-00
  Router Cap : 192.0.2.1, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
LSP ID    : PE-2.00-00
  Router Cap : 192.0.2.2, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
LSP ID    : PE-3.00-00
  Router Cap : 192.0.2.3, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
LSP ID    : PE-4.00-00
  Router Cap : 192.0.2.4, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
```

```
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
LSP ID    : PE-5.00-00
  Router Cap : 192.0.2.5, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
LSP ID    : PE-6.00-00
  Router Cap : 192.0.2.6, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
LSP ID    : PE-7.00-00
  Router Cap : 192.0.2.7, D:0, S:0
    TE Node Cap : B E M  P
    SR Cap: IPv4
       SRGB Base:20000, Range:100
    SR Alg: metric based SPF
Level (2) Capability Count : 7
===============================================================================
*A:PE-1#
```

A similar output occurs for each router in the SR domain.

After enabling the SR context within the IGP instance, the assigned index for each locally configured Prefix SID is advertised. After the advertisement of prefix SIDs, MPLS data plane ingress label mapping (ILM) is programmed with a pop operation. In this context, a show command can be used to display the prefix SIDs, in order, within the SR domain. As an example, on PE-1, this becomes:

```
*A:PE-1# show router isis prefix-sids
===============================================================================
Rtr Base ISIS Instance 0 Prefix/SID Table
===============================================================================
Prefix                          SID        Lvl/Typ    SRMS   AdvRtr
                                                       MT     Flags
-------------------------------------------------------------------------------
192.0.2.1/32                    1          2/Int.     N      PE-1
                                                         0      NnP
192.0.2.2/32                    2          2/Int.     N      PE-2
                                                         0      NnP
192.0.2.3/32                    3          2/Int.     N      PE-3
                                                         0      NnP
192.0.2.4/32                    4          2/Int.     N      PE-4
                                                         0      NnP
192.0.2.5/32                    5          2/Int.     N      PE-5
                                                         0      NnP
192.0.2.6/32                    6          2/Int.     N      PE-6
                                                         0      NnP
192.0.2.7/32                    7          2/Int.     N      PE-7
                                                         0      NnP
-------------------------------------------------------------------------------
No. of Prefix/SIDs: 7
Flags: R  = Re-advertisement
       N  = Node-SID
```

```
       nP = no penultimate hop POP
       E  = Explicit-Null
       V  = Prefix-SID carries a value
       L  = value/index has local significance
===============================================================================
*A:PE-1#
```

The SR OS implementation, by default, sets the node SID (or N–flag) and noPenultimate hop PoP (or nP–flag) inside the prefix SID TLV. Another useful flag that can be set is the re-advertisement (or R-flag). The R-flag is set when a prefix SID is propagated between levels or areas, or redistribution is in place (from another protocol).

Prefix SID information can also be viewed within the IGP database attached to (extended) IP prefix reachability TLVs. For example, on PE-1, this becomes:

```
*A:PE-1# show router isis database level 2 PE-1.00-00 detail
===============================================================================
Rtr Base ISIS Instance 0 Database (detail)
===============================================================================
Displaying Level 2 database
-------------------------------------------------------------------------------
LSP ID    : PE-1.00-00                                  Level     : L2
Sequence  : 0x5                  Checksum  : 0x830c   Lifetime  : 1111
Version   : 1                    Pkt Type  : 20       Pkt Ver   : 1
Attributes: L1L2                 Max Area  : 3        Alloc Len : 1492
SYS ID    : 1920.0000.2001       SysID Len : 6        Used Len  : 248

TLVs :
  Supp Protocols:
    Protocols     : IPv4
  IS-Hostname   : PE-1
  Router ID   :
    Router ID   : 192.0.2.1
  ---snip---
  TE IP Reach   :
    ---snip---
    Default Metric  : 0
    Control Info:   S, prefLen 32
    Prefix   : 192.0.2.1
    Sub TLV   :
      Prefix-SID Index:1, Algo:0, Flags:NnP

Level (2) LSP Count : 1
```

After enabling the SR context within the IGP instance, adjacency SIDs are also automatically assigned and advertised for each formed adjacency over an IP interface. From a data plane perspective, one local adjacency SID consumes one ILM entry, programming a pop operation.

Similar to prefix SIDs, adjacency SID information can be viewed within the IGP database attached to IS neighbor TLVs. Translated into the previous show command on PE-1, this becomes:

```
*A:PE-1# show router isis database level 2 PE-1.00-00 detail
===============================================================================
Rtr Base ISIS Instance 0 Database (detail)
===============================================================================
Displaying Level 2 database
-------------------------------------------------------------------------------
LSP ID    : PE-1.00-00                                  Level     : L2
Sequence  : 0x5                  Checksum  : 0x830c     Lifetime  : 1111
Version   : 1                    Pkt Type  : 20         Pkt Ver   : 1
Attributes: L1L2                 Max Area  : 3          Alloc Len : 1492
SYS ID    : 1920.0000.2001       SysID Len : 6          Used Len  : 248

TLVs :
  Supp Protocols:
    Protocols     : IPv4
  IS-Hostname   : PE-1
  Router ID   :
    Router ID   : 192.0.2.1
  ---snip---
 TE IS Nbrs   :
    Nbr   : PE-2.00
    Default Metric  : 10
    Sub TLV Len     : 19
    IF Addr   : 192.168.12.1
    Nbr IP    : 192.168.12.2
    Adj-SID: Flags:v4VL Weight:0 Label:262143
  TE IS Nbrs   :
    Nbr   : PE-7.00
    Default Metric  : 10
    Sub TLV Len     : 19
    IF Addr   : 192.168.17.1
    Nbr IP    : 192.168.17.2
    Adj-SID: Flags:v4VL Weight:0 Label:262142
  ---snip---
```

The SR OS implementation, by default, sets the value (or V–flag), meaning that the adjacency SID carries a value (as opposed to an index). Also, the local (or L-flag) is set by default, meaning that the adjacency SID has only local significance. The v4-flag set to 0 means that the adjacency SID references to an adjacency with outgoing IPv4 encapsulation.

Another way to display adjacency SID information is using the show router isis adjacency detail command.

```
*A:PE-1# show router isis adjacency "int-PE-1-PE-2" detail
===============================================================================
Rtr Base ISIS Instance 0 Adjacency (detail)
===============================================================================
SystemID    : PE-2                              SNPA        : 4a:c5:01:01:00:02
Interface   : int-PE-1-PE-2                     Up Time     : 0d 00:07:08
State       : Up                                Priority    : 0
Nbr Sys Typ : L2                                L. Circ Typ : L2
```

```
Hold Time   : 22                                  Max Hold    : 27
Adj Level   : L2                                  MT Enabled  : No
Topology    : Unicast

IPv6 Neighbor     : ::
IPv4 Neighbor     : 192.168.12.2
IPv4 Adj SID      : Label 262143
Restart Support   : Disabled
Restart Status    : Not currently being helped
Restart Supressed : Disabled
Number of Restarts: 0
Last Restart at   : Never

===============================================================================


*A:PE-1# show router isis adjacency "int-PE-1-PE-7" detail
===============================================================================
Rtr Base ISIS Instance 0 Adjacency (detail)
===============================================================================
SystemID    : PE-7                                SNPA        : 4a:a4:01:01:00:01
Interface   : int-PE-1-PE-7                       Up Time     : 0d 00:06:39
State       : Up                                  Priority    : 0
Nbr Sys Typ : L2                                  L. Circ Typ : L2
Hold Time   : 23                                  Max Hold    : 27
Adj Level   : L2                                  MT Enabled  : No
Topology    : Unicast

IPv6 Neighbor     : ::
IPv4 Neighbor     : 192.168.17.2
IPv4 Adj SID      : Label 262142
Restart Support   : Disabled
Restart Status    : Not currently being helped
Restart Supressed : Disabled
Number of Restarts: 0
Last Restart at   : Never

===============================================================================
*A:PE-1#
```

Finally, when enabling the SR context within the IGP instance, the SR module resolves received prefixes with prefix SID sub-TLVs present. As a result, MPLS data plane resources are consumed. The ILM is programmed with a swap operation and the label-to-next-hop-label-forwarding-entry (LTN) with a push operation, both pointing to the primary and/or LFA next-hop label forwarding entry (NHLFE). Also, an SR tunnel is added in the tunnel table manager (TTM). As a result, an SR shortest path tunnel is set up to each other router that is part of the SR domain. Now, SR shortest path tunnels can be used for all users of TTM.

Example 1: VPRN service with LFA and RLFA enabled

In the network topology of Figure 198, no LDP and RSVP-TE signaling protocols are enabled. Each router of the SR domain has a full mesh of SR shortest path tunnels to the other routers, and no LDP and RSVP-TE LSPs are present. For example, on PE-1, the TTM looks like:

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.2/32      isis (0)  MPLS  524291    11       192.168.12.2    10
192.0.2.3/32      isis (0)  MPLS  524295    11       192.168.12.2    20
192.0.2.4/32      isis (0)  MPLS  524293    11       192.168.12.2    30
192.0.2.5/32      isis (0)  MPLS  524292    11       192.168.17.2    30
192.0.2.6/32      isis (0)  MPLS  524296    11       192.168.17.2    20
192.0.2.7/32      isis (0)  MPLS  524294    11       192.168.17.2    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

The objective is to configure a VPRN between PE-1 and PE-7, using SR shortest path tunnels as transport tunnel. The configuration is as follows:

```
*A:PE-1# configure service vprn 100 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:10001
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64496:100
            interface "loopback" create
                address 192.0.1.1/32
                loopback
            exit
            no shutdown

*A:PE-7# configure service vprn 100 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:10007
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64496:100
            interface "loopback" create
                address 192.0.1.7/32
                loopback
            exit
            no shutdown
```

Within the VPRN service configuration, a loopback interface is created on both PEs to verify the transport mechanism. Tunnel information displaying the MPLS label value is retrieved using the show router fp-tunnel-table <slot number> command, as follows:

```
*A:PE-1# show router fp-tunnel-table 1 192.0.2.7/32
===============================================================================
Tunnel Table Display
Legend:
B - FRR Backup
===============================================================================
Destination                                             Protocol   Tunnel-ID
    Lbl                 NextHop           Intf/Tunnel
-------------------------------------------------------------------------------
192.0.2.7/32                                            SR-ISIS-0   -
  20007                 192.168.17.2        1/1/2
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

This means that, when traffic arrives on PE-1, the MPLS label 20007 is pushed to reach destination PE-7. Because, in this example, the prefix SID index range global mode is used, the value 20007 comes from the start label on PE-7 (first label of the SRGB, which is 20000, plus the configured index value of node SID PE-7 (7)), so 20007.

Enabling prefix LFA within the IS-IS context on PE-1 will enable LFA/FRR protection. Next-hop LFA protection is present for node PE-4, node PE-5, and the link between PE-4 and PE-5. Translated into CLI and show commands, this becomes:

```
*A:PE-1# configure router isis loopfree-alternate
*A:PE-1# show router isis lfa-coverage
===============================================================================
Router Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
---snip---
IPV4 Unicast     L2      2/6(33%)       3/11(27%)           0/0(0%)
---snip---

*A:PE-1# show router route-table alternative
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
      Alt-NextHop                                                  Alt-
                                                                   Metric
-------------------------------------------------------------------------------
---snip---
192.0.2.4/32                                    Remote  ISIS      00h27m26s  18
       192.168.12.2                                                 30
       192.168.17.2 (LFA)                                           40
```

```
192.0.2.5/32                                    Remote  ISIS      00h23m34s  18
       192.168.17.2                                                    30
       192.168.12.2 (LFA)                                              40
---snip---
192.168.45.0/30                                 Remote  ISIS      00h27m27s  18
       192.168.12.2                                                    40
       192.168.17.2 (LFA)                                              50
---snip---
-------------------------------------------------------------------------------
No. of Routes: <...>
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested


*A:PE-1# show router fp-tunnel-table 1
===============================================================================
Tunnel Table Display

Legend:
B - FRR Backup
===============================================================================
Destination                                                 Protocol   Tunnel-ID
       Lbl              NextHop           Intf/Tunnel
-------------------------------------------------------------------------------
---snip---
192.0.2.4/32                                                SR-ISIS-0   -
       20004            192.168.12.2      1/1/1
       20004            192.168.17.2(B)   1/1/2
192.0.2.5/32                                                SR-ISIS-0   -
       20005            192.168.17.2      1/1/2
       20005            192.168.12.2(B)   1/1/1


*A:PE-1# show router tunnel-table detail
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
---snip---
-------------------------------------------------------------------------------
Destination      : 192.0.2.4/32
NextHop          : 192.168.12.2
Tunnel Flags     : has-lfa exclude-for-igpshortcuts
Age              : 00h00m52s
CBF Classes      : (Not Specified)
Owner            : isis (0)             Encap            : MPLS
Tunnel ID        : 524293               Preference       : 11
Tunnel Label     : 20004                Tunnel Metric    : 30
Tunnel MTU       : 1556
-------------------------------------------------------------------------------
Destination      : 192.0.2.5/32
NextHop          : 192.168.17.2
Tunnel Flags     : has-lfa exclude-for-igpshortcuts
Age              : 00h00m52s
CBF Classes      : (Not Specified)
Owner            : isis (0)             Encap            : MPLS
Tunnel ID        : 524292               Preference       : 11
Tunnel Label     : 20005                Tunnel Metric    : 30
```

```
Tunnel MTU       : 1556
-------------------------------------------------------------------------------
---snip---
Number of tunnel-table entries with LFA : 2
```

When a failure occurs on the primary SR path (only applicable for prefix PE-4/PE-5 and the link between PE-4 and PE-5), the traffic takes the LFA backup SR path to the destination using the same MPLS label value.

To extend the LFA/FRR coverage, for example, to find an LFA protection for node PE-7, which is one of the VPRN service endpoints, RLFA can be enabled. RLFA creates a virtual LFA by using a repair tunnel to carry packets to a point in the network from where they will not be looped back to the source, but forwarded (SPF-based) toward the destination prefix.

The RLFA implementation uses the PQ algorithm. The node where RLFA is configured (PE-1 in this example) computes an extended P-space and a Q-space. The intersection of both spaces is called the PQ-node. This PQ node is the destination node of the repair tunnel using an SR shortest path tunnel. To compute both spaces, SPF is used.

In this example, IS-IS is used as the IGP, using a default metric value of 10 for all links. With the assumption that the link between PE-1and PE-7 is broken, the calculation of both the extended P-space and the Q-space at PE-1 is as follows:

- extended P-space — An SPF computed from node PE-1 and rooted at PE-2. It is used to calculate the set of routers that are reachable without any path transiting the protected link between PE-1 and PE-7. The following nodes belong to the extended P-space: PE-2, PE-3, PE-4, and PE-5.
- Q-space — A reverse SPF computed from PE-1 and rooted from PE-7 (acting as destination proxy). It is used to calculate the set of routers that can reach PE-7 without transiting the protected link between PE-1 and PE-7. The nodes PE-4, PE-5, and PE-6 belong to the Q-space.

Possible PQ-nodes are PE-4 or PE-5, because they are in the intersection of both spaces.

RLFA is configured as follows:

```
*A:PE-1# configure router isis loopfree-alternate remote-lfa

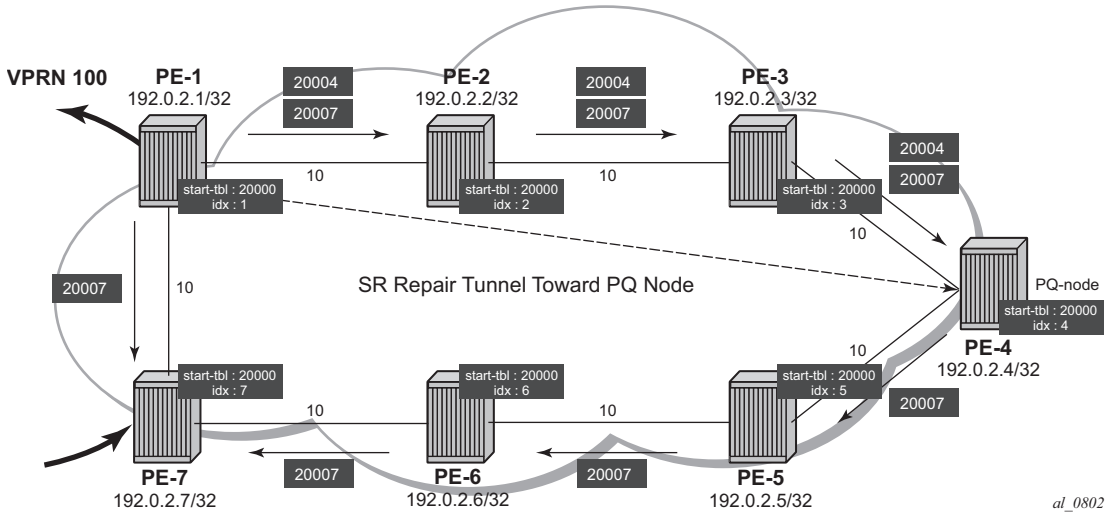
The nodes PE-2, PE-3, PE-6, and PE-7 now have RLFA protection, while PE-4 and PE-5have
LFA protection.
*A:PE-1# show router fp-tunnel-table 1
===============================================================================
Tunnel Table Display
Legend:
B - FRR Backup
===============================================================================
Destination                                             Protocol   Tunnel-ID
    Lbl                 NextHop           Intf/Tunnel
-------------------------------------------------------------------------------
```

```
192.0.2.2/32                                                  SR-ISIS-0    -
  20002                 192.168.12.2        1/1/1
  20002/20005           192.168.17.2(B)     1/1/2
192.0.2.3/32                                                  SR-ISIS-0    -
  20003                 192.168.12.2        1/1/1
  20003/20005           192.168.17.2(B)     1/1/2
192.0.2.4/32                                                  SR-ISIS-0    -
  20004                 192.168.12.2        1/1/1
  20004                 192.168.17.2(B)     1/1/2
192.0.2.5/32                                                  SR-ISIS-0    -
  20005                 192.168.17.2        1/1/2
  20005                 192.168.12.2(B)     1/1/1
192.0.2.6/32                                                  SR-ISIS-0    -
  20006                 192.168.17.2        1/1/2
  20006/20004           192.168.12.2(B)     1/1/1
192.0.2.7/32                                                  SR-ISIS-0    -
  20007                 192.168.17.2        1/1/2
  20007/20004           192.168.12.2(B)     1/1/1
---snip---
```

The main difference between normal prefix LFA and RLFA is that for RLFA a two-MPLS label stack is pushed by the head-end node (PE-1). The top label is the SR-label to reach the PQ node (for example, 20004 for PE-4) and the bottom label is the SR-label to reach the destination node (for example,20007 for PE-7). The notation inside the show command is bottom-label/top-label.

Figure 199 illustrates the RLFA traffic path protecting the link between PE-1 and PE-7:

Figure 199 **RLFA Traffic Path During Protection**



Inside the TTM, a tunnel-flag, has-lfa exclude-for-igpshortcuts, is set for all destination nodes that have LFA protection available:

```
*A:PE-1# show router tunnel-table detail
===============================================================================
Tunnel Table (Router: Base)
===============================================================================
Destination      : 192.0.2.2/32
NextHop          : 192.168.12.2
Tunnel Flags     : has-lfa exclude-for-igpshortcuts
Age              : 00h00m34s
CBF Classes      : (Not Specified)
Owner            : isis (0)             Encap            : MPLS
Tunnel ID        : 524291               Preference       : 11
Tunnel Label     : 20002                Tunnel Metric    : 10
Tunnel MTU       : 1556
-------------------------------------------------------------------------------
Destination      : 192.0.2.3/32
NextHop          : 192.168.12.2
Tunnel Flags     : has-lfa exclude-for-igpshortcuts
Age              : 00h00m34s
CBF Classes      : (Not Specified)
Owner            : isis (0)             Encap            : MPLS
Tunnel ID        : 524295               Preference       : 11
Tunnel Label     : 20003                Tunnel Metric    : 20
Tunnel MTU       : 1556
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
Destination      : 192.0.2.6/32
NextHop          : 192.168.17.2
Tunnel Flags     : has-lfa exclude-for-igpshortcuts
Age              : 00h00m34s
CBF Classes      : (Not Specified)
Owner            : isis (0)             Encap            : MPLS
Tunnel ID        : 524296               Preference       : 11
Tunnel Label     : 20006                Tunnel Metric    : 20
Tunnel MTU       : 1556
-------------------------------------------------------------------------------
Destination      : 192.0.2.7/32
NextHop          : 192.168.17.2
Tunnel Flags     : has-lfa exclude-for-igpshortcuts
Age              : 00h00m34s
CBF Classes      : (Not Specified)
Owner            : isis (0)             Encap            : MPLS
Tunnel ID        : 524294               Preference       : 11
Tunnel MTU       : 1556
Tunnel Label     : 20007                Tunnel Metric    : 10
-------------------------------------------------------------------------------
Number of tunnel-table entries with LFA : 6
---snip---
```

Verification of the loopback address configured within the VPRN service context on PE-7 (using loopback address 192.0.1.7/32) shows that an SR shortest path tunnel is used as the transport mechanism:

```
*A:PE-1# show router 100 route-table 192.0.1.7/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
```

```
Dest Prefix             : 192.0.1.7/32
  Protocol              : BGP_VPN
  Age                   : 00h00m57s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.7
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.7 (SR-ISIS:0 tunnel)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-1#
```

Example 2: TTM preference with VPRN service

The following example is a variant on the previous example. The difference in this example is that, in addition to SR, LDP and RSVP-TE are also enabled between PE-1 and PE-7. A single RSVP LSP is configured originating at PE-1 and terminating at PE-7.

The objective of this example is to show the difference in protocol preference within TTM and how to influence the default behavior. This can be useful in case of migration scenarios from a non-SR environment toward a hybrid environment having LDP/RSVP and SR enabled.

In the following example, LFA/RLFA is no longer configured on the PE-1 node.

Translated into configuration commands, this becomes:

```
*A:PE-1# configure router isis no loopfree-alternate
*A:PE-1# configure router ldp interface-parameters interface "int-PE-1-PE-7"
*A:PE-1# configure router mpls interface "int-PE-1-PE-7"
*A:PE-1# configure router mpls no shutdown
*A:PE-1# configure router rsvp no shutdown
*A:PE-1# configure router mpls
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-7"
                to 192.0.2.7
                primary "dyn"
                exit
                no shutdown
            exit
            no shutdown

*A:PE-7# configure router ldp interface-parameters interface "int-PE-7-PE-1"
*A:PE-7# configure router mpls no shutdown
*A:PE-7# configure router rsvp no shutdown
```

By enabling LDP and RSVP between PE-1 and PE-7, the TTM on both nodes changed. With the VPRN service between PE-1 and PE-7 of example 1, only those two specific service endpoints are displayed:

```
*A:PE-1# show router tunnel-table 192.0.2.7
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.7/32      rsvp      MPLS  1         7        192.168.17.2   10
192.0.2.7/32      ldp       MPLS  65537     9        192.168.17.2   10
192.0.2.7/32      isis (0)  MPLS  524294    11       192.168.17.2   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#

*A:PE-7# show router tunnel-table 192.0.2.1
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65537     9        192.168.17.1   10
192.0.2.1/32      isis (0)  MPLS  524294    11       192.168.17.1   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-7#
```

On node PE-1, an RSVP LSP, an LDP LSP, and an SR shortest path tunnel (using IS-IS) are present. Because the VPRN service has auto-bind-tunnel resolution any enabled, the protocol type with the highest TTM preference (meaning the lowest absolute preference value in TTM) is taken; in this case, the RSVP LSP. This can be verified for the configured loopback address within the VPRN service context, as follows:

```
*A:PE-1# show router 100 route-table 192.0.1.7/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
Dest Prefix             : 192.0.1.7/32
  Protocol              : BGP_VPN
  Age                   : 00h00m48s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.7
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
```

```
    Resolving Next-Hop  : 192.0.2.7 (RSVP tunnel:1)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-1#
```

On node PE-7, only an LDP LSP and an SR shortest path tunnel (using IS-IS) are present. Because the VPRN service has auto-bind-tunnel resolution any enabled, the protocol type with highest TTM preference (meaning the lowest absolute preference value in TTM) is taken; in this case, the LDP LSP. This can be verified for the configured loopback address within the VPRN service context, as follows:

```
*A:PE-7# show router 100 route-table 192.0.1.1/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
Dest Prefix             : 192.0.1.1/32
  Protocol              : BGP_VPN
  Age                   : 00h01m03s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.1
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.1 (LDP tunnel)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-7#
```

Some configuration changes are possible to change this default behavior:

- It is possible to change the auto-bind-tunnel resolution any command into auto-bind-tunnel resolution filter. Because this is a service-specific parameter, the operator has the choice to only configure this on one specific service endpoint. From a migration point of view, a smooth and easy SR migration is possible, having no affect on other deployed services on this node.
- It is possible to change the SR tunnel-table protocol preference on a node. From a migration point of view, this affects all services initiating on this node.

Using the current example, PE-1 implements the auto-bind-tunnel change (option 1), while PE-7 implements the TTM preference change (option 2).

First, a resolution-filter CLI context within VPRN service 100 on node PE-1 must be created. The example uses a resolution-filter context, which uses a filter to only allow SR shortest path tunnels (IS-IS based). Translated into configuration commands, this becomes:

```
*A:PE-1# configure service vprn 100 auto-bind-tunnel resolution-filter sr-isis
```

Then, change the auto-bind-tunnel resolution any command into resolution filter on PE-1. Translated into a configuration command, this becomes:

```
*A:PE-1# configure service vprn 100 auto-bind-tunnel resolution filter
```

As a result, the RSVP LSP is no longer used. Instead, the SR shortest path tunnel is used for the traffic from PE-1 to PE-7:

```
*A:PE-1# show router 100 route-table 192.0.1.7/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
Dest Prefix             : 192.0.1.7/32
  Protocol              : BGP_VPN
  Age                   : 00h00m12s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.7
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.7 (SR-ISIS:0 tunnel)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-1#
```

The VPRN service on node PE-7 is still using the LDP LSP as transport mechanism to reach node PE-1 at this point. Because the previous CLI change is only done within the VPRN service context 100 on PE-1, only the direction from PE-1 to PE-7 is affected.

Another way to influence the default TTM preference is shown as follows on the PE-7 node. Using the default behavior, the LDP LSP is used, because of the preference value of 9. If the SR tunnel table preference value is lowered to a value smaller than LDP, for instance 4, the SR shortest path tunnels originating on this node will always have preference compared to LDP LSP. Translated into a configuration command, this becomes:

```
*A:PE-7# configure router isis segment-routing tunnel-table-pref 4
```

```
*A:PE-7# show router tunnel-table 192.0.2.1
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.1/32      isis (0)  MPLS  524294    4        192.168.17.1   10
192.0.2.1/32      ldp       MPLS  65537     9        192.168.17.1   10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-7#
```

As a result, the LDP LSP is no longer used and the SR shortest path tunnel is the preferred transport tunnel:

```
*A:PE-7# show router 100 route-table 192.0.1.1/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
Dest Prefix             : 192.0.1.1/32
  Protocol              : BGP_VPN
  Age                   : 00h00m25s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.1
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.1 (SR-ISIS:0 tunnel)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-7#
```

At this point, within the VPRN service, the SR shortest path tunnels are used bidirectionally between PE-1 and PE-7.

If, for example, an operator configures explicit SDP binding within the same VPRN service on both endpoints, the explicit SDPs will always have preference. In this example, manual SDPs are configured on nodes PE-1 and PE-7, both using LDP. Translated into configuration commands, this becomes:

```
*A:PE-1# configure service sdp 17 mpls create
           far-end 192.0.2.7
           ldp
           no shutdown

*A:PE-1# configure service vprn 100 spoke-sdp 17 create

*A:PE-7# configure service sdp 71 mpls create
```

```
            far-end 192.0.2.1
            ldp
            no shutdown

*A:PE-7# configure service vprn 100 spoke-sdp 71 create
```

As a result, SR shortest path tunnels are no longer used, but rather LDP-based SDPs are used instead:

```
*A:PE-1# show router 100 route-table 192.0.1.7/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
Dest Prefix             : 192.0.1.7/32
  Protocol              : BGP_VPN
  Age                   : 00h00m34s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.7
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.7 (SDP tunnel:17)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-1#

*A:PE-7# show router 100 route-table 192.0.1.1/32 extensive
===============================================================================
Route Table (Service: 100)
===============================================================================
Dest Prefix             : 192.0.1.1/32
  Protocol              : BGP_VPN
  Age                   : 00h00m35s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.1
    Label               : 262141
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.1 (SDP tunnel:71)
      Metric            : 10
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-7#
```

Conclusion

Segment Routing is a technique using extensions of the existing link state protocols, and using existing MPLS or IPv6 infrastructure as the data plane. It is a source routing technique similar to RSVP-TE, but without the need to run an extra signaling protocol. SR also avoids other scaling restrictions of associated RSVP-TE, such as midpoint state. SR is simple to control and operate because the intelligence and state are part of the packet, not held by the network. Other benefits are that SR can be introduced in an incremental way using different migration scenarios to assure a smooth transition.

Shared Risk Link Groups for RSVP-Based LSP

This chapter provides information about Shared Risk Link Groups for RSVP-Based LSPs.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This feature is applicable to SR OS. See the release notes for a full and up to date list of supported hardware.

Originally, this chapter was written for release 7.0. The CLI in this edition corresponds to release 14.0.R1. No prerequisites are needed.

Overview

Introduction

Shared risk link group (SRLG) is a feature which allows the user to establish a backup secondary label switched path (LSP) path or a fast-reroute (FRR) LSP path which is disjoint from the path of the primary LSP. Links which are members of the same SRLG represent resources which share the same risk. For example, fiber links sharing the same conduit or multiple wavelengths sharing the same fiber.

A typical application of the SRLG feature is to provide an automatic placement of secondary backup LSPs or FRR bypass/detour LSPs that minimizes the probability of fate sharing with the path of the primary LSP.

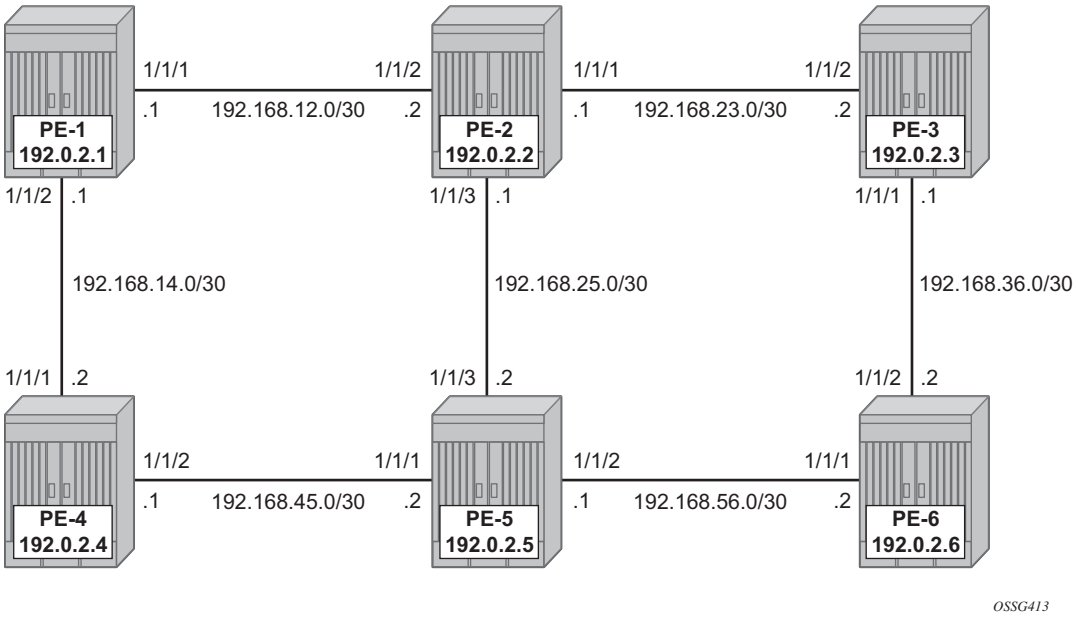
SRLG groups are used to determine which links belong to the same SRLG. The mechanism is similar to multiprotocol label switching (MPLS) admin groups. To advertise SRLG, the information is part of the IGP TE parameters in an opaque link state advertisement (LSA). The SRLG is advertised in a new Shared Risk Link Group TLV (type 138) in IS-IS (RFC 4205, *Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*). It is advertised in a new SRLG sub-TLV (type 16) of the existing Link TLV in OSPF (RFC 4203, *OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*.

For FRR, a choice can be made on what to do when no FRR tunnel can be found with the SRLG constraints. No FRR tunnel might be signaled or a FRR tunnel might be signaled not taking the SRLG constraints into account.

SRLG

Figure 200 displays the example topology for this chapter.

Figure 200 **Example Topology**



A single IGP area (IS-IS in this case) with traffic engineering (TE) enabled is required for the SRLG feature to work properly.

When OSPF is used as the IGP, the functionality is similar.

Configuration

Configuring the IP/MPLS network

IS-IS, MPLS, and RSVP are configured on all interfaces. TE is enabled in IS-IS. Optionally, admin groups “green” and “red” are configured on all nodes. The “green” links are the following: the link between PE-1 and PE-2, the link between PE-2 and PE-3, and the link between PE-3 and PE-6. The “red” links are: the link between PE-1 and PE-4, the link between PE-4 and PE-5, and the link between PE-5 and PE-6. The remaining link is the link between PE-2 and PE-5, which does not belong to an admin group. For more information about admin groups, see chapter RSVP Point-to-Point LSPs.

In addition, ECMP is set to 2, instead of the default value 1 in order to highlight the application of SRLG in the final example: SRLG database.

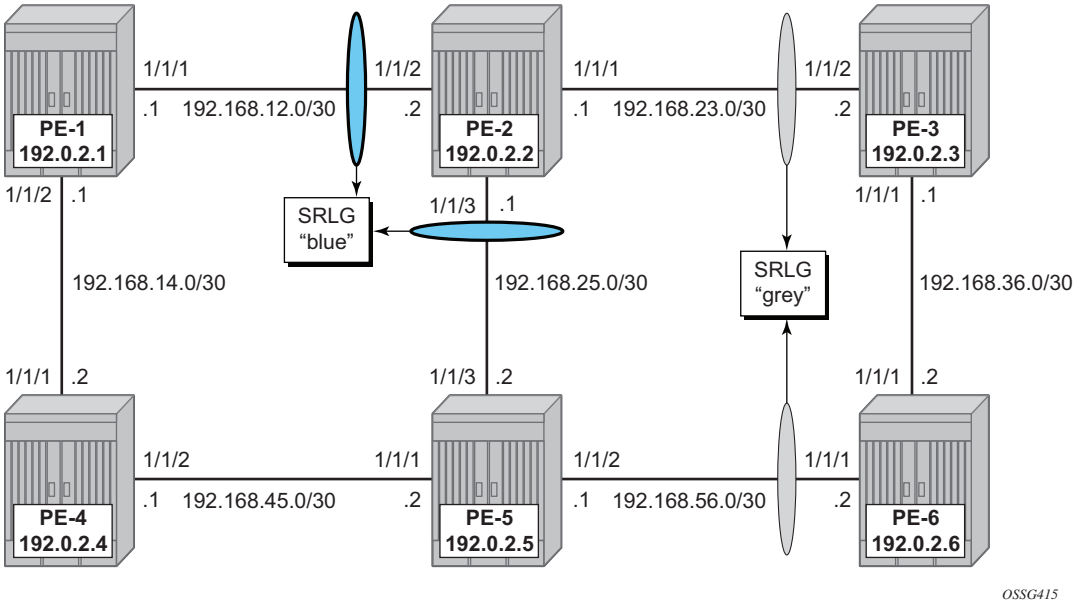
```
*A:PE-1# configure router ecmp 2
```

Define SRLG groups

Define the SRLG groups, and link them to the related MPLS interfaces.

Two SRLG groups are defined, named blue and gray. On following drawing the related IP/MPLS interfaces are indicated.

Figure 201 **SRLG Topology**



The configuration of the blue SRLG group is only mandatory on PE-1, PE2 and PE-5, while the gray SRLG group is only mandatory on PE-2, PE-3, PE-5 and PE6. However, it is good practice to configure both SRLG groups on all nodes, for example as follows on PE-1.

```
*A:PE-1# configure router if-attribute srlg-group blue value 1
*A:PE-1# configure router if-attribute srlg-group gray value 2
```

The IP/MPLS interfaces need to be linked to the related SRLG group, which is a uni-directional indicator, applying only to the egress direction; therefore, it needs to be configured on both sides of the IP/MPLS interface. For example on PE-1, the interface to PE-2 is part of **srlg-group "blue"**. An interface can be part of multiple SRLG groups similar to the admin-group functionality.

```
*A:PE-1>config>router>mpls# info
----------------------------------------------
            entropy-label rsvp-te enable
            interface "system"
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                admin-group "green"
                no shutdown
            exit
            interface "int-PE-1-PE-4"
                admin-group "red"
                no shutdown
            exit
            no shutdown
```

```
*A:PE-1>config>router>mpls# interface "int-PE-1-PE-2"
*A:PE-1>config>router>mpls>if# srlg-group "blue"
```

The same must be done on PE-2, PE-3, PE-5 and PE-6. Afterwards, verify the MPLS configuration for example on PE-2, where the SRLG groups are linked to the interfaces. Admin-groups are configured in parallel to indicate that both can be configured and will work independently.

```
*A:PE-2>config>router>mpls# info
----------------------------------------------
        entropy-label rsvp-te enable
        interface "system"
            no shutdown
        exit
        interface "int-PE-2-PE-1"
            admin-group "green"
            srlg-group "blue"
            no shutdown
        exit
        interface "int-PE-2-PE-3"
            admin-group "green"
            srlg-group "gray"
            no shutdown
        exit
        interface "int-PE-2-PE-5"
            srlg-group "blue"
            no shutdown
        exit
        no shutdown
```

The SRLG configuration can be verified using the following show commands.

To show all SRLG groups on the node:

```
*A:PE-2# show router if-attribute srlg-group
===============================================================================
Interface Srlg Groups
===============================================================================
Group Name                      Group Value      Penalty Weight
-------------------------------------------------------------------------------
blue                            1                0
gray                            2                0
-------------------------------------------------------------------------------
No. of Groups: 2
===============================================================================
*A:PE-2#
```

In the list of MPLS interfaces, admin groups and SRLG groups are indicated.

```
A:PE-2# show router mpls interface
===============================================================================
MPLS Interfaces
===============================================================================
Interface                       Port-id          Adm    Opr    TE-metric
-------------------------------------------------------------------------------
```

```
system                                 system           Up     Up     None
  Admin Groups                         None
  SRLG Groups                          None
int-PE-2-PE-1                          1/1/2            Up     Up     None
  Admin Groups                         green
  SRLG Groups                          blue
int-PE-2-PE-3                          1/1/1            Up     Up     None
  Admin Groups                         green
  SRLG Groups                          gray
int-PE-2-PE-5                          1/1/3            Up     Up     None
  Admin Groups                         None
  SRLG Groups                          blue
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
A:PE-2#
```

To verify the SRLG groups in the IGP TE database, the following command can be used. The output can be extensive, but searching on the SRLG group name will lead to the correct interfaces.

The following output shows the link-state advertisements of PE-2 on PE-1 in this case. The SRLG information is linked to the IP interfaces in a dedicated TE-TLV.

```
*A:PE-1# show router isis database PE-2.00-00 detail
===============================================================================
Rtr Base ISIS Instance 0 Database (detail)
===============================================================================
Displaying Level 1 database
-------------------------------------------------------------------------------
LSP ID    : PE-2.00-00                                 Level     : L1
Sequence  : 0x6                  Checksum  : 0x3cc1   Lifetime  : 1168
Version   : 1                    Pkt Type  : 18       Pkt Ver   : 1
Attributes: L1                   Max Area  : 3        Alloc Len : 508
SYS ID    : 1920.0000.2002       SysID Len : 6        Used Len  : 508

TLVs :

---snip---

TE SRLGs     :
    SRLGs : PE-1.00
    Lcl Addr  : 192.168.12.2
    Rem Addr  : 192.168.12.1
    Num SRLGs       : 1
        1

---snip---

  TE SRLGs     :
    SRLGs : PE-3.00
    Lcl Addr  : 192.168.23.1
    Rem Addr  : 192.168.23.2
    Num SRLGs       : 1
        2

---snip---
```

```
  TE SRLGs     :
    SRLGs : PE-5.00
    Lcl Addr  : 192.168.25.1
    Rem Addr  : 192.168.25.2
    Num SRLGs       : 1
        1

---snip---
```

On-Line Verification

An on-line verification can be done by a **tools perform** command. This will trigger a constrained shortest path first (CSPF) call to the interior gateway protocol (IGP) TE database, and the result will be an explicit route object (ERO) object which can potentially be used to set up a CSPF-based LSP.

The following shows the command syntax.

```
*A:PE-1# tools perform router mpls cspf
  - cspf to <ip-addr> [from <ip-addr>] [bandwidth <bandwidth>] [include-bitmap
<bitmap>] [exclude-bitmap <bitmap>] [hop-limit <limit>] [exclude-address <excl-addr>
[<excl-addr>...(upto 8 max)]] [use-te-metric] [strict-srlg] [srlg-group <grp-
id>...(upto 8 max)] [exclude-node <excl-node-id> [<excl-node-id>..(upto 8 max)]]
[skip-interface <interface-name>] [ds-class-type <class-type>] [cspf-reqtype <req-
type>] [least-fill-min-thd <thd>] [setup-priority <val>] [hold-priority <val>]

 <ip-addr>            : a.b.c.d
 <bandwidth>          : [1..100000] in Mbps
 <bitmap>             : [0..4294967295] - accepted in decimal, hex(0x) or binary(0b)
 <limit>              : [2..255]
 <excl-addr>          : a.b.c.d (outbound interface)
 <use-te-metric>      : keyword
 <strict-srlg>        : keyword
 <grp-id>             : [0..4294967295]
 <excl-node-id>       : [a.b.c.d] (outbound interface)
 <interface-name>     : [max 32 chars]
 <class-type>         : [0..7]
 <req-type>           : all|random|least-fill : keywords
 <thd>                : [1..100]
 <priority>           : [0..7]
```

Where the relevant parameters are:

- **to** — Defines the far-end address of the LSP. This is the system-address of the destination LER
- **srlg-group** — Specifies which SRLG groups should be avoided while building the path to the destination (ERO object)

- **strict-srlg** — Indicates whether the SRLG group is a strict requirement or not. When this parameter is given, only paths without traversing the SRLG will be displayed.

Example:

On PE-1, a CSPF calculation is made with PE-3 as destination, without any SRLG restrictions, as follows:

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3
Req CSPF for all ECMP paths
    from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7, Hold
Priority 0 TE Class: 7

CSPF Path
To        : 192.0.2.3
Path 1    : (cost 20)
    Src:   192.0.2.1   (= Rtr)
   Egr:  192.168.12.1      -> Ingr:  192.168.12.2       Rtr:  192.0.2.2     (met 10)
   Egr:  192.168.23.1      -> Ingr:  192.168.23.2       Rtr:  192.0.2.3     (met 10)
    Dst:   192.0.2.3   (= Rtr)

*A:PE-1#
```

Given a restriction on **srlg-group “blue”** (grp-id =1), the result is as follows:

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 srlg-group 1
Req CSPF for all ECMP paths
    from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7, Hold
Priority 0 TE Class: 7

CSPF Path
To        : 192.0.2.3
Path 1    : (cost 40)
    Src:   192.0.2.1   (= Rtr)
   Egr:  192.168.14.1       -> Ingr:  192.168.14.2      Rtr:  192.0.2.4     (met 10)
   Egr:  192.168.45.1       -> Ingr:  192.168.45.2      Rtr:  192.0.2.5     (met 10)
   Egr:  192.168.56.1       -> Ingr:  192.168.56.2      Rtr:  192.0.2.6     (met 10)
      1 SRLGs:  2
   Egr:  192.168.36.2       -> Ingr:  192.168.36.1      Rtr:  192.0.2.3     (met 10)
    Dst:   192.0.2.3   (= Rtr)

*A:PE-1#
```

The path will be through PE-4, PE-5 and PE-6.

When a strict restriction is requested on **srlg-group “gray**”, no valid CSPF path toward the destination can be found.

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 srlg-group 2 strict-srlg
Req CSPF for all ECMP paths
    from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7, Hold
Priority 0 TE Class: 7

MINOR: CLI No CSPF path to "192.0.2.3" with specified constraints.
```

```
*A:PE-1#
```

Removing the **strict** restriction results in a successful return of CSPF, indicating that the CSPF path is not SRLG disjoint.

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 srlg-group 2
Req CSPF for all ECMP paths
    from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7, Hold
Priority 0 TE Class: 7

CSPF Path
To        : 192.0.2.3 (NOT SRLG DISJOINT)
Path 1    : (cost 20)
    Src:   192.0.2.1   (= Rtr)
   Egr:  192.168.12.1       -> Ingr:  192.168.12.2       Rtr:  192.0.2.2    (met 10)
      1 SRLGs:  1
   Egr:  192.168.23.1       -> Ingr:  192.168.23.2       Rtr:  192.0.2.3    (met 10)
      1 SRLGs:  2
    Dst:   192.0.2.3   (= Rtr)

*A:PE-1#
```

The best practice for debugging is to enable debug-tracing on the CSPF process, with following command:

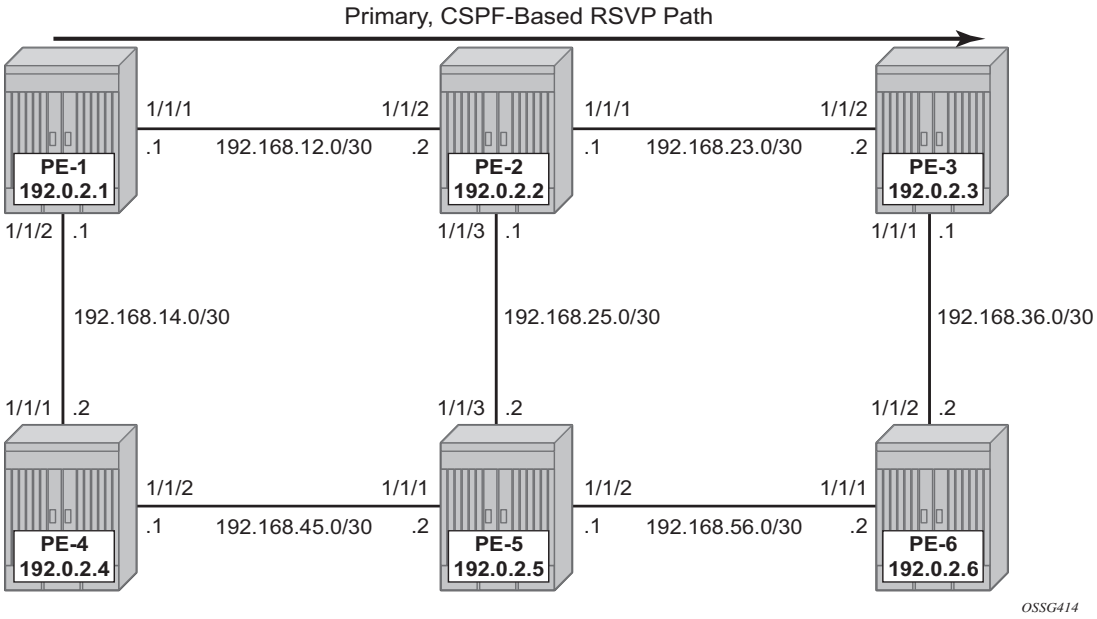
```
*A:PE-1# debug router isis cspf
```

SRLG for FRR

The fast-reroute mechanism used here is facility link-protection. The SRLG feature is independent of the FRR type and works for all combinations (facility versus one-to-one, link versus node protection).

Configure an LSP from PE-1 to PE-3, and enable CSPF.

Figure 202 **Path Primary RSVP_TE LSP**



The configuration of the LSP"LSP-PE-1-PE-3_FRR_facility-link" is based on an empty path, with FRR facility link protection enabled.

```
*A:PE-1# configure router
        mpls
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3_FRR_facility-link"
                to 192.0.2.3
                cspf
                fast-reroute facility
                    no node-protect
                exit
                primary "dyn"
                exit
                no shutdown
            exit
```

To verify the primary path, **oam lsp-trace** command can be used, checking the intermediate nodes.

```
*A:PE-1# oam lsp-trace "LSP-PE-1-PE-3_FRR_facility-link" detail
lsp-trace to LSP-PE-1-PE-3_FRR_facility-
link: 0 hops min, 0 hops max, 116 byte packets
1  192.0.2.2  rtt=0.670ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.23.2 ifaddr=192.168.23.2 iftype=ipv4Numbered MRU=1564
           label[1]=262143 protocol=4(RSVP-TE)
2  192.0.2.3  rtt=1.13ms rc=3(EgressRtr) rsc=1
*A:PE-1#
```

To check if the bypass tunnels are up and running, an indication (@)can be found in the detail output of **show router mpls ls <x> path detail** as seen in the following output.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3_FRR_facility-link" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-3_FRR_facility-link Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-3_FRR_facility-link Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-1-PE-3_FRR_facility-link
Path LSP ID       : 29184
From              : 192.0.2.1             To                  : 192.0.2.3
Admin State       : Up                    Oper State          : Up
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Up
Out Interface     : 1/1/1                 Out Label           : 262143
Path Up Time      : 0d 00:05:33           Path Down Time      : 0d 00:00:00
Retry Limit       : 0                     Retry Timer         : 30 sec
Retry Attempt     : 0                     Next Retry In       : 0 sec
BFD Template      : None                  BFD Ping Interval   : 60
BFD Enable        : False

Adspec            : Disabled              Oper Adspec         : Disabled
CSPF              : Enabled               Oper CSPF           : Enabled
Least Fill        : Disabled              Oper LeastFill      : Disabled
FRR               : Enabled               Oper FRR            : Enabled
FRR NodeProtect   : Disabled              Oper FRR NP         : Disabled
FR Hop Limit      : 16                    Oper FRHopLimit     : 16
FR Prop Admin Gr* : Disabled              Oper FRPropAdmGrp   : Disabled
Propogate Adm Grp : Disabled              Oper Prop Adm Grp   : Disabled
Inter-area        : False

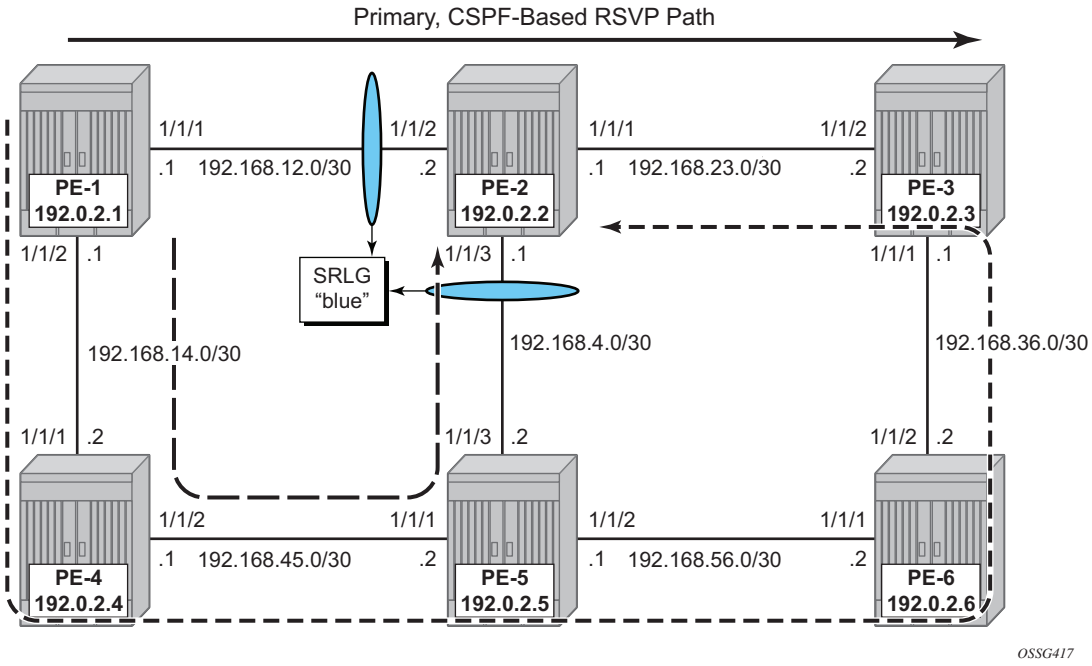
Neg MTU           : 1560                  Oper MTU            : 1560
Bandwidth         : No Reservation        Oper Bandwidth      : 0 Mbps
Hop Limit         : 255                   Oper HopLimit       : 255
Record Route      : Record                Oper Record Route   : Record
Record Label      : Record                Oper Record Label   : Record
Setup Priority    : 7                     Oper Setup Priority : 7
Hold Priority     : 0                     Oper Hold Priority  : 0
Class Type        : 0                     Oper CT             : 0
Backup CT         : None
MainCT Retry      : n/a
    Rem           :
MainCT Retry      : 0
    Limit         :
Include Groups    :                       Oper Include Groups :
None                                            None
Exclude Groups    :                       Oper Exclude Groups :
None                                            None
```

```
Adaptive          : Enabled              Oper Metric          : 20
Preference        : n/a
Path Trans        : 1                    CSPF Queries         : 1
Failure Code      : noError
Failure Node      : n/a
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.168.12.1 (192.0.2.1) @                  Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2) @                  Record Label        : 262143
 -> 192.168.23.2 (192.0.2.3)                    Record Label        : 262143
Computed Hops     :
    192.168.12.1(S)
 -> 192.168.12.2(S)
 -> 192.168.23.2(S)
Resignal Eligible: False
Last Resignal     : n/a                  CSPF Metric          : 20
===============================================================================
* indicates that the corresponding row element may have been truncated.

*A:PE-1#
```

Two links are protected: one bypass tunnel originates in PE-1 protecting the link between PE-1 and PE-2. Another bypass tunnel originates in PE-2 protecting the link between PE-2 and PE-3. The focus is on the bypass tunnel originating in PE-1. When SRLG is enabled, the bypass tunnel originating in PE-1 will have different hops. The expected paths followed by the bypass tunnels originating in PE-1 with and without SRLG are shown in Figure 203.

Figure 203 **FRR Bypass Tunnels Originating in PE-1 With and Without SRLG**



To verify the bypass data path on the point of local repair (PLR) PE-1, the following command can be used.

```
*A:PE-1# show router mpls bypass-tunnel detail

===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.12.2-61441
-------------------------------------------------------------------------------
To              : 192.168.25.1        State               : Up
Out I/F         : 1/1/2               Out Label           : 262143
Up Time         : 0d 00:11:31         Active Time         : n/a
Reserved BW     : 0 Kbps              Protected LSP Count : 1
Type            : Dynamic             Bypass Path Cost    : 30
Setup Priority  : 7                   Hold Priority       : 0
Class Type      : 0
Exclude Node    : None                Inter-Area          : False
Computed Hops   :
    192.168.14.1(S)                   Egress Admin Groups :
                                      red
 -> 192.168.14.2(S)                   Egress Admin Groups :
                                      red
 -> 192.168.45.2(S)                   Egress Admin Groups : None
 -> 192.168.25.1(S)                   Egress Admin Groups : None
Actual Hops     :
    192.168.14.1 (192.0.2.1)          Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)          Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)          Record Label        : 262142
```

```
 -> 192.168.25.1 (192.0.2.2)         Record Label        : 262142
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

===============================================================================
*A:PE-1#
```

The SRLG restriction is not taken into account at this moment at PLR PE-1. The actual hops are PE-4, PE-5 and PE-2 visualized by the path with the long dashes in Figure 203.

To take the SRLG restrictions into account, additional configuration is needed for MPLS.

```
*A:PE-1>config>router>mpls# srlg-
srlg-database  srlg-frr

*A:PE-1>config>router>mpls# srlg-frr
  - no srlg-frr
  - srlg-frr [strict]

 <strict>             : keyword


*A:PE-1# configure router mpls srlg-frr strict
```

The option **strict** should only be used if the logical topology allows this. In other words, one must be sure that an alternative path is possible which avoids SRLG-groups.

Note: Enabling or disabling SRLG for FRR is a system-wide configuration and requires the MPLS routing instance to be manually set to shutdown and then to no shutdown to activate the change. This may cause service outage. Nokia recommends that the operator incorporates the SRLG into the initial network design and implementation to minimize the traffic loss.

```
*A:PE-1# configure router rsvp shutdown
*A:PE-1# configure router rsvp no shutdown
```

The bypass tunnel originating in PLR PE-1can be verified with a previously used command.

```
*A:PE-1# show router mpls bypass-tunnel detail

===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.12.2-61442
-------------------------------------------------------------------------------
To             : 192.168.23.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262143
```

```
Up Time         : 0d 00:00:42          Active Time         : n/a
Reserved BW     : 0 Kbps               Protected LSP Count : 1
Type            : Dynamic              Bypass Path Cost    : 50
Setup Priority : 7                     Hold Priority       : 0
Class Type      : 0
Exclude Node    : None                 Inter-Area          : False
Computed Hops   :
    192.168.14.1(S)                    Egress Admin Groups :
                                       red
 -> 192.168.14.2(S)                    Egress Admin Groups :
                                       red
 -> 192.168.45.2(S)                    Egress Admin Groups :
                                       red
 -> 192.168.56.2(S)                    Egress Admin Groups :
                                       green
 -> 192.168.36.1(S)                    Egress Admin Groups :
                                       green
 -> 192.168.23.1(S)                    Egress Admin Groups : None
Actual Hops     :
    192.168.14.1 (192.0.2.1)           Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)           Record Label        : 262143
 -> 192.168.45.2 (192.0.2.5)           Record Label        : 262142
 -> 192.168.56.2 (192.0.2.6)           Record Label        : 262142
 -> 192.168.36.1 (192.0.2.3)           Record Label        : 262140
 -> 192.168.23.1 (192.0.2.2)           Record Label        : 262141
Last Resignal   :
Attempted At    : n/a                  Resignal Reason     : n/a
Resignal Status: n/a                   Reason              : n/a

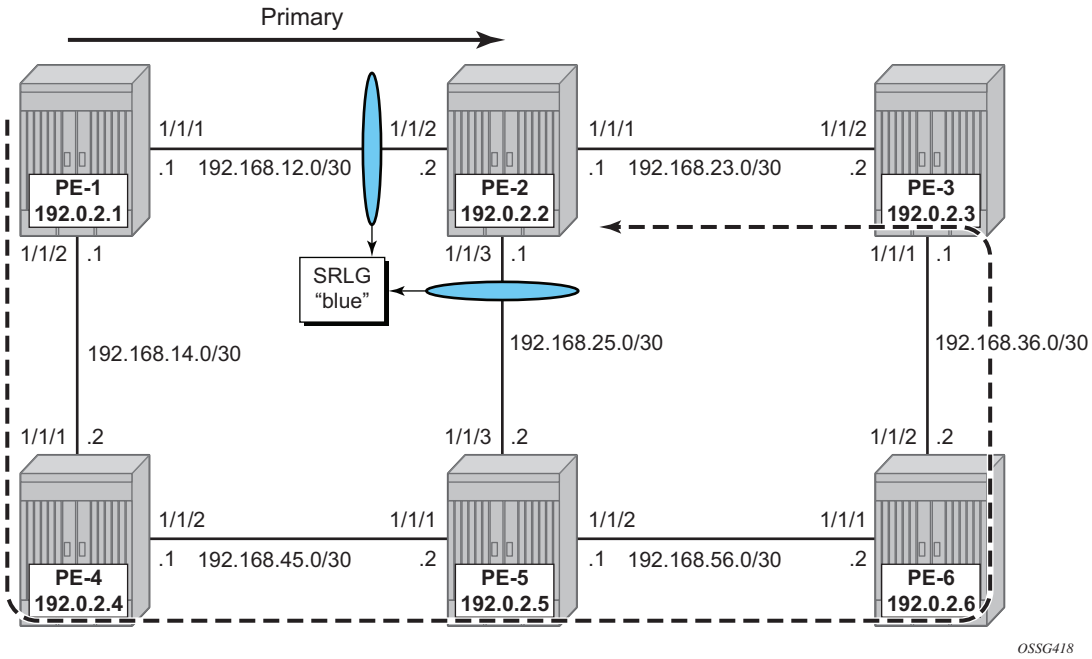
===============================================================================
*A:PE-1#
```

This path taking the SRLG constraints into account is represented by the line with the short dashes in Figure 203 .

SRLG for Standby Path

Where SRLG groups can be constraints for bypass tunnels, they can also be a constraint to set up a secondary path. Looking at the following picture, the secondary path is expected to follow the dotted-line instead of passing over the direct link between PE-5 and PE-2.

Figure 204 **SRLG for Secondary Path**



An LSP is configured with a primary and a secondary path, which have no hops defined. The configuration of the LSP will need a specific indication at the level of the secondary path to enable the restriction on the srlg-groups.

```
*A:PE-1# configure router
        mpls
            path "prim"
                no shutdown
            exit
            path "secon"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2-srlg"
                to 192.0.2.2
                cspf
                primary "prim"
                exit
                secondary "secon"
                    standby
                    srlg
                exit
                no shutdown
            exit
```

Where both paths are empty paths, the ERO object creation solely relies on CPSF without any specific hop.

To verify the data path, the detailed output of the **show router mpls** lsp <..> path command can be used, as well as the **lsp-trace** OAM command. This output shows both ERO objects of the primary and secondary path.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2-srlg" path detail
===============================================================================
MPLS LSP LSP-PE-1-PE-2-srlg Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-2-srlg Path prim
-------------------------------------------------------------------------------
---snip---

ExplicitHops:
    No Hops Specified
Actual Hops      :
    192.168.12.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262143
ComputedHops:
    192.168.12.1(S)
 -> 192.168.12.2(S)
ResigEligib*: False
LastResignal: n/a                               CSPF Metric : 10
-------------------------------------------------------------------------------
LSP LSP-PE-1-PE-2-srlg Path secon
-------------------------------------------------------------------------------
---snip---

ExplicitHops:
    No Hops Specified
Actual Hops      :
    192.168.14.1 (192.0.2.1)                    Record Label        : N/A
 -> 192.168.14.2 (192.0.2.4)                    Record Label        : 262142
 -> 192.168.45.2 (192.0.2.5)                    Record Label        : 262141
 -> 192.168.56.2 (192.0.2.6)                    Record Label        : 262141
 -> 192.168.36.1 (192.0.2.3)                    Record Label        : 262143
 -> 192.168.23.1 (192.0.2.2)                    Record Label        : 262140
ComputedHops:
    192.168.14.1(S)
 -> 192.168.14.2(S)
 -> 192.168.45.2(S)
 -> 192.168.56.2(S)
 -> 192.168.36.1(S)
 -> 192.168.23.1(S)
Srlg        : Enabled
SrlgDisjoint: True
ResigEligib*: False
LastResignal: n/a                               CSPF Metric : 50
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The **lsp-trace** command can be used for secondary path as well. The intermediate LSRs and the MPLS labels used can be clearly seen.

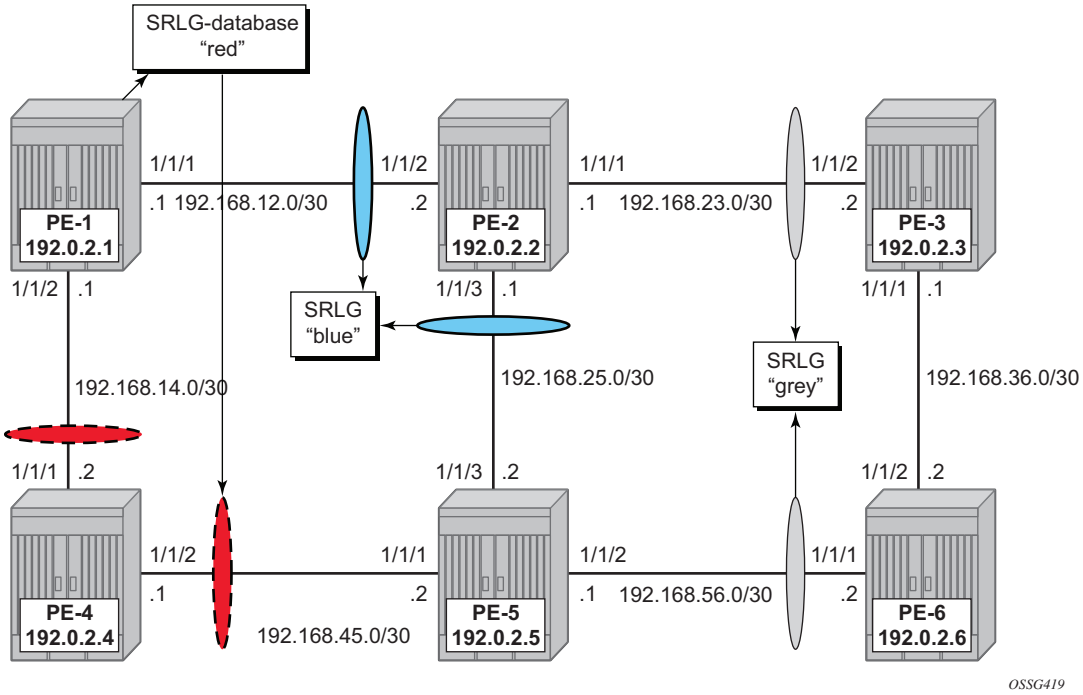
```
*A:PE-1# oam lsp-trace "LSP-PE-1-PE-2-srlg" path "secon" detail
lsp-trace to LSP-PE-1-PE-2-srlg: 0 hops min, 0 hops max, 116 byte packets
1  192.0.2.4  rtt=0.639ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.45.2 ifaddr=192.168.45.2 iftype=ipv4Numbered MRU=1564
           label[1]=262141 protocol=4(RSVP-TE)
2  192.0.2.5  rtt=1.20ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.56.2 ifaddr=192.168.56.2 iftype=ipv4Numbered MRU=1564
           label[1]=262141 protocol=4(RSVP-TE)
3  192.0.2.6  rtt=2.18ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.36.1 ifaddr=192.168.36.1 iftype=ipv4Numbered MRU=1564
           label[1]=262143 protocol=4(RSVP-TE)
4  192.0.2.3  rtt=1.92ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=192.168.23.1 ifaddr=192.168.23.1 iftype=ipv4Numbered MRU=1564
           label[1]=262140 protocol=4(RSVP-TE)
5  192.0.2.2  rtt=3.94ms rc=3(EgressRtr) rsc=1
*A:PE-1#
```

SRLG Database

In case not all IP/MPLS routers in the area support SRLG, a static SRLG database can be created on the systems which will be used as an additional constraint when performing the CSPF calculation to define the path.

An example can be seen in Figure 205 where an additional SRLG group “red” is defined on PE-1, with information related to the interface between PE-4 and PE-5.

Figure 205 **SRLG Database Example**




```
*A:PE-1# configure router if-attribute srlg-group "red" value 3

*A:PE-1# configure router
        mpls
            interface "int-PE-1-PE-4"
                srlg-group "red"
            exit
            srlg-database
                router-id 192.0.2.4
                    interface 192.168.45.1 srlg-group "red"
                    no shutdown
                exit
                router-id 192.0.2.5
                    interface 192.168.45.2 srlg-group "red"
                    no shutdown
                exit
            exit
```

This information is local to PE-1 and will only have effect on CSPF calculations on PE-1, not on the other nodes.

When a CSPF calculation is done for a path from PE-1 to PE-5, the result will be two equal-cost paths, because ECMP equals 2. When adding the **srlg-group “red”** as a restriction, only a single path will be found, passing PE-2.

Conclusion

Interpreting the SRLG information into the TE database makes it possible to protect an LSP even when multiple IP/MPLS interfaces fail as a result of an underlying transmission failure. Transmission failures can occur quite often since not all transmission links are one to one protected.

SRLG groups in MPLS provide a very dynamic and simple way to assure LSP FRR path protection on every PLR throughout the followed LSP path. The SRLG groups are also taken into account when defining the ERO for secondary paths, at least if the configured secondary path is empty.

For interoperability reasons, the SRLG-database is available, because systems can link interfaces to an SRLG with interconnecting systems that do not support the SRLG feature; therefore they cannot advertise the SRLG information through the IGP.

The creation and maintenance of an SRLG database requires operational effort and systems that do not support SRLG will never take any SRLG information into account during CSPF calculation for the creation of FRR bypass or detour tunnels.

Static Point-to-Point LSPs

This chapter provides information about static point-to-point label switched paths (LSPs).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was originally written for release 7.0.R5. The output in this edition corresponds to release 14.0.R1. There are no pre-requisites or conditions on the hardware for this configuration.

Overview

Due to the connectionless nature of the network layer protocol IP, packets travel through the network on a hop-by-hop basis with routing decisions made at each node. As a result, hyperaggregation of data on certain links may occur and it may impact the provider's ability to provide guaranteed service levels across the network end-to-end. To address these shortcomings, multiprotocol label switching (MPLS) was developed. The technology provides the capability to establish connection oriented paths, called label switched paths (LSPs), over a connectionless (IP) network. The LSP offers a mechanism to engineer network traffic independently from the underlying network routing protocol (mostly IP) to improve the network resiliency and recovery options and to permit delivery of new services that are not readily supported by conventional IP routing techniques (Layer 2 IP virtual private networks (VPNs)). These benefits are essential for today's communication network explaining the wide deployment base of the MPLS technology.

RFC 3031, *Multiprotocol Label Switching Architecture,* specifies the MPLS architecture while this document describes the configuration and troubleshooting of static point-to-point LSPs on SR OS. Point-to-point LSPs can also be dynamically established using a label signaling protocol, such as label distribution protocol (LDP) or resource reservation protocol (RSVP). See chapters LDP Point-to-Point LSPs and RSVP Point-to-Point LSPs.

Packet Forwarding

As a packet of a connectionless network layer protocol travels from one router to the next, each router in the network makes an independent forwarding decision by performing the following basic tasks: first analyzing the packet's header, then referencing the local routing table to find the longest match based on the destination address in the IP header, and finally sending out the packet on the selected interface. In other terms, the first function partitions the entire set of possible packets into a set of forwarding equivalence classes (FECs). All packets associated to a particular FEC will be forwarded along the same logical path to the same destination. The second function maps each FEC to a next hop destination router. Each router along the path performs these actions.

On the other hand, in MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network. In turn, the FEC is mapped to an LSP, which is established prior to any data flowing. An MPLS label, representing the FEC to which the packet is assigned, is attached to the packet (push operation) and once labeled, the packet is forwarded to the next hop router along that LSP path. At subsequent hops, there is no further analysis of the network layer header of the packet. Instead, the label is used as an index into a table which specifies the next hop and a new label. The old label is replaced with the new label (swap operation), and the packet is forwarded to its next hop. At the MPLS network egress, the label is removed from the packet (pop operation). If this router is the final destination (based on the remaining packet), the packet is handed to the receiving application (such as a virtual private LAN service (VPLS) domain). If this router is not the final destination of the packet, the packet will be sent into a new MPLS tunnel or forwarded by conventional IP forwarding towards the layer 3 destination.

Terminology

Figure 206 **Generic MPLS Network, MPLS Label Operations**

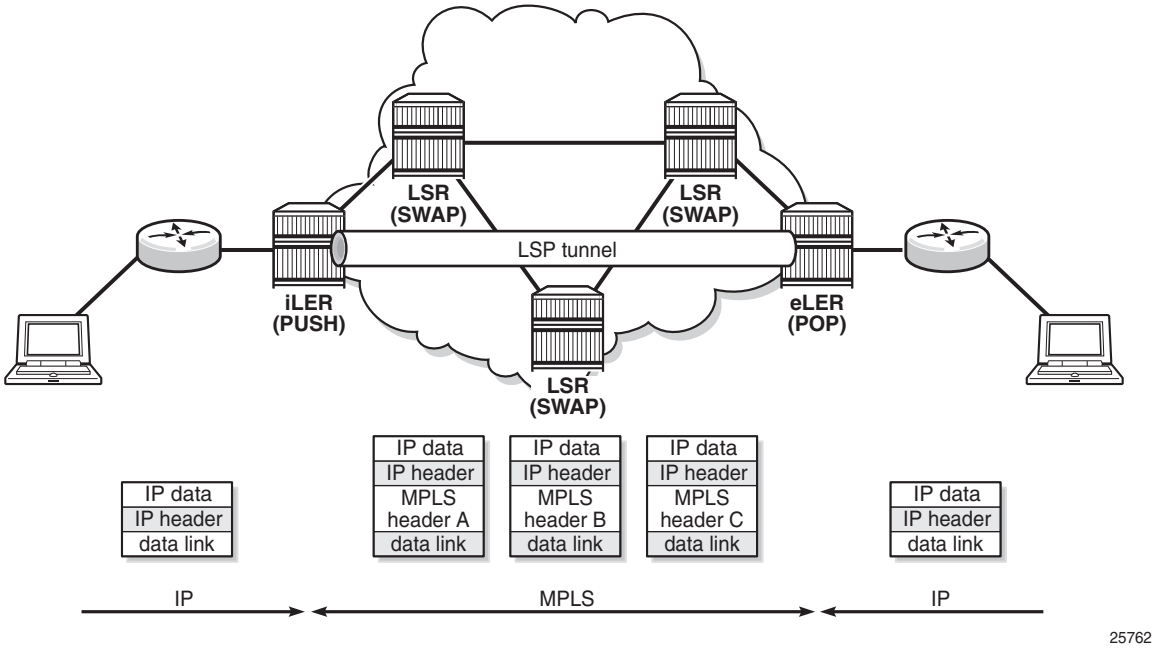


Figure 206depicts a general network topology clarifying the MPLS-related terms. A label edge router (LER) is a device at the edge of an MPLS network, with at least one interface outside the MPLS domain. A router is usually defined as an LER based on its position relative to a particular LSP. The MPLS router at the head-end of an LSP is called the ingress label edge router (iLER). The MPLS router at the tail-end of an LSP is called the egress label edge router (eLER). The iLER receives unlabeled packets from outside the MPLS domain, then applies MPLS labels to the packets, and forwards the labeled packets into the MPLS domain. The eLER receives labeled packets from the MPLS domain, then removes the labels, and forwards unlabeled packets outside the MPLS domain. The last LSR before the eLER can be configured with an implicit-null label (numeric value 3). This LSR will pop the outer label and send MPLS packets without an outer label to the eLER. This is known as penultimate hop popping (PHP). A label switching router (LSR) is a device internal to an MPLS network, with all interfaces inside the MPLS domain. These devices switch labeled packets inside the MPLS domain. In the core of the network, LSRs ignore the network layer (IP) header of the packet and simply forward the packet using the MPLS label swapping mechanism.

A single LSP is unidirectional. In common practice, because the bidirectional nature of most traffic flows is implied, the term LSP often is used to define the pair of LSPs that enable the bidirectional flow. For ease of terminology and discussion however, the LSP in this chapter is referred to as a single entity.

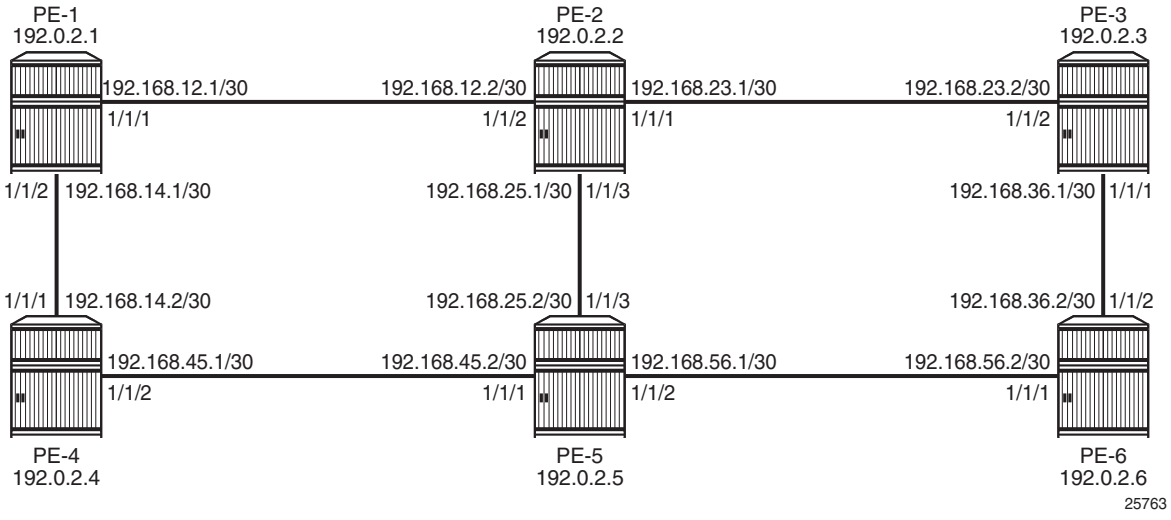
LSP Establishment

Prior to packet forwarding, the LSP must be established. In order to do so, labels need to be distributed for the path. For static LSPs, the label distribution is done manually by the network administrator. Although a high control level of the labels in use is achieved, the LSP cannot enjoy the resilience and recovery functionality the dynamic label signaling protocols can offer.

Example Topology

The example topology is displayed in Figure 207. The setup consists of six 7750 SR nodes located in a single autonomous system.

Figure 207 **MPLS Example Topology**



Configuration

For static LPSs, there is no need for an IGP.

For LSPs that are set up manually, the first step is to enable MPLS on:

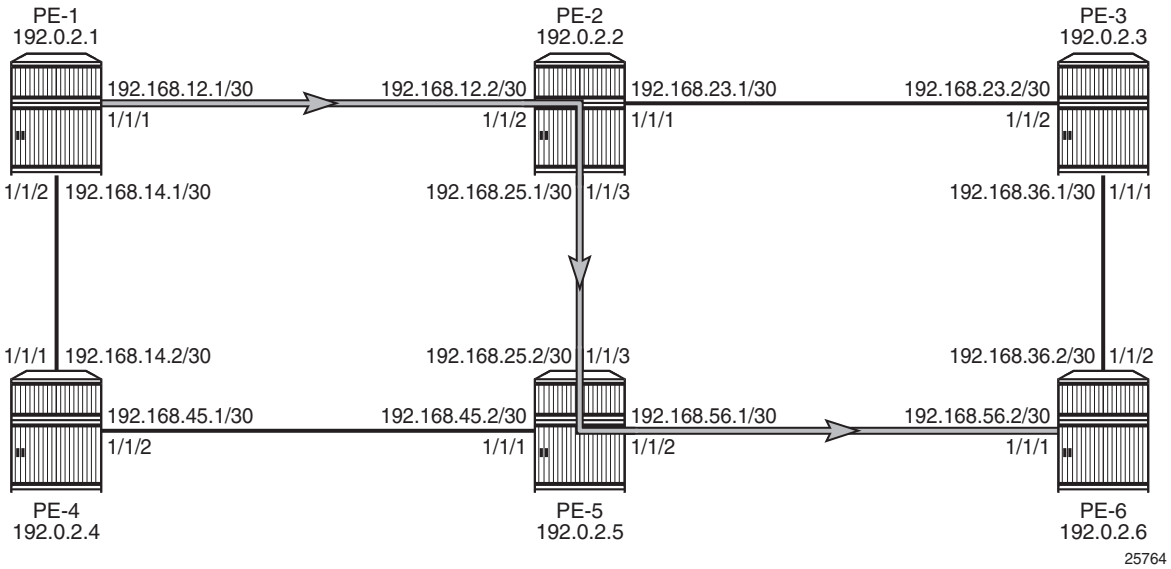
- All network interfaces that will be used to carry LSPs
- System IP address

For manually configured LSPs, any interface used by the static LSP must be added into the MPLS protocol instance, even though RSVP is not actually used to signal labels. For router PE-1, this results in the following configuration:

```
*A:PE-1# configure router
        mpls
            interface "system"
            exit
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-4"
            exit
            no shutdown
```

As an example, a static LSP will be created starting from PE-1, running over PE-2 and PE-5, then terminating on PE-6 as shown in Figure 208.

Figure 208 **Static LSP Running over PE-1, PE-2, PE-5, PE-6**



Verify the acceptable label range for use with static configurations for each node; as follows:

```
*A:PE-1# show router mpls-labels label-range
===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505856      505856
    Seg-Route   0           0           -           0           505856
===============================================================================
*A:PE-1#
```

The label range for static LSPs extends from the value 32 to 18431. To ensure the labels have not yet been allocated to another configuration, use the command:

```
*A:PE-1# show router mpls-labels label 32 18431 in-use
===================================================================
MPLS Labels from 32 to 18431 (In-use)
===================================================================
Label               Label Type          Label Owner
-------------------------------------------------------------------
-------------------------------------------------------------------
In-use labels (Owner: All) in specified range   : 0
In-use labels in entire range                   : 0
===================================================================
*A:PE-1#
```

This command shows the number of incoming labels in use. At the iLER, the number of labels in use will remain 0 even after the static LSP has been configured where the iLER has a push operation for a label. The reason is that the labels shown are relevant to the labels that the router is generating, as for label swap or pop operations. There is no information shown about labels that other routers are advertising. For the push operation, any label can be used, even if it is not within the label range of the router pushing the label. For the originating router PE-1, the label 100 will be used for the push operation on the interface towards PE-2.

Static LSPs are configured within the MPLS configuration context, but do not rely on dynamic label signaling.

The configuration of the MPLS static LSP head-end PE-1 contains:

- The system IP address of the destination router PE-6 (to).
- A push operation of the label 100.
- The interface address facing the current node of the next hop along the static path, which is PE-2 (nexthop).

```
*A:PE-1# configure router
        mpls
            static-lsp "LSP-PE-1-PE-6-static"
```

```
                to 192.0.2.6
                push 100 nexthop 192.168.12.2
                no shutdown
            exit
```

The transit LSRs (PE-2 and PE-5) perform swap operations and forward the packet to the manually defined next-hop. On the LSR under the context of the interface on which the incoming LSP arrives, the correct label is selected (label-map) and in this context a swap operation with a new label and the new next hop (nexthop) is entered.

```
*A:PE-2# configure router
        mpls
            interface "int-PE-2-PE-1"
                label-map 100
                    swap 150 nexthop 192.168.25.2
                    no shutdown
                exit
                no shutdown
            exit
*A:PE-5# configure router
        mpls
            interface "int-PE-5-PE-2"
                label-map 150
                    swap 200 nexthop 192.168.56.2
                    no shutdown
                exit
```

The terminating router PE-6 performs a pop operation and forwards the now unlabeled packets external to the MPLS domain.

```
*A:PE-6# configure router
        mpls
            interface "int-PE-6-PE-5"
                label-map 200
                    pop
                    no shutdown
                exit
```

To verify the operational status of the static LSP configuration, the **show router mpls static-lsp** command is used on the iLER. A static LSP is considered to be operationally up when its next-hop is reachable. Since there is no check whether the end-to-end LSP path is up (the LSP connectivity to the eLER is never verified), it can be that the static LSP path is broken while the iLER displays an operational enabled LSP.

```
*A:PE-1# show router mpls static-lsp
===============================================================================
MPLS Static LSPs (Originating)
===============================================================================
LSP Name     To              Next Hop        Out Label Up/Down Time   Adm  Opr
 ID           Metric          Oper Metric     Out Port
-------------------------------------------------------------------------------
LSP-PE-1-PE- 192.0.2.6       192.168.12.2    100       0d 00:02:19    Up   Up
-6-static
```

```
 1           N/A             N/A             1/1/1
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-1#
```

On the LSR, the **transit** keyword is added to the command. On PE-2:

```
*A:PE-2# show router mpls static-lsp transit
===============================================================================
MPLS Static LSPs (Transit)
===============================================================================
In Label    In Port     Out Label   Out Port    Next Hop            Adm   Opr
-------------------------------------------------------------------------------
100         1/1/2       150         1/1/3       192.168.25.2        Up    Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-2#
```

On LSR PE-5:

```
*A:PE-5# show router mpls static-lsp transit
===============================================================================
MPLS Static LSPs (Transit)
===============================================================================
In Label    In Port     Out Label   Out Port    Next Hop            Adm   Opr
-------------------------------------------------------------------------------
150         1/1/3       200         1/1/2       192.168.56.2        Up    Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-5#
```

On the terminating router (eLER), the keyword **terminate** is added.

```
*A:PE-6# show router mpls static-lsp terminate
===============================================================================
MPLS Static LSPs (Terminate)
===============================================================================
In Label    In Port     Out Label   Out Port    Next Hop            Adm   Opr
-------------------------------------------------------------------------------
200         1/1/1       n/a         n/a         n/a                 Up    Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-6#
```

To track the label action associated with the static LSP configuration, the **show router mpls interface label-map** command can be used on all LSRs and eLERs (not on the iLER).

```
*A:PE-2# show router mpls interface label-map
===============================================================================
MPLS Interfaces (Label-Map)
```

```
===============================================================================
In Label  In I/F     Out Label Out I/F    Next Hop          Type      Adm  Opr
-------------------------------------------------------------------------------
100       1/1/2      150       1/1/3      192.168.25.2      Static    Up   Up
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
*A:PE-2#
*A:PE-6# show router mpls interface label-map
===============================================================================
MPLS Interfaces (Label-Map)
===============================================================================
In Label  In I/F     Out Label Out I/F    Next Hop          Type      Adm  Opr
-------------------------------------------------------------------------------
200       1/1/1      n/a       n/a        n/a               Static    Up   Up
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
*A:PE-6#
```

The **show router mpls status** command is used to verify each of the LSP types, the number of configured LSPs and whether they originate on, transit through or terminate on the router.

```
*A:PE-1# show router mpls status
===============================================================================
MPLS Status
===============================================================================
Admin Status              : Up          Oper Status               : Up
Oper Down Reason          : n/a
FRR Object                : Enabled     Resignal Timer            : Disabled
Hold Timer                : 1 seconds   Next Resignal             : N/A
Srlg Frr                  : Disabled    Srlg Frr Strict           : Disabled
Admin Group Frr           : Disabled
Dynamic Bypass            : Enabled     User Srlg Database        : Disabled
BypassResignalTimer       : Disabled    BypassNextResignal        : N/A
LeastFill Min Thd         : 5 percent   LeastFill Reopti Thd      : 10 percent
Local TTL Prop            : Enabled     Transit TTL Prop          : Enabled
AB Sample Multiplier      : 1           AB Adjust Multiplier      : 288
Exp Backoff Retry         : Disabled    CSPF On Loose Hop         : Disabled
Lsp Init RetryTimeout     : 30 seconds  MBB Pref Current Hops     : Disabled
Logger Event Bundling     : Disabled
RetryIgpOverload          : Disabled

P2mp Resignal Timer       : Disabled    P2mp Next Resignal        : N/A
Sec FastRetryTimer        : Disabled    Static LSP FR Timer       : 30 seconds
P2P Max Bypass Association: 1000
P2PActPathFastRetry       : Disabled    P2MP S2L Fast Retry       : Disabled
In Maintenance Mode       : No
MplsTp                    : Disabled
Pce-report                : None
Next Available Lsp Index  : 2
Entropy Label RSVP-TE     : Enabled
===============================================================================
MPLS LSP Count
===============================================================================
                    Originate           Transit             Terminate
-------------------------------------------------------------------------------
```

```
Static LSPs         1                   0                   0
Dynamic LSPs        0                   0                   0
Detour LSPs         0                   0                   0
P2MP S2Ls           0                   0                   0
MPLS-TP LSPs        0                   0                   0
Mesh-P2P LSPs       0                   N/A                 N/A
One Hop-P2P LSPs    0                   N/A                 N/A
SR-TE LSPs          0                   N/A                 N/A
===============================================================================
```

*A:PE-1#PHP can be used with static LSPs. This is achieved by configuring the penultimate LER to swap the incoming label to implicit-null instead of a specific label value (the label-map must be shutdown to add the **swap** command).

```
*A:PE-5# configure router
        mpls
            interface "int-PE-5-PE-2"
                label-map 150
                    shutdown
                    swap implicit-null-label nexthop 192.168.56.2
                    no shutdown
                exit
```

The previous configuration will cause PE-5 to pop the top label from the incoming labeled frame received from PE-2 and send it to PE-6 without adding another outer label. The result can be seen from the following command (label 3 is never actually pushed onto a frame).

```
*A:PE-5# show router mpls static-lsp transit
===============================================================================
MPLS Static LSPs (Transit)
===============================================================================
In Label    In Port     Out Label   Out Port    Next Hop            Adm   Opr
-------------------------------------------------------------------------------
150         1/1/3       3           1/1/2       192.168.56.2        Up    Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================
*A:PE-5#
```

If the traffic arriving at PE-5 was IP with a single label, then it would arrive at PE-6 as unlabeled IP traffic.

If the static LSP spans a single hop (PE-1 to PE-2), the ingress LER would push the implicit-null instead of pushing a label.

```
*A:PE-1# configure router
        mpls
            static-lsp "LSP-PE-1-PE-2-static"
                to 192.0.2.2
                push implicit-null-label nexthop 192.168.12.2
                no shutdown
            exit
```

In this case, no MPLS action (swap or pop) is required for this LSP on PE-2.

Conclusion

MPLS provides the capability to establish connection-oriented paths over a connectionless network. The static LSP offers a mechanism to engineer network traffic In this chapter, the configuration of static LSPs is given together with the associated show output which can be used to verify and troubleshoot.

Tunneling of ICMP Reply Packets over MPLS LSPs

This chapter provides information about tunneling of ICMP reply packets over MPLS LSPs.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS routers and was initially written for release 13.0.R7. The CLI in this edition corresponds to release 14.0.R1. Internet control message protocol (ICMP) tunneling over multiprotocol label switching (MPLS) label switched paths (LSPs) is supported in SR OS release 12.0.R4 or later. Chassis-mode D must be used.

Overview

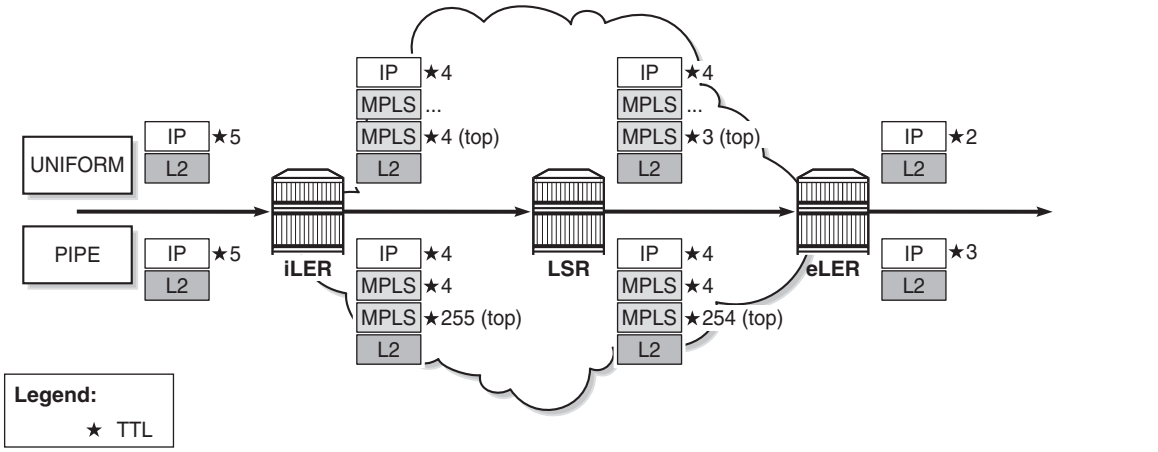
In IP forwarding, time-to-live (TTL) is a well-known mechanism to mitigate the damage in case of a loop. The TTL value in the IP header is decremented by one at each hop and the packet is discarded when the TTL equals 0. TTL is also used in traceroute, where the first batch of echo requests are sent with TTL equal to 1, the second batch of echo requests is sent with TTL equal to 2, and so on. Any intermediate node where the TTL expires (is decremented to 0) sends an ICMP reply of type "Time exceeded" (type 11) to the sender. From the replies, the sequence of hops can be determined.

If ICMP messages are sent in an MPLS tunnel, in pipe mode, the hops in the tunnels are invisible and the TTL is only decremented by the label edge routers (LERs), not by the intermediate label switching routers (LSRs). However, there are two modes for TTL handling, according to *RFC 3443: Time To Live Processing in MPLS Networks*:

1. Uniform mode: the MPLS network is visible from the outside. MPLS nodes use the TTL in the same way as any other IP node.
2. Pipe mode: the MPLS network is invisible from the outside. MPLS use of TTL is independent from IP TTL use. The network appears like a pipe between ingress label edge router (iLER) and egress label edge router (eLER).

Both TTL uses are shown in Figure 209:

Figure 209 **Use of TTL: Uniform versus Pipe**



Independent of the mode, the iLER decrements the TTL in the IP header by one. The iLER adds service and transport MPLS headers.

Note: In an L2 virtual private network (VPN), the TTL in the IP header is kept intact.

- In uniform mode, the iLER sets the TTL of every MPLS header to match the TTL in the IP header and every LSR decrements the MPLS TTLs. The IP header remains unchanged as long as the packet is in the MPLS tunnel. The eLER pops the MPLS labels and decrements the minimum TTL of the headers (which is the TTL in both MPLS headers) by one. This TTL is used in the IP header.
- In pipe mode, the iLER sets the TTL of the top MPLS header to 255 and every LSR decrements that TTL by one. The eLER pops the MPLS labels and decrements the minimum TTL of the headers (which is the IP TTL) by one. This TTL is used in the IP header. There can be uncounted hops in pipe mode, because the LSRs are not counted.

The LERs can be in uniform mode and the LSRs in pipe mode, and vice versa.

The default use of TTL in SR OS is as follows:

- Uniform mode for LSP shortcuts: resource reservation protocol (RSVP) shortcuts, label distribution protocol (LDP) shortcuts, and border gateway protocol (BGP) shortcuts.
- Pipe mode for L2 and L3 VPN services, BGP labeled routes, IPv6 provider edge (6PE) router, and IPv6 on VPN to PE router (6VPE).

However, the use of TTL can be changed by configuration.

Figure 210 shows the use of TTL for an L2 VPN service in pipe mode. The TTL in the IP header is preserved. There is no processing of the IP header for an L2 service. The TTL in the pushed MPLS headers is 255 and the TTL in the top MPLS header is decremented by one in the LSRs. The eLER pops the MPLS labels.

Figure 210 **Use of TTL in an L2 VPN Service in Pipe Mode**

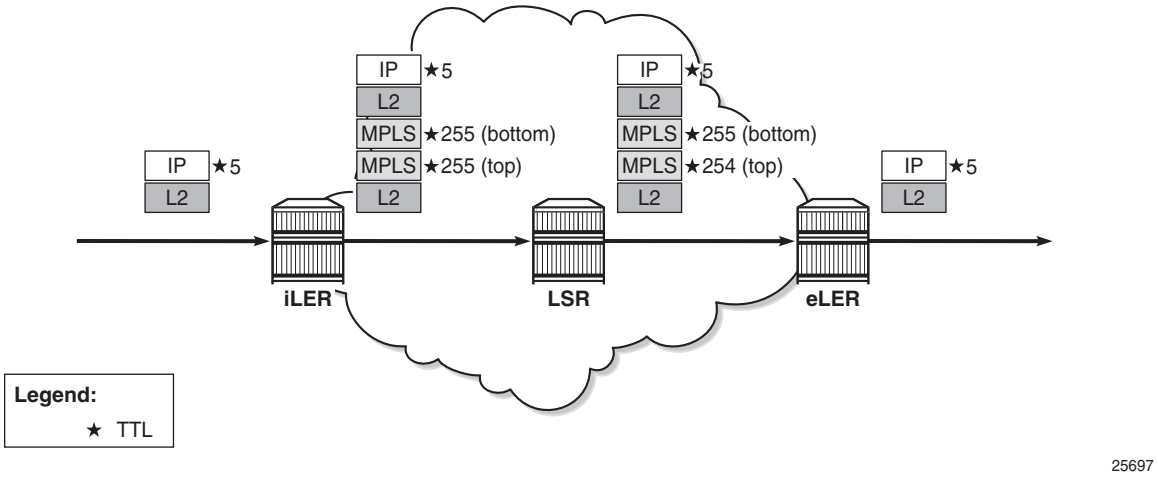
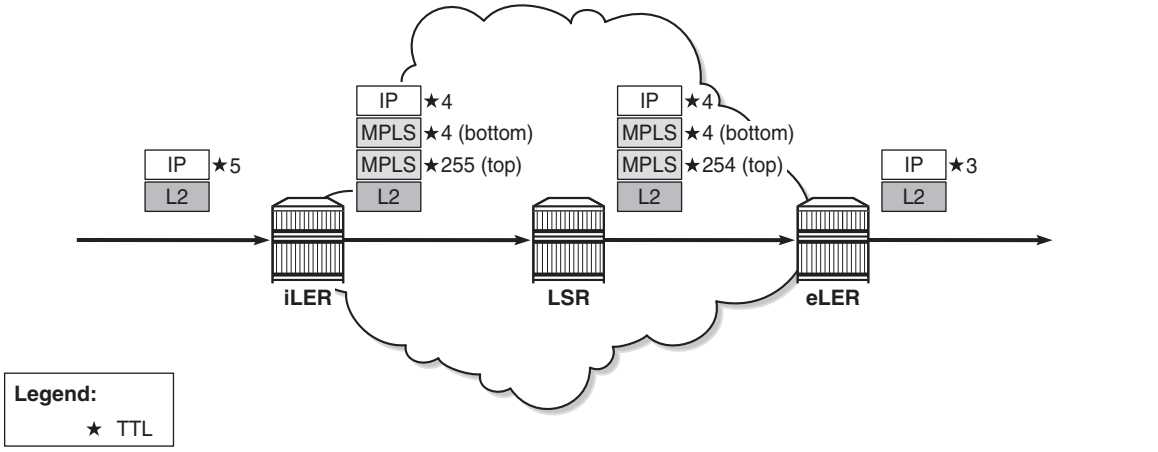


Figure 211 shows the use of TTL for an L3 VPN service in pipe mode. The TTL in the IP header is decremented by the iLER and the eLER, but not by the LSRs. In pipe mode, the bottom MPLS header inherits the IP TTL after it has been decremented by the iLER. The transport MPLS header gets TTL 255 and this TTL is decremented by one at each LSR. The eLER takes the minimum of the TTL of the MPLS headers and the IP TTL and decrements that by one. This will match the IP TTL in the forwarded packet. The MPLS labels are popped. There are uncounted hops, because the LSRs are invisible in pipe mode.

Figure 211 **Use of TTL in an L3 VPN Service in Pipe Mode**

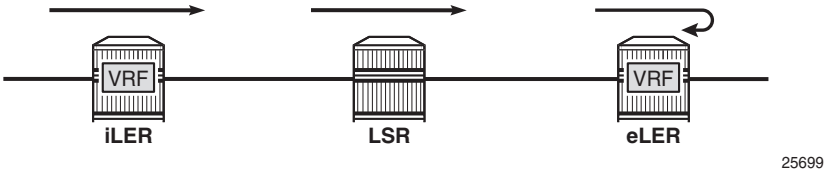


Tunneling of ICMP reply packets over MPLS LSPs provides the ability for a network operator or customer to trace the MPLS network hops in the path, for virtual private routed network (VPRN), 6PE/6VPE, and BGP labeled routes.

ICMP Tunneling over an MPLS LSP

Figure 212 shows the actions performed in iLER, LSR, and eLER, when tunneling ICMP messages over an MPLS LSP:

Figure 212 **Tunneling of ICMP Reply Packets over an MPLS LSP**



- In the iLER, uniform mode is required within the VPRN service. The IP TTL is propagated in the MPLS TTL, both for in-transit and control processing module (CPM) generated IP packets. In this example, it is assumed that a UDP traceroute message is forwarded with source IP address S1.

- In all LSRs, ICMP tunneling is enabled globally on the system, according to RFC 3032, MPLS Label Stack Encoding. When the MPLS TTL expires in an LSR, the LSR generates an ICMP reply with code "Time exceeded" and destination IP address S1. However, the CPM sends this ICMP reply packet in the forward direction of the MPLS LSP tunnel that the packet arrived on. The ICMP reply packet is sent to the eLER, not to the iLER.
- The eLER performs a lookup for the IP address S1 and sends the ICMP reply to S1 toward the iLER.
  - The lookup of the IP address S1 is in the global routing table (GRT) for BGP shortcut, 6PE, and BGP labeled route prefixes.
  - The lookup of IP address S1 is in the virtual routing and forwarding (VRF) table for VPRN and 6VPE prefixes.

TTL Propagation

The TTL propagation can be configured in LERs and LSRs.

TTL Propagation at the iLER

Different commands are used for TTL propagation in a VPRN versus BGP labeled routes. Pipe mode is enabled by default in either case.

TTL Propagation in iLER for VPRN

The TTL propagation of VPN-IPv4 or VPN-IPv6 packets in a VPRN service can be enabled globally as follows:

```
*A:PE-3# configure router ttl-propagate vprn-local
  - vprn-local <ttl-prop-type>

 <ttl-prop-type>      : none|all|vc-only - Default: vc-only


*A:PE-3# configure router ttl-propagate vprn-transit
  - vprn-transit <ttl-prop-type>

 <ttl-prop-type>      : none|all|vc-only - Default: vc-only
```

There are three options for the propagation of TTL in the iLER of a VPRN:

None	No IP TTL propagation to any MPLS header in the stack: the transport and the VC MPLS headers will have a value of 255. This is needed for proper operation of traceroute in an inter-AS option B VPRN.
All	Uniform mode: IP TTL is propagated to all MPLS headers in the stack.
VC-only (default)	Pipe mode: IP TTL is propagated to the VC header, but not to the transport headers in the stack.

For more information about inter-AS option B, see chapter Rosen MVPN Inter-AS Option B.

In inter-AS option B, a traceroute for a VPN IP prefix issued from a customer edge (CE) router results in both ingress autonomous system boundary router (ASBR) and egress ASBR not responding. The traceroute also misses a couple of hops if the target CE node is two or more hops away from the egress PE. The reason is that the VC label TTL inherits the decremented IP TTL at the ingress PE, but is decremented twice in the MPLS network while the IP TTL is only decremented at the ingress and the egress PE nodes. The option "none" for ttl-propagate makes the ASBRs transparent to the traceroute behavior and corrects the uncounted hop issue.

The global configuration can be overruled within each VPRN, as follows

```
*A:PE-3# configure service vprn 1 ttl-propagate local
  - local <ttl-prop-type>

 <ttl-prop-type>      : none|all|vc-only|inherit - Default: inherit


*A:PE-3# configure service vprn 1 ttl-propagate transit
  - transit <ttl-prop-type>

 <ttl-prop-type>      : none|all|vc-only|inherit - Default: inherit
```

TTL Propagation in iLER for BGP Labeled Route

IPv4 and IPv6 packets are forwarded using BGP labeled routes in the GRT, as described in *RFC 3107, Carrying Label Information in BGP-4.* This also applies to 6PE. TTL propagation for RFC 3107 label routes can be configured as follows:

```
*A:PE-3# configure router ttl-propagate label-route-local
  - label-route-local <ttl-prop-type>

 <ttl-prop-type>      : none|all - Default: none


*A:PE-3# configure router ttl-propagate label-route-transit
  - label-route-transit <ttl-prop-type>
```

```
<ttl-prop-type>      : none|all - Default: none
```

There are two options for TTL propagation in the iLER for BGP labeled routes:

None (default)	Pipe mode: No TTL is propagated from the IP header to the MPLS headers in the transport MPLS stack. However, the IP TTL is propagated to the bottom header: the virtual circuit (VC) header
All	Uniform mode: TTL is propagated to all headers in the transport MPLS stack.

If the BGP peer advertises the implicit-null label value for the BGP labeled route (in the case of a third-party implementation), the TTL propagation follows the configuration of the RSVP/LDP LSP shortcut that the BGP labeled route resolves to. This is not controlled by the preceding commands.

TTL Propagation at the LSR

In a VPRN service, there is no TTL propagation to be configured in the LSRs.

TTL Propagation in LSR for BGP Labeled Route

The IP TTL and VC TTL are not decremented by the LSRs. The TTL that is decremented is the minimum of the RSVP/LDP transport TTL and the BGP TTL.

Step 1. The LSR determines the TTL using the following function:

TTL = MIN {incoming transport label stack TTL, incoming swapped/stitched label TTL}

Step 2. The LSR decrements the TTL by one and writes it to the outgoing swapped/stitched BGP label.

This is always performed when an LSR is swapping or stitching a label at any stack depth.

The control plane indicates to the data plane whether a BGP labeled route is stitched or an LDP FEC is being stitched. The same node can perform stitching for one BGP labeled route and swapping for another one. See chapter LDP FEC to BGP Label Route Stitching for more information.

Step 3. The LSR can propagate the decremented TTL to the outgoing transport label stack (if any) that is pushed on top of the BGP swapped/stitched label. This is configured as follows:

```
*A:PE-2# configure router ttl-propagate lsr-label-route
  - lsr-label-route <ttl-prop-type>
```

```
<ttl-prop-type>      : none|all - Default: none
```

There are two options for TTL propagation in the LSR:

None (default)	No TTL propagation of the decremented TTL to the MPLS transport label stack.
All	TTL propagation of the decremented TTL to all LDP/RSVP transport labels.

It is safe to not propagate the TTL to the transport label stack for an ASBR/area border router (ABR)/data path route reflector (RR)/BGP-LDP stitching node. Not propagating the TTL provides isolation of the network domains downstream of the LSRs. Operations, administration, and maintenance (OAM) packets, such as traceroute and ping, sent in the context of a BGP labeled route or VPRN will not expire in LSR nodes within these domains.

A node performing pseudowire (PW) switching terminates the transport label stack in pipe mode; the node ignores the TTL of the incoming transport label stack and propagates the TTL of the VC label. The TTL of the new pushed transport label stack is always 255.

Some Considerations on TTL Propagation in LSR for BGP Labeled Routes

- When an LSR stitches an LDP label to a BGP label, the decremented TTL of the stitched label can be propagated to the LDP/RSVP transport labels with the preceding configuration.
- When an LSR stitches a BGP label to an LDP label, the decremented TTL of the stitched label is automatically propagated to the RSVP label if the outgoing LDP LSP is tunneled over RSVP.
- When the LSR pops a BGP label and forwards the packet using an IGP route (IGP route is preferred over BGP labeled route), the LSR pushes an LDP label on the packet and the TTL behavior is the same as when an LSR stitches a BGP label to an LDP label.
- In a carrier supporting carrier (CSC) VPRN, the ingress CSC CE swaps an iBGP label for an eBGP label and the ingress CSC PE swaps the incoming eBGP label for a VPN-IPv4 label. The reverse operation is performed by the egress CSC PE and the egress CSC CE. In all cases, the decremented TTL of the swapped label is propagated to the LDP/RSVP transport labels.
- SR OS does not support ASBR or data path RR functionality for labeled IPv6 routes in the global routing instance (6PE).

TTL Propagation at the eLER

For packets received with a BGP labeled route and searched for in the GRT, the TTL of the forwarded IP packet is set to MIN{MPLS_TTL-1, IP_TTL-1}, where MPLS_TTL refers to the TTL in the outermost label in the popped stack. This is the same behavior as for LSP shortcuts.

For packets received in the context of VPRN, the TTL of the forwarded IP packet is set to MIN{MPLS_TTL-1, VC_TTL-1, IP_TTL-1}, where MPLS_TTL refers to the TTL in the outermost label in the popped stack and VC_TTL refers to the TTL in the VC label in the popped stack.

Some Considerations on TTL Propagation at the eLER

- When a packet is received in one VPRN instance and is redirected using policy-based routing to be forwarded in another VPRN instance, the TTL is governed by the configuration of the outgoing VPRN instance.
- When a packet is received in a VPRN context but is searched for in the GRT (grt-leaking configured), the behavior of the TTL propagation is governed by:
  - the BGP labeled route configuration when the matching route is an RFC 3107 label route or a 6PE route
  - the LSP shortcut configuration when the matching route is an RSVP or LDP shortcut (default uniform mode)

    For shortcuts, uniform mode is default. Pipe mode can be configured as follows:

```
*A:PE-6# configure router ldp no shortcut-transit-ttl-propagate
*A:PE-6# configure router ldp no shortcut-local-ttl-propagate

*A:PE-6# configure router mpls no shortcut-transit-ttl-propagate
*A:PE-6# configure router mpls no shortcut-local-ttl-propagate
```

Enabling ICMP Tunneling on LSRs

For all scenarios (VPRN and BGP labeled routes), ICMP tunneling needs to be enabled on all LSRs, as follows:

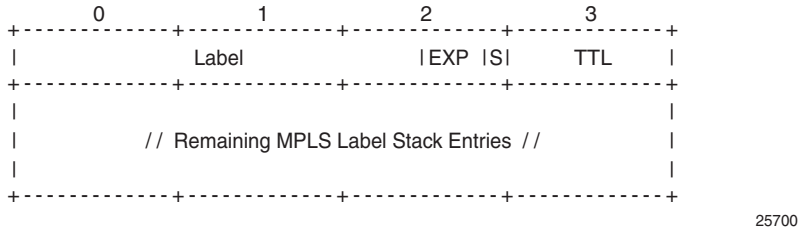
```
*A:PE-2# configure router icmp-tunneling
```

The LSR will generate the ICMP reply packet of type 11 - "Time exceeded", with source IP address set to a local address of the LSR node and appending the IP header and leading octets of the original datagram. The LSR does not perform a lookup for the destination IP address of the ICMP reply packet, which is the source IP address of the sender of the label TTL expiry packet. The CPM injects the ICMP reply packet in the forward direction toward the eLER. The TTL of pushed labels is 255.

There is no need to enable ICMP tunneling on the eLER. The eLER performs a user packet lookup in the data path in the VRF table or GRT and forwards the ICMP reply packet to the destination. If the eLER does not have a route to the destination, the packet is dropped.

RFC 4950: ICMP Extensions for Multiprotocol Label Switching defines an extension object (MPLS label stack object) that permits LSRs to include label stack information to ICMP messages; see Figure 213:

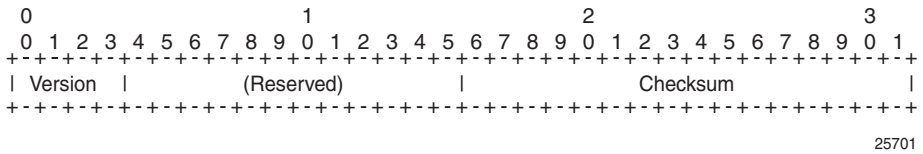
Figure 213 **MPLS Label Stack Object**



The MPLS label stack object is applicable for ICMPv4 and ICMPv6. The MPLS label stack contains the MPLS shim header: label, experimental bits for type of service (ToS), S-bit indicating the bottom of the stack, and TTL. The object can be appended to the ICMP Time Exceeded and ICMP Destination Unreachable messages. The LSR that sends the ICMP reply message will not change the MPLS label stack.

RFC 4884 Extended ICMP to Support Multi-Part Messages defines the ICMP extension header; see Figure 214:

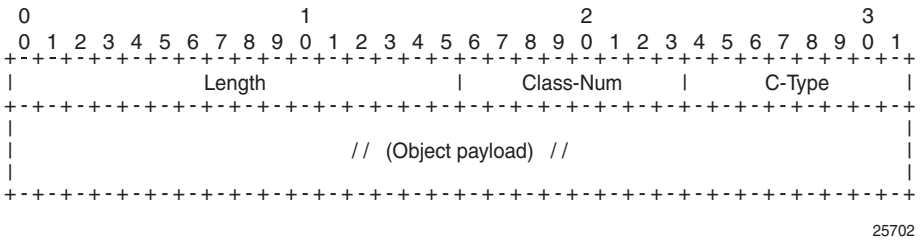
Figure 214 **ICMP Extension Header**



The version of the ICMP extension header is 2. The twelve reserved bits must be set to 0.

An extension object contains 32-bit words, representing an object header and payload, as defined in RFC 4884; see Figure 215:

Figure 215 **ICMP Extension Object: Object Header and Payload**



The length of the object is the length of the header (4 octets) plus the length of the object payload: 4 octets per LSR. The class number identifies the object class; in this case, object class 1 for MPLS label stack class. The C-type defines the object subtype; in this case, the subtype is 1 for an incoming MPLS label stack.

Backward compatibility is guaranteed between the ICMP message with extension header, the ICMP messages without extension header, and the ICMP message with a non-compliant extension header.

Effect of ICMP Tunneling on OAM

ICMP tunneling over MPLS LSPs affects the behavior of some CPM originated OAM packets that are forwarded within a VPRN context.

- ICMP ping and UDP traceroute are sent according to the TTL propagation configured in the VPRN context.
- VPRN-ping and VPRN-trace are not affected.

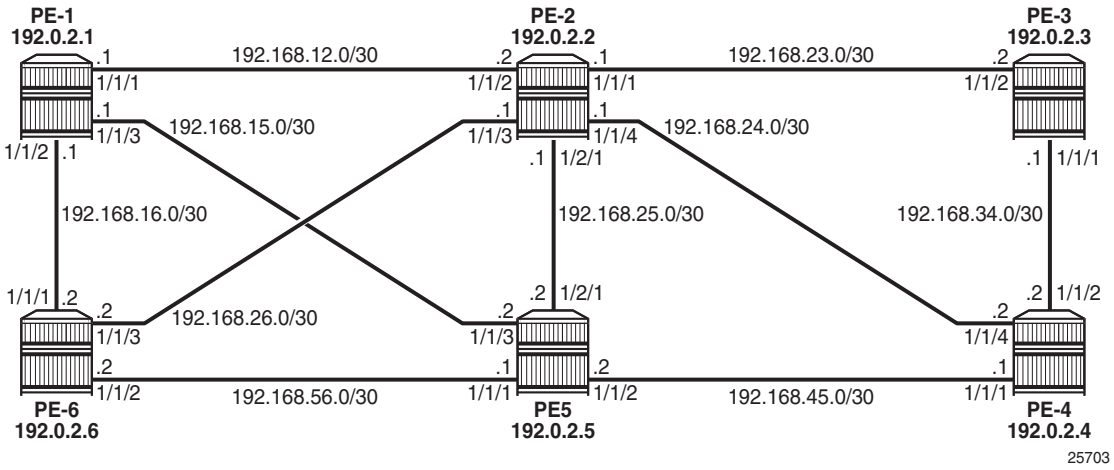
OAM packets forwarded over a BGP labeled route follow the TTL configuration of the iLER.

ICMP tunneling behavior at an LSR only applies to UDP traceroute packets. Other OAM packets expiring at the LSR, such as ICMP ping, VPRN ping, VPRN trace, LSP ping, and LSP trace, follow their specific procedures or are silently dropped.

Configuration

Figure 216 shows the example configuration, which has six 7750 SRs:

Figure 216 **Example Configuration**



Initial Configuration

The nodes have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces. For PE-3:

```
*A:PE-3# configure router
        interface "int-PE-3-PE-2"
            address 192.168.23.2/30
            port 1/1/2
        exit
        interface "int-PE-3-PE-4"
            address 192.168.34.1/30
            port 1/1/1
        exit
        interface "system"
            address 192.0.2.3/32
        exit
```

- IGP: OSPF (alternatively, any IGP could have been used). For PE-3:

```
*A:PE-3# configure router
        ospf
            area 0.0.0.0
                interface "system"
```

```
                exit
                interface "int-PE-3-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-3-PE-4"
                    interface-type point-to-point
                exit
            exit
        exit
```

- Link LDP. For PE-3:

```
*A:PE-3# configure router
        ldp
            interface-parameters
                interface "int-PE-3-PE-2"
                exit
                interface "int-PE-3-PE-4"
                exit
            exit
```

Configure VPRN

A VPRN service is configured on PE-3 and PE-6. The routes are exchanged via BGP. BGP is configured on all nodes with PE-2 as route reflector (RR).

```
*A:PE-3# configure router
        autonomous-system 64496
        bgp
            group "internal"
                family vpn-ipv4
                peer-as 64496
                neighbor 192.0.2.2
                exit
            exit
            no shutdown
```

The configuration on RR PE-2 is as follows:

```
*A:PE-2# configure router
        autonomous-system 64496
        bgp
            cluster 1.1.1.1
            group "internal"
                family vpn-ipv4
                peer-as 64496
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
```

```
                exit
                neighbor 192.0.2.6
                exit
            exit
            no shutdown
```

Import and export policies are configured on PE-3 and PE-6, as follows:

```
*A:PE-3# configure router
        policy-options
            begin
            community "VPN1" members "target:64496:1"
            policy-statement "VPN1-export"
                entry 10
                    from
                        protocol direct
                    exit
                    to
                        protocol bgp-vpn
                    exit
                    action accept
                        community add "VPN1"
                    exit
                exit
            exit
            policy-statement "VPN1-import"
                entry 10
                    from
                        protocol bgp-vpn
                        community "VPN1"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

VPRN 1 is configured on PE-3 and PE-6, as follows:

```
*A:PE-3# configure service
        vprn 1 customer 1 create
            vrf-import "VPN1-import"
            vrf-export "VPN1-export"
            route-distinguisher 64496:13
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            interface "loopback1" create
                address 192.0.1.3/32
                loopback
            exit
            no shutdown
        exit
```

The configuration on PE-6 is similar, with loopback address 192.0.1.6/32.

Default TTL Handling in VPRN

The default configuration for TTL propagation on the iLER corresponds to pipe mode, as follows:

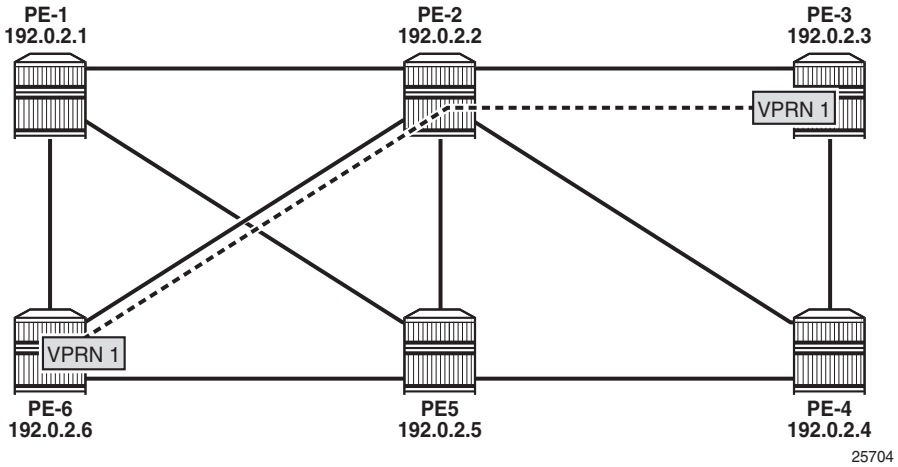
```
*A:PE-3# configure router ttl-propagate
*A:PE-3>config>router>ttl-propagate# info detail
----------------------------------------------
            label-route-local none
            label-route-transit none
            lsr-label-route none
            vprn-local vc-only
            vprn-transit vc-only
----------------------------------------------
```

No ICMP tunneling is enabled in the LSRs, which implies that no ICMP "Time exceeded" messages will be tunneled by the LSR to the eLER. A traceroute message sent in VPRN 1 from PE-3 to the loopback address in VPRN 1 on PE-6 shows that the loopback address is the next hop. There are no intermediate hops detected.

```
*A:PE-3# traceroute router 1 192.0.1.6
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
  1  192.0.1.6 (192.0.1.6)    1.84 ms  1.76 ms  1.75 ms
*A:PE-3#
```

Figure 217 shows the tunnel from iLER PE-3 to eLER PE-6:

Figure 217 **Tunnel from iLER PE-3 to eLER PE-6 via LSR PE-2**



For comparison, a traceroute in the base router toward the system address of PE-6 shows PE-2 as intermediate hop, as follows:

```
*A:PE-3# traceroute 192.0.2.6
traceroute to 192.0.2.6, 30 hops max, 40 byte packets
  1  192.168.23.1 (192.168.23.1)    0.714 ms  0.687 ms  0.695 ms
  2  192.0.2.6 (192.0.2.6)    1.73 ms  1.63 ms  1.61 ms
*A:PE-3#
```

Uniform Mode in iLER and ICMP Tunneling in LSR

In the iLER PE-3, uniform mode is enabled for local VPRNs, as follows:

```
*A:PE-3# configure service vprn 1 ttl-propagate local all
```

This is a specific configuration for VPRN 1 that overrules the global configuration. By default, it is set to inherit the global configuration. By default, the global configuration is pipe mode.

This TTL propagation is only configured on PE-3, not on PE-6. This implies that traceroute messages from the VPRN in PE-3 will have TTL propagation to all MPLS labels (uniform mode), while traceroute messages from the VPRN in PE-6 will have pipe mode.

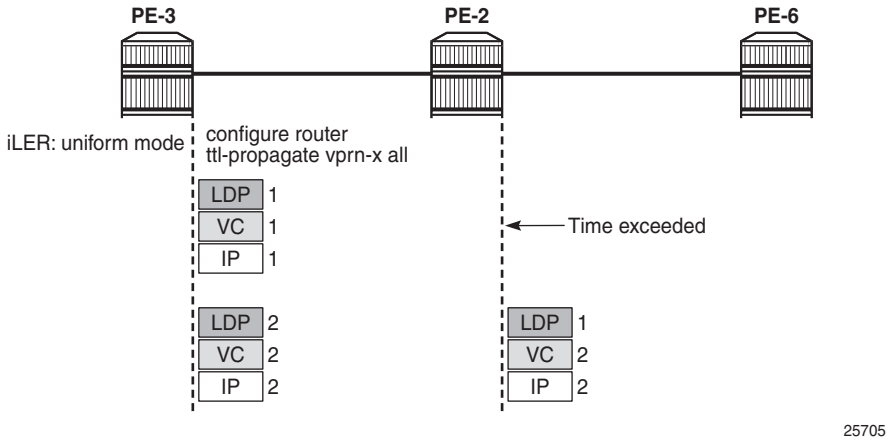
In the LSR PE-2, ICMP tunneling is enabled, as follows:

```
*A:PE-2# configure router icmp-tunneling
```

A UDP traceroute is sent from VPRN 1 on PE-3 to the loopback address in VPRN 1 on PE-6. A message with TTL 1 is sent first. The IP TTL and VC TTL are not decremented by the LSR PE-2. Only the LDP TTL is decremented, so this message times out on the LSR PE-2, which tunnels the ICMP Time Exceeded reply with destination address 192.0.1.3 toward eLER PE-6. The eLER looks up the prefix 192.0.1.3 in the VRF table and sends the ICMP Time Exceeded reply toward VPRN 1 in PE-3. Three UDP traceroute messages with TTL 1 are sent. Then, UDP traceroute messages with TTL 2 are sent. They reach the destination PE-6 before time-out.

Figure 218 shows the TTLs in the UDP traceroute messages:

Figure 218 **UDP Traceroute in VPRN with iLER in Uniform Mode**



In the following output, there is only an MPLS label stack object when TTL expires in the LSR, because ICMP tunneling is occurring. The MPLS label stack object is not used by the eLER. The LSR where ICMP tunneling occurs adds an MPLS label stack object to the ICMP reply message. The MPLS label stack object contains information about the MPLS labels (VC label and LDP transport label) in the stack: MPLS labels, experimental bits for ToS, and TTL. S indicates the bottom of the label stack. In the detailed output of the traceroute command, the MPLS label stack information is shown for the echo requests that timed out in the LSR:

```
*A:PE-3# traceroute router 1 192.0.1.6 detail
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
  1   1  192.168.26.1  (192.168.26.1)  1.66 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262138, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   1, S = 1
  1   2  192.168.26.1  (192.168.26.1)  1.32 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262138, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   1, S = 1
  1   3  192.168.26.1  (192.168.26.1)  1.57 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262138, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   1, S = 1
  2   1  192.0.1.6  (192.0.1.6)  1.89 ms
  2   2  192.0.1.6  (192.0.1.6)  1.54 ms
  2   3  192.0.1.6  (192.0.1.6)  1.86 ms

*A:PE-3#
```

The top label or transport label 262138 is the LDP label pushed by PE-3:

```
*A:PE-3# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
```

```
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                    Op              IngLbl    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                              Push              --      262138
192.168.23.1                              1/1/2

192.0.2.6/32                              Swap            262138    262138
192.168.23.1                              1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-3#
```

The bottom label or service label 262137 is the BGP label, which remains the same end-to-end:

```
*A:PE-3# show router bgp routes vpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.3        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     64496:13:192.0.1.3/32                             100         None
      192.0.2.3                                         None        262137
      No As-Path
u*>i  64496:16:192.0.1.6/32                             100         None
      192.0.2.6                                         None        262137
      No As-Path
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-3#
```

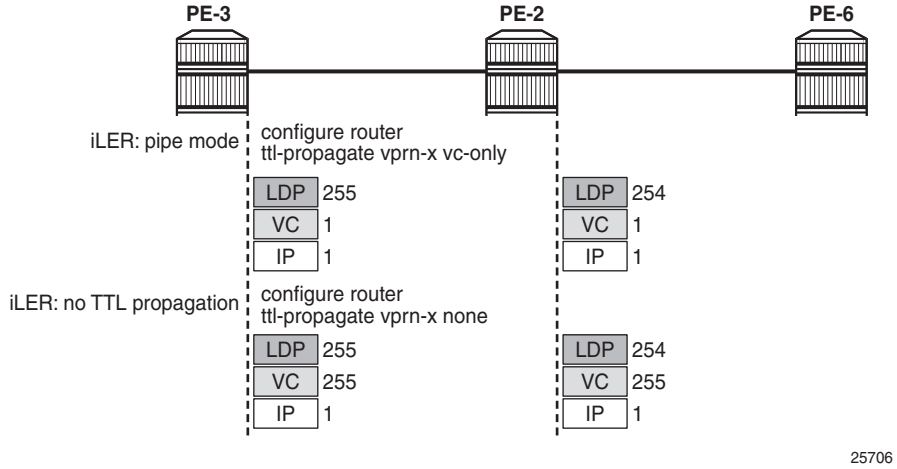
When the iLER is configured in pipe mode (VC-only) or if there is no TTL propagation to any MPLS label (none), the output of the traceroute detail command does not contain the MPLS label stack object information. In pipe mode, the IP TTL is propagated to the VC header, but not to the LDP header. When the TTL propagation is none, the IP TTL is not propagated to VC or LDP. The LSRs are invisible and there will be missing hops.

```
*A:PE-3# configure service vprn 1 ttl-propagate local vc-only
*A:PE-3# traceroute router 1 192.0.1.6 detail
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
  1    1  192.0.1.6  (192.0.1.6)  1.80 ms
  1    2  192.0.1.6  (192.0.1.6)  1.85 ms
  1    3  192.0.1.6  (192.0.1.6)  2.16 ms

*A:PE-3#
```

The reason is that the TTL of the LDP header is 255 at the iLER and it is decremented by one in every LSR. The UDP traceroute message will not time out on the LSR PE-2. The TTL of the VC header is not decremented in the LSR. When the traceroute message does not time out in PE-2, the hop PE-2 is invisible. In a similar way, the traceroute messages will not time out in the LSR when no TTLs are propagated to any MPLS header. Figure 219 shows the TTLs in both cases:

Figure 219 **UDP Traceroute in VPRN without TTL Propagation to LDP**



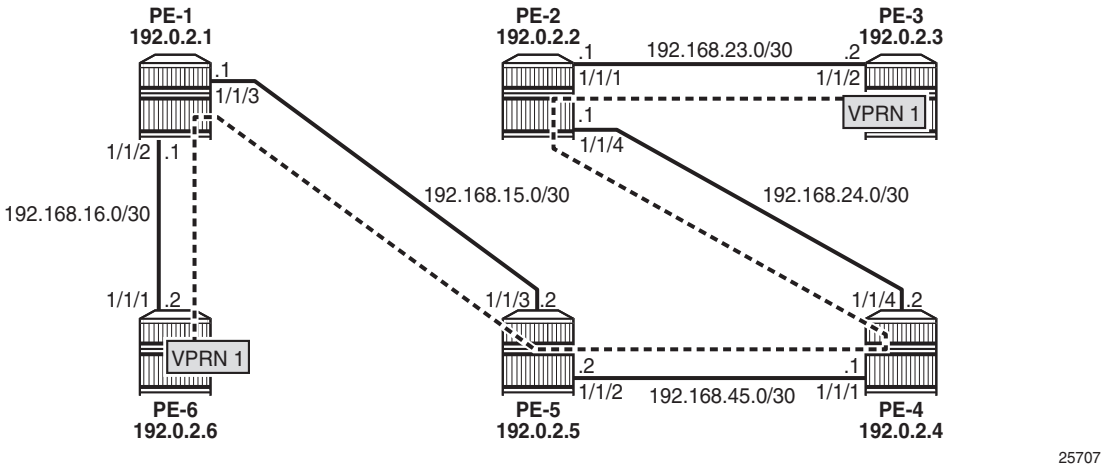
The TTL propagation is restored to uniform mode in the iLER as follows:

```
*A:PE-3# configure service vprn 1 ttl-propagate local all
```

Uniform Mode in iLER and ICMP Tunneling in Multiple LSRs

After shutting down some ports in the nodes, the tunnel from iLER PE-3 to PE-6 has four intermediate hops (LSRs) instead of one, as shown in Figure 220:

Figure 220 **Tunnel from iLER PE-3 to eLER PE-6 with Multiple LSRs**



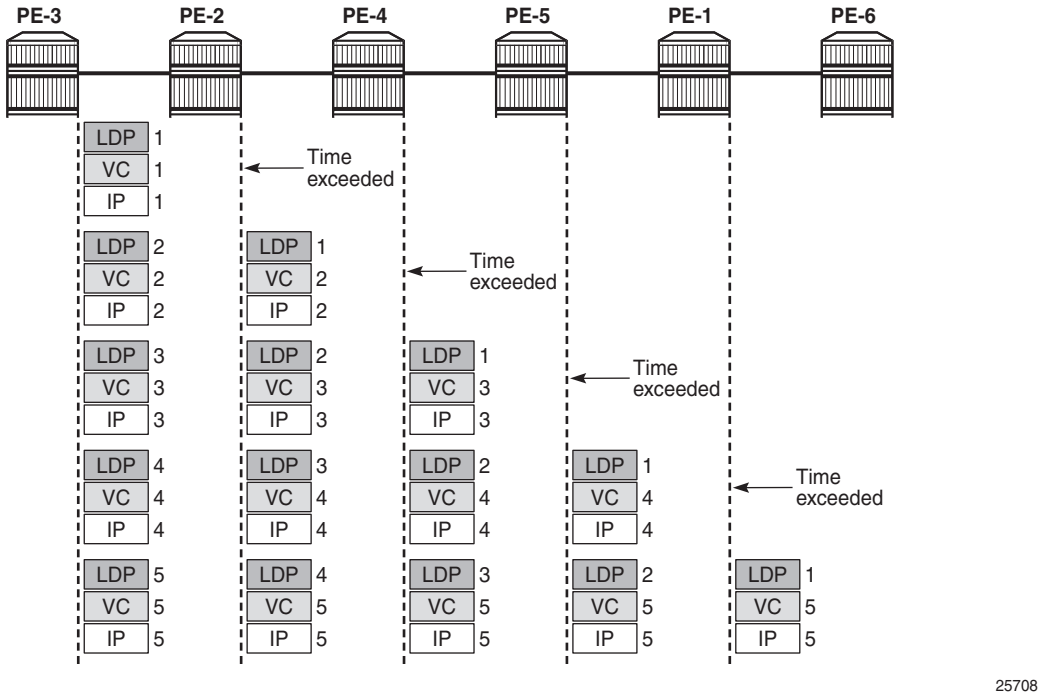

```
*A:PE-2# configure port 1/1/[2..3] shutdown
*A:PE-2# configure port 1/2/1 shutdown
*A:PE-3# configure port 1/1/1 shutdown
*A:PE-5# configure port 1/1/1 shutdown
```

TTL propagation on the iLER PE-3 is in uniform mode. ICMP tunneling needs to be enabled on all LSRs, as follows:

```
*A:PE-1# configure router icmp-tunneling
```

UDP traceroute messages are sent from VPRN 1 on iLER PE-3 to VPRN 1 on PE-6, as shown in Figure 221:

Figure 221 **UDP Traceroute with iLER in Uniform Mode**



The detailed output of the traceroute command shows the MPLS label stack object information as added by the LSR where the ICMP Time Exceeded message was tunneled. For brevity, only the first of the three messages from each intermediate node is shown.

```
*A:PE-3# traceroute router 1 192.0.1.6 detail
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
  1   1  192.168.24.1  (192.168.24.1)  3.90 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262135, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   1, S = 1
  ---snip---
  2   1  192.168.45.1  (192.168.45.1)  2.81 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262137, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   2, S = 1
  ---snip---
  3   1  192.168.15.2  (192.168.15.2)  2.80 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262138, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   3, S = 1
  ---snip---
  4   1  192.168.16.1  (192.168.16.1)  2.97 ms
         returned MPLS Label Stack Object
             entry  1:  MPLS Label =  262138, Exp = 7, TTL =   1, S = 0
             entry  2:  MPLS Label =  262137, Exp = 7, TTL =   4, S = 1
  ---snip---
```

```
  5   1  192.0.1.6  (192.0.1.6)  2.90 ms
  ---snip---

*A:PE-3#
```

The TTL for the bottom MPLS header (BGP) is not decremented in each hop; only the TTL for the transport MPLS header (LDP) is decremented. The bottom label or BGP label of 262137 is not changed end-to-end. The LDP transport label for the different nodes is as follows.

For iLER PE-3, the LDP transport label for traffic toward PE-6 is 262135:

```
*A:PE-3# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push              --      262135
192.168.23.1                            1/1/2

192.0.2.6/32                            Swap            262134    262135
192.168.23.1                            1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-3#
```

For LSR PE-2, the LDP transport label toward PE-6 is 262137:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
---snip---
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push              --      262137
192.168.24.2                            1/1/4
```

```
192.0.2.6/32                                    Swap            262135    262137
192.168.24.2                                    1/1/4

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-2#
```

For LSR PE-4, the LDP transport label toward PE-6 is 262138:

```
*A:PE-4# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
---snip---
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                          Op              IngLbl    EgrLbl
EgrNextHop                                      EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                                    Push              --      262138
192.168.45.2                                    1/1/1

192.0.2.6/32                                    Swap            262137    262138
192.168.45.2                                    1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-4#
```

For LSR PE-5, the LDP transport label toward PE-6 is 262138:

```
*A:PE-5# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
---snip---
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                          Op              IngLbl    EgrLbl
EgrNextHop                                      EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                                    Push              --      262138
192.168.15.1                                    1/1/3

192.0.2.6/32                                    Swap            262138    262138
192.168.15.1                                    1/1/3

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-5#
```

For LSR PE-1, the LDP transport label toward PE-6 is 262143, but this label will not be present in the traceroute detailed output, because this message cannot time out on an LSR where ICMP tunneling takes place:

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
---snip---
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push              --      262143
192.168.16.2                            1/1/2

192.0.2.6/32                            Swap            262138    262143
192.168.16.2                            1/1/2

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

Conclusion

Tunneling of ICMP reply messages over MPLS LSPs provides the ability to trace the hops in an MPLS tunnel. This mechanism applies to VPRN, 6PE/6VPE, and BGP labeled routes.

ICMP tunneling at an LSR applies to UDP traceroute packets that time out at the LSR. The ICMP Time Exceeded message is tunneled by the LSR toward the eLER and the eLER routes the packet to the sender of the traceroute message.

Unnumbered Interfaces in RSVP-TE and LDP

This chapter provides information about Unnumbered Interfaces in RSVP-TE and LDP.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS routers and was initially written for SR OS release 13.0.R7. The CLI in this edition corresponds to release 14.0.R1. SR OS supports unnumbered interfaces in RSVP-TE and LDP in release 11.0.R1 or later.

Overview

Unnumbered interfaces enable IP processing on a point-to-point (P2P) interface without an explicit IP address. Unnumbered interfaces are supported in resource reservation protocol with traffic engineering (RSVP-TE) and label distribution protocol (LDP).

An unnumbered interface is uniquely identified in the network by the tuple (router-id,ifIndex), where the interface index (ifIndex) is unique on the router. The two endpoints of an unnumbered link exchange the ifIndex that they assigned to the link.

The (router-id,ifIndex) tuple is used by the following:

- Intermediate system to intermediate system (IS-IS) or open shortest path first (OSPF) to advertise link information
- RSVP to signal label switched paths (LSPs) over this unnumbered interface
- LDP to establish hello adjacencies and resolve forwarding equivalence classes (FECs)

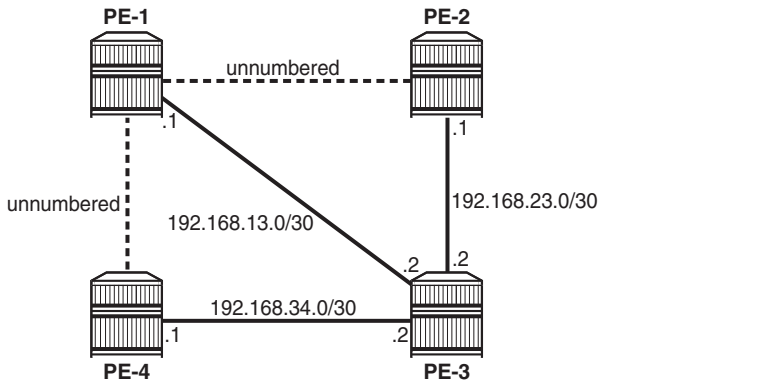
- Operations, administration, and maintenance (OAM) to send or respond to an multiprotocol label switching (MPLS) echo request over an unnumbered interface

The unnumbered interface can "borrow" the IP address of another interface on the node.

The borrowed IP address is used exclusively as the source address for IP packets that are originated from the unnumbered interface. The borrowed IP address defaults to the system loopback interface address, but can be changed manually. The borrowed IP address corresponds to the router ID in the tuple representing the unnumbered interface.

The configuration used in this chapter is shown in Figure 222. There are two unnumbered links: one between PE-1 and PE-2, and another between PE-1 and PE-4. The remaining links are numbered.

Figure 222 **Example Topology for Unnumbered Interfaces in RSVP and LDP**



Configure unnumbered interfaces as follows:

```
configure router interface <itf-name> unnumbered [<ip-int-name|ip-address>]
```

To configure an unnumbered link with the system address as the borrowed IP address, no address needs to be configured:

```
*A:PE-4# configure router
        interface "int-PE-4-PE-1"
            port 1/1/1
            unnumbered
        exit
```

An unnumbered interface has to be a P2P link.

Unnumbered Interfaces in IS-IS

Unnumbered interfaces are identified in IS-IS by a combination of the system ID and an extended local circuit ID, as described in *RFC 5307 IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching*.

Enable debugging on the unnumbered interface on PE-4 (int-PE-4-PE-1) as follows:

```
*A:PE-4# debug router isis packet "int-PE-4-PE-1" detail
```

The following IS-IS hello protocol data unit (PDU) is received from PE-1. The I/F Address is the borrowed IP address; in this case, the system address of PE-1, because the interface is unnumbered.

```
76 2016/01/29 12:50:42.96 UTC MINOR: DEBUG #2001 Base ISIS PKT
"ISIS PKT:
RX ISIS PDU ifId 2 len 52:
  DMAC          : 09:00:2b:00:00:05
  Proto Disc    : 131
  Header Len    : 20
  Version PID   : 1
  ID Length     : 0
  Version       : 1
  Reserved      : 0
  Max Area Addr : 3
  PDU Type      : (11) Point-2-Point IS-IS Hello Pdu
  Circuit Type  : L1
  Source Id     : 19 20 00 00 20 01
  Hold Time     : 27
  Packet length : 52
  Circuit Id    : 0
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols     : IPv4
  I/F Addresses :
    I/F Address   : 192.0.2.1
  3Way Adjacency   :
    State          : UP
    Ext ckt ID     : 4
    NbrSysID       : 19 20 00 00 20 04
    Nbr ext ckt ID : 2
```

The three-way adjacency contains the neighbor extended local circuit ID (Nbr ext ckt ID: 2). This ID is the local interface index of the unnumbered interface (int-PE-4-PE-1), which can be verified as follows:

```
*A:PE-4# show router interface "int-PE-4-PE-1" detail | match "If Index"
If Index         : 2                    Virt. If Index    : 2
Last Oper Chg    : 01/29/2016 12:40:28  Global If Index   : 1
*A:PE-4#
```

On PE-1, the interface toward PE-4 has a different index, as follows:

```
*A:PE-1# show router interface "int-PE-1-PE-4" detail | match "If Index"
If Index         : 4                    Virt. If Index    : 4
Last Oper Chg    : 01/29/2016 12:40:14  Global If Index   : 3
*A:PE-1#
```

For numbered interfaces, such as int-PE-4-PE-3, the I/F Address is the interface address; in that case, 192.168.34.1, for messages received from PE-3 instead of the router ID.

When a shared risk link group (SRLG) is configured in combination with IS-IS and unnumbered interfaces, the least significant bit in the flags field of the SRLG type-length value (TLV) indicates that the interface is unnumbered (0) or numbered (1).

Unnumbered Interfaces in OSPF

For unnumbered interfaces in OSPF, link local and remote identifiers are defined in *RFC 4203 OSPF Extensions in Support of Generalized Multi-Protocol Label Switching*. The OSPF link state advertisement (LSA) is defined in *RFC 2328, OSPF version 2*.

For numbered interfaces, the link data is the IP interface address; for unnumbered interfaces, the link data is the interface index value. The value starts from 1 in the format 0.0.0.1. SR OS recognizes an unnumbered interface when the first byte in the link data has a value of 0; SR OS then treats the link data as an interface index instead of an IP address.

Unnumbered Interfaces in RSVP-TE

Unnumbered IP interfaces can be used as traffic engineering (TE) links for the signaling of RSVP P2P LSPs and point-to-multipoint (P2MP) LSPs.

Fast Reroute (FRR) facility backup over unnumbered interfaces is supported, while FRR one-to-one will only use numbered interfaces in the detour path.

The unnumbered IP address is advertised by IS-IS or OSPF, and constrained shortest path first (CSPF) can include them in the computation of a path.

Unnumbered interfaces of the remote router can be specified in the explicit route object (ERO), and in the record route object (RRO), by a combination of router ID (borrowed IP address) and interface ID, as defined in *RFC 3477 Signaling Unnumbered Links in RSVP-TE*.

The choice of the data interface is indicated in the path message by including the interface identifier of the data channel. In the path message, the IP address equals the local borrowed IP address; in the resv message, the IP address is the remote borrowed IP address. As well as the borrowed IP address, there is also a logical interface handle (LIH). This interface identification is defined in *RFC 3473 Generalized Multi-Protocol Label Switching Signaling Resource Reservation Protocol Traffic Engineering Extensions*.

To see the path and resv messages on PE-4, enter the following debug commands:

```
*A:PE-4# debug router rsvp packet path detail
*A:PE-4# debug router rsvp packet resv detail
```

The path message contains the RSVPHop object with the local interface identifier of the data channel, as follows:

```
960 2016/03/28 11:50:21.83 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.4, To:192.0.2.2
          TTL:255, Checksum:0x3636, Flags:0x0
Session    - EndPt:192.0.2.2, TunnId:1, ExtTunnId:192.0.2.4
SessAttr   - Name:LSP-PE-4-PE-2::dyn
             SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop    - Ctype:3, Addr:192.1.2.4, LIH:2
             RouterId :192.0.2.4, InterfaceId :2
TimeValue  - RefreshPeriod:30
SendTempl  - Sender:192.0.2.4, LspId:31266
SendTSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1564
LabelReq   - IfType:General, L3ProtID:2048
RRO        - Unnumbered: RouterId 192.0.2.4 InterfaceID 2, Flags:0x0
ERO        - Unnumbered RouterId 192.0.2.1, LinkId 4, Strict
             Unnumbered RouterId 192.0.2.2, LinkId 2, Strict
FRRObj     - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
             ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

The ERO and RRO objects are also shown. The unnumbered interfaces are defined by the combination of the router ID and the interface ID.

The resv message also contains the RSVPHop object, but the address is now the remote address of PE-1 instead of the local address of PE-4, as follows:

```
962 2016/03/28 11:50:29.72 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESV Msg
Recv RESV From:192.0.2.1, To:192.0.2.4
          TTL:255, Checksum:0x4e21, Flags:0x0
```

```
Session    - EndPt:192.0.2.2, TunnId:1, ExtTunnId:192.0.2.4
RSVPHop    - Ctype:3, Addr:192.0.2.1, LIH:2
             RouterId :192.0.2.1, InterfaceId :4
TimeValue  - RefreshPeriod:30
Style      - SE
FlowSpec   - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1560, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:192.0.2.4, LspId:31266, Label:262142
RRO        - Unnumbered: RouterId 192.0.2.1 InterfaceID 4, Flags:0x1
             Label:262142, Flags:0x1
             Unnumbered: RouterId 192.0.2.2 InterfaceID 2, Flags:0x0
             Label:262142, Flags:0x1
"
```

To see the patherr and resverr messages on PE-4, enter the following debug command:

```
*A:PE-4# debug router rsvp packet patherr detail
*A:PE-4# debug router rsvp packet resverr detail
```

The patherr message contains the following ErrorSpec object, as defined by RFC 3473. In this case, the error is caused by disabling TE on ingress label egress router (iLER) PE-4. No route can be found to the destination because there is no lookup in the TE database. The LSP will not come up, even if CSPF is disabled on the LSP.

```
1079 2016/03/28 12:05:55.73 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESVERR Msg
Send RESVERR From:192.1.2.4, To:192.0.2.1
             TTL:255, Checksum:0xf8ae, Flags:0x0
Session    - EndPt:192.0.2.2, TunnId:1, ExtTunnId:192.0.2.4
RSVPHop    - Ctype:3, Addr:192.1.2.4, LIH:2
             RouterId :192.0.2.4, InterfaceId :2
ErrorSpec  - Ctype:3, ErrNode:192.1.2.4, Flags:0x0, ErrCode:3, ErrValue:0
             RouterId :192.0.2.4, InterfaceId :2
Style      - SE
FlowSpec   - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
             MPU:20, MTU:1560, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:192.0.2.4, LspId:31266
"
```

Considerations for Unnumbered Interfaces in RSVP-TE

Consider the following for unnumbered interfaces in RSVP-TE:

- With RSVP, TE must be enabled in IS-IS or OSPF. The router ID of the router that advertised an unnumbered interface index is obtained from the TE database. Therefore, if TE is disabled in IS-IS or OSPF, a non-CSPF LSP with the next hop for this path over an unnumbered interface will not come up. The router ID of the neighbor that has the next hop of the path message cannot be searched for.

  - The operational state of the LSP path will remain down with reason "noRouteToDestination".
  - If a path message is received at the LSR in which TE is disabled and the next hop for the LSP path is over an unnumbered interface, a patherr message is sent back to the iLER with error code 24: Routing problem; Error value 5: "No route available toward destination".
- "Only FRR facility protection is supported; FRR one-to-one protection only works for numbered interfaces.
- There is no FRR facility protection if the point of local repair (PLR) is the iLER and the bypass tunnel egress interface is unnumbered.
- Bi-directional forwarding detection (BFD) cannot be enabled on an unnumbered router interface. Therefore, RSVP FRR procedures will not be triggered via a BFD session timeout, but only by physical failures and local interface down events.
- Unnumbered interfaces cannot be configured as hops in a path. This is true for RSVP-TE LSPs, as well as for static LSPs.
- RSVP hello and hello related capabilities, such as graceful restart helper, are not supported.
- SRLG is supported, but the user SRLG DB (user-srlg-db) feature at the iLER is not supported. Unnumbered interfaces cannot be added to the SRLG DB. When the user SRLG DB feature is enabled on the iLER, all unnumbered interfaces are considered as having no SRLG membership.

Unnumbered Interfaces in LDP

LDP can establish hello adjacencies and can resolve unicast and multicast FECs over unnumbered interfaces.

For link LDP, hello adjacencies are brought up using hello packets with source IP address set to the borrowed IP address and a destination IP address set to 224.0.0.2. The borrowed IP address is the system address, by default. Hello packets with the same source IP address are accepted when received over parallel unnumbered interfaces from the same peer LSR ID. The corresponding hello adjacencies are associated with a single LDP session.

The transport address for the TCP connection, which is encoded in the hello packet, will always be set to the LSR ID of the node. The user can configure the local-lsr-id option on the interface and change the value of the LSR ID to either the local interface or some other interface name: loopback or not, numbered or not. The transport address for the LDP session will be updated with the new LSR ID.

For targeted LDP, the source and destination addresses of targeted hello packets are the LDP LSR IDs of the nodes. The user can configure the local-lsr-id option on the targeted session. The transport address for the LDP session and the source IP address of targeted hello messages will be updated to the new LSR ID value.

LDP will advertise/withdraw unnumbered interfaces using the address/address-withdraw messages. The borrowed IP address of the interface is used.

A FEC can be resolved to an unnumbered interface in the same way as it is resolved to a numbered interface. The outgoing interface and the next hop are searched for in the routing table manager (RTM). The next hop consists of the router ID and link identifier of the interface to the peer LSR. All LDP FEC types are supported. LDP FEC equal cost multi-path (ECMP) over a mix of unnumbered and numbered interfaces is supported.

RFC 5036 LDP Specification describes the address list TLV that is used in the LDP address message, and the LDP address withdrawal message. For unnumbered interfaces, the borrowed IP address is used, which is typically the system address of the sender node.

On PE-1, enable debugging for LDP packets from peer 192.0.2.2 as follows:

```
*A:PE-1# debug router ldp peer 192.0.2.2 packet init detail
*A:PE-1# debug router ldp peer 192.0.2.2 packet label detail
```

The following LDP address packets are shown at PE-1:

```
278 2016/02/02 07:59:20.11 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address packet (msgId 6) to 192.0.2.2:0
Protocol version = 1
Address Family = 1  Number of addresses = 2
Address 1 = 192.0.2.1
Address 2 = 192.168.13.1
"


277 2016/02/02 07:59:20.09 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address packet (msgId 6) from 192.0.2.2:0
Protocol version = 1
Address Family = 1  Number of addresses = 2
Address 1 = 192.0.2.2
Address 2 = 192.168.23.1
"
```

The received LDP address packet contains two addresses: the system IP address 192.0.2.2 for the unnumbered interfaces and the interface address 192.168.23.1 for a numbered interface on that node.

Considerations for Unnumbered Interfaces in LDP

All LDP features are supported on unnumbered interfaces, except for the following:

- BFD cannot be enabled on an unnumbered router interface. Therefore, LDP FRR procedures will not be triggered via a BFD session timeout, but only by physical failures and local interface down events.
- Unnumbered interfaces cannot be added into LDP global and peer prefix policies.

Unnumbered Interfaces in OAM

The following applies to unnumbered interfaces in RSVP-TE or LDP.

The downstream mapping object is a TLV that can be included in an echo request, as described in *RFC 4379 Detecting Multi-Protocol Label Switched Data Plane Failures*.

Only one downstream mapping object may appear in an echo request. The presence of a downstream mapping object is a request that a downstream mapping object be included in the echo reply.

For unnumbered interfaces, the address type is 2 (IPv4 unnumbered), the downstream IP address is the borrowed IP address of the downstream LSR, and the downstream interface address is the index assigned by the upstream LSR to the interface.

The downstream detailed mapping object is a TLV that can be included in an echo request, as described in *RFC 6424 Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels*.

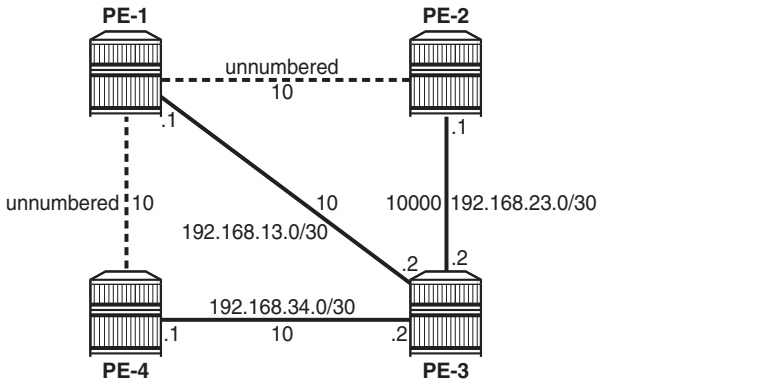
The following output shows the detailed LSP trace for an RSVP LSP from PE-4 to PE-2. Two unnumbered interfaces are used: the first between PE-4 and PE-1 and the second between PE-1 and PE-2. The interface type (iftype) is ipv4Unnumbered.

```
*A:PE-4# oam lsp-trace "LSP-PE-4-PE-2" detail
lsp-trace to LSP-PE-4-PE-2: 0 hops min, 0 hops max, 116 byte packets
1  192.0.2.1  rtt=0.602ms rc=8(DSRtrMatchLabel) rsc=1
     DS 1: ipaddr=0.0.0.0 ifaddr=0 iftype=ipv4Unnumbered MRU=1564
           label[1]=262143 protocol=4(RSVP-TE)
2  192.0.2.2  rtt=1.14ms rc=3(EgressRtr) rsc=1
*A:PE-4#
```

Configuration

The following configuration example is for unnumbered interfaces in RSVP and LDP; see Figure 223. The nodes are 7750 SRs.

Figure 223 **Configuration Example for Unnumbered Interfaces in RSVP and LDP**



All interfaces have a TE metric of 10, while the link between PE-2 and PE-3 has a TE metric of 10000. As such, the preferred path from PE-4 to PE-2 will be over the unnumbered interfaces between PE-4 and PE-1 and between PE-1 and PE-2.

Unnumbered Interfaces

Router interfaces are configured on all nodes, numbered and unnumbered. Initially, the unnumbered interfaces are configured with default settings. The following interfaces are configured on PE-1:

```
*A:PE-1# configure router
        interface "int-PE-1-PE-2"
            port 1/1/1
            unnumbered
        exit
        interface "int-PE-1-PE-3"
            address 192.168.13.1/30
            port 1/1/3
        exit
        interface "int-PE-1-PE-4"
            port 1/1/2
            unnumbered
        exit
        interface "system"
```

```
            address 192.0.2.1/32
        exit
```

There are two unnumbered interfaces: int-PE-1-PE-2 and int-PE-1-PE-4. There is no borrowed IP address configured for the unnumbered interfaces. This implies that the borrowed IP address will be the system address.

For the unnumbered interfaces, the borrowed IP address is indicated between square brackets in the following output:

```
*A:PE-1# show router interface
===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm         Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                    PfxState
-------------------------------------------------------------------------------
int-PE-1-PE-2                    Up          Up/Down     Network 1/1/1
   Unnumbered If[system]                                         n/a
int-PE-1-PE-3                    Up          Up/Down     Network 1/1/3
   192.168.13.1/30                                               n/a
int-PE-1-PE-4                    Up          Up/Down     Network 1/1/2
   Unnumbered If[system]                                         n/a
system                           Up          Up/Down     Network system
   192.0.2.1/32                                                  n/a
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
*A:PE-1#
```

Each interface, numbered or unnumbered, gets an interface index. This interface index can be retrieved as follows:

```
*A:PE-1# show router interface "int-PE-1-PE-2" detail | match "If Index"
If Index         : 2                    Virt. If Index    : 2
Last Oper Chg    : 03/31/2016 08:30:28  Global If Index   : 1
*A:PE-1# show router interface "int-PE-1-PE-3" detail | match "If Index"
If Index         : 3                    Virt. If Index    : 3
Last Oper Chg    : 03/31/2016 08:30:28  Global If Index   : 2
*A:PE-1# show router interface "int-PE-1-PE-4" detail | match "If Index"
If Index         : 4                    Virt. If Index    : 4
Last Oper Chg    : 03/31/2016 08:30:28  Global If Index   : 3
*A:PE-1# show router interface "system" detail | match "If Index"
If Index         : 1                    Virt. If Index    : 1
Last Oper Chg    : 03/31/2016 08:30:28  Global If Index   : 256
*A:PE-1#
```

The unnumbered interface toward PE-2 has ifIndex 2, and the unnumbered interface toward PE-4 has ifIndex 4.

BFD cannot be enabled on unnumbered interfaces. When BFD is configured on an unnumbered interface, the following error is raised:

```
*A:PE-4# configure router
```

```
*A:PE-4>config>router# interface "int-PE-4-PE-1"
*A:PE-4>config>router>if# bfd 100 receive 100 multiplier 3
INFO: BFD #1007 Bfd not allowed
```

An interior gateway protocol (IGP) needs to be configured; in this case, IS-IS is chosen. OSPF could have been used equally well. TE must be enabled for unnumbered interfaces used in RSVP, even when CSPF is disabled. The IS-IS configuration on PE-1 is as follows:

```
*A:PE-1# configure router
            level-capability level-1
            area-id 49.0001
            traffic-engineering
            interface "system"
            exit
            interface "int-PE-1-PE-2"
                interface-type point-to-point
            exit
            interface "int-PE-1-PE-3"
                interface-type point-to-point
            exit
            interface "int-PE-1-PE-4"
                interface-type point-to-point
            exit
```

An unnumbered interface has to be a P2P link.

The TE database contains the router ID and the ifIndex for unnumbered interfaces, as follows:

```
*A:PE-1# show router isis database PE-2.00-00 detail
===============================================================================
Rtr Base ISIS Instance 0 Database (detail)
===============================================================================
Displaying Level 1 database
-------------------------------------------------------------------------------
LSP ID    : PE-2.00-00                                  Level     : L1
Sequence  : 0x2                  Checksum  : 0x5ba3   Lifetime  : 1047
Version   : 1                    Pkt Type  : 18       Pkt Ver   : 1
Attributes: L1                   Max Area  : 3        Alloc Len : 180
SYS ID    : 1920.0000.2002       SysID Len : 6        Used Len  : 180

TLVs :
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols     : IPv4
  IS-Hostname   : PE-2
  Router ID   :
    Router ID   : 192.0.2.2
---snip---
  I/F Addresses :
    I/F Address   : 192.168.23.1
    I/F Address   : 192.0.2.2
  TE IS Nbrs   :
    Nbr   : PE-1.00
```

```
    Default Metric  : 10
    Sub TLV Len     : 10
    LclId    : 2
    RmtId    : 2
  TE IS Nbrs   :
    Nbr   : PE-3.00
    Default Metric  : 10
    Sub TLV Len     : 12
    IF Addr   : 192.168.23.1
    Nbr IP    : 192.168.23.2
  TE IP Reach   :
    Default Metric  : 10
    Control Info:    , prefLen 30
    Prefix   : 192.168.23.0
    Default Metric  : 0
    Control Info:    , prefLen 32
    Prefix   : 192.0.2.2

Level (1) LSP Count : 1
Displaying Level 2 database
-------------------------------------------------------------------------------
Level (2) LSP Count : 0
===============================================================================
*A:PE-1#
```

PE-2 has an unnumbered interface toward PE-1 (Nbr: PE-1.00), with local interface index 2 (LclId: 2) and remote interface index 2 (RmtId: 2). For the numbered interface toward PE-3, the local and remote interface IP addresses are shown (IF Addr and Nbr IP), not the interface index.

Unnumbered Interfaces in RSVP

MPLS and RSVP need to be enabled on the interfaces on the nodes. TE metrics are configured on the MPLS interfaces. For node PE-4, the configuration is as follows:

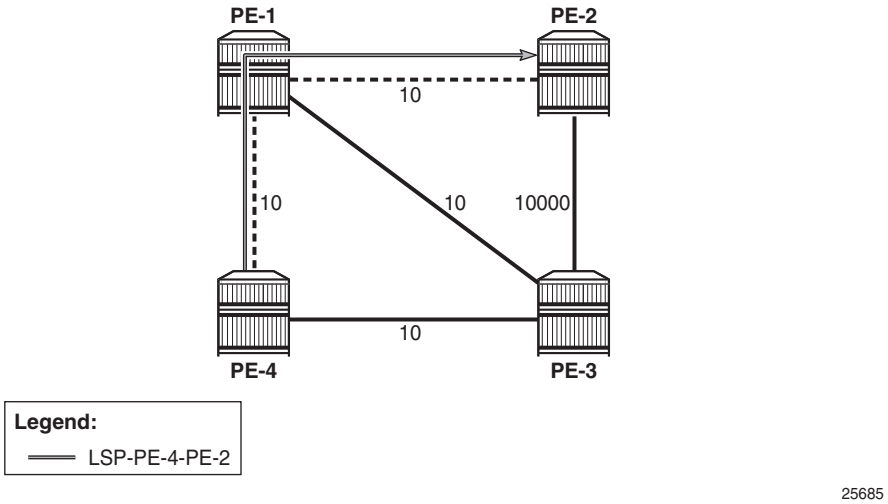
```
*A:PE-4# configure router
        mpls
            interface "system"
            exit
            interface "int-PE-4-PE-1"
                te-metric 10
            exit
            interface "int-PE-4-PE-3"
                te-metric 10
            exit
        exit
        rsvp
            no shutdown
        exit
```

An LSP is configured from PE-4 to PE-2 with CSPF enabled and using the TE metrics, not the IGP metrics. Unnumbered interfaces cannot be configured as hops in a path. A dynamic path "dyn", without any hops, is configured to be used in an LSP from PE-4 to PE-2, as follows:

```
*A:PE-4# configure router
        mpls
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-4-PE-2"
                to 192.0.2.2
                cspf use-te-metric
                primary "dyn"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The LSP from PE-4 to PE-2 will have TE metric 20 when the next hop is PE-1, and TE metric 30 or 10010 when the next hop is PE-3. Figure 224 shows LSP-PE-4-PE-2, which uses only unnumbered interfaces.

Figure 224 **LSP-PE-4-PE-2 on Unnumbered Interfaces**



The following tunnel table shows a next hop 0.0.0.1, which implies that it is an unnumbered interface. The only unnumbered interface at PE-4 is int-PE-4-PE-1. The metric in this tunnel table is 16777215 because the IGP metric is not used.

```
*A:PE-4# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
```

```
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
192.0.2.2/32      rsvp      MPLS  1         7        0.0.0.1        16777215
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-4#
```

The actual and computed hops can be verified, as well as the CSPF metric (TE metric), as follows:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-4-PE-2
Path LSP ID      : 55808
From             : 192.0.2.4            To                  : 192.0.2.2
Admin State      : Up                   Oper State          : Up
Path Name        : dyn                  Path Type           : Primary
Path Admin       : Up                   Path Oper           : Up
Out Interface    : 1/1/1                Out Label           : 262143
---snip---
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.0.2.4, If Index : 2                     Record Label        : N/A
 -> 192.0.2.1, If Index : 4                     Record Label        : 262143
 -> 192.0.2.2, If Index : 2                     Record Label        : 262143
Computed Hops    :
    192.0.2.4, If Index : 2(S)
 -> 192.0.2.1, If Index : 4(S)
 -> 192.0.2.2, If Index : 2(S)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric         : 20
===============================================================================
*A:PE-4#
```

The computed hops are strict hops, as indicated by "(S)". Because the interfaces are unnumbered, the system address and the ifIndex are displayed. The CSPF metric is 20.

Configuring the Borrowed IP Address

The borrowed IP address does not need to be the system address, but the address must exist on the node. When the unnumbered interface is configured with a borrowed IP address that does not exist on the node, the interface goes down. This can be verified by assigning a non-existent address to the unnumbered interface int-PE-1-PE-2, as follows:

```
*A:PE-1# configure router
        interface "int-PE-1-PE-2"
            port 1/1/1
            unnumbered 192.1.2.1
        exit
```

The operational state of this interface goes down, which can be verified as follows:

```
*A:PE-1# show router interface
===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm          Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                    PfxState
-------------------------------------------------------------------------------
int-PE-1-PE-2                    Up           Down/Down   Network 1/1/1
   Unnumbered If[192.1.2.1]                                      n/a
int-PE-1-PE-3                    Up           Up/Down     Network 1/1/3
   192.168.13.1/30                                               n/a
int-PE-1-PE-4                    Up           Up/Down     Network 1/1/2
   Unnumbered If[system]                                         n/a
system                           Up           Up/Down     Network system
   192.0.2.1/32                                                  n/a
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
*A:PE-1#
```

The borrowed IP address is indicated between square brackets. The interface is down because the IP address is not known on PE-1. The down reason code (noIfAddress) can be retrieved as follows:

```
*A:PE-1# show router interface "int-PE-1-PE-2" detail | match "Down Reason Code"
Down Reason Code : noIfAddress
*A:PE-1#
```

The IP address can be configured as a loopback address on PE-1 and assigned to all unnumbered interfaces, as follows:

```
*A:PE-1# configure router
        interface "loopback1"
            address 192.1.2.1/32
            loopback
        exit
        interface "int-PE-1-PE-4"
```

```
            port 1/1/2
            unnumbered 192.1.2.1
        exit
```

When the borrowed IP address is known on node PE-1, the unnumbered interface is operationally up, which can be verified as follows:

```
*A:PE-1# show router interface
===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm         Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                    PfxState
-------------------------------------------------------------------------------
int-PE-1-PE-2                    Up          Up/Down     Network 1/1/1
   Unnumbered If[192.1.2.1]                                      n/a
int-PE-1-PE-3                    Up          Up/Down     Network 1/1/3
   192.168.13.1/30                                               n/a
int-PE-1-PE-4                    Up          Up/Down     Network 1/1/2
   Unnumbered If[192.1.2.1]                                      n/a
loopback1                        Up          Up/Down     Network loopback
   192.1.2.1/32                                                  n/a
system                           Up          Up/Down     Network system
   192.0.2.1/32                                                  n/a
-------------------------------------------------------------------------------
Interfaces : 5
===============================================================================
*A:PE-1#
```

In a similar way, the borrowed IP address on PE-2 is configured as 192.1.2.2 and on PE-4 as 192.1.2.4.

TE Required for Unnumbered Interfaces in RSVP

For unnumbered interfaces, the IGP needs to look up the router ID in the TE database. Therefore, TE must be enabled even if CSPF is disabled. TE is disabled in IS-IS and CSPF is disabled in the LSP on PE-4, as follows:

```
*A:PE-4# configure router isis no traffic-engineering
*A:PE-4# configure router mpls lsp "LSP-PE-4-PE-2" no cspf
```

LSP-PE-4-PE-2 is operationally down with failure code "noRouteToDestination", which can be verified as follows:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4- PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
```

```
    s - Soft Preemption
    S - Strict                          L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-4-PE-2
Path LSP ID       : 55818
From              : 192.0.2.4             To                  : 192.0.2.2
Admin State       : Up                    Oper State          : Down
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Down
Out Interface     : n/a                   Out Label           : n/a
Path Up Time      : 0d 00:00:00           Path Down Time      : 0d 00:00:34
Retry Limit       : 0                     Retry Timer         : 30 sec
Retry Attempt     : 2                     Next Retry In       : 23 sec
BFD Template      : None                  BFD Ping Interval   : 60
BFD Enable        : False

Adspec            : Disabled              Oper Adspec         : N/A
CSPF              : Disabled              Oper CSPF           : N/A
---snip---
Failure Code      : noRouteToDestination
Failure Node      : 192.0.2.4
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    No Hops Specified
Resignal Eligible: False
Last Resignal     : n/a                   CSPF Metric         : N/A
===============================================================================
*A:PE-4#
```

The configuration is restored by enabling TE in IS-IS and CSPF in the LSP context, as follows:

```
*A:PE-4# configure router isis traffic-engineering
*A:PE-4# configure router mpls lsp "LSP-PE-4-PE-2" cspf use-te-metric
```

FRR Facility

FRR facility is enabled on the LSP as follows:

```
*A:PE-4# configure router
        mpls
            lsp "LSP-PE-4-PE-2"
                fast-reroute facility
                exit
            exit
```

The following LSP path detail output shows where an FRR detour is available (@) and in which node a bypass tunnel originates. The letter "n" indicates that a node is protected, as in hop 192.0.2.4. When there is a detour available, but there is no "n", link protection is available, as in hop 192.0.2.1:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-4-PE-2
Path LSP ID       : 55822
From              : 192.0.2.4             To                   : 192.0.2.2
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262141
---snip---
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.0.2.4, If Index : 2 @ n                  Record Label        : N/A
 -> 192.0.2.1, If Index : 4 @                    Record Label        : 262141
 -> 192.0.2.2, If Index : 2                      Record Label        : 262141
Computed Hops     :
    192.0.2.4, If Index : 2(S)
 -> 192.0.2.1, If Index : 4(S)
 -> 192.0.2.2, If Index : 2(S)
Resignal Eligible: False
Last Resignal     : n/a                   CSPF Metric          : 20
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-4#
```

Information about the bypass tunnel originating in PE-4 can be retrieved as follows:

```
*A:PE-4# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-node192.0.2.1-61441
-------------------------------------------------------------------------------
To             : 192.168.23.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262142
Up Time        : 0d 00:02:49         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 10010
```

```
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node    : None                Inter-Area          : False
Computed Hops  :
    192.168.34.2(S)                  Egress Admin Groups : None
 -> 192.168.34.1(S)                  Egress Admin Groups : None
 -> 192.168.23.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.34.2 (192.0.2.4)         Record Label        : N/A
 -> 192.168.34.1 (192.0.2.3)         Record Label        : 262142
 -> 192.168.23.1 (192.0.2.2)         Record Label        : 262139
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-4-PE-2::dyn
From           : 192.0.2.4           To                  : 192.0.2.2
Avoid Node/Hop : 192.0.2.1           Downstream Label    : 262141
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-4#
```

This bypass tunnel, via PE-3 to PE-2, offers node protection for node PE-1. There are no unnumbered interfaces in this path. In a similar way, information about the bypass tunnel to protect the link between PE-1 and PE-2 can be retrieved in PE-1, as follows:

```
*A:PE-1# show router mpls bypass-tunnel protected-lsp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.0.2.2-61441
-------------------------------------------------------------------------------
To             : 192.168.23.1        State               : Up
Out I/F        : 1/1/3               Out Label           : 262143
Up Time        : 0d 00:02:51         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : Dynamic             Bypass Path Cost    : 10010
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.13.1(S)                  Egress Admin Groups : None
 -> 192.168.13.2(S)                  Egress Admin Groups : None
 -> 192.168.23.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.13.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3)         Record Label        : 262143
 -> 192.168.23.1 (192.0.2.2)         Record Label        : 262140
Last Resignal  :
Attempted At   : n/a                 Resignal Reason     : n/a
Resignal Status: n/a                 Reason              : n/a

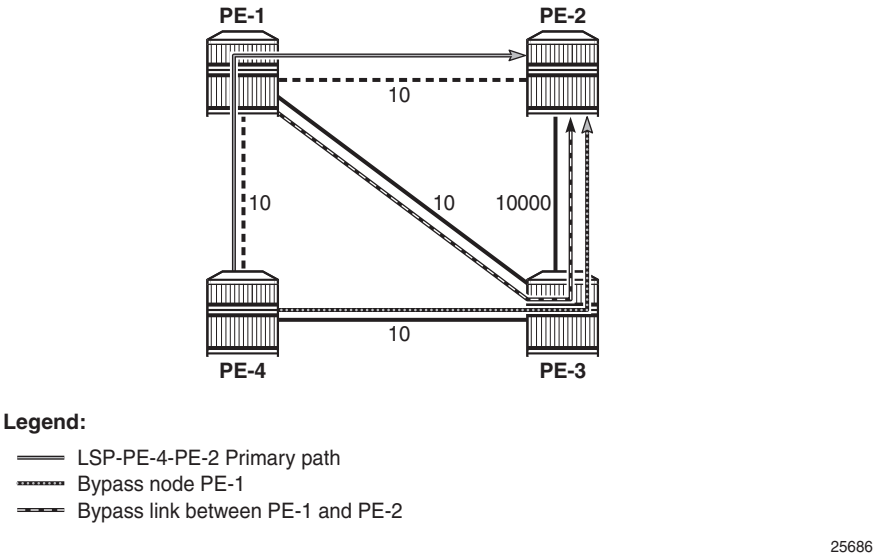
Protected LSPs -
```

```
LSP Name        : LSP-PE-4-PE-2::dyn
From            : 192.0.2.4           To                  : 192.0.2.2
Avoid Node/Hop  : 192.0.2.2           Downstream Label    : 262141
Bandwidth       : 0 Kbps

===============================================================================
*A:PE-1#
```

Figure 225 shows the LSP and the two bypass tunnels: one in PE-4, offering node protection for node PE-1, and another in PE-1, bypassing the link between PE-1 and PE-2.

Figure 225 **LSP and FRR Facility Bypass Tunnels**



For each bypass tunnel, an additional RSVP session is set up. The following output shows that, in PE-4, two LSPs are signaled: the primary LSP and the bypass tunnel for node PE-1.

```
*A:PE-4# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.4       192.0.2.2       1      55822 LSP-PE-4-PE-2::dyn            Up
192.0.2.4       192.168.23.1    61441  2     bypass-node192.0.2.1-61441    Up
-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
*A:PE-4#
```

Similarly, PE-1 has an RSVP session for the primary LSP, but also for the bypass tunnel for the link toward PE-2, as follows:

```
*A:PE-1# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.4       192.0.2.2       1      55822 LSP-PE-4-PE-2::dyn            Up
192.0.2.1       192.168.23.1    61441  2     bypass-link192.0.2.2-61441    Up
-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
*A:PE-1#
```

PE-3 is only used by the bypass tunnels, as follows:

```
*A:PE-3# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.168.23.1    61441  2     bypass-link192.0.2.2-61441    Up
192.0.2.4       192.168.23.1    61441  2     bypass-node192.0.2.1-61441    Up
-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
*A:PE-3#
```

PE-2 terminates the LSP and the bypass tunnels, as follows:

```
*A:PE-2# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.4       192.0.2.2       1      55822 LSP-PE-4-PE-2::dyn            Up
192.0.2.1       192.168.23.1    61441  2     bypass-link192.0.2.2-61441    Up
192.0.2.4       192.168.23.1    61441  2     bypass-node192.0.2.1-61441    Up
-------------------------------------------------------------------------------
Sessions : 3
===============================================================================
*A:PE-2#
```

FRR One-to-One Only Supported on Numbered Interfaces

When FRR one-to-one is enabled on the LSP, the LSP will not use unnumbered interfaces.

FRR is reconfigured on the LSP as follows:

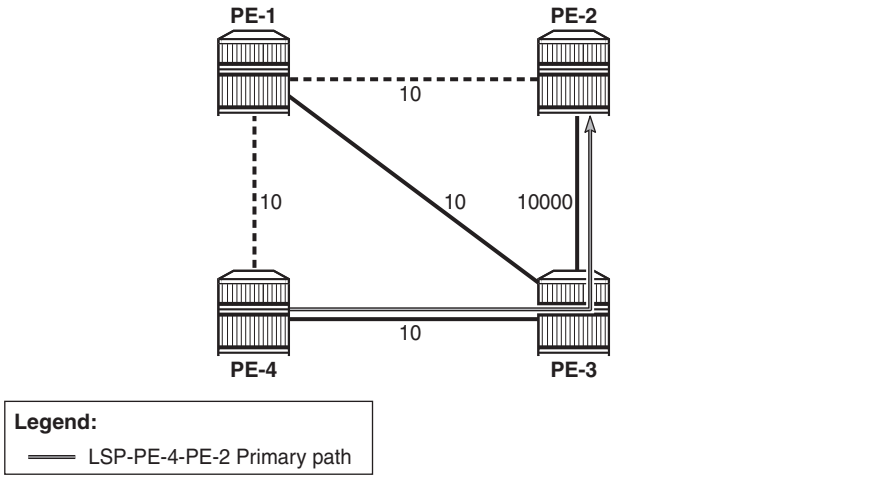
```
*A:PE-4# configure router
        mpls
            lsp "LSP-PE-4-PE-2"
                no fast-reroute
                fast-reroute one-to-one
                exit
            exit
```

The LSP will only come up if it can use numbered interfaces end-to-end. In this case, the LSP will take the path via PE-3 with CSPF metric 10010, as follows:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-4-PE-2
Path LSP ID      : 55828
From             : 192.0.2.4            To                   : 192.0.2.2
Admin State      : Up                   Oper State           : Up
Path Name        : dyn                  Path Type            : Primary
Path Admin       : Up                   Path Oper            : Up
Out Interface    : 1/1/2                Out Label            : 262143
---snip---
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.168.34.2 (192.0.2.4)                    Record Label        : N/A
 -> 192.168.34.1 (192.0.2.3)                    Record Label        : 262143
 -> 192.168.23.1 (192.0.2.2)                    Record Label        : 262142
Computed Hops    :
    192.168.34.2(S)
 -> 192.168.34.1(S)
 -> 192.168.23.1(S)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 10010
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-4#
```

Figure 226 shows the LSP in case of FRR one-to-one. No numbered interfaces are used. Unfortunately, there is no bypass tunnel possible with only numbered interfaces; therefore, there is no protection.

Figure 226 **FRR One-to-One Only Supported on Numbered Interfaces**



If there is no path available with only numbered interfaces, the LSP will remain operationally down with failure code "noCspfRouteToDestination". This can be verified by shutting down port 1/1/2 toward PE-3, as follows:

```
*A:PE-4# configure port 1/1/2 shutdown
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-4-PE-2
Path LSP ID       : 55830
From              : 192.0.2.4             To                  : 192.0.2.2
Admin State       : Up                    Oper State          : Down
Path Name         : dyn                   Path Type           : Primary
Path Admin        : Up                    Path Oper           : Down
Out Interface     : n/a                   Out Label           : n/a
---snip---
Failure Code      : noCspfRouteToDestination
Failure Node      : 192.0.2.4
```

```
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    No Hops Specified
Computed Hops    :
    No Hops Specified
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : N/A
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-4#
```

The port is enabled again and the LSP configuration is restored to FRR facility, as follows:

```
*A:PE-4# configure port 1/1/2 no shutdown
*A:PE-4# configure router
        mpls
            lsp "LSP-PE-4-PE-2"
                no fast-reroute
                fast-reroute facility
                exit
            exit
```

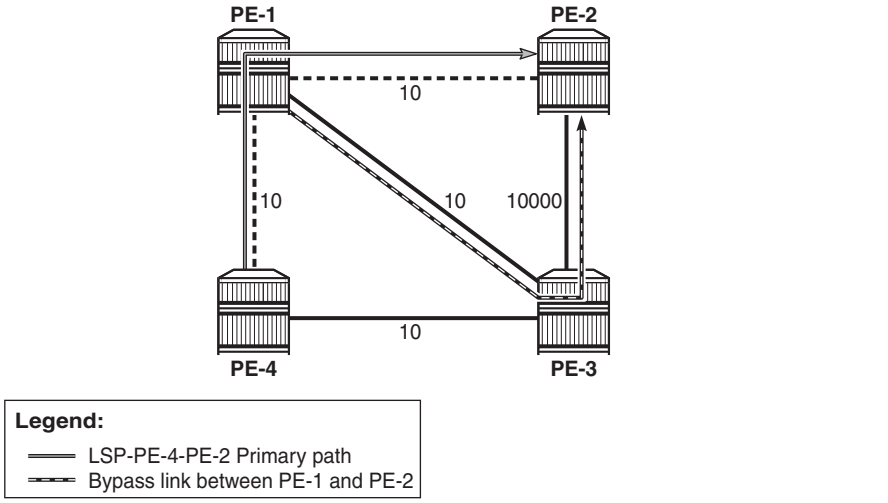
FRR Bypass Not Possible on iLER on Unnumbered Interfaces

FRR facility is not supported on the iLER PE-4 if the bypass is over an unnumbered interface. The interface toward PE-3 is reconfigured as unnumbered, as follows:

```
*A:PE-3# configure router interface "int-PE-3-PE-4" no address
*A:PE-3# configure router interface "int-PE-3-PE-4" unnumbered
*A:PE-3# configure router mpls interface "int-PE-3-PE-4" te-metric 10
*A:PE-4# configure router interface "int-PE-4-PE-3" no address
*A:PE-4# configure router interface "int-PE-4-PE-3" unnumbered
*A:PE-4# configure router mpls interface "int-PE-4-PE-3" te-metric 10
```

When an interface changes from numbered to unnumbered or vice versa, it is no longer known in the MPLS context. Therefore, the interface needs to be added in the MPLS context again. When the interface toward PE-3 is numbered, there is a bypass tunnel in PE-4 to protect node PE-1, but this bypass tunnel cannot be established on an unnumbered interface. The only remaining protection for the LSP is the bypass tunnel originating in PE-1 to protect the link between PE-1 and PE-2, as shown in Figure 227.

Figure 227 **FRR on iLER: No Bypass on Unnumbered Interfaces**



The following output shows that there is only a detour available in PE-1:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-4-PE-2
Path LSP ID       : 55836
From              : 192.0.2.4           To                  : 192.0.2.2
Admin State       : Up                  Oper State          : Up
Path Name         : dyn                 Path Type           : Primary
Path Admin        : Up                  Path Oper           : Up
Out Interface     : 1/1/1               Out Label           : 262143
---snip---
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.0.2.4, If Index : 2                   Record Label        : N/A
 -> 192.0.2.1, If Index : 4 @                 Record Label        : 262143
 -> 192.0.2.2, If Index : 2                   Record Label        : 262143
Computed Hops     :
    192.0.2.4, If Index : 2(S)
 -> 192.0.2.1, If Index : 4(S)
 -> 192.0.2.2, If Index : 2(S)
```

```
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 20
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-4#
```

In iLER PE-4, there is only the LSP tunnel, no bypass tunnel, as follows:

```
*A:PE-4# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.4       192.0.2.2       1      55836 LSP-PE-4-PE-2::dyn            Up
-------------------------------------------------------------------------------
Sessions : 1
===============================================================================
*A:PE-4#
```

The original configuration is restored with numbered interfaces between PE-3 and PE-4, as follows:

```
*A:PE-3# configure router interface "int-PE-3-PE-4" no unnumbered
*A:PE-3# configure router interface "int-PE-3-PE-4" address 192.168.34.1/30
*A:PE-3# configure router mpls interface "int-PE-3-PE-4" te-metric 10
*A:PE-4# configure router interface "int-PE-4-PE-3" no unnumbered
*A:PE-4# configure router interface "int-PE-4-PE-3" address 192.168.34.2/30
*A:PE-4# configure router mpls interface "int-PE-4-PE-3" te-metric 10
```

Admin Groups for Unnumbered Interfaces in RSVP

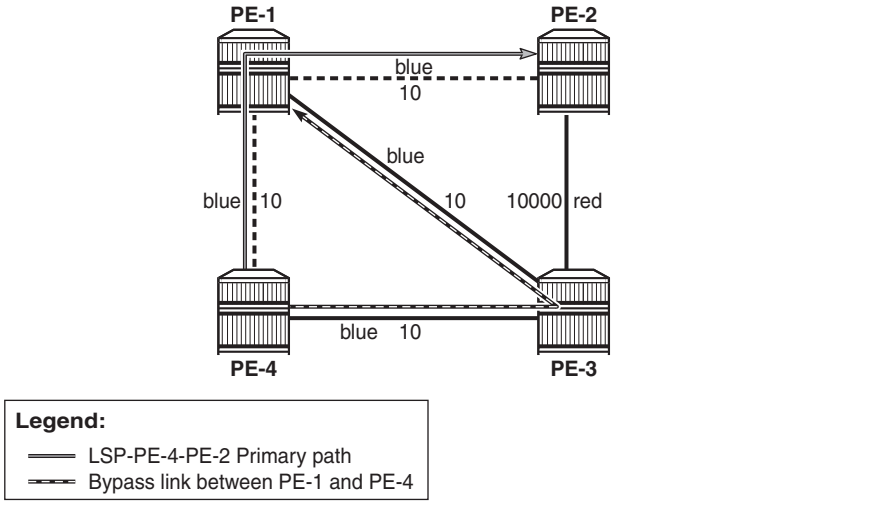
Administrative groups (link-coloring) can be used to calculate a path with the restriction to only include, or exclude, links of a particular admin group (color). Paths can be disjointed from each other, without the need for an explicit hops list. For unnumbered interfaces, an explicit hop list is not an option, but admin groups are.

Two admin groups are configured on all nodes, as follows:

```
*A:PE-4# configure router if-attribute admin-group "red" value 0
*A:PE-4# configure router if-attribute admin-group "blue" value 1
```

Admin group "blue" is assigned to all MPLS interfaces, except for the link between PE-2 and PE-3; see Figure 228.

Figure 228 **FRR Facility and Admin Groups**



The admin groups are assigned to the interfaces in the MPLS context, as follows:

```
*A:PE-2# configure router
        mpls
            interface "int-PE-2-PE-1"
                admin-group "blue"
            exit
            interface "int-PE-2-PE-3"
                admin-group "red"
            exit
```

To ensure that FRR bypass tunnels will only use links belonging to the same admin group, the following is configured on all nodes. It is required on all PLRs.

```
*A:PE-4# configure router mpls admin-group-frr
```

In the LSP context, the admin group "blue" is included. The option propagate-admin-group implies the tunnels must use links belonging to the admin group "blue". This is configured for the LSP tunnel, and for the FRR bypass tunnels, as follows:

```
*A:PE-4# configure router
        mpls
            lsp "LSP-PE-4-PE-2"
                include "blue"
                propagate-admin-group
                fast-reroute facility
                    propagate-admin-group
                exit
            exit
        exit
```

This configuration implies that the link that does not belong to admin group "blue" is excluded, and cannot be used by the LSP nor by a bypass tunnel. Therefore, there will be no bypass tunnel to protect node PE-1 and no bypass tunnel originating in PE-1 protecting the link to PE-2. There will be a bypass tunnel originating in PE-4 to protect the link between PE-4 and PE-1, as shown in Figure 7. The following output shows that there is a detour available for link protection in PE-4:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name         : LSP-PE-4-PE-2
Path LSP ID      : 55842
From             : 192.0.2.4            To                  : 192.0.2.2
Admin State      : Up                   Oper State          : Up
Path Name        : dyn                  Path Type           : Primary
Path Admin       : Up                   Path Oper           : Up
Out Interface    : 1/1/1                Out Label           : 262140
---snip---
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    192.0.2.4, If Index : 2 @                    Record Label        : N/A
 -> 192.0.2.1, If Index : 4                      Record Label        : 262140
 -> 192.0.2.2, If Index : 2                      Record Label        : 262138
Computed Hops    :
    192.0.2.4, If Index : 2(S)
 -> 192.0.2.1, If Index : 4(S)
 -> 192.0.2.2, If Index : 2(S)
---snip---
```

The following output shows that there are two RSVP sessions in PE-4: one for the LSP and one for the bypass tunnel to protect the link between PE-4 and PE-1.

```
*A:PE-4# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                         State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.4       192.0.2.2       1      55842 LSP-PE-4-PE-2::dyn           Up
192.0.2.4       192.168.13.1    61597  8     bypass-link192.0.2.1-61597   Up
-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
```

```
*A:PE-4#
```

The configuration is restored as follows:

```
*A:PE-4# configure router mpls no admin-group-frr
*A:PE-4# configure router
        mpls
            lsp "LSP-PE-4-PE-2"
                no include "blue"
                no propagate-admin-group
                fast-reroute no propagate-admin-group
            exit
```

SRLGs for Unnumbered Interfaces in RSVP

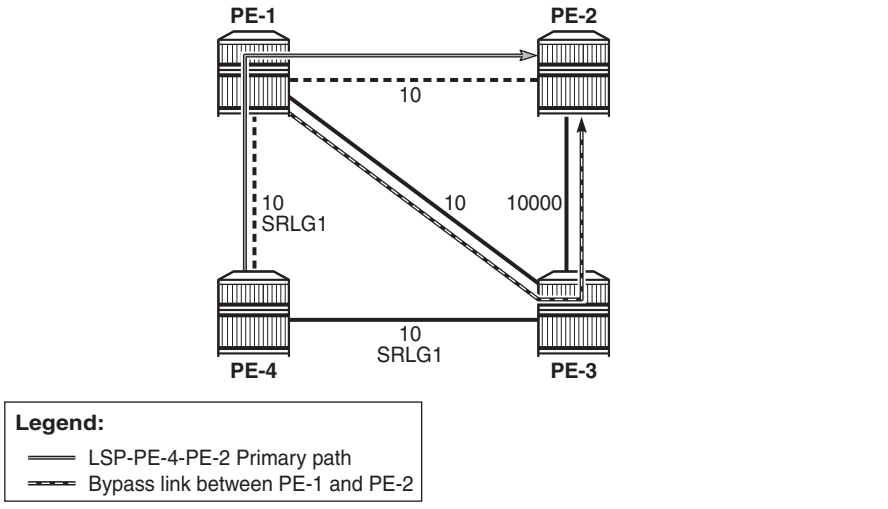
SRLGs allow operators to create automatic secondary LSPs or FRR tunnels that are disjointed from the protected primary tunnel. See chapter Shared Risk Link Groups for RSVP-Based LSP for more information.

One SRLG group is configured on all nodes, as follows:

```
*A:PE-4# configure router if-attribute srlg-group "SRLG1" value 1
```

SRLG "SRLG1" is assigned to the interface between PE-4 and PE-1, and to the interface between PE-4 and PE-3, as shown in Figure 229.

Figure 229 **SRLG-FRR Strict: No Bypass on PE-4**



The SRLG is assigned to the interfaces in the MPLS context, as follows:

```
*A:PE-4# configure router
        mpls
            interface "int-PE-4-PE-1"
                srlg-group "SRLG1"
            exit
            interface "int-PE-4-PE-3"
                srlg-group "SRLG1"
            exit
        exit
```

The configuration on PE-1 and PE-3 is similar.

When SRLG for FRR is enabled in strict mode, CSPF will not establish any detour LSP if there is no path that meets the SRLG constraint. This configuration implies that there is no bypass tunnel in PE-4. The following enables SRLG for FRR in strict mode on all nodes:

```
*A:PE-4# configure router mpls srlg-frr strict
```

Note: Enabling or disabling SRLG for FRR is a system-wide configuration that requires the MPLS routing instance to be manually set to shutdown, then to no shutdown, to activate the change. This can be service affecting. Nokia recommends that the operator include the SRLG in the initial network design and implementation to minimize the traffic loss.

The following output shows that there is only a detour available in PE-1:

```
*A:PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
===============================================================================
MPLS LSP LSP-PE-4-PE-2 Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available              # - Detour In Use
    b - Bandwidth Protected           n - Node Protected
    s - Soft Preemption
    S - Strict                        L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP LSP-PE-4-PE-2 Path dyn
-------------------------------------------------------------------------------
LSP Name          : LSP-PE-4-PE-2
Path LSP ID       : 55850
From              : 192.0.2.4             To                   : 192.0.2.2
Admin State       : Up                    Oper State           : Up
Path Name         : dyn                   Path Type            : Primary
Path Admin        : Up                    Path Oper            : Up
Out Interface     : 1/1/1                 Out Label            : 262143
---snip---
Explicit Hops     :
    No Hops Specified
Actual Hops       :
    192.0.2.4, If Index : 2                     Record Label        : N/A
 -> 192.0.2.1, If Index : 4 @                   Record Label        : 262143
```

```
 -> 192.0.2.2, If Index : 2                      Record Label        : 262143
Computed Hops    :
    192.0.2.4, If Index : 2(S)
 -> 192.0.2.1, If Index : 4(S)
 -> 192.0.2.2, If Index : 2(S)
---snip---
```

The following output shows that PE-1 has two RSVP sessions: one for the LSP and one for the bypass tunnel to protect the link between PE-1 and PE-2.

```
*A:PE-1# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.4       192.0.2.2       1      55850 LSP-PE-4-PE-2::dyn            Up
192.0.2.1       192.168.23.1    61492  16    bypass-link192.0.2.2-61492    Up
-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
*A:PE-1#
```

This was the last example for unnumbered interfaces in RSVP. MPLS and RSVP are disabled in all nodes as follows:

```
*A:PE-1# configure router mpls shutdown
*A:PE-1# configure router rsvp shutdown
```

Unnumbered Interfaces in LDP

Link LDP is configured on PE-4, as follows:

```
*A:PE-4# configure router
        ldp
            interface-parameters
                interface "int-PE-4-PE-1"
                exit
                interface "int-PE-4-PE-3"
                exit
            exit
```

The configuration of link LDP on the other nodes is similar. Link LDP sessions are established, which can be verified as follows:

```
*A:PE-4# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
```

```
-------------------------------------------------------------------------------
192.0.2.1:0         Link      Established   11        12        0d 00:00:13
192.0.2.3:0         Link      Established   11        12        0d 00:00:13
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
---snip---
```

The peer LDP ID is the LSR ID, which is the system address, by default. The IP address configured on the unnumbered interface (such as 192.1.2.1) is not used. The following tunnel table shows a distinction between numbered and unnumbered interfaces in the next hop:

```
*A:PE-4# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      ldp       MPLS  65537     9        0.0.0.1         10
192.0.2.2/32      ldp       MPLS  65538     9        0.0.0.1         20
192.0.2.3/32      ldp       MPLS  65539     9        192.168.34.1    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-4#
```

For destination 192.0.2.1 or 192.0.2.2, the unnumbered interface toward PE-1 is taken. The next hop is represented by 0.0.0.1. When a node has several unnumbered interfaces, the corresponding next hop values are different, as follows, for PE-1:

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.2/32      ldp       MPLS  65537     9        0.0.0.1         10
192.0.2.3/32      ldp       MPLS  65538     9        192.168.13.2    10
192.0.2.4/32      ldp       MPLS  65539     9        0.0.0.3         10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

The LDP active prefix bindings only contain system addresses, no other loopback prefixes, as follows:

```
*A:PE-4# show router ldp bindings active prefixes
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
```

```
            (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262142
Unnumbered                              1/1/1

192.0.2.1/32                            Swap            262142    262142
Unnumbered                              1/1/1

192.0.2.2/32                            Push              --      262143
Unnumbered                              1/1/1

192.0.2.2/32                            Swap            262141    262143
Unnumbered                              1/1/1

192.0.2.3/32                            Push              --      262143
192.168.34.1                            1/1/2

192.0.2.3/32                            Swap            262140    262143
192.168.34.1                            1/1/2

192.0.2.4/32                            Pop             262143      --
  --                                      --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 7
===============================================================================
---snip---
```

For prefixes 192.0.2.1 and 192.0.2.2, the egress next hop is unnumbered. The egress interface for both is 1/1/1. There is no ifIndex. Local addresses are advertised in LDP address messages, such as:

```
6 2016/03/31 09:39:16.85 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address packet (msgId 5) to 192.0.2.1:0
Protocol version = 1
Address Family = 1  Number of addresses = 3
Address 1 = 192.0.2.4
Address 2 = 192.1.2.4
Address 3 = 192.168.34.2
"
5 2016/03/31 09:39:16.78 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address packet (msgId 5) from 192.0.2.1:0
Protocol version = 1
```

```
Address Family = 1  Number of addresses = 3
Address 1 = 192.0.2.1
Address 2 = 192.1.2.1
Address 3 = 192.168.13.1
"
```

The list of advertised local addresses includes the loopback addresses "loopback1": 192.1.2.1 and 192.1.2.4. These loopback addresses did not occur in the preceding lists of LDP sessions or LDP bindings, but they occur in the LDP session local addresses, as follows:

```
*A:PE-4# show router ldp session local-addresses
===============================================================================
LDP Session Local-Addresses
===============================================================================
-------------------------------------------------------------------------------
Session with Peer 192.0.2.1:0,
             Local 192.0.2.4:0
-------------------------------------------------------------------------------
IPv4 Sent Addresses:
                192.0.2.4       192.1.2.4       192.168.34.2
IPv6 Sent Addresses:
                None
IPv4 Recv Addresses:
                192.0.2.1       192.1.2.1       192.168.13.1
IPv6 Recv Addresses:
                None
-------------------------------------------------------------------------------
---snip---
```

If there were only unnumbered addresses and no additional loopback addresses, only the system address and other loopback addresses would be sent or received. The interface addresses in the list of local addresses are from numbered interfaces.

Configuring the Local LSR ID

To use the loopback address "loopback1" in the LDP sessions, the local LSR ID is configured as follows:

```
A:PE-4# configure router
         ldp
            interface-parameters
                interface "int-PE-4-PE-1"
                    ipv4
                        local-lsr-id interface-name "loopback1"
                        transport-address interface
                        no shutdown
                    exit
                exit
            exit
```

The transport address is the system address, by default, but here it is changed to the address of "loopback1", which is 192.1.2.4. The configuration is similar on PE-1. On PE-2, the system addresses are kept and no additional configuration is required.

LDP hello messages are sent from the transport address to 224.0.0.2 to establish hello adjacencies. The transport address is 192.1.2.4 for the unnumbered interface toward PE-1, and 192.0.2.4 (system address) for the numbered interface toward PE-3. LDP hello adjacencies are verified as follows:

```
*A:PE-4# show router ldp discovery
===============================================================================
LDP IPv4 Hello Adjacencies
===============================================================================
Interface Name                   Local Addr                           State
AdjType                          Peer Addr
-------------------------------------------------------------------------------
int-PE-4-PE-1                    192.1.2.4:0                          Estab
link                             192.1.2.1:0

int-PE-4-PE-3                    192.0.2.4:0                          Estab
link                             192.0.2.3:0

-------------------------------------------------------------------------------
No. of IPv4 Hello Adjacencies: 2
===============================================================================
---snip---
```

The LDP hello adjacencies are established, but the LDP session on the unnumbered interface is non-existent, as follows:

```
*A:PE-4# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.3:0         Link      Established   85        87        0d 00:03:29
192.1.2.1:0         Link      Nonexistent   19        20        0d 00:01:11
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
---snip---
```

The LDP session is non-existent because the prefix 192.1.2.1/32 is not in the routing table and the LDP session is to be established between 192.1.2.4 and 192.1.2.1. The following export policy is configured and added in the IS-IS context on the nodes:

```
*A:PE-4# configure router
        policy-options
            begin
            policy-statement "export_ISIS"
                entry 10
                    from
                        protocol direct
```

```
                    exit
                    action accept
                    exit
                exit
                default-action reject
            exit
            commit
        exit
    exit
*A:PE-4# configure router isis export "export_ISIS"
```

The loopback addresses are now exported in IS-IS. When the loopback addresses are in the routing table, the LDP session is established, as follows:

```
*A:PE-4# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.3:0         Link      Established   146       146       0d 00:06:08
192.1.2.1:0         Link      Established   71        72        0d 00:03:50
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
---snip---
```

The local LSR ID can also be configured for targeted LDP sessions, as follows:

```
*A:PE-4# configure router
        ldp
            targeted-session
                peer 192.1.2.1
                    local-lsr-id "loopback1"
                    no shutdown
                exit
            exit
```

The configuration on PE-1 is similar. The LDP adjacency type is now both link and targeted for peer 192.1.2.1, as follows:

```
*A:PE-4# show router ldp session
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.3:0         Link      Established   257       257       0d 00:11:07
192.1.2.1:0         Both      Established   185       186       0d 00:08:49
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
---snip---
```

Even though the LDP sessions are established, there is no LDP prefix binding for the loopback address, as follows:

```
*A:PE-4# show router ldp bindings active prefixes
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262142
Unnumbered                              1/1/1

192.0.2.1/32                            Swap            262142    262142
Unnumbered                              1/1/1

192.0.2.2/32                            Push              --      262143
Unnumbered                              1/1/1

192.0.2.2/32                            Swap            262141    262143
Unnumbered                              1/1/1

192.0.2.3/32                            Push              --      262143
192.168.34.1                            1/1/2

192.0.2.3/32                            Swap            262140    262143
192.168.34.1                            1/1/2

192.0.2.4/32                            Pop             262143      --
  --                                      --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 7
===============================================================================
```

There is no label mapping for prefix 192.1.2.1/32. SDPs are created toward all other nodes, as follows:

```
*A:PE-4# configure service
        sdp 411 mpls create
            far-end 192.0.2.1
            ldp
            no shutdown
        exit
        sdp 412 mpls create
            far-end 192.1.2.1
```

```
            ldp
            no shutdown
        exit
        sdp 421 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 431 mpls create
            far-end 192.0.2.3
            ldp
            no shutdown
        exit
```

The following output shows that, between PE-4 and PE-1, there are two LDP SDPs: one using the system address and another using the loopback address 192.1.2.x. The configuration on the other nodes is similar. All SDPs that have a system address as the far end are operationally up, while the SDP toward 192.1.2.1 is down:

```
*A:PE-4# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
411    0       1556    192.0.2.1        Up   Up          MPLS    L     TLDP
412    0       0       192.1.2.1        Up   Down        MPLS    L     TLDP
421    0       1556    192.0.2.2        Up   Up          MPLS    L     TLDP
431    0       1556    192.0.2.3        Up   Up          MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 4
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
===============================================================================
*A:PE-4#
```

The SDP toward 192.1.2.1 is down because the transport tunnel is down, as follows:

```
*A:PE-4# show service sdp detail
===============================================================================
Services: Service Destination Points Details
===============================================================================
---snip---
-------------------------------------------------------------------------------
 Sdp Id 412  -192.1.2.1
-------------------------------------------------------------------------------
Description           : (Not Specified)
SDP Id                : 412                    SDP Source         : manual
Admin Path MTU        : 0                      Oper Path MTU      : 0
Delivery              : MPLS
Far End               : 192.1.2.1
Tunnel Far End        : 192.1.2.1              LSP Types          : LDP

Admin State           : Up                     Oper State         : Down
Signaling             : TLDP                   Metric             : 0
```

```
Acct. Pol            : None                  Collect Stats      : Disabled
Last Status Change   : 03/31/2016 08:00:52   Adv. MTU Over.     : No
Last Mgmt Change     : 03/31/2016 09:41:55   VLAN VC Etype      : 0x8100
Bw BookingFactor     : 100                   PBB Etype          : 0x88e7
Oper Max BW(Kbps)    : 0                     Avail BW(Kbps)     : 0
Net-Domain           : default               Egr Interfaces     : n/a
Flags                : TranspTunnDown
---snip---
```

The solution is to manually add an LDP prefix binding.

Configuring the FEC Originate

The labels to be used for a manually created LDP prefix binding must be chosen from the range for static labels: from 32 to 18431. This range can be retrieved as follows:

```
*A:PE-1# show router mpls-labels label-range
===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18399       18400
Dynamic         18432       524287      0           505850      505856
    Seg-Route   0           0           -           0           505856
===============================================================================
*A:PE-1#
```

To manually add an LDP prefix binding for the loopback prefixes, configure the following:

On PE-4:

```
*A:PE-4# configure router
        ldp
            fec-originate 192.1.2.1/32 next-hop 192.0.2.4
                          interface "int-PE-4-PE-1" swap-label 101
            fec-originate 192.1.2.4/32 pop advertised-label 104
        exit
```

On PE-1:

```
*A:PE-1# configure router
        ldp
            fec-originate 192.1.2.1/32 pop advertised-label 101
            fec-originate 192.1.2.4/32 next-hop 192.0.2.1
                          interface "int-PE-1-PE-4" swap-label 104
        exit
```

This configuration for unnumbered interfaces includes the interface name, such as int-PE-4-PE-1. This parameter is optional for numbered interfaces.

If the label is chosen from the dynamic range instead of the static range, an error is raised for the pop operation, as follows:

```
*A:PE-1# configure router ldp fec-originate 192.1.2.4/32 pop advertised-label 100001
                                                                                ^
Error: Invalid parameter. Label value not in allowed range
A:PE-1#
```

For interoperability, no error is raised for the swap operation.

As a result, three active LDP bindings are added: one pop operation for the local loopback prefix, and a swap and a push operation for the remote loopback prefix, as follows:

```
*A:PE-4# show router ldp bindings active prefixes
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push              --      262142
Unnumbered                              1/1/1

192.0.2.1/32                            Swap            262142    262142
Unnumbered                              1/1/1

192.0.2.2/32                            Push              --      262143
Unnumbered                              1/1/1

192.0.2.2/32                            Swap            262141    262143
Unnumbered                              1/1/1

192.0.2.3/32                            Push              --      262143
192.168.34.1                            1/1/2

192.0.2.3/32                            Swap            262140    262143
192.168.34.1                            1/1/2

192.0.2.4/32                            Pop             262143      --
  --                                      --

192.1.2.1/32                            Push              --      101
Unnumbered                              1/1/1

192.1.2.1/32                            Swap            262139    101
```

```
Unnumbered                                      1/1/1

192.1.2.4/32(S)                                 Pop             104         --
  --                                              --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 10
===============================================================================
---snip---
```

The following output shows that the SDPs are all operationally up, including the one toward the loopback address:

```
*A:PE-4# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr          Del     LSP   Sig
-------------------------------------------------------------------------------
411    0       1556    192.0.2.1        Up   Up           MPLS    L     TLDP
412    0       1556    192.1.2.1        Up   Up           MPLS    L     TLDP
421    0       1556    192.0.2.2        Up   Up           MPLS    L     TLDP
431    0       1556    192.0.2.3        Up   Up           MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 4
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
===============================================================================
*A:PE-4#
```

LDP FRR Loop-Free Alternate on Unnumbered Interfaces

LDP FRR loop-free alternate (LFA) is supported on unnumbered interfaces and on numbered interfaces. For information about LDP FRR LFA, see chapter MPLS LDP FRR using ISIS as IGP. LDP FRR LFA can be configured as follows:

```
*A:PE-4# configure router isis loopfree-alternate
*A:PE-4# configure router ldp fast-reroute
*A:PE-4# configure router ip-fast-reroute
```

Enabling FRR LFA is a local decision. In this configuration, it is configured on all nodes. The LFA coverage can be retrieved as follows:

```
*A:PE-4# show router isis lfa-coverage
===============================================================================
Rtr Base ISIS Instance 0 LFA Coverage
===============================================================================
Topology         Level   Node           IPv4                IPv6
-------------------------------------------------------------------------------
IPV4 Unicast     L1      3/3(100%)      7/7(100%)           0/0(0%)
IPV6 Unicast     L1      0/0(0%)        0/0(0%)             0/0(0%)
```

```
IPV4 Multicast   L1      0/0(0%)         0/0(0%)             0/0(0%)
IPV6 Multicast   L1      0/0(0%)         0/0(0%)             0/0(0%)
IPV4 Unicast     L2      0/0(0%)         7/7(100%)           0/0(0%)
IPV6 Unicast     L2      0/0(0%)         0/0(0%)             0/0(0%)
IPV4 Multicast   L2      0/0(0%)         0/0(0%)             0/0(0%)
IPV6 Multicast   L2      0/0(0%)         0/0(0%)             0/0(0%)
===============================================================================
*A:PE-4#
```

There is protection for the three other nodes and for all remote prefixes in the routing table, which can be verified as follows:

```
*A:PE-4# show router route-table alternative
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
      Alt-NextHop                                                  Alt-
                                                                   Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h57m41s  15
       int-PE-4-PE-1                                                10
       192.168.34.1 (LFA)                                           20
192.0.2.2/32                                    Remote  ISIS      00h53m52s  15
       int-PE-4-PE-1                                                20
       192.168.34.1 (LFA)                                           30
192.0.2.3/32                                    Remote  ISIS      00h34m39s  15
       192.168.34.1                                                 10
       int-PE-4-PE-1 (LFA)                                          20
192.0.2.4/32                                    Local   Local     00h58m17s  0
       system                                                       0
192.1.2.1/32                                    Remote  ISIS      00h19m17s  15
       int-PE-4-PE-1                                                10
       192.168.34.1 (LFA)                                           20
192.1.2.2/32                                    Remote  ISIS      00h19m17s  15
       int-PE-4-PE-1                                                20
       192.168.34.1 (LFA)                                           30
192.1.2.4/32                                    Local   Local     00h53m39s  0
       loopback1                                                    0
192.168.13.0/30                                 Remote  ISIS      00h34m39s  15
       int-PE-4-PE-1                                                20
       192.168.34.1 (LFA)                                           30
192.168.23.0/30                                 Remote  ISIS      00h34m39s  15
       192.168.34.1                                                 10010
       int-PE-4-PE-1 (LFA)                                          10020
192.168.34.0/30                                 Local   Local     00h34m40s  0
       int-PE-4-PE-3                                                0
-------------------------------------------------------------------------------
No. of Routes: 10
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
*A:PE-4#
```

For unnumbered interfaces, the interface name is shown (int-PE-4-PE-1); for numbered interfaces, the next hop IP address is shown (192.168.34.1). The LFA type is link protection for the three nodes, as follows:

```
*A:PE-4# show router isis topology lfa detail
===============================================================================
Rtr Base ISIS Instance 0 Topology Table
===============================================================================
-------------------------------------------------------------------------------
IS-IS IP paths (MT-ID 0),   Level 1
-------------------------------------------------------------------------------
Node      : PE-1.00                         Metric      : 10
Interface : int-PE-4-PE-1                   SNPA        : none
Nexthop   : PE-1

LFA intf  : int-PE-4-PE-3                   LFA Metric  : 20
LFA nh    : PE-3                            LFA type    : linkProtection

Node      : PE-2.00                         Metric      : 20
Interface : int-PE-4-PE-1                   SNPA        : none
Nexthop   : PE-1

LFA intf  : int-PE-4-PE-3                   LFA Metric  : 30
LFA nh    : PE-3                            LFA type    : linkProtection

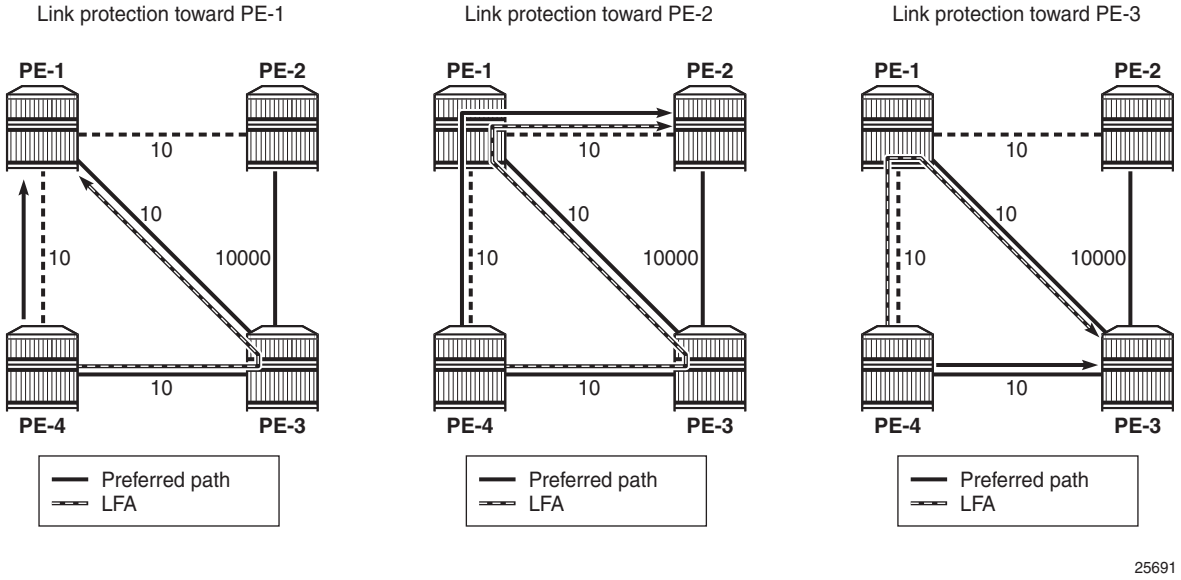
Node      : PE-3.00                         Metric      : 10
Interface : int-PE-4-PE-3                   SNPA        : none
Nexthop   : PE-3

LFA intf  : int-PE-4-PE-1                   LFA Metric  : 20
LFA nh    : PE-1                            LFA type    : linkProtection

===============================================================================
*A:PE-4#
```

The LFA protection is shown in Figure 230.

Figure 230 **LDP FRR LFA Link Protection on PE-4**



The LDP bindings for FRR LFA indicate alternate ("BU") in the list, as follows:

```
*A:PE-4# show router ldp bindings prefixes
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP IPv4 Prefix Bindings
===============================================================================
Prefix                                  IngLbl                  EgrLbl
Peer                                    EgrIntf/LspId
EgrNextHop
-------------------------------------------------------------------------------
192.0.2.1/32                            262142U                 262141BU
192.0.2.3:0                             1/1/2
192.168.34.1

192.0.2.1/32                              --                    262142
192.1.2.1:0                             1/1/1
Unnumbered

192.0.2.2/32                            262141U                 262142BU
192.0.2.3:0                             1/1/2
192.168.34.1

192.0.2.2/32                            262141N                 262143
192.1.2.1:0                             1/1/1
Unnumbered
```

```
192.0.2.3/32                                --                      262143
192.0.2.3:0                               1/1/2
192.168.34.1

192.0.2.3/32                              262140U                 262141BU
192.1.2.1:0                               1/1/1
Unnumbered

192.0.2.4/32                              262143U                   --
192.0.2.3:0                                 --
  --

192.0.2.4/32                              262143U                   --
192.1.2.1:0                                 --
  --

192.1.2.1/32                              262139U                 262139BU
192.0.2.3:0                               1/1/2
192.168.34.1

192.1.2.1/32                                --                      101
192.1.2.1:0                               1/1/1
Unnumbered

192.1.2.4/32                              104U                      --
192.0.2.3:0                                 --
  --

192.1.2.4/32                              104U                      --
192.1.2.1:0                                 --
  --

-------------------------------------------------------------------------------
No. of IPv4 Prefix Bindings: 12
===============================================================================
---snip---
```

Conclusion

Unnumbered interfaces were initially supported for SONET/ATM/FR, and later also on Ethernet access ports. IS-IS adjacencies and OSPF neighbors can be established on unnumbered interfaces. This chapter showed that unnumbered interfaces can be added to RSVP or LDP. Most features that are supported on numbered interfaces are also supported on unnumbered interfaces.

Services Overview

In This Section

This section provides configuration information for the following topics:

- G.8032 Ethernet Ring Protection Multiple Ring Topology
- G.8032 Ethernet Ring Protection Single Ring Topology

G.8032 Ethernet Ring Protection Multiple Ring Topology

This chapter provides information about G.8032 Ethernet ring protection multiple ring topologies.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

Initially, this chapter was written for release SR OS 12.0.R5, but the CLI in this edition is based on release 14.0.R2.

Overview

G.8032 Ethernet ring protection is supported for data service SAPs within a regular VPLS service, a PBB VPLS (I/B-component) or a routed VPLS (R-VPLS). G.8032 is one of the fastest protection schemes for Ethernet networks. This chapter describes the advanced topic of Multiple Ring Control, sometimes referred to as multi-chassis protection, with access rings being the most common form of multiple ring topologies. Single Rings are covered in G.8032 Ethernet Ring Protection Single Ring Topology.This chapter will use a VPLS service to illustrate the configuration of G.8032. For very large ring topologies, Provider Backbone Bridging (PBB) can also be used, but it is not configured in this chapter.

ITU-T G.8032v2 specifies protection switching mechanisms and a protocol for Ethernet layer network (ETH) Ethernet rings. Ethernet rings can provide wide-area multipoint connectivity more economically due to their reduced number of links. The mechanisms and protocol defined in ITU-T G.8032v2 are highly reliable with stable protection and never form loops, which would negatively affect network operation

and service availability. Each ring node is connected to adjacent nodes participating in the same ring using two independent paths, which use ring links (configured on ports or link aggregation groups (LAGs)). A ring link is bounded by two adjacent nodes and a port for a ring link is called a ring port. The minimum number of nodes on a ring is two.

The fundamentals of this ring protection switching architecture are:

- the principle of loop avoidance and
- the utilization of learning, forwarding, and address table mechanisms defined in the ITU-T G.8032v2 Ethernet flow Forwarding Function (ETH_FF) (Control plane).

Loop avoidance in the ring is achieved by guaranteeing that, at any time, traffic may flow on all but one of the ring links. This particular link is called the ring protection link (RPL) and under normal conditions this link is blocked, so it is not used for traffic. One designated node, the RPL owner, is responsible to block traffic over the one designated RPL. Under a ring failure condition, the RPL owner is responsible for unblocking the RPL, allowing the RPL to be used for traffic. The protocol ensures that even without an RPL owner defined, one link will be blocked and it operates as a **break before make** protocol, specifically the protocol guarantees that no link is restored until a different link in the ring is blocked. The other side of the RPL is configured as an RPL neighbor. An RPL neighbor blocks traffic on the link.

The event of a ring link or ring node failure results in protection switching of the traffic. This is achieved under the control of the ETH_FF functions on all ring nodes. A ring automatic protection switching (R-APS) protocol is used to coordinate the protection actions over the ring. The protection switching mechanisms and protocol supports a multi-ring/ladder network that consists of connected Ethernet rings.

Ring Protection Mechanism

The ring protection protocol is based on the following building blocks:

- Ring status change on failure
  - Idle -> Link failure -> Protection -> Recovery -> Idle
- Ring control state changes
  - Idle -> Protection -> Manual Switch -> Forced Switch -> Pending
- Re-use existing ETH OAM
  - Monitoring: ETH continuity check messages (CCM)
  - Failure notification: Y.1731 signal failure

- Forwarding database MAC flush on ring status change
- RPL (Ring Protection Link)
  - Defines blocked link in idle status

When sub-rings are used, they can either connect to a major ring (which is configured in the exact same way as a single ring) or another sub-ring, or to a VPLS service. When connected to a major/sub ring, there is the option to extend the sub-ring control service through the major ring or not. This gives the following three options for sub-ring connectivity:

a. **Sub-ring to a major/sub ring with a virtual channel** — In this case, a data service on the major/sub ring is created which is used to forward the R-APS messages for the sub-ring over the major/sub ring, between the interconnection points of the sub-ring to the major/sub ring. This allows the sub-ring to operate as a fully connected ring and is mandatory if the sub-ring connects two major/sub rings since the virtual channel is the only mechanism that the sub-rings can use to exchange control messages. It also could improve failover times if the sub-ring was large as it provides two paths on the sub-ring interconnection nodes to propagate the fault indication around the sub-ring, whereas without a virtual channel the fault indication may need to traverse the entire sub-ring. Each sub-ring requires its own data service on the major/sub ring for the virtual channel.

b. **Sub-ring to a major/sub ring without a virtual channel** — In this case the sub-ring is not fully connected and does not require any resources on the major/sub ring. This option requires that the R-APS messages are not blocked on the sub-ring over its RPL.

c. **Sub-ring to a VPLS service** — This is similar to the preceding option, but it uses a VPLS service instead of a major ring. In this option, sub-ring failures can initiate the sending of an LDP MAC flush message into the VPLS service when spoke or MPLS mesh SDPs are used in the VPLS service.

Eth-Ring Terminology

The implementation of Ethernet Ring (eth-ring) on SROS uses a VPLS as the construct for a ring flow function (one for ETH_FF (solely for control) and one for each service_FF) and SAPs (on ports or LAGs) as ring links. The control VPLS must be a regular VPLS, but the data VPLS can be a regular VPLS, a PBB (B/I-) VPLS or a routed VPLS. The state of the data service SAPs is inherited from the state of the control service SAPs. Table 17 displays a comparison between the ITU-T and SR/ESS terminologies.

Table 17 **Terminology Comparison**

ITU-T G.8032v2 Terminology	SROS Terminology
ETH_FF	control vpls
Service_FF	data vpls
East Ring Link	path a
West Ring Link	path b
RPL owner	rpl-node owner
RPL Link	path {a|b} rpl-end
MEP	control-mep
ERP control process	eth-ring instance or ring-id
Major Ring	eth-ring
Sub-ring	eth-ring sub-ring
Ring node	Ring Node PE
Ring-ID	Not used; fixed at 1 per G.8032v2

There are various ways that multiple rings can be interconnected and the possible topologies may be large. Customers typically have two forms of networks; access ring edge networks or larger multiple ring networks. Both topologies require ring interconnection.

Figure 231 **G.8032 Major Ring and Sub-Ring**

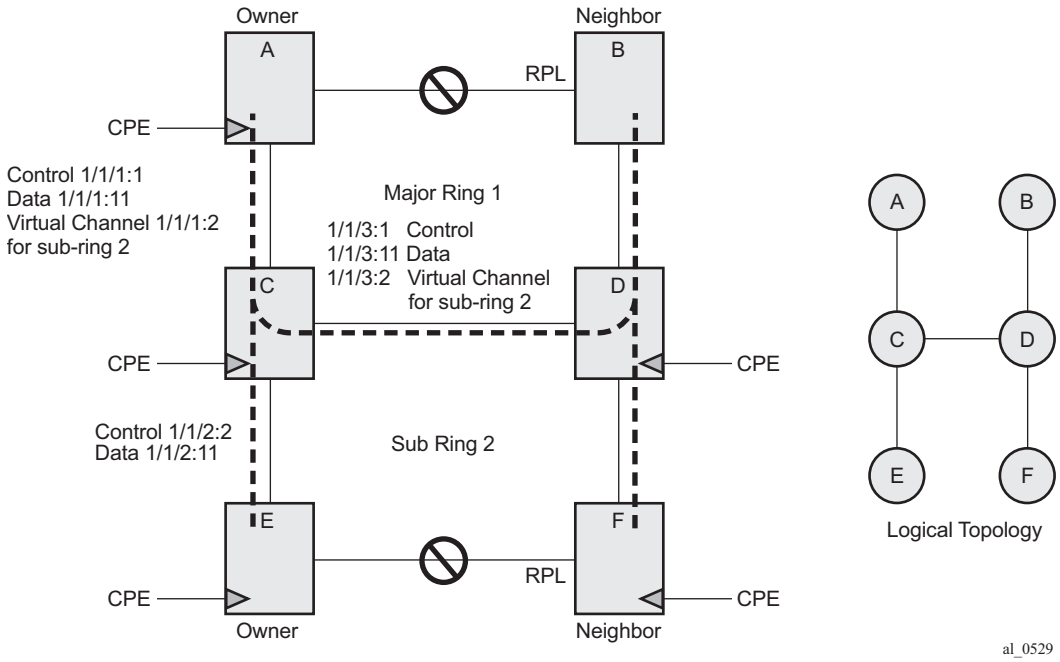


Figure 231 shows a ring of six nodes, with a major ring (regular Ethernet ring) on the top four nodes and a sub-ring on the bottom. A major ring is a fully connected ring. A sub-ring is a partial ring that depends on a major ring or a VPLS topology for part of the ring interconnect. Two major rings can be connected by a single sub-ring or a sub-ring can support other sub-rings.

In the major ring (on nodes A, B, C and D), one path of the RPL owner is designated to be the RPL and the respective SAPs will be blocked in order to prevent a loop. The choice of where to put the RPL is up to the network administrator and can be different for different control instances of the ring allowing an RPL to be used for some other ring's traffic. In the sub-ring, one path is designated as the RPL and will be blocked. Both the major ring and the sub-ring have their own RPL. The sub-ring interconnects to the major ring on nodes C and D and has a virtual channel on the major ring. SROS supports both virtual channel and non-virtual channel rings. Schematics of the physical and logical topologies are also shown in Figure 231.

The G.8032 protocol defines a Ring-ID value (1-255). The SR OS implementation only uses a Ring-ID value of 1, which complies with G.8032v2. The configuration on a node uses a ring instance with a number but all rings use a Ring-ID of 1. This ring instance number is purely local and does not have to match on other ring nodes. Only the VLAN ID must match between SROS ring nodes. For consistency in this example, VPLS instances and Ethernet ring instances are shown as matching for the same ring.

An RPL owner and RPL neighbor are configured for both the major ring and sub-ring. The path and associated link will be the RPL when the ring is fully operational and will be blocked by the RPL owner whenever there is no fault on other ring links. Each ring RPL is independent. If a different ring link fails, then the RPL will be unblocked by the RPL owner. The link shared between a sub-ring and the major ring is completely controlled by the major ring as if the sub-ring were not there. Each ring can completely protect one fault within its ring. When the failed link recovers, it will initially be blocked by one of its adjacent nodes. The adjacent node sends an R-APS message across the ring to indicate the error is cleared and after a configurable time, if reversion is enabled, the RPL will revert to being blocked with all other links unblocked. This ensures that the ring topology when fully operational is predictable.

If a specific RPL owner is not configured (not recommended by G.8032 specification), then the last link to become active will be blocked and the ring will remain in this state until another link fails. This operation makes the selection of the blocked link non-deterministic and is not recommended.

The protection protocol uses a specific control VLAN, with the associated data VLANs taking their forwarding state from the control VLAN. The control VLAN cannot carry data.

Load Balancing with Multiple Ring Instances

Each control ring is independent of the other control rings on the same topology. Therefore since the RPL is used by one control ring, it is often desirable to set up a second control ring that uses a different link as RPL. This spreads out traffic in the topology, but if there is a link failure in the ring, all traffic will be on the remaining links. In the following examples, only a single control ring instance is configured. Other control and data rings could be configured if desired.

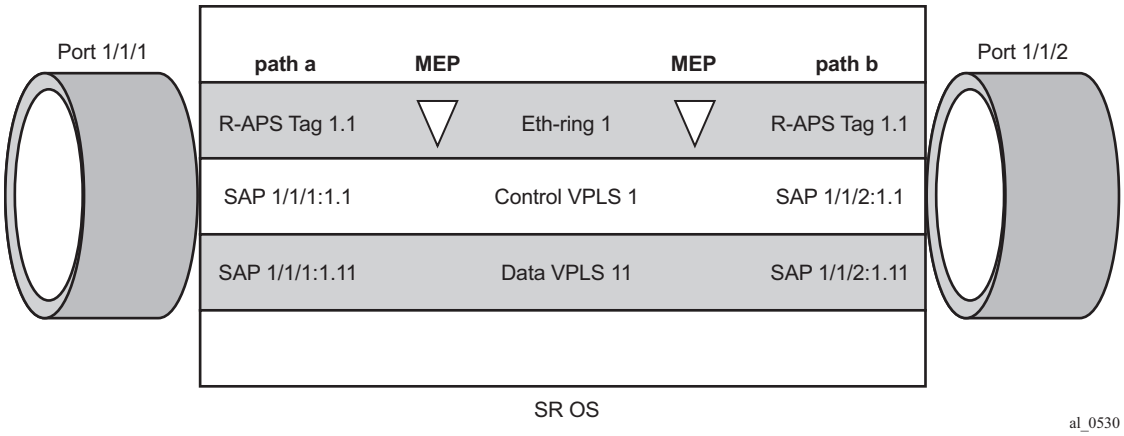
Provider Backbone Bridging (PBB) Support

PBB services also support G.8032 as data services (the services used for the control VPLS must be a regular VPLS). B/I-VPLS rings support both major rings and sub-rings. B-VPLS rings support multi-chassis link aggregation group (MC-LAG) as a dual homing option when aggregating I-VPLS traffic onto a B-VPLS ring. In other words, I-VPLS rings should not be dual-homed into two backbone edge bridge (BEB) nodes where the B-VPLS uses G.8032 to get connected to the rest of the B-VPLS network as the only mechanism which can propagate MAC flushes between an I-VPLS and B-VPLS is an LDP MAC flush.

SROS Implementation

G.8032 is built from VPLS components and each ring consists of the configuration components illustrated in Figure 232.

Figure 232 **G.8032 Ring Components**

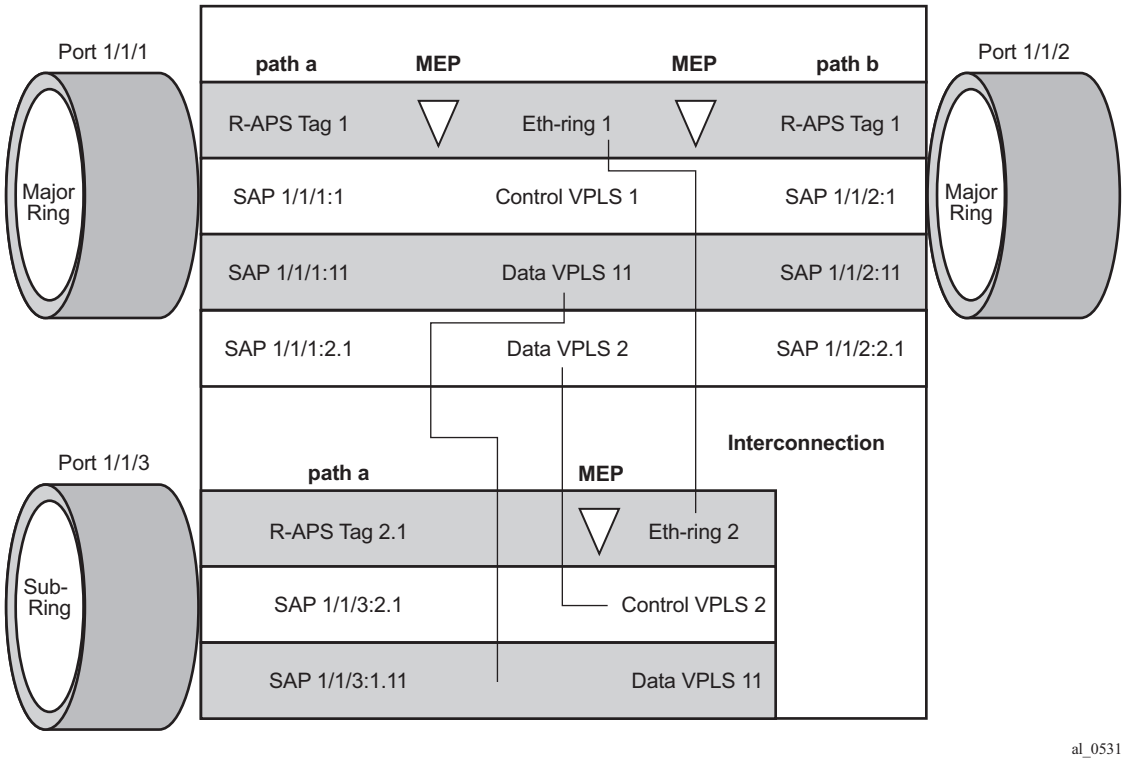


These components consist of:

- The Eth-ring instance which defines the R-APS tags, the MEPs and the ring behavior.
- The control VPLS which has the SAPs that match the R-APS.
- The data VPLS which is linked to the ring. All of the data VPLS SAPs follow the operational state of the control VPLS SAPs in that each blocked SAP controlled by the ring is blocked for all control and data instances.

Figure 233 shows the major ring and sub-ring interconnection components:

Figure 233 **G.8032 Sub-Ring Interconnection Components**



For a sub-ring, the configuration is the same as a single ring except at the junction of the major ring and the sub-ring. The interconnection of a sub-ring and a major ring links the control VPLS of the sub-ring to a data VPLS of the major ring when a virtual link is used. Similarly the data VPLS of the sub-ring is linked to a data VPLS of the major ring. G.8032 Sub-Ring Interconnection Components illustrates the relationship of a sub-ring and a major ring. Since this sub-ring has a virtual channel, the data VPLS 2 has both data SAPs from the sub-ring and data SAPs from the major ring. The virtual channel is also optional and in non-virtual-link cases, no VPLS instance is required (see non-virtual-link in the section Configuration of a Sub-Ring to a VPLS Service (with a Non-Virtual Link)).

In Figure 233, the inner tag values are kept the same for clarity, but in fact any encapsulation that is consistent with the next ring link will work. In other words, ring SAPs can perform VLAN ID translation and even when connecting a sub-ring to a major ring. This also means that other ports may reuse the same tags when connecting independent services.

The R-APS tags (ring automatic protection switching tags) and SAPs on the rings can either be dot1Q or QinQ encapsulated. It is also possible to have the control VPLS using single tagged frames with the data VPLSs using double tagged framed; this requires the system to be configured with the new-qinq-untagged-sap parameter (configure system ethernet new-qinq-untagged-sap), with the ring path raps-tags and control VPLS SAPs configured as qtag.0, and the data VPLSs configured as QinQ SAPs. STP cannot be enabled on SAPs connected to eth-rings.

R-APS messages received from other nodes are normally blocked on the RPL interface but the sub-ring case with non-virtual channel recommends that R-APS messages be propagated over the RPL. Configuring sub-ring non-virtual-link on all nodes on the sub-ring propagation of R-APS messages is mandatory in order to achieve this.

R-APS messages are forwarded out of the egress using forwarding class NC, this should be prioritized accordingly in the SAP egress QoS policy to ensure that congestion does not cause R-APS messages to be dropped which could cause the ring to switch to another path.

Configuration

This section describes the configuration of multiple rings. The eth-ring configuration commands are as follows.

```
configure eth-ring <ring-index>
              ccm-hold-time { [down <down-timeout>] [up <up-timeout>] }
              compatible-version {1|2}
              description <description-string>
              guard-time <time>
              node-id <xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx>
              path {a|b} [ {<port-id>|<lag-id>]} raps-tag <qtag1>[.<qtag2>] ]
                    description <description-string>
                    eth-cfm
                          mep <mep-id> domain <md-index> association <ma-index>
                          ...
                    rpl-end
                    shutdown
              revert-time <time>
              rpl-node {owner|nbr}
              shutdown
              sub-ring {virtual-link|non-virtual-link}
                  interconnect [ring-id <ring-index>|vpls]
                      propagate-topology-change
```

Parameters:

- <ring-index> — This is the number by which the ring is referenced, values: 1 to128.
- ccm-hold-time { [down <down-timeout>] [up <up-timeout>] }
  - **down** — This command specifies the timer which controls the delay between detecting that ring path is down and reporting it to the G.8032 protection module. If a non-zero value is configured, the system will wait for the time specified in the value parameter before reporting it to the G.8032 protection module. This parameter applies only to ring path CCM. It does not apply to the ring port link state. To dampen ring port link state transitions, use the hold-time parameter from the physical member port. This is useful if the underlying path between two nodes is going across an optical system which implements its own protection.
  - **up** — This command specifies the timer which controls the delay between detecting that ring path is up and reporting it to the G.8032 protection module. If a non-zero value is configured, the system will wait for the time specified in the value parameter before reporting it to the G.8032 protection module. This parameter applies only to ring path CCM. It does not apply to the member port link state. To dampen member port link state transitions, use the hold-time parameter from the physical member port.

    Values:
    <down-timeout> : [0..5000] in centiseconds - Default: 0
    <up-timeout> : [0..5000] in deciseconds - Default: 20
    1 centisecond = 10ms
    1 decisecond = 100ms
- **compatible-version** — This command configures eth-ring compatibility version for the G.8032 state machine and messages. The default is version 2 (ITU G.8032v2) and all 7x50 systems use version 2. If there is a need to interwork with third party devices that only support version 1, this can be set to version 1 allowing the reception of version 1 PDUs. Version 2 is encoded as 1 in the R-APS messages. Compatibility allows the reception of version 1 (encoded as 0) R-APS PDUs but, as per the G.8032 specification, higher versions are ignored on reception. For SR OS, messages are always originated with version 2. Therefore if a third party switch supported version 3 (encoded as 2) or higher, interworking is also supported provided the other switch is compatible with version 2.
- **description** <*description-string*> — This configures a text string, up to 80 characters, which can be used to describe the use of the eth-ring.

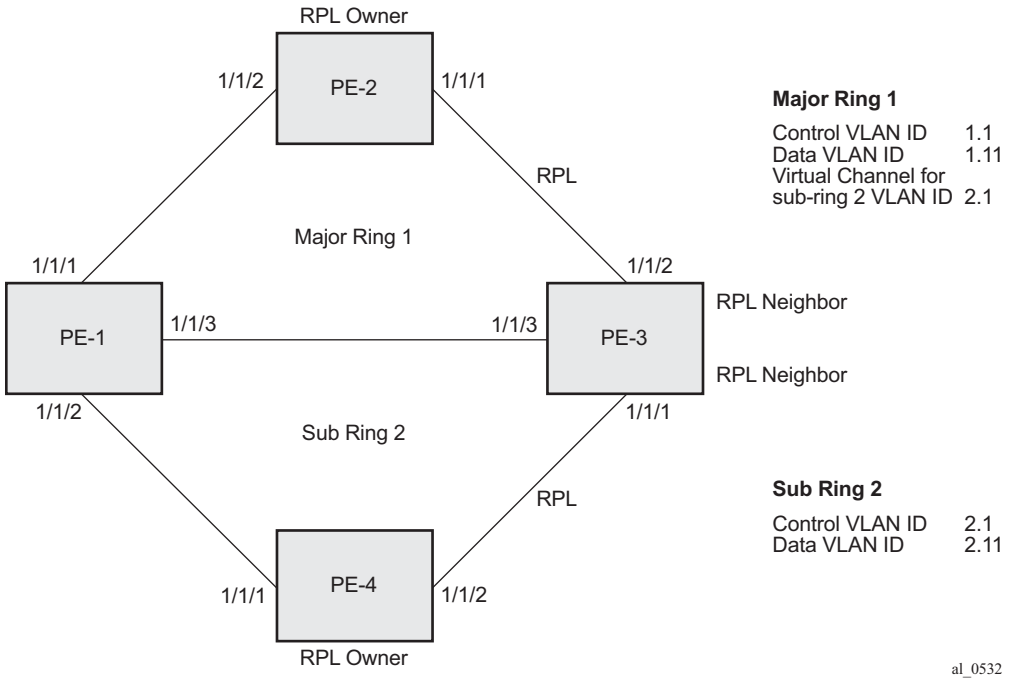
- **guard-time** <*time*> — The forwarding method, in which R-APS messages are copied and forwarded at every Ethernet ring node, can result in a message corresponding to an old request, that is no longer relevant, being received by Ethernet ring nodes. Reception of an old R-APS message may result in erroneous ring state interpretation by some Ethernet ring nodes. The guard timer is used to prevent Ethernet ring nodes from acting upon outdated R-APS messages and prevents the possibility of forming a closed loop. Messages are not forwarded when the guard-timer is running.

  Values: [1..20] in deciseconds - Default: 5
  1 decisecond = 100ms
- **node-id** <xx:xx:xx:xx:xx:xx> or <xx-xx-xx-xx-xx-xx> — This allows the node identifier to be explicitly configured. By default the chassis MAC is used. Not required in typical configurations.
- **path** {**a**|**b**} [{<*port-id*>|*lag-id*} **raps-tag** <*qtag*[.<*qtag*>]] — This parameter defines the paths around the ring, of which there are two in different directions on the ring: an “a” path and a “b” path, except on the interconnection node where a sub-ring connects to another major/sub ring in which case there is one path (either a or b) configured together with the **sub-ring** command. The paths are configured on a dot1Q or QinQ encapsulated access or hybrid port or a LAG with the encapsulation used for the R-APS messages on the ring. These can be either single or double tagged.
  - **description** <*description-string*> — This configures a text string, up to 80 characters, which can be used to describe the use of the path.
  - **eth-cfm** — Configures the associated Ethernet CFM parameters.
    - **mep** <*mep-id*> **domain** <*md-index*> **association** <*ma-index*> — The MEP defined under the path is used for the G.8032 protocol messages, which are based on IEEE 802.1ag/Y.1731 CFM frames.
  - **rpl-end** — When configured, this path is expected to be one end of the RPL. This parameter must be configured in conjunction with the *rpl-node* parameter.
  - **shutdown** — This command shuts down the path.
- **revert-time** <*time*> — This command configures the revert time for an Eth-Ring. Revert time is the time that the RPL will wait before returning to the blocked state. Configuring **no revert-time** disables reversion, effectively setting the revert-time to zero.
  Values: [60..720] in seconds - Default: 300

- **rpl-node** {**owner**|**nbr**} — A node can be designated as either the **owner** of the RPL, in which case this node is responsible for the RPL, or the **nbr** (neighbor), in which case this node is expected to be the neighbor to the RPL owner across the RPL. The **nbr** is optional and is included to be compliant with the specification. This parameter must be configured in conjunction with the **rpl-end** command. On a sub-ring without virtual channel it is mandatory to configure **sub-ring non-virtual-link** on all nodes on the sub-ring to ensure propagation of the R-APS messages around the sub-ring.
- **shutdown** — This command shuts down the ring.
- **sub-ring** {**virtual-link**|**non-virtual-link**} — This command is configured on the interconnection node between the sub-ring and its major/sub ring to indicate that this ring is a sub-ring. The parameter specifies whether it uses a virtual link through the major/sub ring for the R-APS messages or not. A ring configured as a sub-ring can only be configured with a single path.
  - **interconnect** [**ring-id** <*ring-index*>|**vpls**] — A sub-ring connects to either another ring or a VPLS service. If it connects to another ring (either a major ring or another sub-ring), the ring identifier must be specified and the ring to which it connects must be configured with both a path “a” and a path “b”, meaning that it is not possible to connect a sub-ring to another sub-ring on an interconnection node.
    Alternatively, the **vpls** parameter is used to indicate the sub-ring connects to a VPLS service. Interconnection using a VPLS service requires the sub-ring to be configured with **non-virtual-link**.
    - **propagate-topology-change** — If a topology change event happens in the sub-ring, it can be optionally propagated with the use of this parameter to either the major/sub ring it is connected to, using R-APS messages, or to the LDP VPLS SDP peers using an LDP “flush-all-from-me” message if the sub-ring is connected to a VPLS service.

The example topology is shown in Figure 234.

Figure 234 **Ethernet Test Topology**



The configuration is divided into the following sections:

- A sub-ring connected to a major ring using a virtual link through the major ring.
- A sub-ring connected to a major ring without a virtual link.
- A sub-ring connected to a VPLS service (without a virtual link).

Configure a Sub-Ring to a Major Ring with a Virtual Link

To configure an Ethernet ring using R-APS, there will be at least 2 VPLS services required for one Eth-Ring instance, one for the control channel and the other (or more) for data channel(s). The control channel is used for R-APS signaling while the data channel is for user data traffic. The state of the data channels is inherited from the state of the control channel.

The following needs to be configured:

- Encapsulation type for each ring port
- ETH-CFM
- Eth-ring for major ring 1

- Eth-ring for sub-ring 2
- Control channel service and add Eth-ring SAPs
- User data channels

Configure the Encapsulation for the Ring Ports.

Eth-Ring needs an R-APS tag to send/receive G.8032 signaling messages. To configure a control channel, an access SAP configuration is required on each path a/b port. The SAP configuration follows that of the port and must be either dot1Q or QinQ, consequently the control and data packets are either single tagged or double tagged. Single tagged control frames are supported on a QinQ port by configuring the system with the new-qinq-untagged-sap parameter (configure system ethernet new-qinq-untagged-sap), and the ring path raps-tags and control VPLS SAPs configured as qtag.0.

In this example, QinQ tags are used. The commands for the major and sub rings ring, on PE-1 for example, are:

```
*A:PE-1# configure port 1/1/1 ethernet mode access
*A:PE-1# configure port 1/1/2 ethernet mode access
*A:PE-1# configure port 1/1/3 ethernet mode access
*A:PE-1# configure port 1/1/1 ethernet encap-type qinq
*A:PE-1# configure port 1/1/2 ethernet encap-type qinq
*A:PE-1# configure port 1/1/3 ethernet encap-type qinq
```

Configuring ETH-CFM.

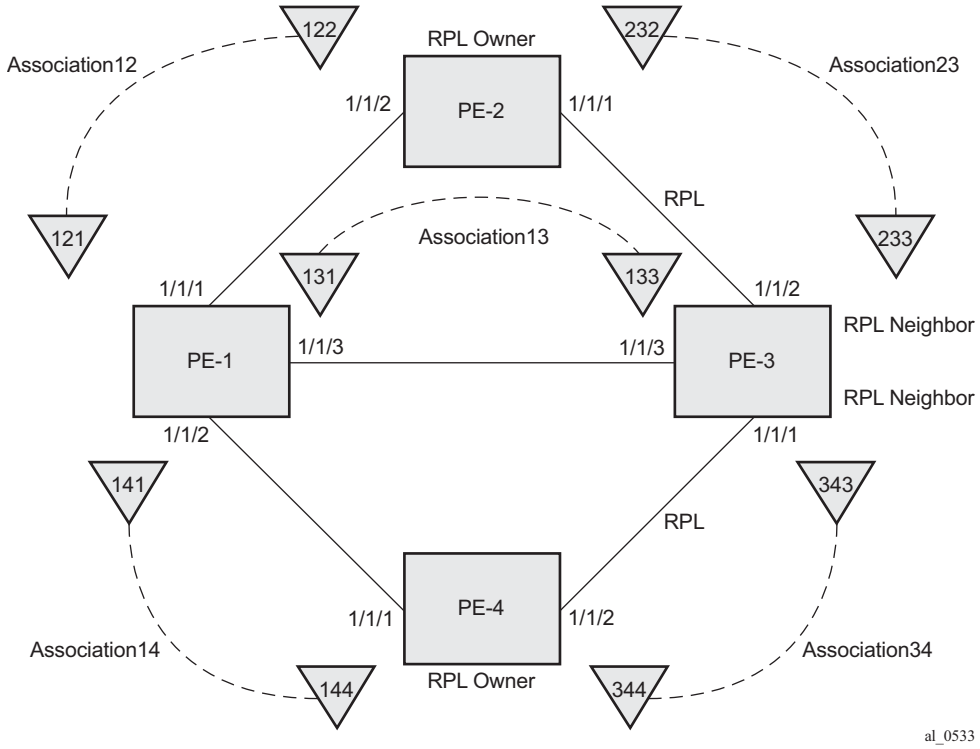
Configuring ETH-CFM domain, association and MEP is required before configuring Ethernet ring. The standard domain format is *none* and the association name should beITU Carrier Code-based (ICC-based - Y.1731), however, the SROS implementation is flexible in that it supports both IEEE and ICC formats. The *eth-ring* MEP requires a CCM interval with values such as 1s, 100ms or 10ms to be configured.

The MEPs used for R-APS control normally will have CCM configured on the control channel path MEPs for failure detection. Alternatively, detecting a failure of the ring may be achieved by running Ethernet in the First Mile (EFM) at the port level if CCM is not possible at 1s, 100ms or 10ms. Also rings can be run without CFM although the ETH-CFM association must be configured for R-APS messages to be exchanged. To omit the failure detecting CCMs, it would be necessary to remove the *ccm-enable* from under the path MEPs and to remove the *remote-mepid* on the corresponding ETH CFM configuration.

Loss-of-signal, in conjunction with other OAM mechanisms, is applicable only when the nodes are directly connected.

Figure 235 shows the details of the MEPs and their associations configured when both the major and sub rings are used. The associations only need to be pair wise unique but for clarity five unique associations are used. Any name format can be used, but it must be consistent on both adjacent nodes.

Figure 235 **ETH-CFM MEP Associations**



The configuration of ETH-CFM for the major and sub rings on each node is as follows. The CCMs for failure detection are configured for 1 second intervals.

Ring node PE-1: Association 12 and 13 are used for the major ring and association 14 is used for the sub-ring.

```
*A:PE-1# configure
    eth-cfm
        domain 1 format none level 2
            association 12 format icc-based name "Association12"
                ccm-interval 1
                remote-mepid 122
            exit
            association 13 format icc-based name "Association13"
                ccm-interval 1
                remote-mepid 133
```

```
            exit
            association 14 format icc-based name "Association14"
                ccm-interval 1
                remote-mepid 144
            exit
        exit
    exit
```

Ring node PE-2: Association 12 and 23 are used for the major ring.

```
*A:PE-2# configure
    eth-cfm
        domain 1 format none level 2
            association 12 format icc-based name "Association12"
                ccm-interval 1
                remote-mepid 121
            exit
            association 23 format icc-based name "Association23"
                ccm-interval 1
                remote-mepid 233
            exit
        exit
    exit
```

Ring node PE-3: Association 23 and 13 are used for the major ring and association 34 is used for the sub-ring.

```
*A:PE-3# configure
 eth-cfm
        domain 1 format none level 2
            association 13 format icc-based name "Association13"
                ccm-interval 1
                remote-mepid 131
            exit
            association 23 format icc-based name "Association23"
                ccm-interval 1
                remote-mepid 232
            exit
            association 34 format icc-based name "Association34"
                ccm-interval 1
                remote-mepid 344
            exit
        exit
    exit
```

Ring node PE-4: Association 14 and 34 are used for the sub-ring.

```
*A:PE-4# configure
    eth-cfm
        domain 1 format none level 2
            association 14 format icc-based name "Association14"
                ccm-interval 1
                remote-mepid 141
            exit
```

```
            association 34 format icc-based name "Association34"
                ccm-interval 1
                remote-mepid 343
            exit
        exit
    exit
```

Configuring Eth-Ring – Major Ring 1

Two paths must be configured to form a ring. In this example, VLAN tag 1.1 is used as control channel for R-APS signaling for the major ring (ring 1) on the ports shown in Figure 234 using the ETH CFM information shown in Figure 235. The revert-time is set to its minimum value and CCM messages are enabled on the MEP. The **control-mep** parameter is required to indicate that this MEP is used for ring R-APS messages.

Ring node PE-1:

```
*A:PE-1# configure
    eth-ring 1
        description "Ethernet Ring 1"
        revert-time 60
        path a 1/1/1 raps-tag 1.1
            description "Ethernet Ring 1 - PathA"
            eth-cfm
                mep 121 domain 1 association 12
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/3 raps-tag 1.1
            description "Ethernet Ring 1 - PathB"
            eth-cfm
                mep 131 domain 1 association 13
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

It is mandatory to configure a MEP in the path context, otherwise this error will be displayed.

```
*A:PE-1>config>eth-ring>path# no shutdown
INFO: ERMGR #1001 Not permitted - must configure eth-cfm MEP first
```

While MEPs are mandatory, enabling CCMs on the MEPs under the paths as a failure detection mechanism is optional as explained earlier.

Ring node PE-2: This is configured as the RPL owner with the RPL being on path “a” as indicated by the **rpl-end** parameter.

```
*A:PE-2# configure
    eth-ring 1
        description "Ethernet Ring 1"
        revert-time 60
        rpl-node owner
        path a 1/1/1 raps-tag 1.1
            description "Ethernet Ring 1 - PathA"
            rpl-end
            eth-cfm
                mep 232 domain 1 association 23
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 1.1
            description "Ethernet Ring 1 - PathB"
            eth-cfm
                mep 122 domain 1 association 12
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

It is not permitted to configure a path as an RPL end without having configured the node on this ring to be either the RPL *owner* or *nbr* otherwise the following error message is reported.

```
*A:PE-2>config>eth-ring>path# rpl-end
INFO: ERMGR #1001 Not permitted - path-type rpl-end is not consistent with eth-ring
'rpl-node' type
```

Ring node PE-3: This is configured as the RPL neighbor with the RPL being on path “b” as indicated by the **rpl-end** parameter.

```
*A:PE-3# configure
```

```
eth-ring 1
    description "Ethernet Ring 1"
    revert-time 60
    rpl-node nbr
    path a 1/1/3 raps-tag 1.1
        description "Ethernet Ring 1 - PathA"
        eth-cfm
            mep 133 domain 1 association 13
                ccm-enable
                control-mep
                no shutdown
            exit
        exit
        no shutdown
    exit
    path b 1/1/2 raps-tag 1.1
        description "Ethernet Ring 1 - PathB"
        rpl-end
        eth-cfm
            mep 233 domain 1 association 23
                ccm-enable
                control-mep
                no shutdown
            exit
        exit
        no shutdown
    exit
    no shutdown
exit
```

The link between PE-2 and PE-3 will be the RPL with PE-2 and PE-3 blocking that link when the ring is fully operational. In this example, the RPL is using path “a” on PE-2 and path “b” on PE-3.

Configuring Eth-Ring – Sub-Ring 2

Ring nodes PE-1, PE-3, and PE-4 form a sub-ring. The sub-ring attaches to the major ring (ring 1). The sub-ring in this case will use a virtual-link. The interconnection ring instance identifier (*ring-id*) is specified and *propagate-topology-change* indicates that sub-ring flushing will be propagated to the major ring. Only one path is specified since the other path is not required at an interconnection node. Sub-rings are almost identical to major rings in operation except that sub-rings send MAC flushes towards their connected ring (either a major or sub ring). Major or sub rings never send MAC flushes to their sub-rings. Therefore a couple of sub-rings connected to a major ring can cause MACs to flush on the major ring but the major ring will not propagate a sub-ring MAC flush to other sub-rings.

Ring node PE-1 provides an interconnection between the major ring (1) and the sub-ring (2). Ring 2 is configured to be a sub-ring which interconnects to ring 1. It will use a virtual link on ring 1 to send R-APS messages to the other interconnection node and topology changes will be propagated from sub-ring 2 to the major ring 1.

```
*A:PE-1# configure
    eth-ring 2
        description "Ethernet Sub-ring 2 on Major Ring 1"
        revert-time 60
        sub-ring virtual-link
            interconnect ring-id 1
                propagate-topology-change
            exit
        exit
        path a 1/1/2 raps-tag 2.1
            description "Ethernet Ring 2 - PathA"
            eth-cfm
                mep 141 domain 1 association 14
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

The configuration of PE-3 is similar to PE-1, but PE-3 is the RPL neighbor, with the RPL end on path “a”, for the RPL between PE-3 and PE-4.

```
*A:PE-3# configure
    eth-ring 2
        description "Ethernet Sub-ring 2 on Major Ring 1"
        revert-time 60
        rpl-node nbr
        sub-ring virtual-link
            interconnect ring-id 1
                propagate-topology-change
            exit
        exit
        path a 1/1/1 raps-tag 2.1
            description "Ethernet Ring 2 - PathA"
            rpl-end
            eth-cfm
                mep 343 domain 1 association 34
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

Ring node PE-4: This node only has configuration for the sub-ring, ring 2. It is also the RPL owner, with path “b” being the RPL end, for the RPL between PE-3 and PE-4.

```
*A:PE-4# configure
    eth-ring 2
        description "Ethernet Sub-ring 2"
        revert-time 60
        rpl-node owner
        exit
        path a 1/1/1 raps-tag 2.1
            description "Ethernet Ring 2 - PathA"
            eth-cfm
                mep 144 domain 1 association 14
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 2.1
            description "Ethernet Ring 2 - PathB"
            rpl-end
            eth-cfm
                mep 344 domain 1 association 34
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

Until the Ethernet Ring instance is attached to a VPLS service, the ring operational status is down and the forwarding status of each port is blocked. This prevents the operator from creating a loop by mis-configuration. This state can be seen on ring node PE-1 as follows:

```
*A:PE-1# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Down
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : Request State: 0xB
                     Sub-Code     : 0x0
                     Status       : 0x20  ( BPR )
                     Node ID      : 4a:c4:ff:00:00:00
```

```
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1.1             Up/Down      normal          blocked
  b  1/1/3    1.1             Up/Down      normal          blocked
===============================================================================
*A:PE-1#
```

Configure the Control Channel VPLS Service

Path “a” and “b” configured in the eth-ring must be added as SAPs into a VPLS service (standard VPLS) using the **eth-ring** parameter. The SAP encapsulation values must match the values of the *raps-tag* configured for the associated path.

G.8032 uses the same raps-tag value on all nodes on the ring, as configured in this example. However, the SR OS implementation relaxes this constraint by requiring the tag to match only on adjacent nodes.

A VPLS service (identifier 1) is configured on PE-1, PE-2 and PE-3 for the control channel for the major ring (ring1), and another VPLS service (identifier 2) is used on PE-1, PE-3 and PE-4 for the sub-ring (ring 2).

Ring node PE-1: Control service for the major ring.

```
*A:PE-1# configure
    service
        vpls 1 customer 1 create
            description "Control VID 1.1 for Ring 1 - Major Ring"
            sap 1/1/1:1.1 eth-ring 1 create
            exit
            sap 1/1/3:1.1 eth-ring 1 create
            exit
            no shutdown
```

Ring node PE-2: Control service for the major ring.

```
*A:PE-2# configure
    service
        vpls 1 customer 1 create
            description "Control VID 1.1 for Ring 1 - Major Ring"
            sap 1/1/1:1.1 eth-ring 1 create
            exit
            sap 1/1/2:1.1 eth-ring 1 create
            exit
            no shutdown
```

Ring node PE-3: Control service for the major ring.

```
*A:PE-3# configure
    service
        vpls 1 customer 1 create
            description "Control VID 1.1 for Ring 1 - Major Ring"
            sap 1/1/2:1.1 eth-ring 1 create
            exit
            sap 1/1/3:1.1 eth-ring 1 create
            exit
            no shutdown
```

SAPs or SDPs can be added to a control channel VPLS on condition the **eth-ring** parameter is present. Trying to add a SAP without this parameter to a control channel VPLS will result in the following message being displayed.

```
*A:PE-1# configure service vpls 1 sap 1/2/1:1 create
MINOR: SVCMGR #1321 Service contains an Ethernet ring control SAP
```

For the sub-ring, the configuration of a split horizon group for the virtual channel on the major ring on the interconnection nodes is recommended. This avoids the looping of control R-APS messages in the case there is a mis-configuration in the major ring.

Ring node PE-1: Control service for the sub-ring. Notice that two SAPs connect to the major ring (ring 1), these being for the virtual channel, and the third SAP connects to the sub-ring (ring 2).

```
*A:PE-1# configure
    service
        vpls 2 customer 1 create
            description "Control/Virtual Channel VID 2.1 for Ring 2"
            split-horizon-group "shg-ring2" create
            exit
            sap 1/1/1:2.1 split-horizon-group "shg-ring2" eth-ring 1 create
                description "Ring 2 Interconnection using Ring 1"
            exit
            sap 1/1/2:2.1 eth-ring 2 create
            exit
            sap 1/1/3:2.1 split-horizon-group "shg-ring2" eth-ring 1 create
                description "Ring 2 Interconnection using Ring 1"
            exit
            no shutdown
```

Ring node PE-2: Control service for the sub-ring. Sub-ring 2 is not present on PE-2, however, its virtual channel on major ring 1 needs to exist throughout ring 1.

```
*A:PE-2# configure
    service
```

```
        vpls 2 customer 1 create
            description "Virtual Channel VID 2.1 for Ring 2"
            sap 1/1/1:2.1 eth-ring 1 create
            exit
            sap 1/1/2:2.1 eth-ring 1 create
            exit
            no shutdown
```

If multiple virtual channels are used (due to the aggregation of multiple sub-rings into the same major ring), their configuration could be simplified on non-interconnection nodes on the major ring. To achieve this on a ring node such as PE-2, a default SAP could be used rather than configuring a VPLS per virtual channel. If QinQ SAPs are used then a default SAP of 1/1/[1,2]:qtag.* could be used but requires all control channels for sub-rings to be using qtag as the outer VLAN ID, or 1/1/[1,2]:* if dot1Q SAPs were used. This is because the SAPs match explicit SAPs definitions first and the default SAP will handle any other traffic.

Ring node PE-3: Control service for the sub-ring. This is similar to the configuration of PE-1.

```
*A:PE-3# configure
    service
        vpls 2 customer 1 create
            description "Control/Virtual Channel VID 2.1 for Ring 2"
            split-horizon-group "shg-ring2" create
            exit
            sap 1/1/1:2.1 eth-ring 2 create
            exit
            sap 1/1/2:2.1 split-horizon-group "shg-ring2" eth-ring 1 create
                description "Ring 2 Interconnection using Ring 1"
            exit
            sap 1/1/3:2.1 split-horizon-group "shg-ring2" eth-ring 1 create
                description "Ring 2 Interconnection using Ring 1"
            exit
            no shutdown
```

Ring node PE-4: Control service for the sub-ring. Both SAPs are configured on the sub-ring
(ring 2).

```
*A:PE-4# configure
    service
        vpls 2 customer 1 create
            description "Control VID 2.1 for Ring 2 Sub-ring"
            sap 1/1/1:2.1 eth-ring 2 create
            exit
            sap 1/1/2:2.1 eth-ring 2 create
            exit
            no shutdown
```

At this point, the Eth-Ring 1 is operationally up and the RPL is blocking successfully on ring node PE-2 port 1/1/1, as expected for the RPL owner/end configuration and on port 1/1/2 on PE-3 as the RPL neighbor.

Show Output

An overview of all of the rings can be shown using the following commands, in this case on
PE-1.

First, the ETH ring status is shown.

```
*A:PE-1# show eth-ring status
===============================================================================
Ethernet Ring (Status information)
===============================================================================
Ring   Admin  Oper       Path Information                MEP Information
ID     State  State  Path         Tag       State     Ctrl-MEP CC-Intvl Defects
-------------------------------------------------------------------------------
1      Up     Up     a - 1/1/1        1.1   Up        Yes      1         -----
                     b - 1/1/3        1.1   Up        Yes      1         -----
2      Up     Up     a - 1/1/2        2.1   Up        Yes      1         -----
                     b - N/A           -        -        -        -----
===============================================================================
Ethernet Tunnel MEP Defect Legend:
R = Rdi, M = MacStatus, C = RemoteCCM, E = ErrorCCM, X = XconCCM
*A:PE-1#
```

It is expected that the state is “up”, even on ring paths which are blocked. The “Defects” column refers to the CFM defects of the MEPs. If there is a problem, these will be flagged.

The following output shows the ring and path forwarding states.

```
*A:PE-1# show eth-ring
===============================================================================
Ethernet Rings (summary)
===============================================================================
Ring Int  Admin Oper            Paths Summary                       Path States
ID   ID   State State                                               a     b
-------------------------------------------------------------------------------
1    -    Up    Up    a - 1/1/1        1.1   b - 1/1/3        1.1   U     U
2    1    Up    Up    a - 1/1/2        2.1   b - Not configured     U     -
===============================================================================
Ethernet Ring Summary Legend:   B - Blocked    U - Unblocked
*A:PE-1#
```

The specific ring information can be shown as follows.

Ring node PE-1:

```
*A:PE-1# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1.1             Up/Up        normal          unblocked
  b  1/1/3    1.1             Up/Up        normal          unblocked
===============================================================================
*A:PE-1#
```

The status around the major ring can also be checked.

Ring node PE-2: Major ring.

```
*A:PE-2# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c5:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplOwner
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : Request State: 0x0
                     Sub-Code     : 0x0
                     Status       : 0x80  ( RB )
                     Node ID      : 4a:c5:ff:00:00:00
Defect Status      :

Sub-Ring Type      : none

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1.1             Up/Up        rplEnd          blocked
  b  1/1/2    1.1             Up/Up        normal          unblocked
```

```
===============================================================================
*A:PE-2#
```

PE-2 is the RPL owner with port 1/1/1 as an RPL end, which is blocked as expected. The *revert-time* is also shown to be the configured value. Detailed information is shown relating to the R-APS PDUs being transmitted on this ring as this node is the RPL owner.

When a revert is pending, the “Time to Revert” will show the number of seconds remaining before the revert occurs.

Ring node PE-3: Major ring.

```
*A:PE-3# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c6:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNeighbor
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :

Sub-Ring Type      : none

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type             Fwd State
-------------------------------------------------------------------------------
  a  1/1/3    1.1             Up/Up        normal           unblocked
  b  1/1/2    1.1             Up/Up        rplEnd           blocked
===============================================================================
*A:PE-3#
```

PE-3 is the RPL neighbor with port 1/1/2 as an RPL end which is blocked as expected.

The information for the sub-ring can also be shown using a similar command.

Ring node PE-1: Sub-ring.

```
*A:PE-1# show eth-ring 2
===============================================================================
Ethernet Ring 2 Information
===============================================================================
Description        : Ethernet Sub-ring 2 on Major Ring 1
Admin State        : Up                 Oper State       : Up
```

```
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :

Sub-Ring Type      : virtualLink        Interconnect-ID  : 1
Topology Change    : Propagate

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/2    2.1             Up/Up        normal          unblocked
  b  -        -                 -/-         -               -
===============================================================================
*A:PE-1#
```

Only path “a” is active and unblocked. The second path, path “b” is not configured as only one path is required on an interconnection node. The “Sub-Ring Type” is shown to be a virtual link interconnecting to ring 1, with topology propagation enabled.

Ring node PE-3: Sub-ring.

```
*A:PE-3# show eth-ring 2
===============================================================================
Ethernet Ring 2 Information
===============================================================================
Description        : Ethernet Sub-ring 2 on Major Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c6:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNeighbor
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :
Sub-Ring Type      : virtualLink        Interconnect-ID  : 1
Topology Change    : Propagate
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    2.1             Up/Up        rplEnd          blocked
  b  -        -                 -/-         -               -
===============================================================================
*A:PE-3#
```

PE-3 is the RPL neighbor with port 1/1/1 as an RPL end, which is blocked as expected.

Ring Node PE-4: Sub-ring.

```
*A:PE-4# show eth-ring 2

===============================================================================
Ethernet Ring 2 Information
===============================================================================
Description        : Ethernet Sub-ring 2
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c7:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplOwner
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : Request State: 0x0
                     Sub-Code     : 0x0
                     Status       : 0xE0  ( RB DNF BPR )
                     Node ID      : 4a:c7:ff:00:00:00
Defect Status      :

Sub-Ring Type      : none

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type             Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    2.1             Up/Up        normal           unblocked
  b  1/1/2    2.1             Up/Up        rplEnd           blocked
===============================================================================
*A:PE-4#
```

PE-4 is the RPL owner with port 1/1/2 as an RPL end, which is blocked as expected.

The details of an individual path can be shown.

```
*A:PE-1# show eth-ring 1 path a

===============================================================================
Ethernet Ring 1 Path Information
===============================================================================
Description        : Ethernet Ring 1 - PathA
Port               : 1/1/1              Raps-Tag         : 1.1
Admin State        : Up                 Oper State       : Up
Path Type          : normal             Fwd State        : unblocked
                                        Fwd State Change : 05/09/2016 08:35:29
Last Switch Command: noCmd
APS Rx PDU         : Request State: 0x0
                     Sub-Code     : 0x0
                     Status       : 0x80  ( RB )
                     Node ID      : 4a:c5:ff:00:00:00

===============================================================================
*A:PE-1#
```

The ring hierarchy created can be shown, either for all rings, or as below for a specific ring.

```
*A:PE-1# show eth-ring 1 hierarchy

===============================================================================
Ethernet Ring 1 (hierarchy)
===============================================================================
Ring Int  Admin Oper            Paths Summary                       Path States
ID   ID   State State                                               a     b
-------------------------------------------------------------------------------
1    -    Up    Up    a - 1/1/1        1.1   b - 1/1/3        1.1   U     U
2    1    Up    Up    a - 1/1/2        2.1   b - Not configured     U     -
===============================================================================
Ethernet Ring Summary Legend:   B - Blocked    U - Unblocked
*A:PE-1#
```

Configuring the User Data Channel VPLS Service

The user data channels are created on a separate VPLS, VPLS 11 in this example, using VLAN tag 1.11. The ring data channels must be on the same ports as the corresponding control channels configured above. The access into the data services can use normal SAPs and/or SDPs, for example the SAP on port 1/2/1 in the following output. Customer data traverses the ring on a data SAP. Multiple parallel data SAPs in different data services can be controlled by one control ring instance (eth-ring 1 in the example).

Ring node PE-1: Two data SAPs correspond to the major ring 1, while the third SAP is the data SAP on the sub-ring 2.

```
*A:PE-1# configure
    service
        vpls 11 customer 1 create
            description "Data VPLS"
            sap 1/1/1:1.11 eth-ring 1 create
            exit
            sap 1/1/2:1.11 eth-ring 2 create
            exit
            sap 1/1/3:1.11 eth-ring 1 create
            exit
            sap 1/2/1:11 create
                description "Sample Customer Service SAP"
            exit
            no shutdown
```

Ring node PE-3 (not shown) would be similar to ring node 1.

Ring node PE-2 is also similar with a single ring data service using VPLS 11 and tag 1.11.

```
*A:PE-2# configure
    service
        vpls 11 customer 1 create
            description "Data VPLS"
            sap 1/1/1:1.11 eth-ring 1 create
            exit
            sap 1/1/2:1.11 eth-ring 1 create
            exit
            sap 1/2/1:11 create
                description "Sample Customer Service SAP"
            exit
            no shutdown
```

Ring node PE- 4: On ring node PE-4 the data VLAN ID is configured as a normal ring data VPLS on ring 2.

```
*A:PE-4# configure
    service
        vpls 11 customer 1 create
            description "Data VPLS"
            sap 1/1/1:1.11 eth-ring 2 create
            exit
            sap 1/1/2:1.11 eth-ring 2 create
            exit
            sap 1/2/1:11 create
                description "Sample Customer Service SAP"
            exit
            no shutdown
```

All of the SAPs which are configured to use Ethernet rings can be shown. The following output is taken from PE-1, where there are:

- two SAPs in VPLS 1 for the control channel of ring 1 (VLAN ID 1.1)
- two SAPs in VPLS 2 on ring 1 for the virtual channel for ring 2 (VLAN ID 2.1).
- one SAP in VPLS 2 on ring 2 for the control channel for ring 2 (VLAN ID 2.1)
- three SAPs in VPLS 11, two on ring 1 and one on ring 2, for the data service (VLAN ID 1.11). This matches the information in Figure 3.

```
*A:PE-1# show service sap-using eth-ring

===============================================================================
Service Access Points (Ethernet Ring)
===============================================================================
SapId              SvcId             Eth-Ring Path Admin Oper  Blocked Control/
                                               State State         Data
-------------------------------------------------------------------------------
1/1/1:1.1          1                 1        a    Up    Up    No      Ctrl
1/1/3:1.1          1                 1        b    Up    Up    No      Ctrl
1/1/1:2.1          2                 1        a    Up    Up    No      Ctrl
1/1/2:2.1          2                 2        a    Up    Up    No      Ctrl
1/1/3:2.1          2                 1        b    Up    Up    No      Ctrl
1/1/1:1.11         11                1        a    Up    Up    No      Data
1/1/2:1.11         11                2        a    Up    Up    No      Data
```

```
1/1/3:1.11         11                1        b    Up    Up    No      Data
-------------------------------------------------------------------------------
Number of SAPs : 8
===============================================================================
*A:PE-1#
```

Statistics are available showing both the CCM and R-APS messages sent and received on a node. An associated **clear** command is available.

```
*A:PE-1# show eth-cfm statistics
===============================================================================
ETH-CFM System Statistics
===============================================================================
Rx Count           : 1201               Tx Count           : 1066
Dropped Congestion : 0                  Discarded Error    : 0
AIS Currently Act  : 0                  AIS Currently Fail : 0
===============================================================================
================================
ETH-CFM System Op-code Statistics
================================
Op-code      Rx Count   Tx Count
--------------------------------
ccm              1018       1018
...
raps              183         48
...
--------------------------------
Total            1201       1066
================================
*A:PE-1#
```

To see an example of the console messages on a ring failure, when the unblocked port (1/1/2) on PE-2 is shut down, the following messages are displayed.

```
*A:PE-2# configure port 1/1/2 shutdown

26 2016/05/09 12:38:16.65 UTC WARNING: SNMP #2004 Base 1/1/2
"Interface 1/1/2 is not operational"

27 2016/05/09 12:38:16.65 UTC MINOR: ERING #2001 Base eth-ring-1
"Eth-Ring 1 path b changed fwd state to blocked"

28 2016/05/09 12:38:16.65 UTC MINOR: ERING #2001 Base eth-ring-1
"Eth-Ring 1 path a changed fwd state to unblocked"
*A:PE-2#
29 2016/05/09 12:38:16.67 UTC MAJOR: SVCMGR #2210 Base
"Processing of an access port state change event is finished and the status of a

ll affected SAPs on port 1/1/2 has been updated."
30 2016/05/09 12:38:19.73 UTC MINOR: ETH_CFM #2001 Base
"MEP 1/12/122 highest defect is now defRemoteCCM"
```

For troubleshooting, the **tools dump eth-ring** <*ring-index*> command displays path information, the internal state of the control protocol, related statistics information and up to the last 16 protocol events (including messages sent and received, and the expiration of timers). An associated **clear** parameter exists, which clears the event information in this output when the command is entered. The following is an example of the output on PE-1.

```
*A:PE-1# tools dump eth-ring 1

ringId 1 (Up/Up): numPaths 2 nodeId 4a:c4:ff:00:00:00
 SubRing: none (interconnect ring 0, propagateTc  No), Cnt 1
  path-a, port 1/1/1 (Up), tag 1.1(Up) status (Up/Up/Fwd)
      cc (Dn/Up): Cnt 3/3 tm 000 04:26:03.850/000 04:31:25.920
      state: Cnt 6 B/F 000 04:26:03.850/000 04:31:28.910, flag: 0x0
  path-b, port 1/1/3 (Up), tag 1.1(Up) status (Up/Up/Fwd)
      cc (Dn/Up): Cnt 2/2 tm 497 02:26:01.820/000 00:25:29.040
      state: Cnt 4 B/F 497 02:26:01.820/000 00:25:31.620, flag: 0x0
  FsmState=  PEND, Rpl = None, revert = 60 s, guard = 5 ds
    Defects =
    Running Timers =
    lastTxPdu = 0x0000 Nr (stopped)
    path-a Normal, RxId(I)= 4a:c5:ff:00:00:00, rx= v1-0x0020 Nr, cmd= None
    path-b Normal, RxId(I)= 4a:c5:ff:00:00:00, rx= v1-0x0020 Nr, cmd= None
  DebugInfo:  aPathSts 5, bPathSts 3, pm (set/clr) 0/0, txFlush 0
    RxRaps: ok 12 nok 0 self 0, TmrExp - wtr 0(0), grd 3, wtb 0
    Flush: cnt 9 (4/5/0) tm 000 04:31:27.710-000 04:31:27.710 Out/Ack 0/1
    RxRawRaps: aPath 2940 bPath 94 vPath 0
    Now: 000 04:31:48.480 , softReset: No - noTx 0

  Seq Event  RxInfo(Path: NodeId-Bytes)
             state:TxInfo (Bytes)            Dir  pA  pB        Time
  === ===== ============================  ===== === === ================
  008   pdu A: 4a:c5:ff:00:00:00-0xb040 Sf(DNF)
              PROT  : 0xb060  Sf(DNF)       Rx<-- Fwd Blk 000 00:22:41.910
  009   pdu B: 4a:c6:ff:00:00:00-0xb020 Sf
              PROT  : 0xb060  Sf(DNF)       RxF<- Fwd Blk 000 00:25:28.930
  010   bUp
              PEND-G: 0x0020  Nr            Tx--> Fwd Blk 000 00:25:31.040
  011   pdu A: 4a:c5:ff:00:00:00-0x0000 Nr
              PEND  : 0x0020  Nr            Rx<-- Fwd Blk 000 00:25:31.620
  012   pdu
              PEND  :                       ----- Fwd Fwd 000 00:25:31.620
  013   pdu B: 4a:c6:ff:00:00:00-0xb060 Sf(DNF)
              PEND  :                       Rx<-- Fwd Fwd 000 00:25:31.930
  014   pdu
              PROT  :                       ----- Fwd Fwd 000 00:25:31.930
  015   pdu B: 4a:c6:ff:00:00:00-0x0020 Nr
              PROT  :                       Rx<-- Fwd Fwd 000 00:25:32.430
  016   pdu
              PEND  :                       ----- Fwd Fwd 000 00:25:32.430
  017   pdu A: 4a:c6:ff:00:00:00-0x0020 Nr
              PEND  :                       Rx<-- Fwd Fwd 000 00:25:32.430
  018   pdu A: 4a:c5:ff:00:00:00-0x0080 Nr(RB )
              PEND  :                       RxF<- Fwd Fwd 000 00:26:32.720
  019   pdu
              IDLE  :                       ----- Fwd Fwd 000 00:26:32.720
  000   pdu B: 4a:c5:ff:00:00:00-0xb020 Sf
```

```
              IDLE  :                        RxF<- Fwd Fwd 000 04:26:01.010
  001   pdu
              PROT  :                        ----- Fwd Fwd 000 04:26:01.010
  002   aDn
              PROT  : 0xb000  Sf             TxF-> Blk Fwd 000 04:26:03.850
  003   pdu A: 4a:c5:ff:00:00:00-0xb020 Sf
              PROT  : 0xb000  Sf             RxF<- Blk Fwd 000 04:31:27.710
  004   aUp
              PEND-G: 0x0000  Nr             Tx--> Blk Fwd 000 04:31:27.840
  005   pdu A: 4a:c5:ff:00:00:00-0x0020 Nr
              PEND  : 0x0000  Nr             Rx<-- Blk Fwd 000 04:31:28.910
  006   pdu
              PEND  :                        ----- Fwd Fwd 000 04:31:28.910
  007   pdu B: 4a:c5:ff:00:00:00-0x0020 Nr
              PEND  :                        Rx<-- Fwd Fwd 000 04:31:28.910

*A:PE-1#
```

Configuration of a Sub-Ring to a Major Ring with a Non-Virtual Link

The differences from the preceding virtual link configuration with a non-virtual link for the sub-ring are:

- The sub-ring configuration on the interconnection nodes, PE-1 and PE-3, is modified to indicate that the sub-ring is not using a virtual link, otherwise it remains the same.
- The sub-ring configuration on the sub-ring node, PE-4, is also modified to indicate that this is part of a sub-ring that is not using a virtual link. This is mandatory on all non-interconnection nodes on the sub-ring in order to ensure the propagation of R-APS messages around the sub-ring.
- The virtual link services and SAPs must be removed from PE-1, PE-2 and PE3, that is:
  - On PE-1 and PE-3, the SAPs in VPLS 2 around the major ring (configured with the parameter *eth-ring 1*) are removed.
  - The service VPLS 2 is removed completely from PE-2.

The new configuration of sub-ring 2 on PE-1 is sas follows, the configuration on PE-3 is similar.

```
*A:PE-1# configure eth-ring 2
*A:PE-1>config>eth-ring# info
----------------------------------------------
    description "Ethernet Sub-ring 2 on Major Ring 1"
        revert-time 60
        sub-ring non-virtual-link
            interconnect ring-id 1
```

```
            propagate-topology-change
        exit
    exit
    path a 1/1/2 raps-tag 2.1
        description "Ethernet Ring 2 - PathA"
        eth-cfm
            mep 141 domain 1 association 14
                ccm-enable
                control-mep
                no shutdown
            exit
        exit
        no shutdown
    exit
    no shutdown
```

The configuration of sub-ring 2 on PE-4 is as follows; note the configuration of the sub-ring non-virtual-link.

```
*A:PE-4# configure eth-ring 2
*A:PE-4>config>eth-ring# info
----------------------------------------------
    description "Ethernet Sub-ring 2"
        revert-time 60
        rpl-node owner
        sub-ring non-virtual-link
        exit
        path a 1/1/1 raps-tag 2.1
            description "Ethernet Ring 2 - PathA"
            eth-cfm
                mep 144 domain 1 association 14
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 2.1
            description "Ethernet Ring 2 - PathB"
            rpl-end
            eth-cfm
                mep 344 domain 1 association 34
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

The SAP usage on PE-1 is as follows with only the control and data SAPs to PE-4 now using sub-ring 2.

```
*A:PE-1# show service sap-using eth-ring
===============================================================================
Service Access Points (Ethernet Ring)
===============================================================================
SapId              SvcId             Eth-Ring Path Admin Oper  Blocked Control/
                                                   State State         Data
-------------------------------------------------------------------------------
1/1/1:1.1          1                 1        a    Up    Up    No      Ctrl
1/1/3:1.1          1                 1        b    Up    Up    No      Ctrl
1/1/2:2.1          2                 2        a    Up    Up    No      Ctrl
1/1/1:1.11         11                1        a    Up    Up    No      Data
1/1/2:1.11         11                2        a    Up    Up    No      Data
1/1/3:1.11         11                1        b    Up    Up    No      Data
-------------------------------------------------------------------------------
Number of SAPs : 6
===============================================================================
*A:PE-1#
```

The information relating to sub-ring 2 is as follows and it can be seen that this is now not using a virtual link, but sub-ring 2 is still connected to major ring 1 and propagation is still enabled from the sub-ring to the major ring. The single ring path (a) is unblocked as the RPL is configured between PE-3 and PE-4.

```
*A:PE-1# show eth-ring 2

===============================================================================
Ethernet Ring 2 Information
===============================================================================
Description        : Ethernet Sub-ring 2 on Major Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :

Sub-Ring Type      : nonVirtualLink     Interconnect-ID  : 1
Topology Change    : Propagate

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/2    2.1             Up/Up        normal          unblocked
  b  -        -                 -/-         -               -
===============================================================================
*A:PE-1#
```

Configuration of a Sub-Ring to a VPLS Service (with a Non-Virtual Link)

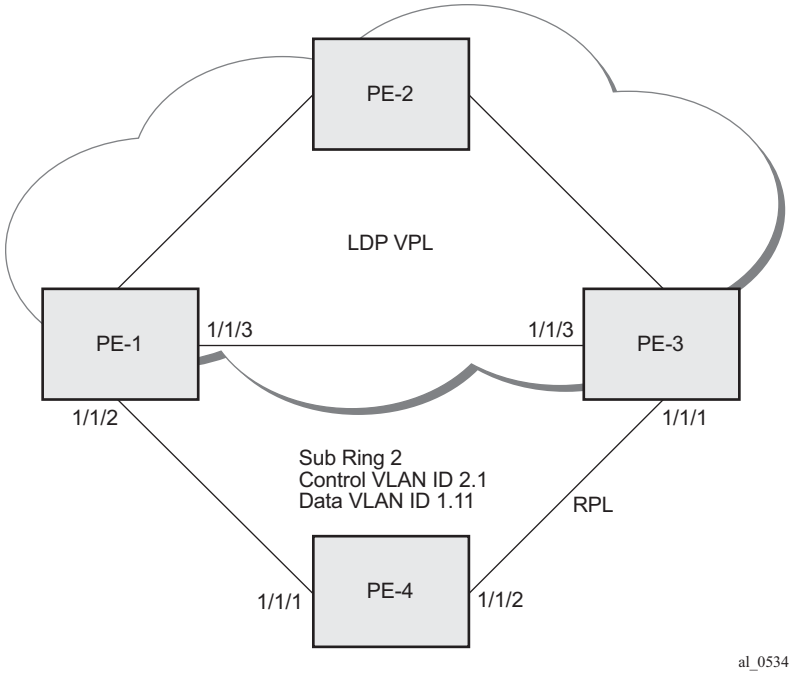
Sub-rings can be connected to VPLS services, in which case a virtual link is not used and is not configurable. While similar to the ring interconnect, there are a few differences.

Flush propagation is from the sub-ring to the VPLS, in the same way as it was for the sub-ring to the major ring. The same configuration parameter is used to propagate topology changes, note that in this case LDP flush messages (flush-all-from-me) are sent into the LDP portion of the network to account for ring changes without the need to configure anything in the VPLS service.

As with other rings, until an Ethernet ring instance is attached to the VPLS service, the ring operational status is down and the forwarding status of each port is blocked. This prevents operator from creating a loop by mis-configuration.

The topology for this case is shown in Figure 236. The configuration is very similar to the sub-ring with a non-virtual link described earlier, but ring 1 is replaced by a VPLS service using LDP signaled mesh SDPs between PE-1, PE-2 and PE-3 to create a fully meshed VPLS service. Both spoke and mesh SDPs using LDP could be used for the VPLS, however, only mesh SDPs have been used in this example.

Figure 236 **Sub-Ring to VPLS Topology**



The differences for the VPLS service connection to the configuration when the sub-ring is connected to a major ring without a virtual link are:

- The sub-ring configuration on the interconnection nodes, PE-1 and PE-3, is modified to indicate that the sub-ring is connected to a VPLS service.
- The sub-ring configuration on the sub-ring node, PE-4, is also modified to indicate that this is part of a sub-ring that is not using a virtual link. This is mandatory on all non-interconnection nodes on the sub-ring in order to ensure the propagation of R-APS messages around the sub-ring.
- The service (VPLS 1) and SAPs relating to the major ring 1 on PE-1, PE-2 and PE-3 are removed. These are replaced by routed IP interfaces configured with a routing protocol and LDP in order to signal the required MPLS labels, together with the necessary SDPs to provide interconnection at a service level.
- The data service (VPLS 11) is configured with mesh SDPs between PE-1, PE-2 and PE-3.

The configuration on PE-1 of the sub-ring 2 is as follows with the interconnect indicating a VPLS service. The configuration on PE-3 is similar.

```
*A:PE-1>config# eth-ring 2
*A:PE-1>config>eth-ring# info
----------------------------------------------
        description "Ethernet Sub-ring 2 on Major Ring 1"
        revert-time 60
        sub-ring non-virtual-link
            interconnect vpls
                propagate-topology-change
            exit
        exit
        path a 1/1/2 raps-tag 2.1
            description "Ethernet Ring 2 - PathA"
            eth-cfm
                mep 141 domain 1 association 14
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

The configuration of sub-ring 2 on PE-4 is as follows; note the configuration of the sub-ring non-virtual-link.

```
*A:PE-4# configure eth-ring 2
*A:PE-4>config>eth-ring# info
----------------------------------------------
        description "Ethernet Sub-ring 2"
        revert-time 60
        rpl-node owner
```

```
        sub-ring non-virtual-link
        exit
        path a 1/1/1 raps-tag 2.1
            description "Ethernet Ring 2 - PathA"
            eth-cfm
                mep 144 domain 1 association 14
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 2.1
            description "Ethernet Ring 2 - PathB"
            rpl-end
            eth-cfm
                mep 344 domain 1 association 34
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

The data service on PE-1 is as follows. The configuration on PE-3 is similar.

```
*A:PE-1# configure
    service
        vpls 11 customer 1 create
            description "Data VPLS"
            sap 1/1/2:1.11 eth-ring 2 create
            exit
            sap 1/2/1:11 create
                description "Sample Customer Service SAP"
            exit
            mesh-sdp 12:11 create
            exit
            mesh-sdp 13:11 create
            exit
            no shutdown
```

The state of the sub-ring is as follows and shows the sub-ring is not using a virtual link, is connected to a VPLS service and has propagation of topology change events enabled. As earlier, the single ring path (a) is unblocked as the RPL is configured between PE-3 and PE-4.

```
*A:PE-1# show eth-ring 2

===============================================================================
Ethernet Ring 2 Information
```

```
===============================================================================
Description        : Ethernet Sub-ring 2 on Major Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :

Sub-Ring Type      : nonVirtualLink     Interconnect-ID  : VPLS
Topology Change    : Propagate

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type             Fwd State
-------------------------------------------------------------------------------
  a  1/1/2    2.1             Up/Up        normal           unblocked
  b  -        -                 -/-         -                -
===============================================================================
*A:PE-1#
```

In this case, if a topology change event occurs in the sub-ring, an LDP flush all-from-me message is sent by PE-1 and PE-3 to their LDP peers. This can be seen by enabling the following debugging for PE-1, where packets 1 and 2 are the flush messages.

```
*A:PE-1# debug router ldp peer 192.0.2.2 packet init
*A:PE-1# debug router ldp peer 192.0.2.3 packet init
*A:PE-1#
*A:PE-1# show debug
debug
    router "Base"
        ldp
            peer 192.0.2.2
                event
                exit
                packet
                    init
                exit
            exit
            peer 192.0.2.3
                event
                exit
                packet
                    init
                exit
            exit
        exit
    exit
exit
*A:PE-1#
```

```
*A:PE-1# configure port 1/1/2 shutdown

100 2016/05/10 07:16:59.16 UTC WARNING: SNMP #2004 Base 1/1/2
"Interface 1/1/2 is not operational"

101 2016/05/10 07:16:59.16 UTC MINOR: ERING #2001 Base eth-ring-2
"Eth-Ring 2 path a changed fwd state to blocked"
*A:PE-1#
1 2016/05/10 07:16:59.16 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address Withdraw packet (msgId 95) to 192.0.2.2:0
 MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/11 Group ID = 0 cBit = 0
"

2 2016/05/10 07:16:59.16 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address Withdraw packet (msgId 82) to 192.0.2.3:0
 MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/11 Group ID = 0 cBit = 0
"

102 2016/05/10 07:16:59.18 UTC MAJOR: SVCMGR #2210 Base
"Processing of an access port state change event is finished and the status of a
ll affected SAPs on port 1/1/2 has been updated."

3 2016/05/10 07:17:02.19 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address Withdraw packet (msgId 79) from 192.0.2.3:0
 "

4 2016/05/10 07:17:02.30 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address Withdraw packet (msgId 80) from 192.0.2.3:0
 "

103 2016/05/10 07:17:02.58 UTC MINOR: ETH_CFM #2001 Base
"MEP 1/14/141 highest defect is now defRemoteCCM"
```

Operational Procedures

Operators may wish to configure rings with or without control over reversion. Reversion can be controlled by timers or the ring can be run without reversion allowing the operator to choose when the ring reverts. To change a ring topology, the **manual** or **force** switch command may be used to block a specified ring path. A ring will still address failures when run without reversion but will not automatically revert to the RPL when resources are restored. A **clear** command can be used to clear the manual or force state of a ring.

The following **tools** commands are available to control the state of paths on a ring.

```
tools perform eth-ring clear <ring-index>
```

```
tools perform eth-ring force <ring-index> path {a|b}
tools perform eth-ring manual <ring-index> path {a|b}
```

In the following output , path “b” of eth-ring 1 is manually blocked and then cleared. Initially, both ports are unblocked.

```
*A:PE-1# show eth-ring 1

===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :

Sub-Ring Type      : none

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1.1             Up/Up        normal          unblocked
  b  1/1/3    1.1             Up/Up        normal          unblocked
===============================================================================
*A:PE-1#
*A:PE-1#
*A:PE-1# tools perform eth-ring manual 1 path b
*A:PE-1# show eth-ring 1

===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : Request State: 0x7
                     Sub-Code     : 0x0
                     Status       : 0x20  ( BPR )
                     Node ID      : 4a:c4:ff:00:00:00
Defect Status      :

Sub-Ring Type      : none

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
```

```
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1.1             Up/Up        normal          unblocked
  b  1/1/3    1.1             Up/Up        normal          blocked
===============================================================================
*A:PE-1#
*A:PE-1#
*A:PE-1# tools perform eth-ring clear 1
*A:PE-1# show eth-ring 1

===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : Ethernet Ring 1
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :

Sub-Ring Type      : none

-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1.1             Up/Up        normal          unblocked
  b  1/1/3    1.1             Up/Up        normal          unblocked
===============================================================================
*A:PE-1#
```

Both the **manual** and **force** command block the path specified, however, the **manual** command fails if there is an existing forced switch or signal fail event in the ring, as seen in the following output. The **force** command will block the port regardless of any existing ring state and there can be multiple force states simultaneously on a ring on different nodes.

```
*A:PE-1# tools perform eth-ring manual 1 path b
INFO: ERMGR #1001 Not permitted -
 The switch command is not compatible to the current state (FS), effective priority
(FS) or rpl-node type (None)
*A:PE-1#
```

Conclusion

Ethernet Ring APS provides an optimal solution for designing native Ethernet services with ring topology. With sub-rings, both multiple rings and access rings increase the versatility of G.8032. G.8032 has been expanded to more of the SR platforms by allowing R-APS with slower MEPs (including CCMs intervals of 1 second). This protocol provides simple configuration, operation and guaranteed fast protection time. The implementation also has a flexible encapsulation that allows dot1Q, QinQ or PBB for the ring traffic. It could be utilized on various services such as mobile backhaul, business VPN access, aggregation and core.

G.8032 Ethernet Ring Protection Single Ring Topology

This chapter provides information about G.8032 Ethernet ring protection single ring topology.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The chapter was initially written for SROS release 8.0.R7, There have been several updates and the CLI in this edition corresponds to release 14.0.R2. This chapter describes ring protection for a single ring topology. Protection for multiple ring topologies is covered in G.8032 Ethernet Ring Protection Multiple Ring Topology.

Overview

G.8032 Ethernet ring protection is supported for data service SAPs within a regular VPLS service, a provider backbone bridging (PBB) VPLS (I/B-component) or a routed VPLS (R-VPLS). G.8032 is one of the fastest protection schemes for Ethernet networks.

ITU-T G.8032v2 specifies protection switching mechanisms and a protocol for Ethernet layer network (ETH) Ethernet rings. Ethernet rings can provide wide-area multi-point connectivity more economically due to their reduced number of links. The mechanisms and protocol defined in ITU-T G.8032v2 achieve highly reliable and stable protection and never form loops, which would negatively affect network operation and service availability. Each ring node is connected to adjacent nodes participating in the same ring using two independent paths, which use ring links (configured on ports or link aggregation groups (LAGs)). A ring link is bounded by two adjacent nodes and a port for a ring link is called a ring port. The minimum number of nodes on a ring is two.

The fundamentals of this ring protection switching architecture are:

- the principle of loop avoidance and
- the utilization of learning, forwarding, and address table mechanisms defined in the ITU-T G.8032v2 Ethernet flow forwarding function (ETH_FF) (Control plane).

Loop avoidance in the ring is achieved by guaranteeing that, at any time, traffic may flow on all but one of the ring links. This particular link is called the ring protection link (RPL) and under normal conditions this link is blocked, so it is not used for traffic. One designated node, the RPL owner, is responsible to block traffic over the one designated RPL. Under a ring failure condition, the RPL owner is responsible for unblocking the RPL, allowing the RPL to be used for traffic. The protocol ensures that even without an RPL owner defined, one link will be blocked and it operates as a “break before make protocol”, specifically the protocol guarantees that no link is restored until a different link in the ring is blocked. The other side of the RPL is configured as an RPL neighbor. An RPL neighbor blocks traffic on the link.

The event of a ring link or ring node failure results in protection switching of the traffic. This is achieved under the control of the ETH_FF functions on all ring nodes. A ring automatic protection switching (R-APS) protocol is used to coordinate the protection actions over the ring. The protection switching mechanisms and protocol supports a multi-ring/ladder network that consists of connected Ethernet rings, however, that is not covered in this chapter.

Ring Protection Mechanism

The Ring Protection protocol is based on the following building blocks:

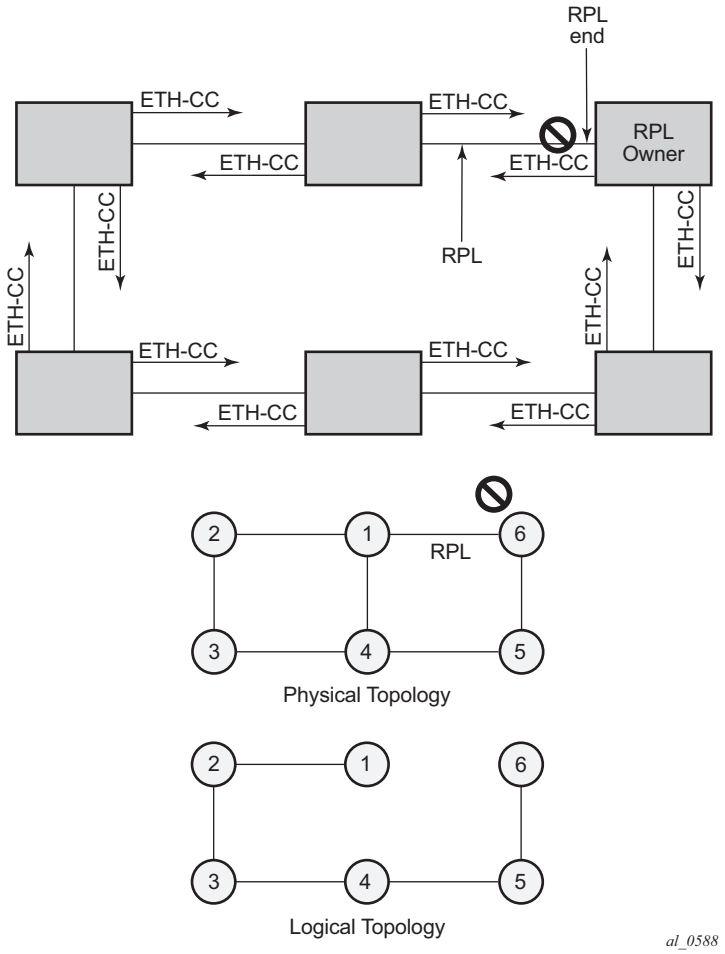
- Ring status change on failure
  - Idle -> Link failure -> Protection -> Recovery -> Idle
- Ring Control State changes
  - Idle -> Protection -> Manual Switch -> Forced Switch -> Pending
- Re-use existing ETH OAM
  - Monitoring: ETH Continuity Check messages
  - Failure Notification: Y.1731 Signal Failure
- Forwarding Database MAC Flush on ring status change
- RPL (Ring Protection Link)
  - Defines blocked link in idle status

Figure 237 shows a ring of six nodes, with the RPL owner on the top right. One link of the RPL owner is designated to be the RPL and will be blocked in order to prevent a loop. Schematics of the physical and logical topologies are also shown.

When an RPL owner and RPL end are configured, the associated link will be the RPL when the ring is fully operational and so be blocked by the RPL owner. If a different ring link fails then the RPL will be unblocked by the RPL owner. When the failed link recovers, it will initially be blocked by one of its adjacent nodes. The adjacent node sends an R-APS message across the ring to indicate the error is cleared and after a configurable time, if reversion is enabled, the RPL will revert to being blocked with all other links unblocked. This ensures that the ring topology is predictable when fully operational.

If a specific RPL owner is not configured, then the last link to become active will be blocked and the ring will remain in this state until another link fails. However, this operation makes the selection of the blocked link non-deterministic.

Figure 237 **G.8032 Operation and Topologies**

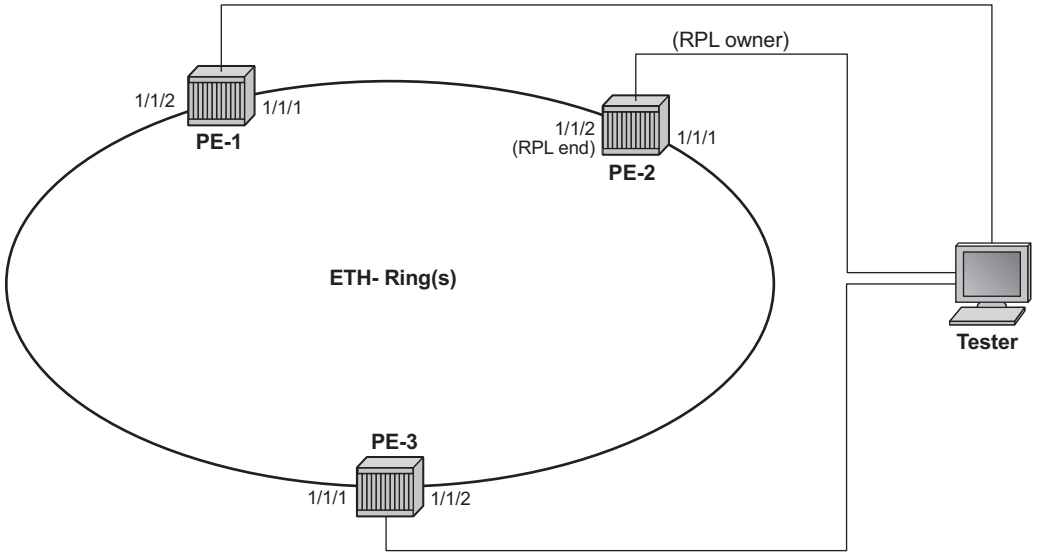


The protection protocol uses a specific control VLAN, with the associated data VLANs taking their forwarding state from the control VLAN.

Configuration

The example topology is shown in Figure 238.

Figure 238 **Example Topology**



The **eth-ring** configuration commands are as follows:

```
configure
    eth-ring <ring-index>
        ccm-hold-time { [down <down-timeout>] [up <up-timeout>] }
        compatible-version <version>
        description <description-string>
        guard-time <time>
        node-id <xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx>
        path {a|b} [ { <port-id>|<lag-id> } raps-tag <qtag>[.<qtag>] ]
            description <description-string>
            eth-cfm
                mep <mep-id> domain <md-index> association <ma-index>
                ...
            rpl-end
            shutdown
        revert-time <time>
        rpl-node {owner|nbr}
        shutdown
        sub-ring {virtual-link|non-virtual-link}
```

Parameters:

- *ring-index* — This is the number by which the ring is referenced, values: 1 to128.
- **ccm-hold-time** {[down <*down-timeout*>] [up <*up-timeout*>]}

  - down — This command specifies the timer that controls the delay between detecting that ring path is down and reporting it to the G.8032 protection module. If a non-zero value is configured, the system will wait for the time specified in the value parameter before reporting it to the G.8032 protection module. This parameter applies only to the ring path continuity check message (CCM). It does *not* apply to the ring port link state. To dampen ring port link state transitions, use the hold-time parameter from the physical member port. This is useful if the underlying path between two nodes is going across an optical system which implements its own protection.
  - up — This command specifies the timer which controls the delay between detecting that the ring path is up and reporting it to the G.8032 protection module. If a non-zero value is configured, the system will wait for the time specified in the value parameter before reporting it to the G.8032 protection module. This parameter applies only to ring path CCM. It does *not* apply to the member port link state. To dampen member port link state transitions, use the hold-time parameter from the physical member port.
  - Values: <down-timeout> : [0..5000] in deciseconds - Default: 0
    <up-timeout> : [0..5000] in deciseconds - Default: 20
    1 centisecond = 10ms
    1 decisecond = 100ms

- compatible version — This command configures eth-ring compatibility version for the G.8032 state machine and messages. The default is version 2 (ITU G.8032v2) and all 7x50 systems use version 2. If there is a need to interwork with third party devices that only support version 1, this can be set to version 1 allowing the reception of version 1 PDUs. Version 2 is encoded as 1 in the R-APS messages. Compatibility allows the reception of version 1 (encoded as 0) R-APS PDUs but, as per the G.8032 specification, higher versions are ignored on reception. For the SR/ESS, messages are always originated with version 2. Therefore if a third party switch supported version 3 (encoded as 2) or higher, interworking is also supported provided the other switch is compatible with version 2.
- description <description-string> — This configures a text string, up to 80 characters, which can be used to describe the use of the eth-ring.
- guard-time <time> — The forwarding method, in which R-APS messages are copied and forwarded at every Ethernet ring node, can result in a message corresponding to an old request, that is no longer relevant, being received by Ethernet ring nodes. Reception of an old R-APS message may result in erroneous ring state interpretation by some Ethernet ring nodes. The guard timer is used to prevent Ethernet ring nodes from acting upon outdated R-APS messages and prevents the possibility of forming a closed loop. Messages are not forwarded when the guard-timer is running.

  Values: [1..20] in deciseconds - Default: 5
  1 decisecond = 100ms

- node-id <xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx> — This allows the node identifier to be explicitly configured. By default the chassis MAC is used. It is not required in typical configurations.
- path {a|b} [{<port-id>|<lag-id>} raps-tag <qtag>[.<qtag>]] — This parameter defines the paths around the ring, of which there are two in different directions on the ring: an “a” path and a “b” path. In addition, it configures the encapsulation used for the R-APS messages on the ring. These can be either single or double tagged.
  - description <description-string> — This configures a text string, up to 80 characters, which can be used to describe the use of the path.
  - eth-cfm — Configures the associated Ethernet connectivity fault management (CFM) parameters.
    - mep <mep-id> domain <md-index> association <ma-index> — The maintenance endpoint (MEP) defined under the path is used for the G.8032 protocol messages, which are based on IEEE 802.1ag/Y.1731 CFM frames.
  - rpl-end — When configured, this path is expected to be one end of the RPL. This parameter must be configured in conjunction with the **rpl-node**.
  - shutdown — This command shuts down the path.
- revert-time <time> — This command configures the revert time for an Eth-Ring. Revert time is the time that the RPL will wait before returning to the blocked state. Configuring “no revert-time” disables reversion, effectively setting the revert-time to zero.

  Values: [60..720] in seconds - Default: 300
- rpl-node {owner|nbr} — A node can be designated as either the owner of the RPL, in which case this node is responsible for the RPL, or the nbr, in which case this node is expected to be the neighbor to the RPL owner across the RPL. The nbr is optional and is included to be compliant with the specification. This parameter must be configured in conjunction with the **rpl-end** parameter.
- shutdown — This command shuts down the ring.
- sub-ring {virtual-link|non-virtual-link} — This is beyond the scope of this chapter, as it is only required for multiple ring topologies.

Prerequisites

Logging

Create following log-id on PE-2 to see major events logged to the console on PE-2. This is an optional step; alternatively, log 99 can be consulted.

```
configure
    log
        log-id 1
            from main
            to console
        exit
```

Configure Encapsulation for Ring Ports

To configure R-APS, there should be at least two VPLS services for one Eth-Ring instance, one for the control channel and the other (or more) for data channel(s). The control channel is used for R-APS signaling while data channel is for user data traffic. The state of the data channels is inherited from the state of the control channel.

- Eth-Ring needs R-APS tags to send and receive G.8032 signaling messages. To configure a control channel, an access SAP configuration is required on each path a/b port. The SAP configuration follows that of the port and must be either *dot1q* or *qinq*, consequently the control and data packets are either single tagged or double tagged. It is also possible to have the control VPLS using single tagged frames with the data VPLSs using double tagged framed; this requires the system to be configured with the **new-qinq-untagged-sap** parameter (**configure system ethernet new-qinq-untagged-sap**), with the ring path raps-tags and control VPLS SAPs configured as qtag.0, and the data VPLSs configured as QinQ SAPs.

  In this example, single tags are used so the commands for ring node PE-1 are:

```
*A:PE-1# configure port 1/1/[1..2] ethernet mode access
*A:PE-1# configure port 1/1/[1..2] ethernet encap-type dot1q
*A:PE-1# configure port 1/1/[1..2] no shutdown
```

Configure ETH-CFM

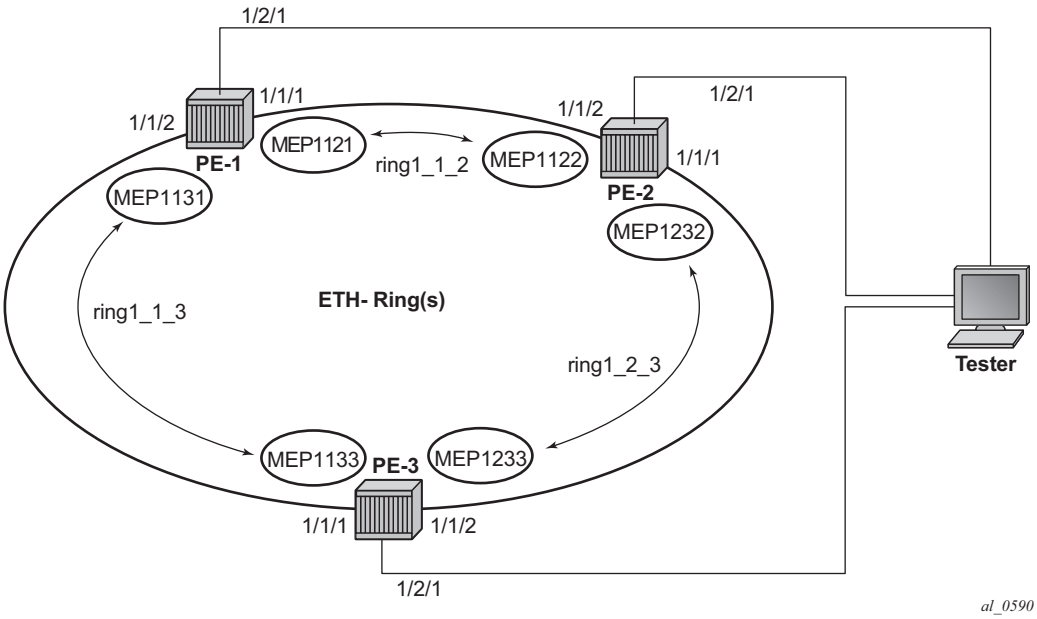
Ethernet Ring requires Eth-CFM domains, associations and MEPs being configured. The domain format should be none and association name should be ITU-T carrier code- based (ICC-based - Y.1731). The minimum CCM interval for the 7x50 is 10ms. The eth-ring MEP requires a CCM interval, such as 10ms, 100ms, or 1s, to be configured.

The MEPs used for R-APS control normally will have CCM configured on the control channel path MEPs for failure detection. Alternatively, detecting a failure of the ring may be achieved by running Ethernet in the First Mile (EFM) at the port level if CCM is not possible at 10ms, 100ms, or 1s. Loss-of-signal, in conjunction with other OAM, is applicable only when the nodes are directly connected.

To omit the failure detecting CCMs, it would be necessary to remove the *ccm-enable* from under the path MEPs and to remove the *remote-mepids* from under the *eth-cfm associations* on all nodes.

Figure 239 shows the Ethernet CFM configuration used here.

Figure 239 **Ethernet CFM Configuration**



The configuration of each node is as follows.

PE-1:

```
configure
```

```
    eth-cfm
        domain 1 format none level 3
            association 1 format icc-based name "ring1_1_2"
                ccm-interval 1
                remote-mepid 1122
            exit
            association 2 format icc-based name "ring1_1_3"
                ccm-interval 1
                remote-mepid 1133
            exit
        exit
    exit
```

PE-2:

```
configure
    eth-cfm
        domain 1 format none level 3
            association 1 format icc-based name "ring1_2_3"
                ccm-interval 1
                remote-mepid 1233
            exit
            association 2 format icc-based name "ring1_1_2"
                ccm-interval 1
                remote-mepid 1121
            exit
        exit
    exit
```

PE-3:

```
configure
    eth-cfm
        domain 1 format none level 3
            association 1 format icc-based name "ring1_1_3"
                ccm-interval 1
                remote-mepid 1131
            exit
            association 2 format icc-based name "ring1_2_3"
                ccm-interval 1
                remote-mepid 1232
            exit
        exit
    exit
```

Configure Eth-Ring

Two paths should be configured to form a ring. In this example, VLAN tag 1 is used as control channel for R-APS signaling in the ring.

PE-1:

```
configure
    eth-ring 1
        path a 1/1/1 raps-tag 1
            eth-cfm
                mep 1121 domain 1 association 1
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 1
            eth-cfm
                mep 1131 domain 1 association 2
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

It is mandatory to configure a MEP in the path context, otherwise the following error will be displayed:

```
*A:PE-1>config>eth-ring>path# no shutdown
INFO: ERMGR #1001 Not permitted - must configure eth-cfm MEP first
*A:PE-1>config>eth-ring>path#
```

While MEPs are mandatory, enabling CCM on the MEP in the pathcontext as a failure detection mechanism is optional.

In order to define the RPL, node PE-2 has been configured as the RPL owner and path “b” as the RPL end. The link between nodes PE-1 and PE-2 will be the RPL with node PE-2 blocking that link when the ring is fully operational.

PE-2:

```
configure
    eth-ring 1
        revert-time 60
        rpl-node owner
        path a 1/1/1 raps-tag 1
            eth-cfm
                mep 1232 domain 1 association 1
                    ccm-enable
                    control-mep
                    no shutdown
```

```
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 1
            rpl-end
            eth-cfm
                mep 1122 domain 1 association 2
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

It is not permitted to configure a path as an RPL end without having configured the node on this ring to be either the RPL *owner* or *nbr* otherwise the following error message is reported.

```
*A:PE-2>config>eth-ring# path b rpl-end
INFO: ERMGR #1001 Not permitted - path-type rpl-end is not consistent with eth-ring
'rpl-node' type
*A:PE-2>config>eth-ring#
```

PE-3:

```
configure
    eth-ring 1
        path a 1/1/1 raps-tag 1
            eth-cfm
                mep 1133 domain 1 association 1
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        path b 1/1/2 raps-tag 1
            eth-cfm
                mep 1233 domain 1 association 2
                    ccm-enable
                    control-mep
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

Until the Ethernet Ring instance is attached to the service (VPLS in this case), the ring operational status is down and the forwarding status of each port is blocked. This prevents operator from creating a loop by mis-configuration. This state can be seen on ring node PE-1 as follows:

```
*A:PE-1# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : (Not Specified)
Admin State        : Up                  Oper State       : Down
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds    RPL Node         : rplNone
Max Revert Time    :  300 seconds        Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds   CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : Request State: 0xB
                     Sub-Code     : 0x0
                     Status       : 0x20  ( BPR )
                     Node ID      : 4a:c4:ff:00:00:00
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1               Up/Down      normal          blocked
  b  1/1/2    1               Up/Down      normal          blocked
===============================================================================
*A:PE-1#
```

Configure Control Channel VPLS Service

Paths a and b defined in the eth-ring must be added as SAPs into a VPLS service (standard VPLS in this example) using the *eth-ring* parameter. The SAP encapsulation values must match the values of the *raps-tag* configured for the associated path.

G.8032 uses the same raps-tag value on all nodes on the ring, as configured in this example. However, the SROS implementation relaxes this constraint by requiring the tag to match only on adjacent nodes.

PE-1:

```
configure
    service
        vpls 1 customer 1 create
            sap 1/1/1:1 eth-ring 1 create
            exit
            sap 1/1/2:1 eth-ring 1 create
```

```
            exit
            no shutdown
        exit
```

PE-2:

```
configure
    service
        vpls 1 customer 1 create
            sap 1/1/1:1 eth-ring 1 create
            exit
            sap 1/1/2:1 eth-ring 1 create
            exit
            no shutdown
        exit
```

PE-3:

```
configure
    service
        vpls 1 customer 1 create
            sap 1/1/1:1 eth-ring 1 create
            exit
            sap 1/1/2:1 eth-ring 1 create
            exit
            no shutdown
        exit
```

A normal SAP or SDP can be added in a control channel VPLS on condition the *eth-ring* parameter is present. Trying to add a SAP or SDP without this parameter into a control channel VPLS will result in the following message being displayed.

```
*A:PE-1# configure service vpls 1 sap 1/2/1:1 create
MINOR: SVCMGR #1321 Service contains an Ethernet ring control SAP
*A:PE-1#
```

Now the Eth-Ring is operationally up and the RPL is blocking successfully on ring node PE-2 port 1/1/2, as expected from the RPL owner/end configuration.

An overview of all of the rings can be shown using the following commands, in this case on node PE-2.

First, the ETH ring status is shown.

```
*A:PE-2# show eth-ring status
===============================================================================
Ethernet Ring (Status information)
===============================================================================
Ring   Admin  Oper       Path Information                MEP Information
ID     State  State  Path         Tag       State     Ctrl-MEP CC-Intvl Defects
```

```
-------------------------------------------------------------------------------
1      Up     Up     a - 1/1/1        1     Up        Yes      1         -----
                     b - 1/1/2        1     Up        Yes      1         -----
===============================================================================
Ethernet Tunnel MEP Defect Legend:
R = Rdi, M = MacStatus, C = RemoteCCM, E = ErrorCCM, X = XconCCM
*A:PE-2#
```

The ring and path forwarding states is shown with following command.

```
*A:PE-2# show eth-ring
===============================================================================
Ethernet Rings (summary)
===============================================================================
Ring Int  Admin Oper            Paths Summary                       Path States
ID   ID   State State                                               a     b
-------------------------------------------------------------------------------
1    -    Up    Up    a - 1/1/1        1     b - 1/1/2        1     U     B
===============================================================================
Ethernet Ring Summary Legend:   B - Blocked    U - Unblocked
*A:PE-2#
```

Specific ring information can be shown on each node separately, as follows.

PE-1:

```
*A:PE-1# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : (Not Specified)
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c4:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplNone
Max Revert Time    :  300 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1               Up/Up        normal          unblocked
  b  1/1/2    1               Up/Up        normal          unblocked
===============================================================================
*A:PE-1#
```

PE-2:

```
*A:PE-2# show eth-ring 1
```

```
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : (Not Specified)
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c5:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplOwner
Max Revert Time    :   60 seconds       Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : Request State: 0x0
                     Sub-Code     : 0x0
                     Status       : 0xA0  ( RB BPR )
                     Node ID      : 4a:c5:ff:00:00:00
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1              Up/Up        normal          unblocked
  b  1/1/2    1              Up/Up        rplEnd          blocked
===============================================================================
*A:PE-2#
```

Node PE-2 is the RPL owner and port 1/1/2 is the RPL end. The *revert-time* shows the configured value.

When a revert is pending, the “Time to Revert” will show the number of seconds remaining before the revert occurs, as follows:

```
*A:PE-2# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : (Not Specified)
Admin State        : Up                 Oper State       : Up
Node ID            : 4a:c5:ff:00:00:00
Guard Time         :    5 deciseconds   RPL Node         : rplOwner
Max Revert Time    :   60 seconds       Time to Revert   :   43 seconds
CCM Hold Down Time :    0 centiseconds  CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type            Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1              Up/Up        normal          unblocked
  b  1/1/2    1              Up/Up        rplEnd          unblocked
===============================================================================
*A:PE-2#
```

On reversion, the following console message is logged.

```
68 2016/05/02 11:22:50.87 UTC MINOR: ERING #2001 Base eth-ring-1
"Eth-Ring 1 path b changed fwd state to blocked"
```

PE-3:

```
*A:PE-3# show eth-ring 1
===============================================================================
Ethernet Ring 1 Information
===============================================================================
Description        : (Not Specified)
Admin State        : Up                  Oper State       : Up
Node ID            : 4a:c6:ff:00:00:00
Guard Time         :    5 deciseconds    RPL Node         : rplNone
Max Revert Time    :  300 seconds        Time to Revert   : N/A
CCM Hold Down Time :    0 centiseconds   CCM Hold Up Time :   20 deciseconds
Compatible Version : 2
APS Tx PDU         : N/A
Defect Status      :
Sub-Ring Type      : none
-------------------------------------------------------------------------------
Ethernet Ring Path Summary
-------------------------------------------------------------------------------
Path Port     Raps-Tag     Admin/Oper      Type             Fwd State
-------------------------------------------------------------------------------
  a  1/1/1    1               Up/Up        normal           unblocked
  b  1/1/2    1               Up/Up        normal           unblocked
===============================================================================
*A:PE-3#
```

Finally, the details of an individual path can be shown.

```
*A:PE-2# show eth-ring 1 path a
===============================================================================
Ethernet Ring 1 Path Information
===============================================================================
Description        : (Not Specified)
Port               : 1/1/1               Raps-Tag         : 1
Admin State        : Up                  Oper State       : Up
Path Type          : normal              Fwd State        : unblocked
                                         Fwd State Change : 05/02/2016 11:17:59
Last Switch Command: noCmd
APS Rx PDU         : Request State: 0x0
                     Sub-Code     : 0x0
                     Status       : 0x20  ( BPR )
                     Node ID      : 4a:c6:ff:00:00:00
===============================================================================
*A:PE-2#
```

Configure User Data Channel VPLS Service

The user data channels are created on a separate VPLS, VPLS 100 in the example. Tag 100 and VPLS 100 are used here. The ring data channels must be on the same ports as the corresponding control channels configured above. The access into the data services can use SAPs and/or SDPs.

PE-1:

```
configure
    service
        vpls 100 customer 1 create
            sap 1/1/1:100 eth-ring 1 create
            exit
            sap 1/1/2:100 eth-ring 1 create
            exit
            sap 1/2/1:100 create
            exit
            no shutdown
        exit
    exit
```

PE-2:

```
configure
    service
        vpls 100 customer 1 create
            sap 1/1/1:100 eth-ring 1 create
            exit
            sap 1/1/2:100 eth-ring 1 create
            exit
            sap 1/2/1:100 create
            exit
            no shutdown
        exit
    exit
```

PE-3:

```
configure
    service
        vpls 100 customer 1 create
            sap 1/1/1:100 eth-ring 1 create
            exit
            sap 1/1/2:100 eth-ring 1 create
            exit
            sap 1/2/1:100 create
            exit
            no shutdown
        exit
    exit
```

All of the SAPs which are configured to use ETH rings can be shown, using PE-1 as an example.

```
*A:PE-1# show service sap-using eth-ring
===============================================================================
Service Access Points (Ethernet Ring)
===============================================================================
SapId              SvcId             Eth-Ring Path Admin Oper  Blocked Control/
                                               State State         Data
-------------------------------------------------------------------------------
1/1/1:1            10                1        a    Up    Up    No      Ctrl
1/1/2:1            10                1        b    Up    Up    No      Ctrl
1/1/1:100          100               1        a    Up    Up    No      Data
1/1/2:100          100               1        b    Up    Up    No      Data
-------------------------------------------------------------------------------
Number of SAPs : 4
===============================================================================
*A:PE-1#
```

Debug

To see an example of the console messages on a ring failure, when the unblocked port (1/1/1) on node PE-2 is shut down, the following messages are displayed.

```
*A:PE-2# configure port 1/1/1 shutdown
15 2016/05/02 09:21:11.15 UTC WARNING: SNMP #2004 Base 1/1/1
"Interface 1/1/1 is not operational"
16 2016/05/02 09:21:11.15 UTC MINOR: ERING #2001 Base eth-ring-1
"Eth-Ring 1 path a changed fwd state to blocked"
17 2016/05/02 09:21:11.15 UTC MINOR: ERING #2001 Base eth-ring-1
"Eth-Ring 1 path b changed fwd state to unblocked"
*A:PE-2#
18 2016/05/02 09:21:11.16 UTC MAJOR: SVCMGR #2210 Base
"Processing of an access port state change event is finished and the status of a
ll affected SAPs on port 1/1/1 has been updated."
19 2016/05/02 09:21:14.85 UTC MINOR: ETH_CFM #2001 Base
"MEP 1/1/1232 highest defect is now defRemoteCCM"
```

For troubleshooting, the **tools dump eth-ring** <*ring-index*> command displays path information, the internal state of the control protocol, related statistics information and up to the last 20 protocol events (including messages sent and received, and the expiration of timers). An associated parameter *clear* exists, clearing the event information in this output when the command is entered. The following is an example of the output on node PE-2 with port 1/1/1 active.

```
*A:PE-2# tools dump eth-ring 1
ringId 1 (Up/Up): numPaths 2 nodeId 4a:c5:ff:00:00:00
 SubRing: none (interconnect ring 0, propagateTc  No), Cnt 0
  path-a, port 1/1/1 (Up), tag 1.0(Up) status (Up/Up/Fwd)
       cc (Dn/Up): Cnt 5/5 tm 000 00:48:02.730/000 00:48:04.430
       state: Cnt 8 B/F 000 00:47:58.630/000 00:48:11.380, flag: 0x0
```

```
  path-b, port 1/1/2 (Up), tag 1.0(Up) status (Up/Up/Blk)
      cc (Dn/Up): Cnt 3/3 tm 000 00:41:33.740/000 00:41:33.960
      state: Cnt 11 B/F 000 00:49:11.680/000 00:47:58.630, flag: 0x0
  FsmState=  IDLE, Rpl = Owner, revert = 60 s, guard = 5 ds
    Defects =
    Running Timers = PduReTx
    lastTxPdu = 0x00a0 Nr(RB )
    path-a Normal, RxId(I)= 4a:c6:ff:00:00:00, rx= v1-0x0020 Nr, cmd= None
    path-b Rpl, RxId(I)= 4a:c6:ff:00:00:00, rx= v1-0x0020 Nr, cmd= None
  DebugInfo:  aPathSts 9, bPathSts 5, pm (set/clr) 0/0, txFlush 0
    RxRaps: ok 40 nok 0 self 1756, TmrExp - wtr 4(2), grd 6, wtb 0
    Flush: cnt 14 (9/5/0) tm 000 00:49:11.680-000 00:49:11.680 Out/Ack 0/1
    RxRawRaps: aPath 954 bPath 976 vPath 0
    Now: 000 01:13:50.160 , softReset: No - noTx 0
  Seq Event  RxInfo(Path: NodeId-Bytes)
             state:TxInfo (Bytes)             Dir  pA  pB        Time
  === ===== ============================  ===== === === ================
  004   bDn
              PROT  : 0xb020  Sf            TxF-> Fwd Blk 000 00:41:33.740
  005   pdu B: 4a:c4:ff:00:00:00-0x0000 Nr
              PROT  : 0xb020  Sf            Rx<-- Fwd Blk 000 00:41:33.840
  006   pdu A: 4a:c4:ff:00:00:00-0x0000 Nr
              PROT  : 0xb020  Sf            Rx<-- Fwd Blk 000 00:41:33.840
  007   pdu B: 4a:c4:ff:00:00:00-0xb040 Sf(DNF)
              PROT  : 0xb020  Sf            Rx<-- Fwd Blk 000 00:41:33.960
  008   pdu A: 4a:c4:ff:00:00:00-0xb040 Sf(DNF)
              PROT  : 0xb020  Sf            Rx<-- Fwd Blk 000 00:41:33.960
  009   bUp
              PEND-G: 0x0020  Nr            Tx--> Fwd Blk 000 00:41:35.980
  010   pdu B: 4a:c4:ff:00:00:00-0x0000 Nr
              PEND  : 0x0020  Nr            Rx<-- Fwd Blk 000 00:41:36.800
  011   pdu A: 4a:c4:ff:00:00:00-0x0000 Nr
              PEND  : 0x0020  Nr            Rx<-- Fwd Blk 000 00:41:36.800
  012  xWtr
              IDLE  : 0x00e0  Nr(RB DNF)    Tx--> Fwd Blk 000 00:42:50.680
  013   aDn
              PROT  : 0xb000  Sf            TxF-> Blk Fwd 000 00:47:58.630
  014   pdu B: 4a:c6:ff:00:00:00-0xb020 Sf
              PROT  : 0xb000  Sf            RxF<- Blk Fwd 000 00:48:01.440
  015   pdu A: 4a:c6:ff:00:00:00-0x0020 Nr
              PROT  : 0xb000  Sf            Rx<-- Blk Fwd 000 00:48:05.780
  016   pdu B: 4a:c6:ff:00:00:00-0x0020 Nr
              PROT  : 0xb000  Sf            Rx<-- Blk Fwd 000 00:48:05.780
  017   aUp
              PEND-G: 0x0000  Nr            Tx--> Blk Fwd 000 00:48:06.380
  018   pdu A: 4a:c6:ff:00:00:00-0x0020 Nr
              PEND-G: 0x0000  Nr            Rx<-- Blk Fwd 000 00:48:06.380
  019   pdu B: 4a:c6:ff:00:00:00-0x0020 Nr
              PEND-G: 0x0000  Nr            Rx<-- Blk Fwd 000 00:48:06.380
  000   pdu A: 4a:c6:ff:00:00:00-0x0020 Nr
              PEND  : 0x0000  Nr            Rx<-- Blk Fwd 000 00:48:11.380
  001   pdu
              PEND  :                       ----- Fwd Fwd 000 00:48:11.380
  002   pdu B: 4a:c6:ff:00:00:00-0x0020 Nr
              PEND  :                       Rx<-- Fwd Fwd 000 00:48:11.390
  003  xWtr
              IDLE  : 0x00a0  Nr(RB )       TxF-> Fwd Blk 000 00:49:11.68

*A:PE-2#
```

Conclusion

Ethernet Ring APS provides optimal solution for designing native Ethernet services with ring topology. This protocol provides simple configuration, operation and guaranteed fast protection time. SROS also has a flexible encapsulation that allows dot1Q, qinq or PBB for the ring traffic. It could be utilized for various services such as mobile backhaul, business VPN access, aggregation and core.

Layer 2 Services and EVPN

In This Section

This section provides configuration information for the following topics:

- BGP Multi-Homing for VPLS Networks
- BGP Virtual Private Wire Services
- BGP VPLS
- EVPN for MPLS Tunnels
- EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS)
- EVPN for PBB over MPLS (PBB-EVPN)
- EVPN for VXLAN Tunnels (Layer 2)
- EVPN for VXLAN Tunnels (Layer 3)
- EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services
- Fully Dynamic VSD Integration Model
- Inter-AS Model C for VLL
- LDP VPLS using BGP-Auto Discovery
- Multi-Chassis Endpoint for VPLS Active/Standby Pseudowire
- Multi-Segment Pseudowire Routing
- P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services
- PBB-Epipe
- PBB-VPLS
- Shortest Path Bridging for MAC

BGP Multi-Homing for VPLS Networks

This chapter describes BGP Multi-Homing (BGP-MH) for VPLS network configurations.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

Initially, the information in this chapter was based on SR OS release 8.0.R5, with additions for release 9.0.R1. The CLI in this edition corresponds to SR OS release 14.0.R2.

Summary

SR OS supports the use of Border Gateway Protocol Multi-Homing for VPLS (hereafter called BGP-MH). BGP-MH is described in draft-ietf-l2vpn-vpls-multihoming, *BGP based Multi-homing in Virtual Private LAN Service*, and provides a network-based resiliency mechanism (no interaction from the provider edge routers (PEs) to multi-tenant units/customer equipment (MTU/CE)) that can be applied on service access points (SAPs) or network (pseudowires) topologies. The BGP-MH procedures will run between the PEs and will provide a loop-free topology from the network perspective (only one logical active path will be provided per VPLS among all the objects SAPs or pseudowires which are part of the same Multi-Homing site).

Each multi-homing site connected to two or more peers is represented by a site-id (2-bytes long) which is encoded in the BGP MH Network Layer Reachability Information (NLRI). The BGP peer holding the active path for a particular multi-homing site will be named as the Designated Forwarder (DF), whereas the rest of the BGP peers participating in the BGP MH process for that site will be named as non-DF and will block the traffic (in both directions) for all the objects belonging to that multi-homing site.

BGP MH uses the following rules to determine which PE is the DF for a particular multi-homing site:

1. A BGP MH NLRI with D flag = 0 (multi-homing object up) always takes precedence over a BGP MH NLRI with D flag = 1 (multi-homing object down). If there is a tie, then:
2. The BGP MH NLRI with the highest BGP LP (Local Preference) wins. If there is a tie, then:
3. The BGP MH NLRI issued from the PE with the lowest PE ID (system address) wins.

The main advantages of using BGP-MH as opposed to other resiliency mechanisms for VPLS are:

- Flexibility: BGP-MH uses a common mechanism for access and core resiliency. The designer has the flexibility of using BGP-MH to control the active/standby status of SAPs, spoke SDPs, Split Horizon Groups (SHGs) or even mesh SDP bindings.
- The standard protocol is based on BGP, a standard, scalable and well-known protocol.
- Specific benefits at the access:
  - It is network-based, independent of the customer CE and, as such, it does not need any customer interaction to determine the active path. Consequently the operator will spend less effort on provisioning and will minimize both operation costs and security risks (in particular, this removes the requirement for spanning tree interaction between the PE and CE).
  - Easy load balancing per service (no service fate-sharing) on physical links.
- Specific benefits in the core:
  - It is a network-based mechanism, independent of the MTU resiliency capabilities and it does not need MTU interaction, therefore operational advantages are achieved as a result of the use of BGP-MH: less provisioning is required and there will be minimal risks of loops. In addition, simpler MTUs can be used.
  - Easy load balancing per service (no service fate-sharing) on physical links.

– Less control plane overhead: there is no need for an additional protocol running the pseudowire redundancy when BGP is already used in the core of the network. BGP-MH just adds a separate NLRI in the L2-VPN family (AFI=25, SAFI=65).

The objective of this section is to provide the required guidelines to configure and troubleshoot BGP-MH for VPLS

Knowledge of the LDP/BGP VPLS (RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*, and RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*) architecture and functionality is assumed throughout this document. For further information refer to the relevant Nokia documentation.

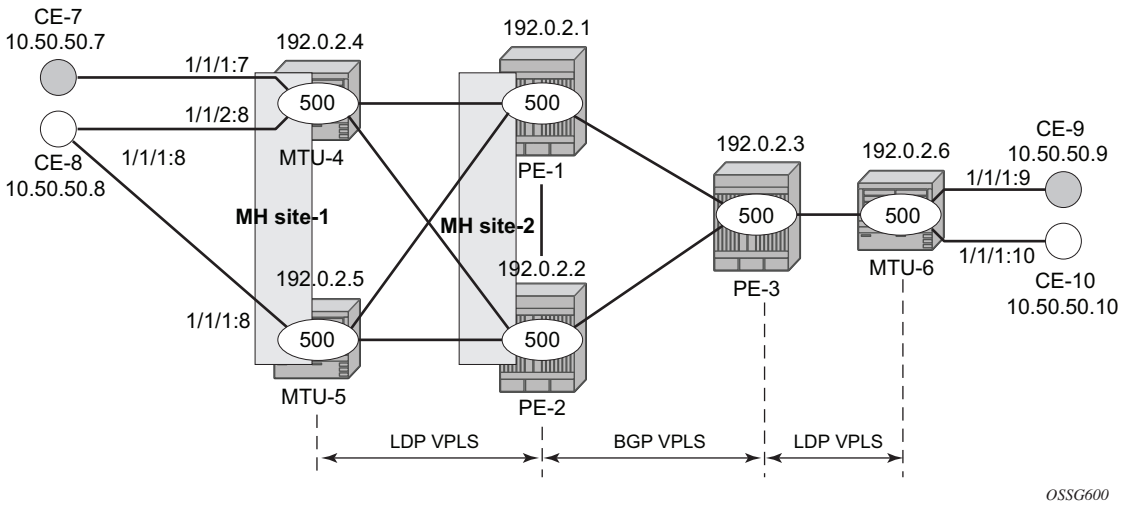
Overview

The following example setup will be used throughout the rest of the chapter.

The initial configuration includes:

- IGP — ISIS, Level 2 on all routers; area 49.0001
- RSVP-TE for transport tunnels
- FRR in the core
- No protection at the access.

Figure 240 **Example Topology**



The setup consists of a three core nodes (PE-1, PE-2 and PE-3) and three multi-tenant unit (MTU) nodes connected to the core. The VPLS service 500 is configured on all the six nodes with the following characteristics:

The VPLS service 500 is configured on all the six nodes with the following characteristics:

- The core VPLS instances are connected by a full mesh of BGP-signaled pseudowires (that is, pseudowires among PE-1, PE-2 and PE-3 will be signaled by BGP VPLS).
- As depicted in the Figure 240, the MTUs are connected to the BGP VPLS core by TLDP pseudowires. While MTU-6 is connected to PE-3 by a single pseudowire, MTU-4 and MTU-5 are dual-homed to PE-1 and PE-2. The following resiliency mechanisms are used on the dual-homed MTUs:
  - MTU-4 is dual-connected to PE-1 and PE-2 by an active/standby pseudowire (A/S pseudowire hereafter).
  - MTU-5 is dual-connected to PE-1 and PE-2 by two active pseudowires, one of them being blocked by BGP MH running between PE-1 and PE-2. The PE-1 and PE-2 pseudowires, set up from MTU-5, will be part of the BGP MH site MH-site-2.
  - MTU-4 and MTU-5 are running BGP MH, being SHG site-1 and sap 1/1/1:8 on MTU-5 part of the same BGP MH site, MH-site-1.
- The CEs are connected to the network in the following way:
  - CE-7, CE-9 and CE-10 are single-connected to the network
  - CE-8 is dual connected to MTU-4 and MTU-5.
  - CE-7 and CE-8 are part of the split-horizon-group (SHG) site-1(SAPs 1/1/4:500 and 1/1/3:500 on MTU-4). Assume that CE-7 and CE-8 have a backdoor link between them so that when MTU-5 is elected as DF, CE1 does not get isolated. This configuration high-lights the use of a SHG within a site configuration.

For each BGP MH site, MH-site-1 and MH-site-2, the BGP MH process will elect a DF, blocking the site objects for the non-DF nodes. In other words, based on the specific configuration explained throughout the section:

- For MH-site-1, MTU-4 will be elected as the DF. The non-DF-MTU-5 will block the SAP 1/1/1:8.
- For MH-site-2, PE-1 will be elected as the DF. The non-DF PE-1 will block the spoke-SDP to MTU-5.

Configuration

This section describes all the relevant configuration tasks for the setup shown in Figure 240. The appropriate associated IP/MPLS configuration is out of the scope of this section. In this particular example the following protocols will be configured beforehand:

- ISIS-TE as IGP with all the interfaces being level-2 (OSPF-TE could have been used instead).
- RSVP-TE as the MPLS protocol to signal the transport tunnels (LDP could have been used instead).
- LSPs between core PEs will be Fast Re-Route protected (facility bypass tunnels) whereas LSP tunnels between MTUs and PEs will not be protected.

Note: The designer can choose whether to protect access link failures by means of MPLS FRR or A/S pseudowire or BGP MH. While FRR provides a faster convergence (around 50ms) and stability (it does not impact on the service layer, therefore, link failures do not trigger MAC flush and flooding) some interim inefficiencies can be introduced compared to A/S pseudowire or BGP MH.

Once the IP/MPLS infrastructure is up and running, the specific service configuration including the support for BGP MH can begin.

Global BGP Configuration

BGP is used in this configuration guide for these purposes:

a. Auto-discovery and signaling of the pseudowires in the core, as per RFC 4761.
b. Exchange of multi-homing site NLRIs and redundancy handling from MTU-5 to the core.
c. Exchange of multi-homing site NLRIs and redundancy handling at the access for CE-7/CE-8.

A BGP route reflector (RR), PE-3, is used for the reflection of BGP updates corresponding to the preceding uses **a** and **b.**

A direct peering is established between MTU-4 and MTU-5 for use **c**. The same RR could have been used for the three cases, however, like in this example, the designer may choose to have a direct BGP peering between access devices. The reasons for this are:

- By having a direct BGP peering between MTU-4 and MTU-5, the BGP updates do not have to travel back and forth.
- On MTU-4 and MTU-5, BGP is exclusively used for multi-homing, therefore there will not be more BGP peers for either MTUs and a RR adds nothing in terms of control plane scalability.

On all nodes, the autonomous-system number must be configured.

```
configure router autonomous-system 65000
```

The following CLI output shows the global BGP configuration required on MTU-4. The 192.0.2.5 address will be replaced by the corresponding peer or the RR system address for PE-1 and PE-2.

```
*A:MTU-4# configure router
        bgp
            family l2-vpn
            rapid-withdrawal
            rapid-update l2-vpn
            group "Multi-Homing"
                neighbor 192.0.2.5
                    peer-as 65000
                exit
            exit
```

In this example, PE-3 is the BGP RR, therefore its BGP configuration will contain a cluster with all its peers included (PE-1 and PE-2):

```
*A:PE-3# configure router
        bgp
            family l2-vpn
            rapid-withdrawal
            rapid-update l2-vpn
            group "internal"
                cluster 1.1.1.1
                neighbor 192.0.2.1
                    peer-as 65000
                exit
                neighbor 192.0.2.2
                    peer-as 65000
                exit
            exit
```

The relevant BGP commands for BGP-MH are in bold. Some considerations about those:

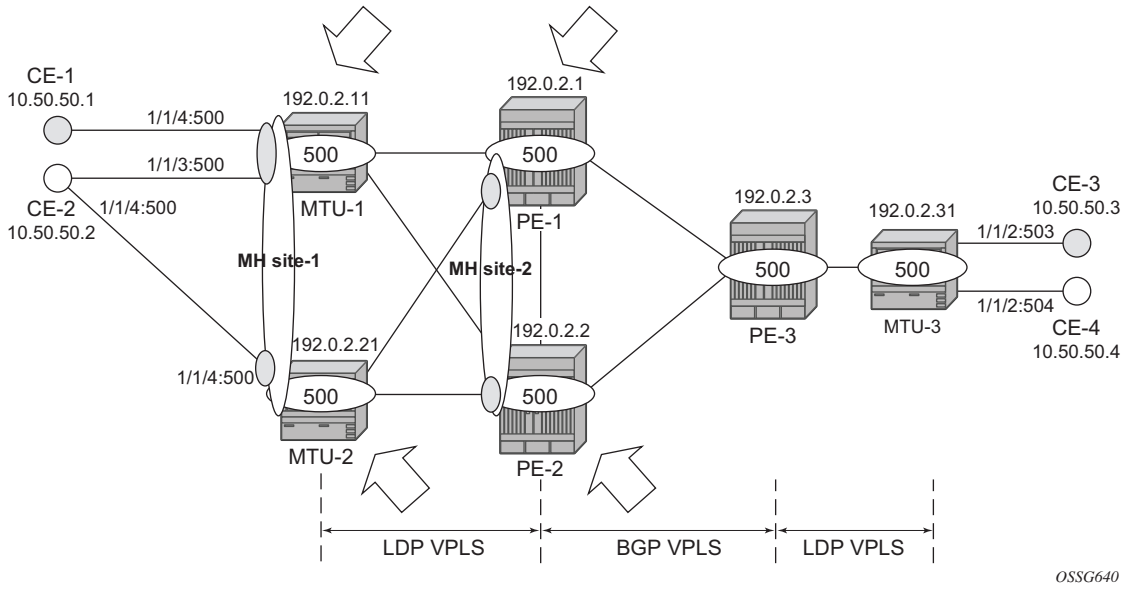
- It is required to specify **family l2-vpn** in the BGP configuration. That statement will allow the BGP peers to agree on the support for the family AFI=25 (Layer 2 VPN), SAFI=65 (VPLS). This family is used for BGP VPLS as well as for BGP MH and BGP AD.

- The **rapid-update l2-vpn** statement allows BGP MH to send BGP updates immediately after detecting link failures, without having to wait for the Minimum Route Advertisement Interval (MRAI) to send the updates in batches. This statement is required to guarantee a fast convergence for BGP MH.
- Optionally, rapid-withdrawal can also be added. Note that, in the context of BGP MH, this command is only useful if a particular multi-homing site is cleared. In that case, a BGP withdrawal is sent immediately without having to wait for the MRAI. A multi-homing site is cleared when the BGP MH site is removed or even the entire VPLS service.

Service Level Configuration

Once the IP/MPLS infrastructure is configured, including BGP, this section shows the configuration required at service level (VPLS 500). The focus is on the nodes involved on BGP MH, that is, MTU-4, MTU-5, PE-1 and PE-2. These nodes are highlighted in Figure 241.

Figure 241 **Nodes Involved in BGP MH**



Core PE Service Configuration

The following CLI excerpt shows the service level configuration on PE-1. The import/export policies configured on the PE nodes are identical:

```
configure
    router
        policy-options
            begin
            community "comm_core" members "target:65000:500"
            policy-statement "vsi500_export"
                entry 10
                    action accept
                        community add "comm_core"
                    exit
                exit
            exit
            policy-statement "vsi500_import"
                entry 10
                    from
                        community "comm_core"
                        family l2-vpn
                    exit
                    action accept
                    exit
                exit
                default-action reject
            exit
            commit
```

The configuration of the SDPs, PW template, and VPLS on PE-1 is as follows:

```
configure
    service
        sdp 12 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            path-mtu 8000
            no shutdown
        exit
        sdp 13 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-1-PE-3"
            path-mtu 8000
            no shutdown
        exit
        sdp 14 mpls create
            far-end 192.0.2.4
            lsp "LSP-PE-1-MTU-4"
            path-mtu 8000
            no shutdown
        exit
        sdp 15 mpls create
            far-end 192.0.2.5
            lsp "LSP-PE-1-MTU-5"
            path-mtu 8000
            no shutdown
        exit
        pw-template 500 use-provisioned-sdp create
```

```
exit
vpls 500 customer 1 create
    bgp
        route-distinguisher 65000:501
        vsi-export "vsi500_export"
        vsi-import "vsi500_import"
        pw-template-binding 500 split-horizon-group "CORE"
        exit
    exit
    bgp-vpls
        max-ve-id 65535
        ve-name 501
            ve-id 501
        exit
        no shutdown
    exit
    site "MH-site-2" create
        site-id 2
        spoke-sdp 15:500
        no shutdown
    exit
    spoke-sdp 14:500 create
    exit
    spoke-sdp 15:500 create
    exit
    no shutdown
exit
```

The following are general comments about the configuration of service 500:

- As seen in the preceding CLI output for PE-1, there are four provisioned SDPs that the service VPLS 500 will use in this example. SDP 14 and SDP 15 are tunnels over which the TLDP FEC128 pseudowires for service 500 will be carried (according to RFC 4762), whereas SDP 12 and SDP 13 are the tunnels for the core BGP pseudowires (based on RFC 4761).
- The BGP context provides the general service BGP configuration that will be used by BGP VPLS and BGP MH:
  - Route distinguisher (notation chosen is based on <AS_number:500 + node_id>)
  - VSI export policies are used to add the export route-targets included in all the BGP updates sent to the BGP peers.
  - VSI import policies are used to control the NLRIs accepted in the RIB, normally based on the route-targets.
  - Both, VSI-export and VSI-import policies can be used to modify attributes like the Local-Preference (LP) that will be used to influence the BGP MH Designated Forwarder (DF) election (LP is the second rule in the BGP MH election process, as previously discussed). The use of these policies will be described later in the section.

– The **pw-template-binding** command maps the previously defined pw-template 500 to the **split-horizon-group "CORE"**. In this way, all the BGP-signaled pseudowires will be part of this split-horizon-group. Although not shown in this example, the **pw-template-binding** command can also be used to instantiate pseudowires within different split-horizon-groups, based on different import route-targets:



Note: Detailed BGP-VPLS configuration is out of the scope of this chapter. For more information, see the Nokia SR OS documentation.

```
*A:PE-1# configure service vpls 500 bgp pw-template-binding ?
  - pw-template-binding <policy-id> [split-horizon-group <group-name>]
                                    [import-rt {ext-community,...(upto 5 max)}]
  - no pw-template-binding <policy-id>

---snip---

*A:PE-1#
```

- The BGP-signaled pseudowires (from PE-1 to PE-2 and PE-3) are set up according to the configuration in the BGP context. Beside those pseudowires, the VPLS 500 also has two more pseudowires signaled by TLDP: spoke-sdp 14:500 (to MTU-4) and spoke-sdp 15:500 (to MTU-5).

The general BGP MH configuration parameters for a particular multi-homing site are shown in the following output:

```
*A:PE-1# configure service vpls ?
  - no vpls <service-id>
  - vpls <service-id> [customer <customer-id>] [create] [vpn <vpn-id>] [m-vpls] [b-
vpls|i-vpls] [etree]

---snip---

*A:PE-1# configure service vpls 500 site ?
  - no site <name>
  - site <name> [create]

 <name>               : [32 chars max]

 [no] boot-timer      - Configure/Override site boot-timer
      failed-thresho* - Configure threshold for the site to be declared down
 [no] mesh-sdp-bindi* - Enable/Disable application to all Mesh-SDP
 [no] monitor-oper-g* - Configure an Operational-Group to monitor
 [no] sap             - Configure a SAP for the site
 [no] shutdown        - Administratively enable/disable the site
 [no] site-activatio* - Configure/Override site activation timer
 [no] site-id         - Configure site identifier
 [no] site-min-down-* - Configure minimum down timer for the site
```

```
 [no] split-horizon-* - Configure a split-horizon-group
 [no] spoke-sdp       - Configure a spoke-SDP
```

Where:

- The site-name is defined by a string of up to 32 characters.
- The site-id is an integer that identifies the multi-homing site and is encoded in the BGP MH NLRI. This ID must be the same one used on the peer node where the same multi-homing site is connected to. That is, MH-site-2 must use the same site-id in PE-1 and PE-2 (value = 2 in the PE-1 site configuration).
- Out of the four potential objects in a site, spoke SDP, SAP, SHG and mesh SDP binding only one can be used at the time on a particular site. To add more than just one sap/spoke-sdp to the same site, a split-horizon-group composed of the sap/spoke-sdp objects must be used in the site configuration. Otherwise, only one object (spoke SDP, SAP, SHG and mesh SDP binding) is allowed per site. A CLI log message warns the operator of such fact:

```
*A:PE-1>config>service>vpls>site# mesh-sdp-binding
MINOR: SVCMGR #5855 only one object is allowed per site
```

- The **failed-threshold** command defines how many objects should be down for the site to be declared down. This command is obviously only valid for multi-object sites (split-horizon-groups and mesh-sdp-bindings). By default, all the objects in a site must be down for the site to be declared as operationally down.

```
*A:PE-1>config>service>vpls>site# failed-threshold ?
  - failed-threshold <[1..1000]>
  - failed-threshold all
```

- The **boot-timer** specifies for how long the service manager waits after a node reboot before running the MH Procedures. The boot-timer value should be configured to allow for the BGP sessions to come up and for the NLRI information to be refreshed/exchanged. In environments with the default BGP MRAI (30 seconds) it is highly recommended to increase this value (for instance, 120 seconds for a normal configuration). The **boot-timer** is only important when a node comes back up and would become the DF. Default value: 10 seconds.

```
*A:PE-1>config>service>vpls>site# boot-timer ?
  - boot-timer <seconds>
  - no boot-timer

 <seconds>            : [0..600]
```

- The **site-activation-timer** command defines the amount of time the service manager will keep the local objects in standby (in the absence of BGP updates from remote PEs) before running the DF election algorithm to decide whether the site should be unblocked. The timer is started when one of the following events occurs only if the site is operationally up:
  - Manual site activation using the **no shutdown** command at the site-id level or at member object(s) level (SAP(s) or pseudowire(s))
  - Site activation after a failure
  - The BGP MH election procedures will be resumed upon expiration of this timer or the arrival of a BGP MH update for the multi-homing site. Default value: 2 seconds.

```
*A:PE-1>config>service>vpls>site# site-activation-timer
  - no site-activation-timer
  - site-activation-timer <seconds>
 <seconds>            : [0..100]
```

- When a BGP MH site goes down, it may be preferred that it stays down for a minimum time. This is configurable by the **site-min-down-timer**. When set to zero, this timer is disabled.

```
*A:PE-1>config>service>vpls>site# site-min-down-timer
  - no site-min-down-timer
  - site-min-down-timer <seconds>

 <seconds>            : [0..100]
```

- The **boot-timer**, **site-activation-timer** and **site-min-down-timer** commands can be provisioned at service level or at global level. The service level settings have precedence and override the global configuration. The no form of the commands at global level, sets the value back to the default values. The **no** form of the commands at service level, makes the timers inherit the global values.

```
*A:PE-1# configure redundancy bgp-multi-homing
  - bgp-multi-homing

 [no] boot-timer      - Configure BGP multi-homing boot-timer
 [no] site-activatio* - Configure BGP multi-homing site activation timer
 [no] site-min-down-* - Configure minimum down timer for the site
```

- The **shutdown** command controls the admin state of the site. Each site has three possible states:
  - Admin state — controlled by the shutdown command.
  - Operational state — controlled by the operational status of the individual site objects.
  - Designated-Forwarder (DF) state — controlled by the BGP MH election algorithm.

The following CLI output shows the three states for a BGP MH site:

```
*A:MTU-5# show service id 500 site "MH-site-1"
===============================================================================
Site Information
===============================================================================
Site Name            : MH-site-1
-------------------------------------------------------------------------------
Site Id              : 1
Dest                 : sap:1/1/1:8          Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : No
DF UpTime            : 0d 00:00:00          DF Chg Cnt       : 1
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
Failed Threshold     : default(all)
Monitor Oper Grp     : (none)
===============================================================================
*A:MTU-5#
```

For this example, configure the site MH-site-2 in PE-1, where the site-id is 2 and the object in the site is spoke-sdp 15:500 (pseudowire established from PE-1 to MTU-5).

The following CLI shows the service configuration for PE-2. The site-id is 2, that is, the same value configured in PE-1. The object defined in PE-2's site is spoke-sdp 25:500 (pseudowire established from PE-2 to MTU-5).

```
configure
    service
        sdp 21 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            path-mtu 8000
            no shutdown
        exit
        sdp 23 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-2-PE-3"
            path-mtu 8000
            no shutdown
        exit
        sdp 24 mpls create
            far-end 192.0.2.4
            lsp "LSP-PE-2-MTU-4"
            path-mtu 8000
            no shutdown
        exit
        sdp 25 mpls create
            far-end 192.0.2.5
            lsp "LSP-PE-2-MTU-5"
            path-mtu 8000
```

```
            no shutdown
        exit
        pw-template 500 use-provisioned-sdp create
        exit
        vpls 500 customer 1 create
            bgp
                route-distinguisher 65000:502
                vsi-export "vsi500_export"
                vsi-import "vsi500_import"
                pw-template-binding 500 split-horizon-group "CORE"
                exit
            exit
            bgp-vpls
                max-ve-id 65535
                ve-name 502
                    ve-id 502
                exit
                no shutdown
            exit
            site "MH-site-2" create
                site-id 2
                spoke-sdp 25:500
                no shutdown
            exit
            spoke-sdp 24:500 create
            exit
            spoke-sdp 25:500 create
            exit
            no shutdown
        exit
```

MTU Service Configuration

The following CLI output shows the service level configuration on MTU-4.

```
configure
    service
        sdp 41 mpls create
            far-end 192.0.2.1
            lsp "LSP-MTU-4-PE-1"
            path-mtu 8000
            no shutdown
        exit
        sdp 42 mpls create
            far-end 192.0.2.2
            lsp "LSP-MTU-4-PE-2"
            path-mtu 8000
            no shutdown
        exit
        vpls 500 customer 1 create
            split-horizon-group "site-1" create
            exit
            bgp
                route-distinguisher 65000:504
```

```
            route-target export target:65000:500 import target:65000:500
        exit
        site "MH-site-1" create
            site-id 1
            split-horizon-group site-1
            no shutdown
        exit
        endpoint "CORE" create
            no suppress-standby-signaling
        exit
        sap 1/1/1:7 split-horizon-group "site-1" create
        exit
        sap 1/1/2:8 split-horizon-group "site-1" create
            eth-cfm
                mep 48 domain 1 association 1 direction down
                    fault-propagation-enable use-if-tlv
                    ccm-enable
                    no shutdown
                exit
            exit
        exit
        spoke-sdp 41:500 endpoint "CORE" create
            precedence primary
        exit
        spoke-sdp 42:500 endpoint "CORE" create
        exit
        no shutdown
    exit
```

The MTU-4 node is configured with the following characteristics:

- The BGP context provides the general BGP parameters for service 500 in MTU-4. The route-target command is now used instead of the vsi-import and vsi-export commands. The intent in this example is to configure only the export and import route-targets. There is no need to modify any other attribute. If the local preference is to be modified (to influence the DF election), a vsi-policy must be configured.
- An A/S pseudowire configuration is used to control the pseudowire redundancy towards the core.
- The multi-homing site, MH-site-1 has a site-id = 1 and a split-horizon-group as an object. The split-horizon-group site-1 is composed of sap 1/1/1:7 and sap 1/1/2:8. As previously discussed, the site will not be declared operationally down until the two SAPs belonging to the site are down. This behavior can be changed by the **failed-threshold** command (for instance, in order to bring the site down when only one object has failed even though the second SAP is still up).
- As an example, a Y.1731 MEP with fault-propagation has been defined in SAP 1/1/2:8. As discussed later in the section, this MEP will signal the status of the SAP (as a result of the BGP MH process) to CE-8.

The service configuration in MTU-5 is as follows:

```
configure
    service
        sdp 51 mpls create
            far-end 192.0.2.1
            lsp "LSP-MTU-5-PE-1"
            path-mtu 8000
            no shutdown
        exit
        sdp 52 mpls create
            far-end 192.0.2.2
            lsp "LSP-MTU-5-PE-2"
            path-mtu 8000
            no shutdown
        exit
        vpls 500 customer 1 create
            bgp
                route-distinguisher 65000:505
                route-target export target:65000:500 import target:65000:500
            exit
            site "MH-site-1" create
                site-id 1
                sap 1/1/1:8
                no shutdown
            exit
            sap 1/1/1:8 create
            exit
            spoke-sdp 51:500 create
            exit
            spoke-sdp 52:500 create
            exit
            no shutdown
        exit
```

Influencing the Designated Forwarder (DF) Decision

As previously explained, assuming that the sites on the two nodes taking part of the same multi-homing site are both up, the two tie-breakers for electing the DF are (in this order):

1. Highest LP
2. Lowest PE ID

The LP by default is 100 in all the routers. Under normal circumstances, if the LP in any router is not changed, MTU-4 will be elected the DF for MH-site-1, whereas PE-1 will be the DF for MH-site-2. Assume in this section that this behavior is changed for MH-site-2 to make PE-2 the DF. Since changing the system address (to make PE-2's ID the lower of the two IDs) is usually not an easy task to accomplish, the vsi-export policy on PE-2 is modified with an LP of 150 with which the MH-site-2 NLRI is announced to PE-1. Because LP 150 is greater than the default 100 in PE-1, PE-2 will be elected as the DF for MH-site-2. The vsi-import policy remains unchanged and the vsi-export policy is modified as follows:

```
configure
    router
        policy-options
            begin
            community "comm_core" members "target:65000:500"
            policy-statement "vsi500_export"
                entry 10
                    action accept
                        community add "comm_core"
                        local-preference 150
                    exit
                exit
            exit
            policy-statement "vsi500_import"
                entry 10
                    from
                        community "comm_core"
                        family l2-vpn
                    exit
                    action accept
                    exit
                exit
                default-action reject
            exit
            commit
```

In PE-1, the import and export policies are not modified. The policies were already applied in the BGP context of VPLS 500, as follows:

```
*A:PE-2# configure
    service
        vpls 500 customer 1 create
            bgp
                route-distinguisher 65000:502
                vsi-export "vsi500_export"
                vsi-import "vsi500_import"
                pw-template-binding 500 split-horizon-group "CORE"
                exit
            exit
---snip---
```

The DF state of PE-2 can be verified as follows:

```
*A:PE-2# show service id 500 site "MH-site-2"
===============================================================================
Site Information
===============================================================================
Site Name            : MH-site-2
-------------------------------------------------------------------------------
Site Id              : 2
Dest                 : sdp:25:500           Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : Yes
DF UpTime            : 0d 00:12:29          DF Chg Cnt       : 2
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
Failed Threshold     : default(all)
Monitor Oper Grp     : (none)
===============================================================================
*A:PE-2#
```

The import and export policies are applied at service 500 level, which means that the LP changes for all the potential multi-homing sites configured under service 500. Therefore, load balancing can be achieved on a per-service basis, but not within the same service.

Another important remark is that these policies are applied on the VPLS 500 for all the potential BGP applications, that is, BGP VPLS, BGP MH and BGP AD. In the example, the LP for the PE-2 BGP updates for BGP MH and BGP VPLS will be set to 150 (this however has no impact on BGP VPLS since a given PE cannot receive two BGP VPLS NLRIs with the same VE-ID, such as a different VE-ID per PE within the same VPLS is required).

The vsi-export policy is restored to its original settings on PE-2, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "vsi500_export"
                entry 10
                    action accept
                        community add "comm_core"
                        no local-preference
                    exit
                exit
            exit
            commit
```

In all the PE nodes, the import and export policies applied in the BGP context of VPLS 500 have identical settings again, and PE-1 is the DF.

Black-Hole Avoidance

SR OS supports the appropriate MAC flush mechanisms for BGP MH, regardless of the protocol being used for the pseudowire signaling:

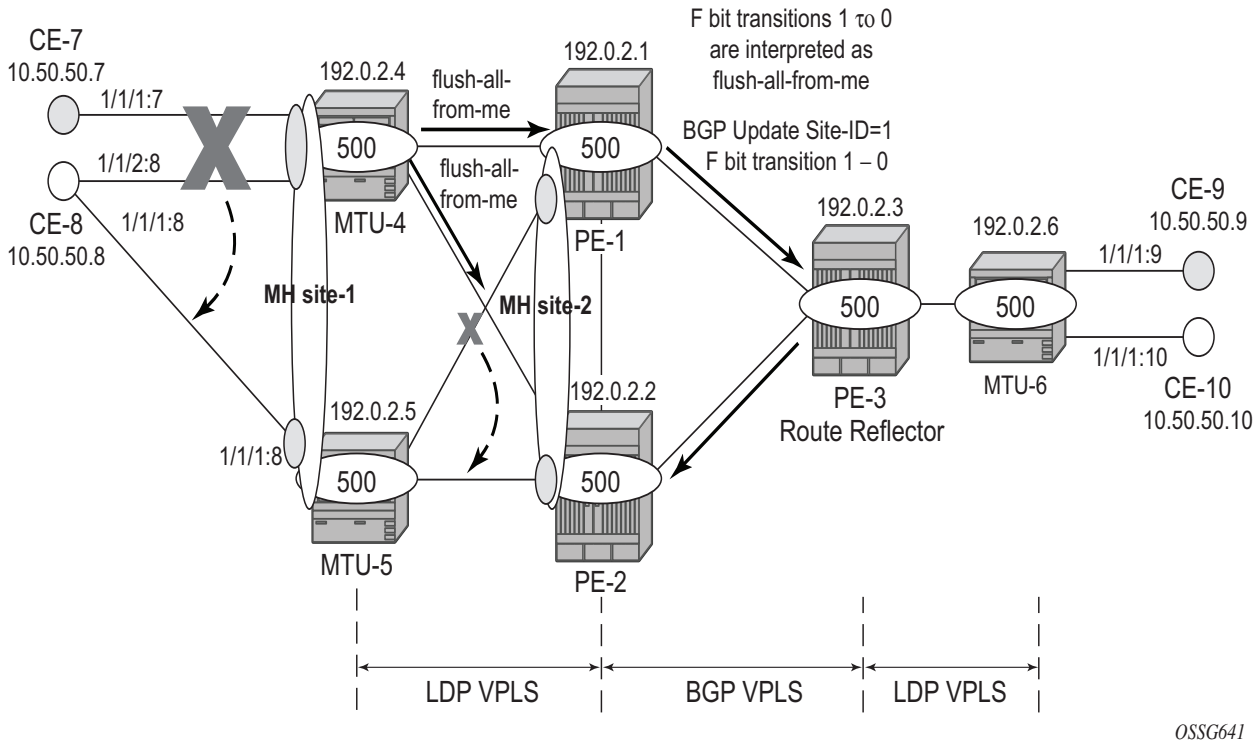
- LDP VPLS — The PE that contains the old DF site (the site that just experienced a DF to non-DF transition) always sends a LDP MAC **flush-all-from-me** to all LDP pseudowires in the VPLS, including the LDP pseudowires associated with the new DF site. No specific configuration is required.
- BGP VPLS — The remote BGP VPLS PEs interpret the F bit transitions from 1 to 0 as an implicit MAC flush-all-from-me indication. If a BGP update with the flag F=0 is received from the previous DF PE, the remote PEs perform MAC flush-all-from-me, flushing all the MACs associated with the pseudowire to the old DF PE. No specific configuration is required.

Double flushing will not happen as it is expected that between any pair of PEs there will exist only one type of pseudowires, either BGP or LDP pseudowire, but not both types.

In the example, assuming MTU-4 and PE-1 are the DF nodes:

- When MH-site-1 is brought operationally down on MTU-4 (so by default, the two SAPs must go down unless the **failed-threshold** parameter is changed so that the site is down when only one SAP is brought down), MTU-4 will issue a **flush-all-from-me** message.
- When MH-site-2 is brought operationally down on PE-1, a BGP update with F=0 and D=1 is issued by PE-1. PE-2 and PE-3 will receive the update and will flush the MAC addresses learned on the pseudowire to PE-1.

Figure 242 **MAC Flush for BGP MH**



Node failures implicitly trigger a MAC flush on the remote nodes, since the TLDP/ BGP session to the failed node goes down.

Access CE/PE Signaling

BGP MH works at service level, therefore no physical ports are torn down on the non-DF but rather the objects are brought down operationally, while the physical port will stay up and used for any other services existing on that port. Due to this reason, there is a need for signaling the standby status of an object to the remote PE or CE.

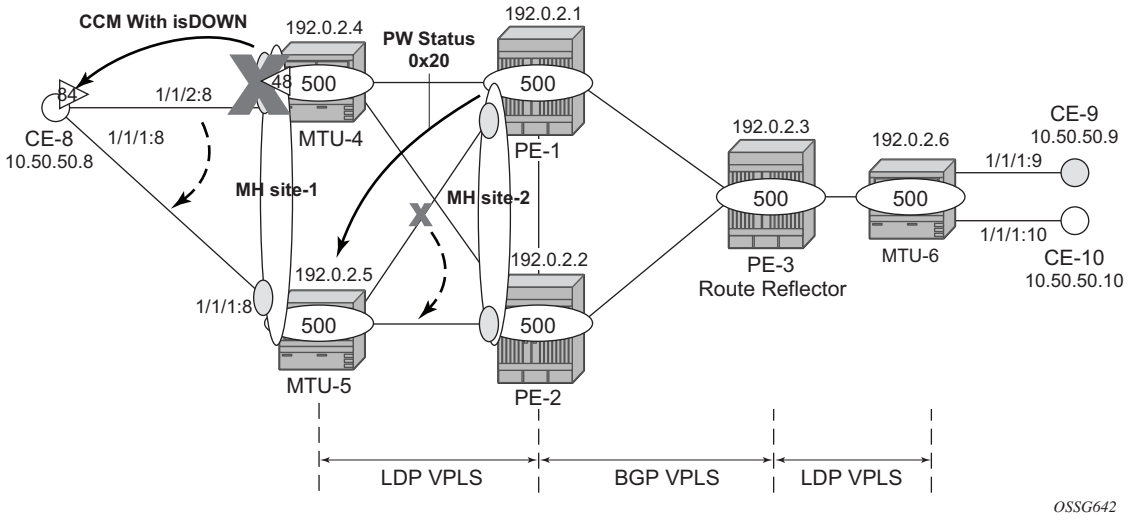
- Access PEs running BGP MH on spoke SDPs and elected non-DF, will signal pseudowire standby status (0x20) to the other end. If no pseudowire status is supported on the remote MTU, a label withdrawal is performed. If there is more than one spoke SDP on the site (part of the same SHG), the signaling is sent for all the pseudowires of the site.

Note: The **configure service vpls x spoke-sdp y:z no pw-status-signaling** parameter allows to send a TLDP label-withdrawal instead of pseudowire status bits, even though the peer supports pseudowire status.

- Multi-homed CEs connected through SAPs to the PEs running BGP MH, are signaled by the PEs using Y.1731 CFM, either by stopping the transmission of CCMs or by sending CCMs with isDown (interface status down encoding in the interface status TLV).

In this example, down MEPs on MTU-4 SAP 1/1/2:8 and CE-8 SAP 1/1/2:8 are configured. In a similar way, other MEPs can be configured on MTU-4 SAP 1/1/1:7, MTU-5 SAP 1/1/1:8, and CE-8 SAP 1/1/1:7 and SAP 1/1/1:8. Figure 243 shows the MEPs on MTU-4 SAP 1/1/2:8 and CE-8. Upon failure on the MTU-4 site MH-site-1, the MEP 48 will start sending CCMs with interface status down.

Figure 243 **Access PE/CE Signaling**



The CFM configuration required at SAP 1/1/2:8 is as follows. Down MEPs will be configured on CE-8 and MTU-5 SAPs in the same way, but in a different association. The option **fault-propagation-enable use-if-tlv** must be added. In case the CE does not understand the CCM interface status TLV, the **fault-propagation-enable suspend-ccm** option can be enabled instead. This will stop the transmission of CCMs upon site failures.

Note: Detailed configuration guidelines for Y.1731 are out of the scope of this chapter.

```
*A:MTU-4# configure
    eth-cfm
        domain 1 format none level 3
            association 1 format icc-based name "Association48"
                bridge-identifier 500
                exit
                ccm-interval 1
                remote-mepid 84
            exit
        exit

*A:MTU-4# configure
    service
        vpls 500 customer 1 create
            sap 1/1/2:8 split-horizon-group "site-1" create
                eth-cfm
                    mep 48 domain 1 association 1 direction down
                        fault-propagation-enable use-if-tlv
                        ccm-enable
                        no shutdown
                    exit
                exit
            exit
```

If CE-8 is a service router, upon receiving a CCM with isDown, an alarm will be triggered and the SAP will be brought down:

```
64 2016/05/26 06:58:30.31 UTC WARNING: SNMP #2004 vprn8 int-CE-8-MTU-4
"Interface int-CE-8-MTU-4 is not operational"
63 2016/05/26 06:58:30.32 UTC MINOR: SVCMGR #2203 vprn8
"Status of SAP 1/1/
2:8 in service 8 (customer 1) changed to admin=up oper=down flags=OamDownMEPFault "
62 2016/05/26 06:58:30.32 UTC MINOR: SVCMGR #2108 vprn8
"Status of interface int-CE-8-MTU-
4 in service 8 (customer 1) changed to admin=up oper=down"
61 2016/05/26 06:58:30.32 UTC MINOR: ETH_CFM #2001 Base
"MEP 1/1/84 highest defect is now defRemoteCCM"
```

On MTU-4, the status of the SAP can be checked:

```
*A:MTU-4# show service id 500 sap 1/1/2:8
===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 500
SAP                : 1/1/2:8                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Down
Flags              : PortOperDown
                     OamDownMEPFault
Multi Svc Site     : None
Last Status Change : 05/26/2016 06:57:24
Last Mgmt Change   : 05/25/2016 11:37:27
===============================================================================
*A:MTU-4#
```

As also depicted in Figure 243, PE-1 will signal pseudowire status standby (code 0x20) when PE-1 goes to non-DF state for MH-site-2. MTU-5 will receive that signaling and, based on the **ignore-standby-signaling** parameter, will decide whether to send the broadcast, unknown unicast and multicast (BUM) traffic to PE-1. In case MTU-5 uses in its configuration **ignore-standby-signaling**. it will be sending BUM traffic on both pseudowires at the same time (which is not normally desired), ignoring the pseudowire status bits. The following output shows the MTU-5 spoke **sdp** receiving the pseudowire status signaling. Although the spoke SDP stays operationally up, the peer Pw Bits field shows **pwFwdingStandby** and MTU-5 will not send any traffic if the **ignore-standby-signaling** parameter is disabled.

```
A:MTU-5# show service id 500 sdp 51:500 detail
===============================================================================
Service Destination Point (Sdp Id : 51:500) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 51:500  -(192.0.2.1)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id          : 51:500                  Type              : Spoke
Spoke Descr     : (Not Specified)
Split Horiz Grp    : (Not Specified)
Etree Root Leaf Tag: Disabled              Etree Leaf AC     : Disabled
VC Type         : Ether                    VC Tag            : n/a
Admin Path MTU  : 8000                     Oper Path MTU     : 8000
Delivery        : MPLS
Far End         : 192.0.2.1
Tunnel Far End  : n/a                      LSP Types         : RSVP
Hash Label      : Disabled                 Hash Lbl Sig Cap  : Disabled
Oper Hash Label : Disabled
Entropy Label   : Disabled

Admin State     : Up                       Oper State        : Up

---snip---

Endpoint        : N/A                      Precedence        : 4
PW Status Sig   : Enabled
Force Vlan-Vc   : Disabled                 Force Qinq-Vc     : Disabled
Class Fwding State : Down
Flags           : None
Time to RetryReset : never                 Retries Left      : 3
Mac Move        : Blockable                Blockable Level   : Tertiary
Local Pw Bits   : None
Peer Pw Bits    : pwFwdingStandby

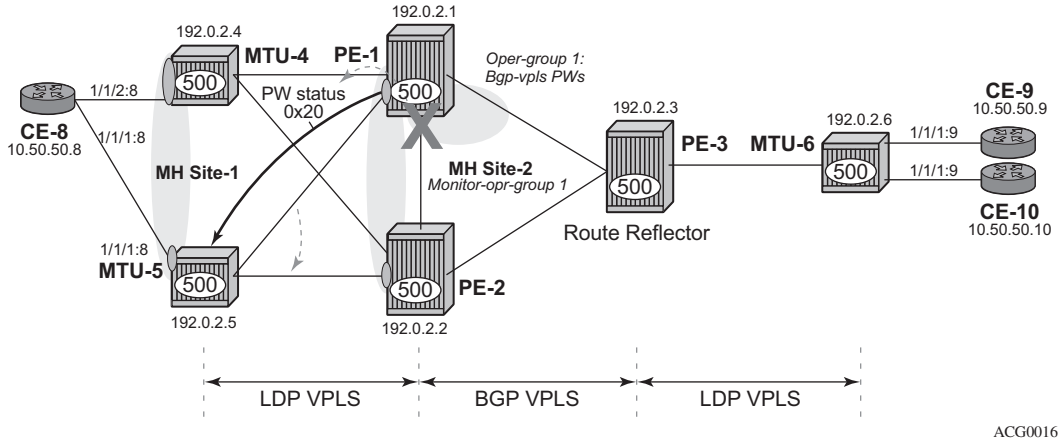
---snip---
```

Operational Groups for BGP-MH

Operational groups (“oper-groups” in the CLI) introduce the capability of grouping objects into a generic group object and associating its status to other service endpoints (pseudowires, SAPs, IP interfaces) located in the same or in different service instances. The operational group status is derived from the status of the individual components using certain rules specific to the application using the concept. A number of other service entities, the monitoring objects, can be configured to monitor the operational group status and to drive their own status based on the oper-group status. In other words, if the operational group goes down, the monitoring objects will be brought down. When one of the objects included in the oper-group comes up, the entire group will also come up, and therefore so will the monitoring objects.

This concept can be used to enhance the BGP-MH solution for avoiding black-holes on the PE selected as the Designated Forwarder (DF), if the rest of the VPLS endpoints fail (pseudowire spoke(s)/pseudowire mesh and/or SAP(s)). Figure 244 illustrates the use of oper-groups together with BGP-MH. On PE-1 (and PE-2) all of the BGP-VPLS pseudowires in the core are configured under the same **oper-group** *group-1*. MH-site-2 is configured as a monitoring object. When the two BGP-VPLS pseudowires go down, **oper-group** *group-1* will be brought down, therefore MH-site-2 on PE-1 will go down as well (PE-2 will become DF and PE-1 will signal standby to MTU-5).

Figure 244 **Oper-Groups and BGP-MH**



In the example above, this feature provides a solution to avoid a black-hole when PE-1 loses its connectivity to the core.

Operational groups are configured in two steps:

1. Identify a set of objects whose forwarding state should be considered as a whole group, then group them under an operational group (in this case **oper-group** *group-1*, which is configured in the **bgp pw-template-bind** context).
2. Associate other existing objects (clients) with the oper-group using the **monitor-group** command (configured, in this case, in the **site** *MH-site-2*).

The following CLI excerpt shows the commands required (**oper-group**, **monitor-oper-group**).

```
*A:PE-1# configure
    service
        oper-group "group-1" create
        exit
        vpls 500
            bgp
                pw-template-binding 500 split-horizon-group "CORE"
                    oper-group "group-1"
                exit
            exit
            site "MH-site-2"
                monitor-oper-group "group-1"
            exit
```

When all the BGP-VPLS pseudowires go down, **oper-group** *group-1* will go down and therefore the monitoring object, **site** *MH-site-2*, will also go down and PE-2 will then be elected as DF. The log 99 gives information about this sequence of events:

```
*A:PE-1# configure service sdp 12 shutdown
*A:PE-1# configure service sdp 13 shutdown

*A:PE-1# show log log-id 99

207 2016/05/26 09:20:52.74 UTC WARNING: SVCMGR #2531 Base BGP-MH
"Service-id 500 site MH-site-2 is not the designated-forwarder"
206 2016/05/26 09:20:52.74 UTC MAJOR: SVCMGR #2316 Base
"Processing of a SDP state change event is finished and the status of all affected S
DP Bindings on SDP 13 has been updated."
205 2016/05/26 09:20:52.74 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 15:500 in service 500 (customer 1) changed to admin=up oper=down
 flags="
204 2016/05/26 09:20:52.74 UTC MINOR: SVCMGR #2326 Base
"Status of SDP Bind 15:500 in service 500 (customer 1) local PW status bits changed
to pwFwdingStandby "
203 2016/05/26 09:20:52.74 UTC MINOR: SVCMGR #2542 Base
"Oper-group group-1 changed status to down"
```

PE-2 becomes the DF.

```
*A:PE-2# show service id 500 site detail
===============================================================================
Site Information
===============================================================================
```

```
Site Name            : MH-site-2
-------------------------------------------------------------------------------
Site Id              : 2
Dest                 : sdp:25:500           Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : Yes
DF UpTime            : 0d 00:03:02          DF Chg Cnt       : 5
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
Failed Threshold     : default(all)
Monitor Oper Grp     : group-1
-------------------------------------------------------------------------------
Number of Sites : 1
===============================================================================
*A:PE-2#
```

The process reverts when at least one BGP-VPLS pseudowire comes back up.

Show Commands and Debugging Options

The main command to find out the status of a given site is the **show service id x site** command.

```
*A:MTU-5# show service id 500 site
===============================================================================
VPLS Sites
===============================================================================
Site                Site-Id   Dest                Mesh-SDP  Admin   Oper  Fwdr
-------------------------------------------------------------------------------
MH-site-1           1         sap:1/1/1:8         no        Enabled up    No
-------------------------------------------------------------------------------
Number of Sites : 1
-------------------------------------------------------------------------------
===============================================================================
*A:MTU-5#
```

A **detail** modifier is available:

```
*A:MTU-5# show service id 500 site detail
===============================================================================
Site Information
===============================================================================
Site Name            : MH-site-1
-------------------------------------------------------------------------------
Site Id              : 1
Dest                 : sap:1/1/1:8          Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : No
DF UpTime            : 0d 00:00:00          DF Chg Cnt       : 1
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
```

```
Failed Threshold     : default(all)
Monitor Oper Grp     : (none)
-------------------------------------------------------------------------------
Number of Sites : 1
===============================================================================
*A:MTU-5#
```

The **detail** view of the command displays information about the BGP MH timers. The values are only shown if the global values are overridden by specific ones at service level (and will be tagged with **Ovr** if they have been configured at service level). The **Timer Remaining** field reflects the count down from the boot/site activation timers down to the moment when this router tries to become DF again. Again, this is only shown when the global timers have been overridden by the ones at service level.

The objects on the non-DF site will be brought down operationally and flagged with **StandByForMHProtocol**. For SAP 1/1/1:8 on non-DF MTU-5:

```
*A:MTU-5# show service id 500 sap 1/1/1:8
===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 500
SAP                : 1/1/1:8                 Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                      Oper State        : Down
Flags              : StandByForMHProtocol
Multi Svc Site     : None
Last Status Change : 05/26/2016 08:21:49
Last Mgmt Change   : 05/25/2016 11:38:29
===============================================================================
*A:MTU-5#
```

For spoke SDP 25:500 on non-DF PE-2:

```
*A:PE-2# show service id 500 sdp 25:500 detail
===============================================================================
Service Destination Point (Sdp Id : 25:500) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 25:500  -(192.0.2.5)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 25:500                  Type              : Spoke
---snip---

Admin State        : Up                      Oper State        : Down
---snip---
Flags              : StandbyForMHProtocol
---snip---
```

The BGP MH routes in the RIB, RIB-In and RIB-Out can be shown by using the corresponding **show router bgp routes** and **show router bgp neighbor x.x.x.x received-routes|advertised-routes** commands. The BGP MH routes are only shown when the operator uses the **l2-vpn family** modifier. Should the operator want to filter only the BGP MH routes out of the l2-vpn routes, the **multi-homing** filter has to be added to the **show router bgp routes** commands.

```
*A:PE-3# show router bgp routes l2-vpn
===============================================================================
 BGP Router ID:192.0.2.3        AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                    Prefix                              MED
      RD                           SiteId                              Label
      Nexthop                      VeId                   BlockSize    LocalPref
      As-Path                      BaseOffset             vplsLabelBa
                                                          se
-------------------------------------------------------------------------------
u*>i  VPLS                         -                      -            0
      65000:501                    -                                   -
      192.0.2.1                    501                    8            100
      No As-Path                   497                    262127
u*>i  MultiHome                    -                      -            0
      65000:501                    2                                   -
      192.0.2.1                    -                      -            100
      No As-Path                   -                      -
u*>i  VPLS                         -                      -            0
      65000:502                    -                                   -
      192.0.2.2                    502                    8            100
      No As-Path                   497                    262127
u*>i  MultiHome                    -                      -            0
      65000:502                    2                                   -
      192.0.2.2                    -                      -            100
      No As-Path                   -                      -
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-3#
```

For the L2 VPN BGP routes toward site 2 (PE-1 and PE-2) in detail:

```
*A:PE-3# show router bgp routes l2-vpn multi-homing siteid 2 detail
===============================================================================
 BGP Router ID:192.0.2.3        AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
```

```
BGP L2VPN-MULTIHOME Routes
===============================================================================
Original Attributes

Route Type     : MultiHome
Route Dist.    : 65000:501
Site Id        : 2
Nexthop        : 192.0.2.1
From           : 192.0.2.1
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65000:500
                 l2-vpn/vrf-imp:Encap=19: Flags=-DF: MTU=0: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h05m30s

Modified Attributes
 ---snip---
-------------------------------------------------------------------------------
Original Attributes

Route Type     : MultiHome
Route Dist.    : 65000:502
Site Id        : 2
Nexthop        : 192.0.2.2
From           : 192.0.2.2
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65000:500
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=0: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m50s
```

```
Modified Attributes
 ---snip---
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-3#
```

The following shows the Layer 2 BGP routes on PE-1:

```
*A:PE-1# show service l2-route-table
  - l2-route-table [detail] [bgp-ad] [multi-homing] [bgp-vpls] [bgp-vpws] [all-routes]

 <detail>             : keyword - display detailed information

*A:PE-1# show service l2-route-table multi-homing
===============================================================================
Services: L2 Multi-Homing Route Information - Summary
===============================================================================
Svc Id     L2-Routes (RD-Prefix)        Next Hop        SiteId     State  DF
-------------------------------------------------------------------------------
500        65000:502                    192.0.2.2       2          up(0)  clear
-------------------------------------------------------------------------------
No. of L2 Multi-Homing Route Entries: 1
===============================================================================
*A:PE-1#
```

Finally, in terms of debugging, the recommendation would be to check the following CLI sources:

In case PE-3 were the RR for MTU-4 and MTU-5 as well as for PE-1 and PE-2, PE-1 would have two more L2-routes for multi-homing in this table, as follows:

```
*A:PE-1# show service l2-route-table multi-homing
===============================================================================
Services: L2 Multi-Homing Route Information - Summary
===============================================================================
Svc Id     L2-Routes (RD-Prefix)        Next Hop        SiteId     State  DF
-------------------------------------------------------------------------------
500        65000:504                    192.0.2.4       1          up(0)  set
500        65000:505                    192.0.2.5       1          up(0)  clear
500        65000:502                    192.0.2.2       2          up(0)  clear
-------------------------------------------------------------------------------
No. of L2 Multi-Homing Route Entries: 3
===============================================================================
*A:PE-1#
```

When oper-groups are configured (as previously shown), the following **show** command helps to find the operational dependencies between monitoring objects and group objects.

```
*A:PE-1# show service oper-group "group-1" detail
===============================================================================
```

```
Service Oper Group Information
===============================================================================
Oper Group       : group-1
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 4 secs
Members          : 2                               Monitoring : 1
===============================================================================

===================================================================
Member SDP-Binds for OperGroup: group-1
===================================================================
SdpId            SvcId      Type IP address      Adm     Opr
-------------------------------------------------------------------
12:4294967292    500        BgpVpls  192.0.2.2       Up      Up
13:4294967293    500        BgpVpls  192.0.2.3       Up      Up
-------------------------------------------------------------------
SDP Entries found: 2
===================================================================
* indicates that the corresponding row element may have been truncated.

===============================================================================
Monitoring Sites for OperGroup: group-1
===============================================================================
SvcId     Site               Site-Id   Dest               Admin   Oper  Fwdr
-------------------------------------------------------------------------------
500       MH-site-2          2         sdp:15:500         Enabled up    Yes
-------------------------------------------------------------------------------
Site Entries found: 1
===============================================================================
*A:PE-1#
```

- **log-id 99** — Provides information about the site object changes and DF changes.
- **debug router bgp** update — Shows the BGP updates for BGP MH, including the sent and received BGP MH NLRIs and flags.

```
*A:MTU-4# debug router bgp update
```

- **debug router ldp** commands — Provides information about the pseudowire status bits being signaled as well as the MAC flush messages.

```
*A:MTU-4# debug router ldp peer 192.0.2.1 packet init detail
*A:MTU-4# debug router ldp peer 192.0.2.1 packet label detail
```

As an example, log-id 99 and debug output is as follows after shutting down MH-site-1 on MTU-4:

```
*A:MTU-4# configure service vpls 500 sap 1/1/1:7 shutdown
*A:MTU-4# configure service vpls 500 sap 1/1/2:8 shutdown

*A:MTU-4# show log log-id 99
===============================================================================
Event Log 99
```

```
===============================================================================
Description : Default System Log
Memory Log contents  [size=500   next event=91  (not wrapped)]

132 2016/05/26 12:05:01.45 UTC WARNING: SVCMGR #2531 Base BGP-MH
"Service-id 500 site MH-site-1 is not the designated-forwarder"
131 2016/05/26 12:05:01.45 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/1/
2:8 in service 500 (customer 1) changed to admin=down oper=down flags=SapAdminDown "
130 2016/05/26 12:05:01.45 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/1/
2:8 in service 500 (customer 1) changed to admin=down oper=down flags=SapAdminDown "
129 2016/05/26 12:05:01.44 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/1/
1:7 in service 500 (customer 1) changed to admin=down oper=down flags=SapAdminDown "
---snip---
```

Log 2 has been configured to log BGP updates and LDP commands.

```
*A:MTU-4# show log log-id 2
===============================================================================
Event Log 2
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=100   next event=11  (not wrapped)]


3 2016/05/26 12:05:01.45 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.4
        [MH] site-id: 1, RD 65000:504
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=0
"
2 2016/05/26 12:05:01.45 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.4
        [MH] site-id: 1, RD 65000:504
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
```

```
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=0
"
1 2016/05/26 12:05:01.46 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address Withdraw packet (msgId 9795) to 192.0.2.1:0
Protocol version = 1
MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/500 Group ID = 0 cBit = 0
"
```

Assuming all the recommended tools are enabled, a DF to non-DF transition can be shown as well as the corresponding MAC flush messages and related BGP processing.

If MH-site-2 is torn down on PE-1, the **debug router bgp update** command would allow us to see two BGP updates from PE-1:

- A BGP MH update for site-id 2 with flag D set (since the site is down).
- A BGP VPLS update for veid=501 and flag D set. This is due to the fact that there are no more active objects on the VPLS, besides the BGP pseudowires.

```
*A:PE-1# configure service vpls 500 spoke-sdp 14:500 shutdown
*A:PE-1# configure service vpls 500 spoke-sdp 15:500 shutdown

*A:PE-1# show log log-id 2

===============================================================================
Event Log 2
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=100   next event=141  (wrapped)]

71 2016/05/26 12:18:23.69 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [MH] site-id: 2, RD 65000:501
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=0
"
70 2016/05/26 12:18:23.69 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
```

```
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 17, veid: 501, vbo: 497, vbs: 8, label-
base: 262122, RD 65000:501
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=1514: PREF=0
"
```

The D flag, sent along with the BGP VPLS update for veid 501, would be seen on the remote core PEs as though it was a pseudowire status fault (although there is no TLDP running in the core).

```
*A:PE-2# show service id 500 all | match Flag
Flags              : PWPeerFaultStatusBits
Flags              : None
Flags              : None
Flags              : None
*A:PE-2#
```

Conclusion

SR OS supports a wide range of service resiliency options as well as the best-of-breed system level HA and MPLS mechanisms for the access and the core. BGP MH for VPLS completes the service resiliency tool set by adding a mechanism that has some good advantages over the alternative solutions:

- BGP MH provides a common resiliency mechanism for attachment circuits (SAPs), pseudowires (spoke SDPs), split horizon groups and mesh bindings
- BGP MH is a network-based technique which does not need interaction to the CE or MTU to which it is providing redundancy to.

The examples used in this section illustrate the configuration of BGP MH for access CEs and MTUs. Show and debug commands have also been suggested so that the operator can verify and troubleshoot the BGP MH procedures.

BGP Virtual Private Wire Services

This chapter describes BGP Virtual Private Wire Service (VPWS) configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for release 11.0.R4. The CLI in the current edition is based on release 14.0.R3. There are no prerequisites for this configuration.

Introduction

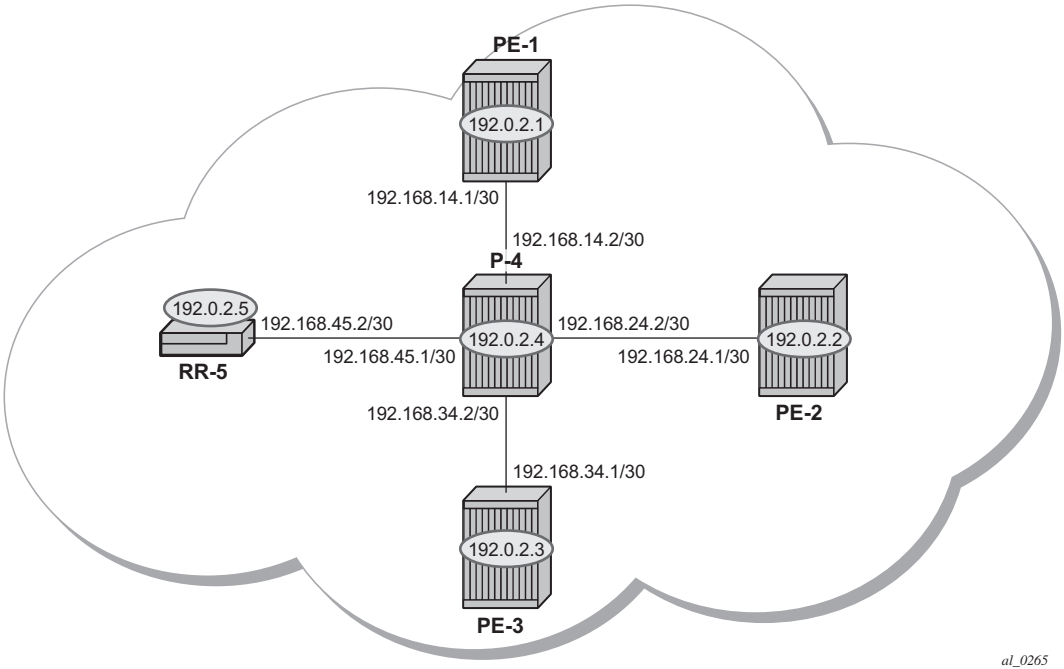
There are currently two IETF standards for the provisioning of virtual private wire services (VPWS). RFC 4447, *Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*, describes label distribution protocol (LDP) VPWS, where VPWS pseudowires are signaled using LDP between provider edge (PE) Routers.

RFC 6624, *Layer 2 Virtual Private Networks Using BGP for Auto-Discovery and Signaling*, describes the use of border gateway protocol (BGP) for signaling of pseudo-wires between such PEs.

The purpose of this chapter is to describe the configuration and troubleshooting for BGP VPWS.

Overview

Figure 245 **Example Topology**



The network topology is displayed in Figure 245. The setup uses five service router (SR) nodes located in the same autonomous system (AS). There are three PE routers connected to a single P router and a route reflector (RR-5) for the AS. The PE routers are all BGP VPWS aware. The Provider (P) router is BGP VPWS unaware and also does not take part in the BGP process.

The following configuration tasks should be completed as a prerequisite:

- IS-IS or OSPF should be configured on each of the network interfaces between the PE/P routers and route reflector.
- MPLS should be configured on all interfaces between PE routers and P routers. It is not required between P-4 and RR-5.
- LDP should be configured on interfaces between PE and P routers. It is not required between P-4 and the RR-5.
- RSVP protocol should be configured on interfaces between PE and P routers. It is not required between P-4 and the RR-5.

BGP VPWS

In this architecture, a VPWS is a collection of two (or three in case of redundancy) BGP VPWS service instances present on different PEs in a provider network.

The PEs communicate with each other at the control plane level by means of BGP updates containing BGP VPWS network layer reachability information (NLRI). Each update contains enough information for a PE to determine the presence of other BGP VPWS instances on peering PEs and to set up pseudowire connectivity for data flow between peers containing the same BGP VPWS service. Therefore, auto-discovery and pseudowire signaling is achieved using a single BGP update message.

Each PE with a BGP VPWS instance is identified by a VPWS edge identifier (VE-ID) and the presence of other BGP VPWS instances is determined using the exchange of standard BGP extended community route targets between PEs.

Each PE will advertise, via the route reflector, the presence of its BGP VPWS instance to all other PEs, along with a block of multiplexer labels (for BGP VPWS there is just one label per block) that can be used to communicate between each instance, plus a BGP next-hop that determines a labeled transport tunnel to be used between PEs.

Each BGP VPWS instance is configured with import and export route target extended communities for topology control, along with VE identification.

The following examples show the configuration of four BGP VPWS scenarios.

- Single homed BGP VPWS
  - using auto-provisioned SDPs
  - using pre-provisioned SDPs
- Dual homed BGP VPWS
  - with single pseudowire
  - with active/standby pseudowire

Configure MP-iBGP

The first step is to configure an MP-iBGP session between each of the PEs and the Route Reflector.

The configuration for PE-1 is as follows:

```
configure router autonomous-system 65536
```

```
configure router
        bgp
            group "INTERNAL"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.5
                exit
            exit
        exit
    exit
```

The configuration for the other PE nodes is exactly the same. The IP addresses can be derived from Figure 245.

The configuration for the Route Reflector (RR-5) is:

```
configure router autonomous-system 65536

configure router
        bgp
            group "INTERNAL"
                family l2-vpn
                peer-as 65536
                cluster 1.1.1.1
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
    exit
```

On RR-5, show that BGP sessions with each PE are established and have a negotiated l2-vpn address family capability.

```
*A:RR-5# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.5        AS:65536       Local AS:65536
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 5           Total Path Memory           : 920
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
```

```
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
===============================================================================
BGP Summary
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
                65536       3    0 00h00m21s 0/0/0 (L2VPN)
                            3    0
192.0.2.2
                65536       3    0 00h00m21s 0/0/0 (L2VPN)
                            3    0
192.0.2.3
                65536       3    0 00h00m21s 0/0/0 (L2VPN)
                            3    0
-------------------------------------------------------------------------------
*A:RR-5#
```

Configuration

Pseudowire Templates

BGP VPWS utilizes pseudowire (PW) templates to dynamically instantiate SDP bindings for a service to signal the egress service de-multiplexer labels used by remote PEs to reach the local PE.

The template determines the signaling parameters of the pseudowire, such as vc-type, vlan-vc-tag, hash-label, filters, etc. The following parameters are recognized by BGP VPWS; the remainder is ignored.

The following commands are supported parameters:

```
configure
  service
     [no] pw-template policy-id [use-provisioned-sdp] [create]
       accounting-policy acct-policy-id
       no accounting-policy
       [no] collect-stats
       [no] controlword
       egress
         filter ipv6 ipv6-filter-id
         filter ip ip-filter-id
         filter mac mac-filter-id
         no filter [ip ip-filter-id] [mac mac-filter-id] [ipv6 ipv6-filter-id]
         qos network-policy-id port-redirect-group queue-group-name
                                                  [instance instance-id]
         no qos
       [no] force-vlan-vc-forwarding
       hash-label [signal-capability]
       no hash-label
       ingress
         filter ipv6 ipv6-filter-id
         filter ip ip-filter-id
         filter mac mac-filter-id
         no filter [ip ip-filter-id] [mac mac-filter-id] [ipv6 ipv6-filter-id]
        qos network-policy-id fp-redirect-group queue-group-name instance instance-id
         no qos
       [no] sdp-exclude group-name
       [no] sdp-include group-name
       vc-type {ether | vlan}
       vlan-vc-tag 0..4094
       no vlan-vc-tag
```

Note that:

- The encapsulation type in the Layer-2 extended community is either 4 (Ethernet VLAN tagged mode) or 5 (Ethernet raw mode), depending on the vc-type parameter.
- The **force-vlan-vc-forwarding** function will add a tag (equivalent to vc-type vlan) and will allow for customer QoS transparency (dot1p + drop eligibility (DE) bits).

The MPLS transport tunnel between PEs can be signaled using LDP or RSVP-TE.

LDP-based SDPs can be automatically instantiated or pre-provisioned. RSVP-TE-based SDPs have to be pre-provisioned. If pre-provisioned pseudowires should be used, the pw-template must be created with the **use-provisioned-sdp** parameter.

```
*A:PE-1# configure service pw-template
  - [no] pw-template <policy-id> [use-provisioned-sdp] [create]
```

Pseudowire Templates for Auto-SDP Creation using LDP

In order to use an LDP transport tunnel for data flow between PEs, it is necessary for link layer LDP to be configured between all PEs/Ps so that a transport label for each PE's system interface is available. For example, on PE-1:

```
configure router
        ldp
            interface-parameters
                interface "int-PE-1-P-4"
                exit
             exit
```

Using this mechanism, SDPs can be auto-instantiated with SDP-ids starting at the higher end of the SDP numbering range, such as 17407. Any subsequent SDPs created use SDP-ids decrementing from this value.

A pseudowire template is required. The example below is created using the default values:

```
configure service
        pw-template 1 create
        exit
```

Pseudowire Templates for Provisioned SDPs using RSVP-TE

RSVP-TE LSPs need to be created between the PE routers on which provisioned SDPs will be used as prerequisite.

The MPLS interface and LSP configuration for PE-1 are:

```
configure router
        mpls
            interface "system"
            exit
            interface "int-PE-1-P-4"
            exit
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "dyn"
                exit
```

```
            no shutdown
        exit
        lsp "LSP-PE-1-PE-3"
            to 192.0.2.3
            primary "dyn"
            exit
            no shutdown
        exit
        no shutdown
```

The MPLS and LSP configuration for PE-2 are similar to that of PE-1 with the appropriate interfaces and LSP names configured.

To use an RSVP-TE tunnel as transport between PEs, it is necessary to bind the RSVP-TE LSP between PEs to an SDP.

The SDP creation on PE-1 toward PE-2 is shown below; similar SDPs are required on each PE to the remote PEs in the service where provisioned SDPs are to be used (only on PE-2 in the following example).

```
configure service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
```

The **signaling bgp** parameter is required. BGP VPWS instances using BGP VPWS signaling are able to use these SDPs. Conversely, SDPs that are bound to RSVP-based LSPs with signaling set to the default value of “tldp” will not be used as SDPs within BGP VPWS.

Single Homed BGP VPWS using Auto-Provisioned SDPs

Figure 246 **Single Homed BGP VPWS using Auto-Provisioned SDPs**

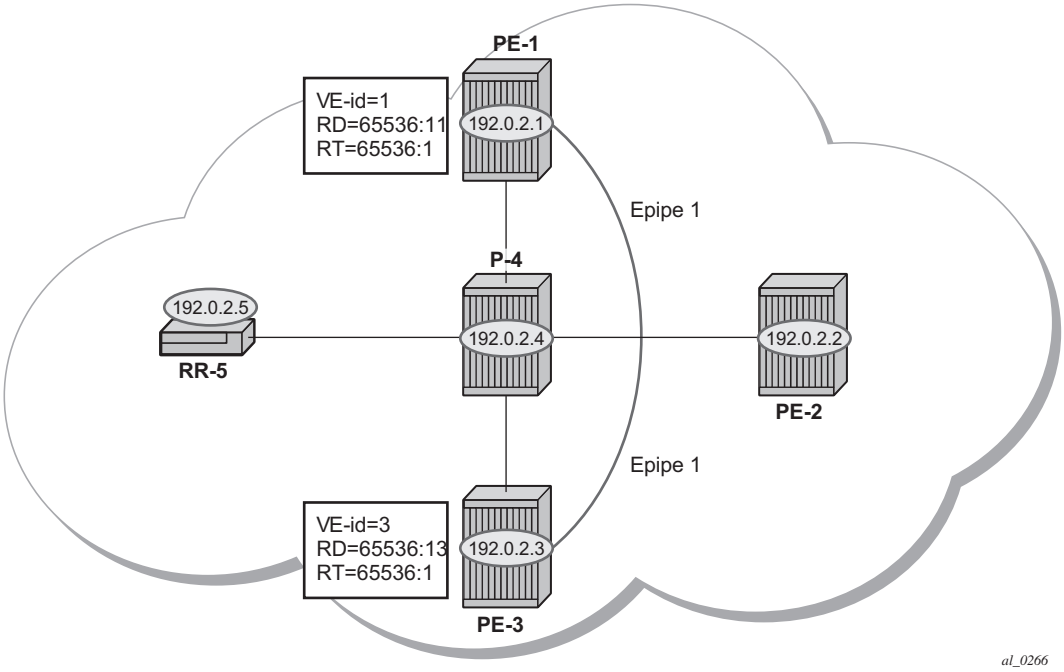


Figure 246 shows a schematic of a single homed BGP VPWS between PE-1 and PE-3 where SDPs are auto-provisioned. In this case, the transport tunnels are LDP signaled.

The following shows the configuration required on PE-1 for a BGP VPWS service using a pseudowire template configured for auto-provisioning of SDPs.

```
*A:PE-1# configure service
        pw-template 1 create
            vc-type vlan
        exit
        epipe 1 customer 1 create
            bgp
                route-distinguisher 65536:11
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
```

```
            remote-ve-name "PE-3"
                ve-id 3
            exit
            no shutdown
        exit
        sap 1/1/4:1 create
        exit
        no shutdown
    exit
```

The **bgp** context specifies parameters that are required for BGP VPWS.

Within the **bgp** context, parameters are configured that are used by the neighboring PEs to determine the membership of a given BGP VPWS; in other words, the auto-discovery of PEs in the same BGP VPWS, the route-distinguisher is configured, along with the route target extended communities. Route target communities are used to determine membership of a given BGP VPWS. The import and export route targets at the BGP level are mandatory. The pw-template binding is then applied and its parameters are used for both the routes sent by this PE and the received routes matching the route target value.

Within the **bgp-vpws** context, the signaling parameters are also configured. These determine the service labels required for the data plane of the VPWS instance.

The VPWS Edge ID (VE-ID) is a numerical value assigned to each PE within a BGP VPWS. This value must be unique for a given BGP VPWS, with the exception of multi-homed scenarios, where two dual-homed PEs can have the same VE-ID and are distinguishable by the site preference (or by the tie breaking rules from the multi-homing draft RFC).

Changes to the pseudowire template are not taken into account once the pseudowire has been set up (changes of route-target are refreshed though). PW-templates can be re-evaluated with the **tools perform service eval-pw-template** command. The **eval-pw-template** checks if all of the bindings using this pw-template policy are still meant to be used this policy. If the template has changed and allow-service-impact is TRUE, then the old binding is removed and it is re-added using the new template.

```
*A:PE-1# tools perform service eval-pw-template 1
eval-pw-template succeeded for Svc 1 Tx L2 ExtComm, Policy 1
eval-pw-template succeeded for Svc 1 17407:4294967295 Policy 1
*A:PE-1#
```

VE-ID and BGP Label Allocations

For a point-to-point VPWS, there are only two members within the BGP VPWS service, so only one label entry is required by each remote service. For dual-homed scenarios, there are two labels for the redundant site, one from each dual-homed PE.

Each PE allocates a label per BGP VPWS instance for the remote PEs, so it signals blocks with one label. It achieves this by advertising three parameters in a BGP update message. For more information about these parameters, see chapter BGP VPLS.

- A Label Base (LB) which is the lowest label in the block.
- A VE Block size (VBS) which is always 1 and cannot be changed.
- A VE Base Offset (VBO) corresponding to the first label in the label block.

PE-3 Service Creation

On PE-3 create a BGP VPWS service using pseudowire template 1. PE-3 has been allocated a VE-ID of 3. For completeness, the pw-template is also shown.

```
*A:PE-3# configure service
        pw-template 1 create
            vc-type vlan
        exit
        epipe 1 customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpws
                ve-name "PE-3"
                    ve-id 3
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:101 create
            exit
            no shutdown
        exit
```

PE-1 Service Operation Verification

Verify that the BGP VPWS service is enabled on PE-1.

```
*A:PE-1# show service id 1 bgp-vpws
===============================================================================
BGP VPWS Information
===============================================================================
Admin State          : Enabled
VE Name              : PE-1                 VE Id            : 1
PW Tmpl used         : 1

Remote-Ve Information
-------------------------------------------------------------------------------
Remote VE Name       : PE-3                 Remote VE Id     : 3
===============================================================================
*A:PE-1#
```

Verify the BGP information used by the BGP VPWS service on PE-1.

```
*A:PE-1# show service id 1 bgp
===============================================================================
BGP Information
===============================================================================
Route Dist           : 65536:11
Oper Route Dist      : 65536:11
Oper RD Type         : configured
Rte-Target Import    : 65536:1              Rte-Target Export: 65536:1
Oper RT Imp Origin   : configured           Oper RT Import   : 65536:1
Oper RT Exp Origin   : configured           Oper RT Export   : 65536:1

PW-Template Id       : 1
BFD Template         : None
BFD-Enabled          : no                   BFD-Encap        : ipv4
Import Rte-Tgt       : None
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

Verify that the service is operationally up on PE-1.

```
*A:PE-1# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 06/01/2016 13:30:00
Last Mgmt Change  : 06/01/2016 13:30:00
Test Service      : No
```

```
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

------------------------------------------------------------------------------
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:1/1/4:1                              q-tag        1578    1578    Up   Up
sdp:17407:4294967295 SB(192.0.2.3)       BgpVpws      0       1552    Up   Up
==============================================================================
*A:PE-1#
```

The SAP and SDP are all operationally up. The indication “SB” next to the SDP-id signifies “Spoke” and “BGP”.

The following output shows the ingress and egress labels for PE-1.

```
*A:PE-1# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17407:4294967295 BgpVpws  192.0.2.3       Up      Up        262137    262137
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The following debug output from PE-1 shows the BGP VPWS NLRI update for Epipe 1 sent by PE-1 to the route reflector (192.0.2.5). This update will then be received by the other PEs.

```
*A:PE-1# debug router bgp update

*A:PE-1# configure log
        log-id 2
            from debug-trace
            to memory
            no shutdown
        exit

*A:PE-1# show log log-id 2
---snip---

4 2016/06/01 13:30:17.85 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
```

```
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
      [VPLS/VPWS] preflen 21, veid: 1, vbo: 3, vbs: 1, label-base: 262137, RD 65536:11,
csv: 0x00000000, type 1, len 1,
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=4: Flags=none: MTU=1514: PREF=0
"
```

Note the presence of the control flags within the extended community which indicate the status of the BGP VPWS instance.

The control flags are described below:

```
0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|D|A|F|Z|Z|Z|C|S| (Z = MUST Be Zero)
+-+-+-+-+-+-+-+-+
```

- D: access circuit down indicator. D is 1 if all access circuits are down, otherwise D is 0.
- A: automatic site id allocation, which is not supported. This is ignored on receipt and set to 0 on sending.
- F: MAC flush indicator, this relates to VPLS. This is set to 0 and ignored on receipt.
- C: presence of a control word. Control word usage is not supported. This is set to 0 on sending (control word not present) and if a non-zero value is received (indicating a control word is required) the pseudowire will not be created.
- S: sequenced delivery. Sequenced delivery is not supported. This is set to 0 on sending (no sequenced delivery) and if a non-zero value is received (indicating sequenced delivery required) the pseudowire will not be created.

The BGP VPWS NLRI is based on that defined for BGP VPLS but is extended with a circuit status vector (CSV). The circuit status vector is used to indicate the status of both the SAP and the spoke-SDP within the local service. As the VE block size used is 1, the most significant bit in the circuit status vector TLV value will be set to 1 if either the SAP or spoke-SDP is down; otherwise, it will be set to 0.

```
6 2016/06/01 13:34:40.86 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
```

```
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 21, veid: 1, vbo: 3, vbs: 1, label-base: 262137,
                    RD 65536:11, csv: 0x00000080, type 1, len 1,
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=4: Flags=D: MTU=1514: PREF=0
"
```

After shutting down the local SAP, the CSV has the most-significant bit set to 1 (0x80).
The BGP VPWS update can be shown using the following command:

```
*A:PE-1# show service l2-route-table bgp-vpws detail
===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

Svc Id          : 1
VeId            : 3
PW Temp Id      : 1
RD              : *65536:3
Next Hop        : 192.0.2.3
State (D-Bit)   : up(0)
Path MTU        : 1514
Control Word    : 0
Seq Delivery    : 0
Status          : active
Tx Status       : active
CSV             : 0
Preference      : 0
Sdp Bind Id     : 17407:4294967295
===============================================================================
*A:PE-1#
```

PE-3 Service Operation Verification

Similar to PE-1, the service operation should be validated on PE-3.

Single Homed BGP VPWS using Pre-Provisioned SDP

It is possible to configure BGP VPWS instances that use RSVP-TE transport tunnels. In this case, the SDPs must be created with the MPLS LSPs mapped and with the signaling set to BGP, as the service labels are signaled using BGP. The pw-template configured within the BGP VPWS instance must use the keyword **use-provisioned-sdp**.

Figure 247 **Single Homed BGP VPWS using Pre-Provisioned SDP**

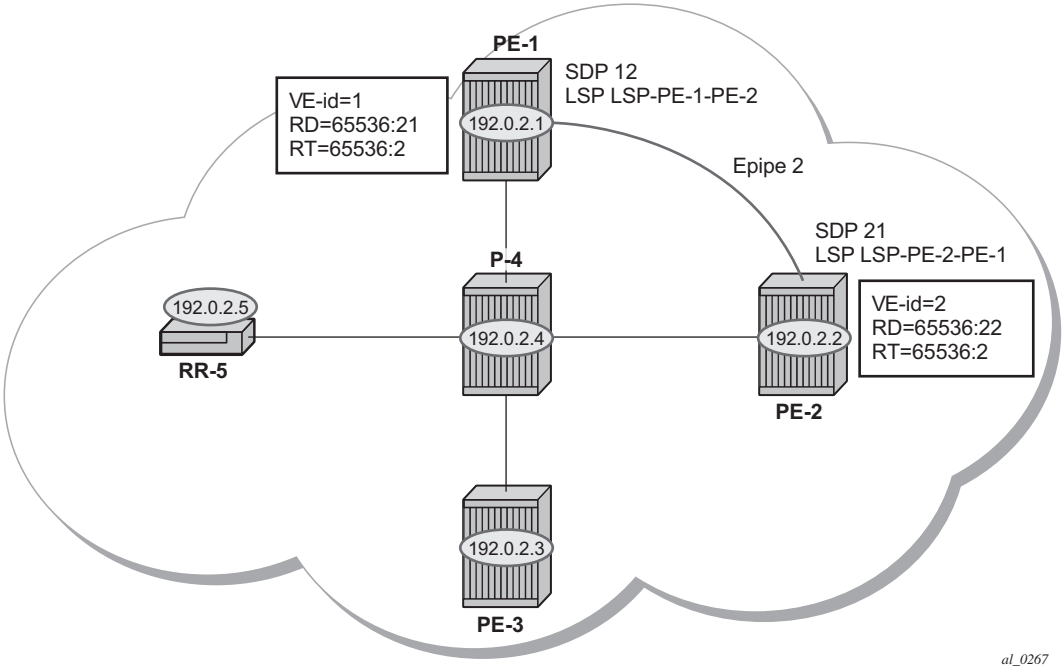


Figure 247 shows a schematic of a BGP VPWS where SDPs are pre-provisioned with RSVP-TE signaled transport tunnels.

SDP on PE-1

```
*A:PE-1# configure service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
```

SDP on PE-2

```
*A:PE-2# configure service
        sdp 21 mpls create
            description "SDP-PE-2-PE-1_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
```

To create a spoke SDP within a service that uses the RSVP-TE transport tunnel, a pseudowire template is required that has the “use-provisioned-sdp” parameter set.

The pw-template is provisioned on both PEs as follows:

```
*A:PE-1# configure service
        pw-template 2 use-provisioned-sdp create
        exit
```

The following output shows the configuration required for a BGP VPWS service using a pseudowire template configured for using pre-provisioned RSVP-TE SDPs.

```
*A:PE-1# configure service
        epipe 2 customer 1 create
            bgp
                route-distinguisher 65536:21
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            sap 1/1/4:2 create
            exit
            no shutdown
```

The route distinguisher and route target extended community values for Epipe 2 are different from that in Epipe 1. This is to differentiate between the two as their visibility is global within the BGP domain. The VE-ID values can be reused in each Epipe instance, as long as they are unique within the instance.

Similarly, on PE-2 the configuration is as follows, where the VE-ID is 2:

```
*A:PE-2# configure service
         epipe 2 customer 1 create
```

```
            bgp
                route-distinguisher 65536:22
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpws
                ve-name "PE-2"
                    ve-id 2
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:2 create
            exit
            no shutdown
```

Verify that the service is operationally up on PE-1.

```
*A:PE-1# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 06/01/2016 13:37:21
Last Mgmt Change  : 06/01/2016 13:37:21
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2                             q-tag        1578    1578    Up   Up
sdp:12:4294967294 S(192.0.2.2)          BgpVpws      0       1552    Up   Up
===============================================================================
*A:PE-1#
```

The SDP-id is the pre-provisioned SDP 12.

For completeness, verify the service is operationally up on PE-2.

```
*A:PE-2# show service id 2 base
===============================================================================
```

```
Service Basic Information
================================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 06/01/2016 13:38:47
Last Mgmt Change  : 06/01/2016 13:38:47
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

--------------------------------------------------------------------------------
Service Access & Destination Points
--------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
--------------------------------------------------------------------------------
sap:1/1/4:2                             q-tag        1578    1578    Up   Up
sdp:21:4294967295 S(192.0.2.1)          BgpVpws      0       1552    Up   Up
================================================================================
*A:PE-2#
```

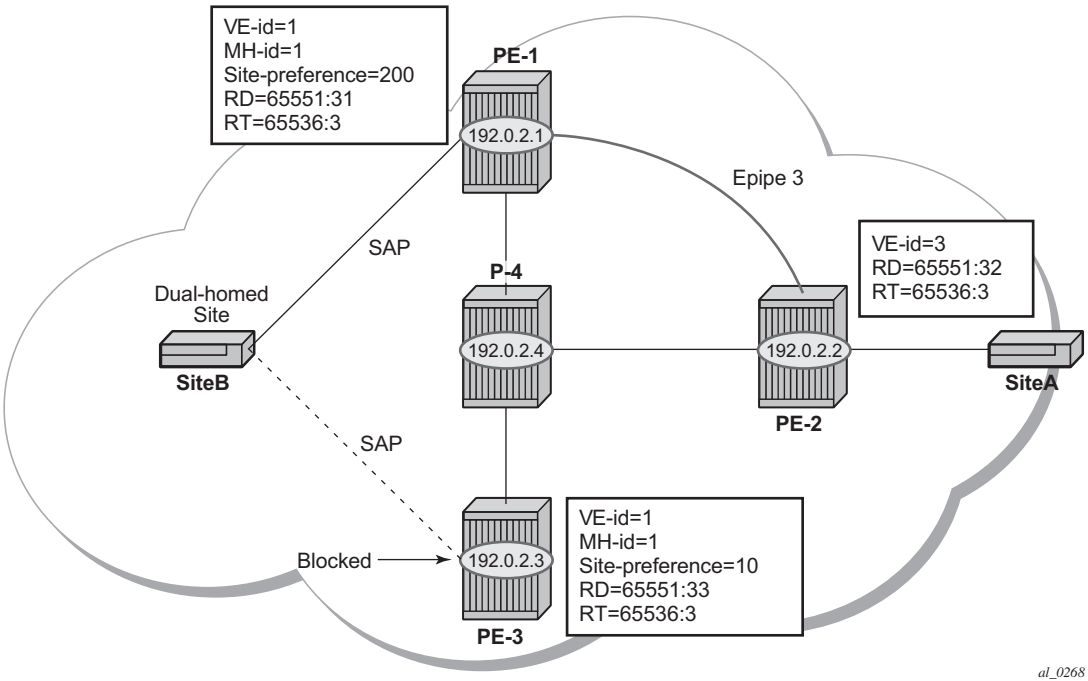
The SDP-id used is the pre-provisioned SDP 21.

Dual Homed BGP VPWS with Single Pseudowire

For access redundancy, an Epipe using a BGP VPWS service can be configured as dual-homed, as described in *draft-ietf-l2vpn-vpls-multihoming-03*. It can be configured with a single pseudowire setup, where the redundant pseudowire is not created until the initially active pseudowire is removed.

The following diagram shows a setup where an Epipe is configured on each PE. Site B is dual-homed to PE-1 and PE-3 with the remote PE-2 connected to site A; each site connection uses a SAP. A single pseudowire using Ethernet Raw Mode encapsulation connects PE-2 to PE-1 or PE-3 (but not both at the same time). The pseudowire is signaled using BGP VPWS over a tunnel LSP between the PEs.

Figure 248 **Dual Homed BGP VPWS with Single Pseudowire**



BGP multi-homing is configured for the dual-homed site B using a site-id=1. The site-preference on PE-1 is set to 200 and to 10 on PE-3, this ensures that PE-1 will be the site's designated forwarder and the pseudowire from PE-2 will be created to PE-1 when PE-1 is fully operational (no pseudowire is created on PE-2 to PE-3). If PE-1 fails, or the multi-homing site fails over to PE-3, then the pseudowire from PE-2 to PE-1 will be removed and a new pseudowire will be created from PE-2 to PE-3.

PE-1 Configuration

```
*A:PE-1# configure service
         pw-template 3 create
         exit
         epipe 3 customer 1 create
            bgp
                route-distinguisher 65536:31
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
```

```
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:3
                site-preference 200
                no shutdown
            exit
            sap 1/1/4:3 create
            exit
            no shutdown
        exit
```

PE-3 Configuration

```
*A:PE-3# configure service
        pw-template 3 create
        exit
        epipe 3 customer 1 create
            bgp
                route-distinguisher 65536:33
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-3"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:3
                site-preference 10
                no shutdown
            exit
            sap 1/1/4:3 create
            exit
            no shutdown
        exit
```

Note that in the above configurations, the remote-ve-name for PE-2 uses VE-ID 2 on both PE-1 and PE-3.

PE-2 Configuration

```
*A:PE-2# configure service
        pw-template 3 create
        exit
        epipe 3 customer 1 create
            bgp
                route-distinguisher 65536:32
```

```
            route-target export target:65536:3 import target:65536:3
            pw-template-binding 3
            exit
        exit
        bgp-vpws
            ve-name "PE-2"
                ve-id 2
            exit
            remote-ve-name "PE-1 or PE-3"
                ve-id 1
            exit
            no shutdown
        exit
        sap 1/1/4:3 create
        exit
        no shutdown
    exit
```

On PE-2, the remote-ve-name is configured as PE-1 or PE-3; this is because both of these PEs are configured with VE-ID 1.

As a result of this configuration, observe that on PE-2, there are multiple route entries for Route-Target 65536:31. In the BGP routing table, there are two entries per partner PE, one for the BGP-MH update (with site-id=1) and the other for the BGP-VPWS update (with VE-ID=1).

```
 *A:PE-2# show router bgp routes l2-vpn rd 65536:31
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                    Prefix                              MED
      RD                           SiteId                              Label
      Nexthop                      VeId                    BlockSize   LocalPref
      As-Path                      BaseOffset              vplsLabelBa
                                                           se
-------------------------------------------------------------------------------
u*>i  MultiHome                    -                       -           0
      65536:31                     1                                   -
      192.0.2.1                    -                       -           200
      No As-Path                   -                       -
u*>i  VPWS                         -                       -           0
      65536:31                     -                                   -
      192.0.2.1                    1                       1           200
      No As-Path                   2                       262135
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-2#
```

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:33
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                 Prefix                          MED
      RD                        SiteId                          Label
      Nexthop                   VeId                BlockSize   LocalPref
      As-Path                   BaseOffset          vplsLabelBa
                                                    se
-------------------------------------------------------------------------------
u*>i  MultiHome                 -                   -           0
      65536:33                  1                               -
      192.0.2.3                 -                   -           10
      No As-Path                -                   -
u*>i  VPWS                      -                   -           0
      65536:33                  -                               -
      192.0.2.3                 1                   1           10
      No As-Path                2                   262136
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-2#
```

The route to PE-1 has the higher site preference, so it is selected as the target for the pseudowire.

```
*A:PE-2# show service l2-route-table bgp-vpws detail
===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

---snip---

Svc Id          : 3
VeId            : 1
PW Temp Id      : 3
RD              : *65536:31
Next Hop        : 192.0.2.1
State (D-Bit)   : up(0)
Path MTU        : 1514
Control Word    : 0
Seq Delivery    : 0
Status          : active
Tx Status       : active
CSV             : 0
Preference      : 200
Sdp Bind Id     : 17407:4294967294
===============================================================================
*A:PE-2#
```

After disabling the SAP in the service on PE-1, BGP UPDATE messages are received. The VPLS/VPWS message received on PE-2 from PE-1 shows in the CSV that the access circuit is down (the CSV has the most-significant bit set to 1 (0x80)), so PE-2 selects the update from PE-3 to create the pseudowire. The BGP-MH update received by PE-2 from PE-1 also shows that the local site is down as indicated by the flags=D.

Note in the following debug output,

- BGP MH (multi-homing) entry uses encap-type=19.
- BGP VPWS entry uses encap-type=5 (Ethernet raw mode).

```
37 2016/06/01 13:41:57.10 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
      [VPLS/VPWS] preflen 21, veid: 1, vbo: 2, vbs: 1, label-base: 262135, RD 65536:31,
                  csv: 0x00000080, type 1, len 1,
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 0
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.1
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:3
        l2-vpn/vrf-imp:Encap=5: Flags=D: MTU=1514: PREF=200
"

36 2016/06/01 13:41:57.10 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [MH] site-id: 1, RD 65536:31
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 0
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.1
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:3
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=200
"
```

The result can be shown on PE-2 as now the spoke SDP is up (active) to PE-3.

```
*A:PE-2# show service l2-route-table bgp-vpws detail
===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

---snip---

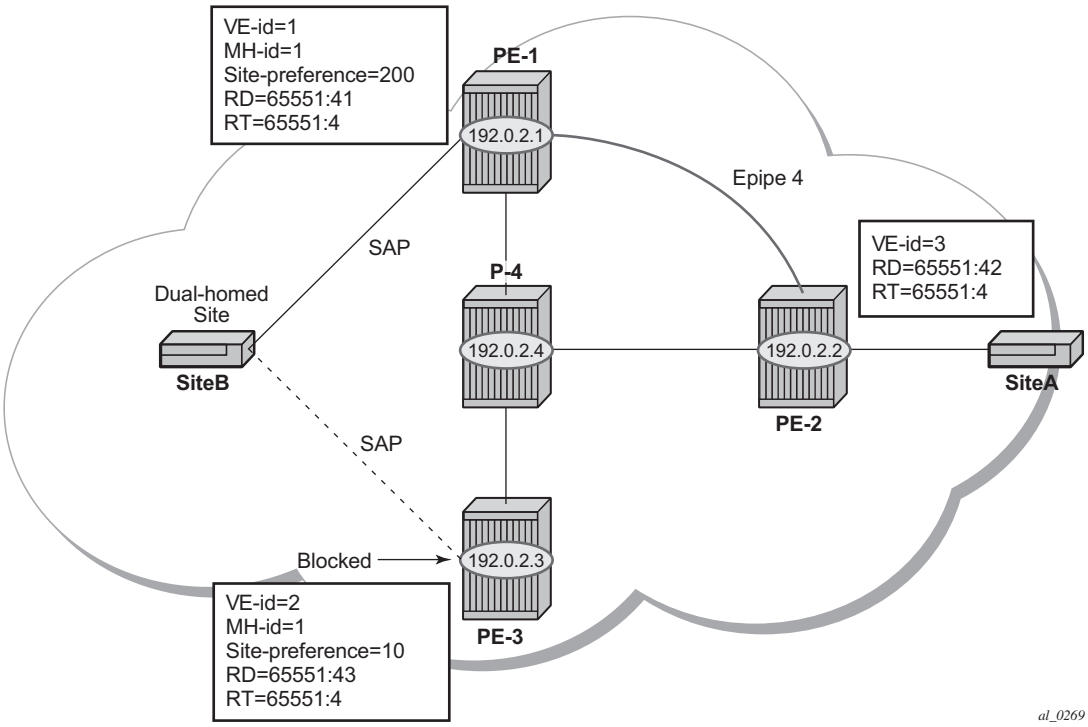
Svc Id          : 3
VeId            : 1
PW Temp Id      : 3
RD              : *65536:33
Next Hop        : 192.0.2.3
State (D-Bit)   : up(0)
Path MTU        : 1514
Control Word    : 0
Seq Delivery    : 0
Status          : active
Tx Status       : active
CSV             : 0
Preference      : 10
Sdp Bind Id     : 17407:4294967293
===============================================================================
*A:PE-2#
```

Dual Homed BGP VPWS with Active/Standby Pseudowire

The second method for BGP VPWS pseudowire redundancy is an active/standby configuration. While in the solution with one pseudowire, the redundant nodes use the same VE-ID for the remote PE and different preferences; in the active/standby solution, the redundant nodes use different VE-IDs for the remote PE and different preferences. The node connecting to both pseudowires (PE-2 in this example) has both remote VE-IDs configured. This allows for faster failover as the standby pseudowire is instantiated in addition to the active pseudowire. If more than two applicable BGP updates are received, at most one standby pseudowire is created (based on the BGP VPWS tie breaking rules).

Figure 249 shows a setup where an Epipe is configured on each PE. Site B is dual-homed to PE-1 and PE-3 with the remote PE-2 connected to site A; each site connection uses a SAP. The active/standby pseudowires using Ethernet raw mode encapsulation connect PE-2 to PE-1 and PE-3. The pseudowires are signaled using BGP VPWS over tunnel LSPs between the PEs.

Figure 249 **Dual Homed BGP VPWS with Active/Standby Pseudowire**



BGP multi-homing (MH) is configured for the dual-homed site B using a site-id=1. The site-preference on PE-1 is set to 200 and to 10 on PE-3; this ensures that PE-1 will be the site's designated forwarder for the MH site. The active pseudowire from PE-2 will be created to PE-1 with the standby pseudowire being created to PE-3. If PE-1 fails, or the multi-homing site fails over to PE-3, then the pseudowire from PE-2 to PE-3 will become active (used as the data path between site A and B).

PE-1 configuration:

```
*A:PE-1# configure service
         pw-template 3 create
         exit
         epipe 4 customer 1 create
            bgp
                route-distinguisher 65536:41
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
```

```
            no shutdown
        exit
        site "SITEB" create
            site-id 1
            sap 1/1/4:4
            site-preference 200
            no shutdown
        exit
        sap 1/1/4:4 create
        exit
        no shutdown
    exit
```

PE-3 configuration:

The local VE-ID is 3 (different from previous example).

```
*A:PE-3# configure service
         pw-template 3 create
         exit
         epipe 4 customer 1 create
            bgp
                route-distinguisher 65536:43
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-3"
                    ve-id 3
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:4
                site-preference 10
                no shutdown
            exit
            sap 1/1/4:4 create
            exit
            no shutdown
        exit
```

PE-2 configuration:

There are two remote VE names configured, PE-1 and PE-3 (this is the maximum number allowed).

```
*A:PE-2# configure service
```

```
pw-template 3 create
exit
epipe 4 customer 1 create
   bgp
       route-distinguisher 65536:42
       route-target export target:65536:4 import target:65536:4
       pw-template-binding 3
       exit
   exit
   bgp-vpws
       ve-name "PE-2"
           ve-id 2
       exit
       remote-ve-name "PE-1"
           ve-id 1
       exit
       remote-ve-name "PE-3"
           ve-id 3
       exit
       no shutdown
   exit
   sap 1/1/4:4 create
   exit
   no shutdown
```

Compared with the single pseudowire solution, both pseudowires are signaled and up on all PEs. The pseudowire with the higher preference is forwarding traffic (to PE-1), while the Tx Status on the other one is set to inactive.

Verified on PE-2:

```
*A:PE-2# show service l2-route-table bgp-vpws detail
===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

---snip---

Svc Id         : 4
VeId           : 1
PW Temp Id     : 3
RD             : *65536:41
Next Hop       : 192.0.2.1
State (D-Bit)  : up(0)
Path MTU       : 1514
Control Word   : 0
Seq Delivery   : 0
Status         : active
Tx Status      : active
CSV            : 0
Preference     : 200
Sdp Bind Id    : 17407:4294967291

Svc Id         : 4
VeId           : 3
PW Temp Id     : 3
```

```
RD              : *65536:43
Next Hop        : 192.0.2.3
State (D-Bit)   : up(0)
Path MTU        : 1514
Control Word    : 0
Seq Delivery    : 0
Status          : active
Tx Status       : inactive
CSV             : 0
Preference      : 10
Sdp Bind Id     : 17406:4294967290
===============================================================================
*A:PE-2#
```

The choice of pseudowire to be used to transmit traffic from PE-2 to PE-1 can also be seen in the endpoint created in the BGP VPWS service. Endpoints are automatically created for the pseudowires within a BGP VPWS service regardless of whether active/standby pseudowires are used; these endpoints are created with a system generated name that ends with the BGP VPWS service id.

```
*A:PE-2# show service id 4 endpoint
===============================================================================
Service 4 endpoints
===============================================================================
Endpoint name                : _tmnx_BgpVpws-4
Description                  : Automatically created BGP-VPWS endpoint
Creation Origin              : bgpVpws
Revert time                  : 0
Act Hold Delay               : 0
Standby Signaling Master     : false
Standby Signaling Slave      : false
Tx Active (SDP)              : 17407:4294967291
Tx Active Up Time            : 0d 00:00:48
Revert Time Count Down       : N/A
Tx Active Change Count       : 3
Last Tx Active Change        : 06/01/2016 13:46:52
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 17406:4294967290 Prec:4                  Oper Status: Up
Spoke-sdp: 17407:4294967291 Prec:4                  Oper Status: Up
===============================================================================
===============================================================================
*A:PE-2#
```

The following command has no effect on an automatically created VPWS endpoint.

```
tools perform service id <service-id> endpoint <endpoint-name> force-switchover
```

Conclusion

BGP VPWS allows the delivery of Layer 2 virtual private wire services to customers where BGP is commonly used. This chapter shows the configuration of single and dual-homed BGP VPWS services together with the associated show output, which can be used to verify and troubleshoot them.

BGP VPLS

This chapter describes advanced BGP VPLS configurations.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 9.0.R3. The CLI in the current edition corresponds to release 14.0.R3. There are no pre-requisites for this configuration.

Summary

There are currently two IETF standards for the provisioning of Virtual Private LAN Services (VPLS). RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*, describes Label Distribution Protocol (LDP) VPLS, where VPLS pseudowires are signaled using LDP between VPLS Provider Edge (PE) routers, either configured manually or auto-discovered using BGP.

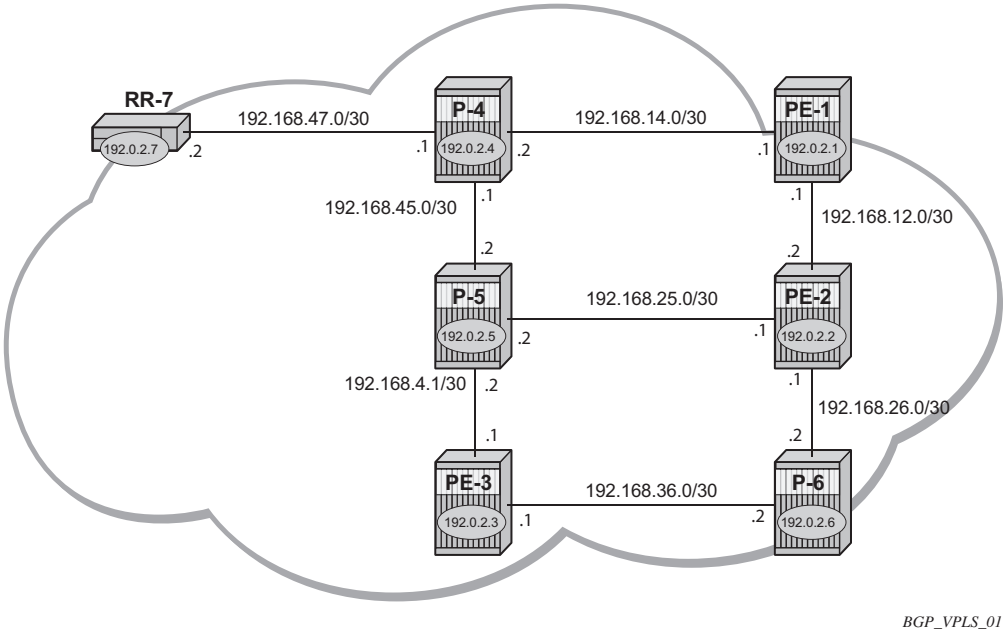
RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*, describes the use of Border Gateway Protocol (BGP) for both the auto-discovery of VPLS PEs and signaling of pseudowires between such PEs.

The purpose of this section is to describe the configuration and troubleshooting for BGP-VPLS.

Knowledge of BGP-VPLS RFC 4761 architecture and functionality is assumed throughout this chapter, as well as knowledge of Multi-Protocol BGP.

Overview

Figure 250 **Network Topology**



The network topology is displayed in Figure 250. The configuration uses seven 7750 Service Router (SR) nodes located in the same Autonomous System (AS). There are three Provider Edge (PE) routers, and RR-7 will act as a Route Reflector (RR) for the AS. The PE routers are all VPLS-aware, the Provider (P) routers are VPLS unaware and do not take part in the BGP process.

The following configuration tasks should be completed as a pre-requisite:

- IS-IS or OSPF on each of the network interfaces between the PE/P routers and RR.
- MPLS should be configured on all interfaces between PE routers and P routers. MPLS is not required between P-4 and RR-7.
- LDP should be configured on interfaces between PE and P routers. It is not required between P-4 and the RR-7.
- The RSVP protocol should be enabled.

BGP VPLS

In this architecture a VPLS instance is a collection of local VPLS instances present on a number of PEs in a provider network. In this context, any VPLS-aware PE is also known as a VPLS Edge (VE) device.

The PEs communicate with each other at the control plane level by means of BGP updates containing BGP-VPLS Network Layer Reachability Information (NLRI). Each update contains enough information for a PE to determine the presence of other local VPLS instances on peering PEs and to set up pseudowire connectivity for data flow between peers containing a local VPLS within the same VPLS instance. Therefore, auto-discovery and pseudowire signaling are achieved using a single BGP update message.

Each PE within a VPLS instance is identified by a VPLS Edge identifier (ve-id) and the presence of a VPLS instance is determined using the exchange of standard BGP extended community route targets between PEs.

Each PE will advertise, via the route reflectors, the presence of each VPLS instance to all other PEs, along with a block of multiplexer labels that can be used to communicate between such instances plus a BGP next hop that determines a labeled transport tunnel between PEs.

Each VPLS instance is configured with import and export route target extended communities for topology control, along with VE identification.

Configuration

The first step is to configure an MP-iBGP session between each of the PEs and the RR.

The configuration for PE-1 is as follows:

```
*A:PE-1# configure router
        autonomous-system 65536
        bgp
            group "INTERNAL"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.7
                exit
            exit
            no shutdown
        exit
```

The BGP configuration for the other PE nodes is identically the same. The IP addresses can be derived from Figure 250.

The configuration for RR-7 is as follows:

```
*A:RR-7# configure router
        autonomous-system 65536
        bgp
            cluster 1.1.1.1
            group "RR-INTERNAL"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

On PE-1, verify that the BGP session with RR-7 is established with address family l2-vpn capability negotiated:

```
*A:PE-1# show router bgp neighbor 192.0.2.7
===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                 : 192.0.2.7
Description          : (Not Specified)
Group                : INTERNAL
-------------------------------------------------------------------------------
Peer AS              : 65536            Peer Port            : 50439
Peer Address         : 192.0.2.7
Local AS             : 65536            Local Port           : 179
Local Address        : 192.0.2.1
Peer Type            : Internal         Dynamic Peer         : No
State                : Established      Last State           : Established
Last Event           : recvKeepAlive
Last Error           : Cease (Connection Collision Resolution)
Local Family         : L2-VPN
Remote Family        : L2-VPN
Hold Time            : 90               Keep Alive           : 30
Min Hold Time        : 0
Active Hold Time     : 90               Active Keep Alive    : 30
Cluster Id           : None
Preference           : 170              Num of Update Flaps  : 0

---snip---
-------------------------------------------------------------------------------
Neighbors : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

On RR-7, show that BGP sessions with each PE are established, and have a negotiated the l2-vpn address family capability.

```
*A:RR-7# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.7        AS:65536       Local AS:65536
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 7           Total Path Memory           : 1288

---snip---
===============================================================================
BGP Summary
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
                65536       4    0 00h00m48s 0/0/0 (L2VPN)
                            4    0
192.0.2.2
                65536       4    0 00h00m48s 0/0/0 (L2VPN)
                            4    0
192.0.2.3
                65536       4    0 00h00m48s 0/0/0 (L2VPN)
                            4    0
-------------------------------------------------------------------------------
*A:RR-7#
```

Configure a full mesh of RSVP-TE LSPs between PE routers.

The MPLS interface and LSP configuration for PE-1 are:

```
*A:PE-1# configure router
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-P-4"
            exit
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "loose"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
                primary "loose"
```

```
            exit
            no shutdown
        exit
        no shutdown
    exit
```

The MPLS and LSP configuration for PE-2 and PE-3 are similar to that of PE-1 with the appropriate interfaces and LSP names configured.

BGP VPLS PE Configuration

Pseudowire Templates

Pseudowire templates are used by BGP to dynamically instantiate SDP bindings, for a given service, to signal the egress service de-multiplexer labels used by remote PEs to reach the local PE.

The template determines the signaling parameters of the pseudowire, control word presence, MAC-pinning, filters etc., plus other usage characteristics such as split horizon groups.

The MPLS transport tunnel between PEs can be signaled using LDP or RSVP-TE.

LDP based pseudowires can be automatically instantiated. RSVP-TE based SDPs have to be pre-provisioned.

Pseudowire Templates for Auto-SDP Creation Using LDP

In order to use an LDP transport tunnel for data flow between PEs, it is necessary for link layer LDP to be configured between all PEs/Ps, so that a transport label for each PE’s system interface is available.

```
*A:PE-1# configure router
        ldp
            interface-parameters
                interface "int-PE-1-PE-2"
                exit
                interface "int-PE-1-P-4
                exit
            exit
        exit
```

Using this mechanism SDPs can be auto-instantiated with SDP-ids starting at the higher end of the SDP numbering range, such as 17407. Any subsequent SDPs created use SDP-ids decrementing from this value.

A pseudowire template is required containing a split horizon group. Each SDP created with this template is contained within a split horizon group so that traffic cannot be forwarded between them.

```
*A:PE-1# configure service
        pw-template 1 create
            split-horizon-group "VPLS-SHG"
            exit
        exit
```

The pseudowire template also has the following options available when used for BGP-VPLS:

```
*A:PE-1# configure service pw-template
  ---snip---
  [no] controlword
  ---snip---
  [no] force-vlan-vc-forwarding
  ---snip---
   vc-type {ether | vlan}
  ---snip---
```

- The control word will determine whether the C flag is set in the Layer 2 extended community and, therefore, if a control word is used in the pseudowire.
- The encap type in the Layer 2 extended community is always 19 (VPLS encap), therefore, the vc-type will always be **ether** regardless of the configured value on the vc-type.
- The **force-vlan-vc-forwarding** command will add a tag (equivalent to **vc-type vlan**) and will allow for customer QoS transparency (dot1p+Drop Eligibility (DE) bits).

Pseudowire Templates for Provisioned SDPs using RSVP-TE

To use an RSVP-TE tunnel as transport between PEs, it is necessary to bind the RSVP-TE LSP between PEs to an SDP.

SDP creation from PE-1 to PE-2 is shown below:

```
*A:PE-1# configure service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
```

```
            no shutdown
        exit
```

The **signaling bgp** parameter is required for BGP-VPLS to be able to use this SDP. Conversely, SDPs that are bound to RSVP-based LSPs with signaling set to the default value of **tldp** will not be used as SDPs within BGP-VPLS.

BGP VPLS Using Auto-Provisioned SDPs

Figure 251 **BGP VPLS Using Auto-Provisioned SDPs**

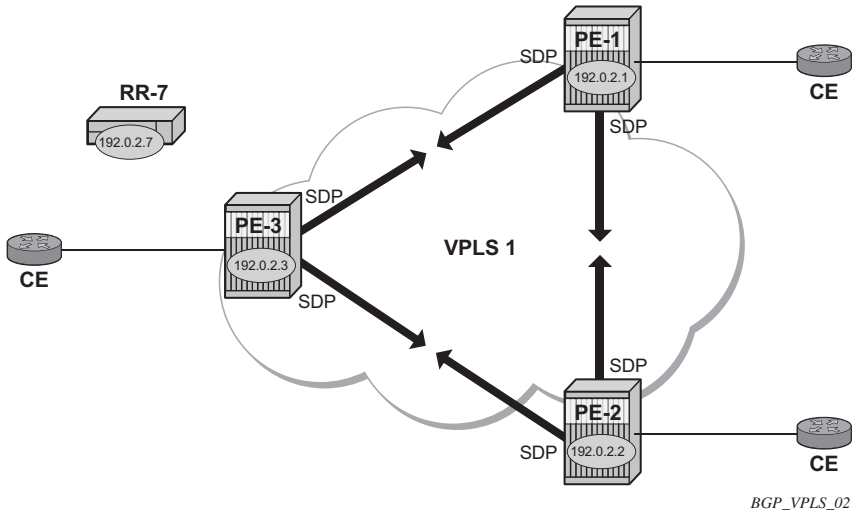


Figure 251 shows a VPLS instance where SDPs are auto-provisioned. In this case, the transport tunnels are LDP signaled.

The following output shows the configuration required on PE-1 for a BGP-VPLS service using a pseudowire template configured for auto-provisioning of SDPs.

```
A:PE-1# configure service
        pw-template 1 create
            split-horizon-group "VPLS-SHG"
            exit
        exit
        vpls 1 customer 1 create
            bgp
                route-distinguisher 65536:1
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpls
```

```
            max-ve-id 10
            ve-name "PE-1"
                ve-id 1
            exit
            no shutdown
        exit
        service-name "VPLS1_PE-1"
        sap 1/1/4:1.0 create
        exit
        no shutdown
    exit
```

The **bgp** context specifies parameters which are valid for all of the VPLS BGP applications, such as BGP-multi-homing, BGP-auto-discovery as well as BGP-VPLS.

Within the **bgp** context, parameters are configured that are used by neighboring PEs to determine membership of a given VPLS instance, such as the auto-discovery of PEs containing the same VPLS instance; the route-distinguisher is configured, along with the route target extended communities.

Route target communities are used to determine membership of a given VPLS instance. The import route target at the BGP level is mandatory. The pseudowire template bind is then applied by the service manager on the received routes matching the route target value.

Within the **bgp-vpls** context, the signaling parameters are configured. These determine the service labels required for the data plane of the VPLS instance.

The VPLS edge ID (ve-id) is a numerical value assigned to each PE within a VPLS instance. This value should be unique for a given VPLS instance, no two PEs within the same instance should have the same ve-id values.

A more specific route target can be applied to a pseudowire template in order to define a specific pseudowire topology, rather than only a full mesh, using the command within the **bgp** context:

pw-template *template-id* [**split-horizon-group** *groupname*] [**import-rt** *import-rt-value* (up to 5 max)]

Changes to the import policies are not taken once the pseudowire has been setup (changes on route-target are refreshed though). Pseudowire templates can be re-evaluated with the command **tools perform service eval-pw-template**. The **eval-pw-template** command checks whether all the bindings using this pseudowire template policy are still meant to use this policy.

If the policy has changed and **allow-service-impact** is true, then the old binding is removed and it is re-added with the new template.

VE-ID and BGP Label Allocations

The choice of ve-id is crucial in ensuring efficient allocation of de-multiplexer labels. The most efficient choice is for ve-ids to be allocated starting at 1 and incrementing for each PE as the following section explains.

The **max-ve-id** *value* determines the range of the ve-id value that can be configured. If a PE receives a BGP-VPLS update containing a ve-id with a greater value than the configured **max-ve-id**, then the update is dropped and no service labels are installed for this ve-id.

The **max-ve-id** command also checks the locally-configured ve-id, and prevents a higher value from being used.

Each PE allocates blocks of labels per VPLS instance to remote PEs, in increments of eight labels. It achieves this by advertising three parameters in a BGP update message,

- A label base (LB) which is the lowest label in the block
- A VE Block size (VBS) which is always eight labels, and cannot be changed
- A VE base offset (VBO).

This defines a block of labels in the range (LB, LB+1, ..., LB+VBS-1).

As an example, if the label base (LB) = 262128, then the range for the block is 262128 to 262135, which is exactly eight labels, as per the block size. (The last label in the block is calculated as 262128+8-1 = 262135)

The label allocated by the PE to each remote PE within the VPLS is chosen from this block and is determined by its ve-id. In this way, each remote PE has a unique de-multiplexer label for that VPLS.

To reduce label wastage, contiguous ve-ids in the range (N..N+7) per VPLS should be chosen, where N>0.

Assuming a collection of PEs with contiguous ve-ids, the following labels will be chosen by PEs from the label block allocated by PE-1 which has a ve-id =1.

Table 18 **VE-IDs and Labels**

VE-ID	Label
2	262129
3	262130
4	262131

Table 18 **VE-IDs and Labels (Continued)**

VE-ID	Label
5	262132
6	262133
7	262434
8	262135

This shows that the label allocated to a given PE is (LB+veid-1). The “1” is the VE block offset (VBO).

This means that the label allocated to a PE router within the VPLS can now be written as (LB + veid - VBO), which means that (ve-id - VBO) calculation must always be at least zero and be less than the block size, which is always 8.

For ve-id ≤ 8, a label will be allocated from this block.

For the next block of 8 ve-ids (ve-id 9 to ve-id 16) a new block of 8 labels must be allocated, so a new BGP update is sent, with a new label base, and a block offset of 9.

Table 19 shows how the choice of ve-ids can affect the number of label blocks allocated, and hence the number of labels:

Table 19 **VE-IDs and Number of Labels**

VE-ID	Block Offset	Labels Allocated
1-8	1	8
9-16	9	8
17-24	17	8
25-32	25	8
33-40	33	8
41-48	41	8
49-56	49	8

This shows that the most efficient use of labels occurs when the ve-ids for a set of PEs are chosen from the same block offset.

If ve-ids are chosen that map to different block offsets, then each PE will have to send multiple BGP updates to signal service labels. Each PE sends label blocks in BGP updates to each of its BGP neighbors for all label blocks in which at least one ve-id has been seen by this PE (it does not advertise label blocks which do not contain an active ve-id, where active ve-id means the ve-id of this PE or any other PE in this VPLS).

The **max-ve-id** must be configured first, and determines the maximum value of the ve-id that can be configured within the PE. The ve-id value cannot be higher than this within the PE configuration, ve-id <= max-ve-id. Similarly, if the ve-id within a received NLRI is higher than the **max-ve-id** *value*, it will not be accepted as valid consequently the max-ve-id configured on all PEs must be greater than or equal to any ve-id used in the VPLS.

Only one ve-id value can be configured. If the ve-id value is changed, BGP withdraws the NLRI and sends a route-refresh.

If the same ve-id is used in different PEs for the same VPLS, a Designated Forwarder election takes place.

Executing the **shutdown** command triggers an MP-UNREACH-NLRI from the PE to all BGP peers.

The **no shutdown** command triggers an MP-REACH-NLRI to the same peers.

PE-2 Service Creation

On PE-2, a VPLS service using pseudowire template 1 is created. In order to make the label allocation more efficient, PE-2 has been allocated a ve-id value of 2. For completeness, the pseudowire template is also shown.

```
*A:PE-2# configure service
        pw-template 1 create
            split-horizon-group "VPLS-SHG"
            exit
        exit
        vpls 1 customer 1 create
            bgp
                route-distinguisher 65536:1
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpls
                max-ve-id 10
                ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
```

```
            exit
            service-name "VPLS1_PE-2"
            sap 1/1/4:1.0 create
            exit
            no shutdown
        exit
```

The **max-ve-id** *value* is set to 10 to allow an increase in the number of PEs that could be a part of this VPLS instance.

PE-3 Service Creation

Create a VPLS instance on PE-3, using a ve-id value of 3.

```
*A:PE-3# configure service
        pw-template 1 create
            split-horizon-group "VPLS-SHG"
            exit
        exit
        vpls 1 customer 1 create
            bgp
                route-distinguisher 65536:1
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpls
                max-ve-id 10
                ve-name "PE-3"
                    ve-id 3
                exit
                no shutdown
            exit
            service-name "VPLS1_PE-3"
            sap 1/1/4:1.0 create
            exit
            no shutdown
        exit
```

PE-1 Service Operation Verification

Verify that the BGP-VPLS site is enabled on PE-1.

```
*A:PE-1# show service id 1 bgp-vpls
===============================================================================
BGP VPLS Information
===============================================================================
Max Ve Id            : 10                   Admin State      : Enabled
VE Name              : PE-1                 VE Id            : 1
```

```
PW Tmpl used         : 1
===============================================================================
*A:PE-1#
```

Verify that the service is operationally up on PE-1.

```
*A:PE-1# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
Name              : VPLS1_PE-1
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/27/2016 12:17:24
Last Mgmt Change  : 05/27/2016 12:18:28
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 1
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:1.0                            qinq         1522    1522    Up   Up
sdp:17406:4294967294 SB(192.0.2.3)       BgpVpls      0       1556    Up   Up
sdp:17407:4294967295 SB(192.0.2.2)       BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The SAP and SDPs are all operationally up. The SB flags signify Spoke and BGP.

The ingress labels for PE-2 and PE-3, the labels allocated by PE-1, can be seen as follows:

```
*A:PE-1# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
```

```
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.3       Up      Up        262130    262128
17407:4294967295 BgpVpls  192.0.2.2       Up      Up        262129    262126
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

As can be seen from the following output, a BGP-VPLS NLRI update is sent to the route reflector (192.0.2.7) and is received by each PE.

The following debug trace from PE-1 shows the BGP NLRI update for VPLS 1 sent by PE-1 to the route reflector.

```
*A:PE-1# debug router bgp update

1 2016/05/27 12:17:27.18 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.7
"Peer 1: 192.0.2.7: UPDATE
Peer 1: 192.0.2.7 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 17, veid: 1, vbo: 1, vbs: 8, label-base: 262128,
                    RD 65536:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
"
```

The control flags within the extended community indicate the status of the VPLS instance.

New control flags have been introduced to allow support for BGP multi-homing, D indicates that all attachment circuits are Down, or the VPLS is shut down. The flags are used in BGP Multi-homing when determining which PEs are designated forwarders.

When flags=none, then all attachment circuits are up. In the preceding example, no flags are present, but should all SAPs become operationally down, then the following debug would be seen:

```
5 2016/05/27 12:21:31.19 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.7
"Peer 1: 192.0.2.7: UPDATE
Peer 1: 192.0.2.7 - Send BGP UPDATE:
```

```
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
       [VPLS/VPWS] preflen 17, veid: 1, vbo: 1, vbs: 8, label-base: 262128, RD 65536:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=1514: PREF=0
```

The BGP VPLS signaling parameters are also present, namely the ve-id of the PE within the VPLS instance, the VBO and VBS, and the label base. The target indicates the VPLS instance, which must be matched against the import route targets of the receiving PEs.

The signaling parameters can be seen within the BGP update with following command:

```
*A:PE-1# show router bgp routes l2-vpn rd 65536:1 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : VPLS
Route Dist.    : 65536:1
VeId           : 1                       Block Size     : 8
Base Offset    : 1                       Label Base     : 262128
Nexthop        : 192.0.2.1
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                     Interface Name : NotAvailable
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65536:1
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                       Dest Class     : 0

-------------------------------------------------------------------------------
```

```
Routes : 4
===============================================================================
*A:PE-1#
```

In this configuration example, PE-1 (192.0.2.1) with ve-id =1 has sent an update with base offset (VBO) =1, block size (VBS) = 8 and label base 262128. This means that labels 262128 (LB) to 262135 (LB+VBS-1) are available as de-multiplexer labels, egress labels to be used to reach PE-1 for VPLS 1.

PE-2 receives this update from PE-1. This is seen as a valid VPLS BGP route from PE-1 through the route reflector with nexthop 192.0.2.1.

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:1
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                    Prefix                              MED
      RD                           SiteId                              Label
      Nexthop                      VeId                    BlockSize   LocalPref
      As-Path                      BaseOffset              vplsLabelBa
                                                           se
-------------------------------------------------------------------------------
u*>i  VPLS                         -                       -           0
      65536:1                      -                                   -
      192.0.2.1                    1                       8           100
      No As-Path                   1                       262128
i     VPLS                         -                       -           0
      65536:1                      -                                   -
      192.0.2.2                    2                       8           100
      No As-Path                   1                       262126
u*>i  VPLS                         -                       -           0
      65536:1                      -                                   -
      192.0.2.3                    3                       8           100
      No As-Path                   1                       262128
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-2#
```

PE-2 uses this information in conjunction with its own ve-id to calculate the egress label toward PE-1, using the condition VBO ≤ ve-id < (VBO+VBS).

The ve-id of PE-2 is in the Label Block covered by VBO =1, thus,

Label calculation = label base + local ve-id - Base offset
= 262128 + 2 - 1
Egress label used = 262129

This is verified using the following command on PE-2 where the egress label toward PE-1 (192.0.2.1) is 262129.

```
*A:PE-2# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.3       Up      Up        262128    262129
17407:4294967295 BgpVpls  192.0.2.1       Up      Up        262126    262129
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

PE-3 also receives this update from PE-1 by the route reflector. This is seen as a valid VPLS BGP route from PE-1 with nexthop 192.0.2.1.

```
*A:PE-3# show router bgp routes l2-vpn rd 65536:1
===============================================================================
 BGP Router ID:192.0.2.3        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                  Prefix                           MED
      RD                         SiteId                           Label
      Nexthop                    VeId                 BlockSize   LocalPref
      As-Path                    BaseOffset           vplsLabelBa
                                                      se
-------------------------------------------------------------------------------
u*>i  VPLS                       -                    -           0
      65536:1                    -                                -
      192.0.2.1                  1                    8           100
      No As-Path                 1                    262128
u*>i  VPLS                       -                    -           0
      65536:1                    -                                -
      192.0.2.2                  2                    8           100
      No As-Path                 1                    262126
i     VPLS                       -                    -           0
      65536:1                    -                                -
      192.0.2.3                  3                    8           100
      No As-Path                 1                    262128
-------------------------------------------------------------------------------
```

```
Routes : 3
===============================================================================
*A:PE-3#
```

The ve-id of PE-3 is also in the label block covered by block offset VBO =1.

Label calculation= label base + local ve-id - VBO
= 262128 + 3 - 1
Egress label used = 262130

This is verified using the following command on PE-3 where egress label toward 192.0.2.1 is 262130.

```
*A:PE-3# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.1       Up      Up        262128    262130
17407:4294967295 BgpVpls  192.0.2.2       Up      Up        262129    262128
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

PE-2 Service Operation Verification

For completeness, verify the service is operationally up on PE-2.

```
*A:PE-2# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
Name              : VPLS1_PE-2
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/27/2016 12:17:02
Last Mgmt Change  : 05/27/2016 12:18:04
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 1
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
```

```
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:1.0                             qinq         1522    1522    Up   Up
sdp:17406:4294967294 SB(192.0.2.3)        BgpVpls      0       1556    Up   Up
sdp:17407:4294967295 SB(192.0.2.1)        BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-2#
```

PE-2 De-Multiplexer Label Calculation

In the same way that PE-1 allocates a label base (LB), block size (VBS), and base offset (VBO), PE-2 also allocates the same parameters for PE-1 and PE-3 to calculate the egress service label required to reach PE-2.

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:1 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
===============================================================================
BGP L2VPN Routes
===============================================================================

---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : VPLS
Route Dist.    : 65536:1
VeId           : 2                      Block Size     : 8
Base Offset    : 1                      Label Base     : 262126
Nexthop        : 192.0.2.2
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65536:1
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.7
Origin         : IGP
```

```
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-2#
```

This is verified using the following command on PE-1 to show the egress label toward PE-2 (192.0.2.2) where the egress label toward PE-2 = 262126 + 1- 1 = 262126.

```
*A:PE-1# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.3       Up      Up        262130    262128
17407:4294967295 BgpVpls  192.0.2.2       Up      Up        262129    262126
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

This is also verified using the following command on PE-3 to show the egress label toward PE-2 (192.0.2.2) where the egress label toward PE-2 = 262126 + 3 - 1 = 262128.

```
*A:PE-3# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.1       Up      Up        262128    262130
17407:4294967295 BgpVpls  192.0.2.2       Up      Up        262129    262128
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

PE-3 Service Operation Verification

Verify that the service is operationally up on PE-3:

```
*A:PE-3# show service id 1 base
===============================================================================
```

```
Service Basic Information
================================================================================
Service Id        : 1                    Vpn Id            : 0
Service Type      : VPLS
Name              : VPLS1_PE-3
Description       : (Not Specified)
Customer Id       : 1                    Creation Origin   : manual
Last Status Change: 05/27/2016 12:17:27
Last Mgmt Change  : 05/27/2016 12:17:29
Etree Mode        : Disabled
Admin State       : Up                   Oper State        : Up
MTU               : 1514                 Def. Mesh VC Id   : 1
SAP Count         : 1                    SDP Bind Count    : 2
Snd Flush on Fail : Disabled             Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled             Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled             Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled             Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

--------------------------------------------------------------------------------
Service Access & Destination Points
--------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
--------------------------------------------------------------------------------
sap:1/1/4:1.0                             qinq         1522    1522    Up   Up
sdp:17406:4294967294 SB(192.0.2.1)        BgpVpls      0       1556    Up   Up
sdp:17407:4294967295 SB(192.0.2.2)        BgpVpls      0       1556    Up   Up
================================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#

*A:PE-3# show service id 1 sdp
================================================================================
Services: Service Destination Points
================================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
--------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.1       Up      Up        262128    262130
17407:4294967295 BgpVpls  192.0.2.2       Up      Up        262129    262128
--------------------------------------------------------------------------------
Number of SDPs : 2
--------------------------------------------------------------------------------
================================================================================
*A:PE-3#
```

PE-3 De-Multiplexer Label Verification

PE-3 also allocates the required parameters for PE-1 and PE-2 to calculate the egress service label required to reach PE-3.

This is verified using the following command on PE-1 to show the egress label toward PE-3 (192.0.2.3) (262128) where egress label toward PE-2 = 262126. The Label Base equals 262128 on PE-3 and 262126 on PE-2.

```
*A:PE-1# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.3       Up      Up        262130    262128
17407:4294967295 BgpVpls  192.0.2.2       Up      Up        262129    262126
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

This is also verified using the following command on PE-2 to show the egress label toward PE-3 (192.0.2.3) which is using auto-provisioned SDP 17406.

```
*A:PE-2# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpVpls  192.0.2.3       Up      Up        262128    262129
17407:4294967295 BgpVpls  192.0.2.1       Up      Up        262126    262129
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

This example has shown that for VPLS instance with 3 PEs, not all labels allocated by a PE will be used by remote PEs as de-multiplexer service labels. There will be some wastage of label space, so there is a necessity to choose ve-ids that keep this waste to a minimum.

The next example will show an even more wasteful use of labels by using a random choice of ve-ids.

BGP VPLS Using Pre-Provisioned SDP

It is possible to configure BGP-VPLS instances that use RSVP-TE transport tunnels. In this case, the SDP must be created with the MPLS LSPs mapped and with signaling set to BGP, as the service labels are signaled using BGP. The pseudowire template configured within the BGP-VPLS instance must use the **use-provisioned-sdp** keyword.

This example also examines the effect of using ve-ids that are not all within the same contiguous block.

Figure 252 **BGP VPLS Using Pre-Provisioned SDP**

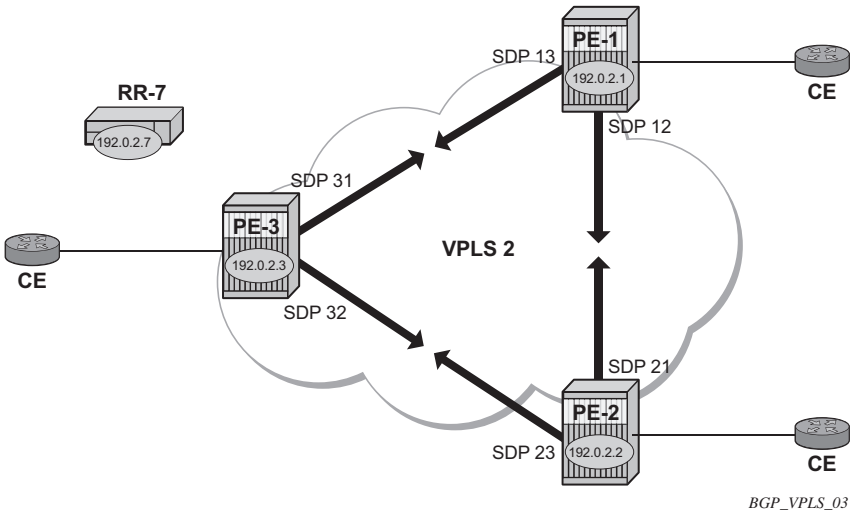


Figure 252 shows an example of a VPLS instance where SDPs are pre-provisioned with RSVP-TE signaled transport tunnels.

SDPs on PE-1

```
*A:PE-1# configure service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
        sdp 13 mpls create
            description "SDP-PE-1-PE-3_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-1-PE-3"
            no shutdown
        exit
```

SDPs on PE-2

```
*A:PE-2# configure service
        sdp 21 mpls create
            description "SDP-PE-2-PE-1_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
        sdp 23 mpls create
            description "SDP-PE-2-PE-3_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-2-PE-3"
            no shutdown
        exit
```

SDPs on PE-3

```
*A:PE-3# configure service
        sdp 31 mpls create
            description "SDP-PE-3-PE-1_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-3-PE-1"
            no shutdown
        exit
        sdp 32 mpls create
            description "SDP-PE-3-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-3-PE-2"
            no shutdown
        exit
```

Pre-provisioned BGP-SDPs can also be used with BGP-VPLS. For reference, they are configured as follows:

```
*A:PE-3# configure service
        sdp 332 mpls create
            signaling bgp
            far-end 192.0.2.2
            no shutdown
        exit
```

To create an SDP within a service that uses the RSVP transport tunnel, a pseudowire template is required that has the **use-provisioned-sdp** parameter set.

Once again, a split horizon group is included to prevent forwarding between pseudowires.

The pseudowire template must be provisioned on all PEs and looks like:

```
*A:PE-1# configure service
        pw-template 2 use-provisioned-sdp create
            split-horizon-group "VPLS-SHG"
            exit
        exit
```

The following output shows the configuration required for a BGP-VPLS service using a pseudowire template configured for using pre-provisioned RSVP-TE SDPs.

```
*A:PE-1# configure service
        vpls 2 customer 1 create
            bgp
                route-distinguisher 65536:2
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:2.0 create
            exit
            no shutdown
        exit
```

The route distinguisher and route target extended community values for VPLS 2 are different from the ones in VPLS 1. The ve-id value for PE-1 can be the same as the one in VPLS 1, but these must be different within the same VPLS instance on the other PEs — PE-2 should not have ve-id = 1.

On PE-2, the configuration is as follows with the ve-id value equal to 20, which will result in a label from a different block:

```
*A:PE-2# configure service
        vpls 2 customer 1 create
            bgp
                route-distinguisher 65536:2
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-2"
                    ve-id 20
                exit
                no shutdown
```

```
            exit
            sap 1/1/4:2.0 create
            exit
            no shutdown
        exit
```

and on PE-3:

```
*A:PE-3# configure service
        vpls 2 customer 1 create
            bgp
                route-distinguisher 65536:2
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-3"
                    ve-id 3
                exit
                no shutdown
            exit
            sap 1/1/4:2.0 create
            exit
            no shutdown
        exit
```

Verify that the service is operationally up on PE-1.

```
*A:PE-1# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/27/2016 12:31:36
Last Mgmt Change  : 05/27/2016 12:32:25
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 2
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
```

```
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2.0                            qinq         1522    1522    Up   Up
sdp:12:4294967292 S(192.0.2.2)           BgpVpls      0       1556    Up   Up
sdp:13:4294967293 S(192.0.2.3)           BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The SDP 12 and 13 are the pre-provisioned SDPs.

For completeness, verify the service is operationally up on PE-2

```
*A:PE-2# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                    Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                    Creation Origin   : manual
Last Status Change: 05/27/2016 12:30:54
Last Mgmt Change  : 05/27/2016 12:32:03
Etree Mode        : Disabled
Admin State       : Up                   Oper State        : Up
MTU               : 1514                 Def. Mesh VC Id   : 2
SAP Count         : 1                    SDP Bind Count    : 2
Snd Flush on Fail : Disabled             Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled             Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled             Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled             Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2.0                            qinq         1522    1522    Up   Up
sdp:21:4294967293 S(192.0.2.1)           BgpVpls      0       1556    Up   Up
sdp:23:4294967292 S(192.0.2.3)           BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-2#
```

Verify the service is operational on PE-3:

```
*A:PE-3# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/27/2016 12:31:16
Last Mgmt Change  : 05/27/2016 12:31:35
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 2
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2.0                            qinq         1522    1522    Up   Up
sdp:31:4294967293 S(192.0.2.1)           BgpVpls      0       1556    Up   Up
sdp:32:4294967292 S(192.0.2.2)           BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

PE-1 De-Multiplexer Label Calculation

In the case of VPLS 1, all ve-ids are in the range of a single label block. In the case of VPLS 2, the ve-ids are in different blocks, for example, the ve-id 20 is in a different block to ve-ids 1 and 3.

As the label allocation is block-dependent, multiple label blocks must be advertised by each PE to encompass this.

Consider PE-1's BGP update NLRIs.

```
*A:PE-1# show router bgp routes l2-vpn rd 65536:2 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP L2VPN Routes
===============================================================================
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : VPLS
Route Dist.    : 65536:2
VeId           : 1                       Block Size     : 8
Base Offset    : 1                       Label Base     : 262120
Nexthop        : 192.0.2.1
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                     Interface Name : NotAvailable
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65536:2
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                       Dest Class     : 0

Route Type     : VPLS
Route Dist.    : 65536:2
VeId           : 1                       Block Size     : 8
Base Offset    : 17                      Label Base     : 262112
Nexthop        : 192.0.2.1
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                     Interface Name : NotAvailable
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65536:2
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
```

```
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
-------------------------------------------------------------------------------
Routes : 8
===============================================================================
*A:PE-1#
```

Two NLRIs updates are sent to the route reflector, with the following label parameters:

1. LB = 262120, VBS = 8, VBO = 1

2. LB = 262112, VBS = 8, VBO = 17

PE-2 has a ve-id of 20. Applying the condition VBO ≤ ve-id < (VBO+VBS)

- Update 1: LB = 262120, VBS = 8, VBO = 1
- VBO ≤ ve-id for ve-id = 20 is TRUE
- ve-id < (VBO+VBS) for ve-id = 20 is FALSE.
- PE-2 cannot choose a label from this block.
- Update 2: LB = 262112, VBS = 8, VBO = 17
- VBO ≤ ve-id for ve-id = 20 is TRUE
- ve-id < (VBO+VBS) for ve-id = 20 is TRUE.
- PE-2 chooses label 262112 + 20 - 17 = 262115 (LB + veid - VBO)

The egress label chosen is verified by examining the egress label toward PE-1 (192.0.2.1) on PE-2.

```
*A:PE-2# show service id 2 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
21:4294967293    BgpVpls  192.0.2.1       Up      Up        262110    262115
23:4294967292    BgpVpls  192.0.2.3       Up      Up        262112    262115
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

PE-3 has a ve-id of 3. Applying the condition VBO ≤ ve-id < (VBO+VBS)

- Update 1: LB = 262120, VBS = 8, VBO = 1
- VBO ≤ ve-id for ve-id = 3 is TRUE

- ve-id < (VBO+VBS) for ve-id = 3 is TRUE.
- PE-3 chooses label 262120 + 3 - 1 = 262122 (LB + veid - VBO)

- Update 2: LB = 262112, VBS = 8, VBO = 17
- VBO ≤ ve-id for ve-id = 3 is FALSE
- ve-id < (VBO+VBS) for ve-id = 3 is TRUE.
- PE-3 cannot choose a label from this block.

The egress label chosen is verified by examining the egress label toward PE-1 (192.0.2.1) on PE-3.

```
*A:PE-3# show service id 2 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
31:4294967293    BgpVpls  192.0.2.1       Up      Up        262120    262122
32:4294967292    BgpVpls  192.0.2.2       Up      Up        262115    262112
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

To illustrate the allocation of label blocks by a PE, against the actual use of the same labels, consider the following. When BGP updates from each PE signal the multiplexer labels in blocks of eight, the allocated label values are added to the in-use pool. First check what label range can be allocated dynamically.

```
*A:PE-1# show router mpls-labels label-range
===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505824      505856
    Seg-Route   0           0           -           0           505856
===============================================================================
*A:PE-1#
```

Verify which labels in the dynamic range are in use. The label pool of PE-1 can be verified as per the following output which shows labels used along with the associated protocol:

```
*A:PE-1# show router mpls-labels label 18432 524287 in-use
===============================================================
MPLS Labels from 18432 to 524287 (In-use)
```

```
===============================================================
Label                Label Type          Label Owner
---------------------------------------------------------------
262110               dynamic             ILDP
262111               dynamic             ILDP
262112               dynamic             BGP
262113               dynamic             BGP
262114               dynamic             BGP
262115               dynamic             BGP
262116               dynamic             BGP
262117               dynamic             BGP
262118               dynamic             BGP
262119               dynamic             BGP
262120               dynamic             BGP
262121               dynamic             BGP
262122               dynamic             BGP
262123               dynamic             BGP
262124               dynamic             BGP
262125               dynamic             BGP
262126               dynamic             BGP
262127               dynamic             BGP
262128               dynamic             BGP
262129               dynamic             BGP
262130               dynamic             BGP
262131               dynamic             BGP
262132               dynamic             BGP
262133               dynamic             BGP
262134               dynamic             BGP
262135               dynamic             BGP
262136               dynamic             ILDP
262137               dynamic             ILDP
262140               dynamic             RSVP
262141               dynamic             RSVP
262142               dynamic             ILDP
262143               dynamic             ILDP
---------------------------------------------------------------
In-use labels (Owner: All) in specified range   : 32
In-use labels in entire range                   : 32
===============================================================
*A:PE-1#
```

This shows that 24 labels have been allocated for use by BGP. Of this number, 16 labels have been allocated for use by PEs within VPLS 2 to communicate with PE-1, the blocks with label base 262112 and with label base 262120.

There are only two neighboring PEs within this VPLS instance, so only two labels will ever be used in the data plane for traffic destined to PE-1. These are 262115 and 262122. The remaining labels have no PE with the associated ve-id that can use them.

Once again, this case emphasizes that to reduce label wastage, contiguous ve-ids in the range (N..N+7) per VPLS should be chosen, where N>0.

Conclusion

BGP-VPLS allows the delivery of Layer 2 VPN services to customers where BGP is commonly used. The examples presented in this chapter show the configuration of BGP-VPLS together with the associated show outputs which can be used for verification and troubleshooting.

EVPN for MPLS Tunnels

This chapter provides information about EVPN for MPLS tunnels.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 13.0.R6, but the CLI in the current edition corresponds to release 14.0.R3. A prerequisite is to read the EVPN for VXLAN Tunnels (Layer 2) chapter.

Overview

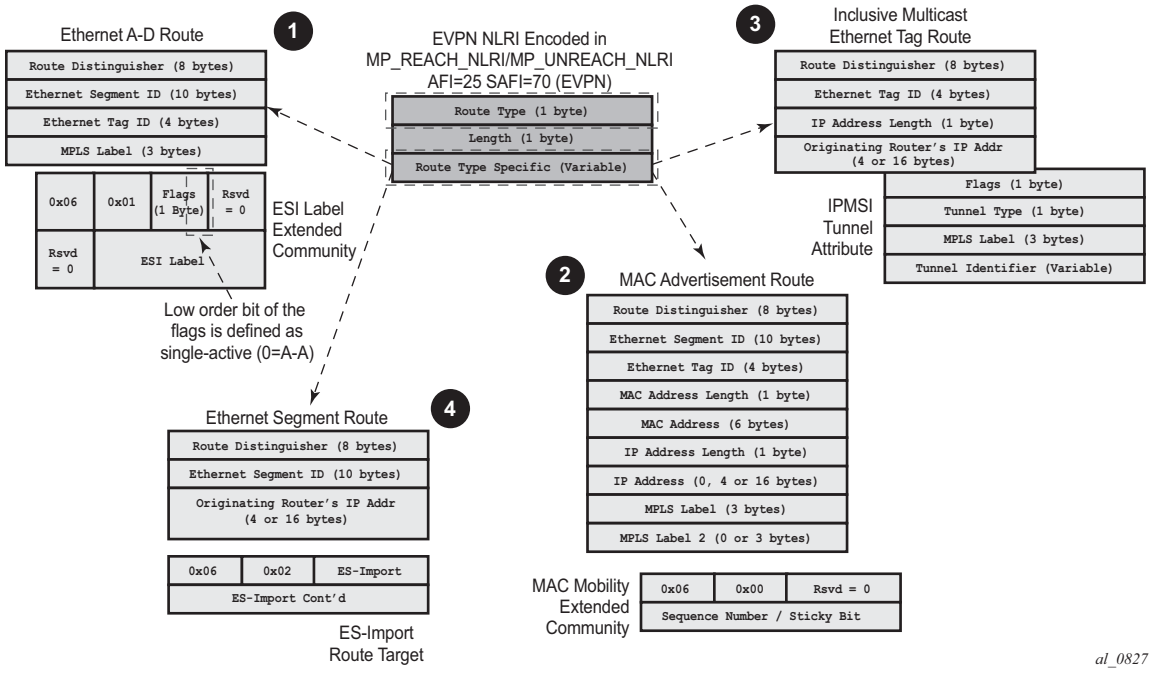
EVPN-MPLS is standardized in RFC 7432, *BGP MPLS-Based Ethernet VPN*, as a Layer 2 VPN technology that can supplement VPLS for ELAN services. Besides the optimizations introduced by EVPN, a significant number of service providers offering ELAN services today are requesting EVPN for their multi-homing capabilities. EVPN supports all-active multi-homing (per-flow load-balancing multi-homing) as well as single-active multi-homing (per-service load-balancing multi-homing). In addition to those superior multi-homing capabilities, EVPN also provides a number of significant benefits, such as:

- IP-VPN-like operation and control for ELAN services.
- Reduction and (in some cases) suppression of the BUM (broadcast, unknown unicast, and multicast) traffic in the network.
- Simple provisioning and management.
- New set of tools to control the distribution of MAC addresses and address resolution protocol (ARP) entries in the network.

While the VPN for virtual extensible local area network (VXLAN) tunnels (Layer-2) chapter focuses on the use of EVPN as a control plane for VXLAN tunnels, this chapter provides configuration guidelines for EVPN when used for MPLS tunnels. Similar to EVPN-VXLAN services, VPLS services with EVPN for MPLS tunnels are referred to as EVPN-MPLS services.

As a reference, the EVPN route types and NLRIs (Network Layer Reachability Information messages) used by the EVPN family in RFC 7432 are shown in Figure 253.

Figure 253 **EVPN Route Types and NLRIs**



When no EVPN multi-homing is used in the network, only the base routes are used. Route types 2 and 3 are considered the base and mandatory routes:

- Route type 2 - MAC/IP route: This route advertises MAC addresses to be installed in the remote FDBs, or MAC/IP address pairs to be installed in the remote proxy-ARP/ND (neighbor discovery) tables.
- Route type 3 - Inclusive multicast route: This route advertises the multicast tree that the advertising PE intends to use for sending BUM traffic for an EVI (EVPN Instance). Ingress replication, P2MP mLDP, and composite tunnels are supported as tunnel types in route type 3 when BGP-EVPN MPLS is enabled. The ingress replication information, as well as the downstream MPLS label (for remote PEs to send BUM traffic to the advertising PE) are encoded in the PTA (Provider Multicast Service Interface Tunnel Attribute).

When EVPN multi-homing is used in an EVI, routes type 1 and 4 are used (where type 1 has two different purposes):

- Route type 1 - Auto-discovery per Ethernet segment (AD per ES) route: This route is advertised per ES from the PE, carries the Ethernet segment identifier (ESI) label (used for split-horizon) in multi-homing mode, and can affect procedures such as the DF election, as well as the aliasing/backup path/mass withdrawal on remote PEs.
- Route-type 1 - Auto-discovery per EVPN instance (AD per-EVI) route: This route allows the remote PEs to provide aliasing and a backup path to the PEs part of the ES.
- Route type 4 - Ethernet segment (ES) route: This route advertises a local configured ES. The exchange of this route can discover remote PEs that are part of the same ES and the Designated Forwarder (DF) election algorithm among them.

The AD per-EVI, MAC/IP, and inclusive multicast routes are considered service-level BGP-EVPN routes. Their RT/RD (Route-Target/Route-Distinguisher) are taken from the VPLS configuration.

The AD per-ES and the ES routes are considered base-level BGP-EVPN routes. However, their RT/RD are taken differently:

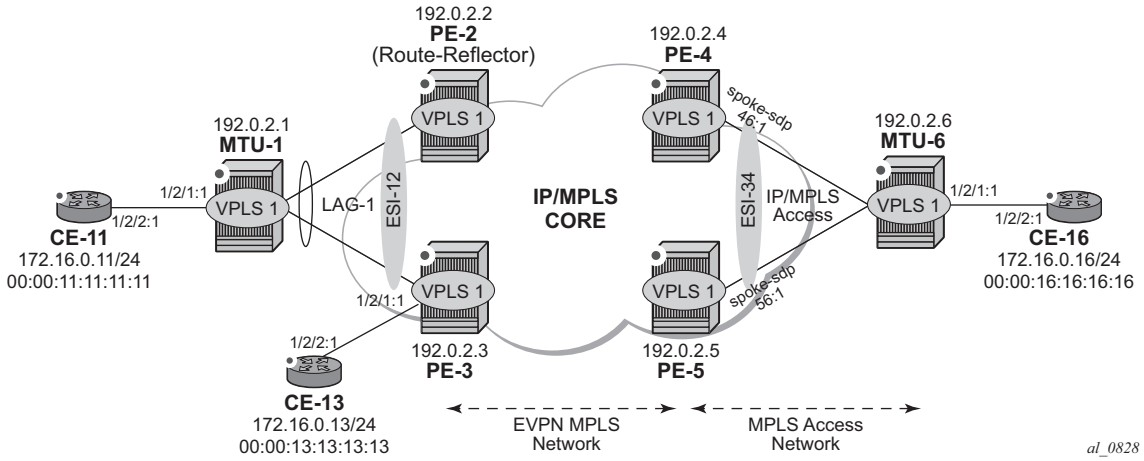
- The ES route RD is taken from the **service>system>bgp-evpn** configuration. The ES route RT is auto-derived from the ethernet-segment.
- The AD per-ES route RD is taken from the system level RD or service level RD. The RT extended community is taken from the service level RT or an RT set for the services defined on the ethernet-segment.

Configuration

This section describes the configuration of EVPN-MPLS for Layer 2 services on SR OS, as well as the available troubleshooting and show commands, and EVPN multi-homing.

Figure 254 shows the topology used throughout this chapter. The network consists of a core with four EVPN PEs (PE-2, PE-3, PE-4, and PE-5) and two MTU devices that are dual-homed to the EVPN network. For MTU-1, all-active multi-homing is used, whereas MTU-6 is connected via single-active multi-homing to the EVPN network. Three CEs are connected to VPLS 1 in MTU-1, PE-3, and MTU-6 in order to test the connectivity.

Figure 254 **EVPN-MPLS for VPLS Services**



As part of the network infrastructure configuration, the following settings and protocols must be added to the configuration before starting with the EVPN-specific configuration for the services:

- The ports interconnecting the four PEs in the core are configured as network ports (or hybrid) and will have router network interfaces defined in them. The ports on PE-2/PE-3 connected to MTU-1 can be access or hybrid ports, whereas the ports on PE-4/PE-5 connected to MTU-6 can be network or hybrid ports. In case of hybrid ports, no LACP can be configured.
- The four PEs in the core (as well as MTU-6 in the access MPLS network) are running IS-IS and establishing point-to-point adjacencies for the exchange of the system IP addresses.
- LDP is used as the MPLS protocol to signal transport tunnel labels among PE-2, PE-3, PE-4, PE-5, and MTU-6. There is no LDP running between MTU-1 and the rest of the network, that is, MTU-1 is a pure Ethernet aggregation device.
- EVPN uses MP-BGP for exchanging reachability at service level. Therefore, BGP peering sessions must be established among the core PEs for the EVPN family. Although typically a separate router is used, in this chapter, PE-2 is used as BGP RR (route reflector) for EVPN routes. For example, the following output shows the configuration of BGP in the RR and one of the BGP clients. The relevant commands for EVPN are shown in bold.

  The configuration on the route reflector PE-2 is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
```

```
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

The BGP configuration on the clients PE-3, PE-4, and PE-5 is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
```

Note: The **def-recv-evpn-encap** command is not used in the preceding configuration because the default MPLS configuration is sufficient to have a correct interpretation of the received EVPN encapsulations.

The default EVPN encapsulation type is MPLS, and can be verified as follows:

```
*A:PE-3# configure router bgp group "internal" neighbor 192.0.2.2
*A:PE-3>config>router>bgp>group>neighbor# info detail | match def-recv
                    def-recv-evpn-encap mpls
```

The EVPN encapsulation type can be configured as MPLS or VXLAN, as follows:

```
*A:PE-3# configure router bgp group "internal" neighbor 192.0.2.2 def-recv-evpn-encap
  - def-recv-evpn-encap <encap-type>

 <encap-type>         : mpls|vxlan
```

EVPN routes type 1 (auto-discovery per-EVI route), type 2 (MAC/IP route), type 3 (inclusive multicast route), and type 5 (IP-prefix route) are always sent with the RFC 5512, *the BGP Encapsulation Subsequent Address Family Identifier (SAFI) and the BGP Tunnel Encapsulation Attribute*, BGP encapsulation extended community that indicates the associated encapsulation of the route. Because the use of this extended community is not mandatory in RFC 7432, the def-recv-evpn-encap command indicates to the system what encapsulation is associated with routes received without any encapsulation. By default, this command is set to mpls. When interoperating with third-party EVPN vendors in mixed MPLS and EVPN-VXLAN networks, this command should be revised accordingly.

EVPN-MPLS Configuration without Multi-Homing

After the base infrastructure (interfaces, IGP, LDP, BGP protocols) is configured, the service and EVPN can be enabled. When no multi-homing is used, the EVPN-MPLS configuration in a VPLS service looks similar to the configuration of EVPN-VXLAN for Layer 2, except for the commands related to the MPLS data plane. The following output shows the VPLS-1 configuration in PE-3 as an example:

```
configure
    service
        vpls 1 customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/2/1:1 create
            exit
            sap lag-1:1 create
            exit
            no shutdown
```

Where the following commands are relevant for a basic EVPN configuration:

- **bgp** enables the context for the BGP configuration relevant to the service. If a manual (non-auto-derived) RD/RT, as well as import/export policies, are needed for the service, the commands in the **bgp** context must be configured. When **bgp-evpn** is enabled in a VPLS instance, other families are supported within the same service (bgp-ad and bgp-mh, not bgp-vpls). This **bgp** context configures the common BGP parameters for all the BGP families in the service. The **pw-template-binding** command is ignored for **bgp-evpn**. Even if the general BGP parameters for the service are auto-derived (as in this example), the **bgp** context must be enabled.

```
*A:PE-3>config>service>vpls# bgp ?
  - bgp
  - no bgp

 [no] pw-template-bi* + Configure pw-template bind policy
 [no] route-distingu* - Configure route distinguisher
 [no] route-target    - Configure route target
 [no] vsi-export      - VSI export route policies
 [no] vsi-import      - VSI import route policies
```

- **bgp-evpn evi <1..65535>** — The EVPN instance or **evi** is a 2-byte identifier used for the auto-derivation of the service RD, service RT, and for the service-carving algorithm when multi-homing is used. The evi can be used for both **bgp-evpn vxlan** and **bgp-evpn mpls** when the user needs to auto-derive the RD and RT for the service. The auto-derivation is always based on:
  - RD system-ip:evi
  - RT autonomous-system:evi

  The configured and operating RD/RT values can be checked with the following show command (in this example, the evi value is 1):

```
*A:PE-3# show service id 1 bgp
===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.3:1
Oper RD Type         : derivedEvi
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64500:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64500:1
PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

Although not required for a basic BGP-EVPN MPLS configuration, some other parameters may be used at the bgp-evpn context level, when EVPN-MPLS services are deployed. Some examples are listed here:

- **bgp-evpn>cfm-mac-advertisement** must be enabled when eth-cfm is used across an EVPN-MPLS service among different PEs. If a maintenance endpoint (MEP) or maintenance domain intermediate point (MIP) is configured in any of the SAP/SDP bindings in the VPLS and has to exchange eth-cfm packets with a remote MEP/MIP across the EVPN-MPLS core, this command must be enabled. In that way, the MEP/MIP MAC address can be advertised in EVPN (otherwise, the MEP/MIP MAC address would not be learned on remote EVPN-MPLS PEs and eth-cfm would not work correctly).
- **bgp-evpn>mac-advertisement** and **bgp-evpn>mac-duplication** — See the EVPN for VXLAN Tunnels (Layer 2) chapter for a description of these two commands.

  bgp-evpn>vxlan must be **shutdown** so that **bgp-evpn mpls** can be enabled.

BGP-EVPN MPLS and BGP-EVPN VXLAN are mutually exclusive, as can be verified as follows:

```
*A:PE-3# configure service vpls 1
*A:PE-3>config>service>vpls# vxlan vni 1 create
*A:PE-3>config>service>vpls>vxlan$ exit
*A:PE-3>config>service>vpls# bgp-evpn
*A:PE-3>config>service>vpls>bgp-evpn# vxlan no shutdown
MINOR: SVCMGR #8025 Evpn mpls is enabled
*A:PE-3>config>service>vpls>bgp-evpn# mpls shutdown
*A:PE-3>config>service>vpls>bgp-evpn# vxlan no shutdown
*A:PE-3>config>service>vpls>bgp-evpn# mpls no shutdown
MINOR: SVCMGR #8026 Evpn vxlan is enabled
```

After the relevant **VPLS** parameters, **BGP** and **bgp-evpn** attributes are added, the specific commands for **bgp-evpn mpls** can be configured:

```
*A:PE-3>config>service>vpls>bgp-evpn>mpls# info
----------------------------------------------
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
----------------------------------------------
```

- **ingress-replication-bum-label** controls whether the system will advertise different service labels for unicast and BUM traffic. If no EVPN multi-homing is configured in the network, this command can be disabled (**no ingress-replication-bum-label**) and the same MPLS label will be advertised for the unicast and BUM traffic for the VPLS instance. If EVPN multi-homing is configured in the PE, this command is strongly recommended to avoid potential transient issues. See the EVPN-MPLS multi-homing section.

- **ecmp** controls the number of remote PEs to which the local PE can load balance the unicast traffic. See the EVPN multi-homing section.
- **auto-bind-tunnel** controls the resolution of EVPN destinations to MPLS transport tunnels. This command is also in VPRN services and works in the same way.
  - If the **auto-bind-tunnel resolution any** is configured, as in the example, EVPN destinations in the service are resolved based on the best tunnel in the tunnel table manager (TTM). For instance, the following command shows the existing EVPN destinations for VPLS 1 in PE-3. The EVPN-MPLS destination (Termination endpoint (TEP) 192.0.2.2, label 262140) is resolved to an LDP transport tunnel because the (best) LDP tunnel to 192.0.2.2 shown in the **show router tunnel-table** is LDP. If there was more than one tunnel type in the TTM to 192.0.2.2, the system would pick the lowest **Pref** (preference) tunnel.

```
*A:PE-3# show service id 1 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.2       262140        0           Yes             06/27/2016 08:09:05
                ldp
192.0.2.4       262140        0           Yes             06/27/2016 08:09:05
                ldp
192.0.2.5       262140        0           Yes             06/27/2016 08:09:05
                ldp
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
---snip---

*A:PE-3# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.2/32      ldp       MPLS  65540     9        192.168.23.1    10
192.0.2.4/32      ldp       MPLS  65537     9        192.168.34.2    10
192.0.2.5/32      ldp       MPLS  65539     9        192.168.35.2    10
192.0.2.6/32      ldp       MPLS  65538     9        192.168.34.2    20
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
```

  - If resolution is set to any, the following tunnel types are selected in order of preference: RSVP, LDP, Segment Routing, and BGP. The user can configure the preference of the segment-routing tunnel type in the TTM for a specific IGP instance.

– If one or more explicit tunnel types are specified using the resolution-filter option, then only these tunnel types will be selected again following the TTM preference.
– The user must set the resolution to filter to activate the list of tunnel-types configured under resolution-filter.

Although not shown in the **bgp-evpn mpls** basic configuration for PE-3, there are other parameters that can be modified:

```
*A:PE-3# configure service vpls 1 bgp-evpn mpls
  - mpls

      auto-bind-tunn* + Configure BGP EVPN mpls auto-bind-tunnel
 [no] control-word    - Enable/disable setting the CW bit in the label message
      ecmp            - Configure maximum ECMP routes information
 [no] entropy-label   - Enable/disable use of entropy-label
 [no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
 [no] ingress-replic* - Use the same label as the one advertised for unicast traffic
 [no] restrict-prote* - Enable/disable protected src MAC restriction
 [no] shutdown        - Administratively Enable/Disable BGP-EVPN mpls
 [no] split-horizon-* - Configure a split-horizon-group
```

- **control-word** enables/disables the insertion of the control-word in the data path. The control-word is disabled by default and is not signaled in EVPN (based on RFC 7432) and has to be consistently configured in all the PEs in the network. The use of the **control-word** prevents packet-reordering from happening in P routers that misinterpret the first nibble of the payload in the packets they receive. In some third-party EVPN vendors, the control-word is enabled by default, so it is recommended to enable it when interoperating with other vendors.
- **entropy-label** enables the use of entropy labels. This is beyond the scope of this chapter.
- **force-vlan-vc-forwarding** allows the system to preserve the vlan-id and pbits of the service-delimiting qtag in a new tag added in the customer frame before sending it to the EVPN core. This command may be used with the **sap ingress vlan-translation** command: the configured translated vlan-id will be sent to the EVPN binds, as opposed to the service-delimiting tag vlan-id. If the ingress SAP/SDP-binding is null encapsulated, the output vlan-id and pbits will be zero.
- **restrict-protected-src** is by default disabled. When enabled, all packets entering the object will be verified not to contain a protected source MAC address. In combination with the parameter **discard-frame**, the packets that contain a protected MAC address will be discarded and an alarm is generated.
- **shutdown** enables/disables the use of MPLS for EVPN. When **mpls no shutdown** is issued, a BGP route-refresh message is sent for the EVPN family.

- **split-horizon-group** <*group-name*> configures an explicit split-horizon-group (SHG) for all the EVPN destinations that can be shared with other SAP/SDP-bindings. See the VPLS to EVPN-MPLS migration and integration section.

After **bgp-evpn mpls** is configured in the service, and **no shutdown**, an inclusive multicast route is sent to the RR. The remote PEs receiving and importing that route will create an EVPN destination to the sending PE. An EVPN destination is identified by a TEP and MPLS label. Use the following show commands to view the service and the EVPN destinations created:

```
show service evpn-mpls
show service id 1 evpn-mpls
show service id 1 bgp-evpn
```

An example of the output is shown for PE-2 when there is no traffic in the network. Therefore, only inclusive multicast routes have been exchanged among the four PEs.

```
*A:PE-2# show service evpn-mpls
===============================================================================
EVPN MPLS Tunnel Endpoints
===============================================================================
EvpnMplsTEP Address EVPN-MPLS Dest      ES Dest             ES BMac Dest
-------------------------------------------------------------------------------
192.0.2.3           1                   0                   0
192.0.2.4           1                   0                   0
192.0.2.5           1                   0                   0
-------------------------------------------------------------------------------
Number of EvpnMpls Tunnel Endpoints: 3
-------------------------------------------------------------------------------
===============================================================================


*A:PE-2# show service id 1 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.3       262140        0           Yes             06/27/2016 08:09:05
                ldp
192.0.2.4       262140        0           Yes             06/27/2016 08:09:05
                ldp
192.0.2.5       262140        0           Yes             06/27/2016 08:09:05
                ldp
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
```

```
===============================================================================
===============================================================================
BGP EVPN-MPLS ES BMAC Dest
===============================================================================
ES BMAC Addr                            Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
*A:PE-2#

*A:PE-2# show service id 1 bgp-evpn
===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled           Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
VXLAN Admin Status : Disabled          Creation Origin    : manual
MAC Dup Detn Moves : 5                 MAC Dup Detn Window: 3
MAC Dup Detn Retry : 9                 Number of Dup MACs : 0
IP Route Advert    : Disabled

EVI                : 1
Ing Rep Inc McastAd: Enabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses            Time Detected
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN MPLS Information
===============================================================================
Admin Status       : Enabled
Force Vlan Fwding  : Disabled          Control Word       : Disabled
Split Horizon Group: (Not Specified)
Ingress Rep BUM Lbl: Enabled           Max Ecmp Routes    : 2
Ingress Ucast Lbl  : 262141            Ingress Mcast Lbl  : 262140
Entropy Label      : Disabled
RestProtSrcMacAct  : none
===============================================================================
===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Resolution         : any
Filter Tunnel Types: (Not Specified)
===============================================================================
```

When traffic is generated, the PEs will start learning MAC addresses and advertising them in BGP so that the remote PEs learn those MAC addresses against EVPN destinations. For instance, when CE-13 sends traffic, PE-3 learns its MAC address and advertises it. The remote PEs (for instance, PE-2) will learn the MAC address and associate it with their EVPN destination to PE-3 (192.0.2.3:262141 in this example):

```
*A:PE-2# show service id 1 fdb detail
===============================================================================
```

```
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:11:11:11:11 sap:lag-1:1              L/106    06/28/16 01:50:54
1         00:00:13:13:13:13 eMpls:                   Evpn     06/28/16 01:50:54
                            192.0.2.3:262141
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

When the **ingress-replication-bum-label** is enabled in the PEs, the advertisement of MAC addresses will create new EVPN destinations, because the label is different from the one previously sent by the inclusive multicast route that created an EVPN destination. In the preceding example, when PE-3 advertises the CE-13 MAC address, PE-2 will create a new binding (see in the following output in bold) that shows one MAC address that is not Mcast (multicast) capable:

```
*A:PE-2# show service id 1 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.3       262140        0           Yes             06/28/2016 01:50:54
                ldp
192.0.2.3       262141        1           No              06/28/2016 02:03:23
                ldp
192.0.2.4       262140        0           Yes             06/27/2016 08:09:05
                ldp
192.0.2.5       262140        0           Yes             06/27/2016 08:09:05
                ldp
-------------------------------------------------------------------------------
Number of entries : 4
-------------------------------------------------------------------------------
===============================================================================
```

When an EVPN-MPLS destination or MAC address is not created/installed correctly, the user may check the BGP-EVPN routes received and the routes kept in the RIB. The routes that the PE receives are shown when **debug router bgp update** is enabled. These routes are shown even before any BGP processing is carried out.

```
*A:PE-2#
4 2016/06/27 11:41:16.27 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
```

```
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:13:13:13:13, IP len: 0, IP: NULL, label1: 4194256
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
"


*A:PE-2# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.          MacAddr           ESI
      Tag                  Mac Mobility      Ip Address
                                             NextHop
                                             Label1
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1          00:00:13:13:13:13 ESI-0
      0                    Seq:0             N/A
                                             192.0.2.3
                                             LABEL 262141

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

If the route is successfully imported, it can be shown in the RIB (**show router bgp routes** commands). The route shown in the debug and the same route in a show command do not necessarily have the same label value. The reason for this expected mismatch is that the debug command shows the complete 24-bit field value because the route is shown before BGP can decide and decipher whether the label value is an MPLS label (high-order 20-bits of the label field) or a VNI (all 24 bits of the Label field for VXLAN).

VPLS to EVPN-MPLS Integration

The SR OS EVPN implementation supports draft-ietf-bess-evpn-vpls-seamless-integ-00 so that EVPN-MPLS and VPLS can be integrated into the same network and within the same service.

The following behavior enables the integration of EVPN and SDP-bindings in the same VPLS network:

- Systems with EVPN endpoints and SDP-bindings to the same far-end bring down the SDP-bindings.
  - SROS will allow the establishment of an EVPN destination and an SDP-binding to the same far-end but the SDP-binding will be kept operationally down. Only the EVPN endpoint will be operationally up. This is true for spoke-sdps (manual and BGP-AD) and mesh-sdps. It is also true between VXLAN and SDP-bindings.
  - If there is an EVPN endpoint to a specified far-end and a spoke-sdp establishment is attempted, the spoke-sdp will be set up but kept down with an operational flag indicating that there is an EVPN route to the same far-end.
  - If there is a spoke-sdp and a valid/used EVPN route arrives, the EVPN endpoint will be set up and the spoke-sdp will be brought down with an operational flag indicating that there is an EVPN route to the same far-end.
  - In the case of an SDP-binding and EVPN endpoint to different far-end IPs on the same remote PE, both links will be up. This can happen if the SDP-binding is terminated in an IPv6 address or IPv4 address different from the system address where the EVPN endpoint is terminated.

  The following example illustrates the preceding description. A spoke SDP is added to the VPLS 1 configuration on PE-2:

```
configure
    service
        sdp 24 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        vpls 1
            spoke-sdp 24:1 create
            exit
        exit
```

The service configuration on PE-4 is as follows:

```
configure
    service
        sdp 42 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        vpls 1 customer 1 create
```

```
            bgp
            exit
            bgp-evpn
                evi 1
                mpls
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 42:1 create
            exit
            spoke-sdp 46:1 create
            exit
            no shutdown
```

Spoke SDP 24:1 is operationally down, as can be verified as follows:

```
*A:PE-2# show service id 1 base

---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm   Opr
-------------------------------------------------------------------------------
sap:lag-1:1                             q-tag        1518    1518    Up    Up
sdp:24:1 S(192.0.2.4)                   Spok         0       8978    Up    Down
===============================================================================
```

Spoke SDP 24:1 is down because of an EVPN route conflict, as indicated by the flags:

```
*A:PE-2# show service id 1 sdp 24 detail | match Flag context all
Flags              : PWPeerFaultStatusBits
                     EvpnRouteConflict
```

- The user can add spoke-sdps and all the EVPN-MPLS endpoints in the same SHG.
  - A CLI command exists in the **bgp-evpn>mpls** context so that the EVPN-MPLS endpoints can be added to an SHG.
  - The **bgp-evpn mpls split-horizon-group** must reference a user-configured split-horizon-group. User-configured split-horizon-groups can be configured within the service context.
  - The same group-name can be associated with saps, spoke-sdps, pw-templates, pw-template-bindings, and EVPN-MPLS endpoints.
  - If the **split-horizon-group** command in **bgp-evpn>mpls** is not used, the default split-horizon-group (in which all the EVPN endpoints are) is still used, but it will not be possible to refer to it on SAPs/spoke-SDPs.

- The system disables the advertisement of MAC addresses learned on spoke SDPs/SAPs that are part of an EVPN split-horizon-group.
  - When the SAPs or spoke SDPs (manual or BGP-AD-discovered) are configured within the same SHG as the EVPN endpoints, MAC addresses will still be learned on them, but will not be advertised in EVPN.
  - The preceding statement is also true if proxy-ARP/ND is enabled and an IP->MAC address pair is learned on a sap/sdp-binding that belongs to the EVPN SHG.
  - The SAPs and/or spoke-SDPs added to an EVPN SHG should not be part of any EVPN multi-homed ES. If that happened, the PE would still advertise the AD per-EVI route for the SAP and/or spoke-SDP, attracting EVPN traffic that could not be forwarded to that SAP and/or SDP-binding.
  - Similar to the preceding statement, an SHG composed of SAPs/SDP-bindings used in a BGP-MH site should not be configured under **bgp-evpn>mpls>split-horizon-group**. This misconfiguration would prevent traffic being forwarded from the EVPN to the BGP-MH site, regardless of the DF/Non-DF state.

An example of a shared SHG configuration on PE-2 is as follows. Because the SAP and evpn-mpls are in the same SHG, no MAC addresses learned over SAP 1/2/1:2 will be advertised in EVPN (not even static MACs).

```
configure
    service
        vpls 2 customer 1 create
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                mpls
                    split-horizon-group "CORE"
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/2/1:2 split-horizon-group "CORE" create
            exit
            sap lag-1:2 create
            exit
            no shutdown
```

EVPN-MPLS Multi-Homing

SR OS supports EVPN multi-homing as per RFC 7432.

The EVPN multi-homing implementation is based on the concept of the ES. An ES is a logical structure that can be defined in one or more PEs and identifies the CE (or access network) multi-homed to the EVPN PEs. An ES is associated with a port, LAG, or SDP object, and is shared by all the services defined on those objects.

Each ES has a unique identifier called ESI (Ethernet Segment Identifier) that is 10 bytes and is manually configured in the current release. The ESI is advertised in the control plane to all the PEs in an EVPN network; therefore, it is very important to ensure that the 10-byte ESI value is unique throughout the entire network. Single-homed CEs are assumed to be connected to an ES with ESI = 0 (single-homed ESs are not explicitly configured).

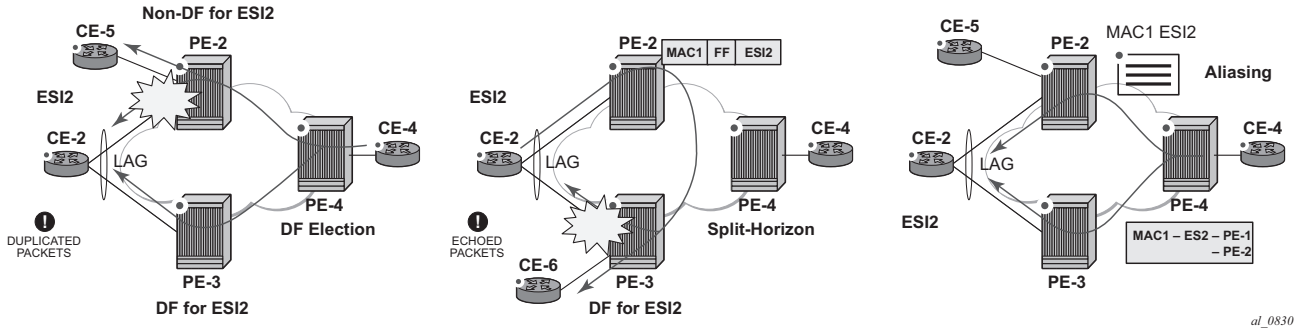
The ES is part of the base BGP-EVPN configuration and is not applied to any EVPN-MPLS service, by default. An ES can be shared by multiple services; the association of a specific SAP or spoke-SDP to an ES is automatically made when the SAP is defined in the same LAG or port configured in the ES, or when the spoke-SDP is defined in the same SDP configured in the ES. The following sections show the configuration of:

- an all-active multi-homing ES with a LAG associated with it
- a single-active multi-homing ES linked to an SDP

All-Active Multi-Homing Concepts

EVPN all-active multi-homing is built around three concepts: DF election, split-horizon (with an ESI-label), and aliasing, as shown in Figure 255, from left to right.

Figure 255 **EVPN-MPLS All-Active Multi-Homing Concepts**



- With DF election, when PE-4 sends BUM traffic to the remote ES (CE-2), only one PE segment sends the BUM packets to the ES (PE-3 is the DF in the preceding example, and is elected to send BUM packets to CE-2). The non-DF, PE-2, removes the LAG SAP from the default multicast list (PE-2 does not bring CE-2 down, because it still needs to send upstream/downstream unicast traffic). PE-2 and PE-3 elect a DF for each service, based on the ES routes and the service-carving algorithm.
- With split-horizon, the PE part of the ES (PE-3 in the preceding example) identifies the BUM packets coming from the PE for the remote (PE-2), but within the same ES (ESI-2), and filters the packets so that they are not sent back to the ES, creating duplication. When PE-2 (non-DF) sends BUM traffic to PE-3 (DF), it uses a special MPLS label in the data path that PE-3 previously advertised for ESI-2 in an AD per-ES route. When PE-3 does an ingress lookup, it recognizes the ESI-label and filters the traffic (PE-3 still sends the BUM traffic to other SAPs/SDP-bindings).
- With aliasing, remote PEs that are not part of the ES can load-balance unicast traffic to all the PEs that are part of the ES, irrespective of from which PE a destination MAC address was learned. PE-4 will create an EVPN destination to ESI-2 that will be resolved to the two next-hops: PE-2 and PE-3. Unicast load-balancing will happen as long as ecmp > 1 is enabled in PE-4.

Nokia recommends the use of **ingress-replication-bum-label** on the PEs that are part of an all-active ES. In an all-active multi-homing scenario, if a specified MAC address (for example, the CE-2 MAC address in the left-hand-side diagram), is not learned yet in a remote PE (for example, PE-4), but is known in the two PEs of the ES (for example, PE-2 and PE-3), the latter PEs might send duplicated packets to the CE.

This issue is solved by the use of **ingress-replication-bum-label** in PE-2 and PE-3. If configured, PE-2/PE-3 will know that the received packet is an unknown unicast packet; therefore, the Non-DF (PE-2) will not send the packet to CE-2 and there will not be duplication.

All-Active Multi-Homing Configuration

The all-active multi-homing configuration example is based on Figure 254.

MTU-1 is connected to the EVPN network using all-active multi-homing. According to RFC 7432, MTU-1 will be able to send traffic to both PEs for VPLS-1. Regular LAG load-balancing is used in MTU-1. Remote PEs such as PE-4 or PE-5 will be able to load-balance the unicast traffic to PE-2 and PE-3. PE-2 and PE-3 will discover that both are part of ESI-12 (due to the exchange of ES routes) and will elect a DF for VPLS-1. The non-DF for VPLS-1, in this case PE-2, will remove lag-1:1 from the VPLS-1 default multicast list. Also, when PE-2 and PE-3 send BUM traffic to each other, they will insert an ESI-label so that they can identify that the source of the BUM packet is ESI-12.

The following output shows the configuration of ESI-12 in PE-2 and PE-3, as well as the LAG interfaces for all-active multi-homing (see Figure 254). The configuration of LAG-1 in MTU-1 is also shown. Per RFC 7432, only a CE/MTU with a LAG can be connected to an all-active multi-homing ES. No other configuration is permitted on the CE for all-active multi-homing.

The configuration of LAG 1 on MTU-1 is as follows:

```
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 32768
        no shutdown
```

LAG 1 is configured as follows on PE-2:

```
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/2
        lacp active administrative-key 1 system-id 00:00:00:00:02:03
        no shutdown
```

LAG 1 is configured as follows on PE-3:

```
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/1
        lacp active administrative-key 1 system-id 00:00:00:00:02:03
        no shutdown
```

Ethernet segment “ESI-12” is configured in the service system bgp-evpn context on PE-2 and PE-3, as follows:

```
configure
```

```
service
    system
        bgp-evpn
            ethernet-segment "ESI-12" create
                esi 01:00:00:00:00:12:00:00:00:01
                es-activation-timer 3
                service-carving
                    mode auto
                exit
                multi-homing all-active
                lag 1
                no shutdown
            exit
```

When configuring an ES, the following must be considered:

- Any EVPN parameter that is not specific to any particular VPLS service, and is common to all the EVIs, is configured in a base BGP-EVPN instance located at **config>service>system>bgp-evpn**. In this base instance, the following attributes may be configured:
  - **ethernet-segments**
  - the base BGP-EVPN instance **route-distinguisher** that will be used for the ES routes. If this **route-distinguisher** is not configured, by default a type-1 RD will be derived as system-ip:0, as shown in the command help:

```
*A:PE-2>config>service>system>bgp-evpn# route-distinguisher
  - no route-distinguisher
  - route-distinguisher <rd>

 <rd>                 : <ip-addr:comm-val>
                        ip-addr        - a.b.c.d
                        comm-val       - [0..65535]
                        default: system-ip:0
```

- The ES must be configured with a name and can contain the following parameters when configured for all-active multi-homing:
  - **esi** — 10-byte identifier that represents the ES in the BGP control plane. The same esi must be configured in all the PEs connected to the same CE/MTU (using a unique value that cannot be associated with any other CE/MTU/access network). RFC 7432 defines five different types of **esi**. In SR OS, the **type** byte, as well as the other 9 bytes can be arbitrarily configured.
  - **multi-homing all-active** — This command indicates that the ES is in all-active mode.

– **lag** <*lag-id*> — The LAG connected to the CE/MTU must be added to the ES. In this example, lag-1 is added to ESI-12, on both PE-2 and PE-3. Although a different lag-id may have been assigned to the same ES on PE-2 and PE-3, PE-2 and PE-3 must have the same configuration on the ES LAG; that is, encap-type. Also, if LACP is added (it is not mandatory), both PEs must have the same admin-key, system-id, and system-priority. MTU-1 will see PE-2 and PE-3 as a single LAG peer. For all-active multi-homing, only the **lag** option is accepted by the system; **port** or **sdp** are not accepted.

– [**no**] **shutdown** — This command controls the administrative state of the ES.

- The preceding parameters are the minimum necessary so that the ES can be activated. In addition to those parameters, there are a few more that the user can configure if requiring values different from the default ones:

  – **es-activation-timer** [**0..100**] can be configured at **redundancy>bgp-evpn-multi-homing>es-activation-timer** or at **service>system>bgp-evpn>eth-seg>es-activation-timer** level (the most specific value is used).

    The **es-activation-timer** operation is as follows:

    - Upon reception of an ES, AD per-ES/EVI route update/withdrawal for a local ESI, the DF-candidate list of IPs is updated and the DF election algorithm is run without waiting for any timer.
    - If the result of the DF election requires the PE to be promoted from non-DF to DF, the **es-activation-timer** will start, and only after its expiration will the PE add the SAP to the default-multicast list. Transitions from non-DF to non-DF, or from DF to non-DF, are immediate and do not wait for any timer.
    - This use of an **es-activation-timer** value minimizes the risks of loops and packet duplication due to **transient** multiple DFs.
    - The same **es-activation-timer** must be configured in all the PEs that are part of the same ESI. The user must configure either a long timer to minimize the risks of loops/duplication, or **es-activation-timer**=0 to speed up the convergence for NDF→ DF transitions. The default value is 3 seconds.

  – **service-carving —** As defined in RFC 7432, service-carving controls the distribution of DF/non-DF roles across the different services defined in an ES.

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving# mode
  - mode {manual|auto}

 <manual|auto>        : auto|manual|off


*A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving# manual
  - manual

 [no] evi             - Configure EVI range
```

```
[no] isid           - Configure ISID range
```

As shown above, **service-carving** has three different modes:

- **service-carving mode auto** (default) — The DF election algorithm will run the function [V(evi) mod N(peers) = i(ordinal)] to know who the DF for a specified service and ESI is. In this example, ESI-12 is configured with mode **auto**; therefore, for VPLS-1 (with EVI-1), PE-3 will be elected as DF because evi(1) mod (2)peers = 1, and the ordinal 1 corresponds to the second lowest IP, PE-3. The algorithm takes the configured **evi** in the service; therefore, the **evi** is mandatory, and for the same service must match in all the PEs that are part of the ES. This guarantees that the election algorithm is consistent across all the PEs of the ESI.
- **service-carving mode manual** — The user can manually decide for which **evi** identifiers the PE is DF or **primary: service-carving mode manual / manual evi <start> [to <*to*>] primary**. The PE will be non-DF for the non-specified EVIs. If **service-carving mode manual** is configured, but no range is defined, all the services are considered to be non-DF. If a range is configured, but the **service-carving** is not **mode manual**, the range has no effect. Only two PEs are supported when **service-carving mode manual** is configured.
- **service-carving mode off** — The lowest originator IP will win the election for a specified service and ES.
- Because the **evi** is used for the service-carving algorithm, it must always be configured in a service with SAPs/SDP bindings created in an ES, irrespective of the service-carving mode (service-carving off, auto, or manual).

Although not configured as part of the ES, the **config>redundancy>bgp-evpn-multi-homing>boot-timer** allows the necessary time for the control plane protocols to come up after the PE has rebooted, and before bringing up the ESs and running the DF algorithm. Some considerations about the boot-timer:

- The **boot-timer** should use a value long enough to allow the IOMs and BGP sessions to come up before exchanging ES routes and run the DF election for each EVI (it is 10 s, by default).
- The **boot-timer** runs per EVI on the ESs in the system. While **system-up-time < boot-timer**, the system will not run the DF election for any EVI. When the boot-timer expires, the DF election for the EVI is run and, if the system is elected DF for the EVI, the **es-activation-timer** will start.
- The system will not advertise ES routes until the boot timer expires. This guarantees that the peer ES PEs do not run the DF election either, until the PE is ready to become the DF, if needed.
- The following show command displays the configured **boot-timer**, as well as the remaining timer if the system is still in boot-stage.

```
*A:PE-2# show redundancy bgp-evpn-multi-homing
===============================================================================
Redundancy BGP EVPN Multi-homing Information
===============================================================================
Boot-Timer              : 10 secs
Boot-Timer Remaining    : 0 secs
ES Activation Timer     : 3 secs
===============================================================================
*A:PE-2#
```

After ESI-12 is configured in PE-2 and PE-3, the lag-1 SAPs in both PEs can be added to the VPLS-1 service. Until the ESI-12 is successfully enabled, the LAG SAPs will be kept down with a StandByForMHProtocol flag. This is illustrated in the following example for PE-2.

```
configure service system bgp-evpn ethernet-segment "ESI-12" shutdown
configure service vpls 1 sap lag-1:1 create
exit

*A:PE-2# show service id 1 sap lag-1:1 detail | match "  Oper State"
Admin State        : Up                       Oper State        : Down

*A:PE-2# show service id 1 sap lag-1:1 detail | match Flag
Flags              : StandByForMHProtocol

*A:PE-2# configure service system bgp-evpn ethernet-segment "ESI-12" no shutdown

*A:PE-2# show log log-id 99

===============================================================================
Event Log 99
===============================================================================
Description : Default System Log
Memory Log contents  [size=500   next event=118  (not wrapped)]

117 2016/06/28 13:52:44.77 UTC MINOR: SVCMGR #2203 Base
"Status of SAP lag-1:1 in service 1 (customer 1) changed to admin=up oper=up flags="
```

All-Active Multi-Homing Operation

To confirm that all-active multi-homing is working correctly for ESI-12, the user can use the following commands:

- **show service system bgp-evpn** — Shows the RD is used for the ES route.
- **show service system bgp-evpn ethernet-segment** — Shows all the ESs configured in the PE and their admin/operational status.
- **show service system bgp-evpn ethernet-segment name ESI-12 evi 1** — Shows the DF candidate PEs for evi 1 and whether the system is DF for evi.

- **show service system bgp-evpn ethernet-segment name ESI-12 all** — Shows all the information related to a specific ESI.

The base BGP-EVPN information includes the RD:

```
*A:PE-2# show service system bgp-evpn

===============================================================================
System BGP EVPN Information
===============================================================================
Eth Seg Route Dist.               : <none>
Eth Seg Oper Route Dist.          : 192.0.2.2:0
Eth Seg Oper Route Dist Type      : default
Ad Per ES Route Target            : evi-rt
===============================================================================
*A:PE-2#
```

The following command shows the configured ESs in the PE and their status:

```
*A:PE-2# show service system bgp-evpn ethernet-segment

===============================================================================
Service Ethernet Segment
===============================================================================
Name                            ESI                             Admin     Oper
-------------------------------------------------------------------------------
ESI-12                          01:00:00:00:00:12:00:00:00:01 Enabled   Up
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
*A:PE-2#
```

The following command shows that PE-2 is not the DF and the DF candidate PEs for EVI 1 are PE-2 and PE-3:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-12" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  06/28/2016 13:52:45
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.2                               06/28/2016 13:52:45
192.0.2.3                               06/28/2016 13:52:47
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
*A:PE-2#
```

The following command shows all information related to ESI-12 on PE-2:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-12" all
===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-12
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:12:00:00:00:01
Multi-homing            : allActive          Oper Multi-homing  : allActive
Source BMAC LSB         : <none>
Lag Id                  : 1
ES Activation Timer     : 3 secs
Exp/Imp Route-Target    : target:00:00:00:00:12:00

Svc Carving             : auto
ES SHG Label            : 262142
===============================================================================
===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   no
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The following command shows all information related to ESI-12 on PE-3:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-12" all
===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-12
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:12:00:00:00:01
Multi-homing            : allActive          Oper Multi-homing  : allActive
Source BMAC LSB         : <none>
Lag Id                  : 1
ES Activation Timer     : 3 secs
Exp/Imp Route-Target    : target:00:00:00:00:12:00

Svc Carving             : auto
ES SHG Label            : 262142
===============================================================================
```

```
===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
---snip---
```

The preceding commands show the ESI-12 configuration on both PEs and the result of the DF election for EVI 1. This output demonstrates two points:

- The ES **Exp/Imp Route-Target** is shown as **target:00:00:00:00:12:00**. This RT is auto-derived from the ESI bytes 2 to 7 (with the type byte being byte 1). Only PE-2 and PE-3 generate this RT and therefore import each other's ES route. This auto-derived ES RT is exported/imported due to auto-created router policies that are applied to the base system BGP-EVPN instance: **_ES_EvpnEthSegRtExp** and **_ES_EvpnEthSegRtImp**. These policies are created when the first ES no shutdown is successfully executed in the system.

  The following output shows the ES route received on PE-2:

```
*A:PE-2#
1 2016/06/28 13:42:54.81 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 70
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-Eth-Seg Len: 23 RD: 192.0.2.3:0
              ESI: 01:00:00:00:00:12:00:00:00:01, IP-Len: 4 Orig-IP-Addr: 192.0.2.3

    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:00:00:00:00:12:00
"


4 2016/06/28 13:42:54.81 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-12, EVI:1, Designated Forwarding state changed to:false"
```

The auto-created policies are as follows:

```
*A:PE-2# show router policy "_ES_EvpnEthSegRtExp"
    entry 1
        from
            family evpn
        exit
        action accept
        exit
    exit


*A:PE-2# show router policy "_ES_EvpnEthSegRtImp"
    entry 3
        from
            community "_ES_ESI-12"
            family evpn
        exit
        action accept
        exit
    exit


*A:PE-2# show router policy community "_ES_ESI-12"
community "_ES_ESI-12" members "target:00:00:00:00:12:00"
```

- The **show service system bgp-evpn ethernet-segment name ESI-12 all** command shows the ESI-label allocated to the PE: **ES SHG Label 262142** in the CLI output for PE-3. In this example, this label is allocated by PE-3 for ESI-12 (a different one is allocated per ESI) and advertised in the AD per-ES route for ESI-12. The following output shows the AD per-ES and AD per-EVI (for evi 1) routes sent by PE-3 and received by PE-2.
  - The AD per-ES route can be identified by the **MAX-ET** in the ethernet-tag field (as per RFC 7432) and carries the ESI-label as well as the multi-homing mode (all-active in this case) in the ESI-label extended community (see Figure 253).

    Prior to release 14.0.R1, a separate AD per-ES route is sent per EVI. In release 14.0.R1, or later, the user can enable the aggregation of AD per-ES routes by using the following command: **config>service>system>bgp-evpn>ad-per-es-route-target evi-rt-set route-distinguisher** *ip-address*. If enabled, a single AD per-ES route with the associated RD and a set of EVI route-targets will be advertised (to a maximum of 128). When there are more than 128 EVIs defined in the ethernet-segment, more than one route will be sent by the system.
  - The AD per-EVI route will have an eth-tag 0 and will carry the service label in the NLRI.

```
*A:PE-2#
2 2016/06/28 13:52:46.80 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 80
```

```
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:1 ESI: 01:00:00:00:00:12:00:00:00:01,
                      tag: MAX-ET Label: 0

    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        esi-label:262142/All-Active
"


3 2016/06/28 13:52:46.80 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 80
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:1 ESI: 01:00:00:00:00:12:00:00:00:01
                      tag: 0 Label: 4194256

    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
"


*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:12:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1         01:00:00:00:00:12:00:00:00:01 192.0.2.3
      0                                                 LABEL 262141

u*>i  192.0.2.3:1         01:00:00:00:00:12:00:00:00:01 192.0.2.3
      MAX-ET                                            LABEL 0
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:12:00:00:00:01
hunt
---snip---
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : N/A
Nexthop        : 192.0.2.3
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
---snip---
Community      : target:64500:1 bgp-tunnel-encap:MPLS
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:12:00:00:00:01
Tag            : 0
Route Dist.    : 192.0.2.3:1
MPLS Label     : LABEL 262141

---snip---

Network        : N/A
Nexthop        : 192.0.2.3
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
---snip---
Community      : target:64500:1 esi-label:262142/All-Active
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:12:00:00:00:01
Tag            : MAX-ET
Route Dist.    : 192.0.2.3:1
MPLS Label     : LABEL 0
---snip---
```

From a service perspective, as soon as CE-11 sends some traffic, the PE learning the CE-11 MAC address will advertise it to the network. The remote PEs (PE-4 and PE-5) will create a new EVPN-MPLS ES destination to ESI-12, with two next-hops: PE-2 and PE-3. The following outputs show the following information:

- PE-4 has learned AD per-EVI/ES routes for ESI-12 from PE-2 and PE-3, as well as the CE-11 MAC address from PE-3 (because MTU-1 picked up its link to PE-3 to send CE-11 frames).

```
*A:PE-4# show router bgp routes evpn auto-disc esi 01:00:00:00:00:12:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
```

```
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                             NextHop
      Tag                                                 Label
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         01:00:00:00:00:12:00:00:00:01 192.0.2.2
      0                                                   LABEL 262141

u*>i  192.0.2.2:1         01:00:00:00:00:12:00:00:00:01 192.0.2.2
      MAX-ET                                              LABEL 0

u*>i  192.0.2.3:1         01:00:00:00:00:12:00:00:00:01 192.0.2.3
      0                                                   LABEL 262141

u*>i  192.0.2.3:1         01:00:00:00:00:12:00:00:00:01 192.0.2.3
      MAX-ET                                              LABEL 0
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

PE-4 has learned MAC address 00:00:11:11:11:11 of CE-11 in ESI-12. The BGP EVPN MAC route has PE-3 as next hop:

```
*A:PE-4# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Ip Address
                                            NextHop
                                            Label1
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1         00:00:11:11:11:11 01:00:00:00:00:12:00:00:00:01
      0                   Seq:0             N/A
                                            192.0.2.3
                                            LABEL 262141
---snip---
```

- In the FDB for VPLS-1, PE-4 has learned the CE-11 MAC address associated with a newly created EVPN-MPLS ES destination:

```
*A:PE-4# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:11:11:11:11 eES:                     Evpn     06/29/16 08:57:00
                            01:00:00:00:00:12:00:00:00:01
---snip---
```

- Due to the aliasing function, the newly created EVPN-MPLS ES destination to ESI-12 has two next-hops (PE-2 and PE-3), to which PE-4 can load-balance the unicast traffic because **ecmp 2** is configured in the VPLS-1 of PE-4.

```
*A:PE-4# show service id 1 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.2       262140        0           Yes             06/28/2016 08:30:01
                ldp
192.0.2.3       262140        0           Yes             06/28/2016 08:30:29
                ldp
192.0.2.5       262140        0           Yes             06/28/2016 08:10:40
                ldp
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:12:00:00:00:01   1                       06/29/2016 08:57:00
---snip---
```

The **show service id 1 evpn-mpls esi 01:00:00:00:00:12:00:00:00:01** command shows the next-hops that the EVPN-MPLS ES destination is resolved to.

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:12:00:00:00:01
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:12:00:00:00:01   1                       06/29/2016 08:57:00
===============================================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.2               262141                  06/29/2016 16:47:00
                        ldp
192.0.2.3               262141                  06/29/2016 16:47:00
                        ldp
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

- PE-3 will show the CE-11 MAC address as learned locally in sap lag-1:1 (because the data plane learning of the CE-11 MAC address happened in PE-3). For PE-2, even though it learned the MAC address from EVPN, it will install it as associated with sap lag-1:1 because the EVPN route came with ESI-12, which is a local ESI. Because of this, whenever PE-2 receives a frame with MAC DA equal to the CE-11 MAC address, it will be able to forward the frame locally to the sap lag-1:1. The following output shows the CE-11 MAC address as it is installed in PE-2 and PE-3:

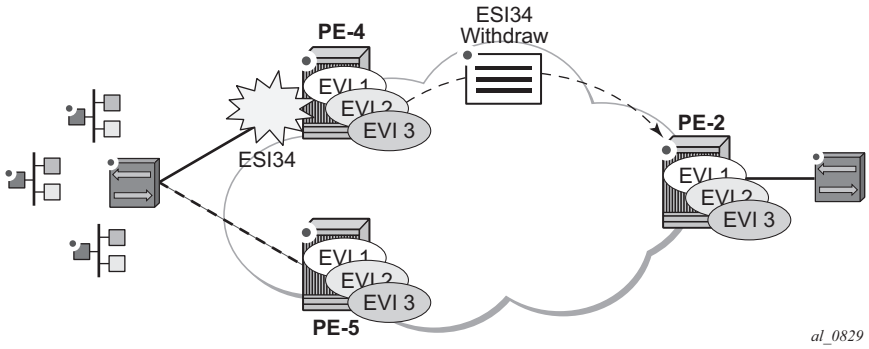
```
*A:PE-2# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:11:11:11:11 sap:lag-1:1              Evpn     06/29/16 08:27:59
---snip---

*A:PE-3# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:11:11:11:11 sap:lag-1:1              L/30     06/29/16 08:57:59
---snip---
```

Single-Active Multi-Homing Concepts

There are two new concepts in EVPN single-active multi-homing: mass-withdraw and backup path. They are illustrated in Figure 256.

Figure 256 **EVPN-MPLS Single-Active Multi-Homing: Mass-Withdraw, Backup Path**



- With mass-withdraw, when ESI-34 goes down, PE-2 does not have to wait for all the MAC routes to be withdrawn to converge all the services. Instead, PE-4 will withdraw the AD per-ES routes (also the AD per-EVI and MAC routes) and that will be used at PE-2 as a notification to stop sending traffic to PE-4 for any MAC address associated with ESI-34.
- With backup path, when PE-2 is notified of the ESI-34 failure due to the withdrawn AD routes, it will not flush any MAC address associated with ESI-34. Instead, it will change the next-hop of the EVPN-MPLS ES destination to the remaining PE in the ESI-34. Backup path only works when there are two PEs in the same ES. If there were more than two PEs in ESI-34, PE-2 would flush all the MAC addresses upon receiving a mass-withdraw notification, because it would not know who the new active PE is.

Single-Active Multi-Homing Configuration

The single-active multi-homing configuration example is based on Figure 254:

MTU-6 is connected to the EVPN network using single-active multi-homing. With the MTU-6 configuration, a VPLS service with active-standby spoke-sdp to PE-4 and PE-5 is configured. In PE-4 and PE-5, the SDP connected to MTU-6 is linked to ESI-34. Both will run the DF election algorithm for EVI 1, and the non-DF PE (PE-4 in this example) will bring down the spoke-sdp and notify MTU-6.

The following output shows the configuration of ESI-34 in PE-4 and PE-5, as well as the SDPs. The configuration of MTU-6 is also shown for completeness. It is important to keep the default **no ignore-standby-signaling** command on MTU-6 spoke-sdps because the PW switchover in MTU-6 will be triggered based on the PW status bits sent by PE-4 and PE-5.

SDP 46 with far-end MTU-6 is configured on PE-4:

```
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
```

Ethernet segment “ESI-34” is configured on PE-4 as follows:

```
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-34" create
                    esi 01:00:00:00:00:34:00:00:00:01
```

```
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
```

On PE-5, SDP 56 is configured as follows:

```
configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
```

Ethernet segment “ESI-34” is configured as follows on PE-5:

```
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-34" create
                    esi 01:00:00:00:00:34:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
```

On MTU-6, the service configuration is as follows:

```
configure
    service
        sdp 64 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        sdp 65 mpls create
            far-end 192.0.2.5
            ldp
            no shutdown
        exit
        vpls 1 customer 1 create
            endpoint "CORE" create
            exit
            sap 1/2/1:1 create
            exit
            spoke-sdp 64:1 endpoint "CORE" create
```

```
            exit
            spoke-sdp 65:1 endpoint "CORE" create
            exit
            no shutdown
```

For a detailed description of the base BGP-EVPN instance and **ethernet-segment** configuration, see the All-Active Multi-Homing Configuration section. The **es-activation-timer**, **esi**, **service-carving**, **boot-timer**, and **shutdown** commands are used in the same way as for all-active multi-homing. Only the differences compared to all-active multi-homing are described here:

- **multi-homing single-active** must be configured so that the ES acts as single-active. Optionally, the **no-esi-label** attribute can be added to the **multi-homing single-active** command. This attribute controls the use of the ESI-label for single-active multi-homing. Although the ESI-label is always used in all-active multi-homing when sending BUM traffic between the PEs in the ES, it is configurable for single-active. However, Nokia recommends to use the default option (using ESI-label) to avoid potential transient issues when there is a DF switchover.
- **sdp** <*sdp-id*> is configured so that the ES can be associated with the SDP connected to MTU-6. Although all-active multi-homing only allows lag associations to the ES, single-active allows LAG, port, and SDP. In this example, SDP is the option, because the access network is MPLS-based.

Similar to the all-active multi-homing case, when configuring the service in PE-4 and PE-5, the service objects are automatically associated with the ESI-34, because they are defined in the SDPs linked to the ESI. The configuration for VPLS 1 on PE-5 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 56:1 create
                no shutdown
            exit
            no shutdown
```

In all-active multi-homing, the non-DF does not bring down the service SAP associated with the ES (it only removes it from the default-multicast-list). However, in single-active multi-homing, the service spoke-SDP (or SAP, if that was the object associated) is brought operationally down. The following output shows the spoke-SDP state in PE-4 (non-DF), as operationally down with the **StandbyForMHProtocol** flag and the **Local Pw Bits** that are signaled to MTU-6:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-34" evi 1
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  06/29/2016 09:16:14
===============================================================================
===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               06/29/2016 09:16:40
192.0.2.5                               06/29/2016 09:16:40
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

Spoke-SDP 46:1 is operationally down on PE-4:

```
*A:PE-4# show service id 1 base

---snip---
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:46:1 S(192.0.2.6)                   Spok         0       8974    Up   Down
===============================================================================
```

Spoke-SDP 46:1 is operationally down with the StandbyForMHProtocol flag:

```
*A:PE-4# show service id 1 sdp 46:1 detail | match Flag
Flags              : StandbyForMHProtocol
```

The local PW bits (pwFwdingStandby) are sent to MTU-6:

```
*A:PE-4# show service id 1 sdp 46:1 detail | match Pw
Local Pw Bits      : pwFwdingStandby
Peer Pw Bits       : None
```

Single-Active Multi-Homing Operation

The same commands used in the All-Active Multi-Homing Operation section can be used for single-active; see that section.

The **show service system bgp-evpn ethernet-segment name ESI-34** command shows an ethernet-segment **Oper Multi-homing** in addition to the configured **Multi-homing** mode. This occurs because, in spite of configuring the ES as all-active, it may operate as single-active if there is a mismatch between the modes advertised by PE-4 and PE-5 in the AD per-ES routes (per RFC 7432). In this example, the configured and the operational value are the same:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-34"
===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-34
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:34:00:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
Source BMAC LSB         : <none>
Sdp Id                  : 46
ES Activation Timer     : 3 secs
Exp/Imp Route-Target    : target:00:00:00:00:34:00

Svc Carving             : auto
ES SHG Label            : 262142
===============================================================================
```

As soon as CE-16 sends some traffic, the DF PE (PE-5) will learn the CE-16 MAC address and will advertise it to the network. The remote PEs (PE-2 and PE-3) will create a new EVPN-MPLS ES destination to ESI-34, but this time with only one next-hop, PE-5, because this is single-active multi-homing. The following outputs show the following information:

- PE-2 has learned AD per-EVI/ES routes for ESI-34 from PE-4 and PE-5, as well as the CE-16 MAC address from an ES EVPN-MPLS destination, which is resolved to PE-5 (the DF for ESI-34).

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:34:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.          ESI                            NextHop
      Tag                                                 Label
-------------------------------------------------------------------------------
```

```
u*>i  192.0.2.4:1         01:00:00:00:00:34:00:00:00:01 192.0.2.4
      0                                                 LABEL 262141

u*>i  192.0.2.4:1         01:00:00:00:00:34:00:00:00:01 192.0.2.4
      MAX-ET                                            LABEL 0

u*>i  192.0.2.5:1         01:00:00:00:00:34:00:00:00:01 192.0.2.5
      0                                                 LABEL 262141

u*>i  192.0.2.5:1         01:00:00:00:00:34:00:00:00:01 192.0.2.5
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

PE-2 has learned the CE-16 MAC address from an ES EVPN-MPLS destination:

```
*A:PE-2# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:11:11:11:11 sap:lag-1:1              Evpn     06/29/16 12:22:41
1         00:00:16:16:16:16 eES:                     Evpn     06/29/16 12:22:41
                            01:00:00:00:00:34:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

The ES EVPN-MPLS destination is resolved to DF PE-5:

```
*A:PE-2# show service id 1 evpn-mpls esi  01:00:00:00:00:34:00:00:00:01
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:34:00:00:00:01   1                       06/29/2016 12:22:41
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.5               262141                  06/29/2016 12:22:41
                        ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

- In this case, the local PEs, PE-4 and PE-5, will learn the CE MAC address from an EVPN-MPLS destination and a local spoke-SDP, respectively.

```
*A:PE-4# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:16:16:16:16 eES:                     Evpn     06/29/16 12:30:14
                            01:00:00:00:00:34:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

The ES EVPN-MPLS destination is resolved to DF PE-5:

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:34:00:00:00:01
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:34:00:00:00:01   1                       06/29/2016 12:30:14
===============================================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.5               262141                  06/29/2016 12:30:14
                        ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

DF PE-5 learns the CE-16 MAC address from a local spoke SDP:

```
*A:PE-5# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:16:16:16:16 sdp:56:1                 L/60     06/29/16 12:37:26
-------------------------------------------------------------------------------
No. of MAC Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

Ethernet-Segment Failures

If either ES fails, a DF re-election will happen and the corresponding AD per-ES/EVI routes will be withdrawn, causing the remote PEs to modify the list of next-hops for the EVPN-MPLS ES destination. The following example illustrates a failure on the SDP between MTU-6 and PE-5 (the DF).

Step 1. A failure occurs in the LSP between MTU-6 and PE-5. This can be any event that brings the SDP down.

```
*A:PE-5#
7 2016/06/29 13:24:54.60 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 56 changed to admin=up oper=down"
```

Step 2. Immediately, PE-5 gives up the DF role and withdraws the ES route, as well as the AD routes and MAC routes. As soon as PE-4 receives any ES or AD withdraw, it will re-run the DF algorithm and, when the es-activation-timer expires, it will become the DF and activate its spoke-SDP.

```
*A:PE-5#
8 2016/06/29 13:24:54.60 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-34, EVI:1, Designated Forwarding state changed to:false"
```

The ES in PE-5 is operational down:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-34"
===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-34
Admin State             : Enabled            Oper State         : Down
ESI                     : 01:00:00:00:00:34:00:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
Source BMAC LSB         : <none>
Sdp Id                  : 56
ES Activation Timer     : 3 secs
Exp/Imp Route-Target    : target:00:00:00:00:34:00

Svc Carving             : auto
ES SHG Label            : 262142
===============================================================================
```

PE-5 is no longer the DF and the only DF candidate is PE-4:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-34" evi 1
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  06/29/2016 13:24:55
===============================================================================
===============================================================================
DF Candidates                          Time Added
-------------------------------------------------------------------------------
```

```
192.0.2.4                                06/29/2016 08:16:40
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

PE-4 becomes the DF and the spoke-SDP 46:1 is brought up.

```
*A:PE-4#
7 2016/06/29 13:24:57.57 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-34, EVI:1, Designated Forwarding state changed to:true"

8 2016/06/29 13:24:57.57 UTC MINOR: SVCMGR #2326 Base
"Status of SDP Bind 46:1 in service 1 (customer 1) local PW status bits changed
to none"

9 2016/06/29 13:24:57.57 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 46:1 in service 1 (customer 1) changed to admin=up oper=up flags="
```

The ES is up in PE-4:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-34"
===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-34
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:34:00:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
Source BMAC LSB         : <none>
Sdp Id                  : 46
ES Activation Timer     : 3 secs
Exp/Imp Route-Target    : target:00:00:00:00:34:00

Svc Carving             : auto
ES SHG Label            : 262142
===============================================================================
```

PE-4 is the DF and there are no other DF candidates:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-34" evi 1
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   yes 06/29/2016 13:24:58
===============================================================================
===============================================================================
DF Candidates                            Time Added
-------------------------------------------------------------------------------
192.0.2.4                                06/29/2016 08:16:40
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

Step 3. The remote PEs, PE-2 and PE-3, receive the AD routes withdrawal and modify the next-hop for the EVPN-MPLS ES destination.

```
52 2016/06/29 13:17:21.60 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 15 Len: 82 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:34:00:00:00:01,
                      tag: 0 Label: 0

        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:34:00:00:00:01,
                      tag: MAX-ET Label: 0
```

The ES EVPN-MPLS destination is resolved to the DF PE-4:

```
*A:PE-2# show service id 1 evpn-mpls esi 01:00:00:00:00:34:00:00:00:01
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:34:00:00:00:01   1                       06/29/2016 13:17:22
===============================================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.4               262141                  06/29/2016 13:17:22
                        ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following must be considered:

- The DF election procedure is revertive, that is, when the failed SDP comes back up, PE-5 will take over again as DF and the network will re-converge.
- The DF election is triggered by the following events:
  - **configure service system bgp-evpn ethernet-segment ESI-34 no shutdown** triggers the DF election for all the services in the ES.
  - A new update/withdrawal of an ES route (containing an ESI configured locally) triggers the DF election for all the services in the ESI.
  - A new update/withdrawal of an AD per-ES route (containing an ESI configured locally) triggers the DF election for all the services associated with the list of RTs received along with the route.
  - A new update of an AD per-ES route with a change in the ESI-label extended community (single-active bit or MPLS label) triggers the DF election for all the services associated with the list of RTs received along with the route.

– A new update/withdrawal of an AD route per-EVI (containing an ESI configured locally) triggers the DF election for that service.

BGP-EVPN Route Selection in EVPN Networks

The selection of the best route for a MAC address is as follows:

- If a PE receives more than one route for the same MAC address, the best MAC route is chosen:
  - If the route key is equal in two or more routes (that is, the mac, mac-length, ip, ip-length, RD, eth-tag), then regular BGP selection applies:
    - If local-pref, AS-path, origin, and MED are equal, the lowest IGP distance to the BGP next-hop is chosen (unless **ignore-nh-metric** is configured). If the BGP next-hop is resolved by an LSP, the cost from the tunnel-table is used.
    - As a last resort tie-breaker, the route with the lowest originator ID, or received from the peer with the lowest BGP Identifier, is chosen (unless **ignore-router-id** is configured and the routes being compared are EBGP routes).
  - If the mac-length, mac, ip-length, ip, eth-tag are equal, and the RD is different, the EVPN selection process is applied in the following order:
    - Conditional static macs (local protected macs)
    - EVPN static macs (remote protected macs)
    - Data plane learned MACs (regular learning on saps/sdp-bindings)
    - EVPN macs with higher SEQ number
    - Lowest IP (next-hop IP of the EVPN NLRI)
    - Lowest eth-tag (will be normally zero)
    - Lowest RD
- After a MAC route is selected, the system checks for an associated ES.
  - If it has an ES, the system uses the MAC address as the EVPN-MPLS ES destination. The ES destination is constructed based on the AD per-EVI routes received for that ES (regardless of MAC address priorities with the ES).
  - The system selects the first ECMP number of AD per-EVI routes arranged by the IP address of PEs (lower IPs are selected first).
  - If the same PE has advertised multiple RDs, the system selects the route with the lowest RD for that PE.

In the example of Figure 254, PE-4 resolves the next-hops for ESI-12 as described in the second choice above, that is, because **ecmp=2**, the two available next-hops are chosen. If **ecmp** is changed to 1, PE-4 will pick up the lower IP (in the BGP next-hop). This is illustrated in the following output:

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:12:00:00:00:01
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:12:00:00:00:01   1                       06/29/2016 13:15:32
===============================================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.2               262141                  06/29/2016 13:13:16
                        ldp
192.0.2.3               262141                  06/29/2016 13:15:32
                        ldp
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

When ECMP equals 1, only the BGP next hop with the lower IP is chosen:

```
*A:PE-4# configure service vpls 1 bgp-evpn mpls ecmp 1
```

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:12:00:00:00:01
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:12:00:00:00:01   1                       06/29/2016 14:00:24
===============================================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.2               262141                  06/29/2016 13:13:16
                        ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

Comparing EVPN Multi-homing and BGP Multi-homing

EVPN-MPLS services support EVPN-MH (EVPN multi-homing) and also BGP-MH as in chapter BGP Multi-Homing for VPLS Networks. While EVPN-MH is the standard way of providing access resiliency in RFC 7432, BGP-MH is also a standard mechanism supported in VPLS or EVPN networks. The following table provides some comparison between both technologies.

Table 20 **Comparing EVPN Multi-homing and BGP Multi-homing**

VPN Requirements	EVPN-MH	BGP-MH	Comments
All-active MH (flow-based load-balancing)	Yes	No	EVPN-MH provides better bandwidth utilization
Single-active MH (service-based load-balancing)	Yes	Yes	
DF PE election - automatic service balancing	Yes Service-carving	No Requires vsi policies and LP manipulation	EVPN-MH provides better automation
DF PE election – manual configuration per service	Yes	No	EVPN-MH allows for manual DF config for EVIs and ISIDs (2 PEs)
Split-horizon indication in the data plane	Yes ESI-label	No	Prevents transient loops when dual-active DFs show up
DF indication in the control plane	No	Yes	BGP MH guarantees one DF at a time. EVPN relies on Timers to ensure one DF at a time
Allows multiple SAPs or SDP-bindings per service on the same site	No	Yes Through the use of SHGs	
Boot timer and site(es)-activation-timers	Yes	Yes	BGP-MH supports more granular configuration (service level)
Support for oper-groups	No	Yes	
Non-DF notification to the CE (MPLS and CFM)	Yes	Yes	Avoids blackholing

In addition to the preceding comparison, the following configuration excerpt compares EVPN-MH with BGP-MH on a bgp-evpn VPLS service and shows that, while EVPN-MH does not have any configuration at service level, BGP-MH is configured within the VPLS context, which gives a more granular control over the redundancy provided. Refer to the BGP Multi-Homing for VPLS Networks chapter for more information about BGP-MH.

```
config>service>system>bgp-evpn# info
----------------------------------------
  ethernet-segment "ESI-34" create
    esi 01:00:00:00:00:34:00:00:00:01
    es-activation-timer 3
    service-carving
      mode auto
      exit
    multi-homing single-active
    sdp 46
    no shutdown
    exit

config>service>vpls# info
----------------------------------------
  bgp
  exit
  bgp-evpn
    evi 1
    vxlan
      shutdown
      exit
    mpls
      ingress-replication-bum-label
      ecmp 2
      auto-bind-tunnel
        resolution any
        exit
      no shutdown
      exit
    exit
  spoke-sdp 46:1 create
    no shutdown
    exit
  no shutdown
----------------------------------------


config>service>vpls# info
----------------------------------------
  bgp
    route-distinguisher 192.0.2.4:1
  exit
  bgp-evpn
    evi 1
    vxlan
      shutdown
      exit
    mpls
      ingress-replication-bum-label
      ecmp 2
      auto-bind-tunnel
        resolution any
        exit
      no shutdown
      exit
    exit
  site "site-1" create
    no shutdown
```

```
    site-id 1
    spoke-sdp 46:1
    site-activation-timer 3
    exit
  spoke-sdp 46:1 create
    no shutdown
    exit
  no shutdown
-------------------------------------
```

Proxy-ARP/ND Configuration for EVPN-MPLS Networks

Although not strictly a BGP-EVPN configuration, **vpls>proxy-arp** and **vpls>proxy-nd** functions are typically enabled along with EVPN-MPLS in order to reduce the amount of flooding in the network. The proxy-ARP/ND agent in the VPLS service will snoop ARP-requests and/or Neighbor Solicitation messages and will reply to those messages locally (if the information is known) without having to flood the requests to the network.

The configuration options for proxy-ARP are the following:

```
*A:PE-2# configure service vpls 1 proxy-arp ?
  - no proxy-arp
  - proxy-arp

 [no] age-time        - Configure aging timer for proxy ARP entries
      dup-detect      - Configure anti-spoofing MAC address information
 [no] dynamic-arp-po* - Configure population of dynamic proxy ARP entries
 [no] garp-flood-evpn - Configure to flood GARP request/replys into EVPN
 [no] send-refresh    - Configure send refresh time
 [no] shutdown        - Administratively enable/disable proxy ARP configuration
 [no] static          - Configure static IP address to MAC address associations
      table-size     - Configure the maximum number of entries in the proxy ARP table
 [no] unknown-arp-re* - Configure to flood unknown ARP request
```

The configuration options for proxy-ND are the following:

```
*A:PE-2# configure service vpls 1 proxy-nd  ?
  - no proxy-nd
  - proxy-nd

 [no] age-time        - Configure aging timer for proxy ND entries
      dup-detect      - Configure anti-spoofing MAC address information
 [no] dynamic-nd-pop* - Configure population of dynamic proxy ND entries
      evpn-nd-advert* - Configure EVPN Neighbor Discovery advertisements
 [no] host-unsolicit* - Configure whether to flood evpn with host neighbor
                        advertisement
 [no] router-unsolic* - Configure whether to flood evpn with router neighbor
                        advertisement
 [no] send-refresh    - Configure send refresh time
 [no] shutdown        - Administratively enable/disable proxy ND configuration
 [no] static          - Configure static IP address to MAC address associations
```

```
     table-size      - Configure the maximum number of entries in the proxy ND table
[no] unknown-ns-flo* - Configure to flood unknown ND solicitation
```

When proxy-ARP/ND is enabled, the following configuration guidelines must be followed:

- **dynamic-arp-populate** or **dynamic-nd-populate** should be used only in networks with a consistent configuration of this command in all PEs.
- When using **dynamic-arp-populate**/**dynamic-nd-populate**, the **age-time** value should be configured to a value equal to three times the **send-refresh** value. This will help reduce the EVPN withdrawals and re-advertisements in the network.
- With large **age-time** values, it would be sufficient to configure the **send-refresh** value to half of the **proxy-ARP**/**ND age-time** or **FDB age-time**.
- In scaled environments (with thousands of services), it is not recommended to set the send-refresh value to less than 300 s. In such scenarios, Nokia recommends using a minimum proxy-ARP/ND **age-time** and FDB **age** of 900 s.
- The use of the following commands reduces or suppresses the ARP/ND flooding in an EVPN network, because EVPN MAC routes replace the function of the regular data plane ARP/ND messages:
  - **no garp-flood-evpn**
  - **no unknown-arp-request-flood-evpn**
  - **no unknown-ns-flood-evpn**
  - **no host-unsolicited-na-flood-evpn**
  - **no router-unsolicited-na-flood-evpn**
- Nokia recommends using the preceding commands only in EVPN networks where the CEs are routers directly connected to an SR OS node acting as the PE. Networks using aggregation switches between the host/routers and the PEs should flood GARP/ND messages in EVPN to make sure the remote caches are updated and BGP does not miss the advertisement of these entries.
- When the **anti-spoof-mac** is used with proxy-ARP/ND, ingress filters (in the access SAPs/SDP-bindings) should be configured to drop all traffic with destination anti-spoof-mac. The same MAC address should be configured in all PEs where dup-detect is active.
- When proxy-ND is used, the configuration of the following commands should be consistent in all the PEs in the network:
  - **router-unsolicited-na-flood-evpn**
  - **host-unsolicited-na-flood-evpn**
  - **evpn-nd-advertise**

- Because EVPN does not propagate the **router** flag in IPv6--> MAC address advertisements, in a mixed network with hosts and routers where **evpn-nd-advertise router** is configured, unsolicited host NA messages should be flooded so that the entire network gets to learn all of the host and router ND entries. In the same way, **evpn-nd-advertise host** should be configured so that unsolicited router NA messages are flooded.

Finally, along with proxy-ARP/ND, **vpls>discard-unknown** may be used in some EVPN-MPLS deployments where all the CEs are routers and they announce themselves to the network by sending GARPs or NAs (Neighbor Solicitation messages). According to RFC 7432, whether or not to flood packets to unknown destination MAC addresses should be an administrative choice, depending on how learning happens between CEs and PEs. **Discard-unknown** provides that administrative choice in case all the MAC addresses in an EVI can be learned even before any traffic is exchanged.

Proxy-ARP/ND along with **discard-unknown** helps reduce the BUM traffic in an EVPN network significantly; however, their use must be analyzed and considered, depending on the type of CEs in the EVI.

An example of proxy-ARP configuration is as follows. This configuration should be added to all PEs. When a new ARP message is received on any of the PEs, they will learn the IP-MAC address pair and will advertise it to the network.

```
configure
    service
        vpls 1
            proxy-arp
                age-time 900
                send-refresh 300
                dynamic-arp-populate
                no shutdown
            exit
```

Enabling proxy-ARP increases the number of MAC/IP routes being sent by the PEs. This is due to the following reasons:

- An additional MAC/IP route will be advertised per new learned IP-MAC address pair, regardless of having advertised the same MAC address already.
- A MAC per VPLS service will be advertised with a system MAC address. That MAC address will be used as MAC SA for proxy-ARP confirm messages when an IP moves to a different PE.

The following output shows the MAC/IP routes on PE-2 when proxy-ARP is enabled in the network.

```
*A:PE-2# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.          MacAddr            ESI
      Tag                  Mac Mobility       Ip Address
                                              NextHop
                                              Label1
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1          00:00:11:11:11:11  01:00:00:00:00:12:00:00:00:01
      0                    Seq:0              10.0.0.1
                                              192.0.2.3
                                              LABEL 262141

u*>i  192.0.2.3:1          00:00:11:11:11:11  01:00:00:00:00:12:00:00:00:01
      0                    Seq:0              N/A
                                              192.0.2.3
                                              LABEL 262141

u*>i  192.0.2.3:1          d8:48:ff:00:03:3a  ESI-0
      0                    Static             N/A
                                              192.0.2.3
                                              LABEL 262141

u*>i  192.0.2.4:1          d8:46:ff:00:03:3a  ESI-0
      0                    Static             N/A
                                              192.0.2.4
                                              LABEL 262141

u*>i  192.0.2.5:1          00:00:16:16:16:16  01:00:00:00:00:34:00:00:00:01
      0                    Seq:0              10.0.0.6
                                              192.0.2.5
                                              LABEL 262141

u*>i  192.0.2.5:1          00:00:16:16:16:16  01:00:00:00:00:34:00:00:00:01
      0                    Seq:0              N/A
                                              192.0.2.5
                                              LABEL 262141

u*>i  192.0.2.5:1          d8:49:ff:00:03:3a  ESI-0
      0                    Static             N/A
                                              192.0.2.5
                                              LABEL 262141

-------------------------------------------------------------------------------
Routes : 7
===============================================================================
```

Troubleshooting and Debug Commands

When troubleshooting an EVPN-MPLS network, the following show commands and debug commands are recommended, as already discussed throughout this chapter:

- **show redundancy bgp-evpn-multi-homing**
- **show router bgp routes evpn (and filters)**
- **show service evpn-mpls [<TEP ip-address>]**
- **show service id bgp-evpn**
- **show service id evpn-mpls (and modifiers)**
- **show service id fdb (and modifiers)**
- **show service system bgp-evpn**
- **show service system bgp-evpn ethernet-segment (and modifiers)**
- **debug router bgp update**
- **log-id 99**

In addition to the preceding commands, the following tools dump commands may also help:

- **tools dump service evpn usage** — This command shows the amount of EVPN-MPLS (and EVPN-VXLAN) destinations consumed in the system.
- **tools dump service system bgp-evpn ethernet-segment** <*name*> **evi <[1..65535]> df** — This command computes the DF election for a specific ESI and EVI. Note: The **show service system bgp-evpn ethernet-segment** commands shows whether the local PE is DF or non-DF for a specific EVI, but it does not show who the DF is if it is not the local PE. In case of more than 2 PEs in the ES, this command may be especially useful.

Some examples are provided below for PE-2. PE-2 is showing seven EVPN-MPLS destinations due to the following:

- Each remote PE consumes one EVPN-MPLS destination for unicast (if they advertise MAC/IP routes to PE-2 and the ingress-replication-bum-label is configured in all the PEs). PE-2 has three remote unicast EVPN-MPLS destinations.
- Each remote PE consumes one EVPN-MPLS destination for multicast (if they advertise inclusive multicast routes to PE-2). PE-2 has three remote multicast EVPN-MPLS destinations.
- Each remote ES consumes one EVPN-MPLS destination (it is only one per ES, regardless of the multi-homing mode and the number of PEs in the ES). PE-2 has one remote ES (ESI-34).

```
*A:PE-2# tools dump service evpn usage
```

```
EVPN usage statistics at 002 06:29:38.940:

MPLS-TEP                                        :             3
VXLAN-TEP                                       :             0
Total-TEP                                       :      3/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)   :             7
Vxlan Dests (TEP, Egress VNI)                   :             0
Total-Dest                                      :      7/196607

Sdp Bind +  Evpn Dests                          :      8/245759
ES L2/L3 PBR                                    :      0/ 32767
```

To compute the DF election for EVI 1:

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "ESI-12" evi 1 df

[06/29/2016 11:40:28] Computed DF: 192.0.2.3 (Remote) (Boot Timer Expired: Yes)
```

Conclusion

SR OS has a full RFC 7432 EVPN-MPLS implementation including single-active and all-active multi-homing. This example has shown how to configure and operate EVPN-MPLS for a simple non multi-homing configuration as well as a multi-homing configuration. Other topics, such as the integration of VPLS objects with EVPN-MPLS and proxy-ARP/ND, have also been discussed.

EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS)

This chapter provides information about EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4. Ethernet Virtual Private Network - Virtual Private Wire Service (EVPN-VPWS) is supported in SR OS release 14.0.R1, and later. EVPN-VPWS in multi-homing scenarios is supported in SR OS release 14.0.R4, and later.

Chapter EVPN for MPLS Tunnels is prerequisite reading.

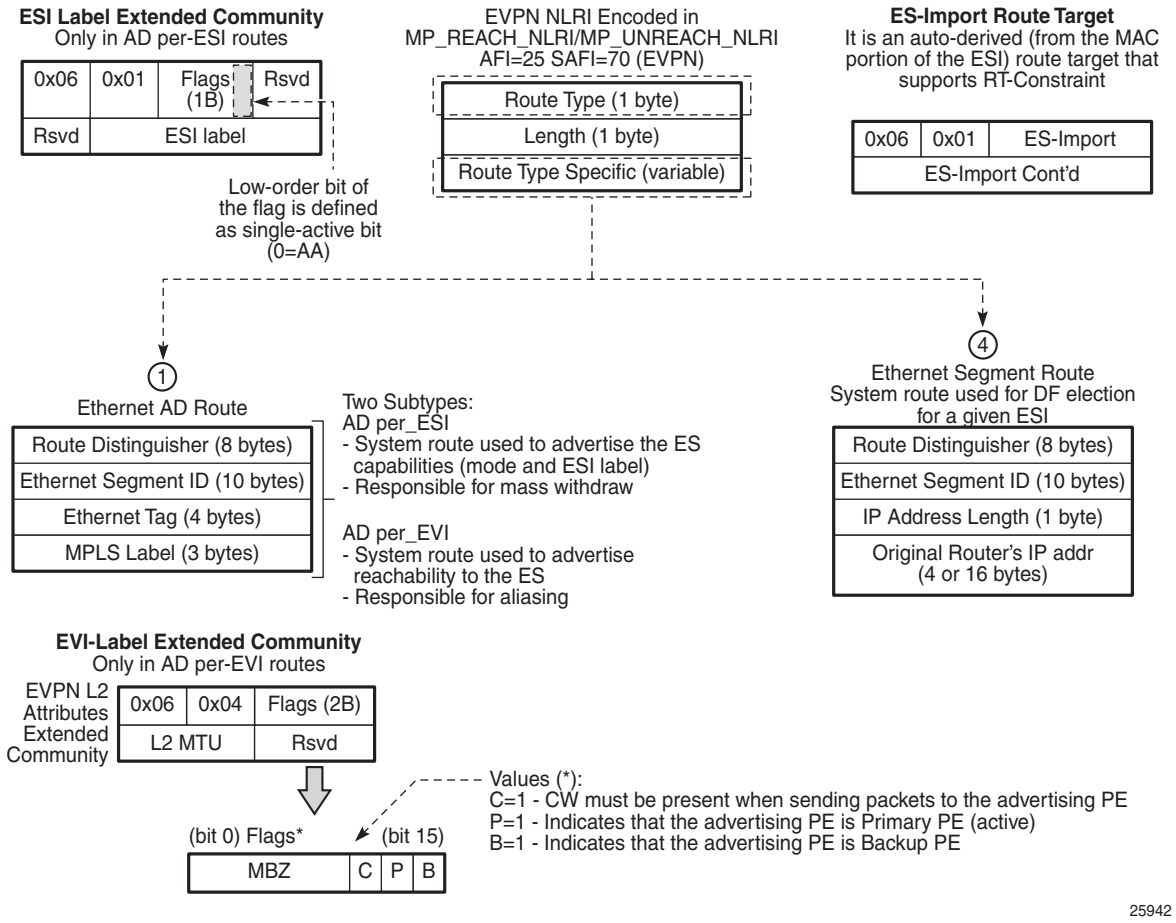
Overview

Service providers prefer an optimized, standardized, and unified control plane for VPNs. EVPN-VPWS is supported in MPLS networks that also run EVPN-MPLS in VPLS services. From a control plane perspective, EVPN-VPWS is a simplified point-to-point version of *RFC7432 - BGP MPLS-Based Ethernet VPN*, because there is no need to advertise MAC routes in VPWS. EVPN-VPWS is described in *draft-ietf-bess-evpn-vpws*.

EVPN-VPWS supports all-active multi-homing (per-flow load-balancing multi-homing) as well as single-active multi-homing (per-service load-balancing multi-homing), using the same Ethernet Segments (ESs) used for EVPN-MPLS VPLS services. EVPN-VPWS uses route-type 1 and route-type 4; it does not use route-types 2, 3, or 5, because Inclusive Multicast, MAC/IP routes, or IP-prefix routes are not required.

Figure 257 shows the encoding of the required extensions for the route-types 1 and 4 for EVPN-VPWS.

Figure 257 **Route Types and NLRIs for EVPN-VPWS**



Two sub-types are defined for route-type 1. Route-type 4 has no sub-types. The route types used for EVPN-VPWS have the following purposes:

- Route-type 1 - Auto-discovery per EVPN instance (AD per-EVI). This route type is used in all EVPN-VPWS scenarios, with or without multi-homing. For EVPN-VPWS, the Ethernet Tag field is encoded with the local attachment circuit (AC) of the advertising PE. This value is configured using the **service>bgp-evpn> local-ac-name>eth-tag** <*value*> command. The route distinguisher (RD),

MPLS label, and the Ethernet segment ID (ESI) are encoded as for EVPN-MPLS. The MPLS label field is used as service label. In case of multi-homing, AD per-EVI routes containing the same ESI are used to provide aliasing and a backup path to the PEs part of the ES. The L2 MTU is encoded with the service MTU configured in the Epipe. The flags used for EVPN-VPWS are:

- Flag C will be set if a control word is configured in the service.
- Flag P will be set if the advertising PE is primary PE.
  - If no multi-homing is used, there is no primary PE (P=0).
  - In all-active multi-homing, all PEs in the ES are primary (P=1).
  - In single-active multi-homing, only one PE per-EVI in the ES is primary (P=1).
- Flag B will be set if the advertising PE is backup PE.
  - The B-flag is only set in case of single-active multi-homing and only for one PE, even if more than two PEs are present in the same single-active ES. The backup PE is the winner of the second designated forwarder (DF) election (excluding the DF). The remaining non-DF PEs send B=0.

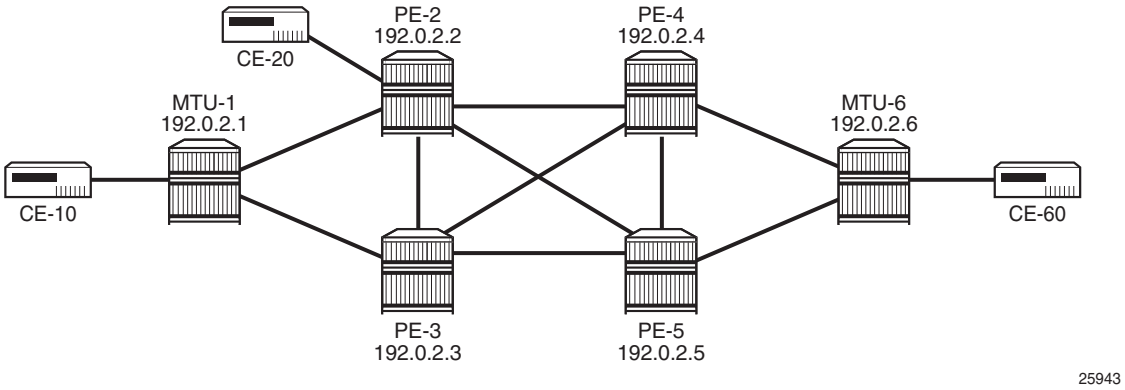
If there is no multi-homing, the ESI, flag P, and flag B will be zero.

- Route-type 1 - AD per Ethernet segment (AD per-ES). Same encoding as for EVPN-MPLS. AD per-ES is only used in multi-homing scenarios where it is advertised per ES from the PE. It carries the ESI label (used for split-horizon, but only for VPLS services and not for Epipe services) and can affect procedures such as the DF election, as well as the aliasing on remote PEs.
- Route-type 4 - ES route. Same encoding as for EVPN-MPLS. Route-type 4 is only used in multi-homing scenarios. This route advertises a local configured ES. The exchange of this route can discover remote PEs that are part of the same ES and the DF election algorithm among them.

Configuration

Figure 258 shows the example topology that will be used throughout this chapter.

Figure 258 **EVPN-VPWS Example Topology**



The example topology consists of six 7750 SR routers with the following initial configuration:

- Network (or hybrid) ports interconnect the core PEs with configured router interfaces.
- MTU-1 is a pure Ethernet aggregator. The ports toward the core PEs are access ports. Likewise, the ports on PE-2 and PE-3 toward MTU-1 are access ports.
- Core PEs and MTU-6 run IS-IS on all interfaces. Point-to-point adjacencies are established for the exchange of system IP addresses.
- Link LDP is configured between all PEs, and toward/from MTU-6.
- EVPN uses BGP for exchanging reachability at service level. Therefore, BGP peering sessions must be established among the core PEs for the EVPN family. Although typically a separate router is used, in this chapter, PE-2 is used as route reflector with the following BGP configuration:

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
```

```
            exit
        exit
    exit
```

The BGP configuration on the other PEs is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
        exit
```

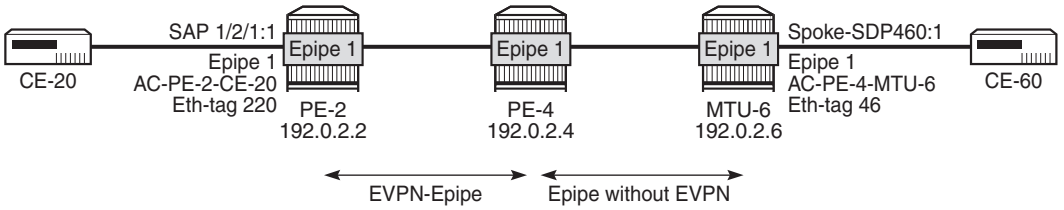
The following EVPN-VPWS scenarios are described in this section:

- EVPN for MPLS tunnels in Epipe services without multi-homing
- EVPN for MPLS tunnels in Epipe services with all-active multi-homing
- EVPN for MPLS tunnels in Epipe services with single-active multi-homing

EVPN for MPLS Tunnels in Epipe Services without Multi-Homing

BGP-EVPN can be enabled in Epipe services with either SAPs or spoke-SDPs at the access, as shown in Figure 259.

Figure 259 **Example Topology for EVPN-VPWS without Multi-Homing**



On PE-2, Epipe 1 is configured as follows:

```
configure
    service
        epipe 1 customer 1 create
            bgp
            exit
            bgp-evpn
                local-ac-name AC-PE-2-CE-20
                    eth-tag 220
                exit
                remote-ac-name AC-PE-4-MTU-6
                    eth-tag 46
                exit
                evi 1
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
```

On PE-4, the service configuration is as follows:

```
configure
    service
        sdp 460 create
            far-end 192.0.2.6
            no shutdown
        exit
        epipe 1 customer 1 create
            bgp
            exit
            bgp-evpn
                local-ac-name AC-PE-4-MTU-6
```

```
            eth-tag 46
        exit
        remote-ac-name AC-PE-2-CE-20
            eth-tag 220
        exit
        evi 1
        mpls
            auto-bind-tunnel
                resolution any
            exit
            no shutdown
        exit
    exit
    spoke-sdp 460:1 create
    exit
    no shutdown
```

Where the following commands are relevant for the EVPN-VPWS configuration:

- **bgp** enables the context for the BGP configuration relevant to the service. The bgp context configures the common BGP parameters for all BGP families in the service, such as route distinguisher and route target. Even if the general BGP parameters for the service are auto-derived, the bgp context must be enabled.

```
*A:PE-2# configure service epipe 1 bgp
  - bgp
  - no bgp
 [no] pw-template-bi* + Configure pw-template bind policy
 [no] route-distingu* - Configure route distinguisher
 [no] route-target    - Configure route target
```

Note: No pw-template binding is supported for EVPN. Although it is included in the configuration, pw-template binding cannot be configured if the service is EVPN.

- The following parameters can be configured in the bgp-evpn context:

```
*A:PE-2# configure service epipe 1 bgp-evpn
  - bgp-evpn
  - no bgp-evpn
 [no] evi             - EVPN Identifier
 [no] local-ac-name   + Configure local attachment-circuit name
      mpls            + Configure BGP EVPN mpls
 [no] remote-ac-name  + Configure remote attachment-circuit name
```

  - The **evi** is a 2-byte identifier used for auto-deriving the service RD, service RT, and for the **service-carving** when multi-homing is used. The auto-derivation of RD and RT is as follows:
    - RD system-ip:evi
    - RT autonomous-system:evi

The EVI values must be unique in the system, regardless of the type of service they are assigned to (Epipe or VPLS).

– The **local-ac-name** and remote-ac-name identify the two attachment circuits connected by the EVPN-VPWS service. The configured eth-tag for the local AC is advertised in the Ethernet Tag field of the AD per-EVI route for the Epipe, along with the corresponding RD, RT, and MPLS label. Both local and remote eth-tags are mandatory to bring up the Epipe service. If the received eth-tag for the Epipe service matches the configured remote AC eth-tag, it will create an EVPN-MPLS destination to the next hop.

The local eth-tag cannot be modified without shutting down bgp-evpn mpls in the Epipe, as shown in the following output:

```
*A:PE-2# configure service epipe 1 bgp-evpn local-ac-name AC-PE-2-CE-20 eth-tag 221
MINOR: SVCMGR #8025 Evpn mpls is enabled
```

Unlike local eth-tags, remote eth-tags can be modified without shutdown.

– The following configuration options are available for Epipes in the **bgp-evpn>mpls** context:

```
*A:PE-2# configure service epipe 1 bgp-evpn mpls
  - mpls
      auto-bind-tunn* + Configure BGP EVPN mpls auto-bind-tunnel
 [no] control-word    - Enable/disable setting the CW bit in the label message
      ecmp            - Configure maximum ECMP routes information
 [no] entropy-label   - Enable/disable use of entropy-label
 [no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
 [no] shutdown        - Administratively Enable/Disable BGP-EVPN mpls
```

This is a subset of the options for VPLS services; see chapter EVPN for MPLS Tunnels.

When the local AC (SAP 1/2/1:1) is up, PE-2 sends a BGP EVPN AD per-EVI route that contains the eth-tag for the local AC:

```
4 2016/07/08 05:48:54.89 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:1 ESI: ESI-0, tag: 220 Label: 4194208
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
        bgp-tunnel-encap:MPLS
"
```

The auto-derived RD is 192.0.2.1:1 and the RT is 64500:1.

When the remote AC on PE-4 (spoke-SDP 460:1) is up, PE-2 receives the following BGP update from PE-4:

```
5 2016/07/08 05:49:31.98 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:1 ESI: ESI-0, tag: 46 Label: 4194208
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
        bgp-tunnel-encap:MPLS
"
```

When the received RT matches and the received Ethernet Tag matches the configured remote AC, the EVPN-MPLS destination (comprised of a Termination Endpoint (TEP) and Egress Label) will be created on PE-2 and PE-4:

```
*A:PE-2# show service id 1 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address               Egr Label               Last Change
                           Transport
-------------------------------------------------------------------------------
192.0.2.4                 262138                  07/08/2016 06:32:10
                          ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The MPLS label in the debug message is not the same as in the service, because the router will strip the extra four lowest bits to get the 20-bit MPLS label. The egress label for the EVPN-MPLS destination on PE-4 is 262138. The 24-bit label value in the BGP update debug is 16 (2^4) times as high: 262138*16 = 4194208. This is because the debug message is shown before the router can parse the label field and see if it corresponds to an MPLS label (20 bits) or a VXLAN VNI (24 bits).

The BGP AD per-EVI routes for Ethernet Tag 46 can be shown with the following command:

```
*A:PE-2# show router bgp routes evpn auto-disc tag 46
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.          ESI                           NextHop
      Tag                                                Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:1          ESI-0                         192.0.2.4
      46                                                 LABEL 262138
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

BGP EVPN information for Epipe 1 can be shown with the following command:

```
*A:PE-2# show service id 1 bgp-evpn
===============================================================================
BGP EVPN Table
===============================================================================
EVI                : 1                  Creation Origin    : manual
===============================================================================
BGP EVPN MPLS Information
===============================================================================
Admin Status       : Enabled
Force Vlan Fwding  : Disabled           Control Word       : Disabled
Max Ecmp Routes    : 0
Local AC Name      : AC-PE-2-CE-20      Eth Tag            : 220
Remote AC Name     : AC-PE-4-MTU-6      Eth Tag            : 46
Entropy Label      : Disabled
===============================================================================
===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Resolution         : any
Filter Tunnel Types: (Not Specified)
===============================================================================
*A:PE-2#
```

Note: Each PE will send its service MTU into the L2 MTU field in the L2-attribute in the AD per-EVI route for the Epipe service. The received L2 MTU will be checked. In case of a mismatch between the received MTU and the configured service MTU, the router will not set up the EVPN destination and, therefore, the service will not come up.

EVPN for MPLS Tunnels in Epipe Services with Multi-Homing

SR OS supports EVPN multi-homing as per *draft-ietf-bess-evpn-vpws*.

The EVPN multi-homing implementation is based on the concept of the Ethernet Segment (ES). An ES is a logical structure that can be defined in one or more PEs and identifies the CE (or access network) multi-homed to the EVPN PEs. An ES is associated with a port, LAG, or SDP object, and is shared by all the services defined on those objects. It can also be shared between Epipe and VPLS services.

Each ES has a unique Ethernet Segment Identifier (ESI) that is 10 bytes and is manually configured in the current release. The ESI is advertised in the control plane to all the PEs in an EVPN network; therefore, it is very important to ensure that the 10-byte ESI value is unique throughout the entire network. Single-homed CEs are assumed to be connected to an ES with ESI = 0 (single-homed ESs are not explicitly configured).

The ES is part of the base BGP-EVPN configuration and is not applied to any EVPN-MPLS service, by default. An ES can be shared by multiple services; the association of a specific SAP or spoke-SDP to an ES is automatically made when the SAP is defined in the same LAG or port configured in the ES, or when the spoke-SDP is defined in the same SDP configured in the ES.

Regardless of the multi-homing mode, the local Ethernet Tag values must match on all the PEs that are part of the same ES. The PEs in the ES will use the AD per-EVI routes from the peer PEs to validate the PEs as DF election candidates for an EVI. The DF election is only relevant for single-active multi-homing ESs. For Epipes defined in an all-active multi-homing ES, there is no DF election required, because all PEs are forwarding traffic and all traffic is treated as unicast.

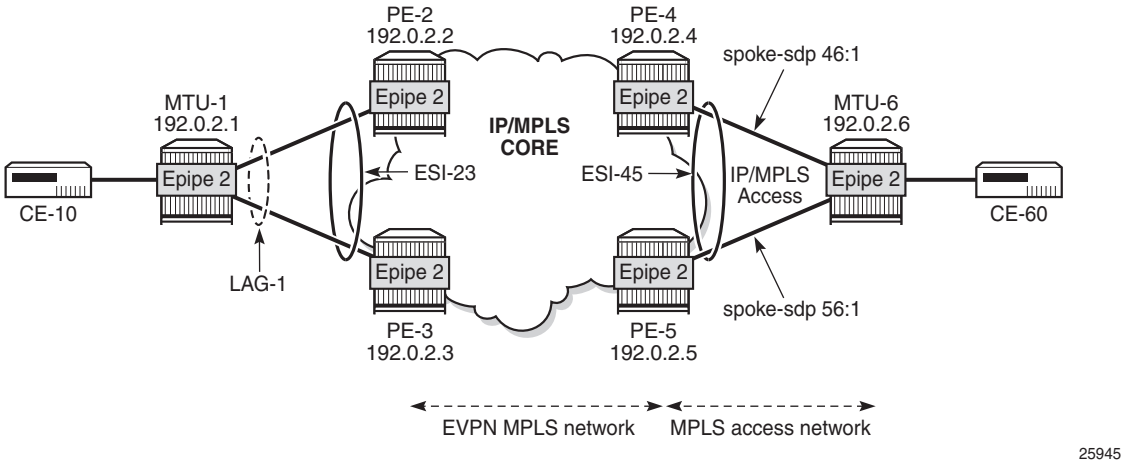
Aliasing is supported when sending traffic to an ES destination. Assuming ECMP is enabled on the ingress PE (and shared queuing or ingress policing), per-flow load-balancing will be performed among all the PEs that advertised P=1. PEs advertising P=0 are not considered as next hops for an ES destination.

The following sections show the configuration of:

- an all-active multi-homing ES with a LAG associated with it
- a single-active multi-homing ES linked to an SDP

Figure 260 shows the example topology has an all-active multi-homing ES "ESI-23" with a LAG associated to it in PE-2 and PE-3. A single-active multi-homing ES "ESI-45" with an SDP associated to it is configured in PE-4 and PE-5.

Figure 260 **Example Topology EVPN-VPWS with Multi-Homing**



EVPN for MPLS Tunnels in Epipe Services with All-Active Multi-Homing

All-active multi-homing allows for per-flow load-balancing. Unlike EVPN-MPLS in VPLS services, EVPN-VPWS has no DF election in all-active multi-homing. All PEs in the ES are active and the remote PE will do per-flow load-balancing. ESI-23 is configured on PE-2 and PE-3 in all-active multi-homing with LAG 1 associated to it. This LAG is used as a SAP in Epipe 2 on both PE-2 and PE-3. The configuration of the ES and Epipe 2 is identical on PE-2 and PE-3, including the local AC and remote AC names and eth-tags:

```
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
        exit
```

```
        epipe 2 customer 1 create
            bgp
            exit
            bgp-evpn
                local-ac-name AC-ESI-23-MTU-1
                    eth-tag 231
                exit
                remote-ac-name AC-ESI-45-MTU-6
                    eth-tag 456
                exit
                evi 2
                mpls
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap lag-1:1 create
            exit
            no shutdown
```

See chapter EVPN for MPLS Tunnels for a detailed explanation of the configuration parameters of the ES.

In EVPN-VPWS multi-homing scenarios, three route types are exchanged: AD per-EVI, AD per-ES, and ES routes. The following ES route (route-type 4) for ESI 01:00:00:00:00:23:00:00:00:01 sent by PE-2 is imported at PE-3:

```
77 2016/07/08 10:48:20.75 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 70
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-Eth-
Seg Len: 23 RD: 192.0.2.2:0 ESI: 01:00:00:00:00:23:00:00:00:01, IP-Len: 4 Orig-IP-
Addr: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:00:00:00:00:23:00
"
```

The target 00:00:00:00:23:00 in the extended community is derived from the ESI (bytes 2 to 7) and is only imported by the PEs that are part of the same ES; that is, PE-2 and PE-3 in this example.

At the same time, the following AD per-ES route (route-type 1) with Maximum Ethernet Tag (MAX-ET, all Fs) and label 0 is sent by PE-2 and imported by the rest of the PEs:

```
78 2016/07/08 10:48:20.75 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 94
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-
AD Len: 25 RD: 192.0.2.3:2 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET Label: 0
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:2
        esi-label:262138/All-Active
"
```

The ESI label 262138 is in the extended community, as well as the indication that the multi-homing is all-active. Epipe services do not require ESI labels because BUM traffic is not recognized as such in EVPN-VPWS services. However, because the ES can be shared by Epipe and VPLS services, the AD per-ES route still includes a non-zero ESI label.

The following AD per-EVI route (route-type 1) with Ethernet Tag 231 sent by PE-2 is received and imported at the rest of the PEs:

```
81 2016/07/08 10:48:20.75 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-
AD Len: 25 RD: 192.0.2.2:2 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 231 Label: 41941
76
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

This route contains the flags for control word, primary, and backup. In all-active multi-homing, all nodes are primary (P=1).

PE-4 has learned AD per-EVI/ES routes for ESI-23 from PE-2 and PE-3, as shown in the following output:

```
*A:PE-4# show router bgp routes evpn auto-disc esi 01:00:00:00:00:23:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.          ESI                            NextHop
      Tag                                                 Label
-------------------------------------------------------------------------------
u*>i  192.0.2.2:2          01:00:00:00:00:23:00:00:00:01 192.0.2.2
      231                                                 LABEL 262136
u*>i  192.0.2.2:2          01:00:00:00:00:23:00:00:00:01 192.0.2.2
      MAX-ET                                              LABEL 0
u*>i  192.0.2.3:2          01:00:00:00:00:23:00:00:00:01 192.0.2.3
      231                                                 LABEL 262137
u*>i  192.0.2.3:2          01:00:00:00:00:23:00:00:00:01 192.0.2.3
      MAX-ET                                              LABEL 0
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-4#
```

For Epipe 2 on PE-4, the EVPN MPLS destination is not pointing at a specific TEP, but ESI-23, as shown in the following output:

```
*A:PE-4# show service id 2 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address               Egr Label               Last Change
                           Transport
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                                 Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01             07/08/2016 10:48:21
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-4#
```

When ECMP > 1 on the ingress PE, multiple TEPs can correspond to a specific ESI (aliasing). In this case, ECMP=2 and PE-4 and PE-5 have two TEP addresses and Egress labels for ESI 01:00:00:00:00:23:00:00:00:01, as shown for PE-4:

```
*A:PE-4# show service id 2 evpn-mpls esi 01:00:00:00:00:23:00:00:00:01
=======================================================
BGP EVPN-MPLS Ethernet Segment Dest
=======================================================
Eth SegId                       Last Change
-------------------------------------------------------
01:00:00:00:00:23:00:00:00:01   07/08/2016 10:48:21
=======================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport
-------------------------------------------------------------------------------
192.0.2.2                262136                   07/08/2016 10:48:21
                         ldp
192.0.2.3                262137                   07/08/2016 10:48:21
                         ldp
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-4#
```

Note: Even if ECMP is configured, the ingress router will not load-balance the traffic unless shared queuing or ingress policing is configured. This is not specific to EVPN but generic to the way Epipes forward traffic.

In all-active multi-homing for EVPN-VPWS, there is no DF election and all PEs in the ES are active. For ESI-23, both PE-2 and PE-3 are active/primary/DF, but there are no DF candidates, because there is no DF election:

```
* A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" evi 2
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   yes 07/11/2016 06:32:49
===============================================================================
===============================================================================
DF Candidates                             Time Added
-------------------------------------------------------------------------------
No entries found
===============================================================================
*A:PE-2#
```

Similarly, on PE-3:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-23" evi 2
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   yes 07/11/2016 06:33:24
===============================================================================
===============================================================================
DF Candidates                             Time Added
-------------------------------------------------------------------------------
No entries found
===============================================================================
*A:PE-3#
```

To confirm that all-active multi-homing is working correctly, the following command shows all information related to a specific ESI; in this case, ESI-23 on PE-2:

```
* A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" all
===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-23
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 262137
Source BMAC LSB         : <none>
Lag Id                  : 1
ES Activation Timer     : 3 secs
Exp/Imp Route-Target    : target:00:00:00:00:23:00
Svc Carving             : auto
===============================================================================
===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
2                   2                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
---snip---
```

EVPN for MPLS Tunnels in Epipe Services with Single-Active Multi-Homing

Single-active multi-homing allows for per-service load-balancing. Single-active multi-homing is configured on PE-4 and PE-5 with ES "ESI-45". Both PEs have an SDP to MTU-6, which is associated with the ES and to the Epipe service. The configuration of the local and remote AC names and eth-tags is identical on PE-4 and PE-5.

On PE-4, the service configuration is as follows:

```
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
        epipe 2 customer 1 create
            bgp
            exit
            bgp-evpn
                local-ac-name "AC-ESI-45-MTU-6"
                    eth-tag 456
                exit
                remote-ac-name "AC-ESI-23-MTU-1"
                    eth-tag 231
                exit
                evi 2
                mpls
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 46:2 create
            exit
            no shutdown
```

On PE-5, the configuration is similar, but with a different SDP:

```
configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
```

```
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
            exit
        exit
        epipe 2 customer 1 create
            bgp
            exit
            bgp-evpn
                local-ac-name "AC-ESI-45-MTU-6"
                    eth-tag 456
                exit
                remote-ac-name "AC-ESI-23-MTU-1"
                    eth-tag 231
                exit
                evi 2
                mpls
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 56:2 create
            exit
            no shutdown
```

Three route types will be exchanged between the core PEs: AD per-EVI, AD per-ES, and ES routes.

PE-4 and PE-5 advertise ES routes that are only imported by them. As an example, the following is the ES route sent by PE-5. It contains a target 00:00:00:00:45:00 in the extended community that is derived from the ESI:

```
247 2016/07/08 10:49:17.85 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 84
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-Eth-
Seg Len: 23 RD: 192.0.2.5:0 ESI: 01:00:00:00:00:45:00:00:00:01, IP-Len: 4 Orig-IP-
Addr: 192.0.2.5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
```

```
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:00:00:00:00:45:00
"
```

The AD per-ES route has a Maximum Ethernet Tag (MAX-TAG) and an ESI label in the extended community. The multi-homing mode is single-active. As in the case of all-active multi-homing, the ESI label is not used in Epipe services.

```
250 2016/07/08 10:49:17.84 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 94
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-
AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: MAX-ET Label: 0
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:2
        esi-label:262136/Single-Active
"
```

The AD per-EVI route contains flags for primary and backup, which will be different for routes received from PE-4 and PE-5. In this case, PE-4 is primary in the single-active multi-homing ES (P=1):

```
256 2016/07/08 10:49:17.84 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 102
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-
AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456 Label: 41941
60
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
```

```
"
```

PE-5 is backup in the single-active multi-homing ES (B=1):

```
258 2016/07/08 10:49:20.78 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-
AD Len: 25 RD: 192.0.2.5:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456 Label: 41941
92
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 1
        bgp-tunnel-encap:MPLS
"
```

The BGP EVPN AD routes can be shown with the following command:

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:45:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.          ESI                             NextHop
      Tag                                                  Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:2          01:00:00:00:00:45:00:00:00:01 192.0.2.4
      456                                                  LABEL 262135
u*>i  192.0.2.4:2          01:00:00:00:00:45:00:00:00:01 192.0.2.4
      MAX-ET                                               LABEL 0
u*>i  192.0.2.5:2          01:00:00:00:00:45:00:00:00:01 192.0.2.5
      456                                                  LABEL 262137
u*>i  192.0.2.5:2          01:00:00:00:00:45:00:00:00:01 192.0.2.5
      MAX-ET                                               LABEL 0
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-2#
```

For each PE in the single-active ES, there are two AD routes: the routes with MAX-ET are AD per-ES routes and the routes with a configured Ethernet Tag are AD per-EVI routes.

The EVPN MPLS destination for Epipe 2 on PE-2 is ESI-45, as shown in the following output:

```
*A:PE-2# show service id 2 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address               Egr Label               Last Change
                           Transport
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                                 Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01             07/08/2016 10:48:09
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

The ESI is resolved to the TEP address of the primary (DF) PE-4, as follows:

```
*A:PE-2# show service id 2 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01
=========================================================
BGP EVPN-MPLS Ethernet Segment Dest
=========================================================
Eth SegId                       Last Change
---------------------------------------------------------
01:00:00:00:00:45:00:00:00:01   07/08/2016 10:48:09
=========================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label               Last Change
                         Transport
-------------------------------------------------------------------------------
192.0.2.4                262135                  07/08/2016 10:48:09
                         ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

The DF election is key for the forwarding and backup functions in single-active multi-homing ESs. The PE elected as DF will be the primary for the ES in the Epipe and will unblock the SAP/spoke-SDP for upstream and downstream traffic. The rest of the PEs in the ES will bring their ES SAPs or spoke-SDPs operationally down.

PE-5 is a non-DF, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45" evi 2
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   no  07/11/2016 06:34:11
===============================================================================
===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               07/11/2016 06:42:10
192.0.2.5                               07/11/2016 06:34:24
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
*A:PE-5#
```

In single-active multi-homing, the service spoke-SDP (or SAP) is brought operationally down on the non-DF, as shown in the following output:

```
*A:PE-5# show service id 2 base
---snip---
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:56:2 S(192.0.2.6)                   Spok         0       8978    Up   Down
===============================================================================
*A:PE-5#
```

The spoke-SDP 56:2 is operationally down with a StandbyForMHProtocol flag:

```
*A:PE-5# show service id 2 sdp 56:2 detail | match Flag
Flags              : StandbyForMHProtocol
```

Two consecutive DF elections take place: the first DF election includes all PEs in the ES for that Epipe and determines which PE is the primary PE (flags P=1, B=0). The second DF election excludes this DF and determines which PE is the backup (P=0, B=1). All other PEs signal flags P=0 and B=0.

When the primary PE fails, AD per-ES/EVI withdrawal messages are sent to the remote PE, which will update its next hop to the backup. The backup PE takes over immediately without waiting for the **es-activation-timer** to bring up its SAP/spoke-SDP.

Ethernet Segment Failures

When the SDP toward the primary (DF) fails, the backup PE needs to take over. An SDP failure is emulated and log 99 on PE-4 shows that SDP 46 is operational down and PE-4 is no longer the DF:

```
89 2016/07/11 10:54:35.10 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 46 changed to admin=up oper=down"
92 2016/07/11 10:54:35.10 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-45, EVI:2, Designated Forwarding state changed to:false"
```

Remote PEs receive route withdrawal updates (unreachable NLRI) from former DF PE-4, for example on PE-2:

```
109 2016/07/11 10:54:24.24 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 15 Len: 82 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-
AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: MAX-ET Label: 0
        Type: EVPN-Eth-
Seg Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:00:00:45:00:00:00:01, IP-Len: 4 Orig-IP-
Addr: 192.0.2.4
        Type: EVPN-
AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456 Label: 0
"
```

The backup PE-5 is promoted to primary (P=1, B=0) and sends BGP updates accordingly. The following AD per-EVI is received on PE-2:

```
112 2016 2016/07/11 10:54:24.25 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-
AD Len: 25 RD: 192.0.2.5:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456 Label: 41941
92
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
```

```
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

PE-5 brings up its spoke-SDP without waiting for the es-activation-timer and takes over immediately. It is now the only DF candidate, and therefore the DF, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45" evi 2
===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   yes 07/11/2016 10:50:53
===============================================================================
===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.5                               07/11/2016 10:51:07
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
*A:PE-5#
```

BGP updates are exchanged and the remote PEs will resolve the ESI to the TEP address 192.0.2.5. For example, on PE-2:

```
*A:PE-2# show service id 2 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01
=======================================================
BGP EVPN-MPLS Ethernet Segment Dest
=======================================================
Eth SegId                       Last Change
-------------------------------------------------------
01:00:00:00:00:45:00:00:00:01   07/11/2016 10:54:24
=======================================================
===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address             Egr Label               Last Change
                        Transport
-------------------------------------------------------------------------------
192.0.2.5               262137                  07/11/2016 10:54:24
                        ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

This process is always revertive; as soon as the SDP 46 is operationally up again, a new DF election is triggered with two DF candidates and PE-4 will be elected as DF.

Troubleshooting and Debugging

The following show and debug commands can be used in EVPN-VPWS:

- show redundancy bgp-evpn-multi-homing
- show router bgp routes evpn (and filters)
- show service evpn-mpls [<TEP ip-address>]
- show service id bgp-evpn
- show service id evpn-mpls (and modifiers)
- show service system bgp-evpn
- show service system bgp-evpn ethernet-segment (and modifiers)
- debug router bgp update
- log-id 99

Most of these commands have been shown in the preceding sections; some commands are shown in this section.

Information about the configured boot timers (before DF election) and ES activation timer (after the system has been elected DF) can be shown as follows:

```
*A:PE-2# show redundancy bgp-evpn-multi-homing
===============================================================================
Redundancy BGP EVPN Multi-homing Information
===============================================================================
Boot-Timer              : 10 secs
Boot-Timer Remaining    : 0 secs
ES Activation Timer     : 3 secs
===============================================================================
*A:PE-2#
```

See chapter EVPN for MPLS Tunnels for a detailed description of these timers.

The following command shows that the BGP route-type 4 (ES route) messages are only imported by the PEs in the same ES; for example, on PE-3:

```
*A:PE-3# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
```

```
-------------------------------------------------------------------------------
u*>i  192.0.2.2:0         01:00:00:00:00:23:00:00:00:01 192.0.2.2
      192.0.2.2
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-3#
```

On PE-4:

```
* A:PE-4# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  192.0.2.5:0         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      192.0.2.5
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-4#
```

The following command shows all the EVPN MPLS destinations toward TEP 192.0.2.4. Epipe 1 has an EVPN MPLS destination toward TEP 192.0.2.4 directly and Epipe 2 has an EVPN MPLS destination to ESI-45, which can be resolved to TEP 192.0.2.4. This is shown in the following output:

```
*A:PE-2# show service evpn-mpls 192.0.2.4
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
Service Id                          Egr Label
-------------------------------------------------------------------------------
1                                   262138
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Service Id          Eth Seg Id                          Egr Label
-------------------------------------------------------------------------------
2                   01:00:00:00:00:45:00:00:00:01       262135
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-MPLS ES BMac Dest
===============================================================================
```

```
Service Id               ES BMac                  Egr Label
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
*A:PE-2#
```

The following command lists all configured ESs on the system:

```
*A:PE-2# show service system bgp-evpn ethernet-segment
===============================================================================
Service Ethernet Segment
===============================================================================
Name                             ESI                             Admin     Oper
-------------------------------------------------------------------------------
ESI-23                           01:00:00:00:00:23:00:00:00:01 Enabled   Up
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
*A:PE-2#
```

In addition to the preceding commands, the following tools dump commands may be useful:

- **tools dump service evpn usage** - This command shows the number of EVPN-MPLS (and EVPN-VXLAN) destinations in the system.
- **tools dump service system bgp-evpn ethernet-segment <name> evi <[1..65535]> df** - This command computes the DF election for a specific ESI and EVI. For all-active, there is no DF election and all PEs forward traffic. For single-active, one PE will be active for a service while another PE will be backup. This command shows the DF (primary), even if it is not the local PE.

The usage of EVPN resources can be shown as follows:

```
*A:PE-2# tools dump service evpn usage
EVPN usage statistics at 000 05:16:09.290:
MPLS-TEP                                          :             1
VXLAN-TEP                                         :             0
Total-TEP                                         :      1/ 16383
Mpls Dests (TEP, Egress Label + ES + ES-BMAC)     :             2
Vxlan Dests (TEP, Egress VNI)                     :             0
Total-Dest                                        :      2/196607
Sdp Bind +  Evpn Dests                            :      2/245759
ES L2/L3 PBR                                      :      0/ 32767
*A:PE-2#
```

On PE-2, there is one MPLS-TEP (192.0.2.4 in Epipe 1 and Epipe 2) and there are two MPLS destinations: 192.0.2.4 and ESI 01:00:00:00:00:45:00:00:00:01. PE-5 is not an MPLS-TEP for PE-2, because it is not a primary and, therefore, not forwarding any traffic.

In all-active multi-homing, the DF election is not applicable:

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "ESI-23" evi 2 df
[07/08/2016 13:32:21] All Active VPWS - DF N/A
```

In single-active multi-homing, the following command shows which PE is the DF:

```
*A:PE-5# tools dump service system bgp-evpn ethernet-segment "ESI-45" evi 2 df
[07/08/2016 13:33:38] Computed DF: 192.0.2.4 (Remote) (Boot Timer Expired: Yes)
[07/08/2016 13:33:38] Computed Backup: 192.0.2.5 (This Node)
```

The command is launched on PE-5, which is a backup. The computed DF is PE-4 and the boot timer has expired, meaning there is no DF re-election pending.

Conclusion

EVPN-VPWS is a simplified point-to-point version of *RFC7432 - BGP MPLS-Based Ethernet VPN*. When used for Epipe and VPLS services, EVPN provides a unified control plane mechanism that simplifies the network deployment and operation. Single-active and all-active multi-homing can be used in Epipes; EVPN-VPWS is a differentiator of EVPN compared to traditional TLDP or BGP Epipe redundancy mechanisms. The Ethernet Segments used for multi-homing can be shared between EVPN VPLS and EVPN Epipes.

EVPN for PBB over MPLS (PBB-EVPN)

This chapter provides information about EVPN for PBB over MPLS (PBB-EVPN).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 13.0.R6. The CLI in the current edition is based on release 14.0.R3.

Important note: A prerequisite is to read the EVPN for MPLS Tunnels chapter.

Overview

EVPN for Provider Backbone Bridging (PBB) over MPLS (hereafter called PBB-EVPN) is specified in RFC 7623, *Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)*. It provides a simplified version of EVPN-MPLS for cases where the network requires very high scalability and does not need all the advanced features supported by EVPN-MPLS (but still requires single-active and all-active multi-homing capabilities). Table 21 provides a comparison between the capabilities of EVPN and PBB-EVPN in SR OS, and may help to choose between them when designing a VPN service.

Table 21 **EVPN and PBB-EVPN SR OS Feature Comparison**

VPN requirements	EVPN	PBB-EVPN	Comments
All-active Multi-Homing (MH) (flow-based load-balancing)	Yes	Yes	Allows better bandwidth utilization
Single-active MH (service-based load-balancing)	Yes	Yes	

Table 21 **EVPN and PBB-EVPN SR OS Feature Comparison (Continued)**

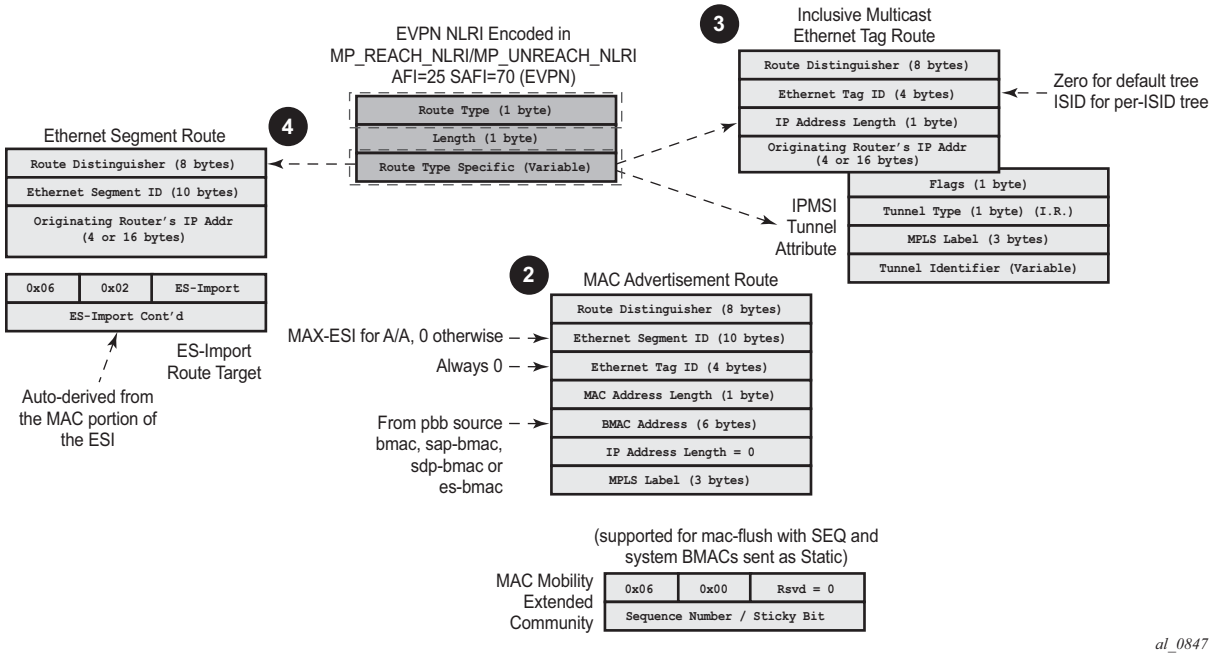
VPN requirements	EVPN	PBB-EVPN	Comments
Ethernet Local Area Network (ELAN) and point-to-point E-Line services	Yes	Yes	
Inter-subnet-forwarding	Yes	No	Allows combined layer 2/layer 3 services. EVPN
Proxy-address resolution protocol / neighbor discovery (Proxy-ARP/ND) and IP-duplication protection	Yes	No	Allows broadcast, unknown and multicast (BUM) traffic reduction and better security
Customer MAC (CMAC) protection	Yes	No	Allows protecting key static CMACs
Data Center integration	Yes	No	Integration with VXLAN and Nuage virtualized services directory (VSD)
Control plane overhead	Medium	Low	PBB-EVPN only advertises Backbone MACs (BMACs) and no route type 1s
Confinement of CMAC learning	No	Yes	CMACs are only learned on PEs with flows using those CMACs
CMAC summarization	No	Yes	Aggregation of CMACs into BMACs

PBB-EVPN is a combination of 802.1ah PBB and RFC7432, *BGP MPLS-Based Ethernet VPN* (EVPN-MPLS), and reuses the PBB-Virtual private LAN service (VPLS) service model, where border gateway protocol BGP-EVPN is enabled in the backbone VPLS (B-VPLS) domain. EVPN is used as the control plane in the B-VPLS domain to control the distribution of BMACs and set up per-backbone service instance identifier (ISID) flooding trees for service instance VPLS (I-VPLS) services. The learning of the CMACs, either on local SAPs/SDP-bindings or associated with remote BMACs, is still performed in the data plane. Only the learning of BMACs in the B-VPLS is performed through BGP.

The SR OS PBB-EVPN implementation supports I-VPLS and PBB-Epipe services, including single-active and all-active multi-homing.

Because PBB-EVPN is based on the same control plane model as EVPN for MPLS, Nokia recommends reading the EVPN for MPLS Tunnels chapter before configuring PBB-EVPN. PBB-EVPN uses a subset of the BGP-EVPN routes described in EVPN for MPLS Tunnels as shown in Figure 261.

Figure 261 **EVPN Route Types**



When no EVPN multi-homing is used in the network, only the base routes are used. Route types 2 and 3 are considered the base and mandatory routes:

- Route type 2 — (B) MAC route — In PBB-EVPN, this route type is used for the advertisement of BMACs that will be installed in the remote forwarding data bases (FDBs). There are no IP addresses advertised in PBB-EVPN. The MAC mobility extended community is used for advertising system BMACs as **protected** (with the sticky bit set) and it is also used for CMAC flush in some single-homing scenarios that will be described later.
- Route type 3 — Inclusive Multicast route — This route type is used for the advertisement of the I-VPLS ISIDs (no Epipes) and the desired multicast tree for each of them. The ISIDs are encoded in the ethernet-tag field of the network layer reachability information (NLRI). When the B-VPLS is created with no shutdown, an Inclusive Multicast route with ISID = 0 is advertised. This is for the creation of the default multicast tree.

When EVPN multi-homing is used in an ISID, route type 4 is used. In PBB-EVPN, there is no route type 1 advertised when multi-homing is used on the ISID services (I-VPLS and Epipes). Only route type 4 (Ethernet segment (ES) route) is used, and in the same way as it is for EVPN-MPLS. See the EVPN for MPLS Tunnels example for more information about ES routes, how they are formed, and how their RT/RD values are populated.

Configuration

This example describes the basic PBB-EVPN configuration first (without multi-homing) and how the flood containment is handled in PBB-EVPN. Flood containment refers to the efficient distribution of the BUM traffic generated for an ISID.

Networks are not always greenfield, so a smooth migration of PBB-EVPN from PBB-VPLS is required to minimize the effect on existing services. This example also describes this migration, starting from a common PBB-VPLS configuration.

Finally, this example describes the configuration of PBB-EVPN multi-homing.

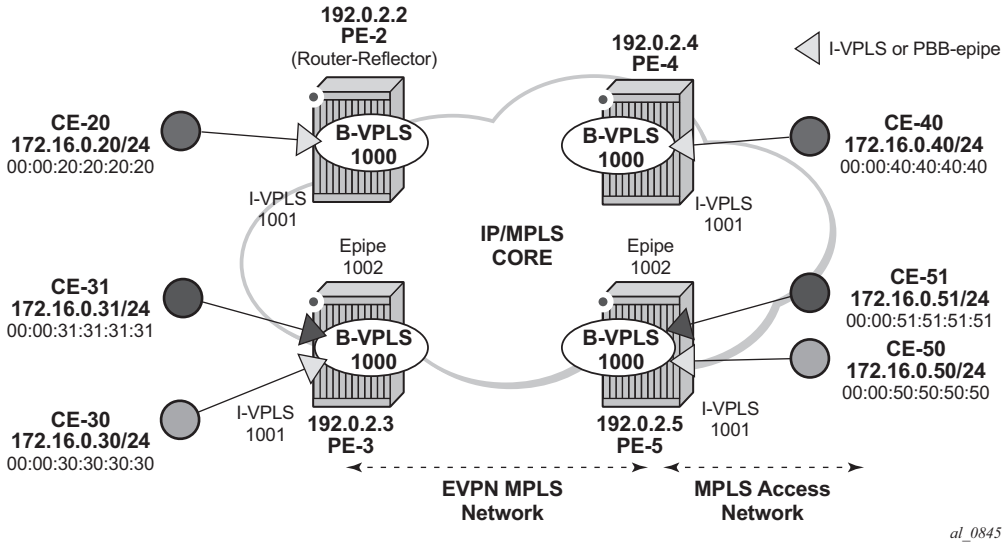
The same setup described in the VPN for MPLS tunnels example is used:

- Four PEs in the core (PE-2, PE-3, PE-4, and PE-5).
- The PEs are interconnected in the same way as explained in EVPN for MPLS Tunnels with the same IP addressing, IS-IS, transport LDP, and BGP peering configuration. There is not any difference with the basic infrastructure. See the EVPN for MPLS Tunnels chapter if more information is required.
- When configuring multi-homing, MTU-1 and MTU-6 are connected to the core.

PBB-EVPN Configuration without Multi-Homing

Figure 262 shows the network setup used in this chapter.

Figure 262 **PBB-EVPN Network without Multi-Homing**



When configuring PBB-EVPN:

- There is no difference at the access side (I-VPLS and Epipe configuration) compared to other PBB technologies supported in SR OS, such as shortest path bridging for MAC (SPBM) or PBB-VPLS.
- The B-VPLS becomes an EVPN-MPLS service, where bgp-evpn mpls is added.

The following output shows an example of a basic configuration in PE-3. B-VPLS 1000 is bgp-evpn enabled and I-VPLS 1001 and Epipe 1002 are linked to B-VPLS 1000.

```
configure
    service
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 1001 customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            sap 1/2/1:1001 create
            exit
            no shutdown
        exit
        epipe 1002 customer 1 create
            pbb
                tunnel 1000 backbone-dest-mac 00:00:00:00:00:05 isid 1002
            exit
            sap 1/2/1:1002 create
            exit
            no shutdown
        exit
```

In the preceding output, there is no new configuration needed for I-VPLS/Epipe services. As for the B-VPLS, the output shows the minimum configuration required. If needed, the following parameters can be modified under **bgp-evpn**:

```
*A:PE-2# configure service vpls 1000 bgp-evpn
  - bgp-evpn
```

```
 - no bgp-evpn

[no] cfm-mac-advert* - Enable/disable the advertisement of MEP, MIP, and VMEP MAC
                       addresses over the BGP EVPN
[no] evi             - EVPN Identifier
[no] ingress-repl-i* - Configure BGP EVPN IMET-IR route advertisement
[no] ip-route-adver* - Configure BGP EVPN IP Route Advertisement
[no] mac-advertisem* - Configure BGP EVPN MAC Advertisement
     mac-duplication + Configure BGP EVPN MAC Duplication
     mpls            + Configure BGP EVPN mpls
[no] unknown-mac-ro* - Configure BGP EVPN Unknown MAC Route
     vxlan           + Configure BGP EVPN vxlan


*A:PE-2# configure service vpls 1000 bgp-evpn mpls
  - mpls

     auto-bind-tunn* + Configure BGP EVPN mpls auto-bind-tunnel
[no] control-word    - Enable/disable setting the CW bit in the label message
     ecmp            - Configure maximum ECMP routes information
[no] entropy-label   - Enable/disable use of entropy-label
[no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
[no] ingress-replic* - Use the same label as the one advertised for unicast traffic
[no] restrict-prote* - Enable/disable protected src MAC restriction
[no] shutdown        - Administratively Enable/Disable BGP-EVPN mpls
[no] split-horizon-* - Configure a split-horizon-group
```

A detailed description of these commands is included in EVPN for MPLS Tunnels. In addition to the preceding commands, the following **service>(b-)vpls>pbb** commands are relevant for PBB-EVPN in the B-VPLS service:

- **force-qtag-forwarding** allows the transparent transport of the customer 802.1p bits across the B-VPLS services.
- **source-bmac** can modify the source BMAC for all the PBB packets containing traffic from non-multi-homed I-VPLS and Epipe services.
- **use-es-bmac** instructs the system to use an ES-specific BMAC for traffic coming from an ES on an I-VPLS or Epipe.
- **use-sap-bmac** instructs the system to use a SAP-specific BMAC for traffic coming from an MC-LAG I-VPLS/Epipe SAP.

Flood Containment for I-VPLS Services

In general, PBB technologies in SR OS support a way to contain flooding for a specified I-VPLS ISID, so that BUM traffic for that ISID only reaches the PEs where the ISID is locally defined. Each PE creates a multicast forwarding information base (MFIB) per I-VPLS ISID on the B-VPLS instance. That MFIB supports SAP/SDP-binding endpoints that can be populated by:

- MMRP in regular PBB-VPLS

- IS-IS in SPBM

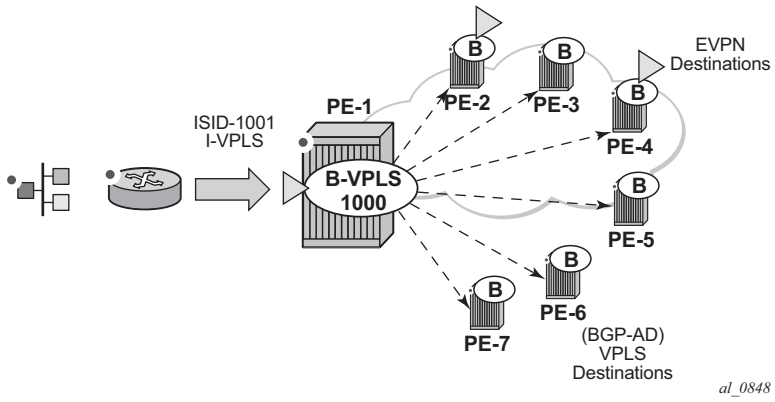
In PBB-EVPN, B-VPLS EVPN destinations can be added to the MFIBs using EVPN Inclusive Multicast Ethernet tag routes when they include the ISID in the Ethernet-tag. By default, when a B-VPLS is successfully enabled (no shutdown), the PE advertises:

- An Inclusive Multicast route for ISID = 0 — This allows the remote PEs to add the advertising PE to the default-multicast-list for the B-VPLS.
- An Inclusive Multicast route for each local ISID defined in the system (a local ISID includes configured I-VPLS and static-ISIDs) — This allows the remote PEs to create MFIB entries in the B-VPLS for the received ISIDs.

Because EVPN destinations, B-SAPs, and B-spoke-SDPs can coexist in the same B-VPLS, be aware of the different flooding lists created and how they are used in a B-VPLS. Figure 263 illustrates this concept with an example for B-VPLS 1000 in PE-1. The assumptions are:

- I-VPLS 1001 is created in PE-1, PE-2, and PE-4 only.
- PE-1, PE-2, PE-3, PE-4, and PE-5 support **bgp-evpn** in B-VPLS 1000.
- PE-6 and PE-7 only support spoke-SDPs.
- PE-1 is connected to all six PEs.

Figure 263 **PBB-EVPN — Flooding Lists**



In this situation, PE-1 creates two flooding lists in B-VPLS 1000:

- Default-multicast-list — composed of:
  - All the EVPN PEs that advertised ISID = 0 (PE-2, PE-3, PE-4, PE-5).
  - All the B-spoke-SDPs (or B-SAPs) (PE-6, PE-7).

  - All the EVPN PEs that advertised ISID 1001 and no ISID 0 (if an isid-policy is created in PE-1 stating **use-def-mcast** for ISID 1001). Note: third-party PEs may not advertise ISID = 0, but only non-zero ISIDs.
- MFIB for ISID 1001 is composed of:
  - All the EVPN PEs that advertised ISID 1001 (PE-2 and PE-4) unless there is an isid-policy in PE-1 stating **use-def-mcast** for ISID 1001.
  - Static-ISIDs defined in manual B-spoke-SDPs and B-SAPs (static-isids cannot be created on BGP-AD auto-discovered B-spoke-SDPs).

Based on the above, when BUM traffic is sent to I-VPLS 1001 on PE-1:

- The traffic is encapsulated in PBB with the group BMAC for ISID 1001 and sent (by default) to the MFIB created for ISID 1001 (PE-2 and PE-4).
- If an isid-policy is added with **use-def-mcast** for ISID 1001, the BUM traffic is encapsulated in PBB with the group BMAC for ISID 1001 and sent to the default-multicast-list, that is, all six remote PEs.

Referring to Figure 262, the following output illustrates the use of the isid-policy in PBB-EVPN. PE-2 does not have any isid-policy configured; when it receives BUM traffic from the local I-VPLS 1001, it uses the MFIB for ISID 1001:

```
configure
    service
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:02
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown


*A:PE-2# show service id 1000 mfib
===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                        Blk
-------------------------------------------------------------------------------
*               01:1E:83:00:03:E9     b-eMpls:192.0.2.3:262134    Local    Fwd
                                      b-eMpls:192.0.2.4:262130    Local    Fwd
                                      b-eMpls:192.0.2.5:262132    Local    Fwd
-------------------------------------------------------------------------------
```

```
Number of entries: 1
===============================================================================
```

An isid-policy can be added to modify this behavior and allow PE-2 to use the default-multicast-list. If I-VPLS 1001 exists in all the remote PEs (as in this example), using the default-multicast list is as efficient as using the MFIB and saves expensive MFIB resources. In the following output, as soon as the isid-policy is added, the MFIB entries for ISID 1001 are removed and PE-2 starts using the default-multicast list. The **tools dump service id 1000 evpn-mpls default-multicast-list** command shows the EVPN destinations that are part of the default-multicast list:

```
configure
    service
        vpls 1000
            isid-policy
                entry 10 create
                    use-def-mcast
                    range 1001 to 2000
                exit
            exit
```

```
*A:PE-2# tools dump service id 1000 evpn-mpls default-multicast-list
-------------------------------------------------------------------------------
TEP Address                             Egr Label
                                        Transport
-------------------------------------------------------------------------------
192.0.2.3                               262134
                                        ldp
192.0.2.4                               262130
                                        ldp
192.0.2.5                               262132
                                        ldp
-------------------------------------------------------------------------------
```

```
*A:PE-2# show service id 1000 mfib
===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                         Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

PBB-VPLS to PBB-EVPN Migration

The principles required for migrating a PBB-VPLS network to PBB-EVPN are explained in section the **VPLS to EVPN-MPLS Integration** of the EVPN for MPLS Tunnels chapter. Those principles are also applicable to EVPN destinations and spoke-SDPs in the B-VPLS and can be summarized in three points:

- Systems with an EVPN destination and SDP-binding to the same far-end IP bring down the SDP-binding. This avoids loops when both constructs exist in the same network.
- SDP-bindings and EVPN destinations can be placed in the same **split-horizon-group (SHG).** When traffic from an SDP-binding/EVPN destination belonging to that SHG is received on a PE, it is never forwarded to another SDP-binding/ EVPN destination on the same SHG.
- MAC addresses learned on an SDP-binding or SAP, that belong to an SHG where EVPN destinations are also created, are not advertised in BGP-EVPN.

Based on those principles, this section describes how to migrate a PBB-VPLS network to PBB-EVPN. The network in Figure 262 represents a regular PBB-VPLS network that needs to be migrated to PBB-EVPN.

In that network, the four PEs are running BGP-AD and TLDP for the discovery and setup of the pseudowires in the B-VPLS instance. The advantage of this configuration is that the migration can be done node-by-node and with minimum impact on customer service.

Initial Configuration

Initially, the network is configured for PBB-VPLS with BGP-AD in B-VPLS 1000. An EVPN family is to be added. At the access, I-VPLS 1001 is connected to the CEs. As an example, the configuration in PE-3 is shown. An equivalent configuration exists in the other three PEs.

Note: The EVPN family is added to the BGP configuration since PBB-EVPN uses this address family. Assuming there are redundant route reflectors (RRs), the addition of EVPN can be done without service impact. In this example the assumption is that the PEs are already configured with the EVPN family.

```
configure
    router
        bgp
            vpn-apply-import
            vpn-apply-export
```

```
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family l2-vpn evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit

configure
    service
        pw-template 1 create
            split-horizon-group "CORE"
            exit
        exit
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
            exit
            bgp
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 64500:1000
                no shutdown
            exit
            no shutdown
        exit
        vpls 1001 customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            sap 1/2/1:1001 create
            exit
            no shutdown
        exit

*A:PE-3# show service id 1000 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1000               Vpn Id            : 0
Service Type      : b-VPLS

---snip---

Oper Backbone Src : 00:00:00:00:00:03
Use SAP B-MAC     : Disabled
i-Vpls Count      : 1
Epipe Count       : 0
Use ESI B-MAC     : Disabled
-------------------------------------------------------------------------------
```

```
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm   Opr
------------------------------------------------------------------------------
sdp:17405:4294967293 SB(192.0.2.5)        BgpAd        0       8974    Up    Up
sdp:17406:4294967294 SB(192.0.2.4)        BgpAd        0       8974    Up    Up
sdp:17407:4294967295 SB(192.0.2.2)        BgpAd        0       8974    Up    Up
==============================================================================
* indicates that the corresponding row element may have been truncated.
```

Multiple MAC registration protocol (MMRP) is not used in the B-VPLS instance. If it were enabled, MMRP would have to be disabled in the network before this migration. If there are ISIDs using B-VPLS SDP-bindings to reach some remote locations and B-VPLS EVPN destinations to reach others, the default-multicast list must be used in the current release (the MFIB cannot be used if there is a mix of both types). Therefore, during the migration process, the ISIDs must be added to the default-multicast-list.

Step 1. Add service-level SHG (if not already there).

From the first node being migrated to PBB-EVPN to all nodes migrated, PBB-VPLS and PBB-EVPN have to coexist within the same meshed network. That is, EVPN-MPLS destinations and SDP-bindings need to be defined in the same **split-horizon-group**. Therefore, if there is no **split-horizon-group** defined in the B-VPLS, the first step is to add it. In our example, the **split-horizon-group** is defined at the **config>service>pw-template>level**; therefore, it has to be added at the B-VPLS level.



Note: When the **service>split-horizon-group** is removed, an eval-pw-template must be performed.



Note: After adding the **split-horizon-group** at the service level, an eval-pw-template must be performed again so that the SDP-bindings take the new SHG configuration.



Note: During the time between the **split-horizon-group** being removed and added back again, the SDP-bindings can forward BUM traffic to each other, so this operation must be done carefully to avoid loops.

Assuming that the first node to be migrated is PE-3, the following output shows the procedure for adding the **split-horizon-group** at the service level.

```
*A:PE-3# configure service pw-template 1 no split-horizon-group
```

```
*A:PE-3# tools perform service id 1000 eval-pw-template 1 allow-service-impact
eval-pw-template succeeded for Svc 1000 17405:4294967293 Policy 1
eval-pw-template succeeded for Svc 1000 17406:4294967294 Policy 1
eval-pw-template succeeded for Svc 1000 17407:4294967295 Policy 1


*A:PE-3# configure
    service
        vpls 1000
            split-horizon-group "CORE" create
            exit
            bgp
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit


*A:PE-3# tools perform service id 1000 eval-pw-template 1 allow-service-impact
eval-pw-template succeeded for Svc 1000 17405:4294967293 Policy 1
eval-pw-template succeeded for Svc 1000 17406:4294967294 Policy 1
eval-pw-template succeeded for Svc 1000 17407:4294967295 Policy 1


*A:PE-3>config>service>vpls# info
----------------------------------------------
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
            exit
            split-horizon-group "CORE" create
            exit
            bgp
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
            bgp-ad
                vpls-id 64500:1000
                no shutdown
            exit
            stp
                shutdown
            exit
            no shutdown
```

Step 2. Add bgp-evpn and ISID-policy configuration to the B-VPLS.

After the B-VPLS is configured with the split-horizon-group, the bgp-evpn configuration and ISID-policy can be added (still in shutdown).

```
configure
    service
        vpls 1000
            bgp-evpn
                evi 1000
                mpls
                    split-horizon-group "CORE"
                    auto-bind-tunnel
                        resolution any
                    exit
                    shutdown
```

```
            exit
        exit
        isid-policy
            entry 10 create
                use-def-mcast
                range 1001 to 3000
            exit
        exit
```

Step 3. Configure bgp-evpn mpls no shutdown on the PE.

When the configuration is ready, the **bgp-evpn mpls** context can be **no shutdown**.

```
*A:PE-3# configure service vpls 1000 bgp-evpn mpls no shutdown
```

The preceding **no shutdown** triggers a route-refresh message for the EVPN family from PE-3, but no changes happen because PE-3 does not create any EVPN destinations until it imports EVPN routes from the other PEs. The three spoke-sdps to the remote PEs are still up.

Step 4. Repeat steps 1 to 3 for the second PE.

The same steps 1 to 3 are repeated for PE-5. When **bgp-evpn mpls no shutdown** is executed, PE-5 sends a route-refresh and gets the BGP-EVPN routes from PE-3. As a result of that, PE-3 brings down the spoke-sdp to PE-5 and creates an EVPN destination to PE-5. The same process happens in PE-5. The following CLI output shows the received routes in PE-3 and spoke-SDP going down.

```
3 2016/07/04 10:58:50.21 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 117
    Flag: 0x90 Type: 14 Len: 47 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-Incl-mcast Len: 17 RD: 64500:1000, tag: 0, orig_addr len: 32,
                              orig_addr: 192.0.2.5
        Type: EVPN-Incl-mcast Len: 17 RD: 64500:1000, tag: 1001, orig_addr len: 32,
                              orig_addr: 192.0.2.5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags [Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 4194112
        Tunnel-Endpoint 192.0.2.5
```

```
"
```

Log 99 shows that spoke SDP 17405:4294967293 is operationally down:

```
187 2016/07/04 10:58:50.20 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 17405:4294967293 in service 1000 (customer 1) changed to admin=
up oper=down flags="
```

Spoke SDP 17405:4294967293 is the spoke SDP toward PE-5 and it is kept down:

```
*A:PE-3# show service id 1000 base
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:17405:4294967293 SB(192.0.2.5)        BgpAd        0       8974    Up   Down
sdp:17406:4294967294 SB(192.0.2.4)        BgpAd        0       8974    Up   Up
sdp:17407:4294967295 SB(192.0.2.2)        BgpAd        0       8974    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The reason why the spoke SDP toward PE-5 is down is an EVPN route conflict:

```
*A:PE-3# show service id 1000 sdp 17405:4294967293 detail | match Flag context all
Flags              : PWPeerFaultStatusBits
                     EvpnRouteConflict
```

An EVPN destination to PE-5 is created:

```
*A:PE-3# show service id 1000 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.5       262132        1           Yes             07/04/2016 10:58:50
                ldp
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

Step 5. Repeat Steps 1 to 3 for the rest of the PEs.

The same process is repeated in all the PEs, node-by-node. The service impact for the I-VPLS 1001 is minimal.

Step 6. (Optional) Remove the ISID policy.

When all the PEs in the B-VPLS 1000 are migrated, the isid-policy can optionally be removed, node-by-node. This forces the B-VPLS instance to start using the MFIB to send I-VPLS BUM traffic to the remote nodes. This has no effect on Epipes (traffic is always unicast for Epipes).

Before removing the ISID-policy and starting to use the MFIB, it is recommended to check that the Inclusive Multicast routes for an ISID to the remote PEs are all active. Otherwise, connectivity for BUM traffic could be interrupted if any of the expected routes are not active. This is illustrated for PE-3.

```
*A:PE-3# show service id 1000 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.2       262134        1           Yes             07/04/2016 12:06:43
                ldp
192.0.2.4       262130        1           Yes             07/04/2016 12:06:49
                ldp
192.0.2.5       262132        1           Yes             07/04/2016 10:58:50
                ldp
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
```

The routes for ISID 1001 are valid and used by BGP (flags **u*>i**):

```
*A:PE-3# show router bgp routes evpn inclusive-mcast tag 1001
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr            NextHop
      Tag
-------------------------------------------------------------------------------
u*>i  64500:1000          192.0.2.5           192.0.2.5
      1001

u*>i  64500:1000          192.0.2.4           192.0.2.4
      1001

u*>i  64500:1000          192.0.2.2           192.0.2.2
      1001
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

There are no entries in the MFIB:

```
*A:PE-3# show service id 1000 mfib
===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                         Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

The ISID policy is removed as follows:

```
configure service vpls 1000 isid-policy no entry 10
```

After removing the ISID-policy, the MFIB is populated with entries for the ISID 1001 group BMAC to the three remote PEs where ISID 1001 is defined:

```
*A:PE-3# show service id 1000 mfib
===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                         Blk
-------------------------------------------------------------------------------
*               01:1E:83:00:03:E9     b-eMpls:192.0.2.2:262134    Local    Fwd
                                      b-eMpls:192.0.2.4:262130    Local    Fwd
                                      b-eMpls:192.0.2.5:262132    Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

Step 7. (Optional) Remove the BGP-AD configuration.

The BGP-AD configuration can stay in the B-VPLS services. However, when the entire network is migrated to PBB-EVPN, all the spoke-sdps will be operationally down and, even if they are not forwarding traffic, they consume resources in the system. Consider removing the BGP-AD configuration and, therefore, the spoke-SDPs.

The following example shows the removal of bgp-ad in PE-4. Be aware that when bgp-ad is removed from the configuration, if the RD/RT was derived from the vpls-id (as in this example), a new RD/RT must be auto-derived for the service. Therefore, new updates will be sent for all the EVPN NLRIs, as shown in the following output.

```
*A:PE-4# show service id 1000 bgp
===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
```

```
Oper Route Dist      : 64500:1000
Oper RD Type         : derivedVpls
Rte-Target Import    : None                  Rte-Target Export: None
Oper RT Imp Origin   : derivedVpls           Oper RT Import   : 64500:1000
Oper RT Exp Origin   : derivedVpls           Oper RT Export   : 64500:1000

PW-Template Id       : 1                     PW-Template SHG  : CORE
Oper Group           : None
Mon Oper Group       : None
BFD Template         : None
BFD-Enabled          : no                    BFD-Encap        : ipv4
Import Rte-Tgt       : None
-------------------------------------------------------------------------------
===============================================================================
```

BGP-AD is disabled as follows:

```
*A:PE-4# configure service vpls 1000 bgp-ad shutdown
```

After BGP-AD is shutdown, the spoke SDP bindings are deleted.

```
175 2016/07/04 12:41:55.27 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 17407:4294967295 in service 1000 (customer 1) changed to admin=down
 oper=down flags=sdpBindAdminDown noIngressVcLabel noEgressVcLabel "

176 2016/07/04 12:41:55.27 UTC MAJOR: SVCMGR #2320 Base
"Service Id 1000, Dynamic bgp-l2vpn SDP Bind Id 17407:4294967295 was deleted."

177 2016/07/04 12:41:55.27 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 17407 changed to admin=down oper=down"

178 2016/07/04 12:41:55.27 UTC MAJOR: SVCMGR #2319 Base
"Dynamic bgp-l2vpn SDP 17407 (192.0.2.2) was deleted."
```

The PW template binding is removed as follows:

```
*A:PE-4# configure service vpls 1000 bgp no pw-template-binding 1
```

The BGP-AD configuration is removed as follows:

```
*A:PE-4# configure service vpls 1000 no bgp-ad
```

Initially, the RD/RT was derived from the vpls-id (64500:1000). After the BGP-AD configuration is removed, a new RD/RT must be auto-derived from the EVI:

```
*A:PE-4# show service id 1000 bgp
===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.4:1000
Oper RD Type         : derivedEvi
Rte-Target Import    : None                  Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi            Oper RT Import   : 64500:1000
Oper RT Exp Origin   : derivedEvi            Oper RT Export   : 64500:1000
```

```
PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

In this case, the system picks up the RD in the following order:

1. Manual RD or auto-RD always take precedence when configured.
2. If no manual/auto-RD, the RD is derived from the **bgp-ad vpls-id**.
3. If no manual/auto-rd/vpls-id configuration, the RD is derived from the **bgp evpn evi**.
4. If no manual/auto-rd/vpls-id/evi configuration, there will not be RD and the service will fail.

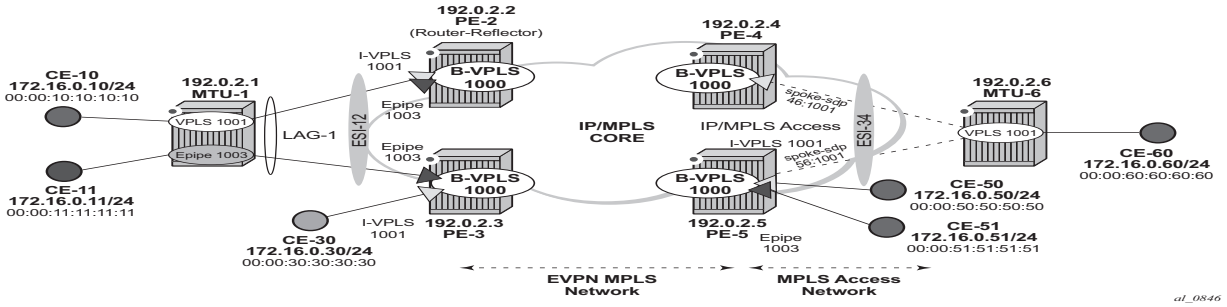
If in the migration from BGP-AD to BGP-EVPN, the advertisement of new updates is not needed, the initial configuration must include manual/auto-RDs. If manual/auto-RDs were not included, a **bgp-ad shutdown** would not cause the change of RD and the consequent BGP updates.

PBB-EVPN Multi-Homing

This section provides configuration guidelines for PBB-EVPN multi-homing. In the same way that EVPN-MPLS supports single-active and all-active multi-homing, PBB-EVPN can also be configured to support both modes. The same **ethernet-segment** that is used for regular EVPN-MPLS service saps and spoke-sdps can be shared with I-VPLS/Epipe SAPs and spoke-SDPs.

Figure 264 shows the example setup used in this section.

Figure 264 **PBB-EVPN Multi-homing**




```
configure port 1/2/[1..2] ethernet mode access
configure port 1/2/[1..2] ethernet encap-type dot
configure port 1/2/[1..2] no shutdown
```

MTU-1 and MTU-6 have been added to the network (compared to Figure 262). I-VPLS 1001 has two new sites that are multi-homed to the PBB-EVPN network. MTU-1 uses all-active multi-homing, whereas MTU-6 is connected to a single-active ES. As with EVPN-MPLS, all-active multi-homing is only supported when a LAG is used at the access. Single-active multi-homing can be supported with regular Ethernet ports (that can form an independent LAG per PE) or SDPs.

Draft-ietf-l2vpn-pbb-evpn describes two types of system BMAC assignments that a PE can implement in a B-VPLS when ESs are present:

- Shared BMAC addresses that can be used for all the single-homed CEs and a number of multi-homed CEs connected to **ethernet-segments**.
- Dedicated BMAC addresses per **ethernet-segment**.

In this chapter and in SR OS terminology:

- A shared-bmac (in IETF) is a **source-bmac** as configured in **service>(b)vpls>pbb>source-bmac**. All the I-VPLS/Epipe traffic coming from single-homed CEs is sent encapsulated in a PBB packet with that **source-bmac**.
- A dedicated-bmac per ES (in IETF) is an **es-bmac** as activated in **service>(b)vpls>pbb>use-es-bmac** and generated from the combination of **vpls>pbb>source-bmac** plus **ethernet-segment>source-bmac-lsb**. If configured, any I-VPLS/Epipe traffic coming from an ES is encapsulated in a PBB packet with the es-bmac as the source BMAC.

The system allows the following user choices per B-VPLS and ES:

- A dedicated **es-bmac** per ES can be used. In that case, the **pbb>use-es-bmac** command is configured in the B-VPLS. In all-active multi-homing, all the PEs that are part of the ES source the PBB packets with the same source **es-bmac**; single-active multi-homing requires the use of a different **es-bmac** per PE.
- A non-dedicated **source-bmac** can be used (this is only possible in single-active multi-homing). In this case, the user does not configure **pbb>use-es-bmac** and the regular **source-bmac** is used for the traffic. A different **source-bmac** has to be advertised per PE.

As discussed, single-active multi-homing can use **source-bmacs** or **es-bmacs**. Using one type or another has a different impact on CMAC flushing, as illustrated in Figure 265.

- If **es-bmacs** are used, as shown on the right-hand side of Figure 265, a less-impacting CMAC flush is achieved, therefore minimizing the flooding after ES failures. In the case of ES failure, PE-1 withdraws the **es-bmac** 00:12 and the remote PE-3 only flushes the CMACs associated with that **es-bmac** (only the CMACs behind the CE are flushed).
- If **source-bmacs** are used, as shown on the left-hand side of Figure 265, in the case of ES failure, a BGP update with higher sequence number is issued by PE-1 and the remote PE-3 flushes all the CMACs associated with the **source-bmac**. Therefore, all the CMACs behind the B-VPLS of the PEs will be flushed, as opposed to only the CMACs behind the CE of the Ethernet Service Instances (ESIs).

Figure 265 **The Use of es-bmac to Minimize CMAC Flush**

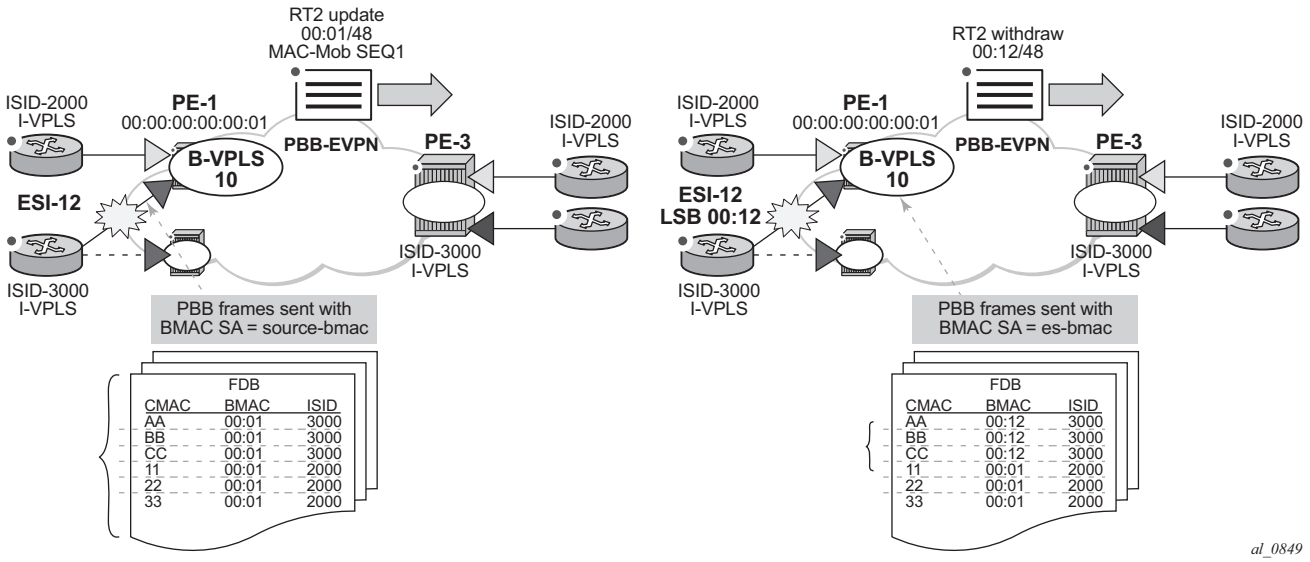


Table 22 shows the PBB-EVPN multi-homing combinations supported in the current release in the topology of Figure 264.

Table 22 **PBB-EVPN Multi-Homing Supported Combinations in SR OS**

CE Connectivity	PE Connectivity	PE Redundancy	BMAC Assignment	I-VPLS Support	Epipe Support
LAG (LACP optional)	LAG SAP	EVPN MH all-active	use-es-bmac (shared BMAC)	Yes	Yes
Ethernet ports (no LAG)	LAG SAP or port SAP	EVPN MH single-active	use-es-bmac (dedicated per PE)	Yes	No
Ethernet ports (no LAG)	LAG SAP or port SAP	EVPN MH single-active	source-bmac (dedicated per PE)	Yes	No
MPLS	spoke-SDP	EVPN MH single-active	source-bmac (dedicated per PE)	Yes	No
MPLS	spoke-SDP	EVPN MH single-active	use-es-bmac (dedicated per PE)	Yes	No

As an example, the configurations of the first, and last two, rows (lag sap all-active, mpls source-bmac, and mpls use-es-bmac, respectively) will be discussed in the following three sections.

PBB-EVPN All-Active Multi-Homing for I-VPLS and Epipes

Figure 264 shows a PBB-EVPN network where ESI-12 is configured as an all-active multi-homing ES on PE-2 and PE-3. Two services are using ESI-12: I-VPLS 1001 and Epipe 1003. The following output shows the relevant configuration in PE-2:

```
configure
    service
        pbb
            mac-name "PE-5" 00:00:00:00:00:05
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-12" create
                    esi 01:00:00:00:00:12:00:00:00:01
                    source-bmac-lsb 12-12 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
```

```
                multi-homing all-active
                lag 1
                no shutdown
            exit
        exit
    exit

    vpls 1000 customer 1 b-vpls create
        service-mtu 2000
        pbb
            source-bmac 00:00:00:00:00:02
            use-es-bmac
        exit
        split-horizon-group "CORE" create
        exit
        bgp
        exit
        bgp-evpn
            evi 1000
            mpls
                split-horizon-group "CORE"
                ecmp 2
                auto-bind-tunnel
                    resolution any
                exit
                no shutdown
            exit
        exit
        no shutdown
    exit
    vpls 1001 customer 1 i-vpls create
        pbb
            backbone-vpls 1000
            exit
        exit
        sap lag-1:1001 create
        exit
        no shutdown
    exit
    epipe 1003 customer 1 create
        pbb
            tunnel 1000 backbone-dest-mac "PE-5" isid 1003
        exit
        sap lag-1:1003 create
        exit
        no shutdown
    exit
```

The following output shows the relevant configuration in PE-3:

```
configure
    service
        pbb
            mac-name "PE-5" 00:00:00:00:00:05
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-12" create
```

```
            esi 01:00:00:00:00:12:00:00:00:01
            source-bmac-lsb 12-12 es-bmac-table-size 8
            es-activation-timer 3
            service-carving
                mode auto
            exit
            multi-homing all-active
            lag 1
            no shutdown
        exit
    exit
exit

vpls 1000 customer 1 b-vpls create
    service-mtu 2000
    pbb
        source-bmac 00:00:00:00:00:03
        use-es-bmac
    exit
    split-horizon-group "CORE" create
    exit
    bgp
    exit
    bgp-evpn
        evi 1000
        mpls
            split-horizon-group "CORE"
            ecmp 2
            auto-bind-tunnel
                resolution any
            exit
            no shutdown
        exit
    exit
    no shutdown
exit
vpls 1001 customer 1 i-vpls create
    pbb
        backbone-vpls 1000
        exit
    exit
    sap 1/2/1:1001 create
    exit
    sap lag-1:1001 create
    exit
    no shutdown
exit
epipe 1003 customer 1 create
    pbb
        tunnel 1000 backbone-dest-mac "PE-5" isid 1003
    exit
    sap lag-1:1003 create
    exit
    no shutdown
exit
```

The preceding configuration shows that Epipe 1003 has a PBB tunnel pointing at the PE-5 source-bmac. Epipe 1003 has the following configuration in PE-5 (the PBB tunnel points at the ESI-12 es-bmac):

```
configure
    service
        pbb
            mac-name "ES-MAC-12" 00:00:00:00:12:12
        exit
        epipe 1003 customer 1 create
            pbb
                tunnel 1000 backbone-dest-mac "ES-MAC-12" isid 1003
            exit
            sap 1/2/1:1003 create
            exit
            no shutdown
        exit
```

Source-bmacs and es-bmacs are distributed in BGP-EVPN. PE-2 and PE-3 will each advertise their own source-bmac in a MAC route with ESI-0 and the shared es-bmac with ESI-MAX (as per the RFC 7623). The es-bmac that each PE uses in a B-VPLS is derived from the configured **service>(b)vpls>pbb>source-bmac** (four high-order bytes) and the ESI-12 configured source-bmac-lsb. In this example, PE-2 and PE-3 will both derive es-bmac 00:00:00:00:12:12. For both PEs to derive the required same es-bmac, the four high-order bytes of the source-bmac must match on both PEs.

The **es-bmac-table-size** parameter modifies the default value (8) for the maximum number of es-bmacs that can be associated with the ethernet-segment across different B-VPLS services. When **source-bmac-lsb** is configured, the associated **es-bmac-table-size** is reserved out of the total FDB space.

The following outputs show the source-bmacs and es-bmac and how they are advertised and installed in the B-VPLS FDB.

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-12" | match BMAC
Source BMAC LSB         : 12-12
```

The following output shows that ES-BMAC is used and that the operational source-BMAC is 00:00:00:00:00:02.

```
*A:PE-2# show service id 1000 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1000               Vpn Id            : 0
Service Type      : b-VPLS
---snip---
Oper Backbone Src : 00:00:00:00:00:02
Use SAP B-MAC     : Disabled
i-Vpls Count      : 1
Epipe Count       : 1
```

```
Use ESI B-MAC     : Enabled
---snip---
```

The source BMAC LSB is configured with the same value on PE-2 and PE-3. The two low-order bytes of the ES-BMAC will be 12:12.

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-12" | match BMAC
Source BMAC LSB         : 12-12
```

On PE-3, ES-BMAC is used and the operational source BMAC is 00:00:00:00:00:03.

```
*A:PE-3# show service id 1000 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1000                Vpn Id            : 0
Service Type      : b-VPLS

--snipped--

Oper Backbone Src : 00:00:00:00:00:03
Use SAP B-MAC     : Disabled
i-Vpls Count      : 1
Epipe Count       : 2
Use ESI B-MAC     : Enabled
```

On PE-2, the FDB for B-VPLS 1000 has an entry for each of the other PEs. PEs do not show their own system BMACs in the FDB:

```
*A:PE-2# show service id 1000 fdb detail
===============================================================================
Forwarding Database, Service 1000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1000      00:00:00:00:00:03 eMpls:                   EvpnS:P  07/04/16 12:49:42
                            192.0.2.3:262136
1000      00:00:00:00:00:04 eMpls:                   EvpnS:P  07/04/16 12:49:41
                            192.0.2.4:262133
1000      00:00:00:00:00:05 eMpls:                   EvpnS:P  07/04/16 12:49:37
                            192.0.2.5:262142
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

On PE-4, the FDB for B-VPLS 1000 has an entry for each of the other PEs and an entry for the ES-BMAC of ES “ESI-12”:

```
*A:PE-4# show service id 1000 fdb detail
===============================================================================
Forwarding Database, Service 1000
```

```
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1000      00:00:00:00:00:02 eMpls:                   EvpnS:P  07/04/16 12:49:56
                            192.0.2.2:262134
1000      00:00:00:00:00:03 eMpls:                   EvpnS:P  07/04/16 12:49:55
                            192.0.2.3:262136
1000      00:00:00:00:00:05 eMpls:                   EvpnS:P  07/04/16 12:49:50
                            192.0.2.5:262142
1000      00:00:00:00:12:12 eES:                     EvpnS:P  07/05/16 06:18:01
                            MAX-ESI
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

On PE-4, there are two BGP routes for ES-BMAC 00:00;00;00:12:12: one with next hop PE-2 and the other with next hop PE-3, as follows:

```
*A:PE-4# show router bgp routes evpn mac mac-address 00:00:00:00:12:12
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Ip Address
                                            NextHop
                                            Label1
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1000      00:00:00:00:12:12 ESI-MAX
      0                   Static            N/A
                                            192.0.2.2
                                            LABEL 262134

u*>i  192.0.2.3:1000      00:00:00:00:12:12 ESI-MAX
      0                   Static            N/A
                                            192.0.2.3
                                            LABEL 262136
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PBB-EVPN all-active multi-homing is based on the same concepts as EVPN-MPLS all-active multi-homing: DF election, split-horizon, and aliasing.

Designated Forwarder (DF) Election

Only the DF PE for an ISID will send multicast traffic to the ES. The DF PE for an ISID can be shown with the following command:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-12" isid 1003
===============================================================================
ISID DF and Candidate List
===============================================================================
Isid          SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1003          1003          0                   yes 07/05/2016 06:35:04
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.2                               07/05/2016 06:35:01
192.0.2.3                               07/05/2016 06:34:59
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

The DF PE identifies multicast traffic by looking at either the destination BMAC or the EVPN label (which can be unicast or multicast).

In the case of Epipes, there are also DF and non-DF PEs. However, traffic is usually unicast (sent to the PBB tunnel backbone-destination-bmac). The non-DF PE will usuallynot discard Epipe traffic to the ES, unless the packet comes with an EVPN multicast label. To avoid packet duplication at the CE for Epipes, it is recommended to either:

- configure **discard-unknown** on all the B-VPLS instances where there are PBB-Epipes. This will prevent the ingress PE from flooding Epipe traffic if the PBB tunnel BMAC is unknown in the FDB.
- configure **ingress-replication-bum-label** so that, when the PBB tunnel BMAC is unknown in the FDB, the ingress PE sends traffic with a multicast label. The non-DF will discard traffic identified as multicast at Epipes.

Ethernet-Segment Split-horizon

In PBB-EVPN all-active multi-homing, the split-horizon function is not based in the ESI label but in a source BMAC check. When BUM traffic is received on an I-VPLS, the PE will encapsulate it in PBB using the **es-bmac** as source BMAC and the group BMAC for the ISID. When the DF PE for the ISID receives that packet, it will not send it back to the ES if the packet is identified as being originated from the ES itself (based on the **es-bmac** shared between the PEs).

Aliasing

Aliasing is based on the advertisement of the same **es-bmac** with MAX-ESI from the PEs part of the same ES. PE-2 and PE-3 advertise the **es-bmac 00:00:00:00:12:12** with MAX-ESI (ESI = all FFs, as per the RFC 7623) and as Static (protected). When the remote PEs, PE-4, and PE-5, receive the two routes for the same BMAC and MAX-ESI, they will create a single EVPN-MPLS destination that will give more than one next-hop (in this case 2), as long as ecmp > 1:

```
*A:PE-4# show service id 1000 evpn-mpls
===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
-------------------------------------------------------------------------------
192.0.2.2       262134        1           Yes             07/04/2016 12:06:35
                ldp
192.0.2.3       262136        1           Yes             07/04/2016 12:06:35
                ldp
192.0.2.5       262142        1           Yes             07/04/2016 12:06:35
                ldp
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
===============================================================================
BGP EVPN-MPLS ES BMAC Dest
===============================================================================
ES BMAC Addr                            Last Change
-------------------------------------------------------------------------------
00:00:00:00:12:12                       07/05/2016 06:18:01
-------------------------------------------------------------------------------
Number of entries: 1
```

```
-------------------------------------------------------------------------------
===============================================================================
```

The EVPN-MPLS ES BMAC destination has two next hops: PE-2 and PE-3.

```
*A:PE-4# show service id 1000 evpn-mpls es-bmac 00:00:00:00:12:12
===============================================================================
BGP EVPN-MPLS ES BMAC Dest
===============================================================================
ES BMAC Addr                          Last Change
-------------------------------------------------------------------------------
00:00:00:00:12:12                     07/04/2016 12:06:35
===============================================================================
===============================================================================
BGP EVPN-MPLS ES BMAC Dest TEP Info
===============================================================================
TEP Address           Egr Label           Last Change
                      Transport
-------------------------------------------------------------------------------
192.0.2.2             262134              07/04/2016 12:06:35
                      ldp
192.0.2.3             262136              07/04/2016 12:06:35
                      ldp
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

A similar output will be obtained in PE-5. Unicast traffic entering I-VPLS 1001 in either PE-4 or PE-5 will be hashed and load balanced to PE-2 and PE-3 if the destination CMAC lookup yields an **es-bmac-dest**:

```
*A:PE-5# show service id 1001 fdb detail pbb
===============================================================================
Forwarding Database, i-Vpls Service 1001
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
00:00:10:10:10:10 eES-BMAC:             1000      00:00:00:00:12:12 L/21
                  00:00:00:00:12:12
00:00:30:30:30:30 b-eMpls:              1000      00:00:00:00:00:03 L/0
                  192.0.2.3:262136
00:00:50:50:50:50 sap:1/2/1:1001        1000      N/A               L/0
00:00:60:60:60:60 sdp:56:1001           1000      N/A               L/0
===============================================================================
```

Verify the FDB of I-VPLS 1001 for ES BMAC destination 00:00:00:00:12:12 as follows:

```
*A:PE-5# show service id 1001 fdb evpn-mpls es-bmac-dest 00:00:00:00:12:12
===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
```

```
1001      00:00:10:10:10:10 eES-BMAC:                L/51     07/05/16 10:57:12
                            00:00:00:00:12:12
-------------------------------------------------------------------------------
No. of Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

If a failure occurs in the ES, the PE will withdraw the **es-bmac** and the remote PEs will remove one next-hop of the ES-BMAC EVPN-MPLS destination.

For PBB-Epipes, aliasing will also work, as long as shared-queuing or policing are enabled on the ingress PE Epipe. In Figure 264, Epipe 1003 on PE-5 requires shared-queuing or policing at the ingress SAP. Otherwise, the traffic will be sent to only one PE of the ES (usually to the lower-IP PE).

For more information about the configuration of the **ethernet-segment** and its parameters, see the EVPN for MPLS Tunnels chapter.

PBB-EVPN Single-Active Multi-Homing for I-VPLS with source-bmacs

ESI-34 is a single-active **ethernet-segment** (see Figure 264) with SDPs linked to it. As indicated in Table 22, only I-VPLS services can be used in this configuration. As discussed in section PBB-EVPN Multi-Homing single-active ES and B-VPLS services can be configured to either use **source-bmacs** or **es-bmacs**. The following configuration shows the former option:

```
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-34" create
                    esi 01:00:00:00:00:34:00:00:00:01
                    source-bmac-lsb 34-04 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
```

```
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:04
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls
                    split-horizon-group "CORE"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 1001 customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            spoke-sdp 46:1001 create
            exit
            no shutdown
        exit
```

The configuration on PE-5 is similar:

```
configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-34" create
                    esi 01:00:00:00:00:34:00:00:00:01
                    source-bmac-lsb 34-05 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
            exit
        exit
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
```

```
            pbb
                source-bmac 00:00:00:00:00:05
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls
                    split-horizon-group "CORE"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 1001 customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            sap 1/2/1:1001 create
            exit
            spoke-sdp 56:1001 create
            exit
            no shutdown
        exit
```

With the preceding configuration, PE-4 and PE-5 will not advertise es-bmacs with MAX-ESI. Therefore, all the remote BMACs on PE-2/PE-3 are associated with regular backbone EVPN-MPLS destinations. The CMACs will be learned in the data plane associated with local SAP/SDP-bindings or remote BMACs. An example for the I-VPLS and B-VPLS FDB in PE-2 follows:

```
*A:PE-2# show service id 1001 fdb detail pbb
===============================================================================
Forwarding Database, i-Vpls Service 1001
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
00:00:10:10:10:10 sap:lag-1:1001        1000      N/A               L/206
00:00:50:50:50:50 b-eMpls:              1000      00:00:00:00:00:05 L/309
                  192.0.2.5:262142
00:00:60:60:60:60 b-eMpls:              1000      00:00:00:00:00:05 L/180
                  192.0.2.5:262142
===============================================================================
```

The B-VPLS FDB on PE-2 looks as follows:

```
*A:PE-2# show service id 1000 fdb detail
===============================================================================
Forwarding Database, Service 1000
```

```
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1000      00:00:00:00:00:03 eMpls:                   EvpnS:P  07/04/16 12:49:42
                            192.0.2.3:262136
1000      00:00:00:00:00:04 eMpls:                   EvpnS:P  07/04/16 12:49:41
                            192.0.2.4:262133
1000      00:00:00:00:00:05 eMpls:                   EvpnS:P  07/04/16 12:49:37
                            192.0.2.5:262142
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

In the preceding example, the DF for ISID 1001 is PE-5. With a failure event on the SDP to MTU-6, PE-5 will not withdraw the advertised **source-bmac** (because it is still being used as source-bmac for other services and even CEs within the same service). PE-5 will send an update of the same source-bmac instead, increasing the sequence number in the MAC mobility extended community. That will be a **flush-all-from-me** indication for the remote PEs (they will flush all the CMACs associated with the updated source-bmac, irrespective of the service).

When the former DF (PE-5) comes back up, PE-4 will become non-DF and will send a CMAC flush indication using the same mechanism as discussed above.

The following example shows a failure of SDP 56 in PE-5 and the corresponding DF switchover and CMAC flush.

```
*A:PE-5#
18 2016/07/05 12:23:57.24 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 56 changed to admin=up oper=down"

20 2016/07/05 12:23:57.24 UTC MINOR: SVCMGR #2095 Base
"Ethernet Segment:ESI-34, ISID:1001, Designated Forwarding state changed to:false"
```

PE-5 sends a BGP update with the same source-BMAC, increasing the sequence number in the MAC mobility extended community:

```
4 2016/07/05 12:23:57.24 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 96
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1000 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:00:00:00:05, IP len: 0, IP: NULL, label1: 4194272
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
```

```
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:1/Static
```

Individual SAP or spoke SDP failures do not trigger any MAC flush or DF re-election. This is as per RFC 7623. In EVPN-MPLS, individual SAP/spoke SDP failures are captured by the AD per-EVI withdrawal, which triggers a DF switchover.

PBB-EVPN Single-Active Multi-Homing for I-VPLS with es-bmacs

As discussed throughout this document, the use of **es-bmacs** for single-active multi-homing can minimize the number of CMACs flushed in a network. A simple change is necessary: activate the **use-es-bmac** command and ensure that the generated es-bmacs in PE-4 and PE-5 are different (the **source-bmac-lsb** in the previous configuration had different values for PE-4 and PE-5 already):

```
*A:PE-4# configure service vpls 1000 pbb use-es-bmac
*A:PE-5# configure service vpls 1000 pbb use-es-bmac
```

On PE-4, the source BMAC LSB in ESI-34 is configured with a value of 34-04:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-34" | match BMAC
Source BMAC LSB         : 34-04
```

On PE-5, the source BMAC LSB in ESI-34 is configured with a value of 34-05:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-34" | match BMAC
Source BMAC LSB         : 34-05
```

The remote PEs (such as PE-2 in the following output) will receive two more BMACs in BGP.

```
*A:PE-2# show service id 1000 fdb detail
===============================================================================
Forwarding Database, Service 1000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1000      00:00:00:00:00:03 eMpls:                   EvpnS:P  07/04/16 12:49:42
                            192.0.2.3:262136
1000      00:00:00:00:00:04 eMpls:                   EvpnS:P  07/04/16 12:49:41
                            192.0.2.4:262133
1000      00:00:00:00:00:05 eMpls:                   EvpnS:P  07/04/16 12:49:37
                            192.0.2.5:262142
1000      00:00:00:00:34:04 eMpls:                   EvpnS:P  07/05/16 13:41:03
                            192.0.2.4:262133
```

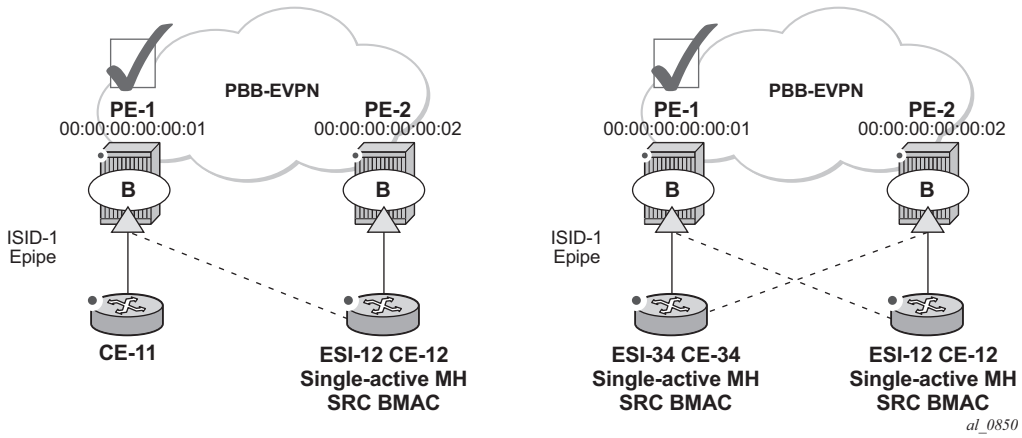
```
1000      00:00:00:00:34:05 eMpls:                    EvpnS:P  07/05/16 13:41:00
                            192.0.2.5:262142
-------------------------------------------------------------------------------
No. of MAC Entries: 5
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

The benefit is that in case of a failure in ESI-34 (as before) the es-bmac is withdrawn and the remote PEs will only flush the CMACs associated with the remote ES-34, as opposed to all the CMACs associated with PE-5.

PBB-EVPN Single-Active Multi-Homing for Epipes

In the network in Figure 264, Epipes can only support single-homing or all-active multi-homing but not single-active. For non-local-switching PBB-Epipes (there is a single SAP per Epipe), only all-active multi-homing is supported. Single-active multi-homing for local-switching enabled PBB-Epipes (two SAPs are defined within the PBB-Epipe instance) is only supported in the following scenarios.

Figure 266 **PBB-EVPN Single-Active Support for Epipes**



Single-active multi-homing is supported for redundancy in a two-node, three or four SAP, scenario, as displayed in Figure 266. In these two cases, the Epipe PBB tunnel will be configured with the source BMAC of the remote PE node. When two SAPs are active in the same Epipe, local-switching is used to exchange frames between the CEs.

All-active multi-homing is not supported for redundancy in this scenario because the PE-1 PBB tunnel cannot point at a locally defined ES-BMAC.

PBB-EVPN Multi-Homing Operation

See the EVPN for MPLS Tunnels chapter for the commands to operate **ethernet-segments**. Consider that there are no AD routes in PBB-EVPN. Also, the DF election algorithm will be based on the ISID values as opposed to EVIs.

Troubleshooting and Debug Commands

When troubleshooting PBB-EVPN networks, most of the troubleshooting commands discussed in EVPN for MPLS Tunnels can be used in the B-VPLS service and the base **service>system>bgp-evpn** instance. Some examples of useful commands are:

- show redundancy bgp-evpn-multi-homing
- show router bgp routes evpn (and filters)
- show service evpn-mpls [<TEP ip-address>]
- show service id bgp-evpn
- show service id evpn-mpls (and modifiers)
- show service id fdb pbb (and modifiers)
- show service system bgp-evpn
- show service system bgp-evpn ethernet-segment (and modifiers)
- debug router bgp update
- log-id 99

In addition, the following **tools dump** commands also discussed in EVPN for MPLS Tunnels can help too:

- tools dump service evpn usage
- tools dump service system bgp-evpn ethernet-segment <name> isid df (Note: **isid** is used instead of **evi**.)

There are two aspects that are specific to PBB-EVPN and not EVPN:

1. Consumption of virtual BMACs in the system— source-bmacs, sap-bmacs, sdp-bmacs, and es-bmacs are system BMACs that use FDB space but are not shown in the FDB together with the rest of the learned MACs. The following command provides information about the virtual system MACs consumed in the system.

```
*A:PE-3# tools dump redundancy src-bmac-lsb
Src-bmac-lsb:      3 (00-03) User: B-Vpls - 1 service(s)
Src-bmac-lsb:   4626 (12-12) User: ES
```

```
Total Src-bmac-lsbs = 2
```

2. Consumption of MFIBs — when ISIDs are not using the default-multicast list in the B-VPLS context for sending BUM traffic, an MFIB is consumed per ISID. The following command provides information about the consumption of MFIBs per system and per B-VPLS.

```
*A:PE-2# tools dump service vpls-pbb-mfib-stats detail

Service Manager VPLS PBB MFIB statistics at 000 05:06:39.170:

Usage per Service
   ServiceId    MFIB User      Count
  ------------+--------------+-------
   1000         Evpn                1
  ------------+--------------+-------
                       Total       1

MMRP
  Current Usage     :     0
  System Limit      :  8191 Full, 40959 ESOnly
  Per Service Limit :  2048 Full,  8192 ESOnly

SPB
  Current Usage     :     0
  System Limit      :  8191
  Per Service Limit :  8191

Evpn
  Current Usage     :     1
  System Limit      : 40959
  Per Service Limit :  8191
```

Conclusion

In addition to a full RFC 7432 EVPN-MPLS implementation, SR OS supports PBB-EVPN as per RFC 7623 for large layer 2 deployments, including single-active and all-active multi-homing. This example has shown how to configure and operate a PBB-EVPN network focusing on the specific aspects of PBB-EVPN compared to EVPN-MPLS.

EVPN for VXLAN Tunnels (Layer 2)

This chapter provides information about Ethernet virtual private network (EVPN) for virtual extensible local area network (VXLAN) tunnels in VPLS services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for release 12.0.R4. The CLI in the current edition is based on release 14.0.R3. Virtual eXtensible Local Area Network (VXLAN) requires IOM3-XP/IMM or higher-based line cards and chassis-mode D. Ethernet Virtual Private Networks (EVPN) is a control plane technology and does not have line card hardware dependencies.

Overview

SR OS supports the EVPN control plane with Virtual eXtensible Local Area Network (VXLAN) data plane in VPLS services.

EVPN is an IETF technology (draft-ietf-l2vpn-evpn) that uses a dedicated BGP address family which allows VPLS services to be operated in a similar way to IP-VPNs, where the MAC addresses, IP addresses and the information to set up the flooding tree are distributed by BGP. EVPN can be used as the control plane for different data plane encapsulations, such as VXLAN and MPLS.

VXLAN (draft-mahalingam-dutt-dcops-vxlan) is an overlay IP tunneling technology used to carry Ethernet traffic over any IP network, and it is becoming the de facto standard for overlay data centers and networks. Compared to other IP overlay tunneling technologies, such as GRE, VXLAN supports multi-tenancy and multi-pathing:

- A tenant identifier, the VXLAN Network Identifier (VNI), is encoded in the VXLAN header and allows each tenant to have an isolated Layer 2 domain.

- VXLAN supports multi-pathing scalability through ECMP. VXLAN uses the outer source UDP port as an entropy field that can be used by the core IP routers to balance the load across different paths.

In SR OS, EVPN and VXLAN can be enabled in VPLS or R-VPLS services. In this chapter, EVPN-VXLAN services will refer to VPLS or R-VPLS services with EVPN and VXLAN enabled. These services can terminate/originate VXLAN tunnels and may have SAPs and/or SDP bindings at the same time. Some other SR OS implementation-specific considerations are the following:

- VXLAN is only supported on network or hybrid ports with null or dot1q encapsulation on Ethernet/LAG/POS/APS interfaces.
- VXLAN packets are originated/terminated with the system IPv4 address, in other words, a system originating VXLAN packets will use the system IP address as source outer IPv4 address and systems will only process VXLAN packets if their destination outer IPv4 address matches its own system IP address.
- Data plane MAC learning is not supported over VXLAN bindings. Only the control plane (EVPN) will be used for populating the FDB with MAC addresses associated to VXLAN bindings.
- EVPN provides support for the following features that are described in this document:
  - The BGP advertisement of the MAC addresses learned on SAPs, SDP-bindings and conditional static MACs to the remote BGP peers. The advertisement of MAC addresses in BGP can optionally be disabled.
  - The optional advertisement of an unknown MAC route, that allows the remote EVPN PEs or Network Virtualization Edge devices (NVEs) to suppress the unknown unicast flooding and send any unknown unicast frame to the owner of the unknown MAC route.
  - Ingress replication of Broadcast, Unknown unicast and Multicast (BUM) packets over VXLAN.
  - A Proxy-ARP table per service populated by the MAC-IP pairs received in BGP MAC advertisements. When an ARP request is received on a SAP or SDP-binding, the system will perform a lookup on this table and will reply to the ARP request if the lookup yields a valid result.
  - MAC mobility and static-mac protection as described in draft-ietf-l2vpn-evpn, as well as MAC duplication detection.
- Multi-homing redundancy for SAPs and SDP-bindings in EVPN-VXLAN services is supported through BGP Multi-homing (L2VPN BGP address family). Only one BGP-MH site is supported in an EVPN-VXLAN service.

One of the main applications for EVPN-VXLAN services in SR OS is the data center gateway (DC GW) function. In such an application, EVPN and VXLAN are expected to be used within the data center and VPLS SDP-bindings or SAPs are expected to be used for the connectivity to the WAN. When the system is used as a DC GW, a VPLS service is configured per Layer 2 domain that has to be extended to the WAN. In those VPLS services, BGP EVPN automatically sets up the VXLAN auto-bindings that connect the DC GW to the data center network virtual edge devices (NVEs). The WAN connectivity is based on regular VPLS constructs where SAPs (null, dot1q and QinQ), spoke-SDPs (FEC type 128 and 129, not BGP-VPLS) and mesh-SDPs are supported. B-VPLS or I-VPLS services are not supported.

Although the DC GW application is one of the most common uses for this feature, this chapter focuses on the configuration and operation of EVPN-VXLAN for Layer 2 services in general, and its integration with regular VPLS services in MPLS networks.

Configuration

This section describes the configuration of EVPN-VXLAN on SR OS as well as the available troubleshooting and show commands. This example focuses on the following configuration aspects:

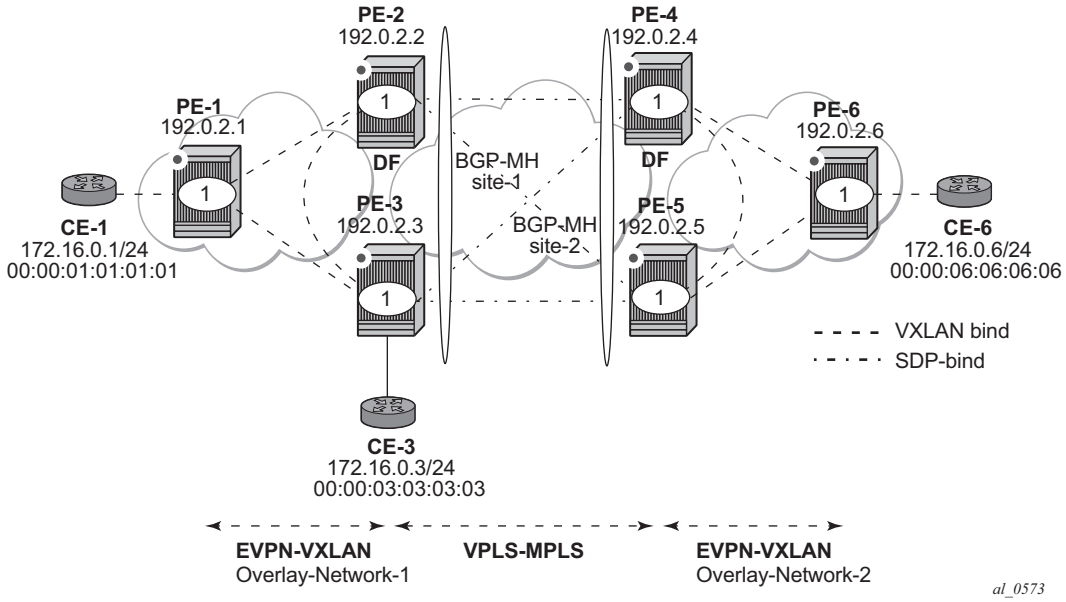
- Enabling EVPN and VXLAN in a VPLS service, including the use of BGP-EVPN, BGP-AD (BGP Auto-discovery) and BGP-MH (BGP Multi-homing) in the same VPLS instance.
- Scaling BGP-MH resiliency with the use of operational groups (oper-groups).
- Use of proxy-arp in EVPN-VXLAN services
- MAC mobility, MAC duplication and MAC protection in EVPN-VXLAN services.

The configuration will be shown for PE-1, PE-2 and PE-3 only; the PEs in Overlay-Network-2 (Figure 267) have an equivalent configuration.

Enabling EVPN-VXLAN in a VPLS Service

Figure 267 shows the topology used in this example.

Figure 267 **EVPN-VXLAN Example Topology**



The example topology shows two overlay (VXLAN) networks interconnected by an MPLS network:

- PE-1, PE-2 and PE-3 are part of Overlay-Network-1
- PE-4, PE-5 and PE-6 are part of Overlay-Network-2

CE-1, CE-3, and CE-6 belong to the same IP subnet, therefore, Layer 2 connectivity must be provided to them.

The example topology can illustrate a data center interconnect (DCI) example, where Overlay-Network-1 and Overlay-Network-2 are two Data Centers interconnected through an MPLS WAN. In this application, CE-1, CE-3, and CE-6 would simulate virtual machines or appliances, PE-2/3/4/5 would act as DC GWs and PE-1/6 as NVEs (or virtual PEs running on compute infrastructure).

The following protocols and objects are configured beforehand:

- The ports interconnecting the six PEs in Figure 267 are configured as network ports (or hybrid) and have router network interfaces defined on them. Only the ports connected to the CEs are configured as access ports.
- The six PEs shown in the Figure 267 are running IS-IS for the global routing table with the four core PEs interconnected using IS-IS Level-2 point-to-point interfaces and each overlay network is using IS-IS Level-1 point-to-point interfaces.

- LDP is used as the MPLS protocol to signal transport tunnel labels among PE-2, PE-3, PE-4 and PE-5. There is no LDP running in the two overlay networks.
- The network port MTU (in all the ports sending/receiving VXLAN packets) must be at least 50-bytes (54 if dot1q encapsulation is used) greater than the service MTU in order to accommodate the size of the VXLAN header.

Once the IGP infrastructure and LDP are enabled in the core, BGP has to be configured. In this example, two BGP families have to be enabled: EVPN within each overlay-network for the exchange of MAC/IP addresses and setting up the flooding domains, and L2-VPN for the use of BGP-MH and BGP-AD in the VPLS-MPLS network.

As an example, the following CLI output shows the relevant BGP configuration of PE-1, which only needs the EVPN family. PE-6 would have a similar BGP configuration. The use of route reflectors (RRs) in these type of scenarios is common. Although this example does not use RRs, an EVPN RR could have been used in Overlay-Network-1 and Overlay-Network-2 and an L2-VPN RR could have been used in the core VPLS-MPLS network.

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

The BGP configuration of PE-2 and PE-3 follows (PE-4 and PE-5 have an equivalent configuration).

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            rapid-update l2-vpn evpn
```

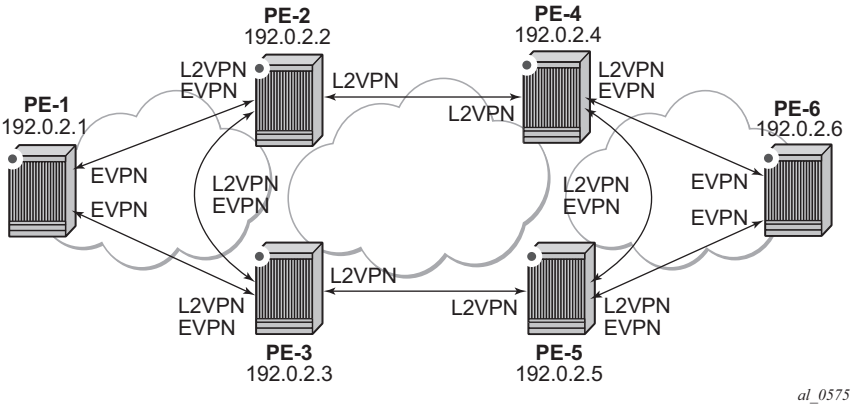
```
            group "DC"
                family l2-vpn evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                family l2-vpn
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

On PE-3:

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            rapid-update l2-vpn evpn
            group "DC"
                family l2-vpn evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
            group "WAN"
                family l2-vpn
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

Figure 268 shows the BGP peering sessions among the PEs and the enabled BGP families. PE-1 will only establish an EVPN peering session with its peers (only the EVPN family is enabled on PE-1), even though PE-2 and PE-3 have EVPN and L2-VPN families configured.

Figure 268 **BGP Adjacencies and Enabled Families**



Once the network infrastructure is running properly, the actual service configuration can be carried out. The following CLI outputs show the configuration of VPLS 1 in PE-1, PE-2 and PE-3 as per the topology illustrated in Figure 267.

VPLS 1 in those three PEs are interconnected using VXLAN bindings, whereas PE-2 and PE-3 are connected to the remote PEs by means of BGP-AD SDP-bindings. Although BGP-AD SDP-bindings are used in this example for the connectivity of the EVPN-VXLAN PEs to a regular VPLS network, SAPs, manual spoke-SDPs or mesh-SDPs could have been used instead. BGP-VPLS cannot be enabled in an EVPN-VXLAN VPLS service.

VPLS 1 configuration of PE-1 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 64500:11
                route-target export target:64500:1 import target:64500:1
            exit
            bgp-evpn
                vxlan
                    no shutdown
                exit
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
```

EVPN-VXLAN is enabled by the configuration of a valid VXLAN Network Identifier (VNI) and the **bgp-evpn>vxlan>no shutdown** command. These two commands, along with the required bgp route-distinguisher (RD) and route-target (RT) information, are the minimum mandatory attributes:

- The VNI is a 24-bit identifier with valid values in the [1..16777215] range. This defines the VNI that SR OS will use in the EVPN routes generated for the VPLS service, and therefore the VNI that the system expects to see in the VXLAN packets destined to that particular VPLS service. The configured VNI determines the VNI that has to be received in the packets for the VPLS service, but not the VNI that will be sent in VXLAN packets to remote PEs for the service. In other words, in this example, VPLS 1 is configured with VNI=1 in all the PEs, however each PE could have used a different VNI. The VNI is a system-wide significant value and two VPLS services cannot be configured with the same VNI.
- The **bgp-evpn>vxlan>no shutdown** command enables the use of EVPN for VXLAN. It requires the previous configuration of the VNI, RD and RT. As soon as this command is executed, EVPN will advertise an inclusive multicast route to all of the BGP EVPN peers (regardless of the existing SAP/SDP-binding operational status). The exchange of inclusive multicast routes allows the establishment of the VXLAN bindings among the PEs.

Upon the reception of the EVPN inclusive multicast routes from PE-2 and PE-3, PE-1 will automatically set up its VXLAN bindings for VPLS 1. A VXLAN binding is represented by an (egress VTEP, egress VNI) pair, where VTEP is a VXLAN Termination End Point. This can be shown with the following show commands:

```
*A:PE-1# show service id 1 vxlan
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1
Creation Origin: manual
RestProtSrcMacAct: none
===============================================================================
VPLS VXLAN service Network Specifics
===============================================================================
Ing Net QoS Policy : none                       Vxlan VNI Id     : 1
Ingress FP QGrp    : (none)                     Ing FP QGrp Inst : (none)
===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
192.0.2.2              1              0            BUM     Up           No
192.0.2.3              1              0            BUM     Up           No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#

*A:PE-1# show service vxlan
===============================================================================
VXLAN Tunnel Endpoints (VTEPs)
===============================================================================
VTEP Address                  Number of Egress VNIs    Oper State
-------------------------------------------------------------------------------
192.0.2.2                     1                        Up
```

```
192.0.2.3                     1                          Up
-------------------------------------------------------------------------------
Number of VTEPs: 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

To actually see this output, the VPLS service needs to be configured on all PEs, with import and export policy “vsi-policy-1” defined on the core PEs; see further. As can be seen in the CLI output, PE-1 has two VXLAN bindings, one to PE-2 and one to PE-3. Both use egress VNI=1 (the actual VNI used in its egress VXLAN packets) and both are part of the flooding multicast list (BUM) for VPLS 1 and are up. There is no layer 2 policy-based routing (L2 PBR).

- The **Mcast= BUM** entry is set when the proper inclusive multicast route is received from the remote VTEP. The VXLAN binding will be used to flood BUM (Broadcast, Unknown unicast, Multicast) packets.
- The **Oper State** is based on the existence of the VTEP in the global routing table.

The VPLS 1 configuration of PE-2 and PE-3 is as follows:

```
configure
    service
        pw-template 1 create
        exit
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.2:1
                vsi-export "vsi-policy-1"
                vsi-import "vsi-policy-1"
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
            bgp-ad
                vpls-id 64500:1
                no shutdown
            exit
            bgp-evpn
                vxlan
                    no shutdown
                exit
            exit
            site "site-1" create
                site-id 1
                split-horizon-group "CORE"
                no shutdown
            exit
            no shutdown
```

On PE-3:

```
configure
    service
        pw-template 1 create
        exit
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.3:1
                vsi-export "vsi-policy-1"
                vsi-import "vsi-policy-1"
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
            bgp-ad
                vpls-id 64500:1
                no shutdown
            exit
            bgp-evpn
                vxlan
                    no shutdown
                exit
            exit
            site "site-1" create
                site-id 1
                split-horizon-group "CORE"
                no shutdown
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
```

In addition to the VNI and **bgp-evpn>vxlan>no shutdown** commands for enabling EVPN-VXLAN in VPLS 1, PE-2 and PE-3 require the configuration of BGP-AD for the discovery and establishment of FEC129 spoke SDPs to the remote PEs in the core, as well as BGP-MH for redundancy. As outlined in Figure 267, there are two BGP-MH sites defined in the network: site-1 is used on PE-2/PE-3 and site-2 is used on PE-4/PE-5. Only one of the two gateway PEs in each overlay network will be the designated forwarder (DF) for VPLS 1, and only the DF will send/receive traffic for VPLS 1 in the overlay network. The following considerations must be taken into account when configuring the connectivity of EVPN-VXLAN services to regular VPLS objects:

- As discussed, in this example, BGP-AD spoke-SDPs are used, but SAPs, manual spoke-SDPs or mesh-SDPs are also supported.
- In this example, BGP-AD spoke-SDPs are auto-instantiated using **pw-template-binding 1 split-horizon-group “CORE”.**
  - This requires the creation of the pw-template 1 (**config>service>pw-template 1 create**).
- The split-horizon-group CORE is added to the BGP-MH site “site-1”. This statement will ensure that all the spoke SDPs automatically established to the remote PEs are part of the BGP-MH site.

- Although the route-targets for the Overlay-Network and the VPLS-MPLS network can have the same value for the same VPLS service, they are usually different. This example assumes the use of RT-DC-1 in Overlay-Network-1 and RT-WAN-1 in the VPLS-MPLS core for VPLS 1. The **vsi-policy-1** allows the system to export and import the right RTs for VPLS 1 on the core PEs:

```
configure
    router
        policy-options
            begin
            community "RT-DC-1" members "target:64500:12"
            community "RT-WAN-1" members "target:64500:11"
            policy-statement "vsi-policy-1"
                entry 10     # to import all the EVPN routes with RT-DC-1
                    from
                        community "RT-DC-1"
                        family evpn
                    exit
                    action accept
                    exit
                exit
                entry 20     # to import all the BGP-AD/MH routes from the WAN
                    from
                        community "RT-WAN-1"
                        family l2-vpn
                    exit
                    action accept
                    exit
                exit
                entry 30      # to export all the EVPN routes with "RT-DC-1"
                    from
                        family evpn
                    exit
                    action accept
                        community add "RT-DC-1"
                    exit
                exit
                entry 40      # to export all the BGP-AD/MH routes with "RT-WAN-1"
                    from
                        family l2-vpn
                    exit
                    action accept
                        community add "RT-WAN-1"
                    exit
                exit
                default-action reject
            exit
            commit
```

Once PE-2 and PE-3 are configured as shown, they will set up the spoke SDPs and will run the DF election algorithm to determine the operational status of those spoke SDPs. See LDP VPLS using BGP-Auto Discovery and BGP Multi-Homing for VPLS Networks for more information about the use of BGP-AD and BGP-MH.

In the configuration for VPLS 1, both gateway PEs, PE-2 and PE-3 will attempt to establish two parallel Layer 2 paths between each other (a BGP-AD spoke SDP and a EVPN VXLAN binding). Since that would create a Layer 2 loop, the SR OS implementation gives priority to the EVPN path and only the VXLAN binding will be active. In other words, when an (egress VTEP, VNI) and a spoke SDP are attempted to be set up to the same far-end IP address at the same time, the VXLAN path will prevail and the spoke SDP will be kept down. The spoke SDP will only be brought up if the VXLAN (egress VTEP, VNI) goes down.

This behavior can be easily observed in this setup by using the following show commands. In PE-2, the spoke SDP to far-end PE-3 will be down with a **EvpnRouteConflict** Flag. The (egress VTEP, VNI) = (192.0.2.3, 1) VXLAN bind will be up.

```
*A:PE-2# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
---snip---
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 1
SAP Count         : 0                   SDP Bind Count    : 3
---snip---
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:17405:4294967293 SB(192.0.2.3)       BgpAd        0       8978    Up   Down
sdp:17406:4294967294 SB(192.0.2.5)       BgpAd        0       8978    Up   Up
sdp:17407:4294967295 SB(192.0.2.4)       BgpAd        0       8978    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-2#


*A:PE-2# show service id 1 all | match Flag context all
Flags              : PWPeerFaultStatusBits
                     EvpnRouteConflict
Flags              : None
Flags              : None
*A:PE-2#


*A:PE-2# show service id 1 vxlan
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1
Creation Origin: manual
RestProtSrcMacAct: none

===============================================================================
VPLS VXLAN service Network Specifics
===============================================================================
Ing Net QoS Policy : none                           Vxlan VNI Id     : 1
```

```
Ingress FP QGrp    : (none)                          Ing FP QGrp Inst : (none)

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
192.0.2.1              1              0            BUM     Up           No
192.0.2.3              1              0            BUM     Up           No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-2#
```

At the non-DF, PE-3, all the spoke SDPs will be down due to BGP-MH:

```
*A:PE-3# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
---snip---
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 1
SAP Count         : 1                   SDP Bind Count    : 3
---snip---
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                             Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/2/1:1                            q-tag        1518    1518    Up   Up
sdp:17405:4294967293 SB(192.0.2.4)     BgpAd        0       8978    Up   Down
sdp:17406:4294967294 SB(192.0.2.5)     BgpAd        0       8978    Up   Down
sdp:17407:4294967295 SB(192.0.2.2)     BgpAd        0       8978    Up   Down
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#


*A:PE-3# show service id 1 all | match Flag context all
Flags              : StandbyForMHProtocol
Flags              : StandbyForMHProtocol
Flags              : StandbyForMHProtocol
                     PWPeerFaultStatusBits
                     EvpnRouteConflict
Flags              : None
```

MAC Learning and unknown-mac-route

Once the VPLS service (VPLS 1) is configured, the network allows the CEs to exchange unicast and BUM traffic over the overlay and VPLS-MPLS service infrastructure. BUM traffic sent by CE-1 will be ingress-replicated to PE-2 and PE-3 by PE-1, and propagated by PE-2 (the DF) to the remote network. From this point on, MAC addresses will be learned on active SAPs and spoke SDPs and advertised in EVPN MAC routes. No data plane MAC learning is carried out on VXLAN bindings. MACs associated with (egress VTEP, VNI) bindings will always be learned through EVPN.

The following CLI output shows the reception of an EVPN MAC route on PE-1 and how the (CE-3) MAC address appears in the FDB for VPLS 1.

```
4 2016/06/08 11:14:06.95 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:03:03:03:03, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
"


*A:PE-1# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:01:01:01:01 sap:1/2/1:1              L/0      06/08/16 11:29:07
1         00:00:03:03:03:03 vxlan:                   Evpn     06/08/16 11:13:02
                            192.0.2.3:1
1         00:00:06:06:06:06 vxlan:                   Evpn     06/08/16 11:24:05
                            192.0.2.2:1
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
*A:PE-1#
```

When a frame destined to 00:00:03:03:03:03 enters SAP 1/2/1:1, it is encapsulated into a VXLAN packet with outer destination IP 192.0.2.3 and VNI 1, and sent on the wire.

In virtualized data center networks where all the MACs are known beforehand (all the virtual machine and appliance MACs are distributed by EVPN before any traffic flows), unknown MAC addresses are always outside the data center. If that is the case, the DC GWs can make use of the **unknown-mac-route** so that the DC NVEs supporting the concept of this route send the unknown unicast traffic only to the DC GW. This minimizes the flooding within the Data Center, as explained in draft-rabadan-l2vpn-dci-evpn-overlay.

In this example, the unknown-mac-route is configured in the gateway PEs (PE-2, PE-3 and PE-4, PE-5) in the following way:

```
*A:PE-2# configure service vpls 1 bgp-evpn unknown-mac-route
22 2016/06/08 12:04:34.03 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:00:00:00:00, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
"
```

Note that:

- Although SROS can generate the unknown-mac-route, it will never honor it and normal flooding applies when an unknown unicast packet arrives at an ingress sap/sdp-binding.
- When unknown-mac-route is configured, it will only be generated when: a) no BGP-MH site is configured within the same VPLS service or b) a site is configured and the site is DF (Designated Forwarder) in the PE. If the site becomes a non-DF site, the unknown-mac-route will be withdrawn.
- If the unknown-mac-route is used in the DC GW and all the NVEs in the DC understand it, the advertisement of MAC addresses can be disabled with the [**no**] **mac-advertisement** command. If so, SR OS will only advertise the unknown-mac-route.

```
*A:PE-2# configure service vpls 1 bgp-evpn unknown-mac-route
*A:PE-2# configure service vpls 1 bgp-evpn no mac-advertisement
```

Scaling BGP-MH Resiliency with the Use of Operational Groups

In Figure 267, VPLS 1 in PE-2/PE-3 is configured with a BGP-MH site that controls which of the two PEs forwards the traffic to the remote PEs (in this case, PE-2 is the DF and the gateway responsible for forwarding packets to the remote PEs).

When new VPLS services are required in PE-2/PE-3, the same BGP-MH configuration can be used. However, if the number of VPLS services grows significantly, the use of individual BGP-MH sites per service will not scale. Since all the services in these two PEs share the same physical topology, the use of oper-groups can provide a simple and scalable way of providing resiliency to as many services as the user needs (up to the maximum number of VPLS services per system).

The way oper-groups can be used to scale this type of deployments is the following (using the network topology in Figure 267 and focusing on Overlay-Network-1):

- A control-VPLS service is defined in PE-2 and PE-3. For instance, VPLS 1.
  - This service is configured with a BGP-MH site in both PEs.
  - An oper-group **control-vpls-1** is created and associated to the pw-template-binding 1 in VPLS 1.
- Data VPLS services are defined in both PEs. For instance: VPLS 2, VPLS 3,... VPLS 999.
  - In all these services, the pw-template-binding is configured with **monitor-oper-group “control-vpls-1”.**
  - The status of the spoke SDPs in the data VPLS services depends on the status of the oper-group. If there is a DF switchover in VPLS 1 and VPLS 1 spoke SDPs go down on PE-2, all the spoke SDPs in all the data VPLS services controlled by **control-vpls-1** in PE-2 will go down too. In the same way, the spoke SDPs in PE-3 will come up.
- To allow per-service load balancing a second control-VPLS service with a different BGP-MH site should be configured.
  - For instance, VPLS 1 might have PE-2 as the DF and VPLS 1000 might be a second control-VPLS service with PE-3 as the DF.
  - Each control-VPLS would control a group of data VPLS services based on the definition and association of a second oper-group.

The following example shows the modification of VPLS 1 as the control-VPLS and the configuration of VPLS 2 as a data-VPLS on PE-2. VPLS 1 controls the VPLS 2 spoke SDP status.

```
configure
```

```
service
    oper-group "control-vpls-1" create
    exit
    vpls 1 customer 1 create
        description "control-VPLS"
        bgp
            pw-template-binding 1 split-horizon-group "CORE"
                oper-group "control-vpls-1"
            exit
        exit
    exit
    vpls 2 customer 1 create
        description "data-VPLS"
        vxlan vni 2 create
        exit
        bgp
            route-distinguisher 192.0.2.2:2
            vsi-export "vsi-policy-2"
            vsi-import "vsi-policy-2"
            pw-template-binding 1
                monitor-oper-group "control-vpls-1"
            exit
        exit
        bgp-ad
            vpls-id 64500:2
            no shutdown
        exit
        bgp-evpn
            unknown-mac-route
            vxlan
                no shutdown
            exit
        exit
        no shutdown
    exit
```

Use of Proxy-ARP in EVPN-VXLAN Services

EVPN-VXLAN services support proxy-ARP functionality that is enabled by the **proxy-arp [no] shutdown** command. The default value is shutdown. When proxy-arp is enabled:

- MAC and IP addresses contained in the received valid EVPN MAC routes are populated in the proxy-ARP table.
- ARP-request messages received on SAPs and SDP-bindings are intercepted and the target IP address is looked up. If the IP address is found, an ARP reply will be issued based on the information found in the proxy-ARP table, otherwise the ARP request would be flooded in the VPLS service (except for the source SAP/SDP binding).

- ARP-reply messages received on SAPs and SDP-bindings are also intercepted and sent to the CPM. These ARP-reply messages are re-injected in the data plane and forwarded based on the FDB information to the destination MAC address. If the destination MAC address is not in the FDB, the ARP-reply message will be flooded in the VPLS service (except for the source SAP/SDP binding).

The following CLI output shows the proxy-ARP configuration in PE-3 and a received valid MAC route that includes the MAC and IP of CE-1. This MAC-IP pair is installed in the proxy-ARP table for VPLS 1.

```
configure service vpls 1 proxy-arp no shutdown


27  2016/06/08 13:50:30.49 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 92
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 37 RD: 192.0.2.1:1 ESI: ESI-0, tag: 0, mac len: 48
              mac: 00:00:01:01:01:01, IP len: 4, IP: 172.16.0.1, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
"



*A:PE-3# show service id 1 proxy-arp detail
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : disabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 1
Static Count      : 0                   EVPN Count        : 1
Dynamic Count     : 0                   Duplicate Count   : 0

Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None

EVPN
-------------------------------------------------------------------------------
Garp Flood        : enabled             Req Flood         : enabled
Static Black Hole : disabled
-------------------------------------------------------------------------------
```

```
===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.1          00:00:01:01:01:01   evpn      active    06/13/2016 12:28:30
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

SR OS does not include a host IP address in any EVPN MAC advertisement for a MAC learned on a SAP or SDP-bind. Host IP addresses are only included in the EVPN MAC advertisements corresponding to R-VPLS IP interfaces. When deployed as DC GW in a Nuage architecture, the Nuage Networks VSC (Virtual Services Controller) or VSG (Virtual Services Gateway) will send virtual machine and host MAC/IP pairs in EVPN MAC routes. See the Nokia Nuage documentation for more information about the Nuage DC architecture. The 7x50 DC GW will populate the proxy-ARP tables with those MAC/IP pairs.

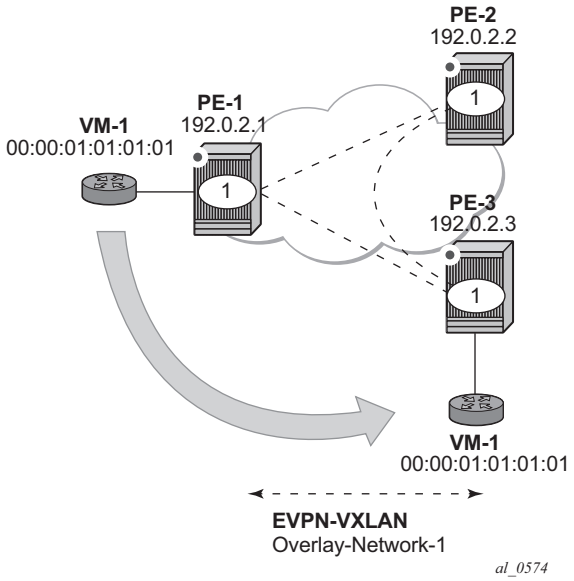
In the CLI excerpt above, assume that PE-1 is replaced by a Nuage VSC that sends the pair <172.16.0.1, 00:00:01:01:01:01> in an EVPN MAC route. PE-3 receives the advertisement and adds the entry to its proxy-ARP table for VPLS 1.

The proxy-arp feature was significantly improved in SR OS release 13.0; see chapter EVPN for MPLS Tunnels.

MAC Mobility, MAC Duplication, and MAC Protection in EVPN

MAC mobility, duplication and protection are fully supported as specified in draft-ietf-l2vpn-evpn. Figure 269 illustrates the concept of mobility (Virtual Machine VM-1 moves from PE-1 to PE-3).

Figure 269 **EVPN MAC Mobility**



MAC mobility is handled in EVPN by the use of sequence numbers in the MAC routes. When 00:00:01:01:01:01 moves from PE-1 to PE-3, SR OS will gracefully handle it in this way:

- 00:00:01:01:01:01 moves to PE-3 SAP 1/2/1:1
- PE-3 advertises 00:00:01:01:01:01 using a higher sequence number (the first time a MAC is advertised, EVPN uses sequence number 0).
- PE-2 at this point has two valid MAC routes for 00:00:01:01:01:01. It picks up the one coming from PE-3 since the sequence number is higher.
- PE-1 receives the MAC route, and since the sequence number is higher than the one for its own route, it updates the FDB and withdraws its own MAC route.

However, if MAC 00:00:01:01:01:01 is constantly learned on the PE-1 and PE-3 SAPs, the preceding process causes an endless exchange of MAC route advertisements and withdraws that has a negative impact on all the PEs in the EVPN network. This issue is known as “MAC duplication” and is originated by a loop at the access or a duplicated MAC address in two hosts of the same service. SR OS solves this issue through the use of the mac-duplication detection feature. MAC-duplication is always enabled with the following default settings:

```
*A:PE-1>config>service>vpls>bgp-evpn# info detail | match mac-duplication context all
-----------------------------------------------
                mac-duplication
                    detect num-moves 5 window 3
                    retry 9
```

Where:

- **num-moves** — Identifies the number of MAC moves in a VPLS service. The counter is incremented when a MAC is locally relearned in the FDB or flushed from the FDB due to the reception of a better remote EVPN route for that MAC. When the threshold is reached for a given MAC, this MAC is put in hold-down state (this 'hold-down' state is described below). Range: <3..10>. Default value: 5.
- **window** — Identifies the timer within which a MAC is considered duplicate if it reaches the configured num-moves. Range: <1..15> minutes. Default value: 3 minutes.
- **Retry —** The timer after which the MAC in hold-down state is automatically flushed and the mac-duplication process starts again. This value is expected to be equal to two times or more than the window. If no retry is configured, this implies that, once mac-duplication is detected, MAC updates for that MAC will be held down until the user intervenes or a network event (that flushes the MAC) occurs. Range: <2..60> minutes. Default value: 9 minutes.

When a MAC is considered a duplicate or in the 'hold-down' state, no further BGP advertisements are issued for this MAC and an alarm is triggered (by the first MAC in hold-down state). The following CLI output shows how PE-3 detects that MAC 00:00:01:01:01:01 is a duplicate (after reaching the **num-moves** in **window**) and the corresponding alarm.

```
142 2016/06/09 09:34:28.59 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 96
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:01:01:01:01, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
        mac-mobility:Seq:5
"


2 2016/06/09 09:34:28.59 UTC MINOR: SVCMGR #2331 Base
"VPLS Service 1 has MAC(s) detected as duplicates by EVPN mac-duplication detect
ion."
```

The **show service id bgp-evpn** command shows the mac-duplication settings and the list of duplicate MACs on hold-down.

```
*A:PE-3# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled            Unknown MAC Route  : Enabled
CFM MAC Advertise  : Disabled
VXLAN Admin Status : Enabled            Creation Origin    : manual
MAC Dup Detn Moves : 3                  MAC Dup Detn Window: 1
MAC Dup Detn Retry : 2                  Number of Dup MACs : 2
IP Route Advert    : Disabled

EVI                : n/a
Ing Rep Inc McastAd: Enabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
00:00:01:01:01:01                            06/09/2016 09:40:29
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The 7x50 stops sending and processing any BGP MAC advertisement routes for that MAC address until:

- The MAC is flushed due to a local event (SAP/SDP-binding associated to the MAC fails) or the reception of a remote withdraw for the MAC (due to a MAC flush at the remote 7x50) or
- The **retry <in_minutes>** timer expires, which flushes the MAC and restart the process.

When the last duplicate MAC address is removed from the duplicate list, the system will show the following message:

```
*A:PE-3#
3 2016/06/09 09:44:28.60 UTC MINOR: SVCMGR #2332 Base
"VPLS Service 1 no longer has MAC(s) detected as duplicates by EVPN mac-duplicat
ion detection."
```

EVPN also provides a mechanism to protect certain MACs that do not move for which connectivity must be guaranteed. These addresses must be protected in case there is an attempt to dynamically learn them in a different place in the EVPN-VXLAN VPLS service (on the same or different PE).

The protected MACs are configured in SR OS as conditional static MACs. A conditional static MAC defined in an EVPN-VXLAN VPLS service is advertised by BGP-EVPN as a static address. An example of the configuration of a conditional static MAC is as follows:

```
configure
    service
```

```
        vpls 1
            static-mac
                mac 00:00:05:05:05:05 create sap 1/2/1:1 monitor fwd-status
            exit
        exit
```

The protected MACs advertised in EVPN are shown in the receiving BGP RIB as Static (MAC mobility extended community with Sequence 0 and sticky bit set) and **EvpnS:P** (Evpn Static: Protected) in the FDB. The advertising PE shows the protected MAC as **CStatic:P** (Conditional Static: Protected) in the FDB:

On the advertising PE:

```
*A:PE-1# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:05:05:05:05 sap:1/2/1:1              CStatic: 06/09/16 10:06:08
                                                     P
1         4a:a3:ff:00:03:3a vxlan:                   EvpnS:P  06/09/16 09:25:49
                            192.0.2.3:1
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

On the receiving PE:

```
*A:PE-3# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:00:05:05:05:05 vxlan:                   EvpnS:P  06/09/16 10:06:34
                            192.0.2.1:1
-------------------------------------------------------------------------------
No. of MAC Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================


*A:PE-3# show router bgp routes evpn mac mac-address 00:00:05:05:05:05 hunt
===============================================================================
 BGP Router ID:192.0.2.3       AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
```

```
===============================================================================
BGP EVPN MAC Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : N/A
Nexthop        : 192.0.2.1
From           : 192.0.2.1
Res. Nexthop   : 192.168.16.1
Local Pref.    : 100                    Interface Name : int-PE-3-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:64500:12 bgp-tunnel-encap:VXLAN
                 mac-mobility:Seq:0/Static
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : N/A
Route Dist.    : 192.0.2.1:1
Mac Address    : 00:00:05:05:05:05
MPLS Label1    : VNI 1                  MPLS Label2    : N/A
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 02h34m57s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-3#
```

The following procedures are supported in order to protect the configured static MAC addresses:

- All the SAP/SDP-bindings are internally configured as MAC protect restrict-protected-src as soon as bgp-evpn is enabled in the VPLS service.
- Local static MACs or remote EVPN static MACs are considered as protected.
- If a frame with a source MAC address matching one of the protected MACs is received on a different SAP/SDP-binding than the owner of the protected MAC, the frame is discarded and an alarm triggered. This MAC protection is not performed for frames received on VXLAN bindings.

- The same throttled alarm mechanism used in MAC protect for restrict-protected-src with discard-frame is used here: the offending frames are captured to a list to be polled by the CPM every ~10min.

In this example, PE-3 has 00:00:05:05:05:05 in its FDB as EvpnS. If SAP 1/2/1:1 receives a frame with source MAC address 00:00:05:05:05:05, the frame is discarded and an alarm triggered:

```
*A:PE-3#
4 2016/06/09 14:05:51.96 UTC MINOR: SVCMGR #2208 Base Slot 1
"Protected MAC 00:00:05:05:05:05 received on SAP 1/2/1:1 in service 1. "
```

Debug and Show Commands

In addition to the previously mentioned **show service id vxlan, show service id bgp-evpn and show service id fdb detail** commands, the following commands provide valuable information when troubleshooting an EVPN-VXLAN VPLS service.

The **show router bgp routes evpn** command supports filtering by route type as well as many other route fields.

```
*A:PE-3# show router bgp routes evpn
  - evpn <evpn-type>

      auto-disc       - Display BGP EVPN Auto-Disc Routes
       eth-seg         - Display BGP EVPN Eth-Seg Routes
       inclusive-mcast - Display BGP EVPN Inclusive-Mcast Routes
       ip-prefix       - Display BGP EVPN IPv4-Prefix Routes
       ipv6-prefix     - Display BGP EVPN IPv6-Prefix Routes
       mac             - Display BGP EVPN Mac Routes


*A:PE-3# show router bgp routes evpn mac
 - mac [hunt|detail] [rd <rd>] [next-hop <ip-address>] [mac-address <mac-address>]
       [community <comm-id>] [tag <tag>] [aspath-regex <reg-exp>]

 <hunt|detail>        : keywords
 <rd>                 : {<ip-addr:comm-val>|
                        <2byte-asnumber:ext-comm-val>|
                        <4byte-asnumber:comm-val>}
 <ip-address>         : a.b.c.d
 <mac-address>        : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx
 <comm-id>            : <as-number1:comm-val1>|<ext-comm>|
                        <well-known-comm>
                        ext-comm        - <type>:{<ip-address:comm-val1>|
                                                 <as-number1:comm-val2>|
                                                 <as-number2:comm-val1>}
                        as-number1      - [0..65535]
                        comm-val1       - [0..65535]
                        type            - target|origin
                        ip-address      - a.b.c.d
                        comm-val2       - [0..4294967295]
```

```
                        as-number2      - [0..4294967295]
                        well-known-comm - null|no-export|no-export-subconfed|
                                          no-advertise
 <tag>                : [0..16777215] | MAX-ET
 <reg-exp>            : [80 chars max]

*A:PE-3# show router bgp routes evpn mac tag 0
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.          MacAddr           ESI
      Tag                  Mac Mobility      Ip Address
                                             NextHop
                                             Label1
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1          00:00:05:05:05:05 ESI-0
      0                    Static            N/A
                                             192.0.2.1
                                             VNI 1

u*>i  192.0.2.2:1          00:00:00:00:00:00 ESI-0
      0                    Seq:0             N/A
                                             192.0.2.2
                                             VNI 1

u*>i  192.0.2.2:2          00:00:00:00:00:00 ESI-0
      0                    Seq:0             N/A
                                             192.0.2.2
                                             VNI 2

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-3#
```

The **tools dump service id vxlan** displays the number of times a service could not add a VXLAN binding or <VTEP, Egress VNI> due to the following limits:

- The per System VTEP limit has been reached
- The per System (egress VTEP, egress VNI) limit has been reached
- The per Service (egress VTEP, egress VNI) limit has been reached
- The per System Bind limit: Total bind limit or VXLAN bind limit has been reached.

Tools dump service evpn usage displays the consumed resources in the system, whereas **tools dump service vxlan dup-vtep-egrvni** displays the (egress VTEP, egress VNI) bindings that have been detected as duplicate attempts, in other words, an attempt to add the same binding to more than one service:

```
*A:PE-3# tools dump service id 1 vxlan

VTEP, Egress VNI Failure statistics at 001 01:06:55.950:

statistics last cleared at 000 00:00:00.000:

Failures: None


*A:PE-3# tools dump service id 1 evpn usage

Evpn Tunnel Interface IP Next Hop: N/A


*A:PE-3# tools dump service evpn usage

EVPN usage statistics at 000 04:45:44.530:

MPLS-TEP                                          :              0
VXLAN-TEP                                         :              2
Total-TEP                                         :      2/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)     :              0
Vxlan Dests (TEP, Egress VNI)                     :              3
Total-Dest                                        :      3/196607

Sdp Bind +  Evpn Dests                            :      9/245759
ES L2/L3 PBR                                      :      0/ 32767


*A:PE-3# tools dump service vxlan dup-vtep-egrvni

Duplicate VTEP, Egress VNI usage attempts at 001 01:08:04.080:

1. 192.0.2.1:100
```

Conclusion

SR OS supports the EVPN control plane for VXLAN tunnels terminated in VPLS services. VXLAN is an overlay IP tunneling mechanism that is being used in data center, data center interconnect and other applications. EVPN is a scalable and flexible control plane that provides control over the MACs being learned and advertised, as well as other mechanisms to optimize Layer 2 services such as proxy-ARP, MAC mobility, MAC duplication detection and MAC protection. SR OS provides a resilient and scalable EVPN-VXLAN solution for Layer 2 services, including interoperability to existing VPLS networks. This chapter showed all of those functions and how they are configured and operated.

EVPN for VXLAN Tunnels (Layer 3)

This chapter provides information about EVPN for VXLAN tunnels (Layer 3).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for release 12.0.R4. The CLI in the current edition is based on release 14.0.R3. Virtual eXtensible Local Area Network (VXLAN) requires IOM3-XP/IMM or higher-based line cards and chassis-mode D. Ethernet Virtual Private Networks (EVPN) is a control plane technology and does not have line card hardware dependencies.

The chapter EVPN for VXLAN Tunnels (Layer 2) example is prerequisite reading.

Overview

As discussed in the EVPN for VXLAN Tunnels (Layer 2) chapter, EVPN and VXLAN can be enabled on VPLS or R-VPLS services in SR OS. While that chapter focuses on the use of EVPN-VXLAN layer 2 services, in other words, how EVPN-VXLAN is configured in VPLS services, this chapter describes how EVPN-VXLAN can be used to provide inter-subnet forwarding in R-VPLS and VPRN services. Inter-subnet forwarding can be provided by regular R-VPLS and VPRN services, however EVPN provides an efficient and unified way to populate forwarding databases (FDBs), address resolution protocol (ARP) tables and routing tables using a single BGP address family. Inter-subnet forwarding in overlay networks would otherwise require data plane learning and the use of routing protocols on a per VPRN basis.

The SR OS solution for inter-subnet forwarding using EVPN is based on building blocks described in draft-sajassi-l2vpn-evpn-inter-subnet-forwarding and the use of the EVPN ip-prefix routes (route type-5) as explained in draft-rabadan-l2vpn-evpn-prefix-advertisement. This example describes three supported common scenarios and provides the CLI configuration and required tools to troubleshoot EVPN-VXLAN in each case. The scenarios configured and explained are:

- EVPN-VXLAN in R-VPLS services
- EVPN-VXLAN in integrated routing bridging (IRB) backhaul R-VPLS services
- EVPN-VXLAN in EVPN tunnel R-VPLS services

In all these scenarios, redundant PEs are usually deployed. If that is the case, the interaction of EVPN, IP-VPN and the routing table manager (RTM) may lead to some routing loop situations that must be avoided by the use of routing policies (this also may happen in traditional IP-VPN deployments when eBGP and MP-BGP interact to populate VPRN routing tables in multi-homed networks). This chapter explains when those routing loops can happen and how to avoid them.

The term IRB interface refers to an R-VPLS service bound to a VPRN IP interface. The terms IRB interface and R-VPLS interface are used interchangeably throughout this example.

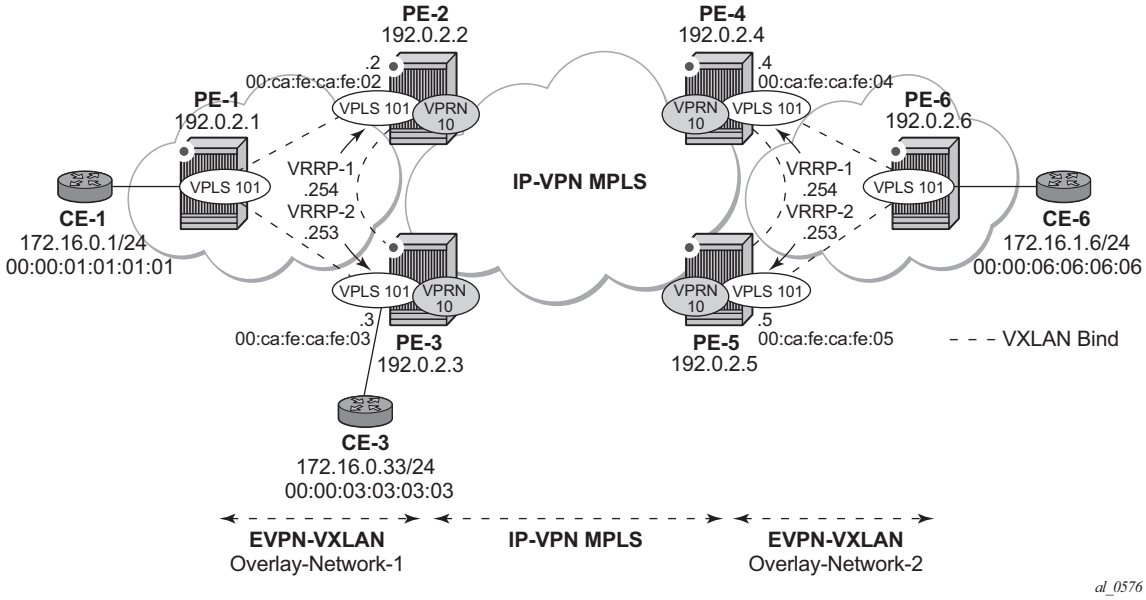
Configuration

This section describes the configuration of EVPN-VXLAN for Layer 3 services on SR OS, as well as the available troubleshooting and show commands. The three scenarios described in the overview are analyzed independently.

EVPN-VXLAN in an R-VPLS Service

Figure 270 shows the topology used in the first scenario.

Figure 270 **EVPN-VXLAN for R-VPLS Services**



The network topology shows two overlay (VXLAN) networks interconnected by an MPLS network:

- PE-1, PE-2 and PE-3 are part of Overlay-Network-1
- PE-4, PE-5 and PE-6 are part of Overlay-Network-2

A Layer 2/Layer 3 service is provided to a customer to connect CE-1, CE-3 and CE-6. In this scenario, Layer 2 connectivity is provided within each overlay network and inter-subnet connectivity (Layer 3) is provided between the overlay networks, therefore VPLS 101 is defined within each overlay network and VPRN 10 connects both Layer 2 services through an IP-VPN MPLS network.

This topology can illustrate a Data Center Interconnect (DCI) example, where Overlay-Network-1 and Overlay-Network-2 are two data centers interconnected through an MPLS WAN. In this application, CE-1, CE-3 and CE-6 would simulate virtual machines or appliances, PE-2/3/4/5 would act as Data Center gateways (DC GW) GWs and PE-1/6 as Network Virtualization Edge devices (or virtual PEs running on a compute infrastructure).

The following protocols and objects are configured beforehand:

- The ports interconnecting the six PEs in Figure 270 are configured as network ports (or hybrid) and will have router network interfaces defined in them. Only the ports connected to the CEs are configured as access ports.

- The six PEs are running IS-IS for the global routing table with the four core PEs interconnected using IS-IS Level-2 point-to-point interfaces and each overlay network using IS-IS Level-1 point-to-point interfaces.
- LDP is used as the MPLS protocol to signal transport tunnel labels among PE-2, PE-3, PE-4 and PE-5. There is no LDP running within each overlay network.
- The network port MTU (in all the ports sending/receiving VXLAN packets) must be at least 50-bytes (54 if dot1q encapsulation is used) greater than the service MTUin order to accommodate the size of the VXLAN header.

Once the IGP infrastructure and LDP in the core are enabled, BGP has to be configured. In this scenario, two BGP families have to be enabled: EVPN within each overlay-network for the exchange of MAC/IP addresses and setting up the flooding domains, and VPN-IPv4 among the four core PEs so that IP-prefixes can be exchanged and resolved to MPLS tunnels in the core.

As an example, the following CLI output shows the relevant BGP configuration of PE-1, which only needs the EVPN family. PE-6 has a similar BGP configuration, that is, only EVPN family is configured for its peers. The use of Route-Reflectors (RRs) in these type of scenarios is common. Although this scenario does not use RRs, an EVPN RR could have been used in Overlay-Network-1 and Overlay-Network-2 and a separate VPN-IPv4 RR could have been used in the core IP-VPN MPLS network.

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

The BGP configuration of PE-2 and PE-3 follows (PE-4 and PE-5 have an equivalent configuration).

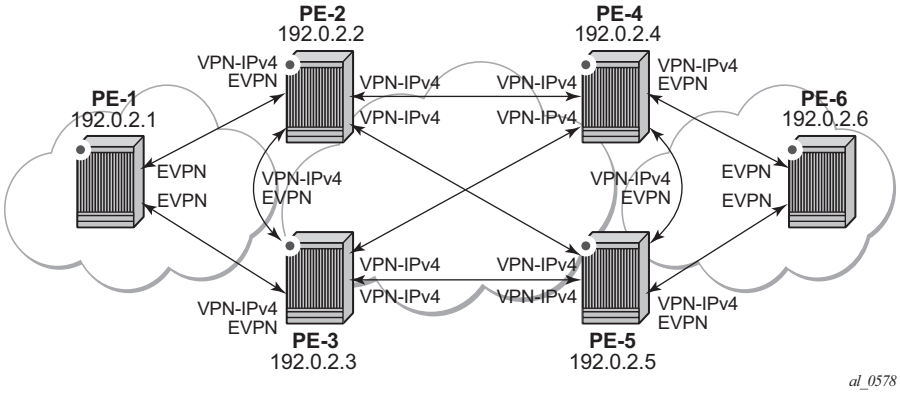
```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
```

```
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family vpn-ipv4 evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                family vpn-ipv4
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit

configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family vpn-ipv4 evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
            group "WAN"
                family vpn-ipv4
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

Figure 271 shows the BGP peering sessions among the PEs and the enabled BGP families. PE-1 and PE-6 only establish an EVPN peering session with their peers (only the EVPN family is enabled on both PEs, even if the peer PEs are VPN-IPv4 capable as well).

Figure 271 **BGP adjacencies and enabled families**



Once the network infrastructure is running properly, the actual service configuration, as illustrated in Figure 270, can be carried out. The following CLI shows the configuration for VPLS 101 and VPRN 10 in PE-1, PE-2 and PE-3. The other overlay network has a similar configuration.

On PE-1:

```
configure
    service
        vpls 101 customer 1 create
            vxlan vni 101 create
            exit
            bgp
                route-distinguisher 192.0.2.1:101
                route-target export target:64500:101 import target:64500:101
            exit
            bgp-evpn
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-101"
            sap 1/2/1:101 create
            exit
            proxy-arp
                no shutdown
            exit
            no shutdown
```

Proxy ARP is disabled (default) on PE-2, as well as on the other core PEs:

```
configure
    service
        vpls 101 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 101 create
            exit
```

```
            bgp
                route-distinguisher 192.0.2.2:101
                route-target export target:64500:101 import target:64500:101
            exit
            bgp-evpn
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-101"
            no shutdown
        exit
        vprn 10 customer 1 create
            ecmp 2
            route-distinguisher 192.0.2.2:10
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:10
            interface "int-1" create
                address 172.16.0.2/24
                mac 00:ca:fe:ca:fe:02
                vrrp 1
                    backup 172.16.0.254
                    priority 254
                    ping-reply
                    traceroute-reply
                    mac 00:ca:fe:ca:fe:54
                exit
                vrrp 2
                    backup 172.16.0.253
                    ping-reply
                    traceroute-reply
                    mac 00:ca:fe:ca:fe:53
                exit
                vpls "evi-101"
                exit
            exit
            no shutdown
```

On PE-3:

```
configure
    service
        vpls 101 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 101 create
            exit
            bgp
                route-distinguisher 192.0.2.3:101
                route-target export target:64500:101 import target:64500:101
            exit
            bgp-evpn
                vxlan
```

```
            no shutdown
        exit
    exit
    service-name "evi-101"
    no shutdown
exit
vprn 10 customer 1 create
    ecmp 2
    route-distinguisher 192.0.2.3:10
    auto-bind-tunnel
        resolution-filter
            ldp
        exit
        resolution filter
    exit
    vrf-target target:64500:10
    interface "int-1" create
        address 172.16.0.3/24
        mac 00:ca:fe:ca:fe:03
        vrrp 1
            backup 172.16.0.254
            ping-reply
            traceroute-reply
            mac 00:ca:fe:ca:fe:54
        exit
        vrrp 2
            backup 172.16.0.253
            priority 254
            ping-reply
            traceroute-reply
            mac 00:ca:fe:ca:fe:53
        exit
        vpls "evi-101"
        exit
    exit
    no shutdown
```

For details about the EVPN and VXLAN configuration on PE-1 VPLS 101, refer to EVPN for VXLAN Tunnels (Layer 2). The configuration of VPLS 101 on PE-2 and PE-3 has the following important aspects:

- The **allow-ip-int-bind** command is required so that the R-VPLS can be bound to VPRN 10.
- The **service-name** command is required and the configured name must match the name configured in the VPRN 10 VPLS interface.
- Even though EVPN and VXLAN are properly configured, **proxy-arp** cannot be enabled in VPLS 101. In an R-VPLS with EVPN-VXLAN, proxy-arp is not supported and the VPRN ARP table is used instead. When an EVPN MAC route that includes an IP address is received in an R-VPLS, the MAC-IP pair encoded in the route is added to the VPRN's ARP table, as opposed to the proxy-arp table.

```
*A:PE-2# configure service vpls 101 proxy-arp no shutdown
MINOR: SVCMGR #8007 Cannot modify proxy arp - service is routed
```

When configuring VPRN 10 on PE-2 and PE-3, the following considerations must be taken into account:

- When trying to enable existing VPRN features on interfaces linked to EVPN-VXLAN R-VPLS interfaces, the following commands are not supported:
  - arp-populate.
  - authentication-policy.

```
*A:PE-2# configure service vprn 10 interface "int-1" authentication-policy "authPol1"
INFO: PIP #1875 Cannot configure auth-policy on routed-vpls interface
```

- Dynamic routing protocols such as IS-IS, RIP, or OSPF are not supported.
- In general, no 7x50 control plane generated packets are sent to the egress VXLAN bindings except for ARP, VRRP, ICMP, BFD, and Eth-CFM.
- As depicted in Figure 270 and shown in the CLI excerpts, VRRP can be configured on the VPRN 10 VPLS interfaces to provide default gateway redundancy to the hosts connected to VPLS 101. Two VRRP instances are configured so that VPLS 101 upstream traffic can be load-balanced to PE-2 and PE-3. With VRRP on EVPN-VXLAN R-VPLS interfaces:
  - Ping and traceroute reply can be configured and are supported. BFD is also supported to speed up the fault detection.
  - **standby-forwarding**, even if it were configured for VRRP, would not have any effect in this configuration: the standby PE will never see any flooded traffic sent to it, therefore this command is not applicable to this scenario.
- When a VPRN 10 VPLS interface is bound to VPLS 101, EVPN advertises all the IP addresses configured for that VPLS interface as MAC routes with a static MAC indication. For the remote EVPN peers, that means that those MAC addresses linked to remote IP interfaces are protected. VRRP virtual IP/MACs are also advertised by EVPN as “static” and so protected. In the example of Figure 270, the VPLS 101 FDB in PE-1 shows the IP interface MACs and VRRP MACs as **EvpnS:P** (Static and protected MAC) as shown in the following output:

```
*A:PE-1# show service id 101 fdb detail
===============================================================================
Forwarding Database, Service 101
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
101       00:00:01:01:01:01 sap:1/2/1:101            L/30     06/17/16 12:21:32
101       00:ca:fe:ca:fe:02 vxlan:                   EvpnS:P  06/17/16 12:21:42
                            192.0.2.2:101
101       00:ca:fe:ca:fe:03 vxlan:                   EvpnS:P  06/17/16 12:21:53
                            192.0.2.3:101
101       00:ca:fe:ca:fe:53 vxlan:                   EvpnS:P  06/17/16 12:21:55
                            192.0.2.3:101
101       00:ca:fe:ca:fe:54 vxlan:                   EvpnS:P  06/17/16 12:21:45
                            192.0.2.2:101
-------------------------------------------------------------------------------
```

```
No. of MAC Entries: 5
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

The VPRN 10 VRRP instances on PE-2 are the following:

```
*A:PE-2# show router 10 vrrp instance
===============================================================================
VRRP Instances
===============================================================================
Interface Name                    VR Id Own Adm  State       Base Pri   Msg Int
                                  IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
int-1                             1     No  Up   Master       254        1
                                  IPv4      Up   n/a         254        No
  Backup Addr: 172.16.0.254
int-1                             2     No  Up   Backup       100        1
                                  IPv4      Up   n/a         100        No
  Backup Addr: 172.16.0.253
-------------------------------------------------------------------------------
Instances : 2
===============================================================================
```

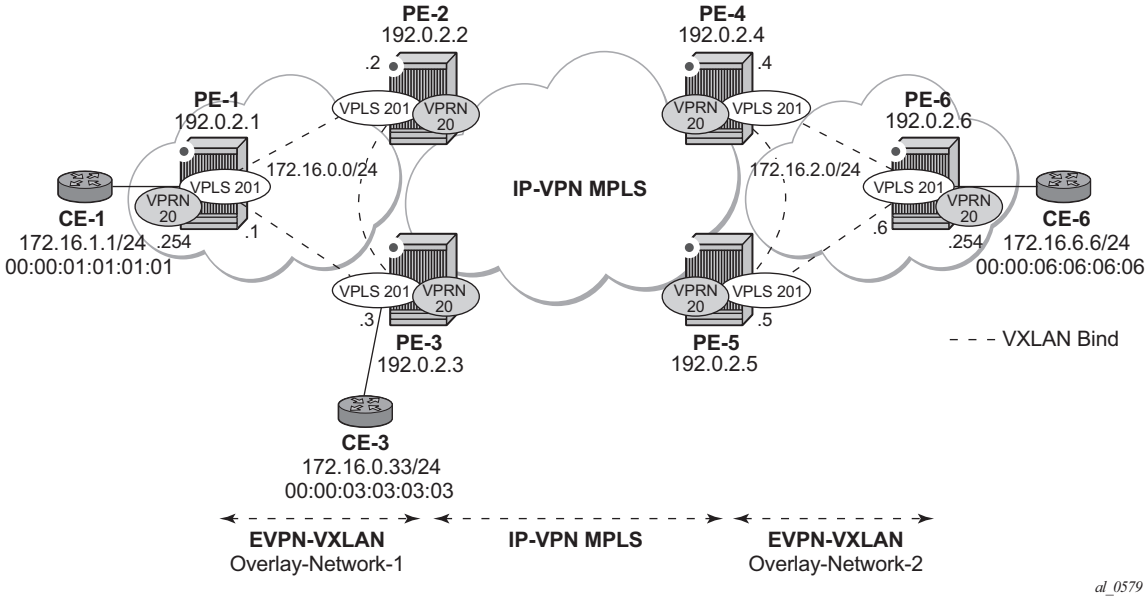
The ARP entries for PE-2 are the following:

```
*A:PE-2# show router 10 arp
===============================================================================
ARP Table (Service: 10)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
172.16.0.2      00:ca:fe:ca:fe:02 00h00m00s Oth[I] int-1
172.16.0.3      00:ca:fe:ca:fe:03 00h00m00s Evp[I] int-1
172.16.0.253    00:ca:fe:ca:fe:53 00h00m00s Oth    int-1
172.16.0.254    00:ca:fe:ca:fe:54 00h00m00s Oth[I] int-1
-------------------------------------------------------------------------------
No. of ARP Entries: 4
===============================================================================
```

EVPN-VXLAN in IRB Backhaul R-VPLS Services

Figure 272 illustrates the second inter-subnet forwarding scenario, where Layer 3 connectivity must be provided not only between the overlay networks but also within each overlay network. In the example depicted in Figure 272, a customer (tenant) has different subnets and connectivity must be provided across all of them (CE-1, CE-3, and CE-6 must be able to communicate), bearing in mind that EVPN-VXLAN is enabled in each overlay network and IP-VPN MPLS is used to inter-connect both overlay networks. VPLS 201 is an IRB Backhaul R-VPLS service since it provides connectivity to the VPRN instances.

Figure 272 **EVPN-VXLAN for IRB Backhaul R-VPLS Services**



From a BGP peering perspective, there is no change in this scenario compared to the previous one: PE-1 and PE-6 only support the EVPN address family. However in this scenario, CE-1 is now connected to an R-VPLS directly linked to the VPRN instances in PE-2/PE-3. As a result of that, IP prefixes must be exchanged between PE-1 and PE-2/PE-3. EVPN is able to advertise not only MAC routes and Inclusive Multicast routes, but also IP prefix routes that contain IP prefixes that can be installed in the attached VPRN routing table.

As an example, the VPRN 20 and VPLS 201 configurations on PE-1, PE-2 and PE-3 are shown. Similar configurations are needed in PE-3, PE-4 and PE-6.

On PE-1, VPRN 20 is configured as follows:

```
configure
    service
        vprn 20 customer 1 create
            route-distinguisher 192.0.2.1:20
            vrf-target target:64500:20
            interface "int-evi-201" create
                address 172.16.0.1/24
                vpls "evi-201"
                exit
            exit
            interface "int-PE-1-CE-1" create
                address 172.16.1.254/24
                sap 1/2/1:20 create
                exit
            exit
            no shutdown
```

On PE-1, VPLS 201 is configured as follows:

```
configure
    service
        vpls 201 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 201 create
            exit
            bgp
                route-distinguisher 192.0.2.1:201
                route-target export target:64500:201 import target:64500:201
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-201"
            no shutdown
```

On PE-2, VPRN 20 is configured with auto-bind-tunnel, as follows:

```
configure
    service
        vprn 20 customer 1 create
            route-distinguisher 192.0.2.2:20
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:20
            interface "int-evi-201" create
                address 172.16.0.2/24
                vpls "evi-201"
                exit
            exit
            no shutdown
```

On PE-2, VPLS 201 is configured as follows:

```
configure
    service
        vpls 201 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 201 create
            exit
            bgp
                route-distinguisher 192.0.2.2:201
                route-target export target:64500:201 import target:64500:201
            exit
            bgp-evpn
                ip-route-advertisement
```

```
            vxlan
                no shutdown
            exit
        exit
        service-name "evi-201"
        no shutdown
```

On PE-3, VPRN 20 is configured with auto-bind-tunnel, as follows:

```
configure
    service
        vprn 20 customer 1 create
            route-distinguisher 192.0.2.3:20
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:20
            interface "int-evi-201" create
                address 172.16.0.3/24
                vpls "evi-201"
                exit
            exit
            no shutdown
```

On PE-3, VPLS 201 is configured as follows:

```
configure
    service
        vpls 201 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 201 create
            exit
            bgp
                route-distinguisher 192.0.2.3:201
                route-target export target:64500:201 import target:64500:201
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-201"
            sap 1/2/1:20 create
            exit
            no shutdown
```

As shown in the CLI excerpt, the configuration in the three nodes (PE-1/2/3) for VPLS 201 and VPRN 20 is very similar. The main difference is the **auto-bind-tunnel** command existing in PE-2/3's VPRN 20. This command allows the VPRN 20 on PE-2/3 to receive IP-VPN routes from the core and resolve them to MPLS tunnels. VPRN 20 on PE-1 does not require such command because all its IP prefixes are resolved to local interfaces or to EVPN peers.

The **ip-route-advertisement** command enables:

- The advertisement of IP prefixes in EVPN, in routes type 5. All the existing IP prefixes in the attached VPRN 20 routing table are advertised in EVPN within the VPLS 201 context (except for the ones associated to VPLS 201 itself).
- The installation of IP prefixes in the attached VPRN 20 routing table with a preference of 169 (bgp-vpn routes for IP-VPN have a preference of 170) and a next-hop of the GW-IP (Gateway IP) address included in the EVPN IP prefix route.

For instance, the following output shows that PE-1 advertises the IP prefix 172.16.1.0/24 as a EVPN route to PE-3 (similar route is sent to PE-2), captured by a **debug router bgp update** session.

```
4 2016/06/20 12:17:31.60 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-IP-Prefix Len: 34 RD: 192.0.2.1:201, tag: 0,
                             ip_prefix: 172.16.1.0/24 gw_ip 172.16.0.1 Label: 201
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:201
        bgp-tunnel-encap:VXLAN
"
```

The VPRN 20 routing table in PE-1 is as follows:

```
*A:PE-1# show router 20 route-table
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                   Local   Local     23h57m48s   0
       int-evi-201                                                    0
172.16.1.0/24                                   Local   Local     23h57m35s   0
```

```
       int-PE-1-CE-1                                                    0
172.16.2.0/24                                 Remote  BGP EVPN  00h00m17s  169
       172.16.0.2                                                       0
172.16.6.0/24                                 Remote  BGP EVPN  00h00m17s  169
       172.16.0.2                                                       0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The subnet 172.16.0.0/24 is used on the interfaces “int-evi-201” in overlay network 1 and subnet 172.16.2.0/24 is used on similar interfaces in overlay network 2. CE-1 has an IP address in subnet 172.16.1.0/24 and CE-6 has an IP address in subnet 172.16.6.0/24. The next hop to reach 172.16.2.0/24 (overlay network 2) or CE-6, is 172.16.0.2 (PE-2), but it could have been PE-3,

There is redundancy in the example setup and therefore loops can occur. This is why routing policies need to be configured on the core PEs (PE-2, PE-3, PE-4, PE-5). These policies are described in section Use of Routing Policies to Avoid Routing Loops in Redundant PEs for routing loop use-case 1.

The routing table on PE-2 shows a BGP EVPN route toward CE-1 (subnet 172.16.1.0/24) via PE-1. The route toward CE-6 uses a tunnel toward PE-4 in overlay network 2.

```
*A:PE-2# show router 20 route-table
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     01d00h08m  0
       int-evi-201                                                      0
172.16.1.0/24                                 Remote  BGP EVPN  00h45m09s  169
       172.16.0.1                                                       0
172.16.2.0/24                                 Remote  BGP VPN   01d00h07m  170
       192.0.2.4  (tunneled)                                           0
172.16.6.0/24                                 Remote  BGP VPN   01d00h07m  170
       192.0.2.4  (tunneled)                                           0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The routing table on PE-3 is as follows:

```
*A:PE-3# show router 20 route-table
```

```
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                   Local   Local     00h09m25s  0
       int-evi-201                                                 0
172.16.1.0/24                                   Remote  BGP EVPN  00h09m25s  169
       172.16.0.1                                                  0
172.16.2.0/24                                   Remote  BGP VPN   00h09m24s  170
       192.0.2.4 (tunneled)                                        0
172.16.6.0/24                                   Remote  BGP VPN   00h09m24s  170
       192.0.2.4 (tunneled)                                        0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-3#
```

When checking the operation of EVPN in this scenario, it is important to observe that the right next hops and prefixes are successfully installed in the VPRN 20 routing table:

- EVPN IP prefixes are sent using a GW-IP matching the primary IP interface address of the R-VPLS for which the routes are sent. For instance, as shown above, IP prefix 172.16.1.0/24 is advertised from PE-1 with GW-IP 172.16.0.1 (the IP address configured for the VPRN 20 VPLS interface in PE-1). In the PE-2/3 VPRN 20 routing tables, IP prefix 172.16.1.0/24 is installed with next hop 172.16.0.1. Traffic arriving at PE-2/3 on VPRN 20 with IP Destination Address (DA) in the 172.16.1.0 subnet matches the mentioned routing table entry. As usual, the next hop is resolved by the ARP table to a MAC and the MAC resolved by the FDB table to an egress VTEP, VNI.
- IP prefixes in the VPRN 20 routing table are advertised in IP-VPN to the remote IP-VPN MPLS peers. Received IP-VPN prefixes are installed in the VPRN 20 routing table using the remote PE system IP address as the next hop, as usual. For instance, 172.16.3.0/24 is installed in PE-2 VPRN 20's routing table with next hop (tunneled) 192.0.2.4 and preference 170.

The following considerations of how the routing table manager (RTM) handles EVPN and IP-VPN prefixes must be taken into account:

- Only VPRN interface primary addresses are advertised as GW-IP in EVPN IP prefix routes. Secondary addresses are never sent as GW-IP addresses.

- EVPN IP prefixes are advertised by default as soon as the **ip-route-advertisement** command is enabled and there are active IP prefixes in the attached VPRN routing table.
- If the same IP prefix is received on a PE via EVPN and IP-VPN at the same time for the same VPRN, by default the EVPN prefix is selected since its preference (169) is better than the IP-VPN preference (170).
- Since EVPN has a better preference compared to IP-VPN, when the VPRNs on redundant PEs are attached to the same R-VPLS service, routing loops may occur. The use case described here is an example where routing loops can occur. Check Use of Routing Policies to Avoid Routing Loops in Redundant PEs to avoid routing loops in redundant PEs for more information.
- When the command **ip-route-advertisement** is enabled, the subnet IP prefixes are advertised in EVPN but not the “host” IP prefixes (/32 prefixes associated with the local interfaces). If the user wants to advertise the host IP prefixes as well, the **incl-host** keyword must be added to the **ip-route-advertisement** command. The following example illustrates this. The host routes can be shown with the **show router route-table all** command. When the **incl-host** keyword is added to PE-1’s VPLS 201, PE-1 advertises the host routes as well and these are installed in the remote PEs’ routing tables.

```
*A:PE-1# show router 20 route-table
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                   Local   Local     01d04h18m   0
       int-evi-201                                                    0
172.16.1.0/24                                   Local   Local     01d04h17m   0
       int-PE-1-CE-1                                                  0
172.16.2.0/24                                   Remote  BGP EVPN  04h20m31s   169
       172.16.0.2                                                     0
172.16.6.0/24                                   Remote  BGP EVPN  04h20m31s   169
       172.16.0.2                                                     0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================


*A:PE-1# show router 20 route-table all
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                Active     Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                   Local   Local     01d04h18m   0
       int-evi-201                                              Y          0
```

```
172.16.0.1/32                                    Local   Host      01d04h18m  0
       int-evi-201                                        Y           0
172.16.1.0/24                                    Local   Local     01d04h17m  0
       int-PE-1-CE-1                                      Y           0
172.16.1.254/32                                  Local   Host      01d04h17m  0
       int-PE-1-CE-1                                      Y           0
172.16.2.0/24                                    Remote  BGP EVPN  04h20m34s  169
       172.16.0.2                                         Y           0
172.16.6.0/24                                    Remote  BGP EVPN  04h20m34s  169
       172.16.0.2                                         Y           0
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
       E = Inactive best-external BGP route
===============================================================================

*A:PE-1# configure service vpls 201 bgp-evpn ip-route-advertisement incl-host

*A:PE-2# show router 20 route-table
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                               Type    Proto     Age        Pref
      Next Hop[Interface Name]                                       Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                    Local   Local     01d04h22m  0
       int-evi-201                                                   0
172.16.1.0/24                                    Remote  BGP EVPN  04h25m22s  169
       172.16.0.1                                                    0
172.16.1.254/32                                  Remote  BGP EVPN  00h03m52s  169
       172.16.0.1                                                    0
172.16.2.0/24                                    Remote  BGP VPN   01d04h22m  170
       192.0.2.4  (tunneled)                                        0
172.16.6.0/24                                    Remote  BGP VPN   01d04h22m  170
       192.0.2.4  (tunneled)                                        0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

- ECMP is fully supported for the VPRN for EVPN IP prefix routes coming from different GW-IP next-hops. However ECMP is not supported for IP prefixes routes belonging to different owners (EVPN and IP-VPN). ECMP is enabled in VPRN 20 on PE-1, as follows:

```
*A:PE-1# configure service vprn 20 ecmp 2
```

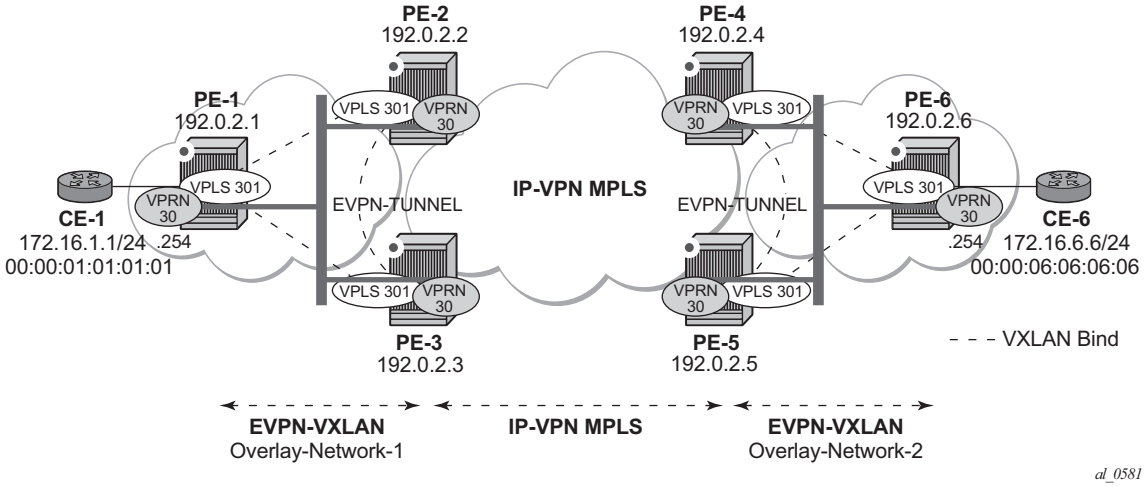
When policies are applied that prevent routing loops, as described in section Use of Routing Policies to Avoid Routing Loops in Redundant PEs, both PE-2 and PE-3 have IP-VPN tunnels for IP prefixes 172.16.2.0/24 and 172.16.6.0/24. In that case, an additional route with a different GW IP as next hop is installed in the routing table for these IP prefixes:

```
*A:PE-1# show router 20 route-table
===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                   Local   Local     00d00h10m   0
       int-evi-201                                                   0
172.16.1.0/24                                   Local   Local     00d00h10m   0
       int-PE-1-CE-1                                                 0
172.16.2.0/24                                   Remote  BGP EVPN  00h00m01s   169
       172.16.0.2                                                    0
172.16.2.0/24                                   Remote  BGP EVPN  00h00m01s   169
       172.16.0.3                                                    0
172.16.6.0/24                                   Remote  BGP EVPN  00h00m01s   169
       172.16.0.2                                                    0
172.16.6.0/24                                   Remote  BGP EVPN  00h00m01s   169
       172.16.0.3                                                    0
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

EVPN-VXLAN in EVPN Tunnel R-VPLS Services

The previous scenario shows how to use EVPN-VXLAN to provide inter-subnet forwarding for a tenant, where R-VPLS services can contain hosts and also offer transit services between VPRN instances. For example, in the use case depicted in Figure 272, VPLS 201 in Overlay-Network-1 is an R-VPLS that can provide intra-subnet connectivity to all the hosts in subnet 172.16.0.0/24 (for example, CE-3 belongs to this subnet) but it can also provide “transit” or “backhaul” connectivity to hosts in subnet 172.16.1.0/24 (for example, CE-1) sending packets to subnets 172.16.2.0/24 or 172.16.6.0/24. In some cases, the R-VPLS where EVPN-VXLAN is enabled does not need to provide intra-subnet connectivity and it is purely a transit or backhaul service where VPRN IRB interfaces are connected. Figure 273 illustrates this use case.

Figure 273 **EVPN-VXLAN in EVPN-tunnel R-VPLS Services**



Compared to the use case in Figure 272, in this case the R-VPLS connecting the IRB interfaces in Overlay-Network-1 (VPLS 301) does not have any connected host. If that is the case, VPLS 301 can be configured as an EVPN tunnel.

EVPN tunnels are enabled using the **evpn-tunnel** command under the R-VPLS interface configured on the VPRN. EVPN tunnels bring the following benefits to EVPN-VXLAN IRB backhaul R-VPLS services:

- Easier and simpler provisioning of the tenant service: if an EVPN tunnel is configured in an IRB backhaul R-VPLS, there is no need to provision the IRB IP addresses in the VPRN. This makes the provisioning easier to automate and saves IP addresses from the tenant IP space.
- Higher scalability of the IRB backhaul R-VPLS: if EVPN tunnels are enabled, BUM traffic is suppressed in the EVPN-VXLAN IRB backhaul R-VPLS service (it is not required). As a result, the number of VXLAN bindings in IRB backhaul R-VPLS services with EVPN tunnels can be much higher.

As an example, the VPRN 30 and VPLS 301 configurations on PE-1, PE-2 and PE-3 are shown. Similar configurations are needed in PE-4, PE-5 and PE-6.

On PE-1:

```
configure
    service
        vprn 30 customer 1 create
            route-distinguisher 192.0.2.1:30
            vrf-target target:64500:30
            interface "int-PE-1-CE-1" create
                address 172.16.0.254/24
                sap 1/1/1:30 create
```

```
                exit
            exit
            interface "int-evi-301" create
                vpls "evi-301"
                    evpn-tunnel
                exit
            exit
            no shutdown

configure
    service
        vpls 301 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 301 create
            exit
            bgp
                route-distinguisher 192.0.2.1:301
                route-target export target:64500:301 import target:64500:301
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-301"
            no shutdown
```

On PE-2:

```
configure
    service
        vprn 30 customer 1 create
            route-distinguisher 192.0.2.2:30
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:30
            interface "int-evi-301" create
                vpls "evi-301"
                    evpn-tunnel
                exit
            exit
            no shutdown

configure
    service
        vpls 301 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 301 create
            exit
            bgp
```

```
                route-distinguisher 192.0.2.2:301
                route-target export target:64500:301 import target:64500:301
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-301"
            no shutdown
```

On PE-3:

```
configure
    service
        vprn 30 customer 1 create
            route-distinguisher 192.0.2.3:30
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:30
            interface "int-evi-301" create
                vpls "evi-301"
                    evpn-tunnel
                exit
            exit
            no shutdown

configure
    service
        vpls 301 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 301 create
            exit
            bgp
                route-distinguisher 192.0.2.3:301
                route-target export target:64500:301 import target:64500:301
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-301"
            no shutdown
```

As shown in the preceding output , the configuration in the three nodes (PE-1/2/3) for VPLS 301 and VPRN 30 is similar to the configuration of VPLS 201 and VPRN 20 in the previous scenario, however, when the **evpn-tunnel** command is added to the VPRN interface, there is no need to configure an IP interface address. The option **evpn-tunnel** can be enabled independently of **ip-route-advertisement** (although no route-type 5 advertisements are sent in that case).

A VPRN supports regular IRB backhaul R-VPLS services as well as EVPN tunnel R-VPLS services. A maximum of eight R-VPLS services with **ip-route-advertisement** enabled per VPRN is supported (in any combination of regular IRB R-VPLS or EVPN tunnel R-VPLS services). EVPN tunnel R-VPLS services do not support SAPs or SDP-binds. No frames are flooded in an EVPN tunnel R-VPLS service, and, in fact no inclusive multicast routes are exchanged in R-VPLS services that are configured as EVPN tunnels.

The **show service id vxlan** command for an R-VPLS service configured as an EVPN tunnel shows <egress VTEP, VNI> bindings excluded from Mcast, in other words, the VXLAN bindings are not used to flood BUM traffic:

```
*A:PE-2# show service id 301 vxlan
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 301
Creation Origin: manual
RestProtSrcMacAct: none

===============================================================================
VPLS VXLAN service Network Specifics
===============================================================================
Ing Net QoS Policy : none                          Vxlan VNI Id     : 301
Ingress FP QGrp    : (none)                        Ing FP QGrp Inst : (none)

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
192.0.2.1              301            1            -       Up           No
192.0.2.3              301            1            -       Up           No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================
```

The process followed upon receiving a route-type 5 on a regular IRB R-VPLS interface (previous scenario) differs from the one for an EVPN tunnel type (this scenario):

- IRB backhaul R-VPLS VPRN interface:

– When a route-type 2 that includes an IP address is received and it becomes active, the MAC/IP information is added to the FDB and ARP tables. This can be checked with the **show>router>arp** command and the **show>service>id>fdb detail** command.

– When a route-type 5 is received on (for instance) PE-2, and becomes active for the R-VPLS service, the IP prefix is added to the VPRN routing table regardless of the existence of a route-type 2 that can resolve the GW IP address. If a packet is received from the WAN side and the IP lookup hits an entry for which the GW IP (IP next-hop) does not have an active ARP entry, the system will ARP to get the MAC. If the ARP is resolved but the MAC is unknown in the FDB table, the system will flood the ARP message into the R-VPLS multicast list. Routes type 5 can be checked in the routing table with the **show>router>route-table** command and the **show>router>fib** command.

• EVPN tunnel R-VPLS VPRN interface:

– When a route-type 2 is received and becomes active, the MAC address is added to the FDB (only). This MAC address is normally a GW-MAC.

– When a route-type 5 is received on (for instance) PE-1, the system looks for the GW-MAC. The IP prefix is added to the VPRN routing table with next hop equal to EVPN-tunnel-GW-MAC; for example, ET-4a:c5:ff:00:00:6a is an EVPN tunnel with GW-MAC 4a:c5:ff:00:00:6a. The GW-MAC is added from the GW-MAC extended community sent along with the route-type 5 for prefix 172.16.6.0/24. If a packet is received from CE-1 and the IP lookup hits an entry for which the next hop is a EVPN tunnel:GW-MAC, the system looks up the GW-MAC in the FDB. Normally a route-type 2 with the GW-MAC has already been received so that the GW-MAC has been added to the FDB. If the GW-MAC is not present in the FDB, the packet will be dropped.

– The IP prefixes with GW-MACs as next hops for the setup in Figure 273 are displayed in the show router route-table command, as follows:

```
*A:PE-1# show router 30 route-table
===============================================================================
Route Table (Service: 30)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                   Local   Local     00h06m15s  0
       int-PE-1-CE-1                                                  0
172.16.6.0/24                                   Remote  BGP EVPN  00h05m31s  169
       int-evi-301 (ET-4a:c5:ff:00:00:6a)                             0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The same routing policies are applied on the core PEs to prevent loops; see Use of Routing Policies to Avoid Routing Loops in Redundant PEs.

The **show service id fdb detail** command can be used to look for the forwarding information for a GW MAC:

```
*A:PE-1# show service id 301 fdb detail

===============================================================================
Forwarding Database, Service 301
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
301       4a:c4:ff:00:00:6a cpm                      Intf     06/22/16 12:26:44
301       4a:c5:ff:00:00:6a vxlan:                   EvpnS:P  06/22/16 13:56:40
                            192.0.2.2:301
301       4a:c6:ff:00:00:6a vxlan:                   EvpnS:P  06/22/16 13:14:25
                            192.0.2.3:301
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================
```

IP prefix routes sent for EVPN tunnel R-VPLS services do not contain a GW-IP (the GW-IP will be zero) but convey a GW-MAC address that is used in the peer VPRN routing table. The following output shows PE-2's VPRN 30 interface MAC address sent to PE-1:

```
*A:PE-2# show router 30 interface "int-evi-301" detail | match MAC
MAC Address      : 4a:c5:ff:00:00:6a    Mac Accounting    : Disabled
*A:PE-2#
```

When ip-route-advertisement is enabled, PE-2 sends route type 5 messages to PE-1, as can be seen in the following BGP update for the route toward subnet 172.16.6.0/24 in overlay network 2, using the MAC as GW-MAC:

```
*A:PE-2# configure service vpls 301 bgp-evpn ip-route-advertisement

8 2016/06/22 12:04:54.36 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 97
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-IP-Prefix Len: 34 RD: 192.0.2.2:301, tag: 0,
                             ip_prefix: 172.16.6.0/24 gw_ip 0.0.0.0 Label: 301
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
```

```
        target:64500:301
        mac-nh:4a:c5:ff:00:00:6a
        bgp-tunnel-encap:VXLAN
"
```

Looking at the VPRN 30 routing table, since IP prefixes are shown with an EVPN tunnel next-hop (GW-MAC) as opposed to an IP next-hop, the user may think that no ARP entries are consumed by VPRN 30. However, internal ARP entries are still consumed in VPRN 30. Although not shown in the show router 30 arp command, the **summary** option shows the consumption of internal ARP entries for EVPN.

```
*A:PE-2# show router 30 route-table
===============================================================================
Route Table (Service: 30)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                   Remote  BGP EVPN  00h31m34s  169
       int-evi-301 (ET-4a:c5:ff:00:00:6a)                             0
172.16.6.0/24                                   Remote  BGP VPN   00h59m09s  170
       192.0.2.4 (tunneled)                                           0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

There are no entries in the ARP table:

```
*A:PE-2# show router 30 arp

===============================================================================
ARP Table (Service: 30)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
```

One internal BGP-EVPN ARP entry is consumed, as can be seen as follows:

```
*A:PE-2# show router 30 arp summary

============================================================
ARP Table Summary (Service: 30)
============================================================
Local ARP Entries    : 1
Static ARP Entries   : 0
Dynamic ARP Entries  : 0
Managed ARP Entries  : 0
Internal ARP Entries : 0
BGP-EVPN ARP Entries : 1
```

```
-------------------------------------------------------------
No. of ARP Entries   : 2
=============================================================
```

The number of BGP-EVPN ARP entries in the **show router 30 arp summary** command matches the number of remote valid GW-MACs for VPRN 30.

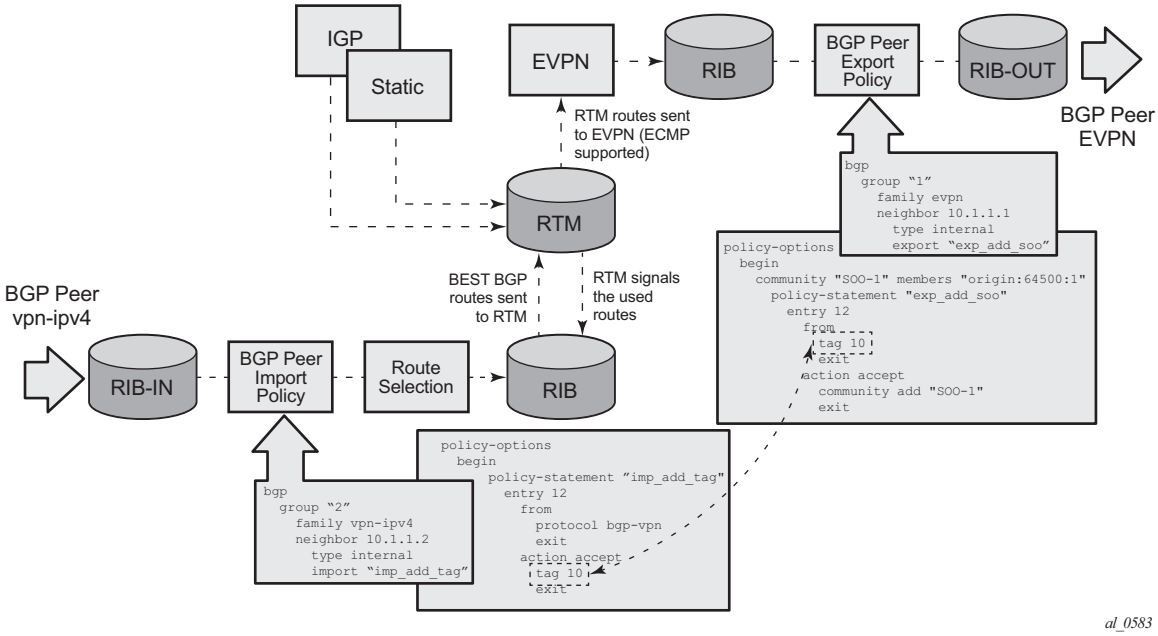
Routing Policies for IP Prefixes in EVPN

Routing policies are supported for IP prefixes imported/exported through BGP EVPN. The default import/export behavior for IP prefixes in EVPN can be modified by the use of routing policies applied either at peer level (**config>router>bgp>group/group>neighbor>import/export**) or VPLS level (**config>service>vpls>bgp>vsi-import/vsi-export**).

When applying routing policies to control the distribution of prefixes between EVPN and IP-VPN, the user must take into account that both families are completely separated as far as BGP is concerned and that when prefixes from a family are imported in the RTM, the BGP attributes are lost to the other family. The use of tags allows the controlled distribution of prefixes across the two families.

Figure 274 illustrates how vpn-ipv4 routes are imported into the RTM and then passed onto EVPN for its own processing. Note that vpn-ipv4 routes can be tagged at ingress and this tag is preserved throughout the RTM and EVPN processing so that the tag can be “matched” by the egress BGP routing policy. In this particular example, egress EVPN routes matching tag 10, are modified to add a site-of-origin community origin:64500:1.

Figure 274 **Routing Policies for Egress EVPN Routes**

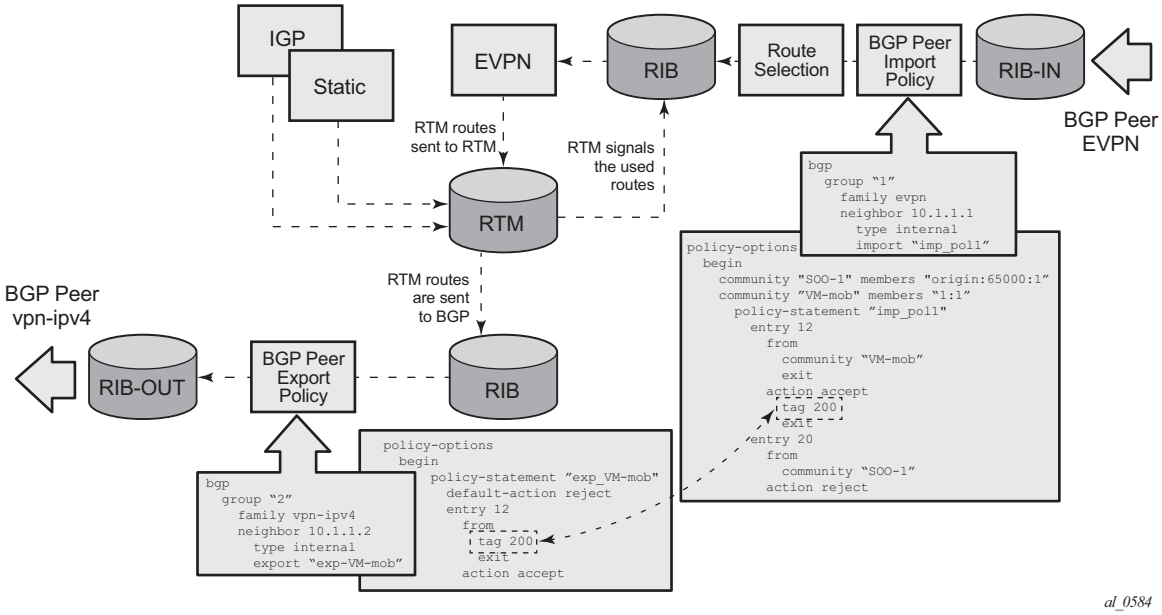


Policy tags can be used to match EVPN IP-prefixes that were learned not only from BGP vpn-ipv4 but also from other routing protocols. The tag range supported for each protocol is different:

```
<tag>  : accepts in decimal or hex
         [0x1..0xFFFFFFFF]H (for OSPF and ISIS)
         [0x1..0xFFFF]H (for RIP)
         [0x1..0xFF]H (for BGP)
```

Figure 275 illustrates the reverse workflow: routes imported from EVPN and exported from RTM to BGP vpn-ipv4. In this example, EVPN routes received with community VM-mob are tagged with TAG 200. At the egress vpn-ipv4 peers, only the routes with TAG 200 are advertised.

Figure 275 **Routing Policies for Ingress EVPN Routes**



The preceding behavior and the use of tags is also valid for **vsi-import** and **vsi-export** policies. The behavior can be summarized in the following statements:

- For EVPN prefix routes received and imported in RTM:
  - Routes can be matched on communities and tags can be added to them. This works at peer level or vsi-import level.
  - Well-known communities (no-export|no-export-subconfed|no-advertise) also require that the routing policies add a tag if the user wants to modify the behavior when exporting to BGP.
  - Routes can be matched based on family evpn.
  - Routes cannot be matched on prefix-list.
- For exporting RTM to EVPN prefix routes:
  - Routes can be matched on tags and based on that, communities added, or routes accepted or rejected, etc. This works at peer level or vsi-export level.
  - Tags can be added for static-routes, rip, ospf, isis and bgp and then be matched in the vsi-export policy for EVPN IP-prefix route advertisement.
  - Tags cannot be added for direct routes.

Use of Routing Policies to Avoid Routing Loops in Redundant PEs

When redundant PE VPRN instances are connected to the same R-VPLS service (IRB backhaul or EVPN tunnel R-VPLS) with the ip-route-advertisement command enabled, routing loops can occur in two different use-cases:

1. Routing loop caused by EVPN and IP-VPN interaction in the RTM.
2. Routing loop caused by EVPN in “parallel” R-VPLS services.

Policy configuration examples for both cases are provided in the following sections.

Routing loop use-case 1: EVPN and IP-VPN interaction

This use case refers to scenarios with redundant PEs and VPRNs connected to the same R-VPLS with **ip-route-advertisement**. The scenarios in Figure 272 (EVPN-VXLAN for IRB Backhaul R-VPLS services) and Figure 273 (EVPN-VXLAN in EVPN tunnel R-VPLS services) are examples of this use case. In both scenarios, the following process causes a routing loop:

1. IP prefix 172.16.6.0/24 is advertised by PE-4 to PE-2 and PE-3.
2. PE-2 imports that prefix in the VPRN routing table and re-advertises the IP prefix in EVPN to PE-1 and PE-3 (the same thing happens in PE-3).
3. PE-3 already has the 172.16.6.0/24 prefix in the VPRN routing table with preference 170 (IP-VPN) but because it receives the IP prefix from EVPN with lower preference (169), the RTM will install the EVPN prefix in the VPRN routing table (the same thing happens in PE-2).
4. PE-3 advertises the EVPN learned IP prefix to all MP-BGP vpn-ipv4 peers (also PE-2).
5. PE-2 receives the IP prefix again from PE-3 and will advertise it in EVPN again, creating a routing loop (PE-3 will do the same thing as well).

This routing loop also happens in traditional multi-homed IP-VPN scenarios where the PE-CE eBGP and MP-BGP vpn-ipv4/v6 protocols interact in the same VPRN RTM, with different router preferences. In either case (EVPN or eBGP interaction with MP-BGP) the issue can be solved by the use of routing policies and site-of-origin communities.

Routing policies are applied to PE-2 and PE-3 (also to PE-4 and PE-5) and allow the redundant PEs to reject their own generated routes in order to avoid the loops. These routing policies can be applied at vsi-import/export level or BGP group/neighbor level. The following output shows an example of routing policies applied at BGP neighbor level for PE-2 (similar policies are applied on PE-3/4/5). Neighbor or group level policies are the preferred way in this kind of use case: a single set of policies is sufficient, as opposed to a set of policies per service (if the policies are applied at vsi-import/export level).

The following policies are applied in the BGP group or BGP group/neighbor context on PE-2:

```
configure
    router
        bgp
            group "DC"
                neighbor 192.0.2.1
                    import "add-tag_to_bgp-evpn_routes"
                exit
                neighbor 192.0.2.3
                    import "reject_based_on_SOO"
                    export "add-SOO_on_export"
                exit
            exit
            group "WAN"
                import "add-tag_to_bgp-vpn_routes"
            exit
```

The routing policies are configured as follows on PE-2:

```
configure
    router
        policy-options
            begin
            community "SOO-PE-2" members "origin:2:1"
            community "SOO-PE-3" members "origin:3:1"
            policy-statement "add-SOO_on_export"
                entry 10
                    from
                        tag 0x2
                    exit
                    action accept
                        community add "SOO-PE-2"
                    exit
                exit
                entry 20
                    from
                        tag 0x3
                    exit
                    action accept
                        community add "SOO-PE-3"
                    exit
                exit
            exit
            policy-statement "reject_based_on_SOO"
```

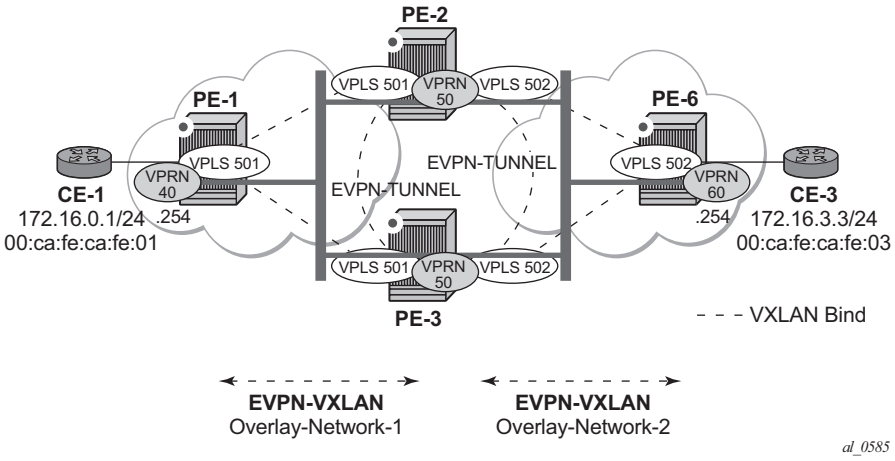
```
        entry 10
            from
                community "SOO-PE-2"
            exit
            action reject
        exit
        entry 20
            from
                community "SOO-PE-3"
            exit
            action reject
        exit
    exit
    policy-statement "add-tag_to_bgp-vpn_routes"
        entry 10
            from
                protocol bgp-vpn
            exit
            action accept
                tag 0x2
            exit
        exit
    exit
    policy-statement "add-tag_to_bgp-evpn_routes"
        entry 10
            from
                family evpn
            exit
            action accept
                tag 0x2
            exit
        exit
    exit
    commit
```

EVPN and MP-BGP routes are tagged at import; on export, a site-of-origin community is added. Routes exchanged between the two redundant PEs are rejected if they are received by a PE with its own site-of-origin.

Routing loop use-case 2: EVPN in parallel R-VPLS services

If a VPRN is connected to more than one R-VPLS with **ip-route-advertisement** enabled, IP prefixes that belong to one R-VPLS are advertised into the other R-VPLS and vice versa. When redundant PEs are used, a routing loop will occur. Figure 276 illustrates this use case. The example shows R-VPLS with an EVPN tunnel configuration, but the same routing loop occurs for regular IRB backhaul R-VPLS services.

Figure 276 **EVPN in Parallel R-VPLS Services**



The configuration of VPRN 50 as well as VPLS 501/502 and the required policies are as follows. For this use case, policies must be applied at vsi-import/export level since more granularity is required when modifying the imported/exported routes.

On PE-2:

```
configure
    service
        vprn 50 customer 1 create
            route-distinguisher 192.0.2.2:50
            interface "int-evi-501" create
                vpls "evi-501"
                    evpn-tunnel
                exit
            exit
            interface "int-evi-502" create
                vpls "evi-502"
                    evpn-tunnel
                exit
            exit
            no shutdown


configure
    service
        vpls 501 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 501 create
            exit
            bgp
                route-distinguisher 192.0.2.2:501
                vsi-export "vsi-export-policy-501"
                vsi-import "vsi-import-policy-501"
            exit
```

```
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-501"
            no shutdown

configure
    service
        vpls 502 customer 1 create
            allow-ip-int-bind
            exit
            vxlan vni 502 create
            exit
            bgp
                route-distinguisher 192.0.2.2:502
                vsi-export "vsi-export-policy-502"
                vsi-import "vsi-import-policy-502"
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
            exit
            service-name "evi-502"
            no shutdown

configure
    router
        policy-options
            begin
            community "exp_RVPLS501" members "origin:2:11" "target:64500:501"
            community "exp_RVPLS502" members "origin:2:11" "target:64500:502"
            community "SOO-PE-2-RVPLS" members "origin:2:11"
            community "SOO-PE-3-RVPLS" members "origin:3:11"
            community "SOO_PE-3_RVPLS501" members "origin:3:11" "target:64500:501"
            community "SOO_PE-3_RVPLS502" members "origin:3:11" "target:64500:502"
            policy-statement "vsi-export-policy-501"
                entry 10
                    from
                        tag 0xc
                    exit
                    action accept
                        community add "SOO_PE-3_RVPLS501"
                    exit
                exit
                entry 20
                    action accept
                        community add "exp_RVPLS501"
                    exit
                exit
            exit
            policy-statement "vsi-export-policy-502"
                entry 10
                    from
```

```
            tag 0xc
        exit
        action accept
            community add "SOO_PE-3_RVPLS502"
        exit
    exit
    entry 20
        action accept
            community add "exp_RVPLS502"
        exit
    exit
exit
policy-statement "vsi-import-policy-501"
    entry 10
        from
            community "SOO-PE-2-RVPLS"
        exit
        action reject
    exit
    entry 20
        from
            community "SOO_PE-3_RVPLS501"
        exit
        action accept
            tag 0xc
        exit
    exit
    default-action accept
    exit
exit
policy-statement "vsi-import-policy-502"
    entry 10
        from
            community "SOO-PE-2-RVPLS"
        exit
        action reject
    exit
    entry 20
        from
            community "SOO_PE-3_RVPLS502"
        exit
        action accept
            tag 0xc
        exit
    exit
    default-action accept
    exit
exit
commit
```

Troubleshooting and Debug Commands

For general information on EVPN and VXLAN troubleshooting and debug commands, please refer to chapter EVPN for VXLAN Tunnels (Layer 2). The following information focuses on specific commands for Layer-3 applications.

When troubleshooting and operating a EVPN-VXLAN scenario with inter-subnet forwarding, it is important to check the IP prefixes and next-hops, as well as ARP tables and FDB tables (**show router x route-table**, **show router x arp**, **show service id y fdb detail**).

ICMP commands can also help checking the connectivity. When traceroute is used on EVPN-VXLAN in EVPN tunnel interfaces, EVPN tunnel interface hops in the traceroute commands are showing the VPRN loopback address or the other non evpn-tunnel interface address. In VPRN services where all of the interfaces are of type EVPN tunnel, ICMP packets fail until an IP address is configured. The following output shows a traceroute from VPRN 30 in PE-1 to CE-6 and from PE-2 to CE-1 (see Figure 273):

```
*A:PE-1# traceroute router 30 172.16.6.6
traceroute to 172.16.6.6, 30 hops max, 40 byte packets
  1  0.0.0.0  * * *
  2  0.0.0.0  * * *
  3  172.16.6.254 (172.16.6.254)    2.33 ms  2.21 ms  2.38 ms
  4  172.16.6.6 (172.16.6.6)    2.72 ms  2.97 ms  2.88 ms


*A:PE-2# traceroute router 30 172.16.1.1
traceroute to 172.16.0.1, 30 hops max, 0 byte packets

Send failed. Unable to find local ip address
```

When troubleshooting R-VPLS services, specifically R-VPLS services configured as EVPN tunnels, the limit of peer PEs per EVPN tunnel service is much higher than for a regular R-VPLS service since the egress <VTEP, VNI> bindings do not have to be added to the multicast flooding list. For this reason, the following **tools dump** command has been added to check the consumed/total EVPN tunnel next hops. The number of EVPN tunnel next hops matches the number of remote GW-MAC addresses per EVPN tunnel R-VPLS service.

```
*A:PE-1# tools dump service id 501 evpn usage

Evpn Tunnel Interface IP Next Hop: 2/8189
```

Finally, when troubleshooting EVPN routes and routing policies, the **show router bgp routes evpn** command and its filters can help:

- Check that the expected routes are received, properly imported and communities/tags added/replaced/removed.

- Check that the expected routes are sent, properly exported and communities added/replaced/removed.

Examples of EVPN IP prefix routes including communities and tags are the following.

```
*A:PE-2# show router bgp routes evpn ?
  - evpn <evpn-type>

      auto-disc       - Display BGP EVPN Auto-Disc Routes
      eth-seg         - Display BGP EVPN Eth-Seg Routes
      inclusive-mcast - Display BGP EVPN Inclusive-Mcast Routes
      ip-prefix       - Display BGP EVPN IPv4-Prefix Routes
      ipv6-prefix     - Display BGP EVPN IPv6-Prefix Routes
      mac             - Display BGP EVPN Mac Routes


*A:PE-2# show router bgp routes evpn ip-prefix
  - ip-prefix [hunt|detail] [rd <rd>] [prefix <ip-prefix/ip-prefix-length>]
              [community <comm-id>] [tag <tag>] [next-hop <ip-address>]
              [aspath-regex <reg-exp>]

 <hunt|detail>        : keywords
 <rd>                 : {<ip-addr:comm-val>|
                        <2byte-asnumber:ext-comm-val>|
                        <4byte-asnumber:comm-val>}
 <ip-prefix/ip-pref*> : ipv4-prefix    - a.b.c.d
                        ipv4-prefix-le - [0..32]
                        ipv6-prefix    - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
                        ipv6-prefix-le - [0..128]
 <comm-id>            : <as-number1:comm-val1>|<ext-comm>|
                        <well-known-comm>
                        ext-comm        - <type>:{<ip-address:comm-val1>|
                                                 <as-number1:comm-val2>|
                                                 <as-number2:comm-val1>}
                        as-number1      - [0..65535]
                        comm-val1       - [0..65535]
                        type            - target|origin
                        ip-address      - a.b.c.d
                        comm-val2       - [0..4294967295]
                        as-number2      - [0..4294967295]
                        well-known-comm - null|no-export|no-export-subconfed|
                                          no-advertise
 <tag>                : [0..16777215] | MAX-ET
 <ip-address>         : a.b.c.d
 <reg-exp>            : [80 chars max]
```

Community origin:2:11 target:64500:502 is added to the outgoing routes (by the routing policy “vsi-export-policy-502”), as can be verified as follows:

```
*A:PE-2# show router bgp routes evpn ip-prefix hunt prefix 172.16.1.0/24 community
origin:69:11
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
```

```
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : N/A
Nexthop        : 192.0.2.2
To             : 192.0.2.1
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : origin:2:11 target:64500:502
                 mac-nh:4a:c5:ff:00:01:33 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 0
Gateway Address: 4a:c5:ff:00:01:33
Prefix         : 172.16.1.0/24
Route Dist.    : 192.0.2.2:502
MPLS Label     : VNI 502
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

---snip---
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-2#
```

Route tag 2 is added by PE-2 to all BGP EVPN routes (according to policy “add-tag_to_bgp-evpn_routes”), as can be verified in the following output:

```
*A:PE-2# show router bgp routes evpn ip-prefix prefix 172.16.1.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.2       AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
```

```
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
Original Attributes

Network        : N/A
Nexthop        : 192.0.2.1
From           : 192.0.2.1
Res. Nexthop   : 192.168.12.1
Local Pref.    : 100                    Interface Name : int-PE-2-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:64500:201 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 0
Gateway Address: 172.16.0.1
Prefix         : 172.16.1.0/24
Route Dist.    : 192.0.2.1:201
MPLS Label     : VNI 201
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m56s

Modified Attributes

Network        : N/A
Nexthop        : 192.0.2.1
From           : 192.0.2.1
Res. Nexthop   : 192.168.12.1
Local Pref.    : 100                    Interface Name : int-PE-2-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:64500:201 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 0
Gateway Address: 172.16.0.1
Prefix         : 172.16.1.0/24
Route Dist.    : 192.0.2.1:201
```

```
MPLS Label     : VNI 201
Route Tag      : 2
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m57s
-------------------------------------------------------------------------------
---snip---
```

Conclusion

SR OS supports not only the EVPN control plane for VXLAN tunnels in Layer 2 applications but also the simultaneous use of EVPN and VXLAN for VPN customers (tenants) with intra and inter-subnet connectivity requirements. R-VPLS services can be configured to provide default gateway connectivity to hosts, IRB backhaul connectivity to VPRN services and EVPN tunnel connectivity to VPRN services. When configured to do so, EVPN can advertise IP prefixes and interact with the VPRN RTM to propagate IP prefix connectivity between EVPN and other routing protocols in the VPRN, including IP-VPN. This example has shown how to configure R-VPLS services for all these functions, as well as how to configure routing policies for EVPN-based IP prefixes.

EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services

This chapter provides information about EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R5.

Chapters EVPN for MPLS Tunnels and EVPN for VXLAN Tunnels (Layer 2) are prerequisite reading.

Overview

When EVPN-MPLS is deployed in the WAN, many service providers are looking for a way to integrate existing Layer 2 EVPN-VXLAN based data center services into the WAN, while keeping the end-to-end advantages of EVPN. The IETF draft document in http://tools.ietf.org/html/draft-ietf-bess-dci-evpn-overlay describes how to provide Layer 2 connectivity for EVPN-overlay data centers in different ways. This chapter follows section 3.4 of that document, in which EVPN-MPLS is used in the same VPLS service that terminates overlay (VXLAN) tunnels.

To provide EVPN-MPLS connectivity to VPLS services terminating EVPN-VXLAN, SR OS supports the configuration of BGP-EVPN MPLS and BGP-EVPN VXLAN at the same time by adding two BGP instances to the service. Two BGP instances are supported in the same VPLS at most. As a rule, BGP-EVPN MPLS can use BGP instance 1 or 2, and BGP-EVPN VXLAN can only use BGP instance 1.

In a service with EVPN-VXLAN and EVPN-MPLS, the **config>service>vpls>bgp-evpn>mpls>bgp-instance 2** command allows the user to associate BGP-EVPN MPLS to a different instance than BGP-EVPN VXLAN, and therefore, have both encapsulations simultaneously enabled in the same service. When the two BGP instances are successfully added to the same VPLS service, the service behaves as follows:

- MAC/IP routes received on one instance will be "consumed" (accepted, imported, and installed in FDB) and re-advertised in the other instance, as long as the route is the best route for a specific MAC or MAC/IP.
- Inclusive multicast routes are independently generated for each BGP instance.
- From a data plane perspective, EVPN-MPLS and EVPN-VXLAN destinations are instantiated in different implicit Split-Horizon-Groups (SHGs) so that traffic can be forwarded between the two SHGs, but not between destinations of the same kind. For example, traffic coming from EVPN-MPLS cannot be forwarded to other destinations in the EVPN-MPLS SHG.

The following example shows a VPLS service configured on PE-2 with two BGP instances and both encapsulations, VXLAN and MPLS, configured at the same time:

```
configure
    service
        vpls 1 customer 1 create
            description "evpn-mpls and evpn-vxlan in the same service"
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 10:2
                route-target target:64500:1
            exit
            bgp 2
                route-distinguisher 10:1
                route-target target:64500:1
            exit
            bgp-evpn
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    bgp-instance 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
```

In the preceding example

- **bgp 1** or simply **bgp** is the default BGP instance.

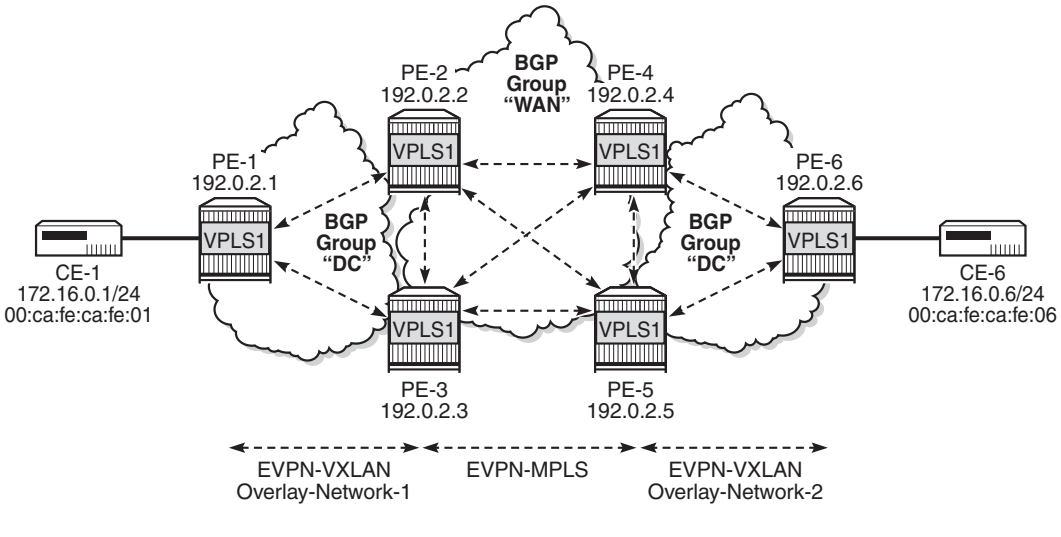
- **bgp 2** is the additional instance that is required when both BGP-EVPN VXLAN and BGP-EVPN MPLS are enabled in the service.
- The same commands supported under instance 1 exist for this second instance, with the following considerations:
  - **pw-template-binding** - the pseudowire (PW) template binding can only exist in instance 1; it is not supported in instance 2. Because no SAPs or SDP-bindings can exist in a VPLS service with two BGP instances, the **pw-template-binding** command is ineffective in this configuration.
  - **route-distinguisher** - the route distinguisher in both BGP instances must be different.
  - **route-target** - the route target in both instances can be the same or different.
  - **vsi-import** and **vsi-export** - import and export policies can also be defined for either BGP instance.
- The **mpls bgp-instance 2** command will assign the second BGP instance to MPLS. VXLAN always uses instance 1 (the default instance). The **bgp-evpn vxlan no shutdown** command will only be allowed if **bgp-evpn** is **shutdown** or if the BGP instance associated with MPLS has a different route distinguisher than the VXLAN instance (and vice versa).
- The **evi** can still be used for auto-derivation of RD/RT on **bgp-instance 1** and auto-derivation of RT (not RD) on **bgp-instance 2**. Auto-RD or an explicitly configured RD is needed in **bgp-instance 2**.

Configuration

Figure 277 shows the example topology that will be used throughout this chapter, as well as the BGP peering topology. PE-1, PE-2, and PE-3 simulate a data center, shown as Overlay-Network-1, where PE-2 and PE-3 are DC GWs. In the same way, PE-4, PE-5, and PE-6 simulate a remote data center, Overlay-Network-2. Inside each DC, EVPN-VXLAN is used.

The two DC GW pairs are connected by EVPN-MPLS; therefore, CE-1 and CE-6 are end-to-end connected by EVPN without any VLAN or PW hand-off, maintaining all the EVPN advantages across the DC Interconnect (DCI) network.

Figure 277 **EVPN-MPLS Interconnect for EVPN-VXLAN - Example Topology**



The example topology consists of six 7750 SR routers with the following initial configuration:

- Hybrid ports (they could have been network type too) are interconnecting the six PEs with configured router interfaces.
- The six PEs are running IS-IS and creating point-to-point adjacencies.
- Link LDP is configured in the core, among PE-2, PE-3, PE-4, and PE-5, while PE-1 and PE-6 are only running VXLAN.
- EVPN uses MP-BGP for exchanging reachability at service level. Therefore, BGP peering sessions must be established among the PEs for the EVPN family. Figure 277 shows the peering sessions established among the six PEs. Although usually a Route-Reflector (RR) is used in each DC and another RR in the WAN, in this example, there are direct peering sessions in each DC and in the WAN.

The following output shows the BGP configuration of PE-2. The BGP configuration on the rest of the DC GWs (PE-3, PE-4, and PE-5) would be similar:

```
configure
    router
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "WAN"
                type internal
                import "drop SOO-DCGW-23"
```

```
            export "allow only mpls and add SOO"
            neighbor 192.0.2.4
            exit
            neighbor 192.0.2.5
            exit
        exit
        group "DC"
            type internal
            import "drop SOO-DCGW-23"
            export "allow only vxlan and add SOO"
            neighbor 192.0.2.1
            exit
            neighbor 192.0.2.3
            exit
        exit
        no shutdown
```

Two different BGP groups are configured: DC and WAN. The DC group contains the DC neighbors (including the peer DC GW) and the WAN group contains the WAN neighbors. This grouping makes the use of policies easier. These policies will be explained in the section The Mandatory Use of BGP Policies in the Multi-Homed Anycast Solution.

The following output shows the BGP configuration of PE-1. PE-6 has a similar BGP configuration.

```
configure
    router
        bgp
            family evpn
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
```

VPLS Service Configuration

After the base infrastructure (interfaces, IGP, LDP in the core, and BGP) is configured, the services can be added. The configuration example in this section will use VPLS 1 as the service to be interconnected across the two DCs.

PE-1 and PE-6 have a regular EVPN-VXLAN configuration; DCI connectivity provided by EVPN-MPLS is completely transparent to them. The configuration of VPLS 1 in PE-1 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

See the EVPN for VXLAN Tunnels (Layer 2) chapter for a complete description of the EVPN-VXLAN commands.

The configuration on PE-2, PE-3, PE-4, and PE-5 (see Figure 277) enables EVPN-VXLAN and EVPN-MPLS in the same VPLS service. As an example, the VPLS 1 configuration on PE-2 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
            exit
            bgp 2
                route-distinguisher 64500:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    ingress-replication-bum-label
                    bgp-instance 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

As described in the Overview section, the preceding configuration enables the router to create EVPN-VXLAN and EVPN-MPLS destinations in the same VPLS service, but in different SHGs. In addition to the **bgp 2** and **bgp-instance 2** commands already described in the Overview section, a new command, **incl-mcast-orig-ip**, is added in the configuration. If configured, this command will change the originating IP address in the inclusive multicast routes (from the default system IP) for both BGP instances. The section Multi-homed Anycast Configuration for Dual BGP-Instance VPLS Services describes why this command is added.

The following section provides a detailed description of the expected behavior for EVPN routes that are imported and exported on dual BGP-instance VPLS services.

EVPN Route Handling in Dual BGP-Instance VPLS Services

This section describes how the BGP-EVPN routes are processed in dual BGP-instance services.

Usually, the router validates the received tunnel encapsulation (from the RFC 5512 Extended Community) with the configured encapsulation of the service/BGP-instance. Therefore, an EVPN-VXLAN route will not get imported into the BGP-EVPN MPLS instance and vice-versa. This is also how the different EVPN route types are handled in dual BGP-instance services:

- **Route Type 1 - Auto-Discovery Routes**

  AD per-EVI routes are never generated by services with two BGP instances in SR OS release 14.0.R5 (because no Ethernet Segment (ES) can be associated with the dual BGP-instance service). However, AD per-EVI routes can still be received from the EVPN-MPLS peers and are processed as usual. Therefore, a VPLS service with two BGP instances will still support aliasing/backup and AD per-ES checking procedures for a remote multi-homed ES, as described in the EVPN for MPLS Tunnels chapter. However, in the example in Figure 277, PE-6 does not have any local multi-homed ES configured; therefore, no AD per-EVI routes are present in this example.
- **Route Type 2 - MAC/IP Routes**

In SR OS release 14.0.R5, no SAP/SDP-bindings are allowed in services with two BGP instances; therefore, all the MACs that are installed in the DC GWs FDBs are coming from EVPN. MAC/IP routes received on one of the two BGP instances will be imported and the MACs added to the FDB according to the existing selection rules. If the MAC is active, (therefore installed in the FDB) it will be re-advertised in the other BGP instance with the new BGP attributes of the other BGP instance (new route target if different, new route distinguisher, and so on). The **mac-advertisement** command will govern the advertisement of any MACs in either BGP instance.

The MAC/IP route redistribution across BGP instances is performed according to the following rules:

- A MAC route is redistributed only if it is the best route according to the EVPN selection rules in the EVPN for MPLS Tunnels chapter.
- Assuming a specific MAC route is the best one and has to be redistributed, the MAC/IP information along with the sticky bit is propagated in the redistribution.
- A change in the MAC/IP route sequence number or sticky bit in one instance is updated in the other instance, as long as that route is the best MAC route for the route key.
- When a MAC moves within the EVPN-VXLAN (or the EVPN-MPLS) network, the MAC route is received on the same BGP instance where it was previously received, but now with a higher sequence number. In this case, the MAC route will be redistributed with the new sequence number. However, a router with two BGP instances in the same service will not detect any duplicate MAC on the EVPN-VXLAN and EVPN-MPLS networks.

As an example, the following output shows the debug of a MAC/IP route received on PE-2, on the BGP instance for EVPN-VXLAN on VPLS 1, and how the route is re-advertised to the BGP instance used for MPLS (with a different next-hop, route distinguisher, label, and BGP tunnel encapsulation):

```
A:PE-2#
7 2016/09/29 08:13:44.60 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:ca:fe:ca:fe:01, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
```

```
        bgp-tunnel-encap:VXLAN
"


8 2016/09/29 08:13:44.60 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 96
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 64500:2 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:ca:fe:ca:fe:01, IP len: 0, IP: NULL, label1: 4194272
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        origin:64500:23
        target:64500:1
        bgp-tunnel-encap:MPLS
"


9 2016/09/29 08:13:44.60 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 96
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 64500:2 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:ca:fe:ca:fe:01, IP len: 0, IP: NULL, label1: 4194272
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        origin:64500:23
        target:64500:1
        bgp-tunnel-encap:MPLS
"
```

- **Route Type 3 - Inclusive Multicast Routes**

  EVPN Inclusive Multicast Ethernet tag (IMET) routes are generated independently for each BGP instance with the correct BGP tunnel encapsulation extended community and the tunnel type associated to the BGP instance; for example, Ingress Replication (IR), P2MP mLDP, or Assisted Replication (AR):

  – On the EVPN-VXLAN BGP instance, IR or AR IMET routes are supported.

    - When **assisted-replication replicator** is enabled and the received VXLAN broadcast and multicast packets contain an IP DA = AR-IP, the DC GW will send the packets back to VXLAN (but not to the VXLAN termination end-point (VTEP) from where the packet is received) in

addition to the EVPN-MPLS destinations.

- If **assisted-replication replicator** is used on the DC GWs, the AR-IP (**configure>service>system>vxlan>assisted-replication-ip**) must be a loopback different from the router's system IP and the configured **bgp-evpn>incl-mcast-orig-ip**. The two AR-IPs in the DC GW pair do not need to be the same IP address.

– On the EVPN-MPLS BGP instance, IR, P2MP mLDP, or composite IMET routes are supported.

– Following is the behavior when the **incl-mcast-orig-ip** command is used:

- The configured IP in the **incl-mcast-orig-ip** command is encoded in the originating IP field of the IMET routes for IR, P2MP, and composite routes for both BGP instances.
- The originating IP field of the IMET AR routes is still derived from the configured **service>system>vxlan>assisted-replication-ip** value.

– The received IMET routes will be processed in the following way depending on their type:

- IMET-IR routes: the EVPN destination (MPLS or VXLAN) is set up based on the NLRI next-hop.
- IMET-P2MP routes: the Provider Tunnel Attribute (PTA) tunnel ID will be used to join the mLDP tree (as mLDP FEC in the LDP mapping messages).
- IMET-P2MP-IR (composite) routes: the PTA tunnel ID is used to join the mLDP tree. The NLRI next-hop is used to build the EVPN destination.
- IMET-AR routes: the NLRI next-hop is used to build the EVPN-VXLAN destination.

– Upon reception of two IMET routes with similar information, the router behaves as follows:

- If the router receives two IMET routes with the same originating IP, different RDs, and different NLRI next-hops, it will set up two EVPN destinations, one to each next-hop.
- If the router gets two IMET routes with the same originating IP, different RDs, but the same next-hop, it will set up only one EVPN destination.
- The router will not set up an EVPN destination to its DC GW peer if the received originating IP matches its own originating IP, regardless of whether the local RD and the remote RD are the same or different. This enables the use of the redundant anycast solution that is described in the following section: Multi-homed Anycast Configuration for Dual BGP-Instance VPLS Services.

- **Route Type 4 - ES Routes**

ESs are supported in routers where dual BGP-instance services exist. However, since dual BGP-instance VPLS services do not support SAP/SDP-bindings, ESs and ES routes are not relevant to these types of services in SR OS release 14.0.R5.

- **Route type 5 - IP-prefix routes**

  R-VPLS services are not supported along with dual BGP instances; therefore, IP-prefix routes are neither generated nor processed by the service.

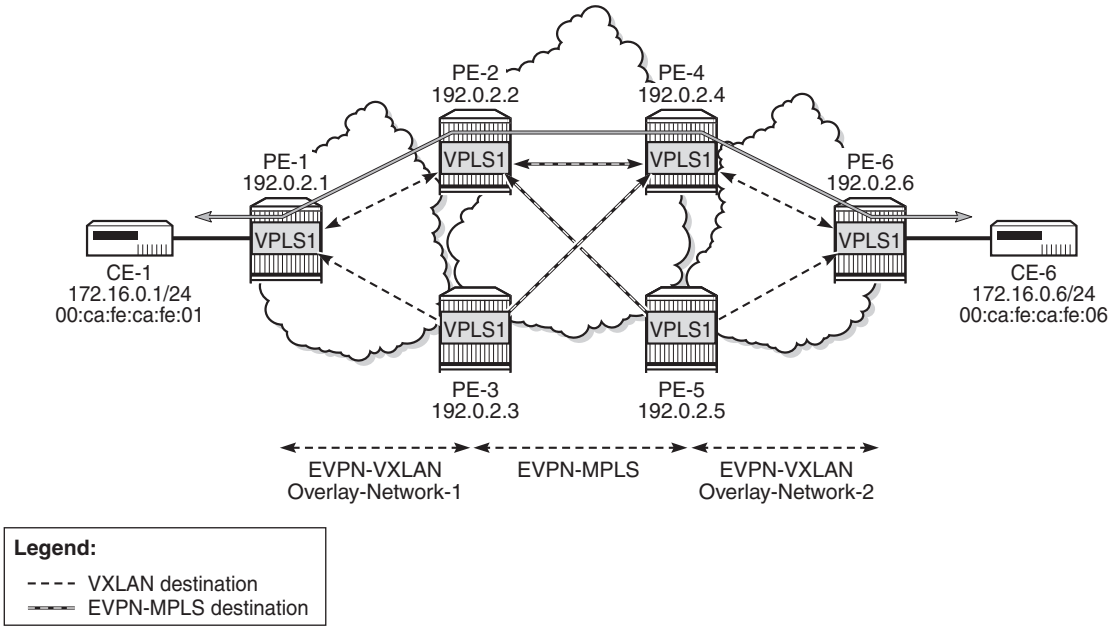
Multi-homed Anycast Configuration for Dual BGP-Instance VPLS Services

Services with EVPN-MPLS and EVPN-VXLAN SHGs are specified in draft-ietf-bess-dci-evpn-overlay and the associated multi-homing solution is also described in the same draft. That multi-homing solution is based on an interconnect ES that allows all-active and single-active multi-homed EVPN networks as well as local attachment circuits in the DC GWs (SAP/SDP-bindings).

In SR OS release 14.0.R5, interconnect ESs are not supported. Therefore, an anycast solution is used to provide redundancy. This anycast solution is based on the two PE DC GWs in the redundant pair being configured to advertised MAC/IP and IMET routes with the same route key, so that the remote PEs will only pick up one of the two anycast DC GWs when sending unicast or BUM traffic, and no loop or packet duplication is created.

Figure 278 is an example of how multi-homing can be achieved for dual BGP-instance VPLS services. The figure also shows the EVPN destinations created and their direction (see the arrows). For instance, only one EVPN multicast destination is created for PE-1, PE-2, or PE-4. Therefore, BUM traffic sent by CE-1 will be sent via PE-2, PE-4, and PE-6 only, and no duplication or loops occur.

Figure 278 **EVPN Destinations Created on Multi-Homed Anycast DC GWs**



The following output shows the VPLS 1 configuration on PE-2 and PE-3 so that this anycast redundancy can be realized. The route distinguishers as well as the **incl-mcast-orig-ip** addresses must match between the two PEs in the redundant pair. VPLS 1 is configured on PE-2 as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
            exit
            bgp 2
                route-distinguisher 64500:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    ingress-replication-bum-label
                    bgp-instance 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
```

```
            exit
        exit
        no shutdown
```

The VPLS 1 configuration on PE-3 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
            exit
            bgp 2
                route-distinguisher 64500:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    ingress-replication-bum-label
                    bgp-instance 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
```

The VPLS 1 configuration on PE-4 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 64501:1
            exit
            bgp 2
                route-distinguisher 64501:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    ingress-replication-bum-label
                    bgp-instance 2
                    auto-bind-tunnel
```

```
                    resolution any
                exit
                no shutdown
            exit
        exit
        no shutdown
```

The VPLS 1 configuration on PE-5 is as follows:

```
configure
    service
        vpls 1 customer 1 create
            vxlan vni 1 create
            exit
            bgp
                route-distinguisher 64501:1
            exit
            bgp 2
                route-distinguisher 64501:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 1
                vxlan
                    no shutdown
                exit
                mpls
                    ingress-replication-bum-label
                    bgp-instance 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
```

Based on the preceding configuration example, the DC GWs behavior in this scenario is as follows:

- PE-2 and PE-3 both send IMET IR routes to the other PEs with the same route key but a different next-hop. The route key in IMET routes comprises [RD, ethernet-tag, originator-IP/length], which in this case will be [64500:1, 0, 23.23.23.23/32] for the EVPN-VXLAN IMET routes and [64500:2, 0, 23.23.23.23/32] for the EVPN-MPLS IMET routes.
- In the same way, PE-2 and PE-3 both send MAC/IP routes to the other PEs with the same route key but a different next-hop. The route key comprises [RD, ethernet-tag, MAC/MAC-length, IP/IP-length].

The configuration of the same **incl-mcast-orig-ip** address and RDs in both DC GWs enables the anycast solution due to the following:

- The configured originating IP (for example, 23.23.23.23 in PE-2 and PE-3) is not required to be a reachable IP address, which forces the remote PEs (or RRs if they exist) to select only one of the two DC GWs for BUM traffic (based on regular BGP selection). In this example, the remote PEs will select the PE-2 IMET route and create only one destination. The following output shows the IMET routes received by PE-1 (only the PE-2 route is used) and the created EVPN-VXLAN destination to PE-2. The same behavior could have been shown in the rest of the PEs.

```
A:PE-1# show router bgp routes evpn inclusive-mcast
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.          OrigAddr
      Tag                  NextHop
-------------------------------------------------------------------------------
u*>i  64500:1              23.23.23.23
      0                    192.0.2.2

*>i   64500:1              23.23.23.23
      0                    192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================


A:PE-1# show service id 1 vxlan
===============================================================================
Vxlan Src Vtep IP: N/A
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1
Creation Origin: manual
Assisted-Replication: none
RestProtSrcMacAct: none

===============================================================================
VPLS VXLAN service Network Specifics
===============================================================================
Ing Net QoS Policy : none                          Vxlan VNI Id     : 1
Ingress FP QGrp    : (none)                        Ing FP QGrp Inst : (none)

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                          Egress VNI  Num. MACs   Mcast Oper  L2
                                                                    State PBR
-------------------------------------------------------------------------------
192.0.2.2                             1           0           BUM   Up    No
-------------------------------------------------------------------------------
```

```
Number of Egress VTEP, VNI : 1
------------------------------------------------------------------------------
==============================================================================
```

- Due to the same RD and originating IP configured on PE-2 and PE3 (similarly in PE-4 and PE-5), the DC GW redundant PEs will never establish an EVPN destination between each other. PE-2 only sets up EVPN multicast destinations to PE-1 and PE-4, as follows:

```
A:PE-2# show service id 1 vxlan
==============================================================================
Vxlan Src Vtep IP: N/A
==============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1
Creation Origin: manual
Assisted-Replication: none
RestProtSrcMacAct: none

==============================================================================
VPLS VXLAN service Network Specifics
==============================================================================
Ing Net QoS Policy : none                          Vxlan VNI Id     : 1
Ingress FP QGrp    : (none)                        Ing FP QGrp Inst : (none)

==============================================================================
Egress VTEP, VNI
==============================================================================
VTEP Address                          Egress VNI  Num. MACs   Mcast Oper  L2
                                                                    State PBR
------------------------------------------------------------------------------
192.0.2.1                             1           0           BUM   Up    No
------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
------------------------------------------------------------------------------
==============================================================================


A:PE-2# show service id 1 evpn-mpls

==============================================================================
BGP EVPN-MPLS Dest
==============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                 Transport
------------------------------------------------------------------------------
192.0.2.4       262141        0           Yes             09/29/2016 08:10:15
                ldp
------------------------------------------------------------------------------
Number of entries : 1
------------------------------------------------------------------------------
==============================================================================
---snip---
```

- Likewise, when the two redundant PEs receive the same MAC/IP route, they will both re-advertise it with the same route key, forcing the remote PEs to pick up only one of the two (based on regular BGP selection) and create only one EVPN destination (if different from the multicast destination). In the following example, PE-6 advertised the CE-6 MAC address, that is, re-advertised by PE-4/PE-5 and then by PE-2/PE-3, but only one of the routes is selected at each hop. The following output shows that PE-1 selects the PE-2 MAC/IP route (see the "used" flag) and uses the existing EVPN destination to PE-2:

```
A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.          MacAddr           ESI
      Tag                  Mac Mobility      Label1
                           Ip Address
                           NextHop
-------------------------------------------------------------------------------
u*>i  64500:1              00:ca:fe:ca:fe:06 ESI-0
      0                    Seq:0             VNI 1
                           N/A
                           192.0.2.2

*>i   64500:1              00:ca:fe:ca:fe:06 ESI-0
      0                    Seq:0             VNI 1
                           N/A
                           192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================

A:PE-1# show service id 1 fdb detail
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
1         00:ca:fe:ca:fe:06 vxlan:                   Evpn     09/29/16 09:09:36
                            192.0.2.2:1
-------------------------------------------------------------------------------
No. of MAC Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static
===============================================================================

A:PE-1# show service id 1 vxlan
```

```
===============================================================================
Vxlan Src Vtep IP: N/A
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1
Creation Origin: manual
Assisted-Replication: none
RestProtSrcMacAct: none

===============================================================================
VPLS VXLAN service Network Specifics
===============================================================================
Ing Net QoS Policy : none                          Vxlan VNI Id     : 1
Ingress FP QGrp    : (none)                        Ing FP QGrp Inst : (none)

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                          Egress VNI  Num. MACs   Mcast Oper  L2
                                                                    State PBR
-------------------------------------------------------------------------------
192.0.2.2                             1           1           BUM   Up    No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
```

- As shown in the preceding outputs, the EVPN destinations are always created to the IMET or MAC/IP route's BGP next-hops, which are still the system IP of the routers (they could have also been a loopback address). The BGP next-hops need to be reachable in their respective network: DC or WAN.

The Mandatory Use of BGP Policies in the Multi-Homed Anycast Solution

BGP policies must be configured in a multi-homed anycast solution, such as the one described in the previous section. Without policies, the following undesired behavior would happen:

- IMET routes with VXLAN encapsulation would be sent to the BGP peers in the MPLS network and IMET routes with MPLS encapsulation sent to BGP peers in the DC. The configured BGP policies will avoid that and make sure that the VXLAN routes are only sent to the DC and MPLS routes only to the WAN.

- MAC/IP routes received in the VXLAN BGP instance of a DC GW would be re-advertised to the redundant DC GW in the MPLS BGP instance and the redundant DC GW would re-advertise the same MAC again into the VXLAN instance, creating a control plane loop. The same thing would happen for MAC/IP routes received in an MPLS BGP instance. The configured BGP policies will prevent a DC GW from re-advertising MAC/IP routes received from the redundant DC GW.

While service-level BGP policies (**config>service>vpls>bgp>vsi-import/export**) may have been configured to prevent these loops and misbehavior, the use of BGP peer-level policies (**config>router>bgp>group>import/export**) is recommended due to the following reasons:

- Simplicity - BGP peer-level policies do not require any extra configuration at the service level, only at the BGP level.
- Scalability - BGP peer-level policies scale better than VSI-level policies, because the number of services where the VSI policies should be configured may be significant.

The following policies are configured in the example used in this chapter. No policies are needed in PE-1 and PE-6; only the DC GWs must be configured.

Following are the policies and how they are applied in PE-2:

```
configure
    router
        policy-options
            begin
            community "mpls" members "bgp-tunnel-encap:MPLS"
            community "vxlan" members "bgp-tunnel-encap:VXLAN"
            community "SOO-DCGW-23" members "origin:64500:23"

/* "drop SOO-DCGW-23" will drop any EVPN route that is received from PE-3,
the other DC GW in the pair. */

            policy-statement "drop SOO-DCGW-23"
                entry 10
                    from
                        community "SOO-DCGW-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit

/* "allow only mpls and add SOO" has a twofold objective: avoids sending EVPN-VXLAN
routes to the MPLS network and marks the advertised EVPN routes with a Site-Of-
Origin extended community that identifies the DC GW pair. */

            policy-statement "allow only mpls and add SOO"
                entry 10
                    from
```

```
                        community "vxlan"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
            exit

/* In the same way, "allow only vxlan and add SOO" avoids sending EVPN-MPLS routes
to the VXLAN network and marks the EVPN routes with a Site-Of-Origin extended
community that identifies the DC GW pair. */

            policy-statement "allow only vxlan and add SOO"
                entry 10
                    from
                        community "mpls"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
            exit
            commit
        exit

/* The policies are properly applied at group level. */

        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "WAN"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only mpls and add SOO"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            group "DC"
```

```
                type internal
                import "drop SOO-DCGW-23"
                export "allow only vxlan and add SOO"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

The same policies are configured on PE-3 (including the addition and filtering of the same Site-Of-Origin because PE-3 is part of the same DC GW pair):

```
configure
    router
        policy-options
            begin
            community "mpls" members "bgp-tunnel-encap:MPLS"
            community "vxlan" members "bgp-tunnel-encap:VXLAN"
            community "SOO-DCGW-23" members "origin:64500:23"
            policy-statement "drop SOO-DCGW-23"
                entry 10
                    from
                        community "SOO-DCGW-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "allow only mpls and add SOO"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
            exit
            policy-statement "allow only vxlan and add SOO"
                entry 10
                    from
                        community "mpls"
                        family evpn
                    exit
                    action drop
                    exit
                exit
```

```
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
            exit
            commit
        exit
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "WAN"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only mpls and add SOO"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            group "DC"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only vxlan and add SOO"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
            no shutdown
        exit
```

PE-4 and PE-5 use the same BGP peer policies, but using a Site Of Origin extended community identifying the PE-4/PE-5 pair instead of the PE-2/PE-3 pair:

```
configure
    router
        policy-options
            begin
            community "mpls" members "bgp-tunnel-encap:MPLS"
            community "vxlan" members "bgp-tunnel-encap:VXLAN"
            community "SOO-DCGW-45" members "origin:64500:45"
---snip---
```

Dual BGP Instance VPLS Service Caveats

When two BGP instances are enabled on the same VPLS service, the following considerations apply:

- SAPs or SDP-bindings are not supported (therefore, no pw-template-binding is needed in the service). Any attempt to add a SAP or SDP-binding to a service with two BGP instances will be blocked by the CLI. For example:

```
configure service vpls 1 sap 1/2/1:1 create
MINOR: SVCMGR #7888 Cannot be configured/enabled with EVPN -
 saps not allowed when both vxlan and evpn-mpls are enabled
```

- Services that are not supported: R-VPLS, M-VPLS, I-VPLS, B-VPLS, or Etree VPLS
  - A consequence of not supporting R-VPLS is that no routes type 5 (IP-Prefix routes) are supported on dual BGP-instance services.
- Proxy-ARP/ND is not supported. ARP request or neighbor solicitation messages are usually only expected on SAP/SDP-bindings though.
- BGP multi-homing is not supported.
- Although the Assisted-Replication feature is supported on dual BGP-instance VPLS services, the Assisted-Replication configuration is only relevant to the VXLAN destinations. See section EVPN Route Handling in Dual BGP-Instance VPLS Services for some considerations about how EVPN handles IMET AR routes.
- The configuration of ESs is not supported for services with dual BGP instances (they cannot be configured because these services do not support SAP/SDP-bindings).

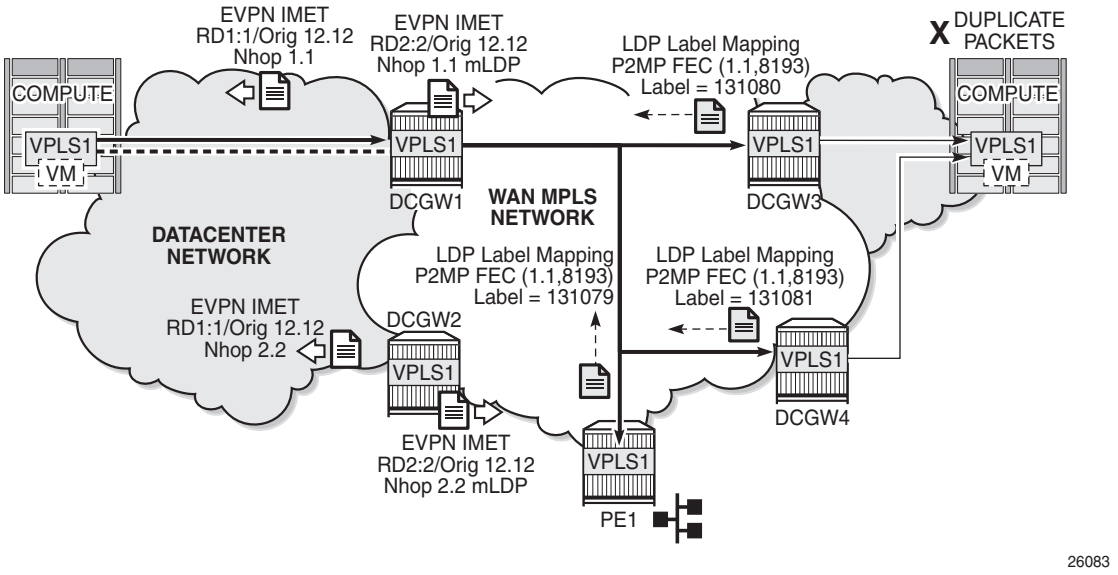
In addition to the preceding restrictions, some commands have a specific behavior when two BGP instances are configured:

- **config>service>vpls>bgp-evpn>[no] mac-advertisement** enables/disables the re-advertisement of MAC/IP routes in a BGP instance for MACs that have been learned in the other BGP instance in the service.
- **config>service>vpls>bgp-evpn>[no] unknown-mac-route** enables/disables the advertisement of the unknown MAC route (MAC 00:..:00) on the BGP-EVPN VXLAN instance. The unknown-mac-route is never sent to the BGP-EVPN MPLS instance.

The Use of Provider-tunnels on Multi-Homed Anycast Solutions

The use of provider-tunnels in dual BGP-instance VPLS services connecting multiple DCs is not recommended. Figure 279 shows the case where the same BGP-EVPN service is configured in redundant anycast DC GWs and mLDP is used in the MPLS instance. In this case, packet duplication may occur if the configuration is not done carefully.

Figure 279 **Use of Provider-tunnels Between Anycast DC GWs Create Packet Duplication**



When mLDP is used along with multiple anycast multi-homing DC GWs to send BUM traffic to remote PEs, but no BUM traffic between DCs is needed, the same originating IP must be used on all the DC GWs; otherwise, packet duplication may happen. In the example in Figure 279, each pair of DC GWs, DCGW1/DCGW2 and DCGW3/DCGW4, is configured with a different originating IP (**config>service>vpls>bgp-evpn> incl-mcast-orig-ip**):

- DCGW3 and DCGW4 will receive the IMET route with the same route key from DCGW1 and DCGW2.
- DCGW3 and DCGW4 will select only one route, which will usually be the same; for example, the DCGW1 IMET route.
- Because of that, both DCGW3 and DCGW4 will join the mLDP tree with root in DCGW1, creating packet duplication when DCGW1 sends BUM traffic.
- Remote PE nodes with a single MPLS instance will join the mLDP tree without any issue.

To avoid the packet duplication shown by the example of Figure 279, the same originating IP may be configured in the four DCGWs, while the RD is still different per pair. By doing that:

- In the example of Figure 279, DCGW3 and DCGW4 will never join any mLDP tree sourced from DCGW1 or DCGW2. This will prevent any packet duplication because a router will ignore IMET routes received with its own originating IP, regardless of the RD.
- PE-1 (a remote EVPN-MPLS PE) will still join the mLDP trees from the two DCs.
- The preceding configuration allows the use of mLDP as long as no BUM traffic is required between the two DCs. If BUM traffic is required between DCs, IR must be used.

Troubleshooting and Debugging

The following show and debug commands can be used in dual BGP-instance VPLS services:

- show router bgp routes evpn (and filters)
- show service evpn-mpls [<TEP ip-address>]
- show service vxlan [<TEP ip-address>]
- show service id bgp-evpn
- show service id evpn-mpls (and modifiers)
- show service id vxlan
- debug router bgp update
- log-id 99

See chapter EVPN for MPLS Tunnels and EVPN for VXLAN Tunnels (Layer 2) for a detailed description of these commands.

Also, in dual BGP-instance VPLS services, the **show service id bgp <bgp-instance>** command may help see the BGP parameters of each individual BGP instance (where instance 1 is always associated with VXLAN):

```
A:PE-2# show service id 1 bgp
  - bgp [<bgp-instance>]

 <bgp-instance>       : [1..2]

A:PE-2# show service id 1 bgp 1
===============================================================================
BGP Information
===============================================================================
```

```
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:1
Oper Route Dist      : 64500:1
Oper RD Type         : configured
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64500:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64500:1
PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================

A:PE-2# show service id 1 bgp 2
===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:2
Oper Route Dist      : 64500:2
Oper RD Type         : configured
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64500:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64500:1
-------------------------------------------------------------------------------
===============================================================================
```

Conclusion

As service providers deploy EVPN-MPLS in the network for Ethernet local area network (ELAN) and Ethernet point-to-point (ELINE) services, the use of EVPN-MPLS to interconnect data centers is becoming a popular option. Based on draft-ietf-bess-dci-evpn-overlay, SR OS supports the connectivity of Layer 2 EVPN-VXLAN services to an EVPN-MPLS network. To implement that EVPN-MPLS Data Center Interconnect (DCI) solution, VPLS services support dual BGP instances, where EVPN-VXLAN and EVPN-MPLS can coexist simultaneously in the same VPLS service. This chapter describes the configuration of such dual BGP-instance VPLS services and how to deploy them in a redundant anycast DC GW configuration.

Fully Dynamic VSD Integration Model

This chapter provides information about fully dynamic virtualized service directory (VSD) integration model.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to 7750 SR-7/12, 7750 SR-a4/8, 7750 SR-12E, 7450 ESS-7/12, XRS-20/16c, and 7750 SR-c4/12. Fully dynamic VSD integration model is a CPM-only feature that requires a multi-core CPM (CPM3 required for 7750 SR-7/12 and 7450 ESS-7/12).

If this feature is used in conjunction with Ethernet Virtual Private Network - Virtual Extensible Local Area Network (EVPN-VXLAN), chassis mode D and IOM3-XP/IMM or higher-based line cards are required for the data plane.

Software requirements for this feature are SR OS release 13.0.R4 or later and Nuage Virtualized Services Platform (VSP) release 3.2.R1 or later. This configuration was tested on SR OS release 13.0.R4 and Nuage VSP release 3.2.R3.

Note: Fully dynamic extensible messaging and presence protocol (XMPP) provisioning is not supported along with the dynamic business services feature in release 13.0. Both features are mutually exclusive.

Note: Provisioning of filter entries from Virtualized Services Directory (VSD) is not supported in SR OS 13.0.

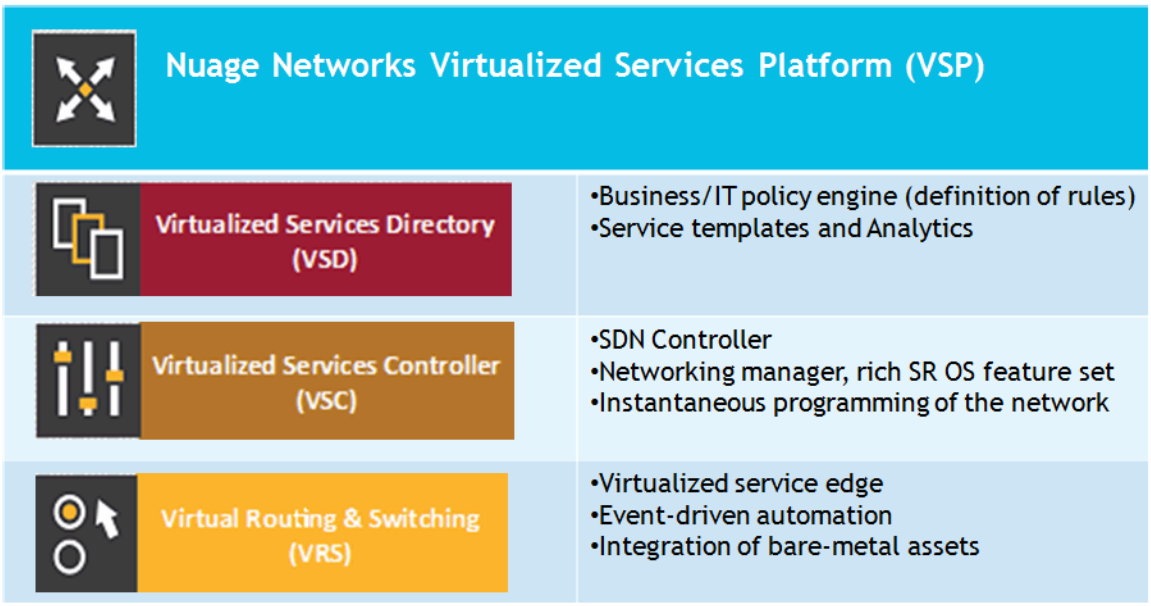
Nuage VSP conceptual knowledge and Nuage VSD operational knowledge are prerequisites. See the Nuage VSP user documentation for more information.

Overview

The Nuage VSP is a Software-Defined Networking (SDN) solution that provides data center (DC) network virtualization and automatically establishes connectivity between compute resources upon their creation. Leveraging programmable business logic and a powerful policy engine, the Nuage VSP provides an open and highly responsive solution that scales to meet the stringent needs of massive multi-tenant DCs. The Nuage VSP can be deployed over an existing DC IP network fabric, and has three main components:

Virtualized Services Directory (VSD), Virtualized Services Controller (VSC), and Virtual Routing and Switching (VRS), as displayed in Figure 280

Figure 280 **Nuage VSP overview**



Virtualized Services Directory (VSD)

The Nuage VSD is a programmable policy and analytics engine. It provides a flexible and hierarchical network policy framework that enables IT administrators to define and enforce resource policies in a user-friendly manner.

The VSD contains a multi-tenant service directory, which supports role-based administration of users, computing, and network resources. The VSD also manages network resource assignments such as IP addresses and ACLs.

For the purpose of service assurance, the VSD allows the definition of sophisticated statistics rules, such as collection frequencies, rolling averages, and samples, as well as Threshold Crossing Alerts (TCA). When a TCA occurs, it will trigger an event that can be exported to external systems through a generic messaging bus. Statistics are aggregated over hours, days, and months, and stored in a Hadoop® analytics cluster to facilitate data mining and performance reporting.

The VSD runs as a number of processes in a virtual machine (VM) environment.

Virtualized Services Controller (VSC)

The Nuage VSC is an SDN controller. It functions as the robust network control plane for DCs, maintaining a full view of per-tenant network and service topologies. Through the VSC, virtual routing and switching constructs are established to program the network forwarding plane, the Nuage VRS, using the OpenFlow protocol.

The VSC communicates with the VSD policy engine using Extensible Messaging and Presence Protocol (XMPP). An ejabberd XMPP server/cluster is used to distribute messages between the VSD and VSC entities. Multiple VSC instances can be interconnected within and across DCs by leveraging Multi-Protocol Border Gateway Protocol (MP-BGP).

The VSC is based on the Service Router Operating System (SR OS) and runs in a virtual machine environment.

Virtual Routing and Switching (VRS)

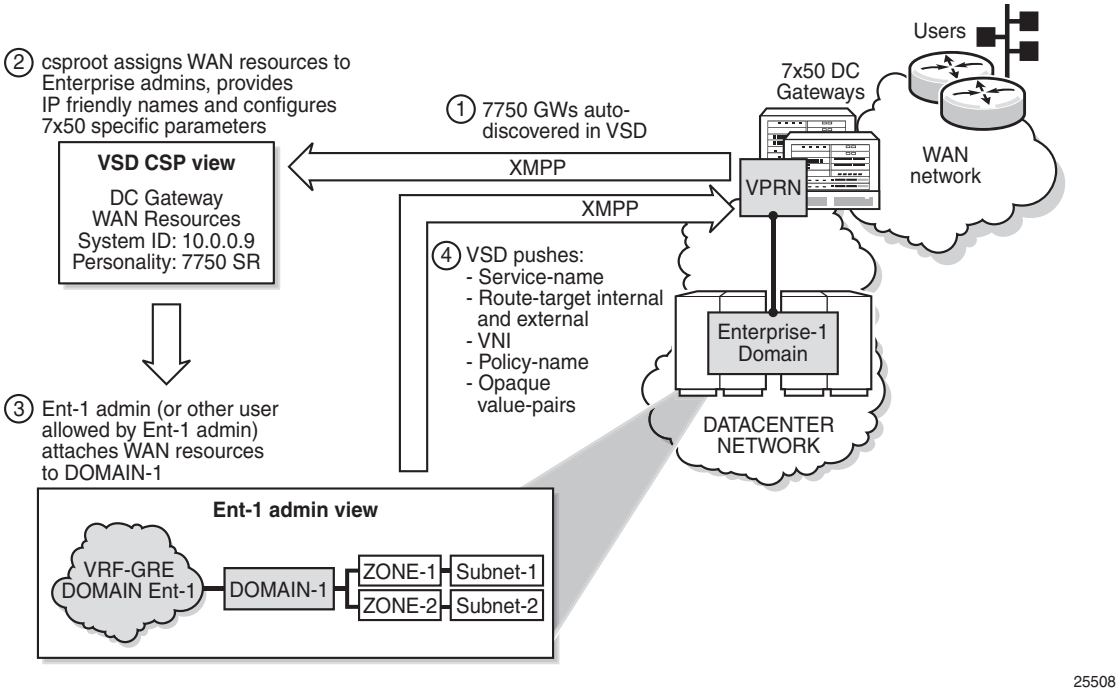
The Nuage VRS component is an enhanced Open vSwitch (OVS) implementation that constitutes the network forwarding plane. It encapsulates and de-encapsulates user traffic, enforcing L2 to L4 traffic policies as defined by the VSD. The VRS tracks Virtual Machine (VM) creation, migration, and deletion events in order to dynamically adjust network connectivity.

DC Gateway automated service provisioning

The first phase of VSD-7x50 integration was introduced in SR OS 12.0.R4. This phase included the development of an XMPP interface on the 7x50 SR and the integration in the Nuage XMPP architecture. This so-called Static + Dynamic (S-D) provisioning model allows the auto-provisioning of VPLS and VPRN route targets, as well as VPLS VNI (VXLAN Network Identifiers) on the 7x50 SR through the XMPP interface and the VSD interaction. The prerequisite in this model is the pre-configuration of the VPLS and VPRN services on the 7x50 through CLI or SNMP. This model is intended to be used in DC Gateways where the WAN and the DC are managed by different administrative entities. The DC administrator will use VSD to "attach" the already configured VPLS or VPRN service to the L2 or L3 domain in the DC.

The second phase of VSD-7x50 integration was introduced in SR OS 13.0R4. This phase supports the Fully Dynamic (F-D) provisioning model. The goal of this model is to avoid the prerequisite of pre-configuring the services on the 7x50 SR existing in the S-D provisioning model, since this model assumes that the service is completely owned by the DC administrator. The entire service will be auto-generated on the 7x50 SR as a result of the interaction with the VSD. Figure 281 shows the workflow of the F-D provisioning model.

Figure 281 **DC Gateway fully dynamic provisioning workflow**



1. As soon as the XMPP server is configured on the 7x50 DC Gateway, it is auto-discovered by the VSD.
2. The Cloud Service Provider (CSP) root user creates WAN services and assigns these resources to Enterprise administrators; for example, to the Ent-1 admin.
3. The Ent-1 admin sees the WAN service in the infrastructure resources and assigns permissions to certain user groups in Ent-1, who can consume these WAN resources by connecting to an L2/L3 domain.
4. As soon as the WAN service is added to an L2/L3 domain, the VSD pushes a list of parameters to the 7x50 DC Gateway, which uses a python script to construct the configuration of the WAN service. The list of parameters sent to the 7750 routers can include:
   - **service-name** (Service ID field in the WAN Service GUI) - used as VSD domain in the CLI
   - **config-type** (Config Type field in the WAN Service GUI) - DYNAMIC for F-D XMPP provisioning
   - **service-type** (based on combination of the Service Type field and IRB check box in the WAN Service GUI) - possible values: L2DOMAIN, L2DOMAIN-IRB, VRF-GRE, or VRF-VXLAN
   - **name** (name of the L2 domain in the VSD to which the WAN service is assigned, or BackHaulSubnet in the case of service-type VRF-VXLAN)
   - **service-policy** (service Policy in the WAN Service GUI field) - should match the python policy configured on the 7x50 DC Gateway
   - **vn-id** (VNI used for the Nuage overlay service) - dynamically supplied for VXLAN WAN services
   - **RT-I** (internal Route Target used for the Nuage overlay service)
   - **RT-E** (ext. Route Target in the WAN Service GUI field)
   - **metadata** (list of opaque parameters supplied in the Metadata section of the WAN Service GUI)

The dynamic provisioning of parameters is provided for the following VSD domain types (configured in the 7x50 DC Gateway):

l2-domain	To attach a service at the gateway to an L2 (Ethernet) domain in the data center with no routing at the gateway, a VPLS service must be associated with a vsd-domain of type l2-domain.
l2-domain-irb	To attach a service at the gateway to an L2 (Ethernet) domain in the data center with routing at the gateway, an R-VPLS service should be associated with a vsd-domain of type l2-domain-irb.
vrf-gre	To attach a service at the gateway to an L3 domain (with GRE transport) in the data center, a VPRN service should be associated with a vsd-domain of type vrf-gre.

vrf-vxlan To attach a service at the gateway to an L3 domain (with VXLAN transport) in the data center, an R-VPLS service (with ip-route-advertisement enabled and linked to an EVPN-tunnel) should be associated with a vsd-domain of type vrf-vxlan.

This chapter will show examples of l2-domain, l2-domain-irb, and vrf-vxlan service type F-D provisioning, and focuses mostly on the 7x50 DC Gateway configuration. For a more detailed F-D provisioning workflow on the VSD UI, refer to the VSP User Guide.

Python script

The XMPP parameters supplied by the VSD are parsed by a python script on the 7x50 DC Gateway that dynamically provisions the VPLS and/or VPRN services provided for the Nuage overlay services.

The python script generates an executable CLI script based on the information received in the XMPP attributes. Three dynamic data service functions can be specified: **setup**, **modify**, or **teardown**. A fourth action, **revert**, is automatically invoked when the modify action fails:

- **setup** function: output = CLI to create a new dynamic data service.
- **teardown** function: output = CLI to delete an existing dynamic data service.
- **modify** function: output = CLI to change the parameters of an existing dynamic data service
- **revert** function: output = CLI to rollback the dynamic data service modify function actions in case of a modify failure

The python script uses the alc.dyn python module that contains a number of functions required to set up dynamic data services. To use the alc.dyn module, it must be imported into the python script:

```
from alc import dyn
```

The alc.dyn module contains a number of functions. Relevant alc.dyn functions for F-D XMPP provisioning are listed here:

- **dyn.action**(dictionary)
- **dyn.add_cli**(string)
- **dyn.select_free-id**(service-id)

The next sections provide a basic description of these functions. The alc.dyn module contains other functions that are not relevant for this feature. For a full list of the alc.dyn functions together with an extensive explanation of each function, refer to the RADIUS-Triggered Dynamic Data Service Provisioning chapter.

The trigger in the python script to execute a specific function is by calling the internal function **dyn.action**(d), where "d" is a python dictionary:

- d = { key : value, key : value, key : value, … , key : value }

For F-D XMPP provisioning, only 1 key:value pair is used and the key string must be set to "script".

The value is a tuple with the following comma separated values:

- (setup-function, modify-function, revert-function, teardown-function)

Setup and teardown functions are mandatory. Modify and revert functions are optional. If a modify function is defined, the revert function must also be defined. If no modify/revert function is required, the keyword "None" should be used instead.

The following two combinations are supported for F-D Dynamic XMPP provisioning:

1. without modify function:

```
d = {"script" : (setup_script, None, None, teardown_script)}
dyn.action(d)
```

2. with modify function (allows for changes in the WAN service on the VSD while the service is assigned to a domain):

```
d = {"script" : (setup_script, modify_script, revert_script, teardown_script)}
dyn.action(d)
```

When the configuration for a new service-name is received from the VSD, the vsd parameters and the opaque parameters string are concatenated into a single dictionary. The setup_script() is called and the dictionary is passed to the function. In this chapter, the dictionary will be named "vsdParams", but any other name would do. Within the python script:

- The VSD UI parameters are referenced as vsdParams['rt'], vsdParams['vni'], vsdParams['servicetype'], and so on.
- The metadata parameters are defined in an opaque string. For example, when the metadata string "rd=1:1,sap=1/1/1:1000" is supplied to the VSD WAN Service GUI, the format in the dictionary will be in the following format: "'metadata': 'rd=1:1,sap=1/1/1:1000 '".

To reference the metadata, the format is changed (trailing space is removed and parameters split up):

```
metadata = vsdParams['metadata']
    metadata = metadata.rstrip()
    metadata = dict(e.split('=') for e in metadata.split(','))
```

The individual metadata parameters can then be referenced in a similar way as the vsd parameters; for example, metadata['rd'], metadata['sap'], and so on.

When the startup script is executed, the *config>service>vsd>domain* is created outside the script context before running the actual script. The teardown script will remove the vsd domain. The domain-name is taken from the service-name supplied by the VSD ("Service ID" field in the WAN Service GUI - used as VSD domain in the CLI). When testing the script with the *tools perform python-script* command, the domain-name is taken from the domain-name command parameter (see Testing the python script section).

When subsequent configuration messages are received from the VSD, the new parameter list is again generated from the VSD message and compared to the last parameter list that was successfully executed.

- If the two strings are identical, no action is taken.
- If there is a difference between the strings, the *modify_script()* function is called. For example, the *modify_script()* function is set up to handle a change in the service-mtu.

If a configuration message is received from the VSD for an existing service-name with no VSD parameters, the *teardown_script()* is called.

If a *setup_script()* fails, the *teardown_script()* is called.

To generate CLI output in the python script, an internal function, **dyn.add_cli**(output-string), is available. It adds the specified output-string to the CLI script. Python enables the use of triple quotes to specify strings that span multiple lines. For example:

```
from alc import dyn
    dyn.add_cli("""
configure
    service
        ies %(svc_id)s customer 1 create
            service-name "%(inst)s"
            description "%(inst)s"
            no shutdown
        exit
    exit
exit
""" % d)
```

An internal function, **dyn.select_free_id**("service-id"), is available to select a free (unused) service identifier in the service-range specified in the dynamic-services context (see the Configuration section). If no service-range is configured, the python script fails when **dyn.select_free_id**("service-id") is called. The service-id is made available again after a successful teardown (removal) of the service.

XMPP

The Extensible Messaging and Presence Protocol (XMPP) is an open technology for real-time communication, using XML (Extensible Markup Language) as the base format for exchanging information. XMPP provides a way to send small pieces of XML from one entity to another in near real time. Although initially intended for Instant Messaging applications, it can be easily extended to be used in a DC environment.

In the Nuage solution, each XMPP client, including the 7x50 SR, is referred to with a JID(JabberID) in the following format: username@xmppserver.domain. The xmppserver.domainpoints to the XMPP server.

The Nuage VSP/7x50 DC Gateway solution uses the XMPP PubSub (Publish Subscribe) extension. This extension allows a user to subscribe to a node so that it can be notified whenever there is new or updated information available. The mechanism is used in this feature to auto-discover the username of the VSD JID. Additionally, the 7x50 will subscribe to a separate PubSub for each DC Gateway, to discover updates on specific domains. Subscriptions are confirmed periodically (every 15 min).

The 7x50 DC Gateway will periodically audit the VSD and request a DIFF list of F-D VSD domains. The VSD keeps a DIFF list of domains, which contains the F-D domain names for which the VSD has not received an info/query (IQ) request from the 7x50 for a long time. The DC Gateway periodically checks the info for each of its deployed dynamic services with an IQ request (every 16-24 min). A DIFF or FULL domain list audit can also be triggered with the *tools perform service vsd fd-domain-sync <full|diff> command.*

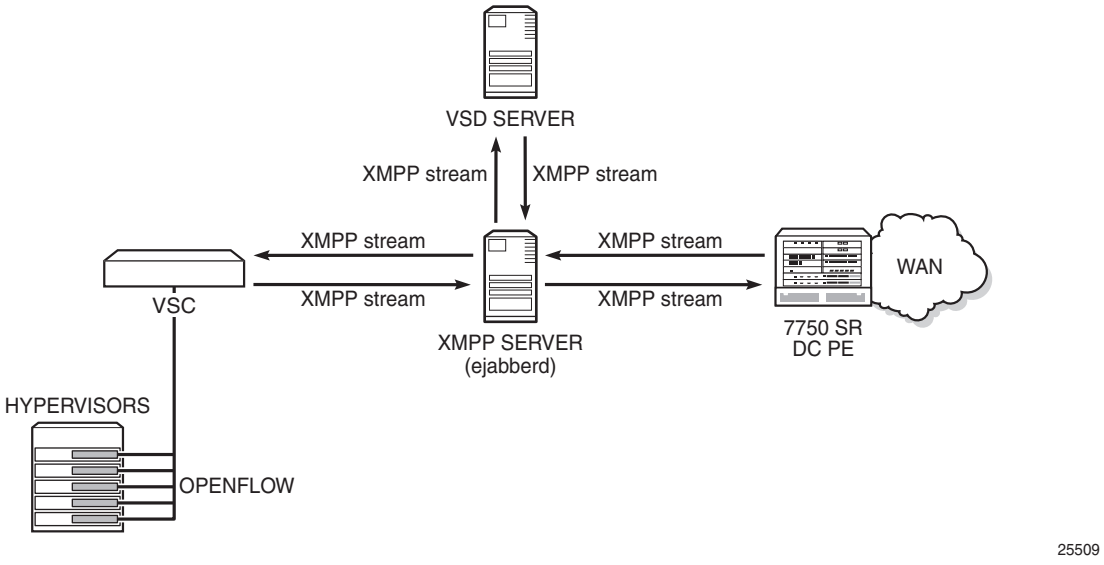
Configuration

This section describes the configuration that is required on the 7x50 DC Gateway for F-D XMPP provisioning.

The following figure shows the basic setup used and also illustrates the XMPP architecture in the data center. Although the VSD and XMPP servers are represented by a single server, a cluster of VSD servers (using the same database) and/or XMPP servers will be a very common configuration in a data center.

It is assumed that underlying IP connectivity and an IGP has already been configured in this setup.

Figure 282 **F-D XMPP provisioning setup**



XMPP configuration

To receive configuration parameters from the VSD, the 7x50 DC Gateway has to establish an XMPP client session with the XMPP server. Only one XMPP server can be configured.

When the XMPP server is properly configured, with no shutdown, the 7750 will try to establish a TCP session with the XMPP server through the management interface first. If it fails to establish communication, the 7750 will use in-band communication and will use its system IP as the source IP address.

To resolve the XMPP server fully qualified domain name (FQDN), provide a DNS server in the boot option file (bof) configuration and configure a dns-domain:

```
*A:pe-9>config>system# show bof
===============================================================================
BOF (Memory)
```

```
===============================================================================
--- snipped ---
    primary-dns      138.203.39.47
    dns-domain       nuage.net
--- snipped ---
===============================================================================
```

Then, configure the system-id of the DC Gateway that will be communicated to the VSD:

```
*A:pe-9>config>system# info
----------------------------------------------
#----------------------------------------------------
echo "vsd Configuration"
#----------------------------------------------------
        vsd
            system-id "pe9"
        exit
```

The next step is to configure the VSD server. The domain-name is the domain portion of the JID. The username is the username portion of the JID acting as an XMPP client. Ensure that the username uses all letters in lowercase (see SR OS 13.0 release notes). If no username is provided, an in-band registration will be provided, using the chassis MAC as username. The use of a password is optional:

```
*A:pe-9>config>system# info
#----------------------------------------------------
echo "Xmpp Configuration"
#----------------------------------------------------
        xmpp
            server vsd domain-name vsd1.nuage.net create username pe9
                no shutdown
            exit
        exit
----------------------------------------------
```

When the XMPP server has been configured, the state should move to "Functional":

```
*A:pe-9# show system xmpp server
===============================================================================
XMPP Server Table
===============================================================================
Name                            User Name          State
 XMPP FQDN                       Last State chgd    Admin State
-------------------------------------------------------------------------------
vsd                             pe9                Functional
 vsd1.nuage.net                  0d 00:37:01        inService
-------------------------------------------------------------------------------
No. of XMPP server's: 1
===============================================================================
```

XMPP Tx/Rx counters and other details can be obtained with the following command:

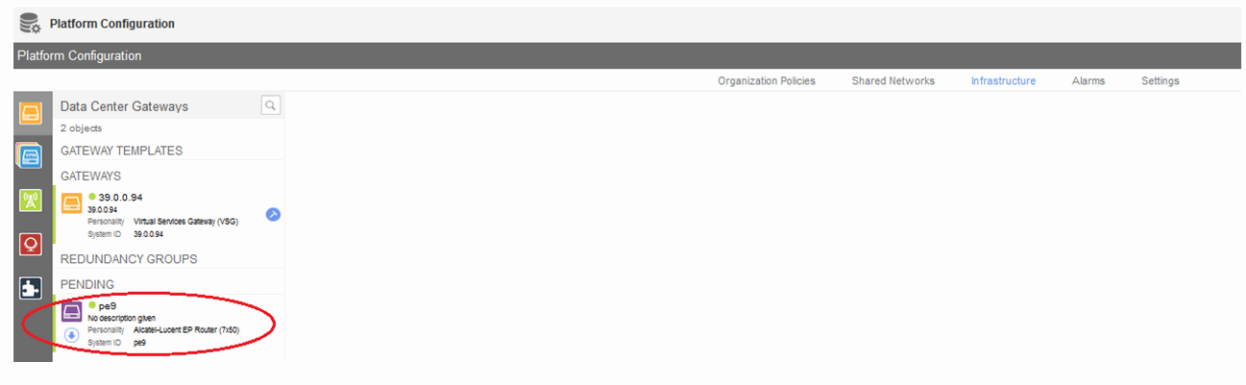
```
*A:pe-9# show system xmpp server "vsd"
===============================================================================
XMPP Server Table
===============================================================================
XMPP FQDN          : vsd1.nuage.net
XMPP Admin User    : pe9
XMPP Oper User     : pe9
State Lst Chg Since: 0d 00:07:43        State              : Functional
Admin State        : Up                 Connection Mode    : outOfBand
Auth Type          : md5
IQ Tx.             : 10                 IQ Rx.             : 10
IQ Error           : 0                  IQ Timed Out       : 0
IQ Min. Rtt        : 20 ms              IQ Max. Rtt        : 80 ms
IQ Ack Rcvd.       : 10
Push Updates Rcvd  : 1                  VSD list Upd Rcvd  : 1
Msg Tx.            : 3                  Msg Rx.            : 3
Msg Ack. Rx.       : 3                  Msg Error          : 0
Msg Min. Rtt       : 0 ms               Msg Max. Rtt       : 80 ms
Sub Tx.            : 1                  UnSub Tx.          : 0
Msg Timed Out      : 0
```

F-D XMPP provisioning uses the PubSub XMPP extension that allows each user to subscribe to a node, to be notified whenever that node gets new pieces of information or updated information.

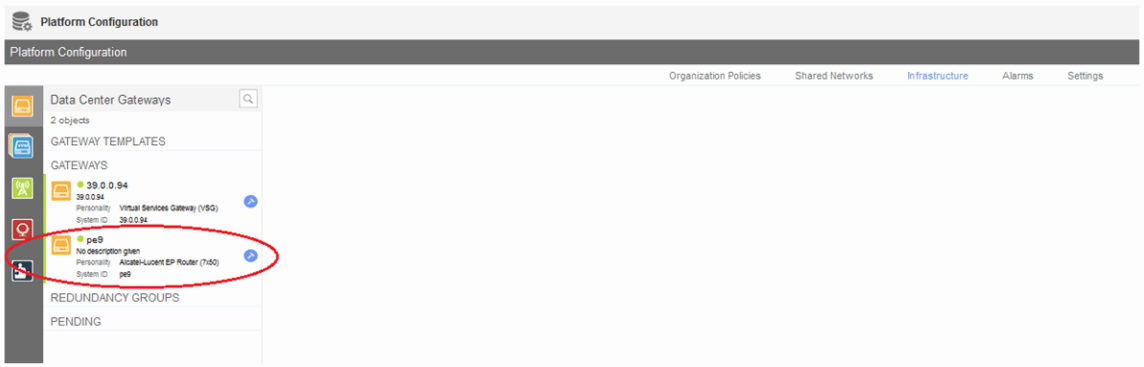
The DC Gateway PubSub subscription state and subscriber name can be shown:

```
*A:pe-9# show system vsd
===============================================================================
VSD Information
===============================================================================
System Id               : pe9
GW Last Audit Tx Time   : 10/13/2015 13:56:34
Gateway Publish-Subscribe Information
-------------------------------------------------------------------------------
Subscribed              : True
Subscriber Name         : nuage_gateway_id_pe9
Last Subscription Time  : 10/13/2015 13:56:34
===============================================================================
```

At the same time, the 7x50 DC Gateway will be announced as a pending gateway in the VSD:



The gateway will be promoted to the available gateways group by clicking on the arrow below the pending gateway icon:



BGP configuration

In the Nuage VSP solution, MP-BGP is used in the control plane to distribute MAC/IP information about the VMs. This information is distributed between the different VSCs, VSGs, and 7x50 DC Gateways. Configure MP-BGP on the 7x50 DC Gateway and the VSC/VSG (in this case, a VSG was used, but VSC is similar):

```
*A:pe-9>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "IP Configuration"
#--------------------------------------------------
        interface "system"
            address 10.0.0.9/32
            no shutdown
        exit
```

```
--- snipped ---
        autonomous-system 65000
--- snipped ---
#--------------------------------------------------
echo "BGP Configuration"
#--------------------------------------------------
            min-route-advertisement 5
            rapid-withdrawal
            rapid-update evpn
            group "Nuage"
                family route-target evpn
                type internal
                neighbor 39.0.0.94
                exit
            exit
            no shutdown


*A:vsc1.nuage.net>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "IP Configuration"
#--------------------------------------------------
--- snipped ---
        interface "system"
            address 39.0.0.94/32
            no shutdown
        exit
        autonomous-system 65000
--- snipped ---
#--------------------------------------------------
echo "BGP Configuration"
#--------------------------------------------------
        bgp
            family route-target evpn
            min-route-advertisement 5
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                type internal
                neighbor 10.0.0.9
                    family evpn
                exit
            exit
            no shutdown
        exit
----------------------------------------------
```

In this setup, the family type "evpn" and "route-target" is used. The former is used to learn the EVPN route updates while the latter is restricting the 7x50 to only learn those MB-BGP routes for which it has a route target configured.

Note: To use vrf-gre domains, configure BGP family "vpn-ipv4" as well. Similarly, to use BGP-MH (for example, in case of redundant 7x50 DC Gateways with L2-domains), the use of BGP family "l2vpn" is required.

Verify that BGP peering is in the operational state:

```
*A:pe-9# show router bgp summary
===============================================================================
 BGP Router ID:10.0.0.9         AS:65000       Local AS:65000
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
--- snipped ---
                  AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                     PktSent OutQ
-------------------------------------------------------------------------------
39.0.0.94
                65000   76287    0 00h04m14s 0/0/0 (RouteTarget)
                         2634    0           0/0/0 (Evpn)
-------------------------------------------------------------------------------
```

Dynamic VSD Services range

F-D XMPP provisioning requires a reserved range of Service-IDs that can be used for dynamic data services. This configured range is no longer available for regular services configured via CLI/SNMP:

```
*A:pe-9>config>service# info
----------------------------------------------
        vsd
            service-range 64000 to 64999
        exit
```

```
*A:pe-9# show service vsd summary
===============================================================================
VSD Information
===============================================================================
Service Range
Start              : 64000                        End           : 64999
===============================================================================
No domain entries found
```

Python script

The python script that will build the dynamic services based on the VSD parameters obtained via XMPP can be stored locally on the CF or on a remote FTP server:

```
*A:pe-9>config>python# info
----------------------------------------------
        python-script "l2domain_services" create
            primary-url "ftp://*:*@138.203.15.48/./l2domain_service.py"
            no shutdown
        exit
```

The script is loaded into memory as soon as a no shutdown is performed and is reloaded with each *shutdown/no shutdown* action. Alternatively, a tools command can be used to reload the script:

```
*A:pe-9# tools perform python-script reload "l2domain_services"
```

In case incorrect python syntax is used in the script, an error message is displayed after a *no shut*down of the script (or a reload with the tools command), with an indication of the line where the error is located.

The details of the python script can be inspected:

```
*A:pe-9# show python python-script "l2domain_services"
===============================================================================
Python script "l2domain_services"
===============================================================================
Description   : (Not Specified)
Admin state   : inService
Oper state    : inService
Action on fail: drop
Protection    : none
Primary URL   : ftp://*:*@138.203.15.48/./l2domain_service.py
Secondary URL : (Not Specified)
Tertiary URL  : (Not Specified)
Active URL    : primary
Last changed  : 10/13/2015 09:54:26
===============================================================================
```

The contents of the python script can also be viewed. The contents of the script are shown in the "Test the python script" section:

```
*A:pe-9# show python python-script "l2domain_services" source-in-use
===============================================================================
Python script "l2domain_services"
===============================================================================
Admin state   : inService
Oper state    : inService
Primary URL   : ftp://*:*@138.203.15.48/./l2domain_service.py
Secondary URL : (Not Specified)
Tertiary URL  : (Not Specified)
Active URL    : primary
-------------------------------------------------------------------------------
Source (dumped from memory)
-------------------------------------------------------------------------------
      1 from alc import dyn
      2
      3 # example of metadata to be added in VSD WAN Service: "rd=1:1,sap=1/1/
1:1000"
      4
--- snipped ---
      8 def setup_script(vsdParams):
--- snipped ---
     70 def modify_script(vsdParams,setup_result):
--- snipped ---
```

```
    106 def revert_script(vsdParams,setup_result):
--- snipped ---
    129 def teardown_script(setupParams):
--- snipped ---
    164 d = {"script" : (setup_script, modify_script, revert_script, teardown_script
)}
    165
    166 dyn.action(d)
================================================================================
```

A list of CLI command nodes that can be used with python script function dyn.add_cli is provided with the *tools dump service vsd-services command-list* command. In general, all the 'leaf' commands under the nodes shown in the tools dump command, can be used with the python script.

Further restriction of CLI commands is possible by creating a separate CLI user for the XMPP interface and associate that user with a profile where the commands are limited.

The CLI user for the XMPP interface is configurable:

```
config>system>security>cli-script>authorization>
        vsd
[no] cli-user <username>
```

Python policy

Python scripts are called by a python policy that will be referred to in the VSD WAN services GUI in the Service Policy field.

Create a python policy (that will be referred to in the VSD GUI) and link the python policy to the python script:

```
*A:pe-9>config>python# info
----------------------------------------------
        python-policy "py-l2" create
            description "Python script to create L2 domains"
            vsd script "l2domain_services"
        exit
```

```
*A:pe-9# show python python-script "l2domain_services" association
================================================================================
Python Script Association
================================================================================
Policy                          Type                                    Dir
--------------------------------------------------------------------------------
py-l2                           vsdAccessRequest                        ingress
--------------------------------------------------------------------------------
```

```
*A:pe-9# show python python-policy "py-l2"
```

```
===============================================================================
Python policy "py-l2"
===============================================================================
Description                : Python script to create L2 domains
-------------------------------------------------------------------------------
Messages
-------------------------------------------------------------------------------
Type                                Dir     Script
-------------------------------------------------------------------------------
vsdAccessRequest                    ingress l2domain_services
-------------------------------------------------------------------------------
```

Test the python script

The python script can be tested separately on the 7x50 DC Gateway; even before connecting it to the Nuage setup.

Some notes about the python script and creating dynamic services:

- The VSD (and *tools evaluate-script* command) will provide some compulsory parameters like:
  - domain-name
  - domain type (l2-domain|vrf-gre|vrf-vxlan|l2-domain-irb)
  - action (setup or teardown)
  - vni
  - RT (internal and external route-targets are provided)
- The VSD (and *tools evaluate-script* command) can provide extra metadata that is supplied in a text string and comma separated. For example, metadata "rd=1:1,sap=1/1/1:1000".
- The RT format supplied by the VSD though XMPP (that is, "x:x") differs from the RT format that can be used with the *tools evaluate-script* command (for example, "target:x:x"). For that reason, it can be useful to add the following check in the script:

```
if not rt.startswith ('target'):
     rt = "target:"+rt
```

- The VSD metadata string includes an empty space at the end. This can be removed with the following python command:

```
metadata = metadata.rstrip()
```

- If a configuration message is received from the VSD for an existing service-name with no VSD parameters, or if a setup_script() fails, the teardown_script() is called.

- At any point in the script, you can add *print* commands to check the status/ content of various parameters.

The python policy and script can then be tested with the tools evaluate-script command. Example syntax has been added into the scripts for convenience.

Before testing the script, it will be useful to enable the following debugging:

```
debug
    python
        python-script "l2domain_services"
            script-all-info
        exit
    exit
    vsd
        scripts
            event
                cli
                errors
                executed-cmd
                warnings
                state-change
            exit
        exit
    exit
exit
```

In this section, a python example script for an l2-domain is used. The contents of the script are shown in the following text format. Examples for l2-domain-irb and vrf-vxlan type domains are shown in dedicated sections:

```
from alc import dyn
# example of metadata to be added in VSD WAN Service: "rd=1:1,sap=1/1/1:1000"
# example of tools cli to test this script: tools perform service vsd evaluate-
script domain-name "l2dom1" type l2-domain action setup policy "py-l2" vni 1234 rt-
i target:1:1 rt-e target:1:1 metadata "rd=1:1,sap=1/1/1:1000"

# teardown example cli: tools perform service vsd evaluate-script domain-
name "l2dom1" type l2-domain action teardown policy "py-l2" vni 1234 rt-
i target:1:1 rt-e target:1:1

def setup_script(vsdParams):

    print ("These are the VSD params: " + str(vsdParams))
    servicetype = vsdParams['servicetype']
    vni = vsdParams['vni']
    rt = vsdParams['rt']

# add "target:" if provisioned by VSD (VSD uses x:x format whereas tools command use
s target:x:x format)
    if not rt.startswith ('target'):
       rt = "target:"+rt
    metadata = vsdParams['metadata']

# remove trailing space at the end of the metadata
    metadata = metadata.rstrip()
```

```
    print ("VSD metadata" + str(metadata))


    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
    vplsSvc_id = dyn.select_free_id("service-id")
    print ("this is the free svc id picked up by the system: " + vplsSvc_id)
    if servicetype == "L2DOMAIN":

      rd = metadata['rd']
      sap_id = metadata['sap']
      print ('servicetype, VPLS id, rt, vni, rd, sap:', servicetype, vplsSvc_id, rt,
 vni, rd, sap_id)
      dyn.add_cli("""
        configure service
           vpls %(vplsSvc_id)s customer 1 create
              description vpls%(vplsSvc_id)s
              proxy-arp
                 dynamic-arp-populate
                 no shutdown
                 exit
              bgp
                 route-distinguisher %(rd)s
                 route-target %(rt)s
              exit
              vxlan vni %(vni)s create
              exit
              bgp-evpn
                  evi %(vplsSvc_id)s
                  vxlan
                      no shutdown
                  exit
              exit
              service-name evi%(vplsSvc_id)s
              sap %(sap_id)s create
              exit
              no shutdown
              exit
            exit
        exit
      """ % {'vplsSvc_id' : vplsSvc_id, 'vni' : vsdParams['vni'], 'rt' : rt, 'rd' :
metadata['rd'], 'sap_id' : sap_id})
      # L2DOMAIN returns setupParams: vplsSvc_id, servicetype, vni, sap
      return {'vplsSvc_id' : vplsSvc_id, 'servicetype' : servicetype, 'vni' : vni, '
sap_id' : sap_id}
#-------------------------------------------------------------------------------
-------------
def modify_script(vsdParams,setup_result):

    print ("These are the setup_result params for modify_script: " + str(setup_resul
t))
    print ("These are the VSD params for modify_script: " + str(vsdParams))

    # remove trailing space at the end of the metadata
    metadata = vsdParams['metadata'].rstrip()

    print ("VSD metadata" + str(metadata))
    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
```

```
    # updating the setup_result dict
    setup_result.update(metadata)
    params = setup_result

    print ("The updated params from metadata and return from the setup result: " + s
tr(params))

    svc_mtu = params['svc-mtu']

    dyn.add_cli("""
      configure service
          vpls %(vplsSvc_id)s
            service-mtu %(svc-mtu)s
             exit
          exit
      exit
    """ %params )


    # Result is passed to teardown_script
    return params
#-------------------------------------------------------------------------------
-------------
def revert_script(vsdParams,setup_result):
    print ("These are the setup_result params for revert_script: " + str(setup_resul
t))
    print ("These are the VSD params for revert_script: " + str(vsdParams))

    # When modify fails, the revert is called and then the teardown is called.
    # It is recommended to revert to same value as used in setup for the attributes
modified in modify_script.

    params = setup_result

    dyn.add_cli("""
      configure service
          vpls %(vplsSvc_id)s
            service-mtu 2000
             exit
          exit
      exit
    """ %params )

    # Result is passed to teardown_script
    return params

#-------------------------------------------------------------------------------
-------------
def teardown_script(setupParams):
    print ("These are the teardown_script setupParams: " + str(setupParams))
    servicetype = setupParams['servicetype']
    if servicetype == "L2DOMAIN":
      dyn.add_cli("""
        configure service
            vpls %(vplsSvc_id)s
               no description
               proxy-arp shut
```

```
                no proxy-arp
                bgp-evpn
                    vxlan
                        shut
                    exit
                    no evi
                    exit
                no vxlan vni %(vni)s
                bgp
                   no route-distinguisher
                   no route-target
                exit
                no bgp
                no bgp-evpn
                sap %(sap_id)s
                   shutdown
                   exit
                no sap %(sap_id)s
                shutdown
                exit
                no vpls %(vplsSvc_id)s
            exit
          exit
        """ % {'vplsSvc_id' : setupParams['vplsSvc_id'], 'vni' : setupParams['vni'], '
sap_id' : setupParams['sap_id']})
       return setupParams

d = {"script" : (setup_script, modify_script, revert_script, teardown_script)}

dyn.action(d)
```

The script can be tested with the following command:

```
*A:pe-9# tools perform service vsd evaluate-script domain-name "l2dom1" type l2-
domain action setup policy "py-l2" vni 1234 rt-i target:1:1 rt-
e target:1:1 metadata "rd=1:1,sap=1/1/1:1000"
1 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=init->waiting-for-setup
"
2 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=waiting-for-setup->generating-setup
"
3 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base Python Output
"Python Output: l2domain_services
These are the VSD params: {'rt': 'target:1:1', 'rte': 'target:1:1', 'domain': ''
, 'servicetype': 'L2DOMAIN', 'vni': '1234', 'metadata': 'rd=1:1,sap=1/1/1:1000 '
}
VSD metadatard=1:1,sap=1/1/1:1000
Modified metadata{'rd': '1:1', 'sap': '1/1/1:1000'}
this is the free svc id picked up by the system: 64000
('servicetype, VPLS id, rt, vni, rd, sap:', 'L2DOMAIN', '64000', 'target:1:1', '
1234', '1:1', '1/1/1:1000')
"
4 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base Python Result
"Python Result: l2domain_services
"
```

```
5 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=generating-setup->executing-setup
"
6 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script cli 1/1
"dyn-script cli 1/1: script:l2dom1(cli 705 dict 0->123)
        configure service
           vpls 64000 customer 1 create
              description vpls64000
              proxy-arp
                 dynamic-arp-populate
                 no shut
                 exit
              bgp
                 route-distinguisher 1:1
                 route-target target:1:1
              exit
              vxlan vni 1234 create
              exit
              bgp-evpn
                  evi 64000
                  vxlan
                      no shut
                  exit
              exit
              service-name evi64000
              sap 1/1/1:1000 create
              exit
              no shutdown
              exit
            exit
        exit
      "
7 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script setup
"dyn-script setup: l2dom1 script:l2dom1 line 2
 configure service"
Success
--- snipped ---
24 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=executing-setup->established
"
```

At this moment a new VSD domain has been created as well as a new service:

```
*A:pe-9# show service vsd domain
===============================================================================
VSD Domain Table
===============================================================================
Name                                    Type           Origin    Admin
-------------------------------------------------------------------------------
l2dom1                                  l2Domain       vsd       inService
-------------------------------------------------------------------------------


*A:pe-9# show service vsd domain "l2dom1" association
============================================================
Service VSD Domain
============================================================
```

```
Svc Id      Svc Type  Domain Type    Domain Admin   Origin
-------------------------------------------------------------
64000       vpls      l2Domain       inService      vsd
-------------------------------------------------------------


*A:pe-9# show service vsd domain "l2dom1"
===============================================================================
VSD Information
===============================================================================
Name               : l2dom1
Description        : l2dom1
Type               : l2Domain                     Admin State   : inService
Last Error To Vsd  : (Not Specified)
Last Error From Vsd: (Not Specified)
Statistics
-------------------------------------------------------------------------------
Last Cfg Chg Evt   : 10/14/2015 16:02:54          Cfg Chg Evts  : 1
Last Cfg Update    : 10/14/2015 16:02:54          Cfg Upd Rcvd  : 1
Last Cfg Done      : 10/14/
2015 16:02:54Cfg Success        : 1                              Cfg Failed    : 0
Last Recd Params   : script = {'domain' : '', 'vn
                   : i' : '1234', 'rt' : 'target:
                   : 1:1', 'rte' : 'target:1:1',
                   : 'servicetype' : 'L2DOMAIN',
                   : 'metadata' : 'rd=1:1,sap=1/1
                   : /1:1000 '}
Last Exec Params   : script = {'domain' : '', 'vn
                   : i' : '1234', 'rt' : 'target:
                   : 1:1', 'rte' : 'target:1:1',
                   : 'servicetype' : 'L2DOMAIN',
                   : 'metadata' : 'rd=1:1,sap=1/1
                   : /1:1000 '}
===============================================================================


*A:pe-9# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
64000        VPLS      Up   Up   1          evi64000
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
```

Note: Service ID 2147483648 and 2147483649 are internal services that are always present on the 7x50. They are not relevant for this feature and will be truncated in other output examples in this document.

```
*A:pe-9# show service id 64000 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64000              Vpn Id            : 0
Service Type      : VPLS
```

```
Name              : evi64000
Description       : vpls64000
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 10/14/2015 16:02:54
Last Mgmt Change  : 10/14/2015 16:02:54
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 64000
SAP Count         : 1                   SDP Bind Count    : 0
--- snipped ---
VSD Domain        : l2dom1
--- snipped ---
-------------------------------------------------------------------------------
BGP Information
-------------------------------------------------------------------------------
Vsi-Import        : None
Vsi-Export        : None
Route Dist        : 1:1
Oper Route Dist   : 1:1
Oper RD Type      : configured
Rte-Target Import : 1:1                 Rte-Target Export : 1:1
Oper RT Imp Origin: configured          Oper RT Import    : 1:1
Oper RT Exp Origin: configured          Oper RT Export    : 1:1
PW-Template Id    : None
-------------------------------------------------------------------------------
--- snipped ---
-------------------------------------------------------------------------------
SAP 1/1/1:1000
-------------------------------------------------------------------------------
Service Id         : 64000
SAP                : 1/1/1:1000              Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                      Oper State        : Up
--- snipped ---
```

Note: You cannot see the dynamic VSD services in the configuration nor can you edit their configuration under normal circumstances (this is discussed further in the next section).

```
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            service-range 64000 to 64999
        exit
----------------------------------------------

*A:pe-9# configure service vpls 64000
MINOR: CLI Modification of services created by a dynamic script is not allowed.
```

The service can be modified by adding/changing a service-mtu to the metadata. This will trigger the modify-script function in the python script. The following basic script is only an example of how a modify-script function operates. The script could be extended to modify other parameters as well; however this is out of the scope of this chapter:

```
*A:pe-9# tools perform service vsd evaluate-script domain-name "l2dom1" type l2-
domain action modify policy "py-l2" vni 1234 rt-i target:1:1 rt-
e target:1:1 metadata "rd=1:1,sap=1/1/1:1000,svc-mtu=2222"
25 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=modify
"dyn-script req=modify: l2dom1
  state=established->waiting-for-modify
"
26 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=modify
"dyn-script req=modify: l2dom1
  state=waiting-for-modify->generating-modify
"
27 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base Python Output
"Python Output: l2domain_services
These are the setup_result params for modify_script: {'servicetype': 'L2DOMAIN',
 'vplsSvc_id': '64000', 'vni': '1234', 'sap_id': '1/1/1:1000'}
These are the VSD params for modify_script: {'rt': 'target:1:1', 'rte': 'target:
1:1', 'domain': '', 'servicetype': 'L2DOMAIN', 'vni': '1234', 'metadata': 'rd=1:
1,sap=1/1/1:1000,svc-mtu=2222 '}
VSD metadatard=1:1,sap=1/1/1:1000,svc-mtu=2222
Modified metadata{'rd': '1:1', 'sap': '1/1/1:1000', 'svc-mtu': '2222'}
The updated params from metadata and return from the setup result: {'rd': '1:1',
 'servicetype': 'L2DOMAIN', 'svc-mtu': '2222', 'sap_id': '1/1/1:1000', 'sap': '1
/1/1:1000', 'vplsSvc_id': '64000', 'vni': '1234'}
"
Success
28 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base Python Result
"Python Result: l2domain_services
"
29 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=modify
"dyn-script req=modify: l2dom1
  state=generating-modify->executing-modify
"
*A:pe-9#
30 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script cli 1/1
"dyn-script cli 1/1: script:l2dom1(cli 123 dict 123->203)
      configure service
          vpls 64000
            service-mtu 2222
             exit
          exit
      exit
    "
31 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script modify
"dyn-script modify: l2dom1 script:l2dom1 line 2
 configure service"
--- snipped ---
36 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=commit
"dyn-script req=commit: l2dom1
  state=waiting-for-commit->established
"
```

The service-mtu has now been changed to 2222:

```
show service id 64000 all | match MTU
MTU               : 2222                Def. Mesh VC Id   : 64000
Admin MTU         : 9212                      Oper MTU          : 9212
```

The service can be removed with the teardown script:

```
*A:pe-9# tools perform service vsd evaluate-script domain-name "l2dom1" type l2-
domain action teardown policy "py-l2" vni 1234 rt-i target:1:1 rt-e target:1:1
37 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=established->waiting-for-teardown
"
38 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=waiting-for-teardown->generating-teardown
"
39 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base Python Output
"Python Output: l2domain_services
These are the teardown_script setupParams: {'servicetype': 'L2DOMAIN', 'svc-mtu'
: '2222', 'sap_id': '1/1/1:1000', 'vplsSvc_id': '64000', 'vni': '1234', 'rd': '1
:1', 'sap': '1/1/1:1000'}
"
40 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base Python Result
"Python Result: l2domain_services
"
41 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=generating-teardown->executing-teardown
"
42 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script cli 1/1
"dyn-script cli 1/1: script:l2dom1(cli 709 dict 203->0)
        configure service
            vpls 64000
               no description
               proxy-arp shut
               no proxy-arp
               bgp-evpn
                   vxlan
                       shut
                   exit
                   no evi
                   exit
               no vxlan vni 1234
               bgp
                  no route-distinguisher
                  no route-target
               exit
               no bgp
               no bgp-evpn
               sap 1/1/1:1000
                  shutdown
                  exit
               no sap 1/1/1:1000
               shutdown
               exit
               no vpls 64000
```

```
            exit
          exit
        "
43 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script teardown
"dyn-script teardown: l2dom1 script:l2dom1 line 2
 configure service"
--- snipped ---
63 2015/10/15 09:51:29.81 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=executing-teardown->stopped
"
```

After the dynamic service has been torn down, the dynamic service and the VSD domain should not be present on the 7x50 DC Gateway:

```
*A:pe-9# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 2
-------------------------------------------------------------------------------
===============================================================================


*A:pe-9# show service vsd domain
No domain entries found
```

Editing dynamic VSD services

As indicated in the previous section, the dynamic VSD services CLI configuration cannot be shown or edited normally. However, under certain circumstances, it might be necessary to inspect/change/remove the configuration of a dynamic VSD service; for example, when the python VSD script was not using the correct syntax and the creation/deletion of the dynamic VSD service failed.

It is possible to edit the dynamic VSD services configuration by entering the *enable-vsd-config* mode. First, create a password, which is required to enter this mode:

```
*A:pe-9# configure system security password
*A:pe-9>config>system>security>password# vsd-password *****
```

Then, enter the enable-vsd-config mode. You will be asked for the previously configured password:

```
*A:pe-9# enable-vsd-config
Password:
```

Now you can edit the dynamic VSD services configuration and change/add/remove configuration:

```
*A:pe-9# configure service
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            domain l2dom1 type l2-domain create
                description "l2dom1"
                no shutdown
            exit
            service-range 64000 to 64999
        exit
        vpls 64000 customer 1 create
            description "vpls64000"
            vxlan vni 1234 create
            exit
            bgp
                route-distinguisher 1:1
            exit
            bgp-evpn
                evi 64000
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            proxy-arp
                dynamic-arp-populate
                no shutdown
            exit
            stp
                shutdown
            exit
            service-name "evi64000"
            sap 1/1/1:1000 create
            exit
            vsd-domain "l2dom1"
            no shutdown
        exit
----------------------------------------------
```

In the enable-vsd-config mode, only dynamic VSD services can be edited, not regular CLI-based services:

```
*A:pe-9# configure service vpls 100 customer 1 create
MINOR: SVCMGR #1201 Invalid service-id - not reserved
```

After inspecting/editing the dynamic VSD services configuration, you should exit this mode again:

```
*A:pe-9# no enable-vsd-config
```

L2 VXLAN

An example python script for F-D XMPP provisioning of an L2 VXLAN type service (l2-domain) was provided in the "Testing the python script" section. In this section, the same script is used for provisioning via the VSD.

To dynamically provision this type of service, a few things must be configured on the VSD: (screenshots of this workflow are available in the VSP User Guide)

Create an L2 WAN service in the VSD:

- under Platform Config/Infrastructure, select the DC Gateway to add a WAN service
- select Service Type "layer 2" (no IRB)
- select "Dynamic" configuration type to allow Fully Dynamic provisioning
- under Service Policy, provide the python policy configured on the DC Gateway
- provide a Name and Service-ID (the Service-ID will be the name of the dynamically created service domain on the DC Gateway)

Add the metadata to the WAN service:

- right-click the WAN service and select "inspect"
- a dialog box appears, select the "Metadata" tag and add the metadata info "rd=1:1,sap=1/1/1:1000"

Add permissions to Enterprise1. The WAN service should now be visible in Enterprise1. Add permissions for a group of users to use this WAN service. Instantiate the L2 domain and attach the WAN service.

As soon as the WAN service is attached to the L2 domain, the VSD will send a notification via XMPP to the DC Gateway about the new Service-ID. The VSD will send an XMPP IQ request to the VSD to obtain the VSD service parameters.

Note: There is an 8 to 12 s delay. The command *tools perform service vsd domain refresh-config* can be used to expedite the request.

As soon as the DC Gateway receives this information, the python policy mentioned in the VSD service parameters is triggered and the VSD parameters are passed to the associated python script. The python script will then construct and execute the same configuration CLI and trigger similar debug information as is shown in the section "Testing the python script".

After the python script has completed successfully, the service can be inspected in a similar way as before.

- show service service-using
- show service id <*service-id*> all
- enter enable-vsd-config mode if required and inspect the service config

MAC addresses can now be learned in the Nuage VSC/VSG and sent via MP-BGP to the DC Gateway:

```
*A:pe-9# show service id 64000 fdb detail
===============================================================================
Forwarding Database, Service 64000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
64000     1e:50:01:01:00:01 vxlan:                   Evpn     10/14/15 16:10:44
                            39.0.0.94:1006636
-------------------------------------------------------------------------------
```

In case HyperVisors (HVs) with VRS are deployed or a Host vPORT has been connected to a VSG, the EVPN MAC/Route (type 2) will also include the IP address of the VM/Host, in which case the DC Gateway can perform a proxy-arp function. For more information about EVPN-VXLAN features, refer to the chapters EVPN for VXLAN Tunnels (Layer 2) and EVPN for VXLAN Tunnels (Layer 3).

```
*A:pe-9# show service id 64000 proxy-arp detail
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : enabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 1
Static Count      : 0                   EVPN Count        : 1
Dynamic Count     : 0                   Duplicate Count   : 0
Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None
EVPN
-------------------------------------------------------------------------------
Garp Flood        : enabled             Req Flood         : enabled
-------------------------------------------------------------------------------
```

```
===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
10.32.78.100        1e:50:01:01:00:01   evpn      active    07/22/2015 10:05:25
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

The svc-ID belonging to the Service-ID configured on the VSD can be easily obtained:

```
*A:pe-9# show service vsd domain "L2-service-1" association
============================================================
Service VSD Domain
============================================================
Svc Id      Svc Type  Domain Type    Domain Admin   Origin
------------------------------------------------------------
64000       vpls      l2Domain       inService      vsd
------------------------------------------------------------
Number of entries: 1
```

The associated RT/RD/VNI information can then be displayed with the *show service id <service-id> all* command, as shown previously.

An overview of the VTEPs that the DC Gateway shares this service with, and the corresponding VNIs is available:

```
*A:pe-9# show service id 64000 vxlan
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1006636
===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
39.0.0.94              1006636        0            Yes     Up           No
-------------------------------------------------------------------------------
```

An overview of all the Service-IDs and associated VNIs that the DC Gateway has in common with a VSC or VSG can be shown with the following command:

```
*A:pe-9# show service vxlan 39.0.0.94
===============================================================================
VXLAN Tunnel Endpoint: 39.0.0.94
===============================================================================
Egress VNI                    Service Id               Oper State
-------------------------------------------------------------------------------
1006636                       64000                    Up
-------------------------------------------------------------------------------
===============================================================================
```

Relevant MAC/IP/VNI/RT/NH information is also in the EVPN BGP RIB:

```
*A:pe-9# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:10.0.0.9         AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.          MacAddr           ESI
      Tag                  Mac Mobility      Ip Address
                                             NextHop
                                             Label1
-------------------------------------------------------------------------------
u*>i  65534:39790          1e:50:01:01:00:01 ESI-0
      1006636              Static            10.32.78.100
                                             39.0.0.94
                                             VNI 1006636
-------------------------------------------------------------------------------
```

The WAN service can be detached from the L2 domain in the VSD GUI, if required.

This triggers similar debug information on the DC Gateway as the *tools perform service vsd evaluate teardown* command shown in the "Testing the python script" section.

After deleting the WAN service in the VSD GUI, the VPLS service and the service domain is removed from the DC Gateway.

L2 VXLAN IRB

An example python script for F-D XMPP provisioning of an L2 VXLAN IRB type service (l2-domain-irb) is as follows:

```
from alc import dyn
# example of metadata to be added in VSD WAN Service: "rd=2:2,sap=1/1/
1:1000,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1,irbGW=10.32.
78.1/24"
# example of tools cli to test this script: tools perform service vsd evaluate-
script domain-name "l2domIRB1" type l2-domain-irb action setup policy "py-l2-
irb" vni  1234 rt-i target:2:2 rt-e target:2:2 metadata "rd=2:2,sap=1/1/
1:1000,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1,irbGW=10.32.
78.1/24"
# teardown example cli: tools perform service vsd evaluate-script domain-
name "l2domIRB1" type l2-domain-irb action teardown policy "py-l2-irb" vni 1234 rt-
i target:2:2  rt-e target:2:2
def setup_script(vsdParams):

    print ("These are the VSD params: " + str(vsdParams))
    servicetype = vsdParams['servicetype']
```

```
    vni = vsdParams['vni']
    rt = vsdParams['rt']
# add "target:" if provisioned by VSD (VSD uses x:x format whereas tools command use
s target:x:x format)
    if not rt.startswith ('target'):
       rt = "target:"+rt
    metadata = vsdParams['metadata']
# remove trailing space at the end of the metadata
    metadata = metadata.rstrip()
    print ("VSD metadata" + str(metadata))
    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
    vplsSvc_id = dyn.select_free_id("service-id")
    vprnSvc_id = dyn.select_free_id("service-id")
    print ("this are the free svc ids picked up by the system: VPLS:" + vplsSvc_id +
 " + VPRN:" + vprnSvc_id)
    if servicetype == "L2DOMAIN-IRB":

      rd = metadata['rd']
      sap_id = metadata['sap']
      vprn_AS = metadata ['vprnAS']
      vprn_RD = metadata ['vprnRD']
      vprn_RT = metadata ['vprnRT']
      vprn_Lo = metadata ['vprnLo']
      irb_GW = metadata ['irbGW']
      print ('servicetype, VPLS id, rt, vni, rd, sap, VPRN id, vprn_AS, vprn_RD, vpr
n_RT, vprn_Lo, irb_GW:', servicetype, vplsSvc_id, rt, vni, rd, sap_id, vprnSvc_id, v
prn_AS, vprn_RD, vprn_RT, vprn_Lo, irb_GW)
      dyn.add_cli("""
        configure service
           vpls %(vplsSvc_id)s customer 1 create
              allow-ip-int-bind
              exit
              description vpls%(vplsSvc_id)s
              bgp
                 route-distinguisher %(rd)s
                 route-target %(rt)s
              exit
              vxlan vni %(vni)s create
              exit
              bgp-evpn
                  evi %(vplsSvc_id)s
                  vxlan
                      no shut
                  exit
              exit
              service-name vpls%(vplsSvc_id)s
              sap %(sap_id)s create
              exit
              no shutdown
              exit
             exit
        exit
        configure service
           vprn %(vprnSvc_id)s customer 1 create
              autonomous-system %(vprn_AS)s
              route-distinguisher %(vprn_RD)s
              vrf-target %(vprn_RT)s
              interface "irbvpls-%(vplsSvc_id)s" create
```

```
                  address %(irb_GW)s
                  vpls "vpls%(vplsSvc_id)s"
                  exit
               exit
               interface "lo1" create
                  address %(vprn_Lo)s/32
                  loopback
               exit
               no shutdown
            exit
         exit
      """ % {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'vni' : vsdParams
['vni'], 'rt' : rt, 'rd' : metadata['rd'], 'sap_id' : sap_id, 'vprn_AS' : vprn_AS, '
vprn_RD' : vprn_RD, 'vprn_RT' : vprn_RT, 'vprn_Lo' : vprn_Lo, 'irb_GW' : irb_GW})
      # L2DOMAIN-
IRB returns setupParams: vplsSvc_id, vprnSvc_id, servicetype, vni, sap, vprn_AS, vpr
n_RD, vprn_RT, vprn_Lo
      return {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'servicetype' :
servicetype, 'vni' : vni, 'sap_id' : sap_id, 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_R
D, 'vprn_RT' : vprn_RT, 'vprn_Lo' : vprn_Lo, 'irb_GW': irb_GW}
#------------------------------------------------------------------------------------
-------------
def teardown_script(setupParams):
    print ("These are the teardown_script setupParams: " + str(setupParams))
    servicetype = setupParams['servicetype']
    if servicetype == "L2DOMAIN-IRB":
      dyn.add_cli("""
        configure service
            vpls %(vplsSvc_id)s
               no description
               bgp-evpn
                   vxlan
                       shut
                   exit
                   no evi
                   exit
               no vxlan vni %(vni)s
               bgp
                  no route-distinguisher
                  no route-target
               exit
               no bgp
               no bgp-evpn
               sap %(sap_id)s
                  shutdown
                  exit
               no sap %(sap_id)s
               shutdown
               exit
            no vpls %(vplsSvc_id)s
            vprn %(vprnSvc_id)s
               interface lo1 shutdown
               no interface lo1
               interface "irbvpls-%(vplsSvc_id)s"
                  no vpls
                  shutdown
                  exit
               no interface "irbvpls-%(vplsSvc_id)s"
               shutdown
```

```
            exit
          no vprn %(vprnSvc_id)s
          exit
        """ % {'vplsSvc_id' : setupParams['vplsSvc_id'], 'vprnSvc_id' : setupParams['v
prnSvc_id'], 'vni' : setupParams['vni'], 'sap_id' : setupParams['sap_id']})
        return setupParams
d = {"script" : (setup_script, None, None, teardown_script)}
dyn.action(d)
```

The python script and policy are configured in a similar way as the previous example:

```
*A:pe-9# configure python
*A:pe-9>config>python# info
----------------------------------------------
        python-script "l2domain-irb_services" create
            primary-url "ftp://*:*@138.203.15.48/./l2domainIRB_service.py"
            no shutdown
        exit
        python-policy "py-l2-irb" create
            description "Python script to create L2-IRB domains"
            vsd script "l2domain-irb_services"
        exit
----------------------------------------------
```

It is also possible to create a python script that covers different domain types. The relevant part of the script is then addressed by using the following if-statement in the script:

```
    servicetype = vsdParams.get('servicetype')
        if servicetype == "L2DOMAIN-IRB":
           --- snipped ---
```

On the VSD, the following has to be provided:

(screenshots of this workflow are available in the VSP User Guide)

Create an L2 WAN service in the VSD:

- under Platform Config/Infrastructure, select the DC Gateway to add a WAN service
- select Service Type "layer 2" and select "IRB"
- select "Dynamic" configuration type to allow Fully Dynamic provisioning
- under Service Policy, provide the python policy configured on the DC Gateway
- provide a Name and Service-ID (the Service-ID will be the name of the dynamically created service domain on the DC Gateway)

Add the metadata to the WAN service:

- right-click the WAN service and select "inspect"

- a pop-up dialog box appears; select the "Metadata" tag and add the metadata info; for example,

  "rd=2:2,sap=1/1/1:1000,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1,irbGW=10.32.78.1/24"

Add permissions to Enterprise1. The WAN service should now be visible in Enterprise1.

Add permissions for a group of users to use this WAN service. Instantiate an L2 domain and attach the WAN service.

After the script has completed, there should be two new services created:

```
*A:pe-9# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
64000        VPLS      Up   Up   1          vpls64000
64001        VPRN      Up   Up   1

*A:pe-9# show service id 64000 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64000               Vpn Id            : 0
Service Type      : VPLS
Name              : vpls64000
Description       : vpls64000
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 07/22/2015 11:15:36
Last Mgmt Change  : 07/22/2015 11:15:36
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 64000
SAP Count         : 1                   SDP Bind Count    : 0
--- snipped ---
VSD Domain        : L2-IRB-Service-1
--- snipped ---

-------------------------------------------------------------------------------
BGP Information
-------------------------------------------------------------------------------
Vsi-Import        : None
Vsi-Export        : None
Route Dist        : 2:2
Oper Route Dist   : 2:2
Oper RD Type      : configured
Rte-Target Import : 65534:6985          Rte-Target Export : 65534:6985
Oper RT Imp Origin: configured          Oper RT Import    : 65534:6985
Oper RT Exp Origin: configured          Oper RT Export    : 65534:6985
PW-Template Id    : None
-------------------------------------------------------------------------------
```

```
--- snipped ---
-------------------------------------------------------------------------------
SAP 1/1/1:1000
-------------------------------------------------------------------------------
Service Id         : 64000
SAP                : 1/1/1:1000              Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                      Oper State        : Up
--- snipped ---
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1006636
===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
39.0.0.94              1006636        0            Yes     Up           No
-------------------------------------------------------------------------------
--- snipped ---

*A:pe-9# show service id 64001 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64001               Vpn Id            : 0
Service Type      : VPRN
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 07/22/2015 11:15:36
Last Mgmt Change  : 07/22/2015 11:15:36
Admin State       : Up                  Oper State        : Up

Route Dist.       : 65000:1             VPRN Type         : regular
Oper Route Dist   : 65000:1
Oper RD Type      : configured
AS Number         : 65000               Router Id         : 10.0.0.9
ECMP              : Enabled             ECMP Max Routes   : 1
--- snipped ---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : irbvpls-64000
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None
IP Addr/mask      : 10.32.78.1/24       Address Type      : Primary
--- snipped ---
Routed VPLS Details
VPLS Name        : vpls64000
Binding Status   : Up
--- snipped ---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : lo1
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None
IP Addr/mask      : 1.1.1.1/32          Address Type      : Primary
```

The dynamically created configuration can be inspected in enable-vsd-config mode (only enter the enable-vsd-config mode when absolutely required):

```
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            domain L2-IRB-Service-1 type l2-domain-irb create
                description "L2-IRB-Service-1"
                no shutdown
            exit
            service-range 64000 to 64999
        exit
        vpls 64000 customer 1 create
            description "vpls64000"
            allow-ip-int-bind
            exit
            vxlan vni 1006636 create
            exit
            bgp
                route-distinguisher 2:2
            exit
            bgp-evpn
                evi 64000
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            stp
                shutdown
            exit
            service-name "vpls64000"
            sap 1/1/1:1000 create
            exit
            vsd-domain "L2-IRB-Service-1"
            no shutdown
        exit
        vprn 64001 customer 1 create
            autonomous-system 65000
            route-distinguisher 65000:1
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:65000:1
            interface "irbvpls-64000" create
                address 10.32.78.1/24
                vpls "vpls64000"
                exit
            exit
            interface "lo1" create
                address 1.1.1.1/32
                loopback
            exit
```

```
            vsd-domain "L2-IRB-Service-1"
            no shutdown
        exit
-----------------------------------------------
```

MAC addresses can now be learned in the Nuage VSP/VSG and sent via MP-BGP to the DC Gateway:

```
*A:pe-9# show service id 64000 fdb detail
===============================================================================
Forwarding Database, Service 64000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
64000     1e:50:01:01:00:01 vxlan:                   Evpn     07/22/15 11:26:15
                            39.0.0.94:1006636
64000     1e:e2:ff:00:f9:3d cpm                      Intf     07/22/15 11:15:36
-------------------------------------------------------------------------------
```

The second MAC address is the GW-MAC that the VM or VSG-connected host will use to reach the interface on the VPRN:

```
*A:pe-9# show router 64001 route-table
===============================================================================
Route Table (Service: 64001)
===============================================================================
Dest Prefix[Flags]                             Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
1.1.1.1/32                                     Local   Local     00h17m10s  0
       lo1                                                           0
10.32.78.0/24                                  Local   Local     00h17m10s  0
       irbvpls-64000                                                 0

*A:pe-9# show router 64001 arp
===============================================================================
ARP Table (Service: 64001)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.32.78.1      1e:e2:ff:00:f9:3d 00h00m00s Oth[I] irbvpls-64000
10.32.78.100    1e:50:01:01:00:01 03h53m22s Dyn[I] irbvpls-64000
1.1.1.1         1e:e2:ff:00:00:00 00h00m00s Oth    lo1
```

Similar commands as shown in the previous section are available to obtain relevant information:

- *show service vsd domain "L2-IRB-Service-1" association* to obtain svc-IDs
- *show service id <id> all* to obtain RT/RD/VNI values
- *show service id <id> vxlan* to obtain VTEPs in the VPLS service
- *show service vxlan <vtep-ip>* to obtain svc-ID and VNI information

- *show router bgp routes evpn mac* to obtain MAC/IP/VNI/RT/NH information

The VM or VSG-connected Host should be able to ping the loopback interface of the VPRN service:

```
*A:ce1# ping 1.1.1.1 source 10.32.78.100
PING 1.1.1.1 56 data bytes
64 bytes from 1.1.1.1: icmp_seq=1 ttl=64 time=1.67ms.
64 bytes from 1.1.1.1: icmp_seq=2 ttl=64 time=1.83ms.
```

L3 VXLAN

An example python script for F-D XMPP provisioning of an L3 VXLAN type service (vrf-vxlan) is:

```
from alc import dyn

# example of metadata to be added in VSD WAN Service: "rd=3:3,vprnAS=65000,vprnRD=65
000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1"

# example of tools cli to test this script: tools perform service vsd evaluate-
script domain-name "l3dom1" type vrf-vxlan action setup policy "py-vrf-
vxlan" vni 1234 rt-i target:3:3 rt-
e target:3:3 metadata "rd=3:3,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprn
Lo=1.1.1.1"

# teardown example cli: tools perform service vsd evaluate-script domain-
name "l3dom1" type vrf-vxlan action teardown policy "py-vrf-vxlan" vni 1234 rt-
i target:3:3 rt-e target:3:3

def setup_script(vsdParams):

    print ("These are the VSD params: " + str(vsdParams))
    servicetype = vsdParams['servicetype']
    vni = vsdParams['vni']
    rt = vsdParams['rt']

# add "target:" if provisioned by VSD (VSD uses x:x format whereas tools command use
s target:x:x format)
    if not rt.startswith ('target'):
       rt = "target:"+rt
    metadata = vsdParams['metadata']

# remove trailing space at the end of the metadata
    metadata = metadata.rstrip()
    print ("VSD metadata" + str(metadata))
    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
    vplsSvc_id = dyn.select_free_id("service-id")
    vprnSvc_id = dyn.select_free_id("service-id")
    print ("this are the free svc ids picked up by the system: VPLS:" + vplsSvc_id +
 " + VPRN:" + vprnSvc_id)
```

```
    if servicetype == "VRF-VXLAN":

      rd = metadata['rd']
      vprn_AS = metadata ['vprnAS']
      vprn_RD = metadata ['vprnRD']
      vprn_RT = metadata ['vprnRT']
      vprn_Lo = metadata ['vprnLo']
      print ('servicetype, VPLS id, rt, vni, rd, VPRN id, vprn_AS, vprn_RD, vprn_RT,
 vprn_Lo:', servicetype, vplsSvc_id, rt, vni, rd, vprnSvc_id, vprn_AS, vprn_RD, vprn
_RT, vprn_Lo)
      dyn.add_cli("""
        configure router policy-options
           begin
             community _VSD_%(vplsSvc_id)s members %(rt)s
             policy-statement vsi_import_%(vplsSvc_id)s
                entry 10
                   from
                      family evpn
                      community _VSD_%(vplsSvc_id)s
                      exit
                   action accept
                   exit
                exit
             exit
             policy-statement vsi_export_%(vplsSvc_id)s
                entry 10
                   from
                      family evpn
                      exit
                      action accept
                      community add _VSD_%(vplsSvc_id)s
                      exit
                   exit
              exit
           commit
        exit

        configure service
           vpls %(vplsSvc_id)s customer 1 create
              allow-ip-int-bind
                 exit
              description vpls%(vplsSvc_id)s
              bgp
                 route-distinguisher %(rd)s
                 vsi-import vsi_import_%(vplsSvc_id)s
                 vsi-export vsi_export_%(vplsSvc_id)s
                 exit
              vxlan vni %(vni)s create
                 exit
              bgp-evpn
                 ip-route-advertisement
                 vxlan
                     no shut
                     exit
                 exit
              service-name vpls%(vplsSvc_id)s
              no shutdown
              exit
             exit
```

```
        exit

        configure service
           vprn %(vprnSvc_id)s customer 1 create
              autonomous-system %(vprn_AS)s
              route-distinguisher %(vprn_RD)s
              vrf-target %(vprn_RT)s
              interface "vpls-%(vplsSvc_id)s" create
                 vpls "vpls%(vplsSvc_id)s" evpn-tunnel
                 exit
              interface "lo1" create
                 address %(vprn_Lo)s/32
                 loopback
                 exit
              no shutdown
              exit
        exit

      """ % {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'vni' : vsdParams
['vni'], 'rt' : rt, 'rd' : metadata['rd'], 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_RD,
 'vprn_RT' : vprn_RT, 'vprn_Lo' : vprn_Lo})
      # VRF-
VXLAN returns setupParams: vplsSvc_id, vprnSvc_id, servicetype, vni, vprn_AS, vprn_R
D, vprn_RT, vprn_Lo
      return {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'servicetype' :
servicetype, 'vni' : vni, 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_RD, 'vprn_RT' : vprn
_RT, 'vprn_Lo' : vprn_Lo}
#-------------------------------------------------------------------------------
-------------

def teardown_script(setupParams):
    print ("These are the teardown_script setupParams: " + str(setupParams))
    servicetype = setupParams['servicetype']
    if servicetype == "VRF-VXLAN":
      dyn.add_cli("""
        configure service
            vpls %(vplsSvc_id)s
               no description
               bgp-evpn
                   vxlan
                       shut
                       exit
                   no evi
                   exit
               no vxlan vni %(vni)s
               bgp
                  no route-distinguisher
                  no route-target
                  exit
               no bgp
               no bgp-evpn
               shutdown
               exit
            no vpls %(vplsSvc_id)s
            vprn %(vprnSvc_id)s
               interface lo1 shutdown
               no interface lo1
               interface "vpls-%(vplsSvc_id)s"
                  vpls "vpls%(vplsSvc_id)s"
```

```
                    no evpn-tunnel
                    exit
                 no vpls
                 shutdown
                 exit
              no interface "vpls-%(vplsSvc_id)s"
              shutdown
          exit
       no vprn %(vprnSvc_id)s
       exit
       configure router policy-options
           begin
           no community _VSD_%(vplsSvc_id)s
           no policy-statement vsi_import_%(vplsSvc_id)s
           no policy-statement vsi_export_%(vplsSvc_id)s
           commit
       exit

     """ % {'vplsSvc_id' : setupParams['vplsSvc_id'], 'vprnSvc_id' : setupParams['v
prnSvc_id'], 'vni' : setupParams['vni']})
     return setupParams

d = {"script" : (setup_script, None, None, teardown_script)}

dyn.action(d)
```

The python script and policy are configured in a similar way as the previous example:

```
*A:pe-9# configure python
*A:pe-9>config>python# info
----------------------------------------------
        python-script "vrf-vxlan_services" create
            primary-url "ftp://*:*@138.203.15.48/./vrf-vxlan_service.py"
            no shutdown
        exit
        python-policy "py-vrf-vxlan" create
            description "Python script to create vrf-vxlan domains"
            vsd script "l3vxlan"
        exit
----------------------------------------------
```

The following steps are required on the VSD to provision the WAN service:

(screenshots of this workflow are available in the VSP User Guide)

Create an L3 WAN service in the VSD:

- under Platform Config / Infrastructure, select the DC Gateway to add a WAN service
- select Service Type "layer 3"
- select "Dynamic" configuration type to allow Fully Dynamic provisioning
- under Service Policy, provide the python policy configured on the DC Gateway

- provide a Name and Service-ID (the Service-ID will be the name of the dynamically created service domain on the DC Gateway)

Add the metadata to the WAN service:

- right-click the WAN service and select "inspect"
- a pop-up dialog box appears; select the "Metadata" tag and add the metadata info; for example,

  "rd=3:3,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1"

Add permissions to Enterprise1. The WAN service should now be visible in Enterprise1.

Add permissions for a group of users to use this WAN service. Instantiate an L3 domain and attach the WAN service.

After the script has completed, there should be two new services:

```
*A:pe-9# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
64000        VPLS      Up   Up   1          vpls64000
64001        VPRN      Up   Up   1


*A:pe-9# show service id 64000 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64000               Vpn Id            : 0
Service Type      : VPLS
Name              : vpls64000
Description       : vpls64000
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 10/14/2015 18:38:44
Last Mgmt Change  : 10/14/2015 18:38:44
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 64000
SAP Count         : 0                   SDP Bind Count    : 0
--- snipped ---
VSD Domain        : L3-service-1
--- snipped ---

-------------------------------------------------------------------------------
BGP Information
-------------------------------------------------------------------------------
Vsi-Import        : vsi_import_64000
Vsi-Export        : vsi_export_64000
Route Dist        : 3:3
```

```
Oper Route Dist   : 3:3
Oper RD Type      : configured
Rte-Target Import : None                Rte-Target Export : None
Oper RT Imp Origin: vsi                 Oper RT Import    : None
Oper RT Exp Origin: vsi                 Oper RT Export    : None
PW-Template Id    : None
-------------------------------------------------------------------------------
--- snipped ---
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 119281

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
39.0.0.94              119281         1            No      Up           No
-------------------------------------------------------------------------------
--- snipped ---
```

The script dynamically creates a VSI-import and VSI-export policy and links it to an RT that was dynamically created by the VSC/VSG:

```
*A:pe-9# show router policy
===============================================================================
Route Policies
===============================================================================
Policy                          Description
-------------------------------------------------------------------------------
vsi_export_64000
vsi_import_64000
-------------------------------------------------------------------------------
Policies : 2
===============================================================================


*A:pe-9# show router policy "vsi_import_64000"
    entry 10
        from
            community "_VSD_64000"
            family evpn
        exit
        action accept
        exit
    exit


*A:pe-9# show router policy "vsi_export_64000"
    entry 10
        from
            family evpn
        exit
        action accept
            community add "_VSD_64000"
        exit
    exit


*A:pe-9# show router policy community "_VSD_64000"
```

```
community "_VSD_64000" members "65534:38619"

*A:pe-9# show service id 64001 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64001               Vpn Id            : 0
Service Type      : VPRN
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 10/14/2015 18:38:44
Last Mgmt Change  : 10/14/2015 18:38:44
Admin State       : Up                  Oper State        : Up

Route Dist.       : 65000:1             VPRN Type         : regular
Oper Route Dist   : 65000:1
Oper RD Type      : configured
AS Number         : 65000               Router Id         : 10.0.0.9
ECMP              : Enabled             ECMP Max Routes   : 1
Auto Bind Tunnel
Resolution        : any
--- snipped ---
Vrf Target        : target:65000:1
--- snipped ---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : vpls-64000
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None

IP Addr/mask      : Not Assigned
--- snipped ---
Routed VPLS Details
VPLS Name        : vpls64000
Binding Status   : Up
--- snipped ---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : lo1
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None
IP Addr/mask      : 1.1.1.1/32          Address Type      : Primary
```

The dynamically created configuration can be inspected in enable-vsd-config mode (only enter the enable-vsd-config mode when required):

```
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            domain L3-service-1 type vrf-vxlan create
                description "L3-service-1"
```

```
                no shutdown
            exit
            service-range 64000 to 64999
        exit
        vpls 64000 customer 1 create
            description "vpls64000"
            allow-ip-int-bind
            exit
            vxlan vni 119281 create
            exit
            bgp
                route-distinguisher 3:3
                vsi-export "vsi_export_64000"
                vsi-import "vsi_import_64000"
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            stp
                shutdown
            exit
            service-name "vpls64000"
            vsd-domain "L3-service-1"
            no shutdown
        exit
        vprn 64001 customer 1 create
            autonomous-system 65000
            route-distinguisher 65000:1
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:65000:1
            interface "vpls-64000" create
                vpls "vpls64000"
                    evpn-tunnel
                exit
            exit
            interface "lo1" create
                address 1.1.1.1/32
                loopback
            exit
            vsd-domain "L3-service-1"
            no shutdown
        exit
-----------------------------------------------
```

The EVPN tunnel NH-MAC addresses can now be learned in the Nuage VSP/VSG and sent via MP-BGP to the DC Gateway:

```
*A:pe-9# show service id 64000 fdb detail
Forwarding Database, Service 64000
===============================================================================
```

```
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
64000     00:00:27:00:00:5e vxlan:                   Evpn     07/22/15 13:38:47
                            39.0.0.94:119281
64000     1e:e2:ff:00:f9:3d cpm                      Intf     07/22/15 13:38:44
-------------------------------------------------------------------------------
```

The first MAC entry address is the tunnel NH-MAC for the VSG and the second is the address for the DC Gateway for EVPN-tunnel service 64000:

```
*A:pe-9# show router 64001 route-table
===============================================================================
1.1.1.1/32                                    Local   Local     00h09m08s  0
       lo1                                                          0
10.32.78.0/24                                 Remote  BGP EVPN  00h09m05s  169
       vpls-64000 (ET-00:00:27:00:00:5e)                          0
10.32.78.100/32                               Remote  BGP EVPN  00h00m04s  169
       vpls-64000 (ET-00:00:27:00:00:5e)                          0
-------------------------------------------------------------------------------
```

The following commands are useful to obtain relevant information:

- *show service vsd domain "L3-service-1" association* to obtain svc-IDs
- *show service id <id> all* to obtain RT/RD/VNI values
- *show service id <id> vxlan* to obtain VTEPs in the VPLS service
- *show service vxlan <vtep-ip>* to obtain svc-id and VNI information

Relevant EVPN tunnel NH-MAC/VNI/RT/NH information is also in the EVPN BGP RIB:

```
*A:pe-9# show router bgp routes evpn mac detail
--- snipped ---

Modified Attributes

Network        : N/A
Nexthop        : 39.0.0.94
From           : 39.0.0.94
Res. Nexthop   : 192.168.39.94
Local Pref.    : 200                    Interface Name : toNuage
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65534:38619 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 39.0.0.94
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 119281
```

```
IP Address     : N/A
Route Dist.    : 65534:29625
Mac Address    : 00:00:27:00:00:5e
MPLS Label1    : VNI 119281            MPLS Label2    : N/A
Route Tag      : 0
--- snipped ---
```

VM/VSG-connected Host and network information is also in the EVPN BGP RIB:

```
*A:pe-9# show router bgp routes evpn ip-prefix detail
--- snipped ---
Modified Attributes

Network        : N/A
Nexthop        : 39.0.0.94
From           : 39.0.0.94
Res. Nexthop   : 192.168.39.94
Local Pref.    : 200                   Interface Name : toNuage
Aggregator AS  : None                  Aggregator     : None
Atomic Aggr.   : Not Atomic            MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65534:38619 ext:30b:220000000000
                 ext:30b:100b00b00000 bgp-tunnel-encap:VXLAN
                 mac-nh:00:00:27:00:00:5e
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 39.0.0.94
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 119281
Gateway Address: 00:00:27:00:00:5e
Prefix         : 10.32.78.100/32
Route Dist.    : 39.0.0.94:10269
MPLS Label     : VNI 119281
Route Tag      : 0
--- snipped ---

Modified Attributes

Network        : N/A
Nexthop        : 39.0.0.94
From           : 39.0.0.94
Res. Nexthop   : 192.168.39.94
Local Pref.    : 200                   Interface Name : toNuage
Aggregator AS  : None                  Aggregator     : None
Atomic Aggr.   : Not Atomic            MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65534:38619 ext:30b:220000000000
                 ext:30b:100b00b00000 bgp-tunnel-encap:VXLAN
                 mac-nh:00:00:27:00:00:5e
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 39.0.0.94
Flags          : Used  Valid  Best  IGP
```

```
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 119281
Gateway Address: 00:00:27:00:00:5e
Prefix         : 10.32.78.0/24
Route Dist.    : 39.0.0.94:10269
MPLS Label     : VNI 119281
Route Tag      : 0
--- snipped ---
```

The VM or VSG-connected Host should be able to ping the loopback interface of the VPRN service:

```
*A:ce1# ping 1.1.1.1 source 10.32.78.100
PING 1.1.1.1 56 data bytes
64 bytes from 1.1.1.1: icmp_seq=1 ttl=64 time=1.67ms.
64 bytes from 1.1.1.1: icmp_seq=2 ttl=64 time=1.83ms.
```

Troubleshooting and debug commands

When testing/troubleshooting F-D XMPP provisioning, the following show/tools/debug commands can be useful:

- tools perform service vsd evaluate-script
- tools perform service vsd fd-domain-sync <full|diff>
- tools perform service vsd domain refresh-config
- tools perform python-script reload
- tools dump service vsd-services command-list

- debug python python-script
- debug vsd scripts event/instance
- debug system xmpp
- debug router bgp update

- show service vsd domain
- show service vsd script
- show service vsd summary
- show system vsd

- show xmpp vsd
- show python python-policy <name> {association}
- show python python-script <name> {association|source-in-use}

- show service vxlan [<vtep-ip>]
- show service service-using {<service-type>}
- show service id route-table
- show service id fdb detail
- show service id proxy-arp detail
- show router [<router-instance>] route-table
- show router [<router-instance>] arp
- show router bgp routes bgp <mac|ip-prefix|inclusive-mcast> {detail}

- log-id 99

Conclusion

The fully dynamic VSD integration model allows for automated provisioning of breakout services on the 7x50 DC Gateway. Different domain types (l2-domain/l2-domain-irb/vrf-vxlan/vrf-gre) are supported. This chapter has shown how to construct, load, and test python scripts for this feature. It also described how to configure WAN services on the VSD and how to verify the dynamically created services.

Inter-AS Model C for VLL

This chapter describes advanced inter-AS model C for VLL configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 8.0.R4. The CLI in this edition corresponds to release 14.0.R3.

Overview

SR OS supports RFC 3107, Carrying Label Information in BGP-4, including VLL/VPLS. BGP SDPs can also be used with PBB-VPLS services.

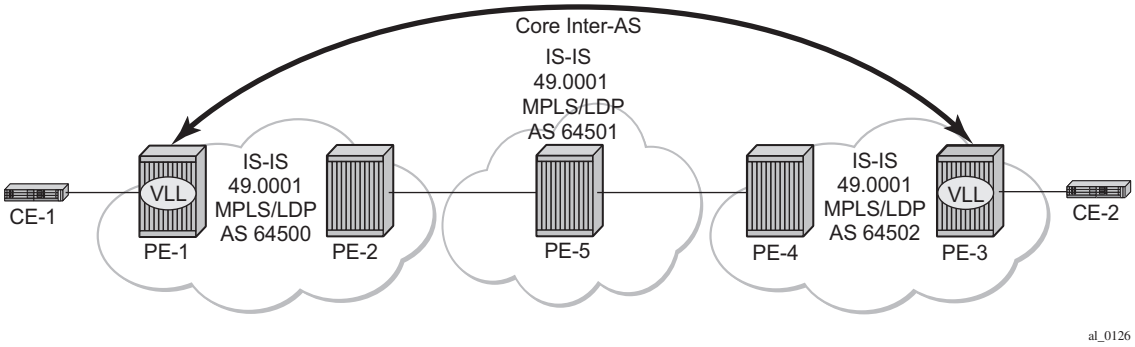
ISPs are looking for mechanisms to implement the VLL and VPLS services outside an autonomous system (AS). Service providers may have inter-AS operation as a consequence of delivering inter- provider VLL/VPLS or because they use multiple ASs as a result of acquisitions and merges.

The objective of this chapter is to describe the interconnection of VLL services across multiple ASs.

Network Setup

Figure 283 shows the network setup used.

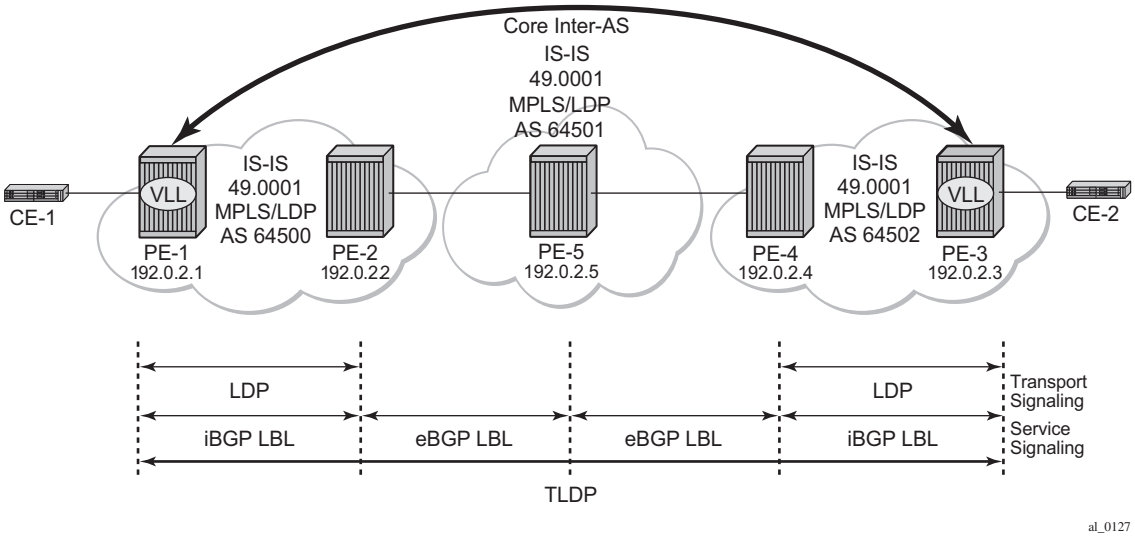
Figure 283 **Network Setup - Inter-AS Model C for VLL**



The network topology displayed in Figure 283 consists of three sites in different ASs with each site using 7750 SRs.

In AS 64500, there are PE-1 and PE-2, AS 64501 has PE-5 and AS 64502 has PE-4 and PE-3. There is a business customer with two remote locations, Site A and Site B, with customer edge (CE) devices CE-1 connected to the AS 64500 via PE-1 and CE-2 connected to the AS 64502 via PE-3. A VLL service is configured between PE-1 and PE-3 to connect site A and site B.

Figure 284 **Inter-AS Model C for VLL**

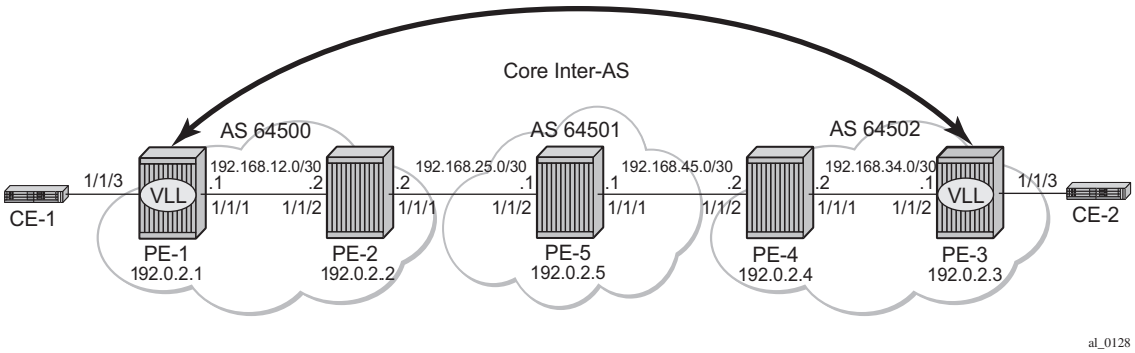


Configuration

This section describes all of the relevant configuration tasks for the detailed setup shown in the Figure 285. In this particular example the following protocols are assumed to be already configured.

- IS-IS as the IGP with all the nodes being level Level1/Level 2.
- LDP as the MPLS protocol to signal the transport tunnels.

Figure 285 **Network Setup Configuration**



BGP Configuration

The following CLI output shows the BGP configuration — iBGP and eBGP — required for the PE routers to implement VLL Inter-AS.

The configuration on PE-5 in AS 64501 is the following:

```
*A:PE-5# configure router
        bgp
            min-route-advertisement 1
            rapid-withdrawal
            group "EBGP"
                family ipv4
                type external
                local-as 64501
                neighbor 192.168.25.1
                    peer-as 64500
                    advertise-label ipv4
                exit
                neighbor 192.168.45.1
                    peer-as 64502
                    advertise-label ipv4
                exit
```

```
            exit
        exit
```

The **advertise-label ipv4** statement must be configured so that MPLS labels are carried with IPv4 network layer reachability information (NLRIs).

The configuration of PE-2 in AS 64500 is as follows:

```
*A:PE-2# configure router
        bgp
            min-route-advertisement 1
            rapid-withdrawal
            group "EBGP"
                family ipv4
                type external
                export "EXPORT-SYSTEMS"
                local-as 64500
                neighbor 192.168.25.2
                    peer-as 64501
                    advertise-label ipv4
                exit
            exit
            group "IBGP"
                family ipv4
                type internal
                neighbor 192.0.2.1
                    next-hop-self
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
```

The configuration of PE-4 in AS 64502 is as follows:

```
*A:PE-4# configure router
         bgp
            min-route-advertisement 1
            export "EXPORT-SYSTEMS"
            rapid-withdrawal
            group "EBGP"
                family ipv4
                type external
                local-as 64502
                neighbor 192.168.45.2
                    peer-as 64501
                    advertise-label ipv4
                exit
            exit
            group "IBGP"
                family ipv4
                type internal
                neighbor 192.0.2.3
                    next-hop-self
                    advertise-label ipv4
```

```
            exit
         exit
         no shutdown
      exit
```

To complement the BGP setup, the configurations of PE-1 and PE-3 (the PEs to which the CEs are connected in AS 64500 and AS 64502), respectively, are displayed:

PE-1:

```
*A:PE-1# configure router
         bgp
            min-route-advertisement 1
            rapid-withdrawal
            group "IBGP"
                family ipv4
                type internal
                neighbor 192.0.2.2
                    next-hop-self
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
```

PE-3:

```
*A:PE-3# configure router
         bgp
            min-route-advertisement 1
            rapid-withdrawal
            group "IBGP"
                family ipv4
                type internal
                neighbor 192.0.2.4
                    next-hop-self
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
```

Policy Configuration

The export policy on the PE-2 and PE-4 peering determines the system addresses leaked to the remote AS. It is worth noting here that it is important to modify the default origin attribute from incomplete to IGP in dual-homing scenarios otherwise the iBGP route will be always preferred over the eBGP route.

This is the configuration on PE-2:

```
*A:PE-2# configure router
        policy-options
            begin
            prefix-list "SYSTEMS-AS64500"
                prefix 192.0.2.1/32 exact
                prefix 192.0.2.2/32 exact
            exit
            policy-statement "EXPORT-SYSTEMS"
                entry 10
                    from
                        prefix-list "SYSTEMS-AS64500"
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
```

This is the configuration on PE-4:

```
*A:PE-4# configure router
        policy-options
            begin
            prefix-list "SYSTEMS-AS64502"
                prefix 192.0.2.3/32 exact
                prefix 192.0.2.4/32 exact
            exit
            policy-statement "EXPORT-SYSTEMS"
                entry 10
                    from
                        prefix-list "SYSTEMS-AS64502"
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
```

Service Configuration

Once BGP is configured, the configuration requires the service to be defined (Epipe 1). The focus here is a VLL service (noting that it is also possible to have a similar configuration with VPLS services).

The following CLI shows the service level configuration on PE-1:

```
*A:PE-1# configure service
        sdp 13 mpls create
            far-end 192.0.2.3
            bgp-tunnel
            no shutdown
        exit
        epipe 1 customer 1 create
            description "Tunnel-PE-1-PE-3"
            sap 1/1/3:1 create
            exit
            spoke-sdp 13:1 create
            exit
            no shutdown
        exit
```

The following CLI shows the service level configuration on PE-3:

```
*A:PE-3# configure service
        sdp 31 mpls create
            far-end 192.0.2.1
            bgp-tunnel
            no shutdown
        exit
        epipe 1 customer 1 create
            description "Tunnel-PE-3-PE-1"
            sap 1/1/3:1 create
            exit
            spoke-sdp 31:1 create
            exit
            no shutdown
        exit
```

Show Commands and Troubleshooting

On PE-3, BGP tunnels exist to the remote AS system addresses that are using LDP as a transport mechanism and the configuration of end-to-end SDPs over which TLDP service labels are exchanged.

The following output shows the SDP and LDP information:

```
*A:PE-1# show service sdp
===============================================================================
```

```
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
13     0       1552    192.0.2.3        Up   Up          MPLS    B     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
===============================================================================
*A:PE-1#

*A:PE-1# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#

*A:PE-1# show router ldp session ipv4
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   95        98        0d 00:04:02
192.0.2.3:0         Targeted  Established   9         10        0d 00:00:27
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
*A:PE-1#
```

The route-table shows that the system IP address of PE-3 is reachable using a BGP tunnel:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     00h06m03s  0
       system                                                          0
192.0.2.2/32                                    Remote  ISIS      00h05m10s  15
       192.168.12.2                                                    10
192.0.2.3/32                                    Remote  BGP       00h00m53s  170
       192.0.2.2 (tunneled)                                            0
192.0.2.4/32                                    Remote  BGP       00h00m53s  170
```

```
       192.0.2.2 (tunneled)                                              0
192.168.12.0/30                              Local   Local     00h06m03s  0
       int-PE-1-PE-2                                                     0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

The following tunnel-table shows the details of the LDP, SDP, and BGP tunnels.

```
*A:PE-1# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.2/32      ldp       MPLS  65537     9        192.168.12.2    10
192.0.2.3/32      sdp       MPLS  13        5        192.0.2.3       0
192.0.2.3/32      bgp       MPLS  262145    12       192.0.2.2       1000
192.0.2.4/32      bgp       MPLS  262146    12       192.0.2.2       1000
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-1#
```

The service details for Epipe 1 are as follows:

```
*A:PE-1# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : Tunnel-PE-1-PE-3
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/27/2016 07:53:04
Last Mgmt Change  : 05/27/2016 06:48:04
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
```

```
sap:1/1/3:1                                 q-tag        1518    1518    Up   Up
sdp:13:1 S(192.0.2.3)                       Spok         0       1552    Up   Up
===============================================================================
*A:PE-1#
```

A ping is used to verify the IP connectivity from PE-1 to the system IP address of PE-3:

```
*A:PE-1# ping 192.0.2.3
PING 192.0.2.3 56 data bytes
64 bytes from 192.0.2.3: icmp_seq=1 ttl=64 time=1.91ms.
64 bytes from 192.0.2.3: icmp_seq=2 ttl=64 time=2.06ms.
64 bytes from 192.0.2.3: icmp_seq=3 ttl=64 time=2.02ms.
64 bytes from 192.0.2.3: icmp_seq=4 ttl=64 time=2.01ms.
64 bytes from 192.0.2.3: icmp_seq=5 ttl=64 time=2.02ms.

---- 192.0.2.3 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.91ms, avg = 2.01ms, max = 2.06ms, stddev = 0.050ms
*A:PE-1#
```

The same commands on PE-3 result in the following output:

```
*A:PE-3# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
31     0       1552    192.0.2.1        Up   Up          MPLS    B     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
===============================================================================
*A:PE-3#

*A:PE-3# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#

*A:PE-3# show router ldp session ipv4
===============================================================================
LDP IPv4 Sessions
```

```
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.1:0         Targeted  Established   12        13        0d 00:00:44
192.0.2.4:0         Link      Established   98        100       0d 00:04:11
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
*A:PE-3#

*A:PE-3# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  BGP       00h01m16s  170
       192.0.2.4 (tunneled)                                         0
192.0.2.2/32                                    Remote  BGP       00h01m16s  170
       192.0.2.4 (tunneled)                                         0
192.0.2.3/32                                    Local   Local     00h06m10s  0
       system                                                       0
192.0.2.4/32                                    Remote  ISIS      00h05m20s  15
       192.168.34.2                                                 10
192.168.34.0/30                                 Local   Local     00h06m11s  0
       int-PE-3-PE-4                                                0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-3#

*A:PE-3# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      sdp       MPLS  31        5        192.0.2.1       0
192.0.2.1/32      bgp       MPLS  262145    12       192.0.2.4       1000
192.0.2.2/32      bgp       MPLS  262146    12       192.0.2.4       1000
192.0.2.4/32      ldp       MPLS  65537     9        192.168.34.2    10
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-3#

*A:PE-3# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
```

```
Description       : Tunnel-PE-3-PE-1
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/27/2016 07:52:31
Last Mgmt Change  : 05/27/2016 06:47:41
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

------------------------------------------------------------------------------
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                             Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:1/1/3:1                            q-tag        1518    1518    Up   Up
sdp:31:1 S(192.0.2.1)                  Spok         0       1552    Up   Up
==============================================================================
*A:PE-3#

*A:PE-3# ping 192.0.2.1
PING 192.0.2.1 56 data bytes
64 bytes from 192.0.2.1: icmp_seq=1 ttl=64 time=1.83ms.
64 bytes from 192.0.2.1: icmp_seq=2 ttl=64 time=2.06ms.
64 bytes from 192.0.2.1: icmp_seq=3 ttl=64 time=2.01ms.
64 bytes from 192.0.2.1: icmp_seq=4 ttl=64 time=2.08ms.
64 bytes from 192.0.2.1: icmp_seq=5 ttl=64 time=2.15ms.

---- 192.0.2.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.83ms, avg = 2.03ms, max = 2.15ms, stddev = 0.107ms
*A:PE-3#
```

On PE-3, the BGP route to the system IP address of PE-1 can been seen with PE-4 as the next hop:

```
*A:PE-3# show router bgp routes
==============================================================================
 BGP Router ID:192.0.2.3        AS:64502       Local AS:64502
==============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

==============================================================================
BGP IPv4 Routes
==============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
------------------------------------------------------------------------------
u*>i  192.0.2.1/32                                     100         None
      192.0.2.4                                        None        262140
```

```
      64501 64500
u*>i  192.0.2.2/32                                        100         None
      192.0.2.4                                           None        262141
      64501 64500
*i    192.0.2.3/32                                        100         10
      192.0.2.4                                           None        262139
      No As-Path
*i    192.0.2.4/32                                        100         None
      192.0.2.4                                           None        262138
      No As-Path
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-3#
```

Again on PE-3, the FIB on slot 1 shows that the system IP address of PE-1 is reachable using BGP over an LDP transport to PE-4:

```
*A:PE-3# show router fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                              Protocol
    NextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                                BGP
    192.0.2.4 (Transport:LDP)
192.0.2.2/32                                                BGP
    192.0.2.4 (Transport:LDP)
192.0.2.3/32                                                LOCAL
    192.0.2.3 (system)
192.0.2.4/32                                                ISIS
    192.168.34.2 (int-PE-3-PE-4)
192.168.34.0/30                                             LOCAL
    192.168.34.0 (int-PE-3-PE-4)
-------------------------------------------------------------------------------
Total Entries : 5
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

The show commands on router PE-5 in AS 64501 are as follows:

```
T*A:PE-5# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.5        AS:64501       Local AS:64501
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 2
Total BGP Paths         : 9           Total Path Memory           : 1720
Total IPv4 Remote Rts   : 8           Total IPv4 Rem. Active Rts  : 4
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0
```

```
Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
===============================================================================
BGP Summary
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.168.25.1
                64500      14    0 00h03m44s 4/2/4 (IPv4)
                           14    0
192.168.45.1
                64502      15    0 00h03m44s 4/2/4 (IPv4)
                           14    0
-------------------------------------------------------------------------------
*A:PE-5#

*A:PE-5# show router bgp routes
===============================================================================
 BGP Router ID:192.0.2.5        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.0.2.1/32                                       None        10
      192.168.25.1                                       None        262141
      64500
i     192.0.2.1/32                                       None        None
      192.168.45.1                                       None        262140
```

```
      64502 64501 64500
u*>i  192.0.2.2/32                                        None        None
      192.168.25.1                                        None        262140
      64500
i     192.0.2.2/32                                        None        None
      192.168.45.1                                        None        262141
      64502 64501 64500
u*>i  192.0.2.3/32                                        None        10
      192.168.45.1                                        None        262139
      64502
i     192.0.2.3/32                                        None        None
      192.168.25.1                                        None        262138
      64500 64501 64502
u*>i  192.0.2.4/32                                        None        None
      192.168.45.1                                        None        262138
      64502
i     192.0.2.4/32                                        None        None
      192.168.25.1                                        None        262139
      64500 64501 64502
-------------------------------------------------------------------------------
Routes : 8
===============================================================================
*A:PE-5#

*A:PE-5# show router bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.168.25.1                  262140         262142         External
192.168.25.1                  262141         262143         External
192.168.45.1                  262138         262140         External
192.168.45.1                  262139         262141         External
-------------------------------------------------------------------------------
Total Labels allocated:   4
===============================================================================
*A:PE-5#

*A:PE-5# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age         Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP       00h03m17s   170
       192.168.25.1                                               0
192.0.2.2/32                                  Remote  BGP       00h03m17s   170
       192.168.25.1                                               0
192.0.2.3/32                                  Remote  BGP       00h03m04s   170
       192.168.45.1                                               0
192.0.2.4/32                                  Remote  BGP       00h03m04s   170
       192.168.45.1                                               0
192.0.2.5/32                                  Local   Local     00h07m44s   0
       system                                                     0
192.168.25.0/30                               Local   Local     00h07m44s   0
```

```
       int-PE-5-PE-2                                                    0
192.168.45.0/30                                Local   Local     00h07m44s  0
       int-PE-5-PE-4                                                    0
-------------------------------------------------------------------------------
No. of Routes: 7
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-5#
```

The commands on PE-2 are as follows:

```
*A:PE-2# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 2           Total Peers                 : 2
Total BGP Paths         : 7           Total Path Memory           : 1320
Total IPv4 Remote Rts   : 4           Total IPv4 Rem. Active Rts  : 2
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0

===============================================================================
BGP Summary
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
```

```
192.0.2.1
                64500      14    0 00h05m07s 0/0/2 (IPv4)
                           16    0
192.168.25.2
                64501      13    0 00h03m34s 4/2/4 (IPv4)
                           15    0
-------------------------------------------------------------------------------
*A:PE-2#

*A:PE-2# show router bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.1                     0              262141         Internal
192.0.2.2                     0              262140         Edge
192.168.25.2                  262140         262139         External
192.168.25.2                  262141         262138         External
-------------------------------------------------------------------------------
Total Labels allocated:   4
===============================================================================
*A:PE-2#

*A:PE-2# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                 Type    Proto     Age        Pref
      Next Hop[Interface Name]                                       Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                       Remote  ISIS      00h07m12s  15
       192.168.12.1                                                      10
192.0.2.2/32                                       Local   Local     00h07m56s  0
       system                                                            0
192.0.2.3/32                                       Remote  BGP       00h02m55s  170
       192.168.25.2                                                      0
192.0.2.4/32                                       Remote  BGP       00h02m55s  170
       192.168.25.2                                                      0
192.168.12.0/30                                    Local   Local     00h07m56s  0
       int-PE-2-PE-1                                                     0
192.168.25.0/30                                    Local   Local     00h07m56s  0
       int-PE-2-PE-5                                                     0
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

The show commands on PE-4:

```
*A:PE-4# show router bgp summary
```

```
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 2           Total Peers                 : 2
Total BGP Paths         : 7           Total Path Memory           : 1320
Total IPv4 Remote Rts   : 4           Total IPv4 Rem. Active Rts  : 2
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
===============================================================================
BGP Summary
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.3
                64502      11    0 00h03m47s 0/0/4 (IPv4)
                           16    0
192.168.45.2
                64501      13    0 00h03m38s 4/2/4 (IPv4)
                           16    0
-------------------------------------------------------------------------------
*A:PE-4#

*A:PE-4# show router bgp routes
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked
 Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```
===============================================================================
BGP IPv4 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  192.0.2.1/32                                      None        None
      192.168.45.2                                      None        262143
      64501 64500
u*>i  192.0.2.2/32                                      None        None
      192.168.45.2                                      None        262142
      64501 64500
i     192.0.2.3/32                                      None        None
      192.168.45.2                                      None        262141
      64501 64502
i     192.0.2.4/32                                      None        None
      192.168.45.2                                      None        262140
      64501 64502
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-4#

*A:PE-4# show router bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.168.45.2                  262142         262141         External
192.168.45.2                  262143         262140         External
192.0.2.3                     0              262139         Internal
192.0.2.4                     0              262138         Edge
-------------------------------------------------------------------------------
Total Labels allocated:   4
===============================================================================
*A:PE-4#
```

Conclusion

The BGP tunnel based SDP binding is allowed for VLL and VPLS services, including PBB-VPLS. Using RFC 3107, it is possible to implement inter-AS Model C VLLs.

The example used in this chapter illustrates the configuration of VLL Inter-AS for access CE sites. Troubleshooting commands also have been shown to verify all the procedures.

LDP VPLS using BGP-Auto Discovery

This chapter provides information about LDP VPLS using BGP-Auto Discovery.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 9.0.R3. The CLI in this edition is based on release 14.0.R3. There are no prerequisites for this configuration.

Summary

MPLS-based virtual private LAN services (VPLS) may have many different provisioning models to allow the signaling of pseudowires between provider edge (PE) routers containing VPLS instances.

Network management system (NMS) provisioning using label distribution protocol (LDP) signaling is a well understood method of provisioning of Layer 2 VPLS services as described in RFC 4762. This relies on the provisioning of pseudowires between VPLS instances using LDP signaling with a common virtual circuit (VC) identifier within the label mapping message to instantiate pseudowires.

Border gateway protocol (BGP) auto discovery (RFC 6074) is an alternative method of provisioning of Layer 2 PE routers containing VPLS service instances to those described above where PEs in a common VPLS instance are automatically discovered using BGP auto discovery (BGP-AD) techniques.

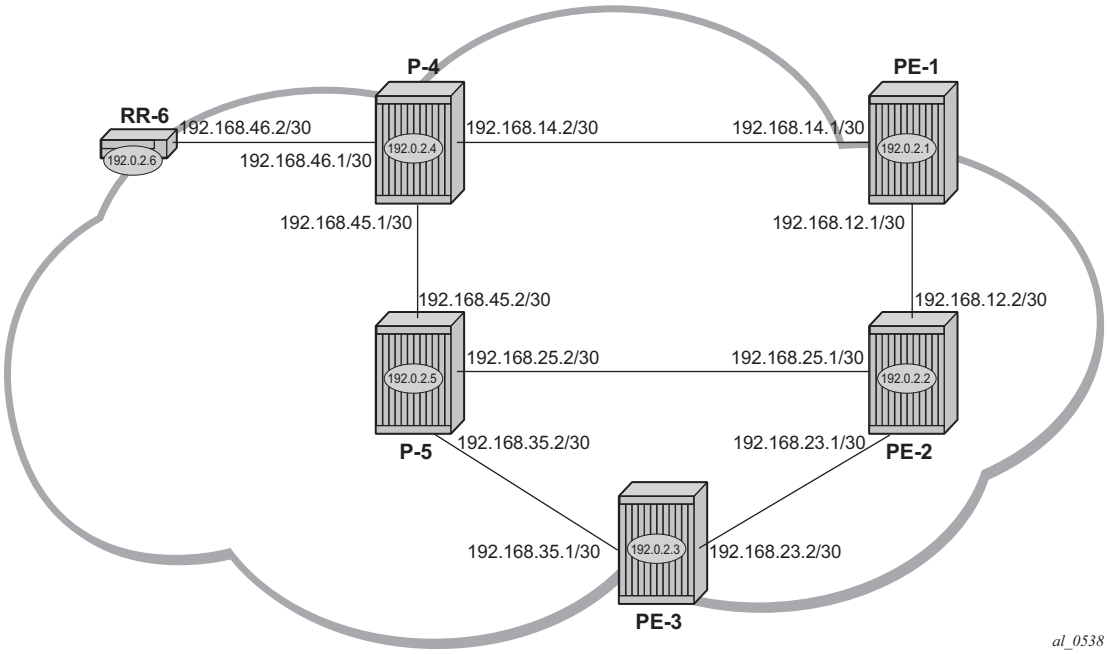
Each PE router advertises the presence of VPLS instances to other PE routers using defined parameters within a BGP update message.

LDP is used as the pseudowire signaling protocol and relies on the auto-discovery of VPLS endpoints to instantiate pseudowires instead of manually provisioning virtual circuits. Locally configured parameters, along with BGP learned parameters, are used to determine local and remote VPLS endpoints, which are used by LDP to signal service labels to peer routers.

Knowledge of BGP-auto-discovery RFC 6074 architecture and functionality, RFC 4447 Pseudo-wire set-up using label distribution protocol is assumed throughout this section, as well as knowledge of multi-protocol BGP (MP-BGP).

Overview

Figure 286 **Example Topology**



The example topology is displayed in Figure 286. The setup uses six 7x50 nodes located in the same autonomous system (AS). There are three PEs and RR-6 will act as a route reflector for the AS. The PE routers are all VPLS aware. The provider (P) routers are VPLS unaware and do not take part in the BGP process. A full mesh VPLS between PE-1, PE-2 and PE-3 is described.

The following configuration tasks should be completed as a pre-requisite:

- IS-IS or OSPF should be enabled on all network interfaces between each of the PE/P routers and route reflector.

- MPLS should be configured on all interfaces between PE and P routers; MPLS is not required between P-4 and RR-6.
- LDP should be configured on interfaces between PE and P routers. It is not required between P-4 and RR-6.
- RSVP protocol is disabled by default, so the RSVP protocol should be enabled.

BGP-AD

In this architecture, a VPLS service is a collection of local VPLS instances present on a number of PEs in a provider network. In this context, VPLS-aware devices are PE routers. Each VPLS instance has a unique identifier known as the VPLS identifier (VPLS-id). All PEs that have this VPLS instance present will have a common VPLS-id configured.

Each VPLS instance within a PE contains a virtual switching instance (VSI). The VPLS attachment circuits and pseudowires are associated with the VSI. Each VSI within a given VPLS has a unique identifier called the VSI identifier (VSI-id) and is a concatenation of the VPLS-id plus an IP address, usually the system IP address.

The PEs communicate with each other at the control plane level by means of BGP updates containing BGP Layer 2 network layer reachability information (NLRI). Each update contains enough information for a PE to determine the presence of other local VPLS instances on peering PEs. In turn, this allows peer PE routers to set up pseudowire connectivity using LDP signaling for data flow between peers containing a local VPLS within the same VPLS instances.

Each update contains parameters usually associated with Multi-Protocol BGP updates:

- NLRI encoded as route-target (usually the VPLS-id) and PE system address.
- Next-Hop — The system IP address of the sending PE router.
- Extended communities — Contains the route target extended community and the VPLS-id as community values.

Each VPLS instance is configured with import and export route target extended communities to create the required pseudowire topology by controlling the distribution of each NLRI.

The purpose of this section is to describe the provisioning of a VPLS instance across three PE routers. A full mesh of pseudowires interconnects the VSI of each PE within the VPLS instance. A single attachment circuit is also configured on each VSI.

Configuration

The first step is to configure an MP-iBGP session using the L2VPN address family between each of the PEs and the route reflector.

The configuration for PE-1 is:

```
configure
     router
        autonomous-system 65536
        bgp
            group "internal"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.6
                exit
            exit
            no shutdown
        exit
    exit
```

The configuration for the other PE nodes is identical. The IP addresses can be derived from Figure 286.

The configuration for route reflector RR-6 is:

```
configure
     router
     autonomous-system 65536
        bgp
            cluster 1.1.1.1
            group "rr-internal"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
    exit
```

On PE-1, verify that the BGP session with RR-6 is established with address family l2-vpn capability negotiated:

```
*A:PE-1# show router bgp neighbor 192.0.2.6

===============================================================================
```

```
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                 : 192.0.2.6
Description          : (Not Specified)
Group                : internal
-------------------------------------------------------------------------------
Peer AS              : 65536            Peer Port            : 51039
Peer Address         : 192.0.2.6
Local AS             : 65536            Local Port           : 179
Local Address        : 192.0.2.1
Peer Type            : Internal         Dynamic Peer         : No
State                : Established      Last State           : Established
Last Event           : recvKeepAlive
Last Error           : Cease (Connection Collision Resolution)
Local Family         : L2-VPN
Remote Family        : L2-VPN
Hold Time            : 90               Keep Alive           : 30
Min Hold Time        : 0
Active Hold Time     : 90               Active Keep Alive    : 30
Cluster Id           : None
Preference           : 170              Num of Update Flaps  : 0
Recd. Paths          : 6
IPv4 Recd. Prefixes  : 0                IPv4 Active Prefixes : 0
IPv4 Suppressed Pfxs : 0                VPN-IPv4 Suppr. Pfxs : 0
VPN-IPv4 Recd. Pfxs  : 0                VPN-IPv4 Active Pfxs : 0
Mc IPv4 Recd. Pfxs.  : 0                Mc IPv4 Active Pfxs. : 0
Mc IPv4 Suppr. Pfxs  : 0                IPv6 Suppressed Pfxs : 0
IPv6 Recd. Prefixes  : 0                IPv6 Active Prefixes : 0
VPN-IPv6 Recd. Pfxs  : 0                VPN-IPv6 Active Pfxs : 0
VPN-IPv6 Suppr. Pfxs : 0
Mc IPv6 Recd. Pfxs.  : 0                Mc IPv6 Active Pfxs. : 0
Mc IPv6 Suppr. Pfxs  : 0                L2-VPN Suppr. Pfxs   : 0
L2-VPN Recd. Pfxs    : 0                L2-VPN Active Pfxs   : 0
MVPN-IPv4 Suppr. Pfxs: 0                MVPN-IPv4 Recd. Pfxs : 0
MVPN-IPv4 Active Pfxs: 0                MDT-SAFI Suppr. Pfxs : 0
MDT-SAFI Recd. Pfxs  : 0                MDT-SAFI Active Pfxs : 0
Flow-IPv4 Suppr. Pfxs: 0                Flow-IPv4 Recd. Pfxs : 0
Flow-IPv4 Active Pfxs: 0                Rte-Tgt Suppr. Pfxs  : 0
Rte-Tgt Recd. Pfxs   : 0                Rte-Tgt Active Pfxs  : 0
Backup IPv4 Pfxs     : 0                Backup IPv6 Pfxs     : 0
Mc Vpn Ipv4 Recd. Pf*: 0                Mc Vpn Ipv4 Active P*: 0
Mc Vpn Ipv4 Suppr. P*: 0
Backup Vpn IPv4 Pfxs : 0                Backup Vpn IPv6 Pfxs : 0
Input Queue          : 0                Output Queue         : 0
i/p Messages         : 3                o/p Messages         : 3
i/p Octets           : 83               o/p Octets           : 83
i/p Updates          : 0                o/p Updates          : 0
MVPN-IPv6 Suppr. Pfxs: 0                MVPN-IPv6 Recd. Pfxs : 0
MVPN-IPv6 Active Pfxs: 0
Flow-IPv6 Suppr. Pfxs: 0                Flow-IPv6 Recd. Pfxs : 0
Flow-IPv6 Active Pfxs: 0
Evpn Suppr. Pfxs     : 0                Evpn Recd. Pfxs      : 0
Evpn Active Pfxs     : 0
MS-PW Suppr. Pfxs    : 0                MS-PW Recd. Pfxs     : 0
MS-PW Active Pfxs    : 0
TTL Security         : Disabled         Min TTL Value        : n/a
Graceful Restart     : Disabled         Stale Routes Time    : n/a
Restart Time         : n/a
```

```
Advertise Inactive   : Disabled         Peer Tracking        : Disabled
Advertise Label      : None
Auth key chain       : n/a
Disable Cap Nego     : Disabled         Bfd Enabled          : Disabled
Flowspec Validate    : Disabled         Default Route Tgt    : Disabled
Aigp Metric          : Disabled         Split Horizon        : Disabled
Damp Peer Oscillatio*: Disabled         Update Errors        : 0
GR Notification      : Disabled         Fault Tolerance      : Disabled
Rem Idle Hold Time   : 00h00m00s
Next-Hop Unchanged   : None
L2 VPN Cisco Interop : Disabled
Local Capability     : RtRefresh MPBGP 4byte ASN
Remote Capability    : RtRefresh MPBGP 4byte ASN
Local AddPath Capabi*: Disabled
Remote AddPath Capab*: Send - None
                     : Receive - None
Import Policy        : None Specified / Inherited
Export Policy        : None Specified / Inherited
Origin Validation    : N/A
EBGP Link Bandwidth  : n/a
IPv4 Rej. Pfxs       : 0                IPv6 Rej. Pfxs       : 0
VPN-IPv4 Rej. Pfxs   : 0                VPN-IPv6 Rej. Pfxs   : 0
Mc IPv4 Rej. Pfxs    : 0                Mc IPv6 Rej. Pfxs    : 0
MVPN-IPv4 Rej. Pfxs  : 0                MVPN-IPv6 Rej. Pfxs  : 0
Flow-IPv4 Rej. Pfxs  : 0                Flow-IPv6 Rej. Pfxs  : 0
L2-VPN Rej. Pfxs     : 0                MDT-SAFI Rej. Pfxs   : 0
Rte-Tgt Rej. Pfxs    : 0                MS-PW Rej. Pfxs      : 0
Mc Vpn Ipv4 Rej. Pfxs: 0                Evpn Rej. Pfxs       : 0
-------------------------------------------------------------------------------
Neighbors : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

On RR-6, show that BGP sessions with each PE are established, and have correctly negotiated the l2-vpn address family capability.

```
A:RR-6# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.6       AS:65536       Local AS:65536
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 7           Total Path Memory           : 1288
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
```

```
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0

===============================================================================
BGP Summary
===============================================================================
Neighbor
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
                65536       3    0 00h00m18s 0/0/0 (L2VPN)
                            3    0
192.0.2.2
                65536       3    0 00h00m18s 0/0/0 (L2VPN)
                            3    0
192.0.2.3
                65536       3    0 00h00m18s 0/0/0 (L2VPN)
                            3    0
-------------------------------------------------------------------------------
A:RR-6#
```

A full mesh of RSVP Label Switched Paths (LSPs) is configured between the PE routers. For reference, the MPLS interface configuration and LSPs for PE-1 to PE-2 and PE-3 is:

```
*A:PE-1# configure router mpls
            interface "int-PE-1-P-4"
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                no shutdown
            exit
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "loose"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
```

```
            primary "loose"
            exit
            no shutdown
        exit
        no shutdown
```

VPLS PE Configuration

Pseudowire-Templates

Pseudowire templates are used by BGP to dynamically instantiate service distribution point (SDP) bindings, for a given service they are used to signal the egress service de-multiplexer labels used by remote PEs to reach the local PE.

The template determines the signaling parameters of the pseudowire, control word presence, plus other usage characteristics such as split horizon groups, MAC-pinning, filters, etc.

The MPLS transport tunnel between PE routers can be signaled using either LDP or RSVP.

LDP based pseudowires can be automatically instantiated. RSVP based SDPs have to be pre-provisioned.

Pseudowire Templates for Auto-SDP Creation using LDP

In order to use an LDP transport tunnel for data flow between PEs, it is necessary for link layer LDP to be configured between all PEs/Ps so that a transport label for each PE's system interface address is available. Using this mechanism SDPs can be auto-instantiated with SDP IDs starting at 17407. Any subsequent SDPs created use SDP-ids decrementing from this value.

A pseudowire template is required which may contain a split-horizon group. Each SDP created with this template is contained within the configured split horizon group so that traffic cannot be forwarded between them.

```
*A:PE-1# configure service
        pw-template 1 create
            split-horizon-group "vpls-shg"
```

```
            exit
        exit
```

A pseudowire template can also be created that does not contain a split-horizon group. The split horizon group can then be specified when the pw-template is included within the service.

```
*A:PE-1# configure service
        pw-template 2 create
        exit
```

Pseudowire Templates for Provisioned SDPs using RSVP

To use an RSVP tunnel as transport between PEs, it is necessary to bind the RSVP LSPs to the SDPs between each PE.

SDP creation from PE-1 to PE-2:

```
*A:PE-1# configure service sdp 12 mpls create
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
```

To create an SDP within a service that uses the RSVP transport tunnel, a pseudowire template is required that has the **use-provisioned-sdp** parameter.

```
A:PE-1# configure service
        pw-template 3 use-provisioned-sdp create
            exit
        exit
```

VPLS BGP-AD using Auto-Provisioned SDPs

Figure 287 **VPLS Instance with Auto-Provisioned SDPs**

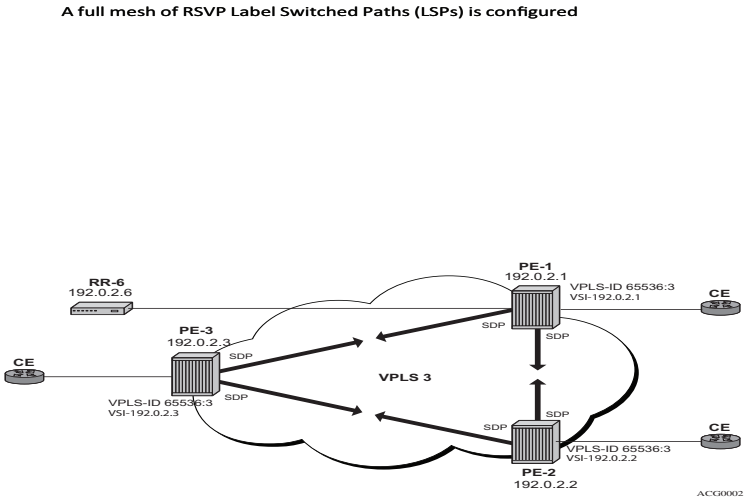


Figure 287 shows a schematic of a VPLS instance where the SDPs are auto-provisioned. SDPs are instantiated by a PE router using LDP signaling upon receipt of BGP Auto-discovery (BGP-AD) updates from peer PE routers.

PE-1 Configuration:

The following output shows the configuration required for a VPLS service using a pseudowire template configured for auto-provisioning of SDPs.

```
On PE-1:
configure
    service
        vpls 3 customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 2 split-horizon-group "vpls-shg"
                                      import-rt "target:65536:3"
                exit
            exit
            bgp-ad
                vpls-id 65536:3
                vsi-id
                    prefix 192.0.2.1
                exit
                no shutdown
            exit
            sap 1/1/4:3.0 create
            exit
            no shutdown
        exit
```

Within the **bgp** context, the pseudowire template is referenced which can be linked to a split-horizon-group and an import route-target, if required.

Within the **bgp-ad** context, the signaling parameters are configured. These are two parameters used by each PE to determine the presence of a given VPLS instance on a PE router. In turn, these are translated into endpoint identifiers for LDP signaling of pseudowires. As previously discussed, these parameters are:

- VPLS-id - a unique identifier of the VPLS instance. Each PE that is a member of a VPLS must share the same VPLS-id. This is inserted as an extended community value in the format AS:n. In this case, the VPLS-id for VPLS 3 is 65536:3. This is a mandatory parameter and if it is not configured, it is not possible to enable bgp-ad using no shutdown.
- Virtual Switching Instance (VSI) prefix — This identifies a specific instance of the VPLS. This must be unique within the VPLS instance, and is encoded using the 4 byte dotted decimal notation. Generally the system address is used as the VSI prefix. If this parameter is not configured, then the system address is used automatically.

The VPLS-id and VSI prefix for VPLS 3 on each PE is shown in Figure 287.

The VPLS-id and VSI prefix are concatenated to form a unique VSI-id. In this case, PE-1 has a VSI-id of 65536:3:192.0.2.1. This uniquely identifies the VPLS instance on each individual PE and is advertised as an L2 VPN BGP update.

A BGP-AD update is transmitted to all other PEs via the Route Reflector as follows:

```
*A:PE-1# show router bgp routes l2-vpn rd 65536:3 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : AutoDiscovery
Route Dist.    : 65536:3
Prefix         : 192.0.2.1
Nexthop        : 192.0.2.1
To             : 192.0.2.6
Res. Nexthop   : n/a
```

```
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65536:3 l2-vpn/vrf-imp:65536:3
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.6
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
A:PE-1#
```

The BGP update shown above is transmitted by PE-1 and has route type auto discovery.

In this L2 VPN update, the VPLS-id is encoded as the L2VPN extended community 65536:3.

The VSI is seen as the prefix 192.0.2.1. The combination of the VPLS-id and the VSI forms the VSI-id and uniquely identifies the VPLS instance within this PE router.

The nexthop is also encoded as the local system IP address 192.0.2.1, which allows remote PEs to identify a suitable transport tunnel to PE-1 and for the targeted-LDP peer for instantiating the SDP.

As can be seen within the update, the VPLS-id 65536:3 is also used to determine the route target extended community and the route distinguisher.

PE-2 Configuration

On PE-2 create a VPLS Service using pseudowire template 1, with VPLS-id 65536:3 and VSI-id prefix 192.0.2.2 (system IP address).

```
On PE-2:
configure
    service
        vpls 3 customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 2 split-horizon-group "vpls-shg"
                                      import-rt "target:65536:3"
                exit
            exit
            bgp-ad
                vpls-id 65536:3
```

```
                vsi-id
                    prefix 192.0.2.2
                exit
                no shutdown
            exit
            sap 1/1/4:3.0 create
            exit
            no shutdown
        exit
```

PE-3 Configuration

Create a VPLS Instance on PE-3: VPLS-id 65536:3 is the same as that of PE-1 and PE-2, with VSI-id of 192.0.2.3 (system IP address).

```
On PE-3:
configure
    service
        vpls 3 customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 2 split-horizon-group "vpls-shg"
                                      import-rt "target:65536:3"
                exit
            exit
            bgp-ad
                vpls-id 65536:3
                vsi-id
                    prefix 192.0.2.3
                exit
                no shutdown
            exit
            sap 1/1/4:3.0 create
            exit
            no shutdown
        exit
```

PE-1 Service Operation Verification

Verify that the service is operationally up on PE-1.

```
*A:PE-1# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/31/2016 11:34:59
Last Mgmt Change  : 05/31/2016 11:35:45
Etree Mode        : Disabled
```

```
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 3
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

------------------------------------------------------------------------------
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:1/1/4:3.0                            qinq         1522    1522    Up   Up
sdp:17406:4294967294 SB(192.0.2.2)       BgpAd        0       1556    Up   Up
sdp:17407:4294967295 SB(192.0.2.3)       BgpAd        0       1556    Up   Up
==============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

As seen from the output, the service is operationally up, with the SAPs and SDPs also up. The **SB** flag indicates that the SDP is of type spoke BGP.

BGP is used to discover the VPLS endpoints and exchange network reachability information. LDP is used to signal the pseudowires between the PEs.

LDP signaling occurs when each PE has discovered the endpoints of the VPLS instance. This compares with the use of the provisioned virtual-circuit IDs used in an NMS provisioned VPLS instances as per RFC 4762.

Verification of the ability of PE-1 to reach the other PE routers with VSIs within the VPLS instance can be seen from the Layer 2 routing table as follows:

```
*A:PE-1# show service l2-route-table bgp-ad
==============================================================================
Services: L2 Route Information - Summary
==============================================================================
Svc Id     L2-Routes (RD-Prefix)                Next Hop        Origin
               Sdp Bind Id                          PW Temp Id
------------------------------------------------------------------------------
3          *65536:3-192.0.2.2                   192.0.2.2       BGP-L2
               17406:4294967294                     2
3          *65536:3-192.0.2.3                   192.0.2.3       BGP-L2
               17407:4294967295                     2
------------------------------------------------------------------------------
No. of L2 Route Entries: 2
==============================================================================
```

```
*A:PE-1#
```

This output shows the presence of the signaled pseudowire SDPs. SDPs from PE-1 to PE-2 and PE-3 are signaled using LDP Forwarding Equivalence Class (FEC) Element 129.

Each PE router uses targeted LDP to signal the local and remote endpoints. If there is an endpoint match, then SDPs are instantiated. This compares with the use of LDP for NMS provisioned SDPs, which uses virtual-circuit IDs to signal pseudowires using LDP FEC Element 128.

In order to signal the SDPs, the following parameters are required:

1. Attachment Group Identifier (AGI): this is used to carry the VPLS-id of the local PE router VPLS instance. The VPLS-id must be identical for all PEs in the same VPLS instance.
2. Source Attachment Individual Identifier (SAII) and Target Attachment Individual Identifier (TAII): These use AII type 1 (RFC 4446) and are used to carry the NLRI (VSI-id minus the RD) of the remote PE router VPLS instance.

The AGI for each PE must be identical. SAII and TAII must be different.

The following shows the service LDP bindings for VPLS 3 on PE-1:

```
*A:PE-1# show router ldp bindings services service-id 3
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, WP - Label Withdraw Pending, C - Cpipe Service
        BU - Alternate For Fast Re-Route, TLV - (Type, Length: Value)
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Service FEC 128 Bindings
===============================================================================
Type                                        VCId       SDPId     IngLbl  LMTU
Peer                                        SvcId                EgrLbl  RMTU
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                          AGII        IngLbl    LMTU
TAII                                          Type        EgrLbl    RMTU
Peer                                          SvcId       SDPId
-------------------------------------------------------------------------------
192.0.2.1                                     null        262135U   1500
```

```
192.0.2.2                                          V-Eth       262133S   1500
192.0.2.2:0                                        3           17406
192.0.2.1                                          null        262136U   1500
192.0.2.3                                          V-Eth       262136S   1500
192.0.2.3:0                                        3           17407
-------------------------------------------------------------------------------
No. of FEC 129s: 2
===============================================================================
*A:PE-1#
```

This shows the two T-LDP bindings for PE-1 toward PE-2 and PE-3 for VPLS 3. The label bindings from this LDP output is identical to the SDP bindings output that follows. The following command can be used to list the SDP IDs and the SDP label bindings:

```
*A:PE-1# show service id 3 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpAd    192.0.2.2       Up      Up        262135    262133
17407:4294967295 BgpAd    192.0.2.3       Up      Up        262136    262136
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The SDP ID for the auto-provisioned SDP toward PE-2 is 17406, the SDP ID toward PE-3 is 17407.

Then the actual AGI, SAII and TAII values are seen in the detailed SDP output.

- AGI — 65536:3
- SAII — Local system IP address 192.0.2.1
- TAII — Remote system IP address 192.0.2.2 or 192.0.2.3

```
*A:PE-1# show service id 3 sdp 17407:4294967295 detail
===============================================================================
Service Destination Point (Sdp Id : 17407:4294967295) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 17407:4294967295  -(192.0.2.3)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id            : 17407:4294967295         Type              : BgpAd
PW-Template Id    : 2
AGI               : 65536:3                  SDP Bind Source   : bgp-l2vpn
Local AII         : 192.0.2.1
Remote AII        : 192.0.2.3
Split Horiz Grp   : vpls-shg
```

```
Etree Root Leaf Tag: Disabled                Etree Leaf AC     : Disabled
VC Type            : Ether                   VC Tag            : n/a
Admin Path MTU     : 0                       Oper Path MTU     : 1556
Delivery           : MPLS
Far End            : 192.0.2.3
---snip---
```

PE-2 Service Operation Verification

For completeness, verify the service is operationally up on PE-2.

```
*A:PE-2# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                     Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                     Creation Origin   : manual
Last Status Change: 05/31/2016 11:34:13
Last Mgmt Change  : 05/31/2016 11:34:59
Etree Mode        : Disabled
Admin State       : Up                    Oper State        : Up
MTU               : 1514                  Def. Mesh VC Id   : 3
SAP Count         : 1                     SDP Bind Count    : 2
Snd Flush on Fail : Disabled              Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled              Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled              Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled              Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:3.0                             qinq         1522    1522    Up   Up
sdp:17406:4294967294 SB(192.0.2.1)        BgpAd        0       1556    Up   Up
sdp:17407:4294967295 SB(192.0.2.3)        BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-2#

*A:PE-2# show service l2-route-table bgp-ad
===============================================================================
Services: L2 Route Information - Summary
===============================================================================
Svc Id     L2-Routes (RD-Prefix)                Next Hop        Origin
               Sdp Bind Id                           PW Temp Id
```

```
-------------------------------------------------------------------------
3          *65536:3-192.0.2.1                    192.0.2.1       BGP-L2
               17406:4294967294                      2
3          *65536:3-192.0.2.3                    192.0.2.3       BGP-L2
               17407:4294967295                      2
-------------------------------------------------------------------------
No. of L2 Route Entries: 2
=========================================================================
*A:PE-2#

*A:PE-2# show router ldp bindings services service-id 3
=========================================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
=========================================================================================
Legend: U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, WP - Label Withdraw Pending, C - Cpipe Service
        BU - Alternate For Fast Re-Route, TLV - (Type, Length: Value)
        LF - Lower FEC, UF - Upper FEC
=========================================================================================
LDP Service FEC 128 Bindings
=========================================================================================
Type                                          VCId       SDPId     IngLbl  LMTU
Peer                                          SvcId                EgrLbl  RMTU
-----------------------------------------------------------------------------------------
No Matching Entries Found
=========================================================================================
=========================================================================================
LDP Service FEC 129 Bindings
=========================================================================================
SAII                                            AGII        IngLbl    LMTU
TAII                                            Type        EgrLbl    RMTU
Peer                                            SvcId       SDPId
-----------------------------------------------------------------------------------------
192.0.2.2                                       null        262133U   1500
192.0.2.1                                       V-Eth       262135S   1500
192.0.2.1:0                                     3           17406
192.0.2.2                                       null        262134U   1500
192.0.2.3                                       V-Eth       262135S   1500
192.0.2.3:0                                     3           17407
-----------------------------------------------------------------------------------------
No. of FEC 129s: 2
=========================================================================================
*A:PE-2#

*A:PE-2# show service id 3 sdp
=========================================================================================
Services: Service Destination Points
=========================================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-----------------------------------------------------------------------------------------
17406:4294967294 BgpAd    192.0.2.1       Up      Up        262133    262135
17407:4294967295 BgpAd    192.0.2.3       Up      Up        262134    262135
-----------------------------------------------------------------------------------------
Number of SDPs : 2
-----------------------------------------------------------------------------------------
```

```
===============================================================================
*A:PE-2#
```

PE-3 Service Operation Verification

Verify service is operationally up on PE-3.

```
*A:PE-3# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/31/2016 11:34:40
Last Mgmt Change  : 05/31/2016 11:34:58
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 3
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:3.0                           qinq         1522    1522    Up   Up
sdp:17406:4294967294 SB(192.0.2.2)      BgpAd        0       1556    Up   Up
sdp:17407:4294967295 SB(192.0.2.1)      BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#


*A:PE-3# show service l2-route-table bgp-ad
===============================================================================
Services: L2 Route Information - Summary
===============================================================================
Svc Id     L2-Routes (RD-Prefix)                Next Hop        Origin
               Sdp Bind Id                          PW Temp Id
-------------------------------------------------------------------------------
3          *65536:3-192.0.2.1                   192.0.2.1       BGP-L2
               17407:4294967295                     2
```

```
3          *65536:3-192.0.2.2                    192.0.2.2       BGP-L2
               17406:4294967294                      2
-------------------------------------------------------------------------------
No. of L2 Route Entries: 2
===============================================================================
*A:PE-3#

*A:PE-3# show router ldp bindings services service-id 3
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, WP - Label Withdraw Pending, C - Cpipe Service
        BU - Alternate For Fast Re-Route, TLV - (Type, Length: Value)
        LF - Lower FEC, UF - Upper FEC
===============================================================================
LDP Service FEC 128 Bindings
===============================================================================
Type                                          VCId       SDPId     IngLbl  LMTU
Peer                                          SvcId                EgrLbl  RMTU
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                            AGII        IngLbl    LMTU
TAII                                            Type        EgrLbl    RMTU
Peer                                            SvcId       SDPId
-------------------------------------------------------------------------------
192.0.2.3                                       null        262136U   1500
192.0.2.1                                       V-Eth       262136S   1500
192.0.2.1:0                                     3           17407
192.0.2.3                                       null        262135U   1500
192.0.2.2                                       V-Eth       262134S   1500
192.0.2.2:0                                     3           17406
-------------------------------------------------------------------------------
No. of FEC 129s: 2
===============================================================================
*A:PE-3#

*A:PE-3# show service id 3 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17406:4294967294 BgpAd    192.0.2.2       Up      Up        262135    262134
17407:4294967295 BgpAd    192.0.2.1       Up      Up        262136    262136
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

BGP AD using Pre-Provisioned SDPs

It is possible to configure BGP-AD instances that use RSVP transport tunnels. In this case, the LSPs and SDPs must be manually created.

Figure 288 **VPLS Instance using Pre-Provisioned SDPs**

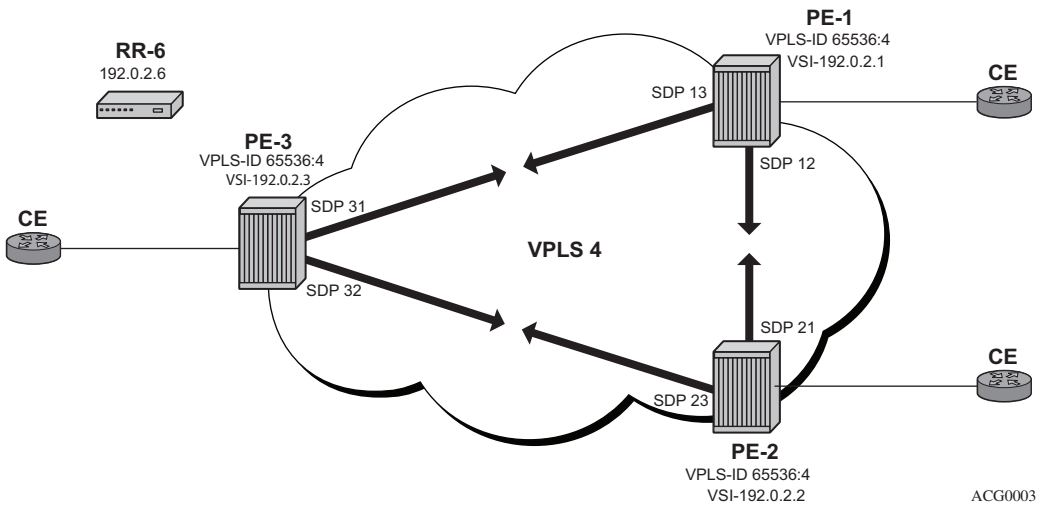


Figure 288 shows a VPLS instance configured across three Provider Edge routers as before.

The SDP configurations for the three PEs are shown below:

SDPs on PE-1

```
configure service
        sdp 12 mpls create
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
        sdp 13 mpls create
            far-end 192.0.2.3
            lsp "LSP-PE-1-PE-3"
            no shutdown
        exit
    exit
```

SDPs on PE-2

```
configure service
       sdp 21 mpls create
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
        sdp 23 mpls create
            far-end 192.0.2.3
            lsp "LSP-PE-2-PE-3"
            no shutdown
        exit
    exit
```

SDPs on PE-3

```
configure service
        sdp 31 mpls create
            far-end 192.0.2.1
            lsp "LSP-PE-3-PE-1"
            no shutdown
        exit
        sdp 32 mpls create
            far-end 192.0.2.2
            lsp "LSP-PE-3-PE-2"
            no shutdown
        exit
    exit
```

The pw-template that is to be used within each VPLS instance must be provisioned on all PEs and must use the keyword **use-provisioned-sdp**. The pw-template looks like:

```
A:PE-1# configure service
        pw-template 3 use-provisioned-sdp create
            exit
        exit
```

This configuration must be repeated on both PE-2 and PE-3.

The following output shows the configuration required for a VPLS service using a pseudowire template configured for pre-provisioned RSVP SDPs.

On PE-1:

```
configure
    service
        vpls 4 customer 1 create
            bgp
                route-distinguisher 65536:4
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3 split-horizon-group "vpls-shg"
                                      import-rt "target:65536:4"
                exit
            exit
```

```
            bgp-ad
                vpls-id 65536:4
                vsi-id
                    prefix 192.0.2.1
                exit
                no shutdown
            exit
            sap 1/1/4:4.0 create
            exit
            no shutdown
        exit
```

Similarly, on PE-2 the configuration is shown below:

On PE-2:

```
configure
    service
        vpls 4 customer 1 create
            bgp
                route-distinguisher 65536:4
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3 split-horizon-group "vpls-shg"
                                      import-rt "target:65536:4"
                exit
            exit
            bgp-ad
                vpls-id 65536:4
                vsi-id
                    prefix 192.0.2.2
                exit
                no shutdown
            exit
            sap 1/1/4:4.0 create
            exit
            no shutdown
        exit
```

On PE-3:

```
configure
    service
        vpls 4 customer 1 create
            bgp
                route-distinguisher 65536:4
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3 split-horizon-group "vpls-shg"
                                      import-rt "target:65536:4"
                exit
            exit
            bgp-ad
                vpls-id 65536:4
                vsi-id
                    prefix 192.0.2.4
                exit
```

```
            no shutdown
        exit
        sap 1/1/4:4.0 create
        exit
        no shutdown
    exit
```

Verify that the service is operationally up on PE-1.

```
*A:PE-1# show service id 4 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 4                   Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/31/2016 11:38:03
Last Mgmt Change  : 05/31/2016 11:38:46
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 4
SAP Count         : 1                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:4.0                           qinq         1522    1522    Up   Up
sdp:12:4294967292 S(192.0.2.2)          BgpAd        0       1556    Up   Up
sdp:13:4294967293 S(192.0.2.3)          BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

The SDP identifiers are the pre-provisioned SDPs, i.e. SDP 12 and 13.

For completeness, verify the service is operationally up on PE-2.

```
*A:PE-2# show service id 4 base
===============================================================================
```

```
Service Basic Information
================================================================================
Service Id        : 4                    Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                    Creation Origin   : manual
Last Status Change: 05/31/2016 11:37:19
Last Mgmt Change  : 05/31/2016 11:37:58
Etree Mode        : Disabled
Admin State       : Up                   Oper State        : Up
MTU               : 1514                 Def. Mesh VC Id   : 4
SAP Count         : 1                    SDP Bind Count    : 2
Snd Flush on Fail : Disabled             Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled             Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled             Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled             Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

--------------------------------------------------------------------------------
Service Access & Destination Points
--------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
--------------------------------------------------------------------------------
sap:1/1/4:4.0                             qinq         1522    1522    Up   Up
sdp:21:4294967292 S(192.0.2.1)            BgpAd        0       1556    Up   Up
sdp:23:4294967293 S(192.0.2.3)            BgpAd        0       1556    Up   Up
================================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-2#
```

Verify service is operational on PE-3.

```
*A:PE-3# show service id 4 base
================================================================================
Service Basic Information
================================================================================
Service Id        : 4                    Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                    Creation Origin   : manual
Last Status Change: 05/31/2016 11:37:45
Last Mgmt Change  : 05/31/2016 11:38:23
Etree Mode        : Disabled
Admin State       : Up                   Oper State        : Up
MTU               : 1514                 Def. Mesh VC Id   : 4
SAP Count         : 1                    SDP Bind Count    : 2
Snd Flush on Fail : Disabled             Host Conn Verify  : Disabled
SHCV pol IPv4     : None
```

```
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:4.0                            qinq         1522    1522    Up   Up
sdp:31:4294967292 S(192.0.2.1)           BgpAd        0       1556    Up   Up
sdp:32:4294967293 S(192.0.2.2)           BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

Conclusion

BGP-auto-discovery coupled with LDP pseudowire signaling allows the delivery of L2 VPN services to customers where BGP is commonly used. This example shows the configuration of BGP-Auto discovery together with the associated show outputs which can be used for verification and troubleshooting.

Multi-Chassis Endpoint for VPLS Active/ Standby Pseudowire

This chapter provides information about multi-chassis endpoint for VPLS active/ standby pseudowire.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SROS release 7.0.R6, but the CLI in this edition is based on release 14.0.R2.

Multi-chassis endpoint peers must be the same chassis type.

Overview

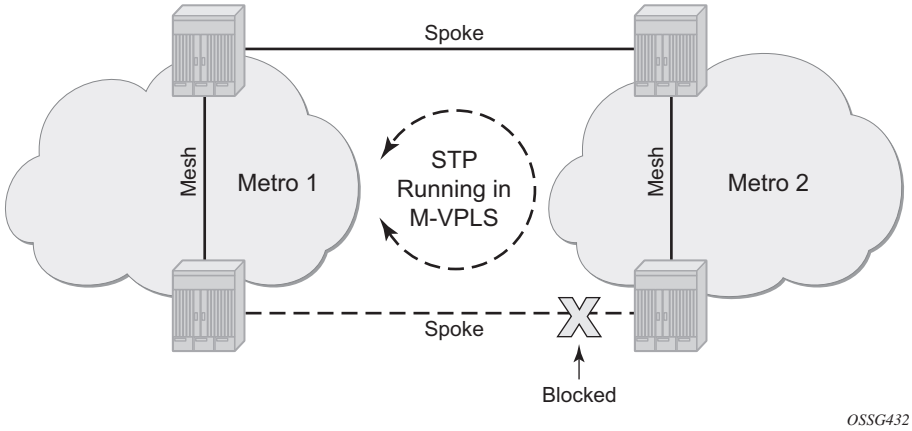
When implementing a large VPLS, one of the limiting factors is the number of T-LDP sessions required for the full mesh of SDPs. Mesh SDPs are required between all PEs participating in the VPLS with a full mesh of T-LDP sessions.

This solution is not scalable, as the number of sessions grows more rapidly than the number of participating PEs. Several options exist to reduce the number of T-LDP sessions required in a large VPLS.

The first option is hierarchical VPLS (H-VPLS) with spoke SDPs. By using spoke SDPs between two clouds of fully meshed PEs, any-to-any T-LDP sessions for all participating PEs are not required.

However, if spoke SDP redundancy is required, STP must be used to avoid a loop in the VPLS. Management VPLS can be used to reduce the number of STP instances and separate customer and STP traffic (Figure 289).

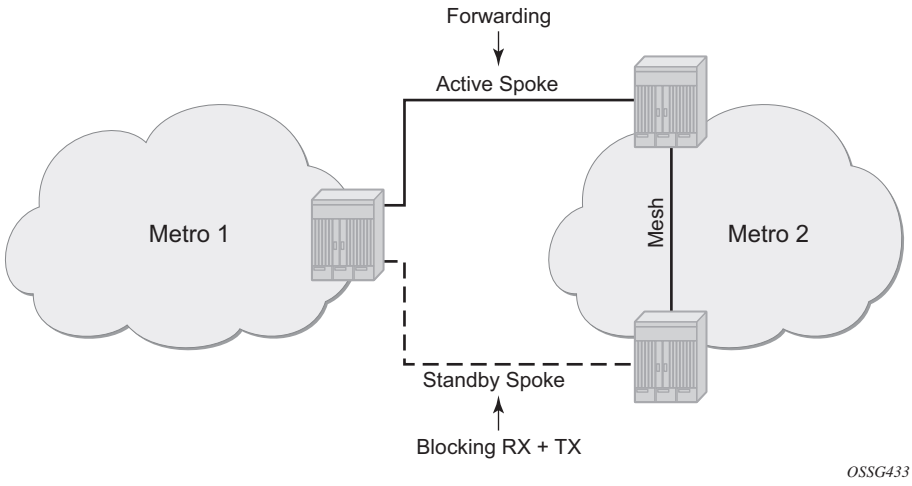
Figure 289 **H-VPLS with STP**



VPLS pseudowire redundancy provides H-VPLS redundant spoke connectivity. The active spoke is in forwarding state, while the standby spoke is in blocking state. Therefore, STP is not needed anymore to break the loop, as illustrated in Figure 290.

However, the PE implementing the active and standby spokes represents a single point of failure in the network.

Figure 290 **VPLS Pseudowire Redundancy**

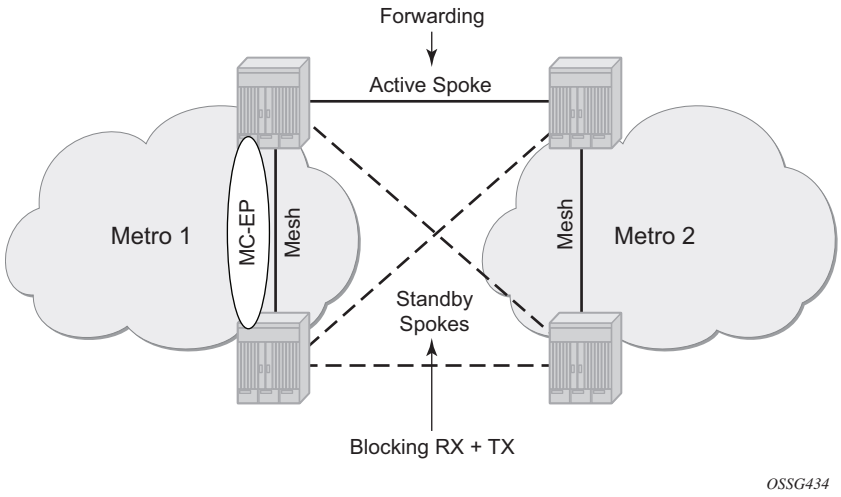


Multi-chassis endpoint (MC-EP) for VPLS active/standby pseudowire expands on the VPLS pseudowire redundancy and allows the removal of the single point of failure.

There is only one spoke in forwarding state, all standby spokes are in blocking state. Mesh and square resiliency are supported.

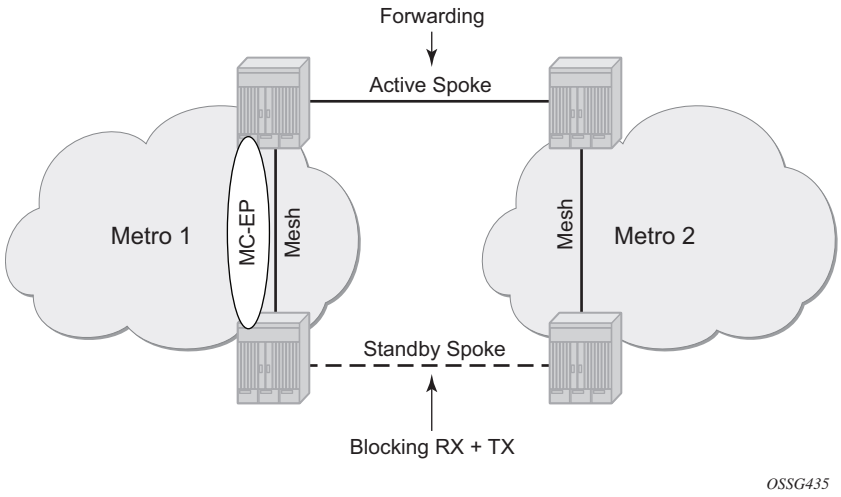
Mesh resiliency can protect against simultaneous node failure in the core and in the MC-EP (double failure), but requires more SDPs (and thus more T-LDP sessions). Mesh resiliency is illustrated in Figure 291.

Figure 291 **Multi-Chassis Endpoint with Mesh Resiliency**



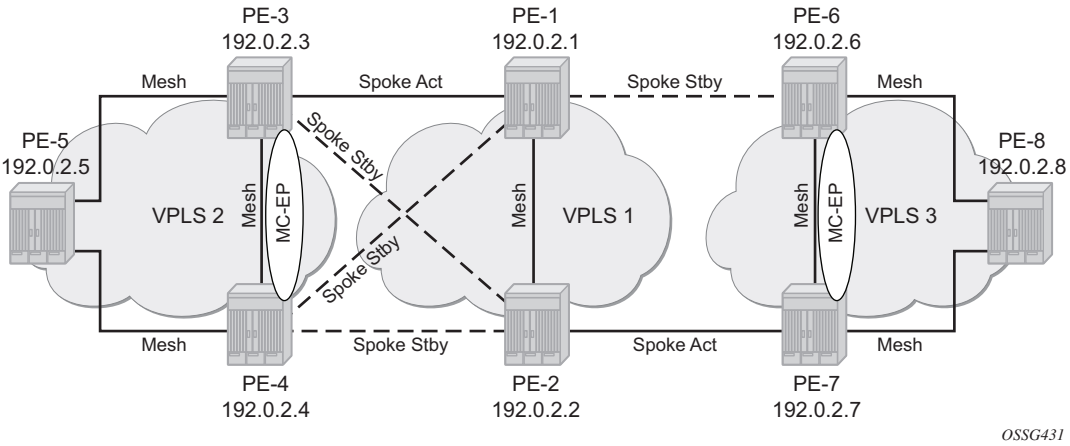
Square resiliency provides single failure node protection, and requires less SDPs (and thus less T-LDP sessions). Square resiliency is illustrated in Figure 292.

Figure 292 **Multi-Chassis Endpoint with Square Resiliency**



Example Topology

Figure 293 **Example Topology**



The network topology is displayed in Figure 293.

The setup consists of:

- Two core nodes (PE-1 and PE-2), and three nodes for each metro area (PE-3, PE-4, PE-5 and PE-6, PE-7, PE-8, respectively).
- VPLS 1 is the core VPLS, used to interconnect the two metro areas represented by VPLS 2 and VPLS 3.
- VPLS 2 will be connected to the core VPLS in mesh resiliency.
- VPLS 3 will be connected to the core VPLS in square resiliency.

Three separate VPLS identifiers are used for clarity, however, the same identifier could be used for each. For interoperation, only the same VC-ID is required to be used on both ends of the spoke SDPs.

The following configuration tasks should be done first:

- IS-IS or OSPF throughout the network.
- RSVP or LDP-signaled LSPs over the paths used for mesh/spoke SDPs.

Configuration

SDP Configuration

On each PE, SDPs are created to match the topology described in Figure 293.

The convention for the SDP naming is: XY where X is the originating node and Y the target node.

An example of the SDP configuration in PE-3 (using LDP):

```
A:PE-3# configure service
        sdp 31 mpls create
            far-end 192.0.2.1
            ldp
            no shutdown
        exit
        sdp 32 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 34 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        sdp 35 mpls create
            far-end 192.0.2.5
            ldp
            no shutdown
        exit
```

Verification of the SDPs on PE-3:

```
*A:PE-3# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
31     0       1556    192.0.2.1        Up   Up          MPLS    L     TLDP
32     0       1556    192.0.2.2        Up   Up          MPLS    L     TLDP
34     0       1556    192.0.2.4        Up   Up          MPLS    L     TLDP
35     0       1556    192.0.2.5        Up   Up          MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 4
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE
```

```
===============================================================================
*A:PE-3#
```

Full Mesh VPLS Configuration

Next, three fully meshed VPLS services are configured.

- VPLS 1 is the core VPLS, on PE-1 and PE-2
- VPLS 2 is the metro 1 VPLS, on PE-3, PE-4 and PE-5
- VPLS 3 is the metro 2 VPLS, on PE-6, PE-7 and PE-8

On PE-1 (similar configuration on PE-2):

```
configure
    service
        vpls 1 customer 1 create
            description "core VPLS"
            mesh-sdp 12:1 create
            exit
            no shutdown
        exit
```

On PE-3 (similar configuration on PE-4 and PE-5):

```
configure
    service
        vpls 2 customer 1 create
            description "Metro 1 VPLS"
            mesh-sdp 34:2 create
            exit
            mesh-sdp 35:2 create
            exit
            no shutdown
        exit
```

On PE-6 (similar configuration on PE-7 and PE-8):

```
configure
    service
        vpls 3 customer 1 create
            description "Metro 2 VPLS"
            mesh-sdp 67:3 create
            exit
            mesh-sdp 68:3 create
            exit
            no shutdown
        exit
```

Verification of the VPLS:

- The service must be operationally up.
- All mesh SDPs must be up in the VPLS service.

On PE-6 (similar on other nodes):

```
*A:PE-6# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/20/2016 08:08:29
Last Mgmt Change  : 05/20/2016 08:08:24
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 3
SAP Count         : 0                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:67:3 M(192.0.2.7)                    Mesh         0       1556    Up   Up
sdp:68:3 M(192.0.2.8)                    Mesh         0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-6#
```

Multi-Chassis Configuration

Multi-chassis will be configured on the MC peers PE-3, PE-4 and PE-6, PE-7. The peer system address is configured, and mc-endpoint will be enabled.

On PE-3 (similar configuration on PE-4, PE-6 and PE-7):

```
configure
    redundancy
        multi-chassis
            peer 192.0.2.4 create
                mc-endpoint
                    no shutdown
                exit
                no shutdown
            exit
```

Verification of the multi-chassis synchronization:

If the multi-chassis synchronization fails, both nodes will fall back to single-chassis mode. In that case, two spoke SDPs could become active at the same time. It is important to verify the multi-chassis synchronization before enabling the redundant spoke SDPs.

```
*A:PE-3# show redundancy multi-chassis mc-endpoint peer 192.0.2.4
===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.4           Peer Name            :
Admin State     : up                  Oper State           : up
Last State chg  :                     Source Addr          :
System Id       : 4a:c6:ff:00:00:00   Sys Priority         : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail     : 3
Passive Mode    : disabled            Psv Mode Oper        : No
Boot Timer      : 300                 BFD                  : disabled
Last update     : 05/20/2016 08:08:44 MC-EP Count          : 0
===============================================================================
*A:PE-3#
```

Mesh Resiliency Configuration

PE-3 and PE-4 will be connected to the core VPLS in mesh resiliency.

- First an endpoint is configured.
- The “no suppress-standby-signaling” is needed to block the standby spoke SDP.
- The multi-chassis endpoint peer is configured. The mc-endpoint ID must match between the two peers.

On PE-3 (similar on PE-4):

```
configure service
        vpls 2
            endpoint "CORE" create
            no suppress-standby-signaling
            mc-endpoint 1
                mc-ep-peer 192.0.2.4
```

```
            exit
        exit
```

After this configuration, the MP-EP Count in the preceding show command changes to 1, as follows:

```
*A:PE-3# show redundancy multi-chassis mc-endpoint peer 192.0.2.4
===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.4           Peer Name           :
Admin State     : up                  Oper State          : up
Last State chg  :                     Source Addr         :
System Id       : 4a:c6:ff:00:00:00   Sys Priority        : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail    : 3
Passive Mode    : disabled            Psv Mode Oper       : No
Boot Timer      : 300                 BFD                 : disabled
Last update     : 05/20/2016 08:10:07 MC-EP Count         : 1
===============================================================================
*A:PE-3#
```

Two spoke SDPs are configured on each peer of the multi-chassis to the two nodes of the core VPLS (mesh resiliency). Each spoke SDP refers to the endpoint CORE.

The precedence is defined on the spoke SDPs as follows:

- Spoke SDP 31 on PE-3 will be active. It is configured as primary (= precedence 0).
- Spoke SDP 32 on PE-3 will be the first backup. It is configured with precedence 1.
- Spoke SDP 41 on PE-4 will be the second backup. It is configured with precedence 2.
- Spoke SDP 42 on PE-4 will be the third backup. It is configured with precedence 3.

On PE-3:

```
configure service
        vpls 2
            spoke-sdp 31:1 endpoint "CORE" create
                precedence primary
            exit
            spoke-sdp 32:1 endpoint "CORE" create
                precedence 1
            exit
```

On PE-4:

```
configure service
        vpls 2
```

```
            spoke-sdp 41:1 endpoint "CORE" create
                precedence 2
            exit
            spoke-sdp 42:1 endpoint "CORE" create
                precedence 3
            exit
```

Verification of the spoke SDPs:

On PE-3 and PE-4, the spoke SDPs must be up.

```
*A:PE-3# show service id 2 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type Far End addr    Adm     Opr       I.Lbl       E.Lbl
-------------------------------------------------------------------------------
31:1             Spok 192.0.2.1       Up      Up        262135      262131
32:1             Spok 192.0.2.2       Up      Up        262134      262131
34:2             Mesh 192.0.2.4       Up      Up        262133      262135
35:2             Mesh 192.0.2.5       Up      Up        262132      262135
-------------------------------------------------------------------------------
Number of SDPs : 4
-------------------------------------------------------------------------------
===============================================================================
*A:PE-3#
```

The endpoints on PE-3 and PE-4 can be verified. One spoke SDP is in Tx-Active mode (31 on PE-1 because it is configured as primary).

```
*A:PE-3# show service id 2 endpoint "CORE" | match "Tx Active"
Tx Active (SDP)              : 31:1
Tx Active Up Time            : 0d 01:16:04
Tx Active Change Count       : 1
Last Tx Active Change        : 05/20/2016 08:10:41
*A:PE-3#
```

There is no active spoke SDP on PE-4.

```
*A:PE-4# show service id 2 endpoint "CORE" | match "Tx Active"
Tx Active                    : none
Tx Active Up Time            : 0d 00:00:00
Tx Active Change Count       : 0
Last Tx Active Change        : 05/20/2016 07:59:47
*A:PE-4#
```

On PE-1 and PE-2, the spoke SDPs are operationally up.

```
*A:PE-1# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
```

```
SdpId            Type Far End addr    Adm     Opr       I.Lbl       E.Lbl
-------------------------------------------------------------------------------
12:1             Mesh 192.0.2.2       Up      Up        262135      262135
13:1             Spok 192.0.2.3       Up      Up        262131      262135
14:1             Spok 192.0.2.4       Up      Up        262130      262133
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

However, because pseudowire signaling has been enabled, only one spoke SDP will be active, the others are set in standby.

On PE-1, spoke SDP 13 is active (no pseudowire bit signaled from PE-3).

And the spoke SDP 14 is signaled in standby by PE-4.

```
*A:PE-1# show service id 1 sdp 13:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : None
*A:PE-1# show service id 1 sdp 14:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
*A:PE-1#
```

On PE-2, both spoke SDPs are signaled in standby.

```
*A:PE-2# configure port 1/1/1 no shutdown
*A:PE-2# show service id 1 sdp 23:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
*A:PE-2# show service id 1 sdp 24:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
*A:PE-2#
```

There is one active and three standby spoke SDPs.

Square Resiliency Configuration

PE-6 and PE-7 will be connected to the core VPLS in square resiliency.

- First an endpoint is configured.
- The “no suppress-standby-signaling” is needed to block the standby spoke SDP.
- The multi-chassis endpoint peer is configured. The mc-endpoint ID must match between the two peers.

On PE-7 (similar on PE-6):

One spoke SDP is configured on each peer of the multi-chassis to one node of the core VPLS (square resiliency). Each spoke SDP refers to the endpoint CORE.

```
configure service
        vpls 3
            endpoint "CORE" create
                no suppress-standby-signaling
                mc-endpoint 1
                    mc-ep-peer 192.0.2.6
                exit
            exit
        exit
    exit
```

The precedence will be defined on the spoke SDPs as follows:

- Spoke SDP 72:1 on PE-7 will be active. It is configured as primary (= precedence 0)
- Spoke SDP 61:1 on PE-6 will be the first backup with precedence 1.

On PE-7:

```
configure service
        vpls 3
            spoke-sdp 72:1 endpoint "CORE" create
                precedence primary
            exit
        exit
    exit
```

On PE-6:

```
configure service
        vpls 3
            spoke-sdp 61:1 endpoint "CORE" create
                precedence 1
            exit
        exit
    exit
```

Verification of the spoke SDPs.

```
*A:PE-7# show service id 3 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type Far End addr    Adm     Opr       I.Lbl       E.Lbl
-------------------------------------------------------------------------------
72:1             Spok 192.0.2.2       Up      Up        262133      262132
76:3             Mesh 192.0.2.6       Up      Up        262135      262135
78:3             Mesh 192.0.2.8       Up      Up        262128      262142
-------------------------------------------------------------------------------
Number of SDPs : 3
```

```
-------------------------------------------------------------------------------
===============================================================================
*A:PE-7#
```

On PE-6 and PE-7, the spoke SDPs must be up.

The endpoints on PE-7 and PE-6 can be verified. One spoke SDP is in Tx-Active mode (72 on PE-7 because it is configured as primary).

```
*A:PE-7# show service id 3 endpoint | match "Tx Active"
Tx Active (SDP)              : 72:1
Tx Active Up Time            : 0d 00:17:24
Tx Active Change Count       : 1
Last Tx Active Change        : 05/20/2016 08:13:18
*A:PE-7#
```

There are no active spoke SDP on PE-6.

```
*A:PE-6# show service id 3 endpoint | match "Tx Active"
Tx Active                    : none
Tx Active Up Time            : 0d 00:00:00
Tx Active Change Count       : 2
Last Tx Active Change        : 05/20/2016 08:13:18
*A:PE-6#
```

The output shows that on PE-1, spoke SDP 16 is signaled with peer in standby mode.

```
*A:PE-1# show service id 1 sdp 16:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
*A:PE-1#
```

On PE-2, the spoke SDP 27 is signaled with peer active (no pseudowire bits).

```
*A:PE-2# show service id 1 sdp 27:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : None
*A:PE-2#
```

There is one active and one standby spoke SDP.

Additional Parameters

Multi-Chassis

```
*A:PE-3# configure redundancy multi-chassis peer 192.0.2.4 mc-endpoint
  - mc-endpoint
  - no mc-endpoint

 [no] bfd-enable      - Configure BFD
 [no] boot-timer      - Configure boot timer interval
 [no] hold-on-neighb* - Configure hold time applied on neighbor failure
 [no] keep-alive-int* - Configure keep alive interval for this MC-Endpoint
 [no] passive-mode    - Configure passive-mode
 [no] shutdown        - Administratively enable/disable the multi-chassis
                        peer end-point
 [no] system-priority - Configure system priority
```

These parameters will be explained in the following sections.

Peer Failure Detection

The default mechanism is based on the keep-alive messages exchanged between the peers.

The keep-alive-interval is the interval at which keep-alive messages are sent to the MC peer. It is set in tenths of a second from 5 to 500), with a default value of 5.

Hold-on-neighbor-failure is the number of keep-alive-intervals that the node will wait for a packet from the peer before assuming it has failed. After this interval, the node will revert to single chassis behavior. It can be set from 2 to 25 with a default value of 3.

BFD Session

BFD is another peer failure detection mechanism. It can be used to speed up the convergence in case of peer loss.

```
*A:PE-3# configure redundancy
        multi-chassis
            peer 192.0.2.4
                mc-endpoint
                    bfd-enable
                exit
            exit
```

It is using the centralized BFD session. BFD must be enabled on the system interface.

```
*A:PE-3# configure router
        interface "system"
            address 192.0.2.3/32
            bfd 100 receive 100 multiplier 3
        exit
```

Verification of the BFD session:

```
*A:PE-3# show router bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id            State                 Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols             Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
system                        Up                    100       100       3
   192.0.2.4                  mcep                  152       84        central
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-3#
```

Boot Timer

The **boot-timer** command specifies the time after a reboot that the node will try to establish a connection with the MC peer before assuming a peer failure. In case of failure, the node will revert to single chassis behavior.

System Priority

The system priority influences the selection of the MC master. The lowest priority node will become the master.

In case of equal priorities, the lowest system-id (=chassis MAC address) will become the master.

VPLS Endpoint and Spoke SDP

Ignore Standby Pseudowire Bits

```
*A:PE-1# configure service vpls 1 spoke-sdp 14:1
---snip---
[no] ignore-standby* - Ignore 'standby-bit' received from LDP peer
---snip---
```

The peer pseudowire status bits are ignored and traffic is forwarded over the spoke SDP.

It can speed up convergence for multicast traffic in case of spoke SDP failure.

Traffic sent over the standby spoke SDP will be discarded by the peer.

In this topology, if the **ignore-standby-signaling** command is enabled on PE-1, it sends MC traffic to PE-3 and PE-4 (and to PE-6). If PE-3 fails, PE-4 can start forwarding traffic in the VPLS as soon as it detects PE-3 being down. There is no signaling needed between PE-1 and PE-4.

Block-on-Mesh-Failure

```
*A:PE-3# configure service vpls 2 endpoint "CORE"
---snip---
[no] block-on-mesh-* - Block traffic on mesh-SDP failure
---snip---
```

In case a PE loses all the mesh SDPs of a VPLS, it should block the spoke SDPs to the core VPLS, and inform the MC-EP peer that can activate one of its spoke SDPs.

If block-on-mesh-failure is enabled, the PE will signal all the pseudowires of the endpoint in standby.

In this topology, if PE3 does not have any valid mesh SDP to the VPLS 2 mesh, it will set the spoke SDPs under endpoint CORE in standby.

When block-on-mesh-failure is activated under an endpoint, it is automatically set under the spoke SDPs belonging to this endpoint.

```
*A:PE-3# configure service vpls 2
*A:PE-3>config>service>vpls# info
----------------------------------------------
            stp
                shutdown
            exit
            endpoint "CORE" create
```

```
                    no suppress-standby-signaling
                    mc-endpoint 1
                        mc-ep-peer 192.0.2.4
                    exit
                exit
                spoke-sdp 31:1 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    precedence primary
                    no shutdown
                exit
                spoke-sdp 32:1 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    precedence 1
                    no shutdown
                exit
                mesh-sdp 34:2 create
                    no shutdown
                exit
                mesh-sdp 35:2 create
                    no shutdown
                exit
                no shutdown
------------------------------------------------
*A:PE-3>config>service>vpls# endpoint "CORE" block-on-mesh-failure
*A:PE-3>config>service>vpls# info
------------------------------------------------
                stp
                    shutdown
                exit
                endpoint "CORE" create
                    no suppress-standby-signaling
                    block-on-mesh-failure
                    mc-endpoint 1
                        mc-ep-peer 192.0.2.4
                    exit
                exit
                spoke-sdp 31:1 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    block-on-mesh-failure
                    precedence primary
                    no shutdown
                exit
                spoke-sdp 32:1 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    block-on-mesh-failure
                    precedence 1
                    no shutdown
                exit
                mesh-sdp 34:2 create
                    no shutdown
                exit
```

```
            mesh-sdp 35:2 create
                no shutdown
            exit
            no shutdown
-----------------------------------------------
```

Precedence

```
*A:PE-3# configure service vpls 2 spoke-sdp 31:1
---snip---
 [no] precedence      - Configure the spoke-sdp precedence
---snip---
```

The precedence is used to indicate in which order the spoke SDPs should be used. The value is from 0 to 4 (0 being primary), the lowest having higher priority. The default value is 4.

Revert-Time

```
*A:PE-3# configure service vpls 2 endpoint "CORE"
---snip---
 [no] revert-time    - Configure the time to wait before reverting to primary spoke-sdp
---snip---
```

If the precedence is equal between the spoke SDPs, there is no revertive behavior. Changing the precedence of a spoke SDP will not trigger a revert. The default is **no revert**.

MAC-Flush Parameters

When a spoke SDP goes from standby to active (due to the active spoke SDP failure), the node will send a **flush-all-but-mine** message.

After a restoration of the spoke SDP, a new **flush-all-but-mine** message will be sent.

```
*A:PE-1# configure service vpls 1 propagate-mac-flush
```

A node configured with **propagate MAC flush** will forward the flush messages received on the spoke-SDP to its other mesh/spoke SDPs.

A node configured with **send flush on failure** will send a **flush-all-from-me** message when one of its SDPs goes down.

```
A:PE-1# configure service vpls 1 send-flush-on-failure
```

Failure Scenarios

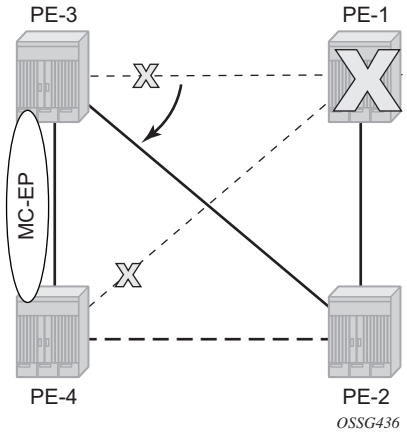
For the subsequent failure scenarios, the configuration of the nodes is as described in the Configuration.

Core Node Failure

When the core node PE-1 fails, the spoke SDPs from PE-3 and PE-4 go down.

Because the spoke SDP 31 between PE-3 and PE-4 was active, the MC master (PE-3 in this case) will select the next best spoke SDP, which will be 32 between PE-3 and PE-2 (precedence 1). See Figure 294.

Figure 294 **Core Node Failure**



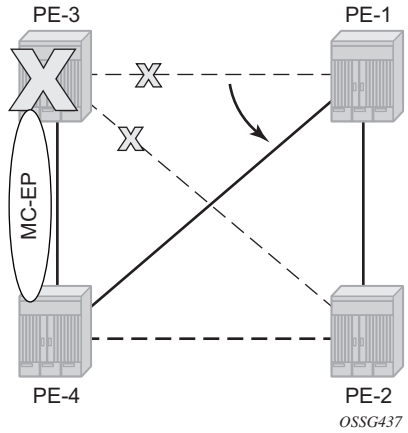

```
*A:PE-3# show service id 2 endpoint
===============================================================================
Service 2 endpoints
===============================================================================
Endpoint name               : CORE
Description                 : (Not Specified)
Creation Origin             : manual
Revert time                 : 0
Act Hold Delay              : 0
Ignore Standby Signaling    : false
Suppress Standby Signaling  : false
Block On Mesh Fail          : true
Multi-Chassis Endpoint      : 1
MC Endpoint Peer Addr       : 192.0.2.4
Psv Mode Active             : No
```

```
Tx Active (SDP)              : 32:1
Tx Active Up Time            : 0d 00:00:12
Revert Time Count Down       : N/A
Tx Active Change Count       : 1
Last Tx Active Change        : 05/20/2016 08:16:48
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 31:1 Prec:0                              Oper Status: Down
Spoke-sdp: 32:1 Prec:1                              Oper Status: Up
===============================================================================
===============================================================================
*A:PE-3#


*A:PE-4# show service id 2 endpoint
===============================================================================
Service 2 endpoints
===============================================================================
Endpoint name                : CORE
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Ignore Standby Signaling     : false
Suppress Standby Signaling   : false
Block On Mesh Fail           : false
Multi-Chassis Endpoint       : 1
MC Endpoint Peer Addr        : 192.0.2.3
Psv Mode Active              : No
Tx Active                    : none
Tx Active Up Time            : 0d 00:00:00
Revert Time Count Down       : N/A
Tx Active Change Count       : 3
Last Tx Active Change        : 05/20/2016 07:59:47
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 41:1 Prec:2                              Oper Status: Down
Spoke-sdp: 42:1 Prec:3                              Oper Status: Up
===============================================================================
===============================================================================
*A:PE-4#
```

Multi-Chassis Node Failure

Figure 295 **Multi-Chassis Node Failure**



When the multi-chassis node PE-3 fails, both spoke SDPs from PE-3 go down.

PE-4 reverts to single chassis mode and selects the best spoke SDP, which will be 41 between PE-4 and PE-1 (precedence 2). See Figure 295.

```
*A:PE-4# show redundancy multi-chassis mc-endpoint peer 192.0.2.3
===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.3           Peer Name            :
Admin State     : up                  Oper State           : down
Last State chg  :                     Source Addr          :
System Id       : 4a:c7:ff:00:00:00   Sys Priority         : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail     : 3
Passive Mode    : disabled            Psv Mode Oper        : No
Boot Timer      : 300                 BFD                  : enabled
Last update     : 05/20/2016 08:13:23 MC-EP Count          : 1
===============================================================================
*A:PE-4#

*A:PE-4# show service id 2 endpoint
===============================================================================
Service 2 endpoints
===============================================================================
Endpoint name                : CORE
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Ignore Standby Signaling     : false
Suppress Standby Signaling   : false
Block On Mesh Fail           : false
Multi-Chassis Endpoint       : 1
MC Endpoint Peer Addr        : 192.0.2.3
Psv Mode Active              : No
```

```
Tx Active (SDP)              : 41:1
Tx Active Up Time            : 0d 00:02:40
Revert Time Count Down       : N/A
Tx Active Change Count       : 1
Last Tx Active Change        : 05/20/2016 08:17:47
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 41:1 Prec:2                              Oper Status: Up
Spoke-sdp: 42:1 Prec:3                              Oper Status: Up
===============================================================================
===============================================================================
*A:PE-4#
```

Multi-Chassis Communication Failure

If the multi-chassis communication is interrupted, both nodes will revert to single chassis mode.

To simulate a communication failure between the two nodes, define a static route on PE-3 that will black-hole the system address of PE-4.

```
configure router
        static-route-entry 192.0.2.4/32
            black-hole
                no shutdown
            exit
        exit
```

Verify that the MC synchronization is down.

```
*A:PE-4# show redundancy multi-chassis mc-endpoint peer 192.0.2.3
===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.3           Peer Name            :
Admin State     : up                  Oper State           : down
Last State chg  :                     Source Addr          :
System Id       : 4a:c7:ff:00:00:00   Sys Priority         : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail     : 3
Passive Mode    : disabled            Psv Mode Oper        : No
Boot Timer      : 300                 BFD                  : enabled
Last update     : 05/20/2016 08:13:23 MC-EP Count          : 1
===============================================================================
*A:PE-4#
```

The spoke SDPs are active on PE-3 and on PE-4.

```
*A:PE-3# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)              : 31:1
```

```
Tx Active Up Time            : 0d 00:05:58
Tx Active Change Count       : 6
Last Tx Active Change        : 05/20/2016 08:19:09

*A:PE-4# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)              : 41:1
Tx Active Up Time            : 0d 00:04:56
Tx Active Change Count       : 3
Last Tx Active Change        : 05/20/2016 08:19:05
```

This can potentially cause a loop in the system. The Passive Mode explains how to avoid this loop.

Passive Mode

As in Multi-Chassis Communication Failure, if there is a failure in the multi-chassis communication, both nodes will assume that the peer is down and will revert to single-chassis mode. This can create loops because two spoke SDPs can become active.

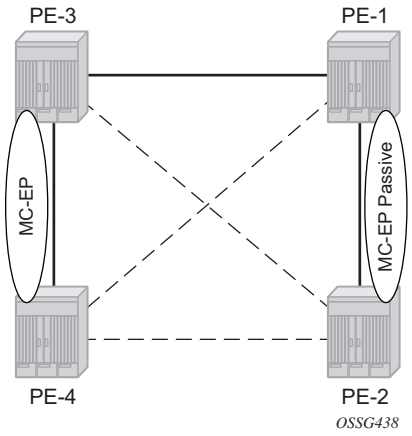
One solution is to synchronize the two core nodes, and configure them in passive mode. See Figure 296.

In passive mode, both peers will stay dormant as long as one active spoke SDP is signaled from the remote end. If more than one spoke SDP becomes active, the MC-EP algorithm will select the best SDP. All other spoke SDPs are blocked locally (in Rx and Tx directions). There is no signaling sent to the remote PEs.

If one peer is configured in passive mode, the other peer will be forced to passive mode as well.

The **no suppress-standby-signaling** and **no ignore-standby-signaling** commands are required.

Figure 296 **Multi-Chassis Passive Mode**



The following output shows the multi-chassis configuration on PE-1 (similar on PE-2).

```
*A:PE-1# configure redundancy
        multi-chassis
            peer 192.0.2.2 create
                mc-endpoint
                    no shutdown
                    passive-mode
                exit
                no shutdown
            exit
        exit
```

The following output shows the VPLS spoke SDPs configuration on PE-1 (similar on PE-2)

```
*A:PE-1# configure service
        vpls 1
            endpoint "METRO1" create
                no suppress-standby-signaling
                mc-endpoint 1
                    mc-ep-peer 192.0.2.2
                exit
            exit
            spoke-sdp 13:1 endpoint "METRO1" create
            exit
            spoke-sdp 14:1 endpoint "METRO1" create
            exit
            no shutdown
        exit
```

To simulate a communication failure between the two nodes, a static route is defined on PE-3 that will black-hole the system address of PE-4.

```
configure router
        static-route-entry 192.0.2.4/32
            black-hole
                no shutdown
            exit
        exit
```

The spoke SDPs are active on PE-3 and on PE-4.

```
*A:PE-3# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)              : 31:1
Tx Active Up Time            : 0d 00:00:28
Tx Active Change Count       : 8
Last Tx Active Change        : 05/20/2016 08:20:24

*A:PE-4# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)              : 41:1
Tx Active Up Time            : 0d 00:00:22
Tx Active Change Count       : 5
Last Tx Active Change        : 05/20/2016 08:20:25
```

PE-1 and PE-2 have blocked one spoke SDP which avoids a loop in the VPLS.

```
*A:PE-1# show service id 1 endpoint "METRO1" | match "Tx Active"
Tx Active (SDP)              : 13:1
Tx Active Up Time            : 0d 00:00:58
Tx Active Change Count       : 5
Last Tx Active Change        : 05/20/2016 08:20:50

*A:PE-2# show service id 1 endpoint  "METRO1" | match "Tx Active"
Tx Active                    : none
Tx Active Up Time            : 0d 00:00:00
Tx Active Change Count       : 2
Last Tx Active Change        : 05/20/2016 08:20:15
```

The passive nodes do not set the pseudowire status bits; therefore, the nodes PE-3 and PE-4 are not aware that one spoke SDP is blocked.

Conclusion

Multi-chassis endpoint for VPLS active/standby pseudowire allows the building of hierarchical VPLS without single point of failure, and without requiring STP to avoid loops.

Care must be taken to avoid loops. The multi-chassis peer communication is important and should be possible on different interfaces.

Passive mode can be a solution to avoid loops in case of multi-chassis communication failure.

Multi-Segment Pseudowire Routing

This chapter describes advanced multi-segment pseudowire routing configurations.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

Multi-segment pseudowire (MS-PW) routing is supported in SR OS release 9.0.R3, and later. This chapter was initially written for SR OS release 10.0.R4. The CLI in this edition is based on release 14.0.R3. There are no specific pre-requisites for this configuration.

Summary

SR OS supports the use of Multi-Segment Pseudowire (MS-PW) routing for Epipe services. MS-PW routing is described in draft-ietf-pwe3-dynamic-ms-pw, also known as dynamic placement of MS-PW and it is an extension of the procedures proposed in RFC 6073 (static MS-PW) to enable multi-segment pseudowires to be dynamically placed. Ultimately, MS-PW Routing provides the capability of setting up MS-PWs without provisioning the switching PEs (S-PEs).

This chapter will go through the configuration process required to set up MS-PW routing and will provide two configuration examples typically deployed by service providers: MS-PW within the same autonomous system (AS) and MS-PW across two different ASs. Different configuration options are shown and described for each example.

Overview

From a data plane perspective, MS-PW routing does not introduce any changes with respect to the existing MS-PW architecture. However, from the control plane perspective, MS-PW routing brings a new information model and set of procedures to set up a MS-PW. These are the building blocks defined by the MS-PW Routing feature:

- A new information model is introduced for dynamic MS-PW based on the FEC129, attachment individual identifier (AII) type 2. Static MS-PW uses FEC128 while VPLS with BGP-AD uses FEC129, but with AII Type 1 instead.
  - FEC129 is suitable for applications where the local PE with a SAII (source attachment individual identifier) must automatically learn the remote TAII (target attachment individual identifier), normally through BGP, before launching the LDP mapping message for the pseudowire setup. The following figure shows the FEC129 structure:

Figure 297 **FEC129 Structure**

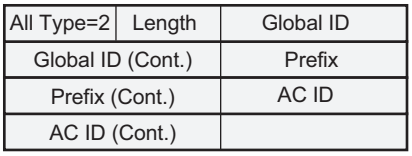
G.Pwid (0x81)	C	Pw Type	Pw Info Length
AGI Type	Length	Value	
AGI Value (Cont.)			
All Type	Length	Value	
SAII Value (Cont.)			
All Type	Length	Value	
TAII Value (Cont.)			

ACG0004A

  - The attachment group identifier (AGI) is not used in dynamic MS-PW signaling. In VPLS, it typically carries the instance identifier. It is zero in dynamic MS-PWs.
  - The SAII and TAII (or pseudowire end-point identifiers) are encoded in FEC129 and can have two different formats: AII type 1 or AII type 2.
- AII type 1 is composed of a fixed 32-bit value unique on the local PE. This AII type is used by VPLS when BGP-AD is needed.

- AII Type 2 is composed of GID:prefix:AC-ID (global-ID:prefix: attachment-circuit-ID) and allows for summarization; by this enhancing scalability in large networks. The GID is normally derived from the AS number, the prefix from the node system address and the AC-ID is the local pseudowire end-point identifier. The combination of the three identifiers gives us a globally unique 96-bit AII value. In general, the same global ID and prefix are assigned for all ACs belonging to the same Terminating PE (T-PE). This is not a strict requirement though.

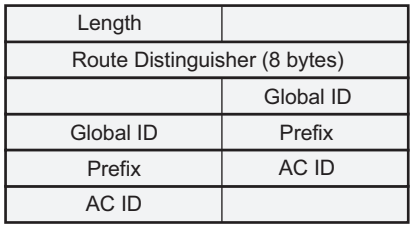
Figure 298 **AII Type 2 Format**



ACG0004B

- A MS-PW routing table must be built in all the T-PEs and S-PEs through one of the following two mechanisms:
  - Multi-protocol BGP (MP-BGP), using a dedicated NLRI and SAFI (pseudowire routing SAFI=6, with AFI=25 L2VPN). The FEC129 AII Type 2 global values are mapped in the pseudowire routing NLRI and advertised by BGP. SR OS supports an NLRI comprising a Length, RD, Global ID and 32-bit prefix, that is, the AC ID is not included in the advertised NLRI. The AC ID is not included as indicated in the draft-ietf-pwe3-dynamic-ms-pw since “the source T-PE knows by provisioning the AC ID on the terminating T-PE to use in signaling. Therefore, there is no need to advertise a “fully qualified” 96 bit address on a per pseudowire attachment circuit basis. Only the T-PE Global ID, Prefix, and prefix length needs to be advertised as part of well known BGP procedures”. This also minimizes the amount of routing information that is advertised in BGP to only what is necessary to reach the far-end T-PE. The MS-PW routing NLRI is shown in Figure 299:

Figure 299 **Pseudowire Routing NLRI (the AC ID is always zero)**



ACG0004C

  - Static routes, configurable via CLI

- Once the MS-PW routing table is populated, targeted LDP (TLDP) will make use of it to signal the MS-PW all the way from the originating T-PE to the terminating T-PE as well as in the reverse direction. The following methods will be used:
  - At the originating T-PE, a longest-match lookup will be performed in the pseudowire routing table for the configured TAII. Based on the lookup outcome, a label mapping message will be sent to the Next Signaling Hop (NSH).



Note: The “originating T-PE” will be the T-PE initiating the MS-PW signaling. Refer to the Active/Passive Signaling and Auto-Configuration section for further information.

  - At the intermediate S-PEs and terminating T-PE, a longest-match lookup between the TAII Type 2 included in the TLDP signaling message and entries installed in the pseudowire routing table will be performed.
  - Alternatively to the pseudowire routing table lookup, TLDP can also use explicit routing, as per section 7.4.2 of draft-ietf-pwe3-dyn-ms-pw. If that is the case, a “path” must be configured at the T-PEs. The originating T-PE will include an ERO (Explicit Route Object) in the TLDP label mapping, containing all the S-PE hops specified in the configured path. Each S-PE along the path will remove its own entry from the ERO and will forward the label mapping message to the next hop.

SR OS supports the information model and all the previously described methods:

- Dynamic placement through MP-BGP, with the pseudowire routing NLRI
- Static routes
- Explicit paths

In addition to the above, the following features are supported on dynamic MS-PW:

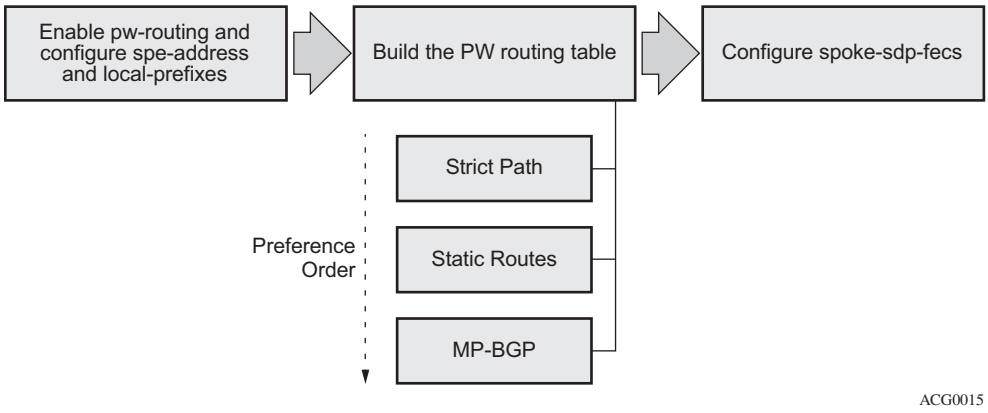
- Auto-configuration of spoke SDPs at T-PE (if enabled on a T-PE, there is no need for configuring the TAII of the remote T-PE. Refer to Active/Passive Signaling and Auto-Configuration. The auto-configuration is typically used in hub-and-spoke scenarios. The TAII would only be configured on the spoke T-PE while the TAII would be automatically provisioned on the hub T-PE if the auto-config parameter is added.
- OAM using virtual circuit connectivity verification vccv-ping and vccv-trace
- Pseudowire redundancy
- Control word
- Hash label (if the chassis dependency requirements are met)
- Standby-signaling-master and standby-signaling-slave commands

- Filters

Configuration

The following flow-chart shows the configuration process to be followed when setting up MS-PW routing. Base IGP and MPLS configuration is assumed to be in place before these configuration tasks can be carried out.

Figure 300 **Configuration Flow Chart**



The following subsections review these three steps, including all the options in detail.

- Pseudowire Routing Enablement
- Building the Pseudowire Routing Table
- Spoke-SDP-FEC Timers

Pseudowire Routing Enablement

The first step in the configuration is to enable pw-routing and configure the required pw-routing basic parameters: the spe-address (in S-PEs and T-PEs) and the local-prefix/prefixes (only required in T-PEs). A new pw-routing context has been added from 9.0 R3 onward. The following CLI examples show the configuration of the spe-address and local-prefixes.

```
*A:PE-1# configure service
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.11 create
                advertise-bgp route-distinguisher 65536:11 community 65535:11
```

```
            exit
            local-prefix 65536:192.0.2.12 create
                advertise-bgp route-distinguisher 65536:12 community 65535:12
            exit
        exit
```

In order to enable support for MS-PW routing on a 7x50 node, a single, globally unique, S-PE ID (known as the spe-address) is first configured under **config>service>pw-routing** on each 7x50 to be used as a T-PE or S-PE. The S-PE address has the format global-id:prefix. It is not possible to configure any local prefixes used for pseudowire routing or to configure spoke SDPs using dynamic MS-PWs at a T-PE unless an S-PE address has already been configured. The S-PE address is used as the address of a node when populating the switching point TLV in the LDP label mapping message and the pseudowire status notification sent for faults at an S-PE. The following CLI output shows the spe-address configuration format:

```
A:PE-1# configure service pw-routing spe-address
  - no spe-address
  - spe-address <global-id:prefix>

 <global-id:prefix>   : <global-id>:{<prefix>|<ipaddress>}
                        global-id - [1..4294967295]
                        prefix    - [1..4294967295]
                        ipaddress - a.b.c.d
```

Where:

- <global-id> is normally the 2 or 4-byte ASN identifying the network (although nothing prevents the operator from configuring any value here)
- <prefix> is normally the system address of the node (although any value in IP address or decimal format can be used)

If an S-PE is capable of dynamic MS-PW signaling, but is not assigned with an S-PE address, then on receiving a dynamic MS-PW label mapping message the S-PE will return a label release with the "LDP_RESOURCES_UNAVAILABLE" (0x38)" status code. The S-PE address cannot be changed unless the dynamic MS-PW configuration is completely removed; therefore Nokia recommends to configure the spe-address carefully and keep it for the life of the services.

The second basic pw-routing context parameter is the local-prefix:

```
A:PE-1# configure service pw-routing local-prefix
  - local-prefix <local-prefix> [create]
  - no local-prefix <local-prefix>

 <local-prefix>       : <global-id>:<ip-addr>|<raw-prefix>
                        ip-addr      - a.b.c.d
                        raw-prefix   - [1..4294967295]
                        global-id    - [1..4294967295]
```

```
 [no] advertise-bgp   - Configure BGP advertisement
```

One or more local (Layer 2) prefixes (up to a maximum of 16), which are formatted in the style of <global-id>:<ipv4-address>, are supported. A local prefix identifies a T-PE in the pseudowire routing domain. When using explicit paths or static-routes, the definition of the local-prefix (or local-prefixes) without any further attribute is enough. However, when BGP is used, the advertise-bgp parameter along with a route-distinguisher (RD) value and an optional BGP community is required.

```
*A:PE-1# configure service pw-routing local-prefix 65536:192.0.2.11 advertise-bgp
  - advertise-bgp route-distinguisher <rd> [community <community>]
  - no advertise-bgp route-distinguisher <rd>

 <rd>                 : <ip-addr:comm-val>|<2byte-asnumber:ext-comm-val>|
                                          <4byte-asnumber:comm-val>
                        ip-addr        - a.b.c.d
                        comm-val       - [0..65535]
                        2byte-asnumber - [1..65535]
                        ext-comm-val   - [0..4294967295]
                        4byte-asnumber - [1..4294967295]
 <community>          : <asnumber:comm-val>
                        asnumber - [1..65535]
                        comm-val - [0..65535]
```

Up to four unique RDs (and communities) can be configured per each local-prefix. Different RDs for the same prefix allow the operator to advertise the same prefix coming from up to four different next-signaling hops (NSH). Route-Reflectors would reflect the four routes in that case, whereas only one would be reflected should the same RD be used.

```
*A:PE-1>config>service>pw-routing>local-prefix# info
----------------------------------------------
                advertise-bgp route-distinguisher 400:20
                advertise-bgp route-distinguisher 500:3
                advertise-bgp route-distinguisher 600:300
                advertise-bgp route-distinguisher 65536:11 community 65535:11

*A:PE-1>config>service>pw-routing>local-prefix# advertise-bgp route-distinguisher
700:100
MINOR: SVCMGR #6072 Maximum number of RD's has been reached
```

For each local prefix, BGP then advertises each global ID/prefix tuple and unique RD and community (if configured) using the MS-PW NLRI, based on the aggregated FEC129 AII Type 2 and the Layer 2 VPN/PW routing AFI/SAFI 25/6, to each BGP neighbor, subject to local BGP policies.

Building the Pseudowire Routing Table

Once the spe-address and the local-prefix(es) have been configured and before configuring the Epipe service itself on the T-PE nodes, we need to populate the pseudowire routing table in all the participating T-PE and S-PE nodes, so that TLDP knows what the Next Signaling Hop (NSH) is and sends LDP label mapping messages.

The pseudowire routing table will be populated with local prefixes, static-routes and BGP routes, where the static-routes have preference over the BGP-learned routes. The pseudowire routing table can be overridden by the explicit paths, should the operator want to configure them. Therefore, when TLDP signals an LDP Label Mapping for a given TAII, it will:

- First check if there is an explicit path configured for that spoke-sdp-fec.
- Otherwise it will look up the TAII prefix into the pseudowire routing table, where static routes take precedence over BGP routes.

An aggregation scheme, similar to that used for classless IPv4 addresses, can be employed in the pseudowire routing table, where a longest match is used to find a route. Except for the default pseudowire route, which is encoded with a 0 mask, masks included in the pw-routing table are:

- /64 for regular prefixes, including a global-id and prefix (as previously mentioned, note that the AC-ID is not included in the BGP NLRI).
- /96 for local prefixes, including the AC-ID, as well as global-id and prefix.

Each S-PE and T-PE must have a pseudowire routing table that contains a reference to the TLDP session to use to signal to a set of next hop S-PEs to reach a given T-PE (or the T-PE if that is the next hop). For Epipes, this table contains aggregated AII Type 2 FECs and may be populated with routes that are learned through MP-BGP or that are statically configured.

Explicit Paths

A set of default explicit routes to a remote T-PE prefix may be configured on a T-PE under **config>services>pw-routing** using the path name command. Explicit paths are used to populate the explicit route TLV used by MS-PW TLDP signaling. Only strict (fully qualified) explicit paths are supported. It is possible to configure explicit paths independently of the configuration of BGP or static routing.

The following CLI excerpt shows an explicit path example for a MS-PW following the PE-1-PE-3-PE-5-PE-2 path (see the diagram in Figure 301). The IP addresses are the system addresses of all the S-PE and T-PE along the path (except for PE-1).

```
*A:PE-1# configure service pw-routing
            path "path-1" create
                hop 1 192.0.2.3
                hop 2 192.0.2.5
                hop 3 192.0.2.2
                no shutdown
            exit
```

Static Routes

In addition to support for BGP routing, static MS-PW routes may also be configured using the **config>services>pw-routing>static-route** command. Each static route comprises of the target T-PE Global-ID and prefix, and the IP address of the TLDP session to the next hop S-PE or T-PE that should be used:

```
A:PE-1# configure service pw-routing static-route
  - no static-route <route-name>
  - static-route <route-name>

 <route-name>         : <global-id>:<prefix>:<next-hop-ip_addr>
                        global-id      - 0..4294967295
                        prefix         - a.b.c.d|0..4294967295
                        ip_addr        - a.b.c.d
```

If a static route <global-id>:<prefix> is set to 0, then this represents the default route.

```
*A:PE-1# configure service pw-routing
    ---snip---
    static-route 0:0.0.0.0:192.0.2.3
    static-route 0:0.0.0.0:192.0.2.4
```

Even though several default-routes can be configured, only one default route is added to the pseudowire routing table. The following command shows the pseudowire routing table content where only one default route (out of the two previously configured ones) is added. The default route added to the pseudowire routing table is the first valid route added to the configuration.

```
A:PE-1# show service pw-routing route-table all-routes
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
```

```
-------------------------------------------------------------------------------
0:0.0.0.0:0/0                                    192.0.2.3      static 19h11m57s
 0:0                                              0:0            yes
...
```

If a static route exists to a given T-PE, then this is used in preference to any BGP route that may exist.

BGP Routes

As already mentioned, the dynamic advertisement of the pseudowire routes is enabled for each prefix and RD using the **advertise-bgp** command in the **config>services>pw-routing>local-prefix** context. A BGP export policy is required in order to export MS-PW routes in MP-BGP. This can be done using a default policy matching all the MS-PW routes, such as the following:

```
*A:PE-1# configure router
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit

*A:PE-1# configure router
        autonomous-system 65536
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "region"
                family ms-pw
                export "export_ms-pw"
                peer-as 65536
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
```

MS-PW routes advertised/received can be debugged and shown on the log sessions (**debug router bgp update**). Adedicated MS-PW address family and NLRI are used to distribute the MS-PW prefixes:

```
28 2015/06/03 07:42:59.69 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 51
    Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
        Address Family MSPW
        NextHop len 4 NextHop 192.0.2.1
       [MSPW] rd: 65536:12, global-id 65536, prefix 192.0.2.12,  ac-id 0, preflen 128
    Flag: 0x40 Type: 1 Len: 1 Origin: 2
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        65535:12
"
```

MS-PW BGP routes can also be displayed in the pseudowire routing table along with the static-routes and the local-prefixes.

```
*A:PE-1# show service pw-routing route-table
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
-------------------------------------------------------------------------------
0:0.0.0.0:0/0                                 192.0.2.3      static 01h09m00s
 0:0                                           0:0            yes
65536:192.0.2.11:0/64                         192.0.2.1      local  17h19m42s
 0:0                                           0:0            yes
65536:192.0.2.11:0/64                         192.0.2.1      local  17h19m42s
 65536:11                                      65535:11       yes
65536:192.0.2.12:0/64                         192.0.2.1      local  17h19m42s
 0:0                                           0:0            yes
65536:192.0.2.12:0/64                         192.0.2.1      local  17h19m42s
 65536:12                                      65535:12       yes
65536:192.0.2.13:0/64                         192.0.2.1      local  00h49m29s
 0:0                                           0:0            yes
65536:192.0.2.14:0/64                         192.0.2.1      local  00h49m29s
 0:0                                           0:0            yes
65536:192.0.2.21:0/64                         192.0.2.3      bgp    00h05m46s
 65536:21                                      65535:11       yes
65536:192.0.2.22:0/64                         192.0.2.4      bgp    00h05m42s
 65536:22                                      65535:12       yes
65536:192.0.2.23:0/64                         192.0.2.3      static 00h49m29s
 0:0                                           0:0            yes
65536:192.0.2.24:0/64                         192.0.2.4      static 00h49m29s
 0:0                                           0:0            yes
-------------------------------------------------------------------------------
Entries found: 11
===============================================================================
```

If there are two (or more) equal cost BGP MS-PW routes with identical <global-ID:prefix> and different RDs in the RIB, they are both tagged as best/used and both will be added to the pseudowire routing table, however, only the one with a higher RD will be shown as “Best” and as a result of that, only that one will be used by TLDP for the NSH.

The pw-routing context at PE-2 contains the following advertise-bgp entries with different RDs for local-prefix 65536:192.0.2.2:

```
*A:PE-2# configure service pw-routing
            local-prefix 65536:192.0.2.2 create
                advertise-bgp route-distinguisher 65536:21
                advertise-bgp route-distinguisher 65536:22
            exit
```

The following CLI output shows an example of two equal cost MS-PW routes. The route 65536:192.0.2.2 with RD 65536:21 and RD 65536:22 are tagged as best/used (u*>):

```
*A:PE-1# show router bgp routes ms-pw aii-type2 65536:192.0.2.2:0
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
u*>?  65536:192.0.2.2         65536:21
      192.0.2.3               65536:192.0.2.2:0/64
      No As-Path
*?    65536:192.0.2.2         65536:21
      192.0.2.4               65536:192.0.2.2:0/64
      No As-Path
u*>?  65536:192.0.2.2         65536:22
      192.0.2.4               65536:192.0.2.2:0/64
      No As-Path
*?    65536:192.0.2.2         65536:22
      192.0.2.3               65536:192.0.2.2:0/64
      No As-Path
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

However, only the one with RD 65536:22 (higher RD) is added as “Best” to the pseudowire routing table and TLDP will use 192.0.2.4 as the NSH:

```
*A:PE-1# show service pw-routing route-table all-routes
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
-------------------------------------------------------------------------------
65536:192.0.2.2:0/64                          192.0.2.3      bgp    00h12m22s
 65536:21                                      65535:11       no
65536:192.0.2.2:0/64                          192.0.2.4      bgp    00h12m17s
 65536:22                                      65535:12       yes
---snip---
```

How does the SR OS TLDP process select the NSH (Next-Signaling Hop) for two identical <global-ID:prefix/RD> tuples?

In case the originating T-PE or any intermediate S-PE receives two (or more) equal cost MS-PW routes with the same RD but from different Next-Hops, all the MS-PW routes will be added to the MS-PW routing table. The following output shows two MS-PW routes with the same <global-ID:prefix/RD> but different NH. Both are added to the MS-PW routing table as “Best”.

```
*A:PE-1# show service pw-routing route-table all-routes
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
-------------------------------------------------------------------------------
---snip---
65536:192.0.2.2:0/64                          192.0.2.3      bgp    01d15h21m
 65536:21                                      65535:21       yes
65536:192.0.2.2:0/64                          192.0.2.4      bgp    01d15h21m
 65536:21                                      65535:21       yes
---snip---
```

If that is the case, TLDP will pick up the NSH out of an ECMP hashing algorithm applied to the <global-ID:prefix:AC-ID> for the SAII and the TAII of the pseudowires pointing at the same prefix. The output of that hashing algorithm will determine what the NSH will be for a given spoke-sdp-fec.

When path diversity for an active and a standby pseudowire (hot standby pseudowire redundancy) is desired and the two pseudowires of the same Epipe end-point are pointing at the same remote <global-ID:prefix> coming from two different NHs, the operator has to make sure TLDP chooses a different NSH for the standby pseudowire. Only in that case hot standby pseudowire redundancy can be achieved. As a rule of thumb, if the SAII/TAII of the active and standby pseudowires are separated by 16 or more AC-ID values, TLDP will select a different NSH for both pseudowires.

For example:

- Given the following SAII/TAII AC-ID values for the active/standby pseudowires on the originating T-PE, TLDP will select the same NSH:
  - Active pseudowire: saii-type2 — 65536:192.0.2.1:1, taii-type2 — 65536:192.0.2.2:1
  - Standby pseudowire: saii-type2 — 65536:192.0.2.1:2, taii-type2 — 65536:192.0.2.2:2
- However, the following SAII/TAII AC-ID values for the active/standby pseudowires on the originating T-PE will allow the ECMP hashing algorithm to make TLDP select different NSHs for the active and the standby pseudowires:
  - Active pseudowire: saii-type2 — 65536:192.0.2.1:1, taii-type2 — 65536:192.0.2.2:1
  - Standby pseudowire: saii-type2 — 65536:192.0.2.1:16, taii-type2 — 65536:192.0.2.2:16

Other AC-ID values greater than 16 (for the standby pseudowire) would also have achieved next hop diversity.

Configuring Dynamic Pseudowires on the T-PEs

Before any LDP signaling can take place, T-LDP sessions must be explicitly configured on T-PEs and S-PEs.

One or more spoke-SDPs may be configured for distributed Epipe VLL services. Dynamic MS-PWs use FEC129 (also known as the Generalized ID FEC) with Attachment Individual Identifier (AII) Type 2 to identify the pseudowire, as opposed to FEC128 (also known as the PW ID FEC) used for traditional single segment pseudowires and for pseudowire switching. FEC129 spoke-SDPs are configured under the spoke-sdp-fec command in the CLI. Spoke-sdp-fecs (or FEC129 spoke-SDPs) are by default fec-type 129 and aii-type 2. Spoke-sdp-fecs can be part of an endpoint and even an ICB (Inter-Chassis Backup) pseudowire.

```
*A:PE-1# configure service epipe 2 spoke-sdp-fec
```

```
  - no spoke-sdp-fec <spoke-sdp-fec-id>
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
  - spoke-sdp-fec <spoke-sdp-fec-id> no-endpoint
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
endpoint <name> [icb]

 <spoke-sdp-fec-id>   : [1..4294967295]
 <fec-type>           : [129..130]
 <aii-type>           : [1..2]
 <name>               : [32 chars max]
 <icb>                : keyword - configure spoke-sdp as inter-chassis backup
```

FEC129 AII Type 2 uses a SAII and a TAII to identify the ends of a pseudowire at the T-PE. The SAII identifies the local end, while the TAII identifies the remote end. The SAII and TAII are each structured as follows:

- Global-ID: this is a 4 byte identifier that uniquely identifies an operator or the local network. Normally this matches the ASN
- Prefix: a 4-byte prefix, which should correspond to one of the local prefixes assigned under pw-routing
- AC-ID: a 4-byte identifier for this end of the pseudowire. This should be locally unique within the scope of the global-id:prefix

In terms of the SDP tunnel being used by each spoke-sdp-fec, pw-routing chooses the MS-PW path in terms of the sequence of S-PEs to use to reach a T-PE. It does not select the SDP to use on each hop, which is instead determined at signaling time. When a label mapping is sent for a pseudowire segment, an LDP SDP will be used to reach the next-hop S-PE/T-PE if such an SDP exists. If not, and an RFC 3107 labeled BGP SDP is available, then that will be used. Otherwise, the label mapping will fail and a label release will be sent.

Note: The RSVP SDPs might be picked at the T-PE through the use of pw-template <policy-id> [use-provisioned-sdp], however there is no way to select an RSVP SDP on an S-PE.

The following CLI output shows one example of two spoke-sdp-fecs belonging to an endpoint:

```
*A:PE-1# configure service
        pw-template 1 create
            controlword
        exit
        epipe 2 customer 1 create
            description "ms-pw epipe with bgp - using 2 prefixes"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
```

```
            sap 1/1/4:2 create
            exit
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.11:1
                taii-type2 65536:192.0.2.21:1
                no shutdown
            exit
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65536:192.0.2.22:1
                no shutdown
            exit
            no shutdown
        exit
```

These are all of the options available in the spoke-sdp-fec context:

```
*A:PE-1# configure service epipe 1 spoke-sdp-fec
  - no spoke-sdp-fec <spoke-sdp-fec-id>
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
  - spoke-sdp-fec <spoke-sdp-fec-id> no-endpoint
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
                                  endpoint <name> [icb]

 <spoke-sdp-fec-id>   : [1..4294967295]
 <fec-type>           : [129..130]
 <aii-type>           : [1..2]
 <name>               : [32 chars max]
 <icb>                : keyword - configure spoke-sdp as inter-chassis backup

 [no] auto-config     - Configure auto-configuration
 [no] path            - Configure path-name
 [no] precedence      - Configure precedence
 [no] pw-template-bi* - Configure Pseudo-Wire template-binding policy
 [no] retry-count     - Configure retry count
 [no] retry-timer     - Configure retry timer
 [no] saii-type2      - Configure Source Attachment Individual Identifier (SAII)
 [no] shutdown        - Administratively enable/disable the spoke SDP FEC binding
      signaling       - Configure Spoke-SDP FEC signaling
 [no] standby-signal* - Enable PW standby-signaling slave
 [no] taii-type2      - Configure Target Attachment Individual Identifier (TAII)
```

Active/Passive Signaling and Auto-Configuration

When an MS-PW is signaled, each T-PE might independently initiate signaling of the MS-PW. This could result in a different path being used in each direction of the pseudowire. To avoid this situation one of the T-PEs will start the pseudowire signaling (active role), while the other T-PE waits to receive the LDP label mapping message before sending the LDP label mapping message for the reverse direction of the pseudowire (passive role).

Enable debugging for LDP messages on PE-2:

```
*A:PE-2# debug router ldp peer 192.0.2.5 packet init detail
*A:PE-2# debug router ldp peer 192.0.2.5 packet label detail
*A:PE-2# debug router ldp peer 192.0.2.6 packet init detail
*A:PE-2# debug router ldp peer 192.0.2.6 packet label detail
```

By default, the T-PE with SAII>TAII will have the active role and will send the label mapping first. When spoke-sdp-fec 21 is first disabled, and then enabled, PE-2 sends a label mapping to PE-5 first (message 77 in following output). Afterwards, it receives a label mapping packet from PE-5 (message 78).

```
*A:PE-2# configure service epipe 2 spoke-sdp-fec 21 shutdown
*A:PE-2# configure service epipe 2 spoke-sdp-fec 21 no shutdown


*A:PE-2# show log log-id 3
===============================================================================
Event Log 3
===============================================================================
78 2016/06/03 12:07:01.09 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Mapping packet (msgId 11220) from 192.0.2.5:0
Protocol version = 1
Label 262131 advertised for the following FECs
Service FEC GENPWE3: ENET(5)
AGI = type: 1, len: 8, val: 00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
Group ID = 0 cBit = 1
Interface parameter Mtu = 1500
Interface parameter VCCV = 0x106
PW status bits = 0x18
Switching hop: System = 192.0.2.3, Remote System = 192.0.2.1
previous segment fec AGI = type: 1, len: 8, val: 00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
S-PE = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.3, AcId: 0
Switching hop: System = 192.0.2.5, Remote System = 192.0.2.3
previous segment fec AGI = type: 1, len: 8, val: 00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
S-PE = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.5, AcId: 0
"
```

```
77 2016/06/03 12:07:01.09 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 8687) to 192.0.2.5:0
Protocol version = 1
Label 262132 advertised for the following FECs
Service FEC GENPWE3: ENET(5)
AGI = type: 1, len: 8, val: 00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
Group ID = 0 cBit = 1
Interface parameter Mtu = 1500
Interface parameter VCCV = 0x306
PW status bits = 0x0
Recv Label Mapping packet (msgId 30) from 192.0.2.3:0
```

For the other T-PE, it is the other way round. PE-1 receives a label mapping packet first before it sends a label mapping packet back.

This default behavior can be modified by the signaling command. When set to master, the T-PE will send a label mapping message regardless of the SAII and TAII. By default the parameter is set to auto (which means the T-PE will trigger label mapping if SAII>TAII).

```
*A:PE-1# configure service epipe 1 spoke-sdp-fec 21 signaling
  - signaling <signaling>

 <signaling>          : auto|master

*A:PE-1# configure service epipe 2 spoke-sdp-fec 21
                shutdown
                signaling master
                no shutdown
                exit
```

The MS-PW routing implementation on SR OS supports single-sided auto-provisioning. This allows it to have “hub” T-PEs where the TAII is not required to be configured and as such simplifies the provisioning. In this case, the spoke T-PE PWs would be configured with specific SAII and TAII as well as signaling master, whereas the hub T-PE PWs would be configured with only the SAII and the auto-config parameter. When the auto-config attribute is set for a spoke-sdp-fec, the T-PE always passively waits for the label mapping to be received before issuing a label mapping message (since it does not know the TAII beforehand). This is a CLI example for a hub T-PE spoke-sdp-fec:

```
*A:PE-2# configure service epipe 2
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                auto-config
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.21:1
```

```
            no shutdown
        exit
```

Spoke-SDP-FEC Timers

MS-PW routing provides a few timers that can be configured at the global pw-routing level or at each specific spoke-sdp-fec level:

```
*A:PE-1# configure service pw-routing
             boot-timer 20
             retry-timer 40
             retry-count 50

*A:PE-1# configure service epipe 2 spoke-sdp-fec 21
               retry-timer 10
               retry-count 10
```

Where:

- Boot-timer (the default is 10 seconds with values 0 — 600 seconds allowed): Configures a hold-off timer for MS-PW routing advertisements and signaling that is used at boot time. This timer helps to make sure all the network infrastructure is up and running before setting up the PWs.
- Retry-timer (the default is 30 seconds with values 10 — 480 seconds allowed): The exponential back-off timer that determines the interval between consecutive retries to re-establish a spoke-SDP. The configured value gives the initial retry time. The attempt fails if a label withdrawal is received. If configured at global and spoke-sdp-fec level, the latter overrides the value set by the global settings.
- Retry-count (the default 30 with values 10 — 10000): Specifies the number of attempts the system should make to re-establish the spoke-SDP after it has failed. After each successful attempt, the counter is reset to zero. When the specified number is reached, no more attempts are made and the spoke-sdp is put into the shutdown state. Use the **no shutdown** command to bring up the path after the retry limit is exceeded. It is present at the pw-routing level as well as the spoke-SDP level. If configured at global and spoke-sdp-fec level, the latter overrides the value set by the global settings.
- The usual endpoint level timers are also available for MS-PW routing:
  - Revert-time <time-value|infinite> (default is 0, values 0-600 sec): configures the time to wait before reverting to the primary spoke-sdp-fec.

– Active-hold-delay (the default is 0, values 0 — 60 deci-seconds): It specifies that the node will delay sending the T-LDP status bits for VLL endpoint when the MC-LAG transitions the LAG subgroup which hosts the SAP from active to standby (MC-Ring or MC-APS are supported too) or when any object in the endpoint, i.e. SAP, ICB, or regular spoke SDP, transitions from up to down operational state. The active-hold-delay range starts from 1 (in units of deci-sec) via CLI, and the only way to get the default value of zero is to use the **no active-hold-delay** command

Standby Signaling

Just as with a regular endpoint with regular spoke-sdps, there can also be standby-signaling-master and standby-signaling-slave parameters for spoke-sdp-fecs.

The standby-signaling-master command is configured in the end-point context and makes sure that standby signaling (TLDP pseudowire status bits 0x20) is sent for the selected standby pseudowire.

```
*A:PE-1# configure service epipe 2 endpoint "CORE" standby-signaling-master
```

It is not allowed to add a SAP associated to an endpoint configured as standby-signaling-master to an Epipe.

```
*A:PE-1>config>service>epipe# sap 1/1/4:2 endpoint "CORE" create
MINOR: SVCMGR #6025 The endpoint has standby-signaling-master configured
```

Standby-signaling-master cannot be set if SAPs have been configured at the end-point (for MC-LAG/Ring/APS or ICB).

```
*A:PE-1>config>service>epipe>endpoint# standby-signaling-master
MINOR: SVCMGR #3805 The command is not allowed in an endpoint with sap
```

The standby-signaling-slave can be configured at endpoint or spoke-sdp-fec level (if the spoke-sdp-fec is not part of an endpoint) but never on both at the same time:

```
*A:PE1>config>service>epipe>endpoint# info
----------------------------------------------
                standby-signaling-slave

*A:PE1>config>service>epipe>spoke-sdp-fec# standby-signaling-slave
MINOR: SVCMGR #2031 Sdp-bind is in an explicit endpoint

*A:PE1>config>service>epipe# info
----------------------------------------------
            sap 1/1/3:3 create
            exit
```

```
            spoke-sdp-fec 11 fec 129 aii-type 2 create
                standby-signaling-slave
```

When this parameter is configured, the node will block the transmit forwarding direction of a spoke SDP based on the pseudowire standby bit received from a TLDP peer.

Spoke-SDP-FEC Templates and Filters

PW-templates are the way to configure the control word for this type of pseudowire as well as ingress/egress filters (ipv4/mac/ipv6). Filters are only supported on the T-PEs, because there is no provisioning of a pw-template (or Epipe at all) on the S-PEs.

```
*A:PE-1# configure service
        pw-template 1 create
            controlword
            egress
                filter ip 1
            exit
        exit


*A:PE-1# configure service
        epipe 2 customer 1 create
---snip---
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                standby-signaling-slave
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65536:192.0.2.22:1
                no shutdown
            exit
```

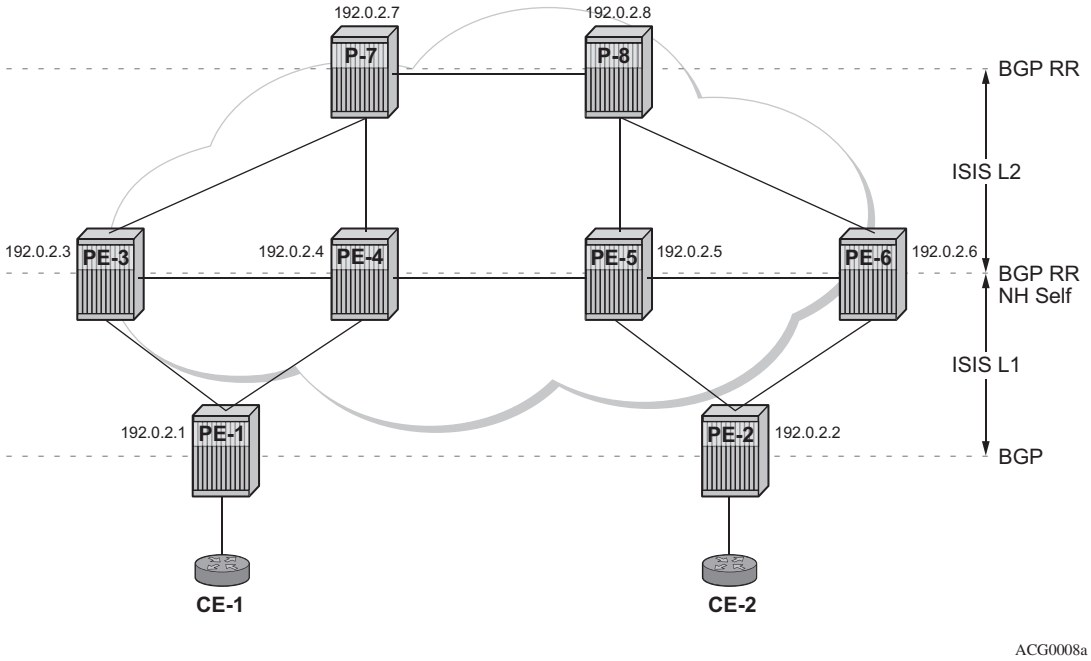
Pw-template changes (just like for VPLS with BGP-AD or BGP-VPLS) are not automatically propagated. A tools perform command is provided to evaluate and distribute the changes at the service level to one or all the services that use that template (if the service ID is omitted, then all the services will be updated).

```
*A:PE-1# tools perform service id 2 eval-pw-template 1 allow-service-impact
```

Intra-AS MS-PW Routing

This section provides a configuration example for an intra-AS scenario. The following example setup will be used for this section.

Figure 301 **Intra-AS MS-PW Example Topology**



Multiple MS-PW routing Epipes are to be configured between PE-1 and PE-2, with PE-3, PE-4, PE-5 and PE-6 being S-PE routers. P-7 and P-8 are pure P routers from a data plane perspective.

All the PEs are pre-configured with IS-IS as the IGP, as shown in the figure: PE-1 and PE-2 are level-1 routers, P-7 and P-8 are level-2 only routers and the rest of the routers are level-1/level-2. Link level LDP is also pre-configured on all the network interfaces and targeted LDP is configured between PE-1 and PE-3/PE-4, between PE-2 and PE-5/PE-6 and among PE-3, PE-4, PE-5 and PE-6. There is no targeted LDP sessions configured on P-7 and P-8.

As outlined in Figure 300, the configuration is a three-step process where the pw-routing context is configured first, then the required configuration so that routing tables get populated accordingly and finally the services themselves.

MS-PW using BGP Routing

In this sub-section, Epipe 2 will be configured between PE-1 and PE-2, where TLDP will use the BGP routes populated in the MS-PW routing table to signal the MS-PW.

The first step is the provisioning of the pw-routing context on all the T-PEs and S-PEs. The spe-address will be configured on all the T-PEs and S-PEs (that is, all the routers except for P-7 and P-8) using the ASN as the global-id and the system address as the prefix. On PE-1 and PE-2 (only) the prefixes used for setting up Epipe 2 are configured. Two prefixes are configured per T-PE so that pseudowire redundancy with path diversity for the standby pseudowire can be carried out. The spe-address and local-prefixes for the T-PEs are shown in the CLI output. The advertise-bgp parameter is required because BGP is used here.

```
*A:PE-1# configure service pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.11 create
                advertise-bgp route-distinguisher 65536:11 community 65535:11
            exit
            local-prefix 65536:192.0.2.12 create
                advertise-bgp route-distinguisher 65536:12 community 65535:12
            exit

*A:PE-2# configure service pw-routing

            spe-address 65536:192.0.2.2
            local-prefix 65536:192.0.2.21 create
                advertise-bgp route-distinguisher 65536:21 community 65535:11
            exit
            local-prefix 65536:192.0.2.22 create
                advertise-bgp route-distinguisher 65536:22 community 65535:12
            exit
```

The second step is the configuration of BGP.

As depicted in Figure 301, BGP is enabled in all the routers. The middle routers (PE-3, PE-4 and PE-5, PE-6) are BGP route-reflectors for PE-1 and PE-2 and they reflect MS-PW routes while changing the next-hop to their own system address. This is required so that TLDP knows where to send the label mapping message for a particular prefix. P-7 and P-8 are regular RRs reflecting routes among all the S-PEs. The BGP configuration of PE-1, PE-3, PE-4 and a P-7 is as follows. Similar commands are configured on the other PEs depending on their T-PE, S-PE or RR function.

The T-PEs have dual-homed BGP sessions to the S-PEs. Example for PE-1:

```
*A:PE-1# configure router
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                    exit
                exit
```

```
            exit
            commit
        exit

*A:PE-1# configure router
        autonomous-system 65536
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "region"
                family ms-pw
                export "export_ms-pw"
                peer-as 65536
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
    exit
```

The S-PEs are reflecting routes and also changing the NH and local preference based on the communities accordingly, so that pseudowire diversity can be ensured.

```
*A:PE-3# configure router
        policy-options
            begin
            community "65535:11" members "65535:11"
            community "65535:12" members "65535:12"
            policy-statement "export_ms-pw_ABR-to-core"
                entry 10
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        local-preference 100
                        next-hop-self
                    exit
                exit
            exit
            policy-statement "export_ms-pw_ABR-to-region"
                entry 10
                    from
                        protocol bgp
```

```
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        local-preference 100
                        next-hop-self
                    exit
                exit
            exit
            commit
        exit


*A:PE-3# configure router
        autonomous-system 65536
        bgp
            rapid-withdrawal
            group "core"
                family ms-pw
                export "export_ms-pw_ABR-to-core"
                peer-as 65536
                neighbor 192.0.2.7
                exit
                neighbor 192.0.2.8
                exit
            exit
            group "region"
                family ms-pw
                cluster 3.3.3.3
                export "export_ms-pw_ABR-to-region"
                peer-as 65536
                enable-peer-tracking
                neighbor 192.0.2.1
                exit
            exit
```

The second S-PE to which PE-1 is connected has the following BGP configuration:

```
*A:PE-4# configure router
        policy-options
            begin
            community "65535:11" members "65535:11"
            community "65535:12" members "65535:12"
            policy-statement "export_ms-pw_ABR-to-core"
                entry 10
                    from
                        protocol bgp
```

```
                            community "65535:12"
                            family ms-pw
                        exit
                        action accept
                            local-preference 150
                            next-hop-self
                        exit
                    exit
                    entry 20
                        from
                            protocol bgp
                            community "65535:11"
                            family ms-pw
                        exit
                        action accept
                            local-preference 100
                            next-hop-self
                        exit
                    exit
                exit
                policy-statement "export_ms-pw_ABR-to-region"
                    entry 10
                        from
                            protocol bgp
                            community "65535:12"
                            family ms-pw
                        exit
                        action accept
                            local-preference 150
                            next-hop-self
                        exit
                    exit
                    entry 20
                        from
                            protocol bgp
                            community "65535:11"
                            family ms-pw
                        exit
                        action accept
                            local-preference 100
                            next-hop-self
                        exit
                    exit
                exit
                commit
            exit


*A:PE-4# configure router
            autonomous-system 65536
            bgp
                rapid-withdrawal
                group "core"
                    family ms-pw
                    export "export_ms-pw_ABR-to-core"
                    peer-as 65536
                    neighbor 192.0.2.7
                    exit
                    neighbor 192.0.2.8
```

```
            exit
        exit
        group "region"
            family ms-pw
            cluster 4.4.4.4
            export "export_ms-pw_ABR-to-region"
            peer-as 65536
            enable-peer-tracking
            neighbor 192.0.2.1
            exit
        exit
    exit
```

Finally this is the BGP configuration for P-7 and P-8. These are pure RRs.

```
*A:P-7# configure router
        autonomous-system 65536
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "core"
                family ms-pw
                cluster 1.1.1.1
                peer-as 65536
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.6
                exit
            exit
        exit
```

After BGP is properly configured and the BGP update exchange takes place, the RIBs are properly populated and the required prefixes uploaded into the MS-PW routing table. An example for PE-1's RIB and pseudowire routing table is shown.

```
*A:PE-1# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
```

```
-------------------------------------------------------------------------------
u*>?  65536:192.0.2.21        65536:21
      192.0.2.3               65536:192.0.2.21:0/64
      No As-Path
*?    65536:192.0.2.21        65536:21
      192.0.2.4               65536:192.0.2.21:0/64
      No As-Path
u*>?  65536:192.0.2.22        65536:22
      192.0.2.4               65536:192.0.2.22:0/64
      No As-Path
*?    65536:192.0.2.22        65536:22
      192.0.2.3               65536:192.0.2.22:0/64
      No As-Path
---snip---
-------------------------------------------------------------------------------

*A:PE-1# show service pw-routing route-table
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
-------------------------------------------------------------------------------
65536:192.0.2.11:0/64                         192.0.2.1      local  01h03m37s
 65536:11                                      65535:11       yes
65536:192.0.2.12:0/64                         192.0.2.1      local  01h03m37s
 65536:12                                      65535:12       yes
---snip---
65536:192.0.2.21:0/64                         192.0.2.3      bgp    00h16m42s
 65536:21                                      65535:11       yes
65536:192.0.2.22:0/64                         192.0.2.4      bgp    00h17m31s
 65536:22                                      65535:12       yes
-------------------------------------------------------------------------------
```

The two prefixes advertised by PE-2 are properly learned by PE-1 through two different next hops. Now, use each one with a different pseudowire and make sure that the active and standby pseudowires follow different paths in the network.

Once the routes are installed in the MS-PW routing table, configure the services on PE-1 and PE-2.

```
*A:PE-1# configure service
        pw-template 1 create
            controlword
        exit
        epipe 2 customer 1 create
            description "ms-pw epipe with bgp - using 2 prefixes"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
            exit
            sap 1/1/4:2 create
            exit
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
```

```
                saii-type2 65536:192.0.2.11:1
                taii-type2 65536:192.0.2.21:1
                no shutdown
            exit
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65536:192.0.2.22:1
                no shutdown
            exit
            no shutdown
        exit


*A:PE-2# configure service
        pw-template 1 create
            controlword
        exit
        epipe 2 customer 1 create
            description "ms-pw epipe with bgp - using 2 prefixes"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
            exit
            sap 1/1/4:2 create
            exit
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.21:1
                taii-type2 65536:192.0.2.11:1
                no shutdown
            exit
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.22:1
                taii-type2 65536:192.0.2.12:1
                no shutdown
            exit
            no shutdown
        exit
```

The following command can be executed to verify that the service and spoke-sdp-fecs are up:

```
*A:PE-1# show service id 2 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : ms-pw epipe with bgp - using 2 prefixes
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 06/03/2016 12:27:05
Last Mgmt Change  : 06/03/2016 12:27:05
Test Service      : No
Admin State       : Up                  Oper State        : Up
```

```
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 2
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

------------------------------------------------------------------------------
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:1/1/4:2                               q-tag        1518    1518    Up   Up
sdp:17405:4294967292 SB(192.0.2.4)        MS-PW        0       1556    Up   Up
sdp:17406:4294967293 SB(192.0.2.3)        MS-PW        0       1556    Up   Up
==============================================================================
*A:PE-1#
```

The sdp-binding identifiers and sdp identifiers are automatically generated by the system.

Use vccv-trace to check that the spoke-sdp-fecs for the active and standby pseudowires follow different and disjoint paths:

```
*A:PE-1# oam vccv-trace spoke-sdp-fec 21

VCCV-TRACE  with 120 bytes of MPLS payload
1  192.0.2.3  rtt=1.75ms rc=8(DSRtrMatchLabel)
2  192.0.2.5  rtt=5.38ms rc=8(DSRtrMatchLabel)
3  192.0.2.2  rtt=8.34ms rc=3(EgressRtr)

*A:PE-1# oam vccv-trace spoke-sdp-fec 22

VCCV-TRACE  with 120 bytes of MPLS payload
1  192.0.2.4  rtt=1.80ms rc=8(DSRtrMatchLabel)
2  192.0.2.6  rtt=5.48ms rc=8(DSRtrMatchLabel)
3  192.0.2.2  rtt=7.83ms rc=3(EgressRtr)
```

MS-PW using Static Routing

In this sub-section, Epipe 3 will be configured between PE-1 and PE-2, where TLDP will use static-routes in the MS-PW routing table to signal the MS-PW.

On PE-1 and PE-2 (only) we will configure the prefixes used for setting up Epipe 3. Those could be the same as used for Epipe 2, however we will use different ones in this example. The **no advertise-bgp** parameter is required now. The static routes for each remote prefix are also configured. Since we will also have pseudowire redundancy for Epipe 3, two prefixes with static-routes pointing at different next-hops will be used:

```
*A:PE-1# configure service pw-routing

            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.13 create
            exit
            local-prefix 65536:192.0.2.14 create
            exit
            static-route 65536:192.0.2.23:192.0.2.3
            static-route 65536:192.0.2.24:192.0.2.4

*A:PE-2# configure service pw-routing

            spe-address 65536:192.0.2.2
            local-prefix 65536:192.0.2.23 create
            exit
            local-prefix 65536:192.0.2.24 create
            exit
            static-route 65536:192.0.2.13:192.0.2.5
            static-route 65536:192.0.2.14:192.0.2.6
```

Static-routes are also required at all S-PEs along the path (keeping the path diversity for the prefixes as well) and for both directions:

```
*A:PE-3# configure service pw-routing
            spe-address 65536:192.0.2.3
            static-route 65536:192.0.2.13:192.0.2.1
            static-route 65536:192.0.2.23:192.0.2.5

*A:PE-4# configure service pw-routing

            spe-address 65536:192.0.2.4
            static-route 65536:192.0.2.14:192.0.2.1
            static-route 65536:192.0.2.24:192.0.2.6
```

Finally, once the MS-PW routing tables are properly populated, the services can be configured and brought up:

```
*A:PE-1# configure service
        pw-template 1 create
            controlword
        exit
        epipe 3 customer 1 create
            description "ms-pw epipe with static routes"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:3 create
            exit
            spoke-sdp-fec 31 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.13:31
                taii-type2 65536:192.0.2.23:31
                no shutdown
```

```
            exit
            spoke-sdp-fec 32 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.14:32
                taii-type2 65536:192.0.2.24:32
                no shutdown
            exit
            no shutdown
        exit

*A:PE-2# configure service
        pw-template 1 create
            controlword
        exit
        epipe 3 customer 1 create
            description "ms-pw epipe with static routes"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:3 create
            exit
            spoke-sdp-fec 31 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.23:31
                taii-type2 65536:192.0.2.13:31
                no shutdown
            exit
            spoke-sdp-fec 32 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.24:32
                taii-type2 65536:192.0.2.14:32
                no shutdown
            exit
            no shutdown
        exit
```

Check the status and path of the spoke-sdp-fecs with the proper show commands and oam vccv-trace/ping commands (see previous sub-section).

MS-PW using Explicit Paths

In this sub-section, Epipe 4 will be configured between PE-1 and PE-2, where TLDP will use explicit paths to signal the MS-PW, overriding the information given by the MS-PW routing table. Although this mode requires the specific configuration of the hops, one by one, the configuration is only done on the T-PEs, as opposed to the static-routes where all the S-PEs must be configured with static routes (a mixed of static-routes and BGP routes can coexist). The local-prefixes shown for Epipe 3 will be re-used here for Epipe 4.

Now path-1 and path-2 will be configured hop by hop, using diverse paths. All the S-PE nodes as well as the terminating T-PE must be included in the path.

```
*A:PE-1# configure service
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.13 create
            exit
            local-prefix 65536:192.0.2.14 create
            exit
            path "path-1" create
                hop 1 192.0.2.3
                hop 2 192.0.2.5
                hop 3 192.0.2.2
                no shutdown
            exit
            path "path-2" create
                hop 1 192.0.2.4
                hop 2 192.0.2.6
                hop 3 192.0.2.2
                no shutdown
            exit
        exit


*A:PE-2# configure service
        pw-routing
            spe-address 65536:192.0.2.2
            local-prefix 65536:192.0.2.23 create
            exit
            local-prefix 65536:192.0.2.24 create
            exit
            path "path-1" create
                hop 1 192.0.2.5
                hop 2 192.0.2.3
                hop 3 192.0.2.1
                no shutdown
            exit
            path "path-2" create
                hop 1 192.0.2.6
                hop 2 192.0.2.4
                hop 3 192.0.2.1
                no shutdown
            exit
        exit
```

Those paths must be specified when configuring the Epipe:

```
*A:PE-1# configure service
        epipe 4 customer 1 create
            description "ms-pw epipe with explicit paths"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:4 create
            exit
            spoke-sdp-fec 41 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                saii-type2 65536:192.0.2.13:41
                taii-type2 65536:192.0.2.23:41
                path "path-1"
                no shutdown
            exit
            spoke-sdp-fec 42 fec 129 aii-type 2 create endpoint CORE
                saii-type2 65536:192.0.2.14:42
                taii-type2 65536:192.0.2.24:42
                path "path-2"
                no shutdown
            exit
            no shutdown
        exit

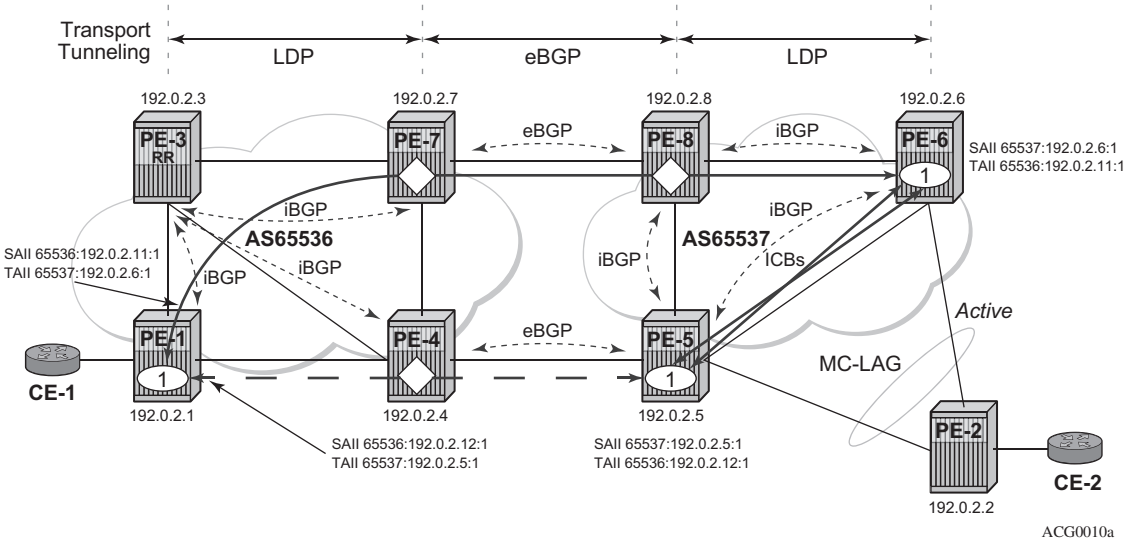

*A:PE-2# configure service
        epipe 4 customer 1 create
            description "ms-pw epipe with explicit paths"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
                revert-time 10
            exit
            sap 1/1/4:4 create
            exit
            spoke-sdp-fec 41 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                saii-type2 65536:192.0.2.23:41
                taii-type2 65536:192.0.2.13:41
                path "path-1"
                no shutdown
            exit
            spoke-sdp-fec 42 fec 129 aii-type 2 create endpoint CORE
                saii-type2 65536:192.0.2.24:42
                taii-type2 65536:192.0.2.14:42
                path "path-2"
                no shutdown
            exit
            no shutdown
        exit
```

Verify the status and path of the spoke-sdp-fecs with the proper **show** commands and **oam vccv-trace/ping** commands (see previous sub-section).

Inter-AS MS-PW Routing

This configuration example for an inter-AS scenario uses BGP tunnels between ASBRs and BGP as the MS-PW routing mechanism. Figure 302 shows the example topology used in this section.

Figure 302 **Inter-AS MS-PW Example Topology**



In this example, only one Epipe is configured (Epipe 1, using MS-PW BGP routing). The T-PEs are PE-1, PE-5 and PE-6. PE-7, PE-8 and PE-4 are S-PEs.

A/S pseudowire redundancy together with MC-LAG at one end will be used, as depicted in the diagram. Inter-chassis backup (ICB) spoke SDPs between PE-5 and PE-6 are required in order to forward the in-flight packets while MC-LAG and A/S pseudowire are converging, in case of network failures. Those ICBs will also be signaled following the MS-PW routing procedures.

The example setup in Figure 302 is pre-configured with the following settings:

- There are two AS (65536 and 65537) which are connected by two ASBR pairs (PE-7/PE-4 and PE-8/PE-5) running eBGP between them. These eBGP sessions will be used to exchange ipv4-labels (to set up the transport BGP-LBL tunnel, according to the RFC 3107) and MS-PW NLRIs.

- Within AS65536, PE-3 is used as an RR to reflect the MS-PW routes. In AS65537 there is a full mesh of iBGP sessions to distribute the MS-PW routes.
- IS-IS is used within each AS.
- LDP is used as a transport MPLS signaling protocol within each AS and a BGP tunnel will be used between the ASBRs (MS-PW routing supports LDP or BGP tunnels as transport).
- A redundant MC-LAG access to PE-6 and PE-5 is configured.

The next section will go through the configuration required to set up a redundant Epipe between CE-1 and CE-2, by combining A/S pseudowire in the network and MC-LAG at the access.

MS-PW using BGP Routing

Epipe 1 will be configured at the end of this section, including the active and redundant pseudowires from PE-1 to PE-5/PE-6, as well as the required ICBs and SAPs at the access.

As discussed, the first step is the provisioning of the pw-routing context. Again, the spe-address must be provisioned in all T-PEs and S-PEs whereas prefixes are mandatory only on the T-PEs involved in the service. The following shows the prefixes configured on PE-1, PE-5 and PE-6. Two prefixes are needed in PE-1 in order to make sure that active and standby pseudowires follow disjoint paths.

```
*A:PE-1# configure service pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.11 create
                advertise-bgp route-distinguisher 65536:11 community 65535:11
            exit
            local-prefix 65536:192.0.2.12 create
                advertise-bgp route-distinguisher 65536:12 community 65535:12
            exit


*A:PE-5# configure service pw-routing
            spe-address 65537:192.0.2.5
            local-prefix 65537:192.0.2.5 create
                advertise-bgp route-distinguisher 65537:5 community 65535:5
            exit


*A:PE-6# configure service pw-routing
            spe-address 65537:192.0.2.6
            local-prefix 65537:192.0.2.6 create
                advertise-bgp route-distinguisher 65537:6 community 65535:6
            exit
```

Once the spe-addresses and prefixes have been provisioned, BGP must be configured accordingly. A simple BGP export-policy is used to export all the local MS-PW prefixes. The configuration on PE-1 is as follows:

```
*A:PE-1# configure router
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit

*A:PE-1# configure router
        autonomous-system 65536
        bgp
            min-route-advertisement 1
            rapid-withdrawal
            group "intra-AS"
                family ms-pw
                export "export_ms-pw"
                peer-as 65536
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
    exit
```

The configuration on PE-6 is as follows:

```
*A:PE-6# configure router
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit

*A:PE-1# configure router
        autonomous-system 65537
        bgp
            min-route-advertisement 1
            enable-peer-tracking
```

```
            rapid-withdrawal
            group "intra-AS"
                family ms-pw
                export "export_ms-pw"
                peer-as 65537
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.8
                exit
            exit
            no shutdown
        exit
    exit
```

At the ASBR, the BGP policies are a little more complex since the following tasks must be accomplished:

- ASBR IPv4 system addresses must be exported to the peer ASBR to establish the RFC 3107 BGP tunnel between ASBRs.
- BGP export policies must be used so that MS-PW NLRI exchange can be controlled and attributes like MED (towards the remote AS) and/or local-preference (towards the local AS) can be modified.
- Finally, BGP import policies must also be used to modify the MS-PW route NH (next-hops) since the TLDP next signaling hop must match a peer TLDP system address.

The prefixes 65536:192.0.2.11 and 65537:192.0.2.6 must be preferred in the PE-7/PE-8 pair whereas the prefixes 65536:192.0.2.12 and 65537:192.0.2.5 must be preferred in the PE-4/PE-5 pair, so that the pseudowires are established as depicted in Figure 302. The preference can be propagated by using the BGP MED (use the local preference (LP) within the AS (LP is not relevant to eBGP)). The following CLI excerpt shows an example of how to modify MED and LP, as well as changing the NH with an import policy. The configuration on ASBR PE-4 is as follows:

```
*A:PE-4# configure router
        policy-options
            begin
            prefix-list "system"
                prefix 192.0.2.4/32 exact
            exit
            community "65535:5" members "65535:5"
            community "65535:6" members "65535:6"
            community "65535:11" members "65535:11"
            community "65535:12" members "65535:12"
            policy-statement "ASBR to ASBR"
                entry 10
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
```

```
            action accept
                origin igp
                metric set 50
            exit
        exit
        entry 20
            from
                protocol bgp
                community "65535:11"
                family ms-pw
            exit
            action accept
                origin igp
                metric set 100
            exit
        exit
    exit
    policy-statement "ASBR to region"
        entry 10
            from
                protocol bgp
                community "65535:5"
                family ms-pw
            exit
            action accept
                origin igp
                local-preference 150
                next-hop-self
            exit
        exit
        entry 20
            from
                protocol bgp
                community "65535:6"
                family ms-pw
            exit
            action accept
                origin igp
                next-hop-self
            exit
        exit
    exit
    policy-statement "export_ipv4_system"
        entry 10
            from
                prefix-list "system"
            exit
            action accept
                origin igp
            exit
        exit
    exit
    policy-statement "import ms-pw NH change"
        entry 10
            from
                protocol bgp
                family ms-pw
            exit
            action accept
```

```
                        next-hop 192.0.2.5
                    exit
                exit
            exit
            commit
        exit

*A:PE-4# configure router
        autonomous-system 65536
        bgp
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            group "inter-AS"
                family ipv4 ms-pw
                import "import ms-pw NH change"
                export "export_ipv4_system" "ASBR to ASBR"
                local-as 65536
                peer-as 65537
                neighbor 192.168.45.2
                    advertise-label ipv4
                exit
            exit
            group "intra-AS"
                family ms-pw
                export "ASBR to region"
                peer-as 65536
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
    exit
```

The configuration on ASBR PE-7 is as follows:

```
*A:PE-7# configure router
        policy-options
            begin
            prefix-list "system"
                prefix 192.0.2.7/32 exact
            exit
            community "65535:5" members "65535:5"
            community "65535:6" members "65535:6"
            community "65535:11" members "65535:11"
            community "65535:12" members "65535:12"
            policy-statement "ASBR to ASBR"
                entry 10
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        metric set 50
                    exit
                exit
```

```
        entry 20
            from
                protocol bgp
                community "65535:12"
                family ms-pw
            exit
            action accept
                origin igp
                metric set 100
            exit
        exit
    exit
    policy-statement "ASBR to region"
        entry 10
            from
                protocol bgp
                community "65535:6"
                family ms-pw
            exit
            action accept
                origin igp
                local-preference 150
                next-hop-self
            exit
        exit
        entry 20
            from
                protocol bgp
                community "65535:5"
                family ms-pw
            exit
            action accept
                origin igp
                next-hop-self
            exit
        exit
    exit
    policy-statement "export_ipv4_system"
        entry 10
            from
                prefix-list "system"
            exit
            action accept
                origin igp
            exit
        exit
    exit
    policy-statement "import ms-pw NH change"
        entry 10
            from
                protocol bgp
                family ms-pw
            exit
            action accept
                next-hop 192.0.2.8
            exit
        exit
    exit
    commit
```

```
        exit


*A:PE-7# configure router
       autonomous-system 65536
       bgp
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            group "inter-AS"
                family ipv4 ms-pw
                import "import ms-pw NH change"
                export "export_ipv4_system" "ASBR to ASBR"
                local-as 65536
                peer-as 65537
                neighbor 192.168.78.2
                    advertise-label ipv4
                exit
            exit
            group "intra-AS"
                family ms-pw
                export "ASBR to region"
                peer-as 65536
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
    exit
```

PE-5 and PE-8 have similar configurations to the ones shown. However, PE-5 is a T-PE as well as an ASBR, therefore a local MS-PW prefix must be exported as opposed to only remote prefixes (that is, some export entries for the local MS-PW routes will not contain **protocol bgp** in the matching criteria).

After BGP is properly configured and the updates get exchanged, the RIBs are populated and the prefixes uploaded onto the MS-PW routing table as shown for PE-1

```
*A:PE-1# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
```

```
---snip---
u*>i  65537:192.0.2.5         65537:5
      192.0.2.4               65537:192.0.2.5:0/64
      65537
u*>i  65537:192.0.2.6         65537:6
      192.0.2.7               65537:192.0.2.6:0/64
      65537
-------------------------------------------------------------------------------
Routes : 4


*A:PE-1# show service pw-routing route-table
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
-------------------------------------------------------------------------------
65536:192.0.2.11:0/64                         192.0.2.1      local  30d22h25m
 0:0                                           0:0            yes
65536:192.0.2.11:0/64                         192.0.2.1      local  30d22h24m
 65536:11                                      65535:11       yes
65536:192.0.2.12:0/64                         192.0.2.1      local  30d22h19m
 0:0                                           0:0            yes
65536:192.0.2.12:0/64                         192.0.2.1      local  30d22h19m
 65536:12                                      65535:12       yes
65537:192.0.2.5:0/64                          192.0.2.4      bgp    02h43m25s
 65537:5                                       65535:5        yes
65537:192.0.2.6:0/64                          192.0.2.7      bgp    02h45m49s
 65537:6                                       65535:6        yes
-------------------------------------------------------------------------------
Entries found: 6
===============================================================================
*A:PE-1#
```

For PE-6:

```
*A:PE-6# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.6        AS:65537       Local AS:65537
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
u*>i  65536:192.0.2.11        65536:11
      192.0.2.8               65536:192.0.2.11:0/64
      65536
u*>i  65536:192.0.2.12        65536:12
```

```
      192.0.2.5               65536:192.0.2.12:0/64
      65536
u*>i  65537:192.0.2.5         65537:5
      192.0.2.5               65537:192.0.2.5:0/64
      No As-Path
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-6#


*A:PE-6# show service pw-routing route-table
===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                          Next-Hop       Owner  Age
 Route-Distinguisher                           Community      Best
-------------------------------------------------------------------------------
65536:192.0.2.11:0/64                         192.0.2.8      bgp    02h52m08s
 65536:11                                      65535:11       yes
65536:192.0.2.12:0/64                         192.0.2.5      bgp    02h38m51s
 65536:12                                      65535:12       yes
65537:192.0.2.5:0/64                          192.0.2.5      bgp    02h38m51s
 65537:5                                       65535:5        yes
65537:192.0.2.6:0/64                          192.0.2.6      local  28d02h16m
 0:0                                           0:0            yes
65537:192.0.2.6:0/64                          192.0.2.6      local  28d02h16m
 65537:6                                       65535:6        yes
-------------------------------------------------------------------------------
Entries found: 5
===============================================================================
*A:PE-6#
```

As can be seen in the show commands, the two PE-1 prefixes are learned on PE-5 and PE-6 through different and disjoint paths, and the PE-5 and PE-6 prefixes are learned by PE-1 through two different and disjoint paths.

After configuring the PW routing context and configuring BGP, the last step is the service configuration on the three T-PEs, as follows. TLDP sessions must have been previously and explicitly configured between the T-PEs and S-PEs (i.e., between PE-1 and PE-4/7, between PE-4 and PE-5, PE-7 and PE-8 and between PE-6 and PE-5/8).

```
*A:PE-1# configure router ldp
            targeted-session
                peer 192.0.2.4
                exit
                peer 192.0.2.7
                exit
            exit


*A:PE-1# configure service
        pw-template 1 create
            controlword
```

```
            exit
            epipe 1 customer 1 create
                description "ms-pw epipe with bgp, inter-AS, MC-LAG redundancy"
                endpoint "CORE" create
                    description "end-point for epipe A/S PW redundancy"
                exit
                sap 1/1/4:1 create
                exit
                spoke-sdp-fec 11 fec 129 aii-type 2 create endpoint CORE
                    precedence primary
                    pw-template-bind 1
                    saii-type2 65536:192.0.2.11:1
                    taii-type2 65537:192.0.2.6:1
                    no shutdown
                exit
                spoke-sdp-fec 12 fec 129 aii-type 2 create endpoint CORE
                    pw-template-bind 1
                    saii-type2 65536:192.0.2.12:1
                    taii-type2 65537:192.0.2.5:1
                    no shutdown
                exit
                no shutdown
            exit


        *A:PE-5# configure service
            pw-template 1 create
                controlword
            exit
            epipe 1 customer 1 create
                description "ms-pw epipe with bgp, inter-AS, MC-LAG redundancy"
                endpoint "CORE" create
                    description "end-point for epipe A/S PW redundancy"
                exit
                endpoint "ACCESS" create
                exit
                sap lag-1:1 endpoint "ACCESS" create
                exit
                spoke-sdp-fec 11 fec 129 aii-type 2 create endpoint CORE
                    pw-template-bind 1
                    saii-type2 65537:192.0.2.5:1
                    taii-type2 65536:192.0.2.12:1
                    no shutdown
                exit
                spoke-sdp-fec 12 fec 129 aii-type 2 create endpoint CORE icb
                    pw-template-bind 1
                    saii-type2 65537:192.0.2.5:2
                    taii-type2 65537:192.0.2.6:2
                    no shutdown
                exit
                spoke-sdp-fec 13 fec 129 aii-type 2 create endpoint ACCESS icb
                    pw-template-bind 1
                    saii-type2 65537:192.0.2.5:3
                    taii-type2 65537:192.0.2.6:3
                    no shutdown
                exit
                no shutdown
            exit
```

```
*A:PE-6# configure service
        pw-template 1 create
            controlword
        exit
        epipe 1 customer 1 create
            description "ms-pw epipe with bgp, inter-AS, MC-LAG redundancy"
            endpoint "CORE" create
                description "end-point for epipe A/S PW redundancy"
            exit
            endpoint "ACCESS" create
            exit
            sap lag-1:1 endpoint "ACCESS" create
            exit
            spoke-sdp-fec 11 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65537:192.0.2.6:1
                taii-type2 65536:192.0.2.11:1
                no shutdown
            exit
            spoke-sdp-fec 12 fec 129 aii-type 2 create endpoint CORE icb
                pw-template-bind 1
                saii-type2 65537:192.0.2.6:3
                taii-type2 65537:192.0.2.5:3
                no shutdown
            exit
            spoke-sdp-fec 13 fec 129 aii-type 2 create endpoint ACCESS icb
                pw-template-bind 1
                saii-type2 65537:192.0.2.6:2
                taii-type2 65537:192.0.2.5:2
                no shutdown
            exit
            no shutdown
        exit
```

The following show commands can be executed to check the status of the Epipe 1 and the pseudowire status signaling received:

```
*A:PE-1# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : ms-pw epipe with bgp, inter-AS, MC-LAG redundancy
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 06/02/2016 09:52:53
Last Mgmt Change  : 06/02/2016 09:56:07
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 2
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled
-------------------------------------------------------------------------------
```

```
Service Access & Destination Points
------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:1/1/4:1                               q-tag        1518    1518    Up   Up
sdp:17406:4294967294 SB(192.0.2.7)        MS-PW        0       9190    Up   Up
sdp:17407:4294967295 SB(192.0.2.4)        MS-PW        0       9190    Up   Up
==============================================================================

*A:PE-1# show service id 1 endpoint
==============================================================================
Service 1 endpoints
==============================================================================
Endpoint name                : CORE
Description                  : end-point for epipe A/S PW redundancy
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Standby Signaling Master     : false
Standby Signaling Slave      : false
Tx Active (SDP-FEC)          : 11
Tx Active Up Time            : 0d 00:04:59
Revert Time Count Down       : N/A
Tx Active Change Count       : 2
Last Tx Active Change        : 06/02/2016 09:56:07
------------------------------------------------------------------------------
Members
------------------------------------------------------------------------------
Sdp-fec: 11 Prec:0                              Oper Status: Up
Sdp-fec: 12 Prec:4                              Oper Status: Up
==============================================================================
```

PE-5 will have the MC-LAG standby interface and as such the SAP will be operationally down and will drive the standby signaling to the remote T-PEs:

```
*A:PE-5# show service id 1 base
==============================================================================
Service Basic Information
==============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : ms-pw epipe with bgp, inter-AS, MC-LAG redundancy
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 06/02/2016 09:52:58
Last Mgmt Change  : 06/02/2016 09:52:58
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 3
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

------------------------------------------------------------------------------
Service Access & Destination Points
------------------------------------------------------------------------------
```

```
Identifier                                  Type         AdmMTU  OprMTU  Adm  Opr
------------------------------------------------------------------------------
sap:lag-1:1                                 q-tag        1518    1518    Up   Down
sdp:17406:4294967293 SB(192.0.2.6)          MS-PW        0       9190    Up   Up
sdp:17406:4294967294 SB(192.0.2.6)          MS-PW        0       9190    Up   Up
sdp:17407:4294967295 SB(192.0.2.4)          MS-PW        0       9190    Up   Up
==============================================================================

*A:PE-5# show service id 1 all | match Flag
Flags              : None
Flags              : None
Flags              : None
Flags              : PortOperDown StandByForMcProtocol
```

The following commands are useful on the S-PEs in order to find the PWs automatically created as well as the SDPs automatically used for those PWs.

```
*A:PE-7# show service sdp-using
===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                         State
-------------------------------------------------------------------------------
2147483647 17406:4294967294   MS-PW  192.0.2.1            Up    262136  262138
2147483647 17407:4294967295   MS-PW  192.0.2.8            Up    262137  262137
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
*A:PE-7#
```

As it can be seen in the preceding output, two PWs (type MS-PW) have been automatically created over two also automatically created SDPs: 17406 and 17407. SDP 17406 is built over an LDP tunnel whereas SDP 17407 runs over a BGP tunnel.

```
*A:PE-7# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop         Metric
-------------------------------------------------------------------------------
192.0.2.1/32      sdp       MPLS  17406     5        192.0.2.1       0
192.0.2.1/32      ldp       MPLS  65539     9        192.168.37.1    20
192.0.2.3/32      ldp       MPLS  65538     9        192.168.37.1    10
192.0.2.4/32      ldp       MPLS  65537     9        192.168.47.1    10
192.0.2.8/32      sdp       MPLS  17407     5        192.0.2.8       0
192.0.2.8/32      bgp       MPLS  262145    12       192.168.78.2    1000
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
*A:PE-7#
```

```
*A:PE-7# show service sdp 17406 detail | match "Active LSP"
Mixed LSP Mode       : Enabled               Active LSP Type    : LDP

*A:PE-7# show service sdp 17407 detail | match "Active LSP"
Mixed LSP Mode       : Enabled               Active LSP Type    : BGP
```

In addition to all of the recommended show commands, **vccv-ping** and **vccv-trace** are two extremely useful commands in this environment. **vccv-trace** can even help to trace the traffic going through the ICBs under failure situations.

Conclusion

Service Providers are always seeking highly scalable VLL services that can be deployed with the lowest operational cost. The SR OS supports MS-PW routing according to the draft-ietf-pwe3-dynamic-ms-pw. MS-PW routing allows the Service Provider to deploy Epipe services without having to provision services in the core of the network. In other words, MS-PW enables end-point provisioning in highly scalable seamless MPLS networks, through the use of BGP. Alternatively, static MS-PW routes or explicit paths can also be used.

The examples used in this chapter illustrate the configuration of MS-PW routing in intra-AS and inter-AS scenarios. Show and OAM commands have also been suggested so that the operator can verify and troubleshoot the MS-PW routing paths and procedures.

P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services

This chapter provides information about P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0.R4.

Point-to-multipoint (P2MP) multicast label distribution protocol (mLDP) tunnels for broadcast, unknown unicast, and multicast (BUM) traffic in EVPN-MPLS networks are supported in release 14.0.R1, and later. IGMP-snooping support for EVPN-MPLS services is supported in release 14.0.R4, and later.

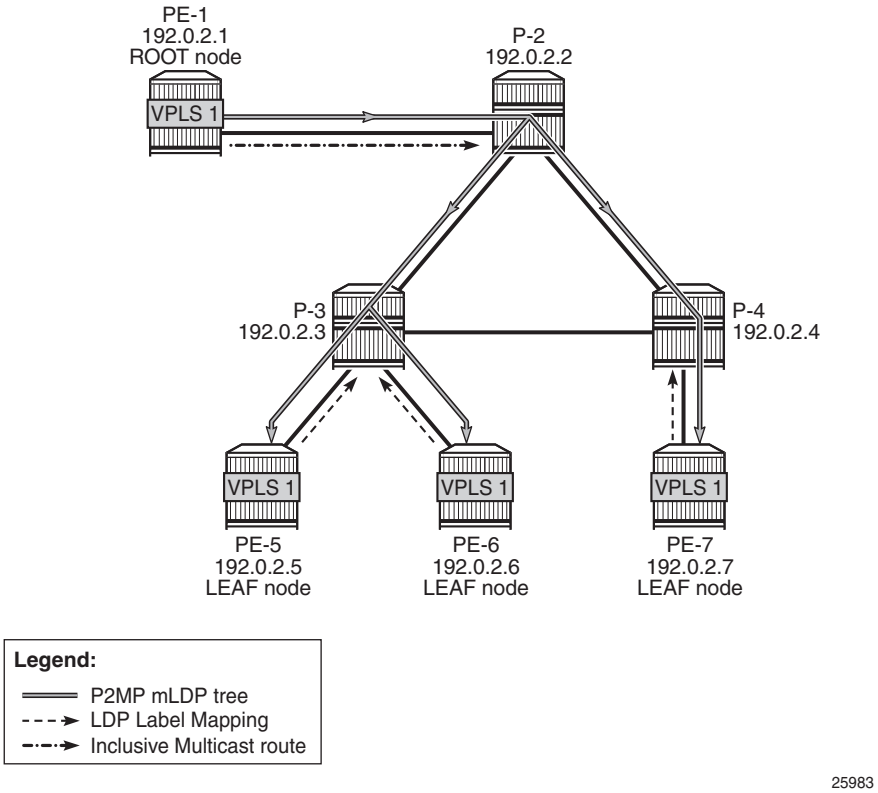
Overview

Service providers are moving their existing VPN services to EVPN. Providers using P2MP LSPs for VPLS services expect the same capabilities in EVPN. Before SR OS release 14.0.R1, only ingress-replication was supported. This works well for broadcast and unknown unicast traffic, but it is inefficient for multicast. Ingress replication does not use a multicast mechanism. Instead, the parent node makes n individual copies and unicasts each copy through an MPLS or IP tunnel to each child node.

BUM traffic will be sent from a root node to a number of leaf nodes, but leaf nodes are also allowed to send BUM traffic to root nodes. If most BUM traffic is flowing from a few root nodes to leaf nodes, it would be inefficient to promote all leaf nodes to root-and-leaf nodes due to the amount of P2MP tunnels that would need to be set up. Another solution is to use a combination of P2MP mLDP and ingress replication (IR) tunnels in the service. The root nodes send BUM traffic using P2MP tunnels while the leaf nodes use IR tunnels to send BUM traffic to the root nodes. This avoids the need to set up a P2MP tree from each leaf, while it still allows leaf nodes to send BUM traffic to the root nodes.

Figure 303 shows a multicast mLDP tree with root node PE-1 and leaf nodes PE-5, PE-6, and PE-7.

Figure 303 **P2MP mLDP Tree with Root Node PE-1 and Leaf Nodes PE-5, PE-6, and PE-7**



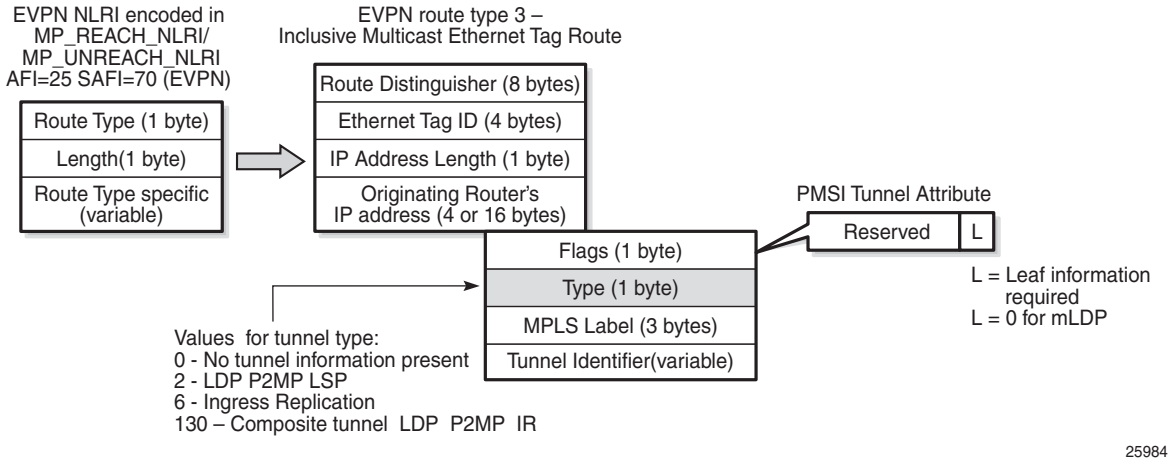
The inclusive multicast Ethernet tag (IMET) route (EVPN route type 3) sent by root node PE-1 contains the required information to set up an mLDP tree, such as the root node IP address and an opaque value. As described in chapter Multicast Label Distribution Protocol, the mLDP tree is set up from the leaf nodes toward the root.

The LDP label mapping message contains the root node address, an opaque value, and an MPLS label. The leaf nodes send an LDP label mapping message to their upstream next hop toward the root node of the tree. Each transit node that has received such LDP label mapping message will generate a new LDP label mapping message to its upstream next hop toward the root. This is repeated until the root node receives an LDP label mapping message and the multicast tree is completed.

Figure 303 shows a P2MP mLDP tree rooted in PE-1, which is optimal for multicast traffic. However, no P2MP mLDP tree needs to be rooted in PE-5, PE-6, and PE-7 for the reverse direction. These three PEs can use IR to send traffic to the root (and to the other leaf nodes if needed).

EVPN route type 3 is used for setting up the flooding tree for a specified VPLS service. EVPN route type 3 includes the provider multicast service interface (PMSI) tunnel attribute (PTA), which can have different formats depending on the tunnel type; see Figure 304.

Figure 304 **BGP-EVPN Route Type 3 with PTA**



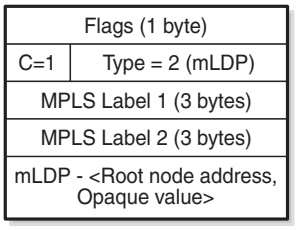
The following route values are used for EVPN-MPLS services:

- The route distinguisher (RD) is taken from the RD of the VPLS service, which can be configured in the BGP context or auto-derived from the BGP-EVPN EVI value. In this case, the RD is auto-derived from the EVI, resulting in a value of 192.0.2.1:1 for VPLS 1 on PE-1.
- The Ethernet tag ID equals 0.
- The IP address length equals 32.
- The originating router's IP address carries the IPv4 system address.

- The PTA can have different formats depending on the tunnel type enabled in the service. The SR OS EVPN-MPLS implementation supports the following tunnel types (SR OS supports different tunnel types for EVPN-VXLAN):
  - Tunnel type 2 - P2MP mLDP
    - The route is referred to as an inclusive multicast Ethernet tag point-to-multipoint (IMET-P2MP).
    - Flags: leaf not required.
    - The MPLS label is zero.
    - The tunnel identifier includes the root node address and an opaque number. This is the tunnel identifier that the leaf nodes use to join to the mLDP P2MP tree.
  - Tunnel type 6 - Ingress replication (IR)
    - The route is referred to as an inclusive multicast Ethernet tag ingress replication (IMET-IR).
    - Flags: leaf not required.
    - The MPLS label is a non-zero, downstream allocated label. This MPLS label is allocated to the service and will be the same for unicast MAC/IP routes for the same service, unless ingress-replication-bum-label is configured in the service.
    - The tunnel identifier is the tunnel endpoint and is equal to the originating IP address.
  - Tunnel type 130 - Composite tunnel: Type: C-bit (composite) + type 2 (mLDP)
    - The route is referred to as an IMET-P2MP-IR.
    - Flags: leaf not required.
    - MPLS label 1 equals zero.
    - MPLS label 2 is a non-zero, downstream allocated label (as any other IR label). The leaf nodes use the label to set up an EVPN-MPLS binding to the root and add it to the default multicast list.
    - The mLDP tunnel identifier is the root node address and an opaque number. This is the tunnel identifier that the leaf nodes will use to join the mLDP P2MP tree.

Figure 3 shows the PTA for tunnel type 130.

Figure 305 **PTA for Composite Tunnel IMET-P2MP-IR**



The composite bit C is set, indicating that the PTA identifies two tunnels: the transmit tunnel is a P2MP mLDP tunnel and the receive tunnel is an IR tunnel.

IMET-P2MP-IR Routes

The composite tunnel type is an optimized solution that combines mLDP and IR within the same EVPN service so that each root node sends BUM traffic using the P2MP tunnel while each leaf-only node sends BUM traffic to the root node using IR.

- PEs configured as **root-and-leaf** can send all BUM traffic over P2MP mLDP tunnels while they receive BUM traffic either over P2MP mLDP tunnels (from other root-and-leaf nodes) or over ingress-replication tunnels (from leaf-only nodes).
- PEs configured as **no root-and-leaf** (default setting) can use IR to send BUM traffic to root nodes and other leaf-only nodes, while receiving BUM traffic over either P2MP mLDP tunnels (from root nodes) or ingress-replication tunnels (from leaf-only nodes).

The root PEs will signal an IMET-P2MP-IR route, indicating that they intend to transmit BUM traffic using an mLDP P2MP tunnel, while they can receive traffic over an IR EVPN-MPLS binding. Composite tunnels reduce the number of P2MP mLDP tunnels that the PE/P routers in the EVI need to handle, because no full mesh of P2MP tunnels among all the PEs in the EVI is required. This is important (in terms of scaling) in services where there are just a pair of root nodes sending BUM in P2MP tunnels and hundreds of leaf nodes that only need to send BUM traffic to the root nodes using IR tunnels.

Configuration

Initial Configuration

The PE and P nodes have the following initial configuration:

- The ports between the routers are configured as network ports and have router interfaces configured.
- IS-IS is enabled on all the router interfaces.
- LDP is enabled on all the router interfaces.
- BGP is enabled on all PEs with route reflector (RR) P-2. The BGP configuration on RR P-2 is as follows:

```
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            min-route-advertisement 1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.6
                exit
                neighbor 192.0.2.7
                exit
            exit
        exit
```

Configure EVPN P2MP mLDP in VPLS Service

On the root node PE-1, VPLS 1 is configured as follows:

```
configure
    service
        vpls 1 customer 1 create
            bgp
```

```
            exit
            bgp-evpn
                ingress-repl-inc-mcast-advertisement       # default setting
                evi 1
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            sap 1/1/3 create
            exit
            no shutdown
```

The configuration options in the BGP-EVPN context of the VPLS are as follows:

```
*A:PE-1# configure service vpls 1 bgp-evpn
  - bgp-evpn
  - no bgp-evpn
 [no] accept-ivpls-e* - Configure to accept non-zero ethernet-
tag MAC routes and process for CMAC flushing
 [no] cfm-mac-advert* - Enable/
disable the advertisement of MEP, MIP, and VMEP MAC addresses over the BGP EVPN
 [no] evi             - EVPN Identifier
 [no] incl-mcast-ori* - Configure original IP address
 [no] ingress-repl-i* - Configure BGP EVPN IMET-IR route advertisement
 [no] ip-route-adver* - Configure BGP EVPN IP Route Advertisement
 [no] mac-advertisem* - Configure BGP EVPN MAC Advertisement
      mac-duplication + Configure BGP EVPN MAC Duplication
      mpls            + Configure BGP EVPN mpls
 [no] unknown-mac-ro* - Configure BGP EVPN Unknown MAC Route
      vxlan           + Configure BGP EVPN vxlan
```

By default, the advertisement of the inclusive multicast route with IR is enabled (**ingress-repl-inc-mcast-advertisement**). However, if it is disabled, the router will not send the IMET-IR or IMET-P2MP-IR routes, regardless of the service being enabled for BGP EVPN-MPLS or BGP EVPN-VXLAN.

For information about the other parameters in the BGP-EVPN context of the VPLS, see chapters EVPN for VXLAN Tunnels (Layer 2) and EVPN for MPLS Tunnels.

The configuration options in the **provider-tunnel inclusive** context are as follows:

```
*A:PE-1# configure service vpls 1 provider-tunnel inclusive
  - inclusive
 [no] data-delay-int* - Configure data delay interval
```

```
 [no] mldp            - Enable/Disable MLDP
 [no] owner           - Configure provider-tunnel owner
 [no] root-and-leaf   - Configure LSP node type
 [no] rsvp            + Configure RSVP parameters
 [no] shutdown        - Administratively enable/disable the service
```

- The **data-delay-interval** is configured in seconds in the range from 3 to 180 seconds. A node configured as root-and-leaf will send all BUM packets (data plane and control plane: ARP, CCMs, and so on) to its provider tunnel after the delay-data-interval has expired. This timer keeps the provider tunnel operationally down until its expiration, and, during that time, the router can use the EVPN-MPLS destinations typically used for IR.
- mLDP is enabled by adding the keyword **mldp** and enabling the provider tunnel (no shutdown).
- The owner must be **bgp-evpn-mpls** if MPLS is enabled in the EVPN.

```
*A:PE-1# configure service vpls 1 provider-tunnel inclusive owner
  - no owner
  - owner {bgp-ad|bgp-vpls|bgp-evpn-mpls}
```

Only one of the three possible owner protocols will support the provider tunnel in the service and needs to be set before the provider tunnel can be enabled. By default, no owner is configured. The following error is raised when a user wants to enable the provider tunnel without an owner:

```
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# no shutdown
INFO: SVCMGR #6732 No owner configured for provider-tunnel
```

After the provider tunnel has an owner and is enabled, the owner can only be changed when the provider tunnel is disabled.

```
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# owner bgp-vpls
INFO: SVCMGR #6721 Provider tunnel is not shutdown
```

After the owner is set, the corresponding protocol is checked to see if it is enabled in the service configuration.

```
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# shutdown
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# owner bgp-vpls
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# no shutdown
MINOR: SVCMGR #6730 provider-tunnel not allowed - bgp-vpls not enabled
```

- If **ingress-repl-inc-mcast-advertisement** is enabled and the PE is root-and-leaf, the router will send an IMET-P2MP-IR route; if no root-and-leaf (default) is configured, the router will send an IMET-IR route. However, if **ingress-repl-inc-mcast-advertisement** is disabled and the PE is root-and-leaf, the router will only send IMET-P2MP routes. Leaf-only nodes will not send any IMET routes at all in case no IR multicast advertisement is allowed.

Root-and-leaf nodes will only send BUM traffic to the P2MP tunnel as long as it is active. If the P2MP tunnel goes operationally down, it will start sending BUM traffic to IR tunnels (EVPN-MPLS destinations shown in the **show service id 1 evpn-mpls** command).

- If a provider tunnel is configured on a node, the router can join P2MP trees as a leaf, by generating an LDP label mapping message including the corresponding P2MP mLDP FEC. If no provider tunnel is configured, the node will not join P2MP mLDP trees, and can only use IR for BUM.
- If one node is configured as root, all other nodes must be configured with provider tunnels; otherwise, they will not receive BUM traffic sent on P2MP tunnels. The configuration of leaf-only node PE-5 is as follows, the main difference with the configuration for the root being the **no root-and-leaf** (default setting):

```
configure
    service
        vpls 1 customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    mldp
                    no shutdown
                exit
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
```

As described, the tunnel types for BUM traffic are controlled by **ingress-repl-inc-mcast-advertisement** and the provider-tunnel context (**root-and-leaf**). The IMET route sending behavior is summarized in Table 23.

Table 23 **IMET routes and Tunnel Types advertised based on the configuration**

IMET route set	Root + Leaf PE	Leaf-only	No provider-tunnel
IR-mcast advertisement	Composite P2MP + IR	IR	IR
No IR-mcast advertisement	P2MP	-	-

Information about the provider tunnel can be retrieved as follows:

```
*A:PE-1# show service id 1 provider-tunnel
===============================================================================
Service Provider Tunnel Information
===============================================================================
Type               : inclusive          Root and Leaf      : enabled
Admin State        : inService          Data Delay Intvl   : 15 secs
PMSI Type          : ldp                LSP Template       :
Remain Delay Intvl : 0 secs             LSP Name used      : 8193
PMSI Owner         : bgpEvpnMpls
===============================================================================
*A:PE-1#
```

Note: The same IMET-P2MP route cannot be imported by two services at the same time. If two VPLS services (where a provider tunnel is enabled) have the same import route-target, only one service will join the mLDP tree (whichever comes first).

EVPN P2MP mLDP Operation

After the root and leaf nodes are configured as shown, the root node sends BGP EVPN composite IMET-P2MP-IR routes, as follows:

```
1 2016/07/13 12:26:10.08 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 100
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-Incl-mcast Len: 17 RD: 192.0.2.1:1, tag: 0,
                              orig_addr len: 32, orig_addr: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 25 PMSI:
        Tunnel-type Composite LDP P2MP IR (130)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label1 Ag  0
        MPLS Label2 IR 4194176
        Root-Node 192.0.2.1, LSP-ID 0x2001
"
```

The PTA for tunnel type 130 (composite tunnel) has two MPLS labels, of which MPLS label 1 equals zero. MPLS label 2 is used by the downstream nodes to set up the EVPN-MPLS destination to the root node and add it to the default multicast list. The actual MPLS label only uses the high-order 20 bits out of the 24 bits advertised in the MPLS label. Therefore, the value 4194176 needs to be divided by 16 to have the MPLS label: 4194176/16 = 262136. This is due to the debug message being shown before the router can parse the label field and see whether it corresponds to an MPLS label (20 bits) or a VXLAN VNI (24 bits).

The tunnel identifier field contains the root node address 192.0.2.1 and the opaque value 0x2001, which corresponds to decimal value 8193. With this tunnel identifier, the leaf nodes can join the mLDP multicast tree toward the root node by sending LDP label mapping messages that contain the root node IP address and the opaque value.



Note: When static P2MP mLDP tunnels and dynamic P2MP mLDP tunnels used by BGP-EVPN coexist on the same router, Nokia recommends that the static tunnels to use a tunnel ID less than 8193. If a tunnel ID is statically configured with a value equal to or greater than 8193, BGP-EVPN may attempt to use the same tunnel ID for services with an enabled provider tunnel and fail to set up an mLDP tunnel.

The root node PE-1 receives IMET-IR routes from all leaf nodes, as shown for the BGP update sent by leaf node PE-5 (via RR P-2):

```
2 2016/07/13 12:26:16.08 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 98
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-Incl-mcast Len: 17 RD: 192.0.2.5:1, tag: 0,
                              orig_addr len: 32, orig_addr: 192.0.2.5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 4194176
        Tunnel-Endpoint 192.0.2.5
"
```

The PTA tunnel type 6 for IR has only one MPLS label, which corresponds to the MPLS label 262136 allocated for the service. The tunnel identifier is the tunnel endpoint 192.0.2.5, which is the system address of the originating leaf node.

On leaf node PE-5, three BGP EVPN inclusive multicast routes have been learned and are used, as follows:

```
*A:PE-5# show router bgp routes evpn inclusive-mcast
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr            NextHop
      Tag
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1         192.0.2.1           192.0.2.1
      0
u*>i  192.0.2.6:1         192.0.2.6           192.0.2.6
      0
u*>i  192.0.2.7:1         192.0.2.7           192.0.2.7
      0
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-5#
```

The details of the BGP EVPN inclusive multicast route sent by root node PE-1 to leaf node PE-5 are as follows:

```
*A:PE-5# show router bgp routes evpn inclusive-mcast rd 192.0.2.1:1 detail
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : N/A
Nexthop        : 192.0.2.1
From           : 192.0.2.2
Res. Nexthop   : 192.168.35.1
Local Pref.    : 100                    Interface Name : int-PE-5-P-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
```

```
Connector      : None
Community      : target:64500:1 bgp-tunnel-encap:MPLS
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.2
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
ESI            : N/A
Tag            : 0
Originator IP  : 192.0.2.1              Route Dist.    : 192.0.2.1:1
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h48m32s
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Composite LDP P2MP IR
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label1 Ag : LABEL 0
MPLS Label2 IR : LABEL 262136
Root-Node      : 192.0.2.1              LSP-ID         : 8193
-------------------------------------------------------------------------------
---snip---
```

The MPLS label is 262136, as described. The LSP ID equals 8193, which corresponds to the hexadecimal value 0x2001 in the preceding BGP update message sent by the root node PE-1.

To set up the mLDP tree, leaf node PE-5 has generated an LDP label mapping message to the next hop router toward the root, P-3. The label mapping message includes the root address 192.0.2.1, the opaque value 8193, and MPLS label 262132, as follows:

```
*A:PE-5# show router ldp bindings active p2mp ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                 Interface
RootAddr                                Op              IngLbl    EgrLbl
EgrNH                                   EgrIf/LspId
-------------------------------------------------------------------------------
8193                                    73740
192.0.2.1                               Pop             262132      --
  --                                      --

-------------------------------------------------------------------------------
```

```
No. of Generic IPv4 P2MP Active Bindings: 1
---snip---
```

P-3 has received two label mapping messages: one from PE-5 and one from PE-6. P-3 has sent one label mapping message to its upstream next hop P-2 with label 262134, as follows:

```
*A:P-3# show router ldp bindings active p2mp ipv4 opaque-type generic
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
-------------------------------------------------------------------------------
8193                                     Unknw
192.0.2.1                                Swap            262134    262132
192.168.35.2                             1/1/3

8193                                     Unknw
192.0.2.1                                Swap            262134    262128
192.168.36.2                             1/1/4

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
*A:P-3#
```

P-2 has received two label mapping messages: one from P-3 and one from P-4. P-2 has sent a label mapping message toward the root node PE-1 with label 262136, as follows:

```
*A:P-2# show router ldp bindings active p2mp ipv4 opaque-type generic
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
-------------------------------------------------------------------------------
8193                                     Unknw
192.0.2.1                                Swap            262136    262134
```

```
192.168.23.2                                1/1/2

8193                                        Unknw
192.0.2.1                                   Swap            262136    262135
192.168.24.2                                1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
*A:P-2#
```

When the LDP label reaches the root node PE-1, the mLDP tree is complete and it can be used for BUM traffic.

The following tools command shows the provider tunnels for VPLS 1 on root and leaf nodes. On root node PE-1, there is one originating inclusive provider tunnel and there are no terminating inclusive provider tunnels, as follows:

```
*A:PE-1# tools dump service id 1 provider-tunnels
===============================================================================
VPLS 1 Inclusive Provider Tunnels Originating
===============================================================================
ipmsi (LDP)                                 P2MP-ID  Root-Addr
-------------------------------------------------------------------------------
8193                                        8193    192.0.2.1
-------------------------------------------------------------------------------
===============================================================================
VPLS 1 Inclusive Provider Tunnels Terminating
===============================================================================
ipmsi (LDP)                                 P2MP-ID  Root-Addr
-------------------------------------------------------------------------------
No Tunnels Found
-------------------------------------------------------------------------------
*A:PE-1#
```

On leaf node PE-5, no originating inclusive provider tunnels are established; only one terminating provider tunnel, as follows:

```
*A:PE-5# tools dump service id 1 provider-tunnels
===============================================================================
VPLS 1 Inclusive Provider Tunnels Originating
===============================================================================
ipmsi (LDP)                                 P2MP-ID  Root-Addr
-------------------------------------------------------------------------------
No Tunnels Found
-------------------------------------------------------------------------------
===============================================================================
VPLS 1 Inclusive Provider Tunnels Terminating
===============================================================================
ipmsi (LDP)                                 P2MP-ID  Root-Addr
-------------------------------------------------------------------------------
                                            8193    192.0.2.1
-------------------------------------------------------------------------------
*A:PE-5#
```

The inclusive provider tunnels are identified by the combination of the P2MP ID (opaque value) and the root address. These parameters are in every label mapping message and they are included in the PTA tunnel identifier for tunnel type 130 (IMET-P2MP-IR) and for tunnel type 2 (IMET-P2MP).

In VPLS 1 on root node PE-1, an SDP of type VplsPmsi is auto-created, as follows:

```
*A:PE-1# show service id 1 sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
17407:4294967290 VplsPmsi not applicable  Up      Up        None      3
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The detailed information about this SDP includes the traffic statistics: ingress/egress and forwarding/dropped, as follows:

```
*A:PE-1# show service id 1 sdp detail
===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 17407:4294967290  -(not applicable)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 17407:4294967290         Type              : VplsPmsi
Split Horiz Grp    : (Not Specified)
Etree Root Leaf Tag: Disabled                 Etree Leaf AC     : Disabled
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 9194                     Oper Path MTU     : 9194
Delivery           : MPLS
Far End            : not applicable
---snip---
PMSI Owner         : bgpEvpnMpls

Admin State        : Up                       Oper State        : Up
---snip---
Statistics           :
I. Fwd. Pkts.      : 0                        I. Dro. Pkts.     : 0
I. Fwd. Octs.      : 0                        I. Dro. Octs.     : 0
E. Fwd. Pkts.      : 15062951                 E. Fwd. Octets    : 903777580
---snip---
```

IGMP Snooping

When IGMP snooping is disabled and a multicast stream enters VPLS 1 on the root node, this stream is sent to all the leaf nodes, even if no receivers join the multicast group on the leaf nodes. In this example, a receiver connected to PE-5 joins a multicast group, but there are no receivers for any multicast group on PE-6 and PE-7. By default, IGMP is disabled and the multicast stream is flooded to all leaf PEs, as can be verified with the following monitor command on PE-6 where no receivers have joined any multicast stream:

```
*A:PE-6# monitor port 1/2/1 repeat 3
===============================================================================
Monitor statistics for Port 1/2/1
===============================================================================
                                                      Input                  Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 10 sec (Mode: Delta)
-------------------------------------------------------------------------------
Octets                                                    0               12722732
Packets                                                   0                 187099
Errors                                                    0                      0
-------------------------------------------------------------------------------
At time t = 20 sec (Mode: Delta)
-------------------------------------------------------------------------------
Octets                                                    0               12692472
Packets                                                   0                 186654
Errors                                                    0                      0
-------------------------------------------------------------------------------
---snip---
```

This implies that bandwidth is wasted, which can be prevented by enabling IGMP snooping. IGMP snooping ensures that multicast traffic will only be sent to the receivers that joined a multicast group. IGMP snooping can be enabled in VPLS 1 on all PEs, as follows:

```
configure service vpls 1 igmp-snooping no shutdown
```

A receiver connected to PE-5 has sent an IGMP report while PE-6 has no receivers that joined a multicast group. The traffic counters are monitored on the outgoing port to the (potential) receivers. On PE-5, traffic is sent to the receiver, as follows:

```
*A:PE-5# monitor port 1/2/1 repeat 3
===============================================================================
Monitor statistics for Port 1/2/1
===============================================================================
                                                      Input                  Output
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 10 sec (Mode: Delta)
```

```
-------------------------------------------------------------------------------
Octets                                                0             13053892
Packets                                               0               191969
Errors                                                0                     0
-------------------------------------------------------------------------------
---snip---
```

On PE-6, no traffic is sent to any receiver, as follows:

```
*A:PE-6# monitor port 1/2/1 repeat 3
===============================================================================
Monitor statistics for Port 1/2/1
===============================================================================
                                                    Input                Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 10 sec (Mode: Delta)
-------------------------------------------------------------------------------
Octets                                                0                     0
Packets                                               0                     0
Errors                                                0                     0
-------------------------------------------------------------------------------
---snip---
```

IGMP snooping can be enabled in EVPN-MPLS services with IR or provider-tunnel mLDP trees. When IGMP snooping is enabled on the VPLS, all the EVPN-MPLS destinations are added to the MFIB as a single router interface. IGMP queries and reports are properly forwarded to and from EVPN-MPLS destinations.

The following shows the EVPN-MPLS destinations as part of the MFIB when IGMP-snooping is enabled:

```
*A:PE-5# show service id 1 mfib
===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                         Blk
-------------------------------------------------------------------------------
*               *                     sap:1/2/1:1                 Local    Fwd
                                      eMpls:192.0.2.1:262136      Local    Fwd
                                      eMpls:192.0.2.6:262136      Local    Fwd
                                      eMpls:192.0.2.7:262136      Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
*A:PE-5#
```

Connected to SAP 1/2/1:1, PE-5 has a receiver that joined the multicast stream. EVPN-MPLS is added as a single logical IGMP snooping interface and treated as an mrouter, also on the other leaf nodes, as follows:

```
===============================================================================
```

```
IGMP Snooping Base info for service 1
===============================================================================
Admin State : Up
Querier     : 172.16.0.5 on SAP 1/2/1:1
-------------------------------------------------------------------------------
Port                      Oper MRtr Pim  Send Max   Max  Max   MVR       Num
Id                        Stat Port Port Qrys Grps  Srcs Grp   From-VPLS Grps
                                                      Srcs
-------------------------------------------------------------------------------
sap:1/2/1:1               Up   Yes  No   No   None  None None  Local     0
evpn-mpls                 Up   Yes  N/A  N/A  N/A   N/A  N/A   N/A       N/A
===============================================================================
*A:PE-5#
```

On leaf node PE-5, the receiving host connected to SAP 1/2/1:1 has IP address 172.16.0.5, as follows:

```
*A:PE-5# show service id 1 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1
===============================================================================
MRouter          Port Id                   Up Time         Expires   Version
-------------------------------------------------------------------------------
172.16.0.5       1/2/1:1                   0d 04:58:11    238s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
*A:PE-5#
```

On leaf node PE-6, SAP 1/2/1:1 has no receiving host connected, but EVPN-MPLS is always added as an mrouter, as follows:

```
*A:PE-6# show service id 1 igmp-snooping base
===============================================================================
IGMP Snooping Base info for service 1
===============================================================================
Admin State : Up
Querier     : 172.16.0.5 on evpn-mpls
-------------------------------------------------------------------------------
Port                      Oper MRtr Pim  Send Max   Max  Max   MVR       Num
Id                        Stat Port Port Qrys Grps  Srcs Grp   From-VPLS Grps
                                                      Srcs
-------------------------------------------------------------------------------
sap:1/2/1:1               Up   No   No   No   None  None None  Local     0
evpn-mpls                 Up   Yes  N/A  N/A  N/A   N/A  N/A   N/A       N/A
===============================================================================
*A:PE-6#
```

On PE-6, the only mrouter in the list is the receiving host connected to PE-5, with port ID EVPN-MPLS instead of a local SAP, as follows:

```
*A:PE-6# show service id 1 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1
```

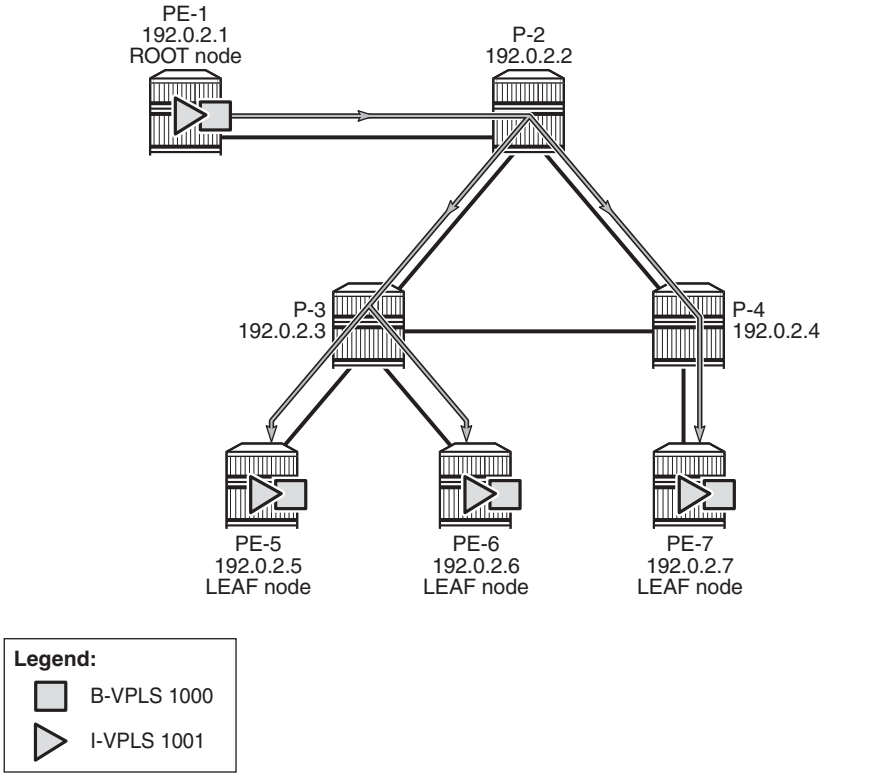
```
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
-------------------------------------------------------------------------------
172.16.0.5       evpn-mpls                  0d 04:45:40    241s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
*A:PE-6#
```

PBB-EVPN and P2MP mLDP

PBB-EVPN is described in chapter EVPN for PBB over MPLS (PBB-EVPN).

Figure 306 shows the setup for P2MP mLDP in PBB-EVPN.

Figure 306 **P2MP mLPD in PBB-EVPN**



P2MP mLDP tunnels can also be used in PBB-EVPN services. In release 14.0, the use of **provider-tunnel inclusive mldp** is only for the default multicast list; no per-ISID IMET-P2MP routes are supported.

The B-VPLS still uses multicast forwarding information bases (MFIBs) for ISIDs using IR.

If an ISID policy is configured in the B-VPLS, a range of ISIDs configured with **use-def-mcast** will use the P2MP tree, and a range of ISIDs configured with **advertise-local** will make the router advertise IMET-IR routes for the local ISIDs in the range.

PE-1 is configured as root-and-leaf. The configuration for B-VPLS and I-VPLS is as follows:

```
configure
    service
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:01
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            isid-policy
                entry 10 create
                    use-def-mcast
                    no advertise-local
                    range 1001 to 2000
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 1001 customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            sap 1/1/3 create
            exit
            no shutdown
        exit
```

In this example, ISIDs in the range from 1001 to 2000 will use the P2MP tree (use-def-mcast) and the router does not advertise the IMET-IR routes for the local ISIDs included in that range (no advertise-local). Any other local ISID will advertise an IMET-IR and will use the MFIB to forward BUM packets to the remote EVPN-MPLS bindings created by IMET-IR routes.

The configuration on the leaf nodes PE-5, PE-6, and PE-7 is similar to the one for the root node, except for the **no root-and-leaf** setting (which is default), as follows:

```
configure
    service
        vpls 1000 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:05
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            isid-policy
                entry 10 create
                    use-def-mcast
                    no advertise-local
                    range 1001 to 2000
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    mldp
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 1001 customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            sap 1/2/1:1001 create
            exit
            no shutdown
        exit
```

A VPLS-PMSI SDP is auto-created in the B-VPLS at the root node, as follows:

```
*A:PE-1# show service id 1000 sdp
===============================================================================
```

```
Services: Service Destination Points
==============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
------------------------------------------------------------------------------
17407:4294967289 VplsPmsi not applicable  Up      Up        None      3
------------------------------------------------------------------------------
Number of SDPs : 1
------------------------------------------------------------------------------
==============================================================================
*A:PE-1#
```

The default multicast list for the B-VPLS 1000 can be retrieved on root and leaf nodes, for instance for leaf node PE-5, as follows:

```
*A:PE-5# tools dump service id 1000 evpn-mpls default-multicast-list
------------------------------------------------------------------------
TEP Address                             Egr Label
                                        Transport
------------------------------------------------------------------------
192.0.2.1                               262140
                                        ldp
192.0.2.6                               262142
                                        ldp
192.0.2.7                               262139
                                        ldp
------------------------------------------------------------------------
*A:PE-5#
```

IGMP snooping can be enabled in the I-VPLS 1001 on all PEs, as follows:

```
configure service vpls 1001 igmp-snooping no shutdown
```

After IGMP snooping is enabled, the multicast stream is not flooded anymore to any receivers until they send an IGMP report for the multicast stream.

On each PE, the logical interface B-EVPN-MPLS is added as a single IGMP snooping interface and treated as an mrouter, as follows:

```
*A:PE-5# show service id 1001 igmp-snooping base
==============================================================================
IGMP Snooping Base info for service 1001
==============================================================================
Admin State : Up
Querier     : 172.16.0.55 on SAP 1/2/1:1001
------------------------------------------------------------------------------
Port                    Oper MRtr Pim  Send Max   Max  Max   MVR       Num
Id                      Stat Port Port Qrys Grps  Srcs Grp   From-VPLS Grps
                                                    Srcs
------------------------------------------------------------------------------
b-evpn-mpls             Up   Yes  N/A  N/A  N/A   N/A  N/A   N/A       N/A
sap:1/2/1:1001          Up   Yes  No   No   None  None None  Local     0
==============================================================================
*A:PE-5#
```

PE-5 has a receiver that sent an IGMP report for a multicast group in I-VPLS 1001 on SAP 1/2/1:1001 and this SAP is an mrouter port. On PE-6, there is no receiver that sent IGMP reports; therefore, the only mrouter port corresponds to the B-EVPN-MPLS logical interface, as follows:

```
*A:PE-6# show service id 1001 igmp-snooping base
===============================================================================
IGMP Snooping Base info for service 1001
===============================================================================
Admin State : Up
Querier     : 172.16.0.55 on evpn-mpls
-------------------------------------------------------------------------------
Port                      Oper MRtr Pim  Send Max   Max  Max   MVR       Num
Id                        Stat Port Port Qrys Grps  Srcs Grp   From-VPLS Grps
                                                      Srcs
-------------------------------------------------------------------------------
b-evpn-mpls               Up   Yes  N/A  N/A  N/A   N/A  N/A   N/A       N/A
sap:1/2/1:1001            Up   No   No   No   None  None None  Local     0
===============================================================================
*A:PE-6#
```

PE-5 has a local mrouter 172.16.0.55 on SAP 1/2/1:1001, as follows:

```
*A:PE-5# show service id 1001 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1001
===============================================================================
MRouter          Port Id                   Up Time        Expires   Version
-------------------------------------------------------------------------------
172.16.0.55      1/2/1:1001                0d 01:15:41    213s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
*A:PE-5#
```

On PE-6, mrouter 172.16.0.55 is not local; therefore, the EVPN-MPLS logical interface is used, as follows:

```
*A:PE-6# show service id 1001 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1001
===============================================================================
MRouter          Port Id                   Up Time        Expires   Version
-------------------------------------------------------------------------------
172.16.0.55      evpn-mpls                 0d 00:02:22    237s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
*A:PE-6#
```

Conclusion

Service providers are migrating their existing VPN services to EVPN and expect at least the same capabilities in EVPN, including the forwarding of BUM traffic. Ingress replication is a good mechanism for broadcast and unknown unicast traffic in EVPN networks, but not efficient for multicast applications. EVPN P2MP mLDP offers efficiency for multicast, using composite tunnels combining the benefits of P2MP mLDP and IR.

PBB-Epipe

This chapter provides information about Provider Backbone Bridging (PBB) — Ethernet Virtual Leased Line in an MPLS-based network which is applicable to SR OS.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SROS release 7.0.R5. The CLI in this edition corresponds to release 14.0.R2. There are no specific prerequisites.

Overview

The draft-ietf-l2vpn-pbb-vpls-pe-model-00*, Extensions to VPLS PE model for Provider Backbone Bridging,* describes the PBB-VPLS model supported by SR OS. This model expands the VPLS PE model to support PBB as defined by the IEEE 802.1ah.

The PBB model is organized around a B-component (backbone instance) and an I-component (customer instance). In Nokia's implementation of the PBB model, the use of an Epipe as I-component is allowed for point-to-point services. Multiple I-VPLS and Epipe services can be all mapped to the same B-VPLS (backbone VPLS instance).

The use of Epipe scales the E-Line services as no MAC switching, learning or replication is required in order to deliver the point-to-point service. All packets ingressing the customer SAP are PBB-encapsulated and unicasted through the B-VPLS tunnel using the backbone destination MAC of the remote PBB PE. All the packets egressing the B-VPLS destined for the Epipe are PBB de-encapsulated and forwarded to the customer SAP.

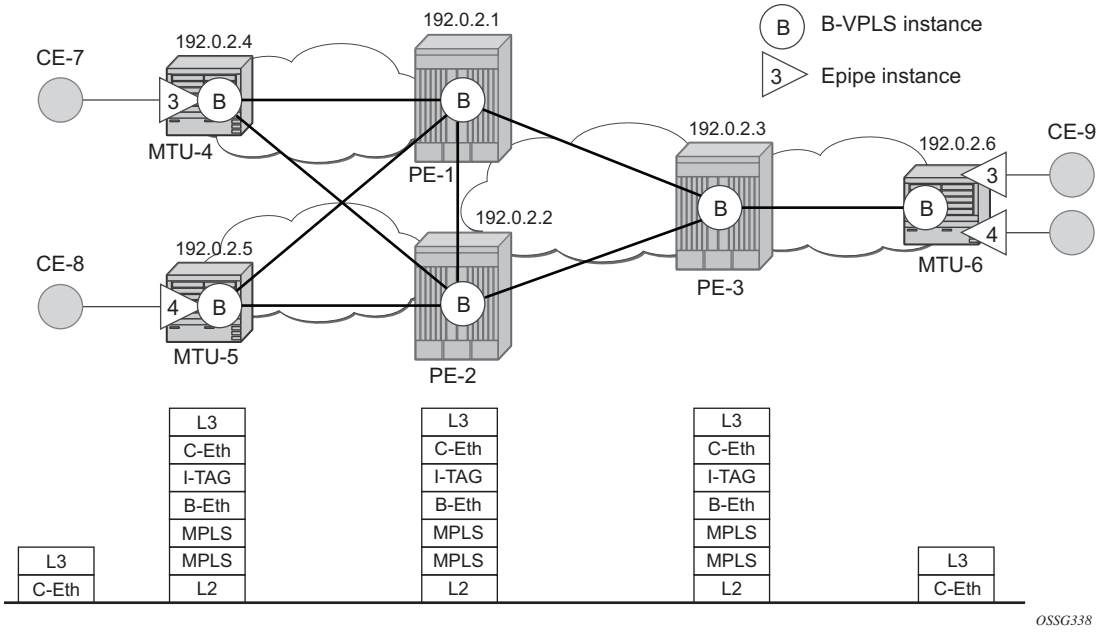
Some use cases for PBB-Epipe are:

- Get a more efficient and scalable solution for point-to-point services:
  - Up to 8K VPLS services per box are supported (including I-VPLS or B-VPLS) and using I-VPLS for point-to-point services takes VPLS resources as well as unnecessary customer MAC learning. A better solution is to connect a PBB-Epipe to a B-VPLS instance, where there is no customer MAC switching/learning.
- Take advantage of the pseudowire aggregation in the M:1 model:
  - Many Epipe services may use only a single service and set of pseudowires over the backbone.
- Have a uniform provisioning model for both point-to-point (Epipe) and multipoint (VPLS) services.
  - Using the PBB-Epipe, the core MPLS/pseudowire infrastructure does not need to be modified: the new Epipe inherits the existing pseudowire and MPLS structure already configured on the B-VPLS and there is no need for configuring new tunnels or pseudowire switching instances at the core.

Knowledge of the PBB-VPLS architecture and functionality on the service router family is assumed throughout this section. For additional information, refer to the relevant Nokia user documentation.

The following network setup will be used throughout the rest of the chapter.

Figure 307 **Network Topology**



The setup consists of a three 7x50 SR/ESS (PE-1, PE-2 and PE-3) core and three Multi-Tenant Unit (MTU) nodes connected to the core. A backbone VPLS instance (B-VPLS 101) will be defined in all the six nodes, whereas two Epipe services will be defined as illustrated in Figure 307 (Epipe 3 in nodes MTU-4 and MTU-6, Epipe 4 in nodes MTU-5 and MTU-6). Those Epipe services will be multiplexed into the common B-VPLS 101, using the I-Service ID (ISID) field within the I-TAG as the demultiplexer field required at the egress MTU to differentiate each specific customer. I-VPLS and Epipe services can be mapped to the same B-VPLS.

The B-VPLS domain constitutes a H-VPLS network itself, with spoke SDPs from the MTUs to the core PE layer. Active/standby (A/S) spoke SDPs can be used from the MTUs to the PEs (like in the MTU-4 and MTU-5 cases) or single non-redundant spoke SDPs (like MTU-6).

The protocol stack being used along the path between the CEs is represented in Figure 307.

Configuration

This section describes all the relevant PBB-Epipe configuration tasks for the setup shown in Figure 307. The appropriate B-VPLS and associated IP/MPLS configuration is out of the scope of this document. In this particular example the following protocols will be configured beforehand in the core:

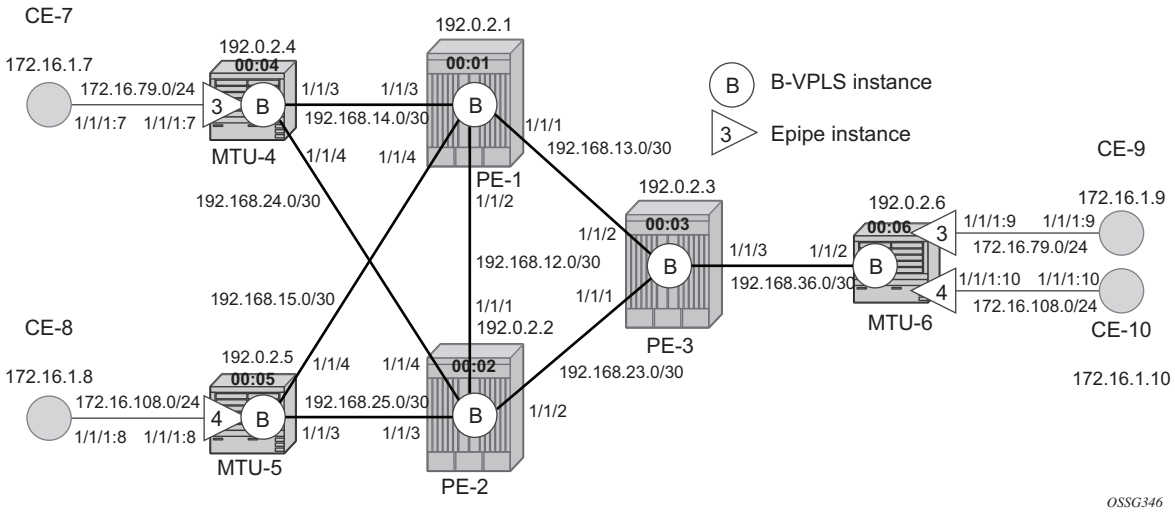
- ISIS-TE as IGP with all the interfaces being level-2. Alternatively, OSPF could have been used.
- RSVP-TE as the MPLS protocol to signal the transport tunnels.
- LSPs between core PEs will be fast re-route protected (facility bypass tunnels) whereas LSP tunnels between MTUs and PEs will not be protected.
- The protection between MTU-4, MTU-5 and PE-1, PE-2 will be based on the A/S pseudowire protection configured in the B-VPLS.
- BGP is configured for auto-discovery, BGP-AD (Layer 2 VPN family), since FEC 129 will be used to establish the pseudowires between PEs in the core (FEC 128 between MTU and PE nodes).

Once the IP/MPLS infrastructure is up and running, the service configuration tasks described in the following sections can be implemented.

PBB Epipe Service Configuration

In this particular example, the Epipes 3 and 4 are using the B-VPLS 101 in the core. The same B-VPLS which is multiplexing the Epipe services into a common service provider infrastructure can also be used to connect the I-VPLS instances existing in the network for multipoint services.

Figure 308 **Setup Detailed View**



B-VPLS and PBB Configuration

First, configure the B-VPLS instance that will carry the PBB traffic. There is no specific requirement on the B-VPLS to support Epipes. The following shows the B-VPLS configuration on MTU-4 and PE-1.

On MTU-4

```
configure
    service
        vpls 101 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:04:04:04:04:04
            exit
            endpoint "core" create
                no suppress-standby-signaling
            exit
            spoke-sdp 41:101 endpoint "core" create
                precedence primary
            exit
            spoke-sdp 42:101 endpoint "core" create
```

```
            exit
            no shutdown
        exit
```

On PE-1

```
configure
    service
        pw-template 1 use-provisioned-sdp create
            split-horizon-group "CORE"
            exit
        exit
        vpls 101 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:01:01:01:01:01
            exit
            bgp
                route-target export target:65000:101 import target:65000:101
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 65000:101
                no shutdown
            exit
            spoke-sdp 14:101 create
            exit
            spoke-sdp 15:101 create
            exit
            no shutdown
        exit
```

The relevant B-VPLS commands are in **bold**.

The keyword **b-vpls** is given at creation time and therefore it cannot be added to an existing regular VPLS instance. Besides the **b-vpls** keyword, the B-VPLS is a regular VPLS instance in terms of configuration, with the following exceptions:

- The B-VPLS service MTU must be at least 18 bytes greater than the Epipe MTU of the multiplexed instances. In this example, the I-VPLS instances will have the default service MTU (1514 bytes) hence any MTU equal or greater than 1532 bytes must be configured. In this particular example, a MTU of 2000 bytes is configured in the B-VPLS instance throughout the network.
- The source B-MAC is the MAC that will be used as a source when the PBB traffic is originated from that node. It is possible to configure a source B-MAC per B-VPLS instance (if there are more than one B-VPLS) or a common source B-MAC that will be shared by all the B-VPLS instances in the node. A common B-MAC is configured as follows:

```
*A:MTU-4# configure service pbb source-bmac 00:04:04:04:04:04
*A:MTU-5# configure service pbb source-bmac 00:05:05:05:05:05
```

```
*A:MTU-6# configure service pbb source-bmac 00:06:06:06:06:06
```

The following considerations will be taken into account when configuring the B-VPLS:

- B-VPLS SAPs:
  - Ethernet DOT1Q and NULL encapsulations are supported.
  - Default SAP types are blocked in the CLI for the B-VPLS SAP.
- B-VPLS SDPs:
  - For MPLS, both mesh and spoke SDPs with split horizon groups are supported.
  - Similar to regular pseudowire, the outgoing PBB frame on an SDP (for example, Bpseudowire) contains a BVID Qtag only if the pseudowire type is Ethernet VLAN (vc-type=vlan). If the pseudowire type is Ethernet (vc-type=ether), the BVID qtag is stripped before the frame goes out.
  - BGP-AD is supported in the B-VPLS, therefore, spoke SDPs in the B-VPLS can be signaled using FEC 128 or FEC 129. In this example, BGP-AD and FEC 129 are used. A split-horizon group has been configured to emulate the behavior of mesh SDPs in the core.
- While Multiple MAC Registration Protocol (MMRP) is useful to optimize the flooding in the B-VPLS domain and build a flooding tree on a per I-VPLS basis, it does not have any effect for Epipes since the destination B-MAC used for Epipes is always the destination B-MAC configured in the Epipe and never the group B-MAC corresponding to the ISID.
- If a local Epipe instance is associated with the B-VPLS, local frames originated or terminated on local Epipe(s) are PBB encapsulated or de-encapsulated using the PBB Etype provisioned under the related port or SDP component.

By default, the PBB Etype is 0x88e7 (which is the standard one defined in the 802.1ah, indicating that there is an I-TAG in the payload) but this PBB Etype can be changed if required due to interoperability reasons. This is the way to change it at port and/or SDP level:

```
A:MTU-4# configure port 1/1/3 ethernet pbb-etype
  - pbb-etype <0x0600..0xffff>
  - no pbb-etype

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex

A:MTU-4# configure service sdp 41 pbb-etype
  - no pbb-etype [<0x0600..0xffff>]
  - pbb-etype <0x0600..0xffff>

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex
```

The following commands are useful to check the actual PBB etype.

```
A:MTU-4# show service sdp 41 detail | match PBB
Bw BookingFactor     : 100                   PBB Etype          : 0x88e7
A:MTU-4#

A:MTU-4# show port 1/1/3 | match PBB
PBB Ethertype      : 0x88e7
A:MTU-4#
```

Before the next step, the Epipe configuration, the operator can optionally configure MAC names under the PBB context. MAC names will simplify the Epipe provisioning later on and in case of any change on the remote node MAC address, only one configuration modification is required as opposed as one change per affected Epipe (potentially thousands of Epipes which are terminated onto the same remote node). The MAC names are configured in the service PBB CLI context:

```
*A:MTU-4# configure service pbb mac-name
  - mac-name <name> <ieee-address>
  - no mac-name <name>

 <name>               : 32 char max
 <ieee-address>       : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx

configure
    service
        pbb
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
```

It is not required to configure a node with its own MAC address, so on MTU-4, the line defining the mac-name MTU-4 can be omitted.

Epipe Configuration

Once the common B-VPLS is configured, the next step is the provisioning of the customer Epipe instances. For PBB-Epipes, the I-component or Epipe is composed of an I-SAP and a PBB tunnel endpoint which points to the backbone destination MAC address (B-DA).

The following outputs show the relevant CLI configuration for the two Epipe instances represented in Setup Detailed View. The Epipe instances are configured on the MTU devices, whereas the core PEs are kept as customer-unaware nodes.

Epipes 3 and 4 are configured on MTU-6 as follows :

```
configure
    service
        epipe 3 customer 1 create
            description "pbb epipe number 3"
            pbb
                tunnel 101 backbone-dest-mac "MTU-4" isid 3
            exit
            sap 1/1/1:9 create
            exit
            no shutdown
        exit
        epipe 4 customer 1 create
            description "pbb epipe number 4"
            pbb
                tunnel 101 backbone-dest-mac "MTU-5" isid 4
            exit
            sap 1/1/1:10 create
            exit
            no shutdown
        exit
```

The following shows the relevant configuration on MTU-4 and MTU-5.

On MTU-4:

```
configure
    service
        epipe 3 customer 1 create
            description "pbb epipe number 3"
            pbb
                tunnel 101 backbone-dest-mac "MTU-6" isid 3
            exit
            sap 1/1/1:7 create
            exit
            no shutdown
        exit
```

On MTU-5:

```
configure
    service
        epipe 4 customer 1 create
            description "pbb epipe number 4"
            pbb
                tunnel 101 backbone-dest-mac "MTU-6" isid 4
            exit
            sap 1/1/1:8 create
            exit
            no shutdown
        exit
```

All Ethernet SAPs supported by a regular Epipe are also supported in the PBB Epipe. Spoke SDPs are not supported in PBB-Epipes, for example, no spoke SDP is allowed when PBB tunnels are configured on the Epipe.

The PBB tunnel links the SAP configured to the B-VPLS 101 existing in the core. The following parameters are accepted in the PBB tunnel configuration:

```
*A:MTU-5# configure service epipe 4 pbb tunnel

  - no tunnel
  - tunnel <service-id> backbone-dest-mac <mac-name> isid <ISID>
  - tunnel <service-id> backbone-dest-mac <ieee-address> isid <ISID>

 <service-id>         : [1..2148007978]|<svc-name:64 char max>
 <mac-name>           : 32 char max
 <ieee-address>       : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx
 <ISID>               : [0..16777215]
```

Where:

- The service-id matches the B-VPLS ID.
- The **backbone-dest-mac** can be given by a MAC name (as in this configuration example) or the MAC itself. It is recommended to use MAC names, as explained in the previous section.
- The ISID must be specified.

Flood Avoidance in PBB-Epipes

As already discussed in the previous section, when provisioning a PBB Epipe, the remote **backbone-dest-mac** must be explicitly configured on the PBB tunnel so that the ingress PBB node can build the 802.1ah encapsulation.

If the configured remote backbone-destination-mac is not known in the local FDB, the Epipe customer frames will be 802.1ah encapsulated and flooded into the B-VPLS until the MAC is learned. As previously discussed, MMRP does not help to minimize the flooding because the PBB Epipes always use the configured **backbone-destination-mac** for flooding traffic as opposed to the group B-MAC derived from the ISID.

Flooding could be indefinably prolonged in the following cases:

- Configuration mistake of the **backbone-destination-mac**. The service will not work but the operator will not detect the mistake since the customer traffic is not dropped at the source node. Every single frame is turned into an unknown unicast PBB frame and hence flooded into the B-VPLS domain.

- Change the **backbone-smac** in the remote PE B-VPLS instance.
- There is only unidirectional traffic in the Epipe service. In this case, the backbone-dest-mac will never be learned in the local SFIB and the frames will always be flooded into the B-VPLS domain.
- The remote node owning the **backbone-destination-mac** simply goes down.

In any of those cases, the operator can easily check whether the PBB Epipe is flooding into the B-VPLS domain, just by looking at the flood flag in the following command output:

```
*A:MTU-4# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : pbb epipe number 3
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/19/2016 09:11:55
Last Mgmt Change  : 05/19/2016 09:11:55
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 0
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:7                              q-tag        1518    1518    Up   Up

-------------------------------------------------------------------------------
PBB Tunnel Point
-------------------------------------------------------------------------------
B-vpls     Backbone-dest-MAC Isid     AdmMTU OperState Flood Oper-dest-MAC
-------------------------------------------------------------------------------
101        MTU-6             3        2000   Up        Yes   00:06:06:06:06:06
-------------------------------------------------------------------------------
Last Status Change: 05/19/2016 09:11:55
Last Mgmt Change  : 05/19/2016 09:11:55
===============================================================================
*A:MTU-4#
```

In this particular example, the PBB Epipe 3 is flooding into the B-VPLS 101, as the flood flag indicates. The operator can also confirm that the operational destination B-MAC for the pbb-tunnel, MTU-6, has not been learned in the B-VPLS FDB:

```
*A:MTU-4# show service id 101 fdb pbb
```

```
===============================================================================
Forwarding Database, b-Vpls Service 101
===============================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
A:MTU-4#
```

In small B-VPLS environments (up to 20 B-VPLSs, each with 10 MC-LAGs), it is possible to configure the PBB V-VPLS MAC notification mechanism to send notification messages at regular intervals (using the renotify parameter), rather than being only event-driven. This can avoid flooding into the B-VPLS.

Flooding Cases 1 and 2 — Wrong backbone-dest-mac

Flooding cases 1 and 2 should be fixed after detecting the flooding (see previous commands) and checking the FDBs and PBB tunnel configurations.

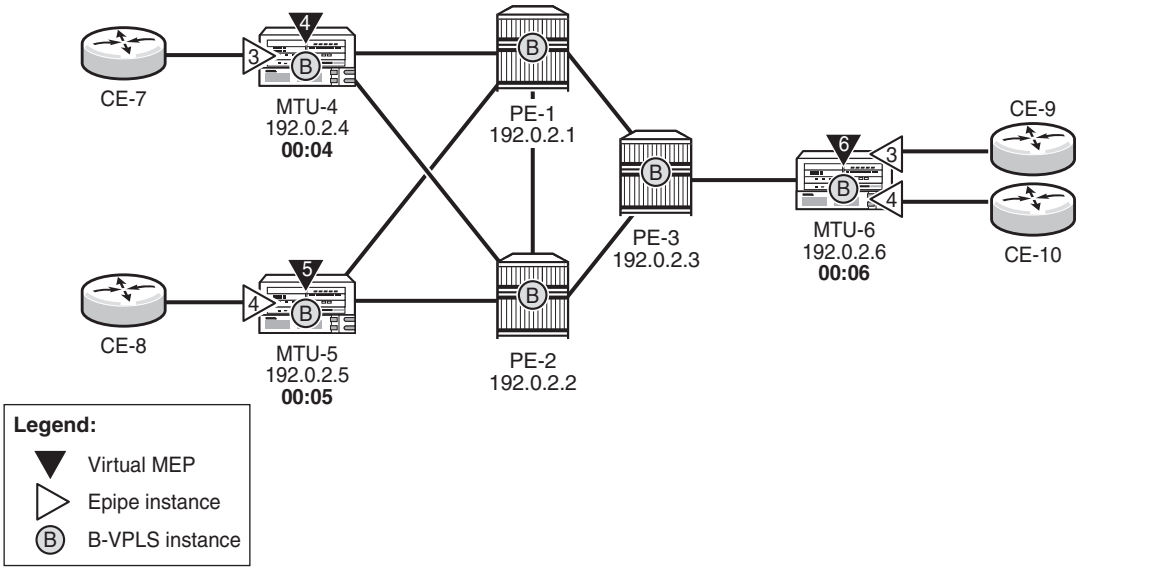
Flooding Case 3 — Unidirectional Traffic: Virtual MEP and CCM Configuration

For flooding case 3 (unidirectional traffic), Nokia recommends the use of ETH-CFM (802.1ag/Y.1731 Connectivity Fault Management) virtual Maintenance End Points (MEPs). By defining a virtual MEP per node terminating a PBB-Epipe, configuring the MEP mac-address to be the source-bmac value and activating continuity check messages (ccm), a twofold effect is achieved:

- The **pbb-tunnel backbone-destination-mac** will always be learned at the local FDB, as long as the remote virtual MEP is active and sending **cc** messages. As a result, there will not be flooding even if we have unidirectional traffic.
- An automatic proactive OAM mechanism exists to detect failures on remote nodes, which ultimately cause unnecessary flooding in the B-VPLS domain.

In the following network example, the virtual MEPs in B-VPLS 101: MEP4, MEP5 and MEP6 are configured.

Figure 309 **Virtual MEPs for Flooding Avoidance**



The following configuration example uses MTU-4. First, the general ETH-CFM configuration is made:

```
configure
    eth-cfm
        domain 1 format none level 3
            association 1 format icc-based name "B-VPLS-000101"
                bridge-identifier 101
                exit
                remote-mepid 5
                remote-mepid 6
            exit
        exit
    exit
```

Then the actual virtual MEP configuration is made:

```
configure
    service
        vpls 101
            eth-cfm
                mep 4 domain 1 association 1
                    ccm-enable
                    mac-address 00:04:04:04:04:04
                    no shutdown
                exit
            exit
        exit
```

The MAC address configured for the MEP4 matches the MAC address configured as the **source-bmac** on MTU-4, which is the **backbone-destination-mac** configured on the Epipe 3 pbb-tunnel on MTU-6. The source-bmac address on MTU-4 is 00:04:04:04:04:04, as follows:

```
configure
    service
        pbb
            source-bmac 00:04:04:04:04:04
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
        exit
```

The backbone-dest-mac configured on MTU-6 uses MAC name “MTU-4”, which corresponds to MAC address 00:04:04:04:04:04, as follows:

```
configure
    service
        pbb
            source-bmac 00:06:06:06:06:06
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
        exit
        epipe 3 customer 1 create
            description "pbb epipe number 3"
            pbb
                tunnel 101 backbone-dest-mac "MTU-4" isid 3
            exit
            sap 1/1/1:9 create
            exit
            no shutdown
        exit
```

Once MEP4 has been configured, check that MTU-6 is receiving CC messages from MEP4 with the following command:

```
*A:MTU-6# show eth-cfm mep 6 domain 1 association 1 all-remote-mepids
===============================================================================
Eth-CFM Remote-Mep Table
===============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-------------------------------------------------------------------------------
4          True  False Absent   Absent 00:04:04:04:04:04 05/19/2016 09:14:46
5          True  False Absent   Absent 00:05:05:05:05:05 05/19/2016 09:14:46
===============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
*A:MTU-6#
```

As a result of the **CC** messages coming from MEP4, the MTU-4 MAC is permanently learned in the B-VPLS 101 FDB on node MTU-6.

```
*A:MTU-6# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : pbb epipe number 3
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/19/2016 09:13:00
Last Mgmt Change  : 05/19/2016 09:13:00
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 0
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:9                              q-tag        1518    1518    Up   Up

-------------------------------------------------------------------------------
PBB Tunnel Point
-------------------------------------------------------------------------------
B-vpls     Backbone-dest-MAC Isid     AdmMTU OperState Flood Oper-dest-MAC
-------------------------------------------------------------------------------
101        MTU-4             3        2000   Up        No    00:04:04:04:04:04
-------------------------------------------------------------------------------
Last Status Change: 05/19/2016 09:13:00
Last Mgmt Change  : 05/19/2016 09:13:00
===============================================================================
*A:MTU-6#
```

Flooding Case 4 — Remote Node Failure

If the node owner of the **backbone-dest-mac** fails or gets isolated, the node where the PBB Epipe is initiated will not detect the failure; that is, if MTU-4 fails, the Epipe 3 remote end will also fail but MTU-6 will not detect the failure and as a result of that, MTU-6 will flood the traffic to the network (flooding will occur after MTU-4 MAC is removed from the B-VPLS FDBs, due to either the B-VPLS flushing mechanisms or aging).

In order to avoid/reduce flooding in this case, the following mechanisms are recommended:

- Provision virtual MEPs in the B-VPLS instances terminating PBB Epipes, as already explained. This will guarantee there is no unknown B-MAC unicast being flooded under normal operation.
- CCM timers should be provisioned based on how long the service provider is willing to accept flooding.

```
*A:MTU-6# configure eth-cfm domain 1 association 1 ccm-interval
 - ccm-interval <interval>
 - no ccm-interval

 <interval>           : {10ms|100ms|1|10|60|600} - default 10 seconds
```

- It is possible to also provision discard-unknown in the B-VPLS on the MTUs, i.e. MTU-4, MTU-5 and MTU-6, so that flooded traffic due to the destination MAC being unknown in the B-VPLS is discarded immediately at the MTU. It is important to configure this in conjunction with the CC messages from the virtual MEPs to ensure that the remote B-MACs are learned in both directions. If for any reason the remote B-MACs are not in the MTU B-VPLS, no traffic will be forwarded at all on the PBB-Epipe.
- As soon as the MTU node recovers, it will start sending CC messages and the backbone-mac will be learned on the backbone and MTU nodes again.

```
*A:PE-1# configure service vpls 101 discard-unknown
*A:PE-2# configure service vpls 101 discard-unknown
*A:PE-3# configure service vpls 101 discard-unknown
*A:MTU-4# configure service vpls 101 discard-unknown
*A:MTU-5# configure service vpls 101 discard-unknown
*A:MTU-6# configure service vpls 101 discard-unknown
```

With the recommended configuration in place, in case MTU-4 fails, the **backbone-dest-mac** configured on the pbb-tunnel for Epipe 3 on MTU-6 will be removed from the B-VPLS 101 on all the nodes (either by MAC flush mechanisms on the B-VPLS or by aging). From that point on, traffic originated from CE-9 will be discarded at MTU-6 and won't be flooded further.

As soon as MTU-4 comes back up, MEP4 will start sending CCM and as such the MTU-4 MAC will be learned throughout the B-VPLS 101 domain and in particular in PE-1, PE-3 and MTU-6 (CCM PDUs use a multicast address). From the moment MTU-4 MAC is known on the backbone nodes and MTU-6, the traffic won't be discarded any more, but forwarded to MTU-4.

PBB-Epipe Show Commands

The following commands can help to check the PBB Epipe configuration and their related parameters.

For the B-VPLS service:

```
*A:MTU-4# show service id 101 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 101                 Vpn Id            : 0
Service Type      : b-VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/19/2016 09:09:37
Last Mgmt Change  : 05/19/2016 09:15:29
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 2000                Def. Mesh VC Id   : 101
SAP Count         : 0                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>
Oper Backbone Src : 00:04:04:04:04:04
Use SAP B-MAC     : Disabled
i-Vpls Count      : 0
Epipe Count       : 1
Use ESI B-MAC     : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:41:101 S(192.0.2.1)                 Spok         8000    8000    Up   Up
sdp:42:101 S(192.0.2.2)                 Spok         8000    8000    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:MTU-4#
```

For the Epipe service:

```
*A:MTU-4# show service id 3 base
===============================================================================
Service Basic Information
===============================================================================
```

```
Service Id        : 3                   Vpn Id            : 0
Service Type      : Epipe
Name              : (Not Specified)
Description       : pbb epipe number 3
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/19/2016 09:11:55
Last Mgmt Change  : 05/19/2016 09:11:55
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 0
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:7                              q-tag        1518    1518    Up   Up

-------------------------------------------------------------------------------
PBB Tunnel Point
-------------------------------------------------------------------------------
B-vpls     Backbone-dest-MAC Isid     AdmMTU OperState Flood Oper-dest-MAC
-------------------------------------------------------------------------------
101        MTU-6             3        2000   Up        No    00:06:06:06:06:06
-------------------------------------------------------------------------------
Last Status Change: 05/19/2016 09:11:55
Last Mgmt Change  : 05/19/2016 09:11:55
===============================================================================
*A:MTU-4#
```

The following command shows all the Epipe instances multiplexed into a particular B-VPLS and its status.

```
*A:MTU-4# show service id 101 epipe
===============================================================================
Related Epipe services for b-Vpls service 101
===============================================================================
Epipe SvcId         Oper ISID           Admin               Oper
-------------------------------------------------------------------------------
3                   3                   Up                  Up
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
*A:MTU-4#
```

To check the virtual MEP information, show the local virtual MEPs configured on the node:

```
*A:MTU-4# show eth-cfm cfm-stack-table all-virtuals
===============================================================================
```

```
CFM Stack Table Defect Legend:
R = Rdi, M = MacStatus, C = RemoteCCM, E = ErrorCCM, X = XconCCM
A = AisRx, L = CSF LOS Rx, F = CSF AIS/FDI rx, r = CSF RDI rx
===============================================================================
CFM Virtual Stack Table
===============================================================================
Service            Lvl Dir Md-index   Ma-index  MepId  Mac-address     Defect
-------------------------------------------------------------------------------
101                  3 U          1          1    4 00:04:04:04:04:04   -------
===============================================================================
*A:MTU-4#
```

The following command shows all the information related to the remote MEPs configured in the association, for example, the remote virtual MEPs configured in MTU-5 and MTU-6:

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1 all-remote-mepids
===============================================================================
Eth-CFM Remote-Mep Table
===============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-------------------------------------------------------------------------------
5          True  False Absent   Absent 00:05:05:05:05:05 05/19/2016 09:13:39
6          True  False Absent   Absent 00:06:06:06:06:06 05/19/2016 09:13:39
===============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
*A:MTU-4#
```

The following command shows the detail information and status of the local virtual MEP configured in MTU-4:

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1
===============================================================================
Eth-Cfm MEP Configuration Information
===============================================================================
Md-index          : 1                         Direction         : Up
Ma-index          : 1                         Admin             : Enabled
MepId             : 4                         CCM-Enable        : Enabled
SvcId             : 101
Description       : (Not Specified)
FngAlarmTime      : 0                         FngResetTime      : 0
FngState          : fngReset                  ControlMep        : False
LowestDefectPri   : macRemErrXcon             HighestDefect     : none
Defect Flags      : None
Mac Address       : 00:04:04:04:04:04         Collect LMM Stats : disabled
CcmLtmPriority    : 7                         CcmPaddingSize    : 0 octets
CcmTx             : 16                        CcmSequenceErr    : 0
CcmIgnoreTLVs     : (Not Specified)
Fault Propagation: disabled                   FacilityFault     : n/a
MA-CcmInterval    : 10                        MA-CcmHoldTime    : 0ms
MA-Primary-Vid    : Disabled
Eth-1Dm Threshold: 3(sec)                     MD-Level          : 3
Eth-Ais           : Disabled
Eth-Ais Tx defCCM: allDef
```

```
Eth-Tst          : Disabled
Eth-CSF          : Disabled
Redundancy:
    MC-LAG State : n/a
CcmLastFailure Frame:
    None
XconCcmFailure Frame:
    None
===============================================================================
*A:MTU-4#
```

When there is a failure on a remote Epipe node, as discussed, the source node keeps sending traffic. The 802.1ag/Y.1731 virtual MEP configured can help to detect and troubleshoot the problem. For instance, when a failure happens in MTU-6 (node goes down or the B-VPLS instance is shut down), the virtual MEP show commands will show the following information:

```
*A:MTU-6# configure service vpls 101 shutdown

*A:MTU-4# show eth-cfm mep 4 domain 1 association 1
===============================================================================
Eth-Cfm MEP Configuration Information
===============================================================================
Md-index         : 1                         Direction         : Up
Ma-index         : 1                         Admin             : Enabled
MepId            : 4                         CCM-Enable        : Enabled
SvcId            : 101
Description      : (Not Specified)
FngAlarmTime     : 0                         FngResetTime      : 0
FngState         : fngDefectReported         ControlMep        : False
LowestDefectPri  : macRemErrXcon             HighestDefect     : defRemoteCCM
Defect Flags     : bDefRDICCM bDefRemoteCCM
Mac Address      : 00:04:04:04:04:04         Collect LMM Stats : disabled
CcmLtmPriority   : 7                         CcmPaddingSize    : 0 octets
CcmTx            : 39                        CcmSequenceErr    : 0
CcmIgnoreTLVs    : (Not Specified)
Fault Propagation: disabled                  FacilityFault     : n/a
MA-CcmInterval   : 10                        MA-CcmHoldTime    : 0ms
MA-Primary-Vid   : Disabled
Eth-1Dm Threshold: 3(sec)                    MD-Level          : 3
Eth-Ais          : Disabled
Eth-Ais Tx defCCM: allDef
Eth-Tst          : Disabled
Eth-CSF          : Disabled
Redundancy:
    MC-LAG State : n/a
CcmLastFailure Frame:
    None
XconCcmFailure Frame:
    None
===============================================================================
*A:MTU-4#
```

The bDefRemoteCCMdefect flag clearly shows that there is a remote MEP in the association which has stopped sending CCMs. In order to find out which node is affected, see the following output:

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1 all-remote-mepids
===============================================================================
Eth-CFM Remote-Mep Table
===============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-------------------------------------------------------------------------------
5          True  True  Absent   Absent 00:05:05:05:05:05 05/19/2016 09:13:39
6          False False Absent   Absent 00:00:00:00:00:00 05/19/2016 09:17:58
===============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
*A:MTU-4#
```

CCMs are no longer received from virtual MEP 6 (the one defined in MTU-6) since 05/19/2016 09:17:58. This conveys which node has failed and when it failed.

Conclusion

Point-to-Point Ethernet services can use the same operational model followed by PBB VPLS for multipoint services. In other words, Epipes can be linked to the same B-VPLS domain being used by I-VPLS instances and use the existing H-VPLS network infrastructure in the core. The use of PBB Epipes reduces dramatically the number of services and pseudowires in the core and therefore allows the service provider to scale the number of ELINE services in the network.

The example used in this document shows the configuration of the PBB Epipes as well as all the related features which are required for this environment. Show commands have also been suggested so that the operator can verify and troubleshoot the service.

PBB-VPLS

This chapter provides information about provider backbone bridging (PBB) in a multi-protocol label switching (MPLS) based network.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SROS and was initially written for release 7.0.R6. The CLI in this edition is based on release 14.0.R2. In release 9.0.R1, and later, the chassis requirements to allow the use of MC-LAG to dual-home a customer site into a PBB-based service have been relaxed and no longer require the node to be in chassis mode D. However, all network ports and B-SAPs must reside on slots populated with either an IOM3-XP or IMM. All other SAPs may reside on any IOM or IMM populated slot.

Note: Although it can be used in an MPLS-based PBB network as explained in this document, the MAC notification feature for dual-homed access is normally used in native PBB networks.

Summary

The draft-ietf-l2vpn-pbb-vpls-pe-model-00, *Extensions to LDP Signaling for PBB-VPLS*, describes the PBB-VPLS model supported by SR OS. This model expands the VPLS PE model to support PBB as defined by the IEEE 802.1ah.

PBB-VPLS combines the best of the PBB and VPLS technologies to deliver the most scalable multi-point Layer 2 VPN in the market. PBB-VPLS inherits all the benefits derived from MPLS (for example, sub-50ms FRR protection, traffic engineering, no need for MSTP in the backbone) while greatly increasing the scalability of the network by providing MAC hiding, service multiplexing and pseudowire aggregation.

The SR OS PBB-VPLS implementation also includes support for:

- MMRP (Multiple MAC Registration Protocol, application within IEEE 802.1ak) for flood containment in the backbone instances, as specified in Section 6 of the draft-ietf-l2vpn-pbb-vpls-pe-model.
- Extensions to LDP signaling for PBB-VPLS, according to draft-balus-l2vpn-pbb-ldp-ext-00. These extensions will avoid network black-hole issues, as described in the Section 3 of the mentioned draft.

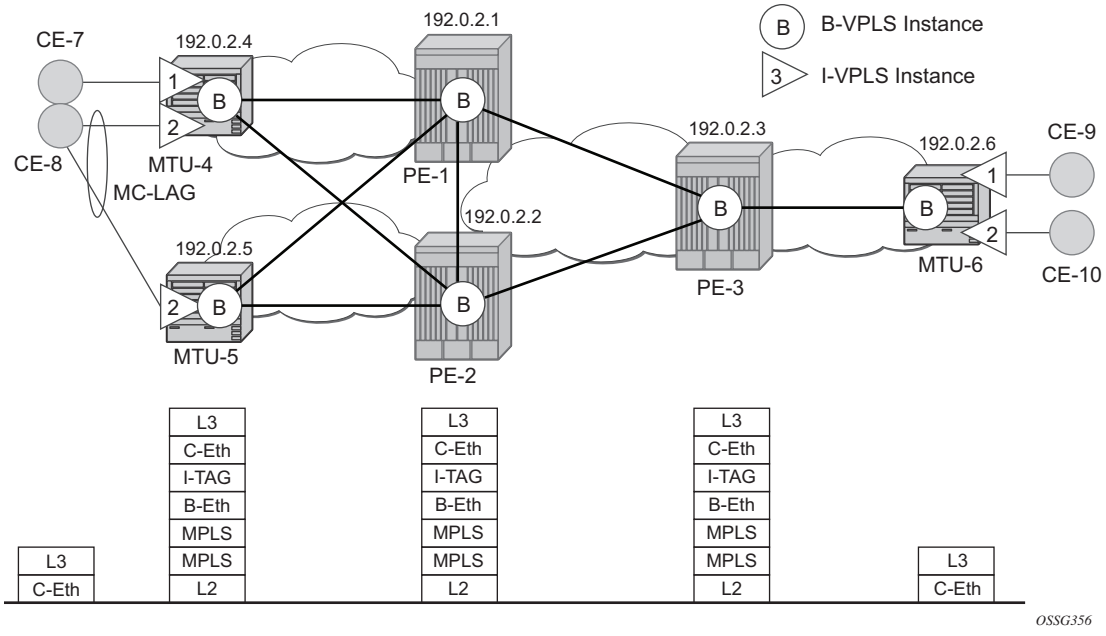
The objective of this section is to provide the required guidelines to configure and troubleshoot a PBB-VPLS network.

Knowledge of the VPLS and H-VPLS (RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signalin*g) architecture and functionality is assumed throughout this document. The most relevant concepts will be briefly explained throughout the document, taking the network setup shown in the next section as an example. For further information, see the relevant Nokia documentation.

Overview

The following network setup will be used throughout the rest of the chapter.

Figure 310 **Example Topology**



The setup consists of three core nodes (PE-1, PE-2 and PE-3) and three MTU (Multi-Tenant Unit) nodes connected to the core. A backbone VPLS instance (B-VPLS 100) will be defined in all the six nodes, whereas a few customer I-VPLS instances will be defined on the three MTU nodes.

Those I-VPLS instances will be multiplexed into the common B-VPLS, using the ISID field within the I-TAG as the demultiplexer field at the egress MTU to differentiate each specific customer.

The B-VPLS domain constitutes an H-VPLS network itself, with spoke SDPs from the MTUs to the core PE layer. Active/standby spoke SDPs can be used from the MTUs to the PEs (for example, in the MTU-4 and MTU-5 cases) or single non-redundant spoke SDPs (for example, MTU-6). CE-8 is dual-connected to the service provider network through MC-LAG.

The protocol stack being used along the path between the CEs is represented in Figure 310.

Configuration

This section describes all the relevant PBB-VPLS configuration tasks for the setup shown in Figure 310. Note that the appropriate associated IP/MPLS configuration is out of the scope of this example. In this particular example, the following protocols will be configured beforehand:

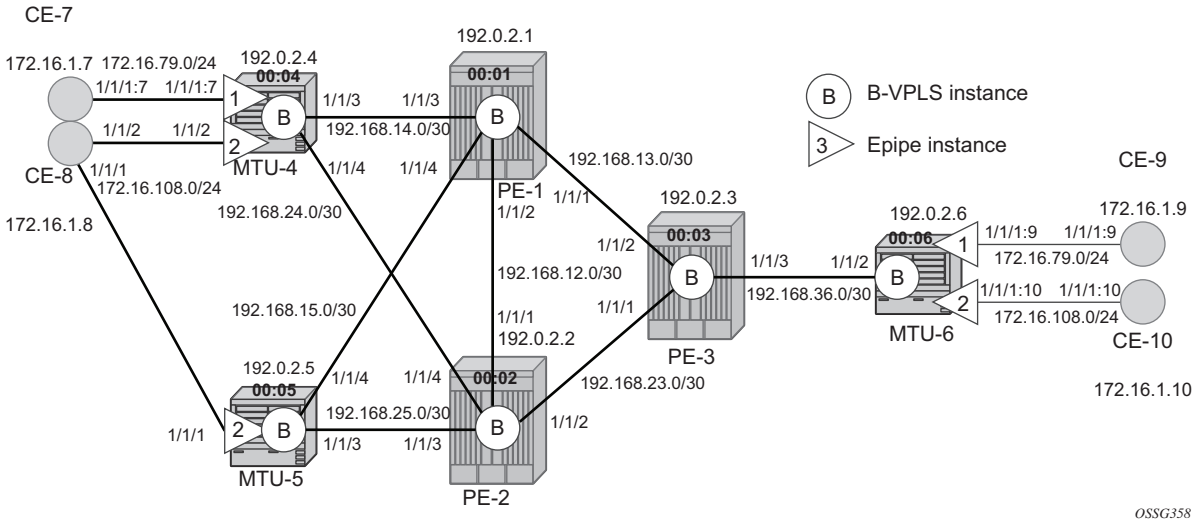
- ISIS-TE as IGP with all the interfaces being Level-2 (OSPF-TE could have been used instead).
- RSVP-TE as the MPLS protocol to signal the transport tunnels (LDP could have been used instead).
- LSPs between core PEs will be fast re-route protected (facility bypass tunnels) whereas LSP tunnels between MTUs and PEs will not be protected.
- The protection between MTU-4, MTU-5 and PE-1, PE-2 will be based on the A/S pseudowire protection configured in the B-VPLS.
- BGP is configured for auto-discovery (Layer 2-VPN family), since FEC 129 will be used for the pseudowires between PEs in the core.

Once the IP/MPLS infrastructure is up and running, the service configuration tasks described in the following sections can be implemented.

PBB-VPLS M:1 Service Configuration

This section explains the process to configure PBB-VPLS services in a M:1 fashion, **M** being the number of customer I-VPLS services multiplexed into the same B-VPLS instance (instance 100). An alternative configuration is 1:1, where each customer I-VPLS has its own B-VPLS. MTU-4 and PE-1 will be picked to show the relevant CLI configuration commands. The bold digits separated by colons **00:xx** are abbreviations for the backbone MAC addresses.

Figure 311 **MTU-4 and PE-1 Nodes as Configuration Examples**



B-VPLS Configuration

The first step is to configure the B-VPLS instance that will carry the PBB traffic. The following shows the B-VPLS configuration.

Configuration examples:

```
On MTU-4:
configure
    service
        vpls 100 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:04:04:04:04:04
            exit
            endpoint "core" create
                no suppress-standby-signaling
            exit
            spoke-sdp 41:100 endpoint "core" create
```

```
                precedence primary
            exit
            spoke-sdp 42:100 endpoint "core" create
            exit
            no shutdown
        exit

On PE-1:
configure
    service
        pw-template 1 use-provisioned-sdp create
            split-horizon-group "CORE"
            exit
        exit
        vpls 100 customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:01:01:01:01:01
            exit
            bgp
                route-target export target:65000:100 import target:65000:100
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 65000:100
                no shutdown
            exit
            spoke-sdp 14:100 create
            exit
            spoke-sdp 15:100 create
            exit
            no shutdown
        exit
```

The relevant B-VPLS commands are in **bold**.

The keyword **b-vpls** is given at creation time and therefore it cannot be added to a regular existing VPLS instance. Besides the **b-vpls** keyword, the B-VPLS is a regular VPLS instance in terms of configuration, with the following exceptions:

- The B-VPLS service MTU must be at least 18 bytes greater than the I-VPLS MTU of the multiplexed instances. In this example, the I-VPLS instances will have the default service MTU (1500 bytes); therefore, any MTU equal to or greater than 1518 bytes must be configured. In this particular example, a MTU of 2000 bytes is configured in the B-VPLS instance throughout the network.

- The source B-MAC is the MAC that will be sourced when the PBB traffic is originated from that node. A source B-MAC per B-VPLS instance can be configured (if there are more than one B-VPLS) or a common source B-MAC that will be shared by all the B-VPLS instances in the box. If no specific source B-MAC is provisioned, the system MAC address is used as the source B-MAC. When using the access multi-homing feature for native PBB, the source-bmac must be a configured one and never the chassis mac address. The way to configure a common B-MAC for all the B-VPLS instances on MTU-4 is shown below:

```
configure
    service
        pbb
            source-bmac 00:04:04:04:04:04
```

The following considerations will be taken into account when configuring the B-VPLS:

- B-VPLS SAPs:
  - Ethernet dot1q and null encapsulations are supported
  - Default SAP (:*) types are blocked in the CLI for the B-VPLS SAP
- B-VPLS SDPs:
  - For MPLS, both mesh and spoke SDPs with split horizon groups are supported.
  - Similar to regular pseudowires, the outgoing PBB frame on an SDP (for example, B-pseudowire) contains a BVID qtag only if the pseudowire type is Ethernet VLAN. If the pseudowire type is **Ethernet**, the BVID qtag is stripped before the frame goes out.
  - BGP-AD is supported in the B-VPLS; therefore, spoke SDPs in the B-VPLS can be signaled using FEC 128 or FEC 129. In this example, BGP-AD and FEC 129 are used. A split-horizon group has been configured to emulate the behavior of mesh-SDPs in the core.
- If a local I-VPLS instance is associated with the B-VPLS, “local frames” originated/terminated on local I-VPLS(s) are PBB encapsulated/de-encapsulated using the PBB etype provisioned under the related port or SDP component.

By default, the PBB etype is 0x88e7 (which is the standard one defined in 802.1ah for the I-TAG) but this PBB etype can be changed if required due to interoperability reasons. This is the way to change it at port and/or SDP level:

```
A:MTU-4# configure port 1/1/3 ethernet pbb-etype
  - pbb-etype <0x0600..0xffff>
  - no pbb-etype
```

```
 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex

A:MTU-4# configure service sdp 41 pbb-etype
  - no pbb-etype [<0x0600..0xffff>]
  - pbb-etype <0x0600..0xffff>

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex
```

The following commands are useful to check the actual PBB etype.

```
A:MTU-4# show service sdp 41 detail | match PBB
Bw BookingFactor     : 100                   PBB Etype          : 0x88e7

A:MTU-4# show port 1/1/3 | match PBB
PBB Ethertype      : 0x88e7
```

I-VPLS Configuration

Once the common B-VPLS is configured, the next step is to provision the customer I-VPLS instances. The following shows the relevant configuration on MTU-4 for the two I-VPLS instances represented in Figure 311. The I-VPLS instances are configured on the MTU devices, whereas the core PEs are customer-unaware nodes.

```
configure
    service
        vpls 1 customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap 1/1/1:7 create
            exit
            no shutdown
        exit
        vpls 2 customer 1 i-vpls create
            pbb
                backbone-vpls 100 isid 2
                exit
            exit
            sap lag-1 create
            exit
            no shutdown
        exit
```

The relevant I-VPLS commands are in **bold**.

The keyword **i-vpls** is given at creation time and therefore it cannot be added to a regular existing VPLS instance. After creating the I-VPLS instance, it has to be linked to its corresponding transport B-VPLS instance. That link is given by the **backbone-vpls** *b-vpls isid* isid command. If no ISID (20 bit customer identification in the ITAG) is specified, the system will take the VPLS instance identifier as the ISID value.

The following considerations will be taken into account when configuring the I-VPLS:

- I-VPLS SAPs:
  - SAPs can be defined on ports with any Ethernet encapsulation type (null, dot1q and qinq)
  - The I-VPLS SAPs can coexist on the same port with SAPs for other business services, for example, VLL and VPLS SAPs.
- I-VPLS SDPs:
  - GRE and MPLS SDPs are supported.
  - No mesh-SDPs are supported, only spoke SDP. Mesh-SDPs can be emulated by using split horizon groups.

Existing SAP processing rules still apply for the I-VPLS case; the SAP encapsulation definition on Ethernet ingress ports defines which VLAN tags are used to determine the service that the packet belongs to:

- Null encapsulation defined on ingress — Any VLAN tags are ignored and the packet goes to a default service for the SAP;
- Dot1q encapsulation defined on ingress — only first VLAN tag is considered;
- QinQ encapsulation defined on ingress — both VLAN tags are considered; wildcard for the inner VLAN tag is supported.
- For dot1q/qinq encapsulations, traffic encapsulated with VLAN tags for which there is no definition is discarded.
- Any VLAN tag used for service selection on the I-SAP is stripped before the PBB encapsulation is added. Appropriate VLAN tags are added at the remote PBB PE when sending the packet out on the egress SAP.

Up to 8000 VPLS instances can be defined per system. That number includes I-VPLS, B-VPLS and regular VPLS.

MMRP for Flooding Optimization

When the M:1 model is used (as in this example), any I-VPLS broadcast/multicast/unknown frame is flooded throughout the B-VPLS domain regardless of the nodes where the originating I-VPLS is defined. In other words, in our example in Figure 310, any broadcast/multicast/unknown frame coming from CE-7 would be flooded in the B domain and would reach PE-2 and MTU-5, even though that traffic only needs to go to PE-3 and MTU-6. In order to build customer-based flooding trees and optimize the flooding, MMRP (Multiple MAC Registration Protocol) must be configured on the B-VPLS.

MMRP can be enabled with its default settings just by executing a **mrp no shutdown** command on all nodes:

```
configure service vpls 100 mrp no shutdown
```

There are certain B-VPLS MRP settings that can be modified. These are the default values:

```
*A:MTU-4>config>service>vpls>mrp# info detail
----------------------------------------------
                mmrp
                    no end-station-only
                    attribute-table-size 2048
                    attribute-table-low-wmark 90
                    attribute-table-high-wmark 95
                    no flood-time
                    no shutdown
                exit
                no shutdown
----------------------------------------------
*A:MTU-4>config>service>vpls>mrp#
```

These attributes can be changed in order to control the number of MMRP attributes per B-VPLS and optimize the convergence time in case of failures in the B-VPLS:

- Controlling the number of attributes per B-VPLS

  The MMRP exchanges create one entry per attribute (group B-MAC) in the B-VPLS where MMRP protocol is running. When the first registration is received for an attribute, an MFIB entry is created for it. The *attribute-table-size* allows the user to control the number of MMRP attributes (group B-MACs) created on a per B-VPLS basis, between 1 and 2048. Based on the configured size, high and low watermarks can be set (in percentage) so that alarms can be triggered upon exceeding the watermarks. This ensures that no B-VPLS will take up all the resources from the total pool. The maximum number of attributes per B-VPLS is 2048 and 4000 can be configured globally on the system.
- Optimizing the convergence time

Assuming that MMRP is used in a certain B-VPLS, under failure conditions the time it takes for the B-VPLS forwarding to resume may depend on the data plane and control plane convergence plus the time it takes for MMRP exchanges to stabilize the flooding trees on a per ISID basis. In order to minimize the convergence time, the PBB SR OS implementation offers the selection of a mode where B-VPLS forwarding reverts for a short time to flooding so that MMRP has enough time to converge. This mode can be selected through configuration using the **flood-time** *value* command where value represents the amount of time in seconds (between 3 and 600) that flooding will be enabled. If this behavior is selected, the forwarding plane starts with B-VPLS flooding for a configurable time period, then it reverts back to the MFIB entries installed by MMRP. The following B-VPLS events initiate the switch from per I-VPLS (MMRP) MFIB entries to BVPLS flooding:

- Reception or local triggering of a TCN (Spanning Tree Topology Change Notification)
- B-SAP failure
- Failure of a B-SDP binding
- Pseudowire activation in a primary/standby H-VPLS resiliency solution
- SF/CPM switchover due to STP reconvergence

The IEEE 802.1ak standard, which defines MRP, requires the implementation of different state machines with associated timers that can be tuned. A full MRP participant maintains the following state machines:

- Registrar state machine
- Applicant state machine
- LeaveAll state machine
- PeriodicTransmission state machine

The two first state machines are maintained for each attribute in which the participant is interested, while the two latter are global to all the attributes.

The job of the registrar function is to record declarations of the attribute made by other participants on the LAN. A registrar does not send any protocol messages, as the applicant looks after the interests of all would-be participants.

The job of the applicant is twofold: first, to ensure that this participant’s declaration is correctly registered by other participants’ registrars, and next, to prompt other participants to register again after one withdraws a declaration.

The associated timers can be tuned on a per SAP/SDP basis:

```
A:MTU-4>config>service>vpls>spoke-sdp# mrp
  - mrp
```

```
 [no] join-time       - Configure timer value in 10th of seconds for sending
                        join-messages
 [no] leave-all-time  - Configure timer value in 10th of seconds for refreshing
                        all attributes
 [no] leave-time      - Configure timer value in 10th of seconds to hold
                        attribute in leave-state
 [no] mrp-policy      - Configure mrp-policy
 [no] periodic-time   - Configure timer value in 10th of seconds for
                        re-transmission of attribute declarations
 [no] periodic-timer  - Control re-transmission of attribute declarations

A:MTU-4>config>service>vpls>spoke-sdp#


A:MTU-4>config>service>vpls>spoke-sdp>mrp# info detail
----------------------------------------------
                    join-time 2
                    leave-time 30
                    leave-all-time 100
                    periodic-time 10
                    no periodic-timer
                    no mrp-policy
----------------------------------------------
A:MTU-4>config>service>vpls>spoke-sdp>mrp#
```

A brief description of the MRP SAP/SDP attributes follows:

- Join-time — This command controls the interval between transmit opportunities that are applied to the applicant state machine. An instance of this join period timer is required on a per-port, per-MRP participant basis. For additional information, refer to IEEE 802.1ak-2007 section 10.7.4.1.
- Leave-time — This command controls the period of time that the registrar state machine will wait in the leave state before transitioning to the MT state when it is removed. An instance of the timer is required for each state machine that is in the leave state. The leave period timer is set to the value leave-time when it is started. A registration is normally in “in” state where there is an MFIB entry and traffic being forwarded. When a “leave all” is performed (periodically around every 10-15 seconds per SAP/SDP binding – see leave-all-time-below), a node sends a message to its peer indicating a leave all is occurring and puts all of its registrations in leave state. The peer refreshes its registrations based on the leave all PDU it receives and sends a PDU back to the originating node with the state of all its declarations. Refer to IEEE 802.1ak-2007 section 10.7.4.2.
- Leave-all-time — This command controls the frequency with which the leaveall state machine generates leaveall PDUs. The timer is required on a per-port, per-MRP participant basis. The leaveall period timer is set to a random value, T, in the range leavealltime<T<1.5*leave-all-time when it is started. Refer to IEEE 802.1ak-2007, section 10.7.4.3.

- Periodic-time — This command controls the frequency the periodic transmission state machine generates periodic events if the periodic transmission timer is enabled. The timer is required on a per-port basis. The periodic transmission timer is set to one second when it is started.
- Periodic-timer — This command enables or disables the periodic transmission timer.

The following command shows the MRP configuration and statistics on a per SAP/SDP basis within the B-VPLS:

```
*A:MTU-4# show service id 100 all | match MRP post-lines 10
Sdp Id 41:100 MRP Information
-------------------------------------------------------------------------------
Join Time          : 0.2 secs                Leave Time        : 3.0 secs
Leave All Time     : 10.0 secs               Periodic Time     : 1.0 secs
Periodic Enabled   : false
Mrp Policy         : N/A
Rx Pdus            : 2542                    Tx Pdus           : 2963
Dropped Pdus       : 0
Rx New Event       : 0                       Rx Join-In Event  : 1986
Rx In Event        : 0                       Rx Join Empty Evt : 1533543
Rx Empty Event     : 3                       Rx Leave Event    : 126
SDP MMRP Information
-------------------------------------------------------------------------------
MAC Address       Registered        Declared
-------------------------------------------------------------------------------
01:1e:83:00:00:01 Yes               Yes

01:1e:83:00:00:02 Yes               Yes
-------------------------------------------------------------------------------
Number of MACs=2 Registered=2 Declared=2
-------------------------------------------------------------------------------
Sdp Id 42:100 MRP Information
-------------------------------------------------------------------------------
Join Time          : 0.2 secs                Leave Time        : 3.0 secs
Leave All Time     : 10.0 secs               Periodic Time     : 1.0 secs
Periodic Enabled   : false
Mrp Policy         : N/A
Rx Pdus            : 0                       Tx Pdus           : 0
Dropped Pdus       : 0
Rx New Event       : 0                       Rx Join-In Event  : 0
Rx In Event        : 0                       Rx Join Empty Evt : 0
Rx Empty Event     : 0                       Rx Leave Event    : 0
SDP MMRP Information
-------------------------------------------------------------------------------
MAC Address       Registered        Declared
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of MACs=0 Registered=0 Declared=0
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
* indicates that the corresponding row element may have been truncated.
Service MRP Information
===============================================================================
```

```
Admin State          : enabled
-------------------------------------------------------------------------------
MMRP
-------------------------------------------------------------------------------
Admin Status         : enabled              Oper Status      : up
Register Attr Cnt    : 2                    Declared Attr Cnt: 2
End-station-only     : disabled
Max Attributes       : 2048                 Attribute Count  : 2
Hi Watermark         : 95%                  Low Watermark    : 90%
Failed Registers     : 0                    Flood Time       : Off
-------------------------------------------------------------------------------
MVRP
-------------------------------------------------------------------------------
MRP SAP Table
===============================================================================
SAP                            Join      Leave     Leave All Periodic
                               Time(sec) Time(sec) Time(sec) Time(sec)
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
MRP SDP-BIND Table
===============================================================================
SDP-BIND                       Join      Leave     Leave All Periodic
                               Time(sec) Time(sec) Time(sec) Time(sec)
-------------------------------------------------------------------------------
41:100                         0.2       3.0       10.0      1.0
42:100                         0.2       3.0       10.0      1.0
===============================================================================
===============================================================================

-------------------------------------------------------------------------------
*A:MTU-4#
```

The following command is useful to check the MRP configuration and status.

```
*A:MTU-4# show service id 100 mrp
===============================================================================
Service MRP Information
===============================================================================
Admin State          : enabled
-------------------------------------------------------------------------------
MMRP
-------------------------------------------------------------------------------
Admin Status         : enabled              Oper Status      : up
Register Attr Cnt    : 2                    Declared Attr Cnt: 2
End-station-only     : disabled
Max Attributes       : 2048                 Attribute Count  : 2
Hi Watermark         : 95%                  Low Watermark    : 90%
Failed Registers     : 0                    Flood Time       : Off
-------------------------------------------------------------------------------
MVRP
-------------------------------------------------------------------------------
Admin Status         : disabled             Oper Status      : down
Max Attr             : 4095                 Failed Register  : 0
Register Attr Count  : 0                    Declared Attr    : 0
Hi Watermark         : 95%                  Low Watermark    : 90%
Hold Time            : disabled             Attr Count       : 0
-------------------------------------------------------------------------------
```

```
===============================================================================
MRP SAP Table
===============================================================================
SAP                             Join      Leave     Leave All Periodic
                                Time(sec) Time(sec) Time(sec) Time(sec)
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
MRP SDP-BIND Table
===============================================================================
SDP-BIND                        Join      Leave     Leave All Periodic
                                Time(sec) Time(sec) Time(sec) Time(sec)
-------------------------------------------------------------------------------
41:100                          0.2       3.0       10.0      1.0
42:100                          0.2       3.0       10.0      1.0
===============================================================================
===============================================================================
*A:MTU-4#
```

In the example throughout the document, as soon as MMRP is enabled, an optimized flooding tree will be built for ISID 1, since the I-VPLS 1 is only defined in MTU-4 and MTU-6, but not in MTU-5. A good way to track the flooding tree for a particular ISID is the following command:

```
*A:MTU-4# show service id 100 mmrp mac
-------------------------------------------------------------------------------
SAP/SDP                                 MAC Address       Registered  Declared
-------------------------------------------------------------------------------
sdp:41:100                              01:1e:83:00:00:01 Yes         Yes
sdp:41:100                              01:1e:83:00:00:02 Yes         Yes
-------------------------------------------------------------------------------
Number of Entries=2 SAPs=0 SDPs=2
-------------------------------------------------------------------------------
*A:MTU-4#

*A:MTU-5# show service id 100 mmrp mac
-------------------------------------------------------------------------------
SAP/SDP                                 MAC Address       Registered  Declared
-------------------------------------------------------------------------------
sdp:52:100                              01:1e:83:00:00:01 Yes         No
sdp:52:100                              01:1e:83:00:00:02 Yes         No
-------------------------------------------------------------------------------
Number of Entries=2 SAPs=0 SDPs=2
-------------------------------------------------------------------------------
*A:MTU-5#
```

The group B-MAC ending in **01** corresponds to the I-VPLS 1 whereas the one ending in **02** to the I-VPLS 2. MMRP PDUs for the two attributes are sent throughout the loop-tree topology (not over STP blocked ports or standby spoke SDPs and observing the split horizon rules). The two attributes are registered on every B-VPLS virtual port; however, the tree is only built on those ports where the attribute is also declared, and not only registered . For instance, the spoke-sdp 52:100 in MTU-5 will not be part of the ISID 1 or ISID 2 flooding trees. Neither attribute is declared since: I-VPLS 1 doesn't exist on MTU-5 and I-VPLS 2 is operationally down on MTU-5 (MC-LAG SAP is in standby state hence the I-VPLS down).

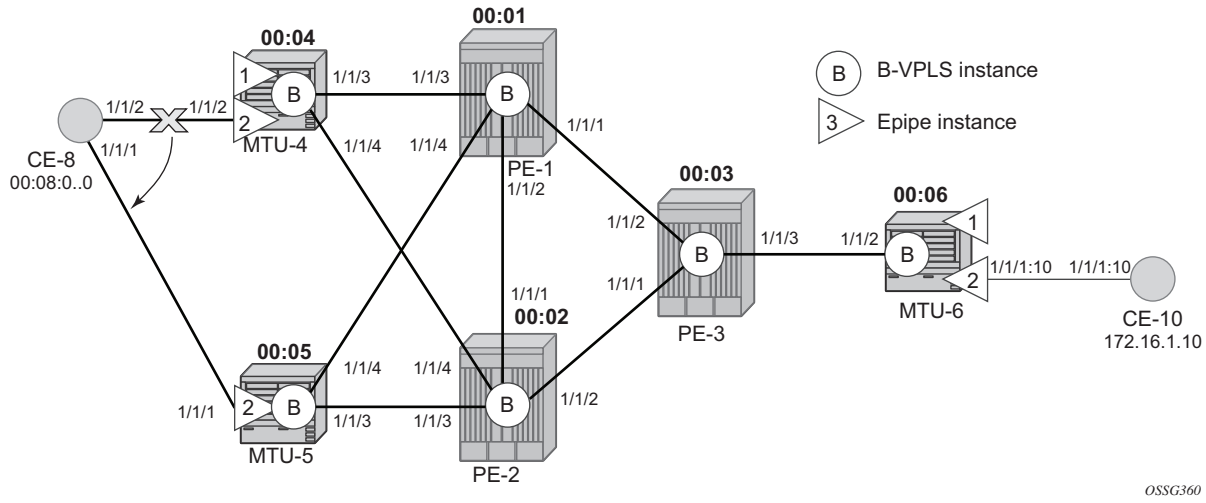
Note that as soon as a group B-MAC attribute is registered on a particular port, an MFIB entry is added for that B-MAC on that port, regardless of the declaration state for that attribute on the port. For instance, neither B-MAC is declared on MTU-5 however the two MFIB entries are created as soon as the attributes are registered:

```
A:MTU-5# show service id 100 mfib
===============================================================================
Multicast FIB, Service 100
===============================================================================
Source Address  Group Address         Sap/Sdp Id               Svc Id   Fwd/Blk
-------------------------------------------------------------------------------
*               01:1E:83:00:00:01     b-sdp:52:100             Local    Fwd
*               01:1E:83:00:00:02     b-sdp:52:100             Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
A:MTU-5#
```

MAC Flush: Avoiding black-holes

Both the I-VPLS and B-VPLS components inherit the MAC flush capabilities of a regular VPLS clearing the related C-MAC and respectively B-MAC FIBs. All types of MAC flush – flush-all-but-mine and flush-all-from-me – are supported together with the related CLI. In addition to these features, some extensions have been added so that MAC flush can be triggered on the B-VPLS based on some events happening on the I-VPLS. The following diagram shows a potential scenario where black-holes can occur if the proper configuration is not added.

Figure 312 **black-hole**



Under normal conditions the I-VPLS 2 FIB on MTU-6 shows that CE-8 MAC address is learned through B-MAC 00:04 (MTU-4's B-MAC):

```
A:MTU-6# show service id 2 fdb pbb
===============================================================================
Forwarding Database, i-Vpls Service 2
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
00:08:00:00:00:00 b-sdp:63:100          100       00:04:04:04:04:04 L/60
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/60
===============================================================================
A:MTU-6#
```

When a failure happens in the CE-8 MC-LAG active link, the link to MTU-5 takes over. However, the FIB on MTU-6 still points at MTU-4's B-MAC and that will still be the B-MAC used in the PBB encapsulation. Therefore a black-hole occurs until either bidirectional traffic is sent or the FIB aging timer expires.

The following command will solve the black-hole:

```
On MTU-4:
*A:MTU-4# configure service vpls 2 pbb send-bvpls-flush
  - send-bvpls-flush {[all-but-mine] [all-from-me]}
  - no send-bvpls-flush
 <all-but-mine>       : keyword
 <all-from-me>        : keyword

configure service vpls 2 pbb send-bvpls-flush all-from-me
```

By enabling **send-bvpls-flush all-from-me** on I-VPLS 2, a failure on the MC-LAG active link on I-VPLS 2 will trigger an LDP MAC flush **flush-all-from-me** into the B-VPLS that will flush the FIB in MTU-6 for I-VPLS 2, avoiding the black-hole. A MC-LAG failure is emulated as follows:

```
*A:MTU-4# configure lag 1 shutdown

On MTU-4:

2 2016/05/24 06:03:05.19 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address Withdraw packet (msgId 263) to 192.0.2.1:0
Protocol version = 1
MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/100 Group ID = 0 cBit = 0
"

On MTU-6:

1 2016/05/24 06:03:17.21 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address Withdraw packet (msgId 206) from 192.0.2.3:0
Protocol version = 1
MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/100 Group ID = 0 cBit = 0
Number of PBB-BMACs = 1
BMAC 1 = 00:04:04:04:04:04
Number of PBB-ISIDs = 1
ISID 1 = 2
Number of Path Vectors : 3
Path Vector(  1) = 192.0.2.4
Path Vector(  2) = 192.0.2.1
Path Vector(  3) = 192.0.2.3
```

Immediately after receiving the MAC flush, the CE-8 MAC is flushed:

```
A:MTU-6# show service id 2 fdb pbb
===============================================================================
Forwarding Database, i-Vpls Service 2
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/30
===============================================================================
A:MTU-6#
```

The CE-8 MAC is learned again, but this time linked to the B-MAC 00:05, which is the B-MAC of MTU-5:

```
*A:MTU-6# show service id 2 fdb pbb
===============================================================================
Forwarding Database, i-Vpls Service 2
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
```

```
00:08:00:00:00:00 b-sdp:63:100          100       00:05:05:05:05:05 L/0
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/120
===============================================================================
*A:MTU-6#
```

The following I-VPLS events are propagated into the B-VPLS depending on the **flush-all-but-mine** or **flush-all-from-me** keywords used in the configuration:

If the **flush-all-but-mine** keyword is configured (positive flush), the following events in the I-VPLS trigger a MAC flush into the B-VPLS:

1. TCN event in one or more of the related I-VPLS/M-VPLS.
2. Pseudowire/SDP binding activation with active/standby pseudowire (standby to active or down to up).
3. Reception of an LDP MAC withdraw flush-all-but-mine in the related I-VPLS.

If the **flush-all-from-me** keyword is configured (negative flush) the following events in the I-VPLS trigger a MAC flush into the B-VPLS:

1. MC-LAG active link failure (in our example).
2. Failure of a local SAP – requires **send-flush-on-failure** to be enabled in I-VPLS.
3. Failure of a local pseudowire/SDP binding – requires **send-flush-on-failure** to be enabled in I-VPLS.
4. Reception of an LDP MAC withdraws flush-all-from-me in the related I-VPLS.

In addition to this and regardless of what type, MAC flush has been optimized to avoid flushing in the core PEs, flushing only the C-MACs mapped to a certain B-MAC (belonging to a specific ISID FIB) and the ability to indicate to core PEs which messages should always be forwarded endpoint-to-endpoint toward all PBB PEs regardless of the propagate-mac-flush setting in B-VPLS. All of this is implemented without the need of any additional CLI commands and it is part of draft-balus-l2vpn-pbb-ldp-ext-00.

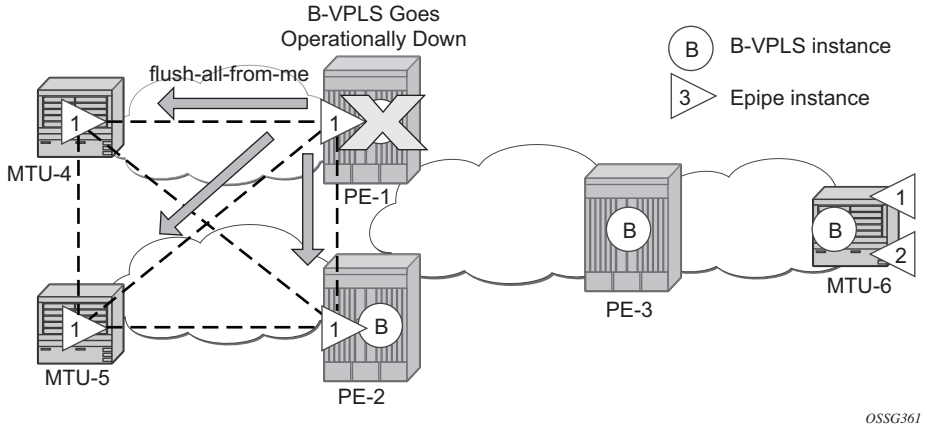
Another extension supported to avoid black-holes within this mix of I- and B-VPLS environments is the block-on-mesh-failure feature in PBB. When the VPLS mesh exists only in I-VPLS or in B-VPLS, and the block-on-mesh-failure feature is enabled, the regular VPLS behavior will apply (when all the mesh SDPs go down an LDP notification with pseudowire status bits = 0x01 - Pseudo Wire Not Forwarding – is sent over the spoke SDPs). When the active/standby pseudowire resiliency is implemented in I-VPLS such that the PBB PE performs the role of a PE-rs, the B-VPLS core replaces the pseudowire (SDP binding) mesh. The block-on-mesh-notification (LDP notification indicating pseudowire not forwarding) will be sent to the MTUs only when the related B-VPLS is operationally down. The B-VPLS core is operationally down only when all of its SAPs and SDPs are down.

The final feature that can be enabled in an I-VPLS with CLI is the send-flush-on-bvpls-failure feature.

```
*A:MTU-4# configure service vpls 2 pbb send-flush-on-bvpls-failure
  - no send-flush-on-bvpls-failure
  - send-flush-on-bvpls-failure
```

This feature is required to avoid black-holes when there is a full-mesh of pseudowires in the I-VPLS domain and the B-VPLS instance can go operationally down. The following figure shows a typical scenario where this feature is needed (normally when PBB-VPLS and multi-chassis end point are combined together).

Figure 313 **Send Flush on BVPLS Failure Example**



Access Dual-Homing and MAC Notification

Although this section is focused on PBB in a MPLS based network, Nokia PBB implementation also allows the operator to use a native Ethernet infrastructure in the PBB core. Native Ethernet tunneling can be emulated using Ethernet SAPs to interconnect the related B-VPLS instances. In those cases, there is no LDP signaling available; therefore, there is no MAC flush sent when the active link in a multi-homed access device fails.

The SR OS supports a mechanism to avoid potential black-holes in native Ethernet PBB networks. In addition to the source B-MAC associated with each B-VPLS, an additional B-MAC is associated with each MC-LAG supporting Multi-homed I-VPLS SAPs. The nodes that are in a multi-homed MC-LAG configuration share a common B-MAC on the related MC-LAG interfaces. When the **mac-notification no shutdown** command is executed, an Ethernet CFM notification message is sent from the node holding the active link. That message will be flooded in the B-VPLS

domain using the MC-LAG SAP B-MAC as the source MAC address. The remote nodes will learn the customer MAC addresses behind the MC-LAG and will link them to this new SAP B-MAC. MC-LAG will keep track of the active link for each particular LAG associated to a SAP B-MAC. Should MC-LAG detect any new active link in a node, a new CFM notification message will be flooded from the new active node.

The following caveats and considerations must be taken into account:

- The MAC notification for access dual-homed devices no longer requires “chassis mode D”.
  - All network ports and B-SAPs must reside on slots populated with either an IOM3-XP or IMM. All other SAPs may reside on any IOM or IMM populated slot. This is enforced by the CLI.
  - MC-LAG SAPs must be on IOM3-XP/IMM.
  - Other SAPs may be on IOM1/IOM2 (non-PBB services, PBB Epipe SAP with no other SAP on local MC-LAG, PBB I-VPLS SAPs even if some of the other I-SAPs are on the MC-LAG).
- Only MC-LAG is supported as dual-home mechanism.
- This mechanism is supported for native PBB and/or MPLS-based PBB-VPLS. Although it is mostly beneficial when native PBB is used in the core, it can also help to optimize the re-learning process in a MPLS-based core in case of MC-LAG failures, in addition to the existing LDP MAC flush procedures.

The example of this configuration shows the setup being used in this configuration example. MAC-notification will be configured in MTU-4 and MTU-5 for the dual-homed CE-8.

The first step is to configure the sap-bmac that will be used for the mac-notification messages. The source-bmac-lsb (source backbone MAC least significant bits) command has been added to the mc-lag branch so that the operator can decide the two last octets to be used in the sap-bmac. Those two last octets can be derived from the lacp-key (if the use-lacp-key statement is used) or can be specifically defined.

```
*A:MTU-4>config>redundancy>mc>peer>mc-lag# lag
  - lag <lag-id> lacp-key <admin-key> system-id <system-id> [remote-lag <remote-lag-
    id>] system-priority <system-priority> source-bmac-lsb use-lacp-key
  - lag <lag-id> lacp-key <admin-key> system-id <system-id> [remote-lag <remote-lag-
    id>] system-priority <system-priority> source-bmac-lsb <MAC-Lsb>
  - lag <lag-id> lacp-key <admin-key> system-id <system-id> [remote-lag <remote-lag-
    id>] system-priority <system-priority>
  - lag <lag-id> [remote-lag <remote-lag-id>]
  - no lag <lag-id>

 <lag-id>             : [1..200]
 <admin-key>          : [1..65535]
 <system-id>          : xx:xx:xx:xx:xx:xx     - xx [00..FF]
 <remote-lag-id>      : [1..800]
 <system-priority>    : [1..65535]
```

```
 <MAC-Lsb>            : [1..65535] or xx-xx or xx:xx
```

There must be a different sap-bmac per MC-LAG. The use of the lacp-key as a default for two least significant octets makes the operations simpler. In our example, the sap-bmac last two octets will come from the lacp-key:

```
configure
    redundancy
        multi-chassis
            peer 192.0.2.5 create
                mc-lag
                    lag 1 lacp-key 15 system-id 00:00:00:00:00:01 system-
priority 65535                                      source-bmac-lsb use-lacp-key
                    no shutdown
                exit
                no shutdown
```

Therefore, the sap-bmac will be formed in the following way:

```
[4 first bytes of the source bmac + 2 bytes from source-bmac-lsb]
```

Enable the mac-notification in B-VPLS 100 on all MTUs as follows:

```
configure service vpls 100 mac-notification no shutdown
```

The **mac-notification** command activates the described mechanism and has the following parameters:

```
*A:MTU-4# configure service vpls 100 mac-notification
  - mac-notification

 [no] count           - Configure count for MAC-notification messages
 [no] interval        - Configure interval for MAC-notification messages
 [no] renotify        - Configure re-notify interval for MAC-notification messages
 [no] shutdown        - Configure admin state for MAC-notification messages

*A:MTU-4#
```

Where:

- interval <value> controls how often the subsequent MAC notification messages are sent. Default = 100 ms. Required values: 100 ms – 10 sec, in increments of 100 ms.
- count <value> controls how often the MAC notification messages are sent. Default = 3. Range: 1-10.

The “count” and “interval” parameters can also be configured at the service context. The settings configured at the B-VPLS service context take precedence though.

```
*A:MTU-4# configure service mac-notification
  - mac-notification

 [no] count           - Configure count for MAC-notification messages
 [no] interval        - Configure interval for MAC-notification messages

*A:MTU-4#
```

Finally, the B-VPLS is instructed to use the SAP B-MAC. The **use-sap-bmac** statement enables the use of the source B-MAC allocated to the multi-homed SAPs (assigned to the MC-LAG) in the related I-VPLS service (could be Epipe service as well). The command will fail if the value of the source-bmac assigned to the B-VPLS is the hardware (chassis) B-MAC. In other words, the source-bmac must be a configured one. The **use-sap-bmac** statement is by default off.

```
*A:MTU-4# configure service
        vpls 100
            pbb
                source-bmac 00:aa:aa:aa:aa:04
                use-sap-bmac
            exit all

*A:MTU-5# configure service
        vpls 100
            pbb
                source-bmac 00:aa:aa:aa:aa:05
                use-sap-bmac
            exit all

A:MTU-6# show service id 2 fdb pbb
===============================================================================
Forwarding Database, i-Vpls Service 2
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
00:08:00:00:00:00 b-sdp:63:100          100       00:aa:aa:aa:00:0f L/0
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/0
===============================================================================
A:MTU-6#
```

As soon as the **mac-notification no shutdown** command is executed, an Ethernet CFM notification message is sent from MTU-4, which is the node where the active MC-LAG link resides. The CFM message will have the source mac “00:aa:aa:aa:00:0f” (4 first bytes of the configured source bmac + 2 bytes from the configured source-bmac-lsb, which is 15 in hex) and will be flooded throughout the B-VPLS domain. Should the link between CE-8 and MTU-4 fail, the MC-LAG protocol will activate the redundant link and MTU-5 will immediately issue a CFM message with the shared sourced sap-bmac that will be flooded in the B-VPLS domain.

PBB and IGMP Snooping

IGMP snooping can be enabled on I-VPLS SAPs and SDPs (it cannot be enabled on B-VPLS). The 7x50 can keep track of IGMP joins received over individual B-SDPs or B-SAPs, and it starts flooding the Multicast Group (and only the multicast group) to all B-components (using the group B-MAC for I-SID) as soon as the first IGMP join for that multicast group is received in one of the B-SAP/SDP components.

The first IGMP join message received over the local B-VPLS will add all the B-VPLS SAP/SDP components into the related multicast table associated with the I-VPLS context. When the querier is connected to a remote I-VPLS instance, over the B-VPLS infrastructure, its location is identified by the B-VPLS SDP/SAP on which the query was received and also by the source B-MAC address used in the PBB header for the query message, the B-MAC associated with the B-VPLS instance on the remote PBB PE.

The following excerpt shows an I-VPLS with IGMP snooping enabled and some static groups added on a SAP. The location of the querier is configured by adding the B-MAC where the querier is connected to (in this example MTU-6) and adding the two B-VPLS spoke-sdps as mrouter ports (B-VPLS mrouter ports are added in the I-VPLS backbone-vpls context).

```
On MTU-4
configure
    service
        vpls 1 customer 1 i-vpls create
            pbb
                backbone-vpls 100
                    igmp-snooping
                        mrouter-dest "MTU-6"
                    exit
                    sdp 41:100
                        igmp-snooping
                            mrouter-port
                        exit
                    exit
                    sdp 42:100
                        igmp-snooping
                            mrouter-port
                        exit
                    exit
                exit
            exit
            igmp-snooping
                no shutdown
            exit
            sap 1/1/1:7 create
                igmp-snooping
                    static
                        group 228.0.0.1
                            starg
                        exit
                        group 228.0.0.2
```

```
                        starg
                    exit
                    group 239.0.0.1
                        source 172.16.99.99
                    exit
                exit
            exit
        exit
        no shutdown

---snip---
```

As in regular VPLS instances, mrouter ports are added to all the multicast groups:

```
*A:MTU-4# show service id 1 mfib
===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                        Blk
-------------------------------------------------------------------------------
*               *                     b-sdp:41:100                100      Fwd
                                      b-sdp:42:100                100      Fwd
*               228.0.0.1             sap:1/1/1:7                 Local    Fwd
                                      b-sdp:41:100                100      Fwd
                                      b-sdp:42:100                100      Fwd
*               228.0.0.2             sap:1/1/1:7                 Local    Fwd
                                      b-sdp:41:100                100      Fwd
                                      b-sdp:42:100                100      Fwd
172.16.99.99    239.0.0.1             sap:1/1/1:7                 Local    Fwd
                                      b-sdp:41:100                100      Fwd
                                      b-sdp:42:100                100      Fwd
-------------------------------------------------------------------------------
Number of entries: 4
===============================================================================
*A:MTU-4#
```

When the “show service id x mfib” command is issued in a B-VPLS, the group B-MAC entries are shown, whereas when the same command is issued in a I-VPLS, the IGMP (S,G) and (*,G) entries for the I and B components are shown if IGMP snooping is enabled.

```
*A:MTU-4# show service id 100 mfib
===============================================================================
Multicast FIB, Service 100
===============================================================================
Source Address  Group Address         Sap/Sdp Id                  Svc Id   Fwd
                                                                        Blk
-------------------------------------------------------------------------------
*               01:1E:83:00:00:01     b-sdp:41:100                Local    Fwd
*               01:1E:83:00:00:02     b-sdp:41:100                Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

```
*A:MTU-4#
```

MMRP Policies and ISID-Based Filtering for PBB Inter-Domain Expansion

As discussed in the MMRP for Flooding Optimization, MMRP is used in the backbone VPLS instances to build per I-VPLS flooding trees. Each I-VPLS has an associated group B-MAC in the B-VPLS, which is derived from the ISID, and is advertised by MMRP throughout the whole B-VPLS context, regardless of whether a certain I-VPLS is present in one or all the B-VPLS PEs.

In an inter-domain environment, the same B-VPLS can be defined in different domains and as such MMRP will advertise all the group B-MACs in every domain, wasting resources in all the PEs no matter if a particular ISID, and hence its group B-MAC, are not required in one of the domains. When MMRP is enabled in a particular PE, data plane and control plane resources are consumed and they must be taken into consideration when designing PBB-VPLS networks:

- Control plane – MRRP processing takes CPU cycles and the number of attributes that can be advertised is not unlimited
- Data plane – each group B-MAC registration takes one MFIB entry (the MFIB is shared between MMRP and IGMP/PIM snooping)

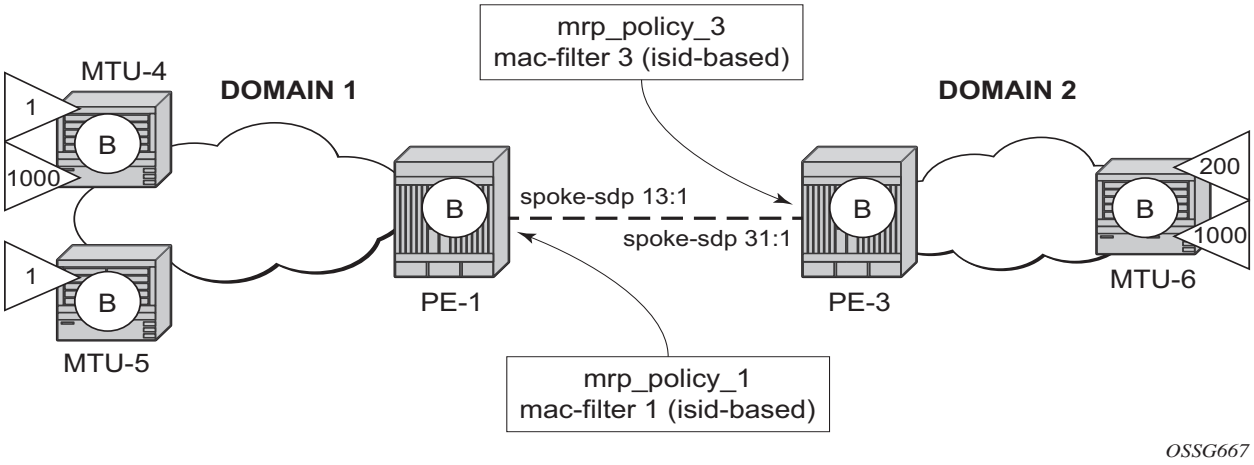
The 7x50 supports MMRP policies and ISID-based filters so that control plane and data plane resources can be saved when I-VPLS instances are not defined in all the domains.

The following figure illustrates an example of usage for MMRP policies and ISID-based filters that will be configured in this section. “Domain 1” and “domain 2” will have a range of local ISIDs each and a range of “inter-domain” ISIDs:

- Domain 1 local ISIDs: from 1 to 100
- Domain 2 local ISIDs: from 101 to 200
- Inter-domain ISIDs: from 1000 to 2000

By applying the MMRP policies indicated in the figure, domain 1 attributes will be prevented from being declared and registered in domain 2 and vice versa, domain 2 attributes from being declared and registered in domain 1. The egress mac-filters will drop any traffic sourced from a local ISID preventing it to be transmitted to the remote domain.

Figure 314 **Inter-Domain B-VPLS and MMRP Policies/ISID-Based Filters Example**



MMRP Policies

The following shows the MMRP policy configuration on node PE-1. This policy will block any registration/declaration except those for ISIDs 1000-2000. Packets will be compared against the configured matching ISIDs as long as the pbb-etype matches the one configured on the port or SDP.

On PE-1:

```
configure
    service
        mrp
            mrp-policy "mrp_policy_1" create
                description "allow-inter-domain-isids"
                default-action block
                entry 10 create
                    action allow
                    match
                        isid 1000 to 2000
                    exit
                exit
            exit
        exit
    exit
```

Once the MMRP policy is configured, it must be applied on the corresponding SAP or sdp-binding. An mrp-policy can be applied to a B-VPLS SAP, B-VPLS spoke-sdp or B-VPLS mesh-sdp:

On PE-1:

```
configure
    service
        vpls 100
            spoke-sdp 14:100 create
                mrp
                    mrp-policy "mrp_policy_1"
                exit
            exit
            spoke-sdp 15:100 create
                mrp
                    mrp-policy "mrp_policy_1"
                exit
            exit
            no shutdown
        exit
```

In the same way, mrp_policy_3 will be configured in PE-3.

Some additional considerations about the MMRP policies:

- Different entries within the same mrp-policy can have overlapping ISID ranges. The entries will be evaluated in the order of their IDs and the first match will cause the implementation to execute the associated action for that entry and then to exit the mrp-policy.
- If no ISID is specified in the match condition then:
  - If the action is “end-station”, no entry is added and the action is block.
  - If the action is different from “end-station”, every ISID is considered for that action.
- The mrp-policy specifies either a forward or a drop action for the group B-MAC attributes associated with the ISIDs specified in the match criteria.

```
*A:PE-1>config>serv>mrp>mrp-policy>entry# action
 - action <action>
 - no action

 <action>             : none|block|allow|end-station


*A:PE-1>config>serv>mrp>mrp-policy>entry#
```

- There is an additional action called e**nd-station.** This action specifies that an end-station emulation is present on the SAP/SDP-binding where the policy has been applied. The matching ISIDs will not get declared/registered in the SAP/SDP-binding (just like the **block** action), however those attributes will get mapped as static MMRP entries on the SAP/SDP-binding, which implicitly get instantiated in the data plane as MFIB entries associated with that SAP/SDP-binding for the related group B-MAC. When the action is “end-station”, the default-action must be block:

```
*A:PE-3>config>serv>mrp>mrp-policy# default-action allow
MINOR: SVCMGR #5904 Mrp-policy default-action must be block when end-station action
exists
```

- The **end-station** action can be used in the inter-domain gateways when, for instance, we do not want MMRP control plane exchanges between domains. The following output shows how to define the static MMRP entries 1000-2000 in PE-3 without receiving any declaration for any of those attributes or having any of those locally configured.

  On PE-3:

```
configure
    service
        mrp
            mrp-policy "mrp_policy_3" create
                default-action block
                entry 10 create
                    action end-station
                    match
                        isid 1000 to 2000
                    exit
                exit
            exit
        exit
    exit
```

```
*A:PE-3# show service id 100 mfib
===============================================================================
Multicast FIB, Service 100
===============================================================================
Source Address  Group Address         Sap/Sdp Id              Svc Id   Fwd/Blk
-------------------------------------------------------------------------------
*               01:1E:83:00:03:E8     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:03:E9     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:03:EA     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:03:EB     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:03:EC     b-sdp:36:100            Local    Fwd

---snip---

*               01:1E:83:00:07:CD     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:07:CE     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:07:CF     b-sdp:36:100            Local    Fwd
*               01:1E:83:00:07:D0     b-sdp:36:100            Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1001
===============================================================================
*A:PE-3#
```

- The mrp-policy can be applied to multiple B-VPLS services as long as the scope of the policy is **template** (the scope can also be **exclusive**).

- Any changes made to the existing policy will be applied immediately to all services where this policy is applied. For this reason, when many changes are required on a mrp-policy, it is recommended that the policy be copied to a work-in-progress policy. That work-in-progress policy can be modified until complete and then written over the original mrp-policy. You can use the **config mrp-policy copy** command to work with the policies in this manner. The **renum** command can also help to change the entries sequence order.

```
A:PE-3# configure service mrp copy
  - copy <src-mrp-policy> to <dst-mrp-policy>

 <src-mrp-policy>     : [32 chars max]
 <dst-mrp-policy>     : [32 chars max]

A:PE-3#

A:PE-3# configure service mrp mrp-policy "mrp_policy_3" renum
  - renum <src-entry-id> to <dst-entry-id>

 <src-entry-id>       : [1..65535]
 <dst-entry-id>       : [1..65535]

A:PE-3#
```

- The **no** form of the **mrp-policy** command deletes the mrp-policy. An mrp policy cannot be deleted until it is removed from all the SAPs/SDP-bindings where it is applied.

ISID-Based Filters

The MMRP policies help to control the exchange of group B-MAC attributes across domains. Based on the registration state of a specific group B-MAC on a SAP/SDP-binding, the broadcast/ unknown-unicast/multicast traffic for a particular I-VPLS will be allowed or dropped. However, to avoid that any local ISID packet is flooded to the remote B-VPLS domain, all the packets tagged with the local ISIDs at the gateway PEs need to be filtered at the data plane. ISID- based filters will prevent the local ISIDs from sending any packet with unicast B-MAC to the remote domain. This is particularly useful for PBB-Epipe services across domains, where all the frames use unicast B-MACs and MMRP policies cannot help because they only act on group B-MAC packets.

The following CLI output shows how to configure an ISID-based filter that drops all the traffic sourced from the local ISIDs on PE-1 (the default action is drop and it does not show up in the configuration).

```
*A:PE-1# configure filter
        mac-filter 1 create
            description "drop_local_isids"
```

```
            type isid
            entry 10 create
                match frame-type 802dot3
                    isid 1000 to 2000
                exit
                log 101
                action
                    forward
                exit
            exit
```

Once the filter is configured, it must be applied on a B-VPLS SAP or SDP-binding and always at egress.

```
*A:PE-1# configure service
        vpls 100
            spoke-sdp 14:100 create
                egress
                    filter mac 1
                exit
            exit
            no shutdown
            spoke-sdp 15:100 create
                egress
                    filter mac 1
                exit
            exit
```

Some additional comments about ISID-based filters:

- The **type isid** statement must be added before introducing any ISID in the match command, otherwise the system will show an error:

```
*A:PE-1>config>filter>mac-filter>entry>match$ isid 1000 to 2000
MINOR: FILTER #1533 The match criteria entered are not compatible with the Mac
filter type - On a normal filter no ISID or VID match criteria are allowed

*A:PE-1>config>filter>mac-filter$             type isid
MINOR: FILTER #1561 Cannot change filter type when filter contains entries
```

- Once the operator sets the “type isid”, the filter cannot be applied at ingress. Only egress ISID-based filters are allowed:

```
*A:PE-1>config>service>vpls>mesh-sdp# ingress filter mac 1
MINOR: SVCMGR #2050 Can not apply filter of type 'isid' on ingress
```

- Like any filter or MMRP policy, the filter can be applied to multiple B-VPLS services as long as the scope of the policy is “template” (the scope can also be “exclusive”).

- The following command shows the filter configuration and packets that have matched the filter (field “Egr. Matches”):

```
*A:PE-1# show filter mac 1
===============================================================================
Mac Filter
===============================================================================
Filter Id           : 1                             Applied        : Yes
Scope               : Template                      Def. Action    : Drop
Entries             : 1                             Type           : isid
Description         : drop_local_isids
-------------------------------------------------------------------------------
Filter Match Criteria : Mac
-------------------------------------------------------------------------------
Entry               : 10                            FrameType      : Ethernet
Description         : (Not Specified)
Log Id              : 101
ISID                : 1000..2000
Primary Action      : Forward
Ing. Matches        : 0 pkts
Egr. Matches        : 5 pkts (580 bytes)

===============================================================================
*A:PE-1#
```

- Like any other filter, the matching packets can be logged. An example follows (the Ethertype is 0x88e7, which is the default standard etype for PBB):

```
*A:PE-1# show filter log 101
===============================================================================
Filter Log
===============================================================================
Admin state : Enabled
Description : Default filter log
Destination : Memory
Wrap        : Enabled
-------------------------------------------------------------------------------
Maximum entries configured : 1000
Number of entries logged   : 5
-------------------------------------------------------------------------------
2016/05/24 08:40:41  Mac Filter: 1:10  Desc:
Interface: int-PE-1-MTU-4  Direction: Egress  Action: Forward
VID match: 0
Src MAC: 00-06-06-06-06-06  Dst MAC: 00-aa-aa-aa-00-0f  EtherType: 88e7
Hex: 00 00 03 e9 00 08 00 00 00 00 00 10 00 00 00 00
     08 00 45 00 00 54 00 24 00 00 40 01 4a 61 ac 10
     6c 02 ac 10 6c 01 00 00 fe 88 c0 09 00 01 57 44*
2016/05/24 08:40:42  Mac Filter: 1:10  Desc:
Interface: int-PE-1-MTU-4  Direction: Egress  Action: Forward
VID match: 0
Src MAC: 00-06-06-06-06-06  Dst MAC: 00-aa-aa-aa-00-0f  EtherType: 88e7
Hex: 00 00 03 e9 00 08 00 00 00 00 00 10 00 00 00 00
     08 00 45 00 00 54 00 25 00 00 40 01 4a 60 ac 10
     6c 02 ac 10 6c 01 00 00 fd 9d c0 09 00 02 57 44*

---snip---
```

```
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

B-VPLS and I-VPLS Show and Debug Commands

For the following output, the MRP policies and ISID-based MAC filters have been removed from the spoke SDPs on PE-1 and PE-3. The following commands can help to check the B-VPLS and I-VPLS configuration and their related parameters. The first is for the B-VPLS on MTU-4:

```
*A:MTU-4# show service id 100 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 100                 Vpn Id            : 0
Service Type      : b-VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/24/2016 05:27:03
Last Mgmt Change  : 05/24/2016 06:49:29
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 2000                Def. Mesh VC Id   : 100
SAP Count         : 0                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>
Oper Backbone Src : 00:aa:aa:aa:aa:04
Use SAP B-MAC     : Enabled
i-Vpls Count      : 2
Epipe Count       : 0
Use ESI B-MAC     : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:41:100 S(192.0.2.1)                 Spok         8000    8000    Up   Up
sdp:42:100 S(192.0.2.2)                 Spok         8000    8000    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:MTU-4#
```

For the I-VPLS on MTU-4:

```
*A:MTU-4# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : i-VPLS
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 05/24/2016 05:38:08
Last Mgmt Change  : 05/24/2016 07:28:35
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 1
SAP Count         : 1                   SDP Bind Count    : 0
Snd Flush on Fail : Enabled             Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
VSD Domain        : <none>
b-Vpls Id         : 100                 Oper ISID         : 1
b-Vpls Status     : Up
Snd Flush in bVpls: None
Flsh On bVpls Fail: Disabled            Prop Flsh fr bVpls: Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:7                              q-tag        1518    1518    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:MTU-4#
```

The following command shows all the I-VPLS instances multiplexed into a particular B-VPLS.

```
*A:MTU-4# show service id 100 i-vpls
===============================================================================
Related i-Vpls services for b-Vpls service 100
===============================================================================
i-Vpls SvcId        Oper ISID           Admin               Oper
-------------------------------------------------------------------------------
1                   1                   Up                  Up
2                   2                   Up                  Up
-------------------------------------------------------------------------------
Number of Entries : 2
```

```
-------------------------------------------------------------------------------
===============================================================================
*A:MTU-4#
```

Some useful commands to check the I and B VPLS FIBs correlating C-MACs and B-MACs:

```
*A:MTU-4# show service id 1 fdb pbb
===============================================================================
Forwarding Database, i-Vpls Service 1
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
-------------------------------------------------------------------------------
00:07:00:00:00:00 sap:1/1/1:7           100       N/A               L/0
00:09:00:00:00:00 b-sdp:41:100          100       00:06:06:06:06:06 L/0
===============================================================================
*A:MTU-4#
*A:MTU-4# show service id 100 fdb pbb
===============================================================================
Forwarding Database, b-Vpls Service 100
===============================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
-------------------------------------------------------------------------------
00:06:06:06:06:06 sdp:41:100            2          0          L/0
4a:c4:ff:00:00:00 sdp:41:100            0          0          L/0
===============================================================================
*A:MTU-4#
```

The following command shows the base MAC notification parameters as well as the source B-MAC configured at the service PBB level. Those values are overridden by any potential mac-notification or source B-MAC values configured under the B-VPLS service context.

```
*A:MTU-4# show service pbb base
===============================================================================
PBB MAC Information
===============================================================================
MAC-Notif Count                   : 3
MAC-Notif Interval                : 1
Source BMAC                       : 00:04:04:04:04:04
===============================================================================
*A:MTU-4#
```

If mac-notification is used in a particular B-VPLS, the configured least significant bits for the sap-bmac on a particular MC-LAG can be shown by using the detailed view of the **show lag** command:

```
*A:MTU-4# show lag 1 detail
===============================================================================
LAG Details
===============================================================================
Description        : N/A
-------------------------------------------------------------------------------
```

```
Details
-------------------------------------------------------------------------------
Lag-id              : 1                     Mode                : access
Adm                 : up                    Opr                 : up

---snip---

MC Peer Address     : 192.0.2.5             MC Peer Lag-id      : 1
MC System Id        : 00:00:00:00:00:01     MC System Priority  : 65535
MC Admin Key        : 15                    MC Active/Standby   : active
MC Lacp ID in use   : true                  MC extended timeout : false
MC Selection Logic  : local master decided
MC Config Mismatch  : no mismatch
Source BMAC LSB     : use-lacp-key          Oper Src BMAC LSB   : 00:0f

---snip---
===============================================================================
*A:MTU-4#
```

The following debug commands allow the operator to check the LDP label mapping, label withdrawal, messages and also the MAC-flush messages for regular VPLS, for I-VPLS and B-VPLS including the PBB extensions and TLVs.

```
*A:MTU-4# show debug
debug
    router "Base"
        ldp
            peer 192.0.2.1
                event
                exit
                packet
                    init detail
                    label detail
                exit
            exit
            peer 192.0.2.2
                event
                exit
                packet
                    init detail
                    label detail
                exit
            exit
        exit
    exit
exit
*A:MTU-4#
```

The following debug commands can help the operator to troubleshoot MMRP.

```
*A:MTU-4# debug service id 100 mrp
  - mrp
  - no mrp

      all-events      - Enable/disable MRP debugging for all events
```

```
  [no] applicant-sm    - Enable/disable MRP debugging for applicant state machine
                         changes
  [no] leave-all-sm    - Enable/disable MRP debugging for leave all state machine
                         changes
  [no] mmrp-mac        - Enable/disable MRP debugging for a particular MAC address
  [no] mrpdu           - Enable/disable MRP debugging for Rx/Tx MRP PDUs
  [no] mvrp-vlan       - Enable/disable debugging for a particular vlan
  [no] periodic-sm     - Enable/disable MRP debugging for periodic state machine
                         changes
  [no] registrant-sm   - Enable/disable MRP debugging for registrant state machine
                         changes
  [no] sap             - Enable/disable MRP debugging for a particular SAP
  [no] sdp             - Enable/disable MRP debugging for a particular SDP

*A:MTU-4#
```

Conclusion

PBB-VPLS allows the service providers to scale VPLS services by multiplexing customer I-VPLS instances into one or more B-VPLS instances. This multiplexing dramatically reduces the number of services, pseudowires and MAC addresses in the core and therefore allows the service provider to scale Layer 2 multi-point networks and provide services across international backbones.

The example used in this section shows the configuration of the customer and backbone VPLS instances as well as all the related features which are required for this environment. Show and debug commands have also been suggested so that the operator can verify and troubleshoot the service.

Shortest Path Bridging for MAC

This chapter describes advanced shortest path bridging for MAC configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

Shortest path bridging for MAC (SPBM) is supported in SR OS release 10.0.R4, or later. SPB Static MAC, static backbone-service instance identifiers (ISIDs) and ISID-policies for SPB are supported in release 11.0.R4, or later. This chapter was initially written for SR OS release 11.0.R4, but the CLI in this edition is based on release 14.0.R3.

Overview

SPB enables a next generation control plane for provider backbone bridges (PBB) and PBB-VPLS that adds the stability and efficiency of link state to unicast and multicast services (Epipes and I-VPLSs). In addition, SPBM provides resiliency, load-balancing and multicast optimization without the need for any other control plane in the B-VPLS (for example, there is no need for spanning tree, or G.8032, or multiple MAC registration protocol (MMRP)).

SPBM exploits the complete knowledge of backbone addressing, which is a key consequence of the PBB hierarchy, by advertising and distributing the backbone MACs (BMACs) through a link-state protocol, namely IS-IS. An immediate effect of this is that the old “flood-and-learn” can at last be turned off in the backbone and every B-VPLS node in the network will know what destination BMAC addresses are expected and valid. As a result of that, receiving an unknown unicast BMAC on a B-VPLS SAP/PW is indicative of an error, whereupon the frame is discarded (due to the reverse path forwarding check – RPFC – performed in SPBM) instead of flooded. Furthermore, SPBM allows condensing all the relevant information distribution (unicast and multicast) into a single control protocol: IS-IS.

SPBM can be easily enabled on the existing B-VPLS instances being used for multiplexing I-VPLS/Epipe services, providing the benefits summarized below:

- Per-service flood containment (for I-VPLS services) without the need for an additional protocol such as MMRP.
- Loop avoidance in the B-VPLS domain without the need for MSTP or other technologies.
- No unknown BMAC flooding in the B-VPLS domain.
- No need for MAC notification mechanisms or vMEPs in the B-VPLS to update the B-VPLS forwarding databases (FDBs) (vMEPs can still be configured though for OAM purposes).

Some other characteristics of the SPB implementation in the SR OS are:

- The SR OS SPB implementation always uses multi-topology (MT) topology instance zero. However, up to four logical instances (that is, SPB instances in different B-VPLS services) are supported if different topologies are required for different services.
- Area addresses are not used and SPB is assumed to be a single area. SPB must be consistently configured on nodes in the system. SPB Regions information and IS-IS hello logic that detect mismatched configuration are not supported. IS-IS area is always zero.
- SPB uses all-intermediate-systems 09-00-2B-00-00-05 destination MAC to communicate.
- SPB Source ID is always zero.
- SPB uses a separate instance of IS-IS from the base IP IS-IS. IS-IS for SPB is configured in the SPB context under the B-VPLS component. Up to four ISIS-SPB instances are supported, where the instance identifier can be any number between 1024 and 2047. The instance number is not in TLVs.
- Two equal cost tree (ECT) algorithms (IEEE 802.1aq) per SPB instance are supported: low-path-id and high-path-id algorithms.
- SPB link state protocol data units (link state packets) contain BMACs, ISIDs (for multicast services) and link and metric information for an IS-IS database.
  - Epipe ISIDs are not distributed in SR OS SPB allowing high scalability of PBB Epipes.
  - I-VPLS ISIDs are distributed in SR OS SPB and the respective multicast group addresses (composed of PBB-OUI plus ISID) are automatically populated in a manner that provides automatic pruning of multicast to the subset of the multicast tree that supports an I-VPLS with a common ISID. This replaces the function of MMRP and is more efficient than MMRP.

- Multiple ISIS-SPB adjacencies between two nodes are not supported as per the IEEE 802.1aq standard specification. If multiple links between two nodes exist, LAG must be used.

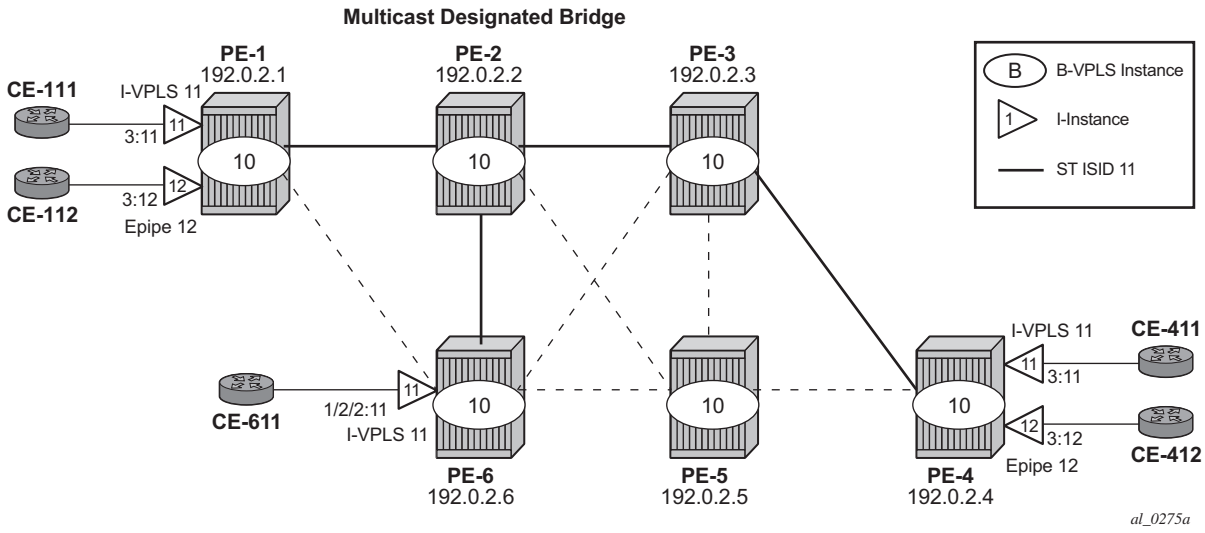
Configuration

This section will describe the configuration of SPBM on SR OS as well as the available troubleshooting commands.

Basic SPBM Configuration

Figure 315 shows the topology used as an example of a basic SPBM configuration.

Figure 315 **Basic SPBM Topology**



Assume the following protocols and objects are configured beforehand:

- The six PEs shown in Figure 315 are running IS-IS for the global routing table with all the interfaces being Level-2.
- LDP is used as the MPLS protocol to signal transport tunnel labels.
- LDP SDPs are configured among the six PEs, in the way it is depicted in Figure 315 (dashed lines and bold lines among PEs).

Once the network infrastructure is properly running, the actual service configuration can be carried out. In the example, B-VPLS 10 will provide backbone connectivity for the services I-VPLS 11 and Epipe 12.

The SPBM configuration is only relevant to the B-VPLS instance and can be added to an existing B-VPLS, assuming that such a B-VPLS does not contain any non-SPB-compatible configuration parameters. The following parameters are not supported in SPB-enabled B-VPLS instances:

- Mesh SDPs (only SAPs or spoke-sdps are supported in SPB-enabled B-VPLS)
- Spanning Tree Protocol (STP)
- Split-Horizon Groups
- Non-conditional Static-macs (configured under sap/spoke-sdps, see section about static BMAC configuration)
- G.8032
- Propagate-mac-flush and send-flush-on-failure
- Maximum number of MACs (max-nbr-mac-addr)
- Bridge Protocol Data Unit (BPDU) translation
- Layer 2 Protocol Termination (L2PT)
- MAC-pinning
- Oper-groups
- MAC-move
- Any BGP, BGP-AD (BGP auto-discovery) or BGP-VPLS (BGP virtual private LAN services) parameters
- Endpoints
- Local/remote age
- MAC-notification
- MAC-protect
- Multiple MAC Registration Protocol (MMRP)
- Provider-tunnel
- Temporary flooding

Assuming all the parameters mentioned are not configured in the B-VPLS (B-VPLS 10 in the example), SPBM can be enabled. The SPBM parameters are all configured in the **config>service>vpls(b-vpls)>spb** and **config>service>vpls(b-vpls)>spoke-sdp/sap>spb** contexts:

```
*A:PE-1# configure service vpls 10 spb ?
  - no spb
  - spb [<isis-instance>] [fid <fid>] [create]
 <isis-instance>      : [1024..2047]
```

```
 <fid>                : [1..4095]

      level           + Configure SPB level information
 [no] lsp-lifetime    - Configure LSP lifetime
 [no] lsp-refresh-in* - Configure LSP refresh interval
 [no] overload        -
 Configure the local router so that it appears to be overloaded
 [no] overload-on-bo* -
 Configure the local router so that it appears to be overloaded
   at boot up
 [no] shutdown        - Administratively enable or disable the operation of ISIS
      timers          + Configure ISIS timers
*A:PE-1# configure service vpls 10 spb timers ?
  - timers
 [no] lsp-wait        -  [no] spf-wait        - ?

*A:PE-1# configure service vpls 10 spoke-sdp 35:10 spb ?
  - no spb
  - spb [create]
 <create>             : keyword
      level           + Configure SPB level information
 [no] lsp-pacing-int* - Configure the interval between LSP packets are sent from the
interface
 [no] retransmit-int* - Configure the minimum interval between LSP packets
retransmission for the given interface
 [no] shutdown        - Administratively Enable/disable the interface

*A:PE-1# configure service vpls 10 spoke-sdp 35:10 spb level 1 ?
  - level <[1..1]>
 [no] hello-interval  - Configure hello-interval for this interface
 [no] hello-multipli* - Configure hello-multiplier for this level
 [no] metric          - Configure IS-IS interface metric for IPv4 unicast
```

The parameters configured in the spb context refer to the SPB IS-IS and they should be configured following the same considerations as for the IS-IS base instance:

- spb [<isis-instance>] [fid <fid>] [create]
  - <isis-instance> identifies the SPB ISIS process. Up to four different ISIS SPB processes can be run in a system (range 1024 to 2047).
  - <fid> or *forwarding identifier* identifies the standard SPBM B-VID which is signaled in IS-IS with each advertised BMAC. Each B-VPLS has a single configurable FID.
- spb>lsp-lifetime <seconds> : [350..65535]
- spb>lsp-refresh-interval <seconds> : [150..65535]
- spb>overload [timeout <seconds>] : [60..1800]
- spb>overload-on-boot [timeout <seconds>] : [60..1800]
- spb>timers>lsp-wait <lsp-wait> [<lsp-initial-wait> [<lsp-second-wait>]]
  - <lsp-wait> : [10..120000] – milliseconds

  - <lsp-initial-wait> : [10..100000] – milliseconds
  - <lsp-second-wait> : [10..100000] – milliseconds
- spb>timers>spf-wait <spf-wait> [<spf-initial-wait> [<second-wait>]]
  - <spf-wait> : [10..120000] – milliseconds
  - <spf-initial-wait> : [10..100000] – milliseconds
  - <second-wait> : [10..100000] – milliseconds
- spoke-sdp/sap>spb>lsp-pacing-interval <milli-seconds> : [0..65535]
  - spoke-sdp/sap>spb>retransmit-interval <seconds> : [1..65535]
  - spoke-sdp/sap>spb>level 1>hello-interval <seconds> : [1..20000]
  - spoke-sdp/sap>spb>level 1>hello-multiplier <multiplier> : [2..100]

In the same way lsp-wait (initial-wait) and spf-wait (initial wait) can be tuned in the base router IS-IS instance to minimize the convergence time (to 0 and 10 respectively), the equivalent SPB IS-IS parameters should also be adjusted so that failover time is minimized at the service level.

The following parameters are specific to SPBM (note that only IS-IS level 1 is supported for SPB):

- spb>level 1>bridge-priority <bridge-priority> : [0..15]
  - This parameter will influence the election of the multicast designated bridge through which all the ST (single trees) for the multicast traffic will be established. The default value will be lowered on that node where the multicast designated bridge function is desired, normally because that node is the best connected node. In the example , PE-2 is the multicast designated bridge for B-VPLS 10 and therefore, PE-2 will be the root of the STs for the I-VPLS instances in that B-VPLS. Default value = 8.
- spb>level 1>ect-algorithm fid-range <fid-range> {low-path-id|high-path-id}
  - This command defines the ect-algorithm used and the FIDs assigned. Two algorithms are supported: low-path-id and high-path-id. They can provide the required path diversity for an efficient load balancing in the B-VPLS. Default = fid-range 1-4095 low-path-id
- spb>level 1>forwarding-tree-topology unicast {spf|st}
  - This command configures the type of tree that will be used for unicast traffic: shortest path tree or single tree. The multicast traffic (that encapsulated I-VPLS broadcast, unicast and multicast (BUM) traffic always uses the ST path (single tree path). Using SPF for unicast traffic can produce some packet re-ordering for unicast traffic compared to BUM traffic as different trees are used, therefore, when the B-VPLS transports I-VPLS traffic and the unicast and multicast trees do not follow the same path, it is recommended to use ST paths for unicast and multicast. Default value = spf.

- spoke-sdp/sap>spb>level 1>metric <ipv4-metric> : [1..16777215]
  - This command configures the metric for each SPB interface (spoke-SDP or SAP). This value helps influence the SPF calculation in order to pick a certain path for the traffic to a remote system BMAC.When the SPB link metric advertised by two peers is different, the maximum value is chosen according to the RFC 6329. Default metric = 10.

As an example, the following CLI output shows the relevant configuration of PE-1 and PE-2 (the multicast designated bridge). SPB has to be created and enabled (no shutdown) at B-VPLS service level first and then created and enabled under each and every SAP/spoke-sdp in the B-VPLS. Non-SPB-enabled SAPs/spoke-sdps can exist in the SPB B-VPLS only if conditional static-macs are configured for them (refer to Static BMACs and Static ISIDs Configuration). As for regular B-VPLS services, the service MTU has to be changed from the default value (1500) to a number 18-bytes greater than the I-VPLS service MTU in order to allow for the PBB encapsulation.

```
*A:PE-1# configure service
        pbb
            source-bmac 00-00-00-01-01-01
            mac-name "PE-1" 00-00-00-01-01-01
            mac-name "PE-2" 00-00-00-02-02-02
            mac-name "PE-3" 00-00-00-03-03-03
            mac-name "PE-4" 00-00-00-04-04-04
            mac-name "PE-5" 00-00-00-05-05-05
            mac-name "PE-6" 00-00-00-06-06-06
        exit
        vpls 10 customer 1 b-vpls create
            service-mtu 2000
            spb 1024 fid 10 create
                overload-on-boot timeout 60
                timers
                    spf-wait 2000 spf-initial-wait 50000 spf-second-wait 100000
                    lsp-wait 8000 lsp-initial-wait 10 lsp-second-wait 1000
                exit
                no shutdown
            exit
            spoke-sdp 12:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 16:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        vpls 11 customer 1 i-vpls create
            pbb
                backbone-vpls 10
                exit
```

```
            exit
            sap 1/1/3:11 create
            exit
            no shutdown
        exit
        epipe 12 customer 1 create
            pbb
                tunnel 10 backbone-dest-mac "PE-4" isid 12
            exit
            sap 1/1/3:12 create
            exit
            no shutdown
        exit
```

As discussed, the bridge-priority will influence the election of the multicast designated bridge. By making PE-2's bridge-priority zero, it ensures that PE-2 becomes the root of all the STs for B-VPLS 10 as long as the priority for the rest of the PEs is larger than zero. In case of a tie, the PE owning the lowest system BMAC will be elected as multicast designated bridge. Figure 315 shows the ST for I-VPLS 11 (see a thicker continuous line representing the ST). PE-2 is the root of the ST tree.

```
*A:PE-2# configure service
        pbb
            source-bmac 00:00:00:02:02:02
            mac-name "PE-1" 00:00:00:01:01:01
            mac-name "PE-2" 00:00:00:02:02:02
            mac-name "PE-3" 00:00:00:03:03:03
            mac-name "PE-4" 00:00:00:04:04:04
            mac-name "PE-5" 00:00:00:05:05:05
            mac-name "PE-6" 00:00:00:06:06:06
        exit
        vpls 10 customer 1 b-vpls create
            service-mtu 2000
            spb 1024 fid 10 create
                level 1
                    bridge-priority 0
                exit
                overload-on-boot timeout 60
                timers
                    spf-wait 2000 spf-initial-wait 50000 spf-second-wait 100000
                    lsp-wait 8000 lsp-initial-wait 10 lsp-second-wait 1000
                exit
                no shutdown
            exit
            spoke-sdp 21:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 23:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
```

```
            spoke-sdp 25:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 26:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The rest of the nodes will be configured accordingly. SPB instance 1024 will set up SPF (shortest path first) trees for unicast traffic and a ST (single tree) per ISID with PE-2 as the root-bridge (because it has the lowest bridge-priority 0 configured) for BUM traffic. The ECT algorithm chosen for the B-VPLS FID (10) is the low-path-id (default one).

Once SPBM is configured on all the six nodes, the six system BMACs and the ISID 11 will be advertised by SPB IS-IS.

The following show commands can help understand the IS-IS configuration for SPB 1024 and the BMACs populated by IS-IS:

- show service id spb base: provides the SPB configuration and parameters for a particular SPB B-VPLS.
- show service id 10 spb fdb: provides the B-VPLS FDB that has been populated by IS-IS, for the unicast and multicast entries.

```
A:PE-1# show service id 10 spb base
===============================================================================
Service SPB Information
===============================================================================
Admin State        : Up                 Oper State         : Up
ISIS Instance      : 1024               FID                : 10
Bridge Priority    : 8                  Fwd Tree Top Ucast : spf
Fwd Tree Top Mcast : st
Bridge Id          : 80:00.00:00:00:01:01:01
Mcast Desig Bridge : 00:00.00:00:00:02:02:02
===============================================================================
Router Base ISIS Instance 1024 Interfaces
===============================================================================
Interface                       Level CircID  Oper State   L1/L2 Metric
-------------------------------------------------------------------------------
sdp:12:10                       L1    65536   Up           10/-
sdp:16:10                       L1    65537   Up           10/-
-------------------------------------------------------------------------------
Interfaces : 2
===============================================================================
FID ranges using ECT Algorithm
-------------------------------------------------------------------------------
```

```
1-4095    low-path-id
===============================================================================
A:PE-1#
*A:PE-1# show service id 10 spb fdb
===============================================================================
User service FDB information
===============================================================================
MAC Addr          UCast Source          State   MCast Source          State
-------------------------------------------------------------------------------
00:00:00:02:02:02 12:10                 ok      12:10                 ok
00:00:00:03:03:03 12:10                 ok      12:10                 ok
00:00:00:04:04:04 12:10                 ok      12:10                 ok
00:00:00:05:05:05 12:10                 ok      12:10                 ok
00:00:00:06:06:06 16:10                 ok      12:10                 ok
-------------------------------------------------------------------------------
Entries found: 5
===============================================================================
*A:PE-1#
```

It can be seen that the unicast (SPF) tree and the multicast (ST) tree differ with respect to PE-6.

The following commands help check the unicast and multicast topology for B-VPLS 10:

- **show service id 10 spb routes** provides a detailed view of the unicast and multicast routes computed by SPF. As shown below, the SPB unicast and multicast routes match on PE-2 since this node is the multicast designated bridge. Unicast and multicast routes will differ on most other nodes.
- **show service id 10 spb mfib** and **show service id 10 mfib** shows information of the MFIB entries generated in the B-VPLS as well as the outgoing interface (OIF) associated with those MFIB entries.

```
*A:PE-2# show service id 10 spb routes
===============================================================
MAC Route Table
===============================================================
FID  MAC Addr                                    Ver.   Metric
      NextHop If                   SysID
---------------------------------------------------------------

Fwd Tree: unicast
---------------------------------------------------------------
10   00:00:00:01:01:01                          4      10
      sdp:21:10                   PE-1
10   00:00:00:03:03:03                          6      10
      sdp:23:10                   PE-3
10   00:00:00:04:04:04                          8      20
      sdp:23:10                   PE-3
10   00:00:00:05:05:05                          12     10
      sdp:25:10                   PE-5
10   00:00:00:06:06:06                          14     10
      sdp:26:10                   PE-6
```

```
Fwd Tree: multicast
------------------------------------------------------------------
10   00:00:00:01:01:01                            4      10
      sdp:21:10                    PE-1
10   00:00:00:03:03:03                            6      10
      sdp:23:10                    PE-3
10   00:00:00:04:04:04                            8      20
      sdp:23:10                    PE-3
10   00:00:00:05:05:05                            12     10
      sdp:25:10                    PE-5
10   00:00:00:06:06:06                            14     10
      sdp:26:10                    PE-6
------------------------------------------------------------------
No. of MAC Routes: 10
==================================================================
==================================================================
ISID Route Table
==================================================================
FID  ISID                                         Ver.
      NextHop If                   SysID
------------------------------------------------------------------
10   11                                           17
      sdp:21:10                    PE-1
      sdp:23:10                    PE-3
      sdp:26:10                    PE-6
------------------------------------------------------------------
No. of ISID Routes: 1
==================================================================
*A:PE-2#


*A:PE-2# show service id 10 spb mfib
================================================================================
User service MFIB information
================================================================================
MAC Addr          ISID     Status
--------------------------------------------------------------------------------
01:1E:83:00:00:0B 11       Ok
--------------------------------------------------------------------------------
Entries found: 1
================================================================================
*A:PE-2#


*A:PE-2# show service id 10 mfib
================================================================================
Multicast FIB, Service 10
================================================================================
Source Address  Group Address         Sap/Sdp Id               Svc Id   Fwd/Blk
--------------------------------------------------------------------------------
*               01:1E:83:00:00:0B     b-sdp:21:10              Local    Fwd
                                      b-sdp:23:10              Local    Fwd
                                      b-sdp:26:10              Local    Fwd
--------------------------------------------------------------------------------
Number of entries: 1
================================================================================
*A:PE-2#
```

SPB multicast trees (STs) are pruned for each particular I-VPLS ISID, based on the advertisement of I-VPLS ISIDs in SPB IS-IS by each individual PE. Multicast B-VPLS traffic not belonging to any particular I-VPLS follows the *default tree*. The default tree is an ST for the B-VPLS which is not pruned and therefore reaches all the PE nodes in the B-VPLS. For instance, Ethernet-CFM CCM messages sent from vMEPs configured on the SPB B-VPLS will use the default tree. The default tree does not consume MFIB entries and can be checked in each node through the use of the following command:

```
*A:PE-5# tools dump service id 10 spb default-multicast-list
saps : { }
spoke-sdps : { 52:10 }
```

PE-5 is not part of the tree for I-VPLS 11. However, as with any SPB node part of B-VPLS 10, PE-5 is part of the default tree. Refer to Configuration of ISID-Policies in SPB B-VPLS to see more use-cases for the default tree.

The following tools commands allow the operator to easily see the forwarding path (unicast and multicast) followed by the traffic to a remote node, with the aggregate metric from the source.

```
*A:PE-1# tools dump service id 10 spb fid 10 forwarding-path destination PE-4
forwarding-tree unicast
Hop  BridgeId                Metric From Src
0    PE-1                    0
1    PE-2                    10
2    PE-3                    20
3    PE-4                    30

*A:PE-1# tools dump service id 10 spb fid 10 forwarding-path destination PE-4
forwarding-tree multicast
Hop  BridgeId                Metric From Src
0    PE-1                    0
1    PE-2                    10
2    PE-3                    20
3    PE-4                    30
```

In large networks or networks where IP multicast, PBB and PBB-SPB services coexist, the data plane MFIB entries is a hardware resource that should be periodically checked. The tools dump service vpls-mfib-stats shows the total number of hardware MFIB entries (40k entries, in chassis-mode d) and the entries being used by IP multicast or PBB (MMRP or SPB). The tools dump service vpls-pbb-mfib-stats shows the breakdown between MFIB entries populated by MMRP, SPB, or by EVPN, and the individual limits, system-wide and per service:

```
*A:PE-2# tools dump service vpls-mfib-stats
Service Manager VPLS MFIB info at 000 00:45:28.450:
```

```
Statistics last cleared at 000 00:00:00.000

            Statistic            |    Count
---------------------------------+-------------
             HW limit SG entries |      40959   # total number of MFIB entries
              Current SG entries |          1
       Limit Non PBB SG entries  |      16383   # IP Multicast MFIB limit
     Current Non PBB SG entries  |          0
---snip---


*A:PE-2# tools dump service vpls-pbb-mfib-stats detail

Service Manager VPLS PBB MFIB statistics at 000 00:45:28.540:

Usage per Service
   ServiceId    MFIB User      Count
  ------------+--------------+-------
   10           spb                1
  ------------+--------------+-------
                      Total       1

MMRP
  Current Usage     :     0
  System Limit      :  8191 Full, 40959 ESOnly
  Per Service Limit :  2048 Full,  8192 ESOnly

SPB
  Current Usage     :     1
  System Limit      :  8191
  Per Service Limit :  8191

Evpn
  Current Usage     :     0
  System Limit      : 40959
  Per Service Limit :  8191
```

Finally, the following debug commands can help monitor the SPB IS-IS process and the protocol PDU exchanges:

- debug service id <svcId> spb
- debug service id <svcId> spb adajacency
- debug service id <svcId> spb interface
- debug service id <svcId> spb l2db
- debug service id <svcId> spb lsdb
- debug service id <svcId> spb packet <detail>
- debug service id <svcId> spb spf

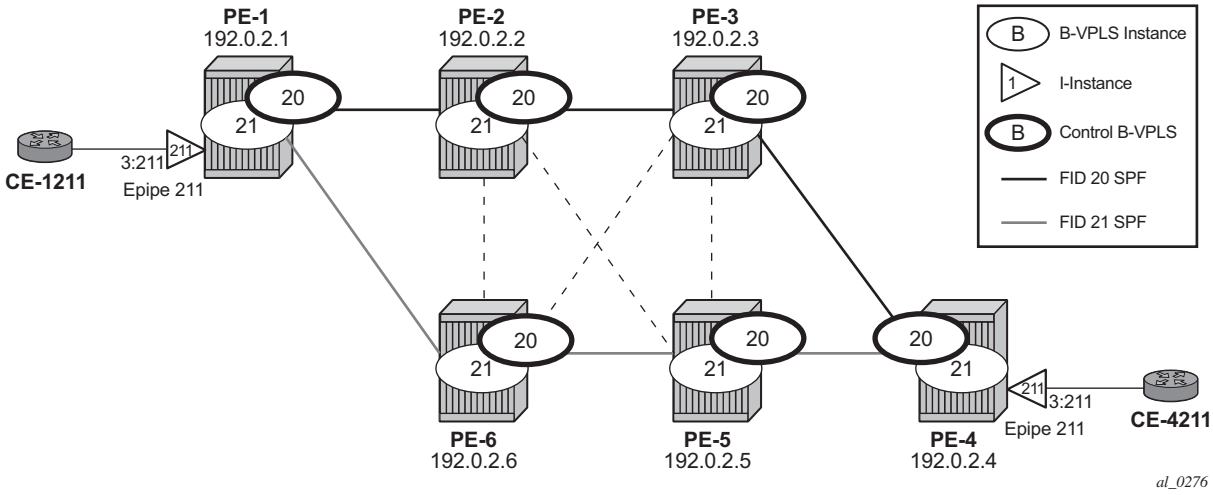
Control and User B-VPLS Configuration

The SR OS implementation of SPB allows a single SPB IS-IS instance to control the paths and FDBs of many B-VPLS instances. This is done by using the control B-VPLS, user B-VPLS and fate-sharing concepts.

The control B-VPLS will be SPB-enabled and configured with all the related SPB IS-IS parameters. Although the control B-VPLS might or might not have I-VPLS/Epipes directly attached, it must be configured on all the nodes where SPB forwarding is expected to be active. SPB uses the logical instance and a forwarding ID (FID) to identify SPB locally on the node. That FID must be consistently configured on all the nodes where the B-VPLS exists. User B-VPLS are other instances of B-VPLS that are usually configured to separate the traffic for manageability reasons, QoS or ECT different treatment.

Figure 316 illustrates the control B-VPLS (B-VPLS 20) and user B-VPLS (B-VPLS 21) concept (in this case there is only one user B-VPLS but there might be many B-VPLS sharing fate with the same control B-VPLS). Both B-VPLSs must share the same topology and both B-VPLSs must share exactly the same interfaces. The user B-VPLS, which is linked to the control B-VPLS by its FID, follows (that is, inherits the state of) the control B-VPLS but may use a different ECT path in case of equal metric paths, like in this example: FID 20, that is, the control B-VPLS, follows the low-path-id ECT, whereas FID 21, for example, the user B-VPLS, follows the high-path-id ECT.

Figure 316 **Control and User B-VPLS Test Topology**



The configurations of B-VPLSs 20 and 21, on PE-1 and PE-2, are as follows. The **spbm-control-vpls 20 fid 21 in B-VPLS 21** command associates FID 21 to the user B-VPLS and links the B-VPLS to its control B-VPLS 20.

```
*A:PE-1# configure service
        vpls 20 customer 1 b-vpls create
            service-mtu 2000
            spb 1025 fid 20 create
                level 1
                    ect-algorithm fid-range 21-4095 high-path-id
                exit
                no shutdown
            exit
            spoke-sdp 12:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 16:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        vpls 21 customer 1 b-vpls create
            service-mtu 2000
            spbm-control-vpls 20 fid 21
            spoke-sdp 12:21 create
                no shutdown
            exit
            spoke-sdp 16:21 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 211 customer 1 create
            pbb
                tunnel 21 backbone-dest-mac "PE-4" isid 211
            exit
            sap 1/1/3:211 create
            exit
            no shutdown
        exit


*A:PE-2# configure service
        vpls 20 customer 1 b-vpls create
            service-mtu 2000
            spb 1025 fid 20 create
                level 1
                    ect-algorithm fid-range 21-4095 high-path-id
                exit
                no shutdown
            exit
            spoke-sdp 21:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
```

```
        spoke-sdp 23:20 create
            spb create
                no shutdown
            exit
            no shutdown
        exit
        spoke-sdp 25:20 create
            spb create
                no shutdown
            exit
            no shutdown
        exit
        spoke-sdp 26:20 create
            spb create
                no shutdown
            exit
            no shutdown
        exit
        no shutdown
    exit
    vpls 21 customer 1 b-vpls create
        service-mtu 2000
        spbm-control-vpls 20 fid 21
        spoke-sdp 21:21 create
            no shutdown
        exit
        spoke-sdp 23:21 create
            no shutdown
        exit
        spoke-sdp 25:21 create
            no shutdown
        exit
        spoke-sdp 26:21 create
            no shutdown
        exit
        no shutdown
    exit
```

If there is a mismatch between the topology of a user B-VPLS and its control B-VPLS, only the user B-VPLS links and nodes that are in common with the control B-VPLS will function.

User B-VPLS instances supporting only unicast services (PBB-Epipes) may share the FID with the other B-VPLS (control or user). This is a configuration shortcut that reduces the LSP advertisement size for B-VPLS services but results in the same separation for forwarding between the B-VPLS services. In the case of PBB-Epipes only BMACs are advertised per FID but BMACs are populated per B-VPLS in the FIB. If I-VPLS services are to be supported on a B-VPLS that B-VPLS must have an independent FID.

Although user B-VPLS 21 does not have any SPB setting (other than the spbm-control-vpls), the spoke-sdps use the same SDPs as the parent control B-VPLS 20. The **show service id <user b-vpls> spb fate-sharing** command shows the control spoke-sdp/saps that control the user spoke-sdp/saps.

```
*A:PE-1# show service id 21 spb fate-sharing
===============================================================================
User service fate-shared sap/sdp-bind information
===============================================================================
Control   Control Sap/             FID       User      User Sap/
SvcId     SdpBind                            SvcId     SdpBind
-------------------------------------------------------------------------------
20        12:20                    21        21        12:21
20        16:20                    21        21        16:21
===============================================================================
*A:PE-1#
```

SPBM Access Resiliency Configuration

The following example shows how to configure an I-VPLS/Epipe attached to an SPB-enabled B-VPLS when access resiliency is used.

Multi-chassis LAG (MC-LAG) is the only resiliency mechanism supported for PBB-Epipes. The MC-LAG active node will advertise the MC-LAG BMAC (or sap-bmac) in SPB IS-IS. In case of failure, when the standby node takes over, it will advertise the MC-LAG sap-bmac. Without SPB, the MC-LAG solution for PBB-Epipe required the use of mac-notification and periodic mac-notification. SPB provides a faster and more efficient solution without the need for any extra mac-notification mechanism. In the example described in this section, Epipe 31 uses MC-LAG access resiliency to get connected to the B-VPLS 30 on nodes PE-2 and PE-6.

As far as I-VPLS access resiliency is concerned, the same mechanisms supported for regular B-VPLS are supported for SPB-enabled B-VPLS, except for G.8032. A very important aspect of the I-VPLS resiliency is a proper mac-flush propagation when there is a failure at the I-VPLS access links.

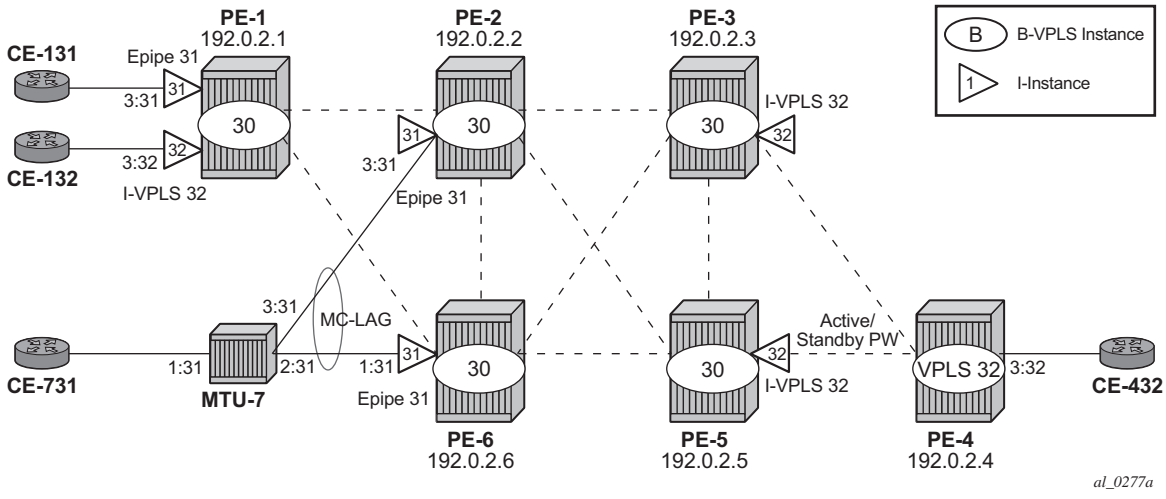
If the SPB-enabled B-VPLS uses B-SAPs for its connectivity to the backbone, there is no mac-flush propagation (because there is no TLDP). In this case, if MC-LAG is used and there is an MC-LAG switchover, the new active chassis will keep using the same source BMAC, such as the sap-bmac, and it will advertise it in the B-VPLS domain so that the remote FDBs can be properly updated. No mac-flush is required in this case.

When the B-VPLS uses spoke-sdps for its backbone connectivity, the traditional LDP MAC flush propagation mechanisms and commands can be used as follows:

- **send-flush-on-failure** works as expected when SPB is used at the B-VPLS. When configured, a flush-all-from-me event is triggered upon a SAP or spoke-sdp failure in the I-VPLS.
- **send-bvpls-flush** works as expected when SPB is used at the B-VPLS. Two variants are configurable: all-from-me/all-but-mine. Any I-VPLS SAP/spoke-sdp failure is propagated to the I-VPLS on the peers to flush their respective customer MACs (CMACs). It works only in conjunction with send-flush-on-failure configuration on I-VPLS. The associated ISID list is passed along with the LDP mac-flush message, which is flushed/retained according to the **all-from-me/all-but-me** flag.
- **send-flush-on-bvpls-failure** works as expected when SPB is used at the B-VPLS. A local B-VPLS failure is propagated to the I-VPLS, which then triggers a LDP mac-flush if it has any spoke-sdp on it.
- **propagate-mac-flush-from-bvpls** does not work when SPB is used at the B-VPLS (because failures within the B-VPLS are handled by SPB) and its configuration is blocked.

In the example described later in this section, I-VPLS 32 uses active/standby spoke-sdp resiliency to get connected to the B-VPLS 30 on nodes PE-3 and PE-5.

Figure 317 **Access Resiliency Test Topology**



As an example of MC-LAG connectivity, the Epipe 31 configuration is shown. Just like for regular PBB-VPLS, a sap-bmac is used as source BMAC for the Epipe traffic from PE-2/PE-6 to PE-1. A sap-bmac is a virtual BMAC formed from the configured source-bmac plus the MC-LAG LACP-key (if configured this way) and owned by the MC-LAG active chassis. The following CLI output shows the configuration of MC-

LAG as well as the generation of the sap-bmac. Once it is properly configured and the MC-LAG and Epipe are up and running, SPB IS-IS will distribute the sap-bmac throughout the B-VPLS, as it does for the system BMACs and OAM vMEP MACs. In this example, PE-2 is the MC-LAG active node, therefore the sap-bmac for Epipe 31 is generated from PE-2.

```
*A:PE-2# configure lag 1
        mode access
        encap-type dot1q
        port 1/1/3
        lacp active administrative-key 32768
        no shutdown
        exit

*A:PE-2# configure redundancy
        multi-chassis
            peer 192.0.2.6 create
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:02:06 system-priority 65535
                                     source-bmac-lsb use-lacp-key
                    no shutdown
                exit
                no shutdown
            exit
        exit

*A:PE-2# configure service
        vpls 30 customer 1 b-vpls create
            service-mtu 2000
            pbb
                use-sap-bmac
            exit
            spb 1026 fid 30 create
                level 1
                    bridge-priority 0
                exit
                no shutdown
            exit
            spoke-sdp 21:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 23:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 25:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 26:30 create
                spb create
```

```
                no shutdown
            exit
            no shutdown
        exit
        no shutdown
    exit
    epipe 31 customer 1 create
        pbb
            tunnel 30 backbone-dest-mac "PE-1" isid 31
        exit
        sap lag-1:31 create
        exit
        no shutdown
    exit
```

```
*A:PE-6# show service id 30 spb fdb
===============================================================================
User service FDB information
===============================================================================
MAC Addr          UCast Source          State   MCast Source          State
-------------------------------------------------------------------------------
00:00:00:01:01:01 61:30                 ok      62:30                 ok
00:00:00:02:00:01 62:30                 ok      62:30                 ok
00:00:00:02:02:02 62:30                 ok      62:30                 ok
00:00:00:03:03:03 63:30                 ok      62:30                 ok
00:00:00:05:05:05 65:30                 ok      62:30                 ok
-------------------------------------------------------------------------------
Entries found: 5
===============================================================================
```

The configuration for I-VPLS 32 on nodes PE-4 and PE-3 is as follows.

```
*A:PE-4# configure service
    vpls 32 customer 1 create     # Ordinary VPLS, no I-VPLS (no B-VPLS present)
        endpoint "CORE" create
            no suppress-standby-signaling
        exit
        sap 1/1/3:32 create
        exit
        spoke-sdp 43:32 endpoint "CORE" create
            precedence primary
            no shutdown
        exit
        spoke-sdp 45:32 endpoint "CORE" create
            no shutdown
        exit
        no shutdown
    exit

*A:PE-3# configure service
    vpls 30 customer 1 b-vpls create
        service-mtu 2000
        spb 1026 fid 30 create
            no shutdown
        exit
        spoke-sdp 32:30 create
```

```
            spb create
                no shutdown
            exit
            no shutdown
        exit
        spoke-sdp 35:30 create
            spb create
                no shutdown
            exit
            no shutdown
        exit
        spoke-sdp 36:30 create
            spb create
                no shutdown
            exit
            no shutdown
        exit
        no shutdown
    exit
    vpls 32 customer 1 i-vpls create
        send-flush-on-failure
        pbb
            backbone-vpls 30
            exit
            send-flush-on-bvpls-failure
            send-bvpls-flush all-from-me
        exit
        spoke-sdp 34:32 create
            no shutdown
        exit
        no shutdown
    exit
```

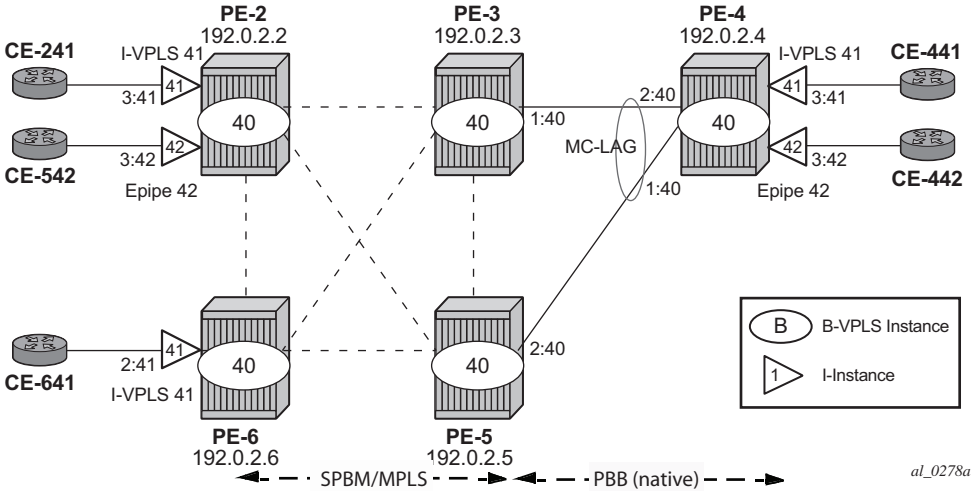
As discussed, **send-flush-on-failure** and **send-bvpls-flush all-from-me** are configured in the I-VPLS. When the active spoke-sdp goes down on PE-3, a flush-all-from-me message will be propagated through the backbone and will flush the corresponding CMACs associated to the I-VPLS 32 in node PE-1. MAC flush-all-from-me messages are automatically propagated in the core up to the remote I-VPLS 32 on node PE-1 (there is no need for any propagate-mac-flush in the intermediate nodes). Note that the *send-flush-on-bvpls-failure* command works as expected. The command *propagate-mac-flush-from-bvpls* is never used when the B-VPLS is SPB-enabled (the command is not allowed).

Static BMACs and Static ISIDs Configuration

SR OS supports the interworking between SPB-enabled B-VPLS and non-SPB B-VPLS instances. SPB networks can be connected to non-SPB capable nodes, for example third party vendor PBB switches or 7210 SAS nodes. This is possible through the use of conditional static BMACs and static ISIDs on the nodes doing the interworking function. Conditional static BMACs and static ISIDs can be associated to non-SPB B-VPLS SAPs or spoke-sdps.

The following example shows an SPB-enabled B-VPLS (40) on nodes PE-2, PE-6, PE-3 and PE-5. Node PE-4 supports PBB, but not SPB and it is connected by a MC-LAG to nodes PE-3 and PE-5. Services I-VPLS 41 and Epipe 42 have end-points on node PE-4. In this example, nodes PE-3 and PE-5 are acting as interworking nodes. They will be configured with the BMAC of PE-4 so that the MC-LAG active node advertises the non-SPB capable node BMAC into SPB IS-IS. The BMAC will be configured as a conditional static BMAC so that an SPB node, such as PE-3 or PE-5, will only advertise PE-4's BMAC if its connection to PE-4 is active. Besides the conditional static BMAC, nodes PE-3/PE-5 should advertise the I-VPLS ISIDs defined in PE-4. Epipe ISIDs are not advertised in SPB IS-IS, therefore it is not necessary to create a static ISID for Epipe 42.

Figure 318 **Access Resiliency Test Topology**



The commands to configure conditional static BMACs and static ISIDs are as follows.

```
*A:PE-3# configure service vpls 40 static-mac mac
  - mac <ieee-address> [create] sap <sap-id> monitor {fwd-status}
  - no mac <ieee-address>
  - mac <ieee-address> [create] spoke-sdp <sdp-id:vc-id> monitor {fwd-status}
```

```
*A:PE-3# configure service vpls 40 sap lag-1:40 static-isid range
  - no range <range-id>
  - range <range-id> isid <isid-value> [to <isid-value>] [create]

 <range-id>           : [1..8191]
 <isid-value>         : [1..16777215]
 <create>             : keyword
```

The **monitor fwd-status** attribute identifies this to be a conditional MAC and is mandatory for static BMACs. This parameter instructs SR OS to advertise the BMAC only if the corresponding SAP/spoke-sdp is in forwarding state.

The configuration of the conditional static BMAC and static ISID is as follows. The values for spf-wait are the default ones.

```
*A:PE-3# configure service
       vpls 40 customer 1 b-vpls create
            service-mtu 2000
            spb 1027 fid 40 create
                timers
                    spf-wait 10000 spf-initial-wait 10 spf-second-wait 1000
                exit
                no shutdown
            exit
            sap lag-1:40 create
                static-isid
                    range 1 create isid 41
                exit
            exit
            spoke-sdp 32:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 35:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 36:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            static-mac
                mac 00:00:00:04:04:04 create sap lag-1:40 monitor fwd-status
            exit
            no shutdown
        exit

*A:PE-5# configure service
        vpls 40 customer 1 b-vpls create
            service-mtu 2000
```

```
            spb 1027 fid 40 create
                timers
                    spf-wait 10000 spf-initial-wait 10 spf-second-wait 1000
                exit

                no shutdown
            exit
            sap lag-1:40 create
                static-isid
                    range 1 create isid 41
                exit
            exit
            spoke-sdp 52:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 53:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 56:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            static-mac
                mac 00:00:00:04:04:04 create sap lag-1:40 monitor fwd-status
            exit
            no shutdown
        exit
```

The configuration of the conditional static BMAC is different from the legacy static-mac command, configured within the sap/sdp-binding context. The latter static-mac is not conditional and it is always added to the FDB. The conditional static BMAC is added to the FDB based on the SAP/sdp-binding state (the conditional static BMAC is tagged in the FDB as **CStatic**, for Conditional Static).

```
*A:PE-3# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      up      No       0         1        active
===============================================================================


*A:PE-3# show service id 40 fdb pbb
===============================================================================
Forwarding Database, b-Vpls Service 40
===============================================================================
```

```
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
------------------------------------------------------------------------
00:00:00:02:02:02 sdp:32:40             0          0          Spb
00:00:00:04:04:04 sap:lag-1:40          0          0          CStatic
00:00:00:05:05:05 sdp:35:40             0          0          Spb
00:00:00:06:06:06 sdp:36:40             0          0          Spb
========================================================================


*A:PE-5# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      down    No       0         0        standby
===============================================================================


*A:PE-5# show service id 40 fdb pbb
========================================================================
Forwarding Database, b-Vpls Service 40
========================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
------------------------------------------------------------------------
00:00:00:02:02:02 sdp:52:40             0          0          Spb
00:00:00:03:03:03 sdp:53:40             0          0          Spb
00:00:00:04:04:04 sdp:53:40             0          0          Spb
00:00:00:06:06:06 sdp:56:40             0          0          Spb
========================================================================
```

The static-isid command identifies a set of ISIDs for I-VPLS services that are external to SPBM. These ISIDs are advertised as supported locally on this node unless altered by an isid-policy. Although the preceding example shows the use of the static-isid associated to a MC-LAG SAP, regular SAPs or spoke SDPs are also supported. ISIDs declared in this way become part of the ISID multicast and consume MFIBs. Multiple SPBM static-isid ranges are allowed under a SAP/spoke SDP. ISIDs are advertised as if they were attached to the local BMAC. Only remote I-VPLS ISIDs need to be defined. In the MFIB, the backbone group MACs are then associated with the active SAP or spoke SDP.

Once the conditional static BMAC for PE-4 and the static-isid 41 (for I-VPLS 41) are configured as discussed, the advertised BMAC and ISID can be checked in the remote SPB nodes:

```
*A:PE-6# show service id 40 spb fdb
===============================================================================
User service FDB information
===============================================================================
MAC Addr          UCast Source          State   MCast Source          State
-------------------------------------------------------------------------------
00:00:00:02:02:02 62:40                 ok      62:40                 ok
00:00:00:03:03:03 63:40                 ok      62:40                 ok
00:00:00:04:04:04 63:40                 ok      62:40                 ok
```

```
00:00:00:05:05:05 65:40                 ok      62:40                 ok
-------------------------------------------------------------------------------
Entries found: 4
===============================================================================


*A:PE-6# show service id 40 spb mfib
===============================================================================
User service MFIB information
===============================================================================
MAC Addr          ISID     Status
-------------------------------------------------------------------------------
01:1E:83:00:00:29 41       Ok
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================


*A:PE-6# show service id 40 mfib
===============================================================================
Multicast FIB, Service 40
===============================================================================
Source Address  Group Address         Sap/Sdp Id               Svc Id   Fwd/Blk
-------------------------------------------------------------------------------
*               01:1E:83:00:00:29     b-sdp:62:40              Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The group address terminates in hex 29, which corresponds to ISID 41.

The configured static-isids can be displayed with the following command (a range 41-100 has been added to the sap lag-1:40 to demonstrate this output):

```
*A:PE-5# configure service
        vpls 40
            sap lag-1:40 create
                static-isid
                    range 1 create isid 41 to 100
                exit
            exit


*A:PE-5# show service id 40 sap lag-1:40 static-isids
===================================
Static Isid Entries
===================================
Entry          Range
-----------------------------------
1              41-100
===================================
```

Configuration of ISID-Policies in SPB B-VPLS

ISID policies are an optional aspect of SPBM which allow additional control of the advertisement of ISIDs and creation of MFIB entries for I-VPLS (Epipe services do not trigger ISID advertisements or the creation of MFIB entries). By default, if no ISID-policies are used, SPBM automatically advertises and populates MFIB entries for I-VPLS and static-isids. ISID-policies can be used on any SPB-enabled node with locally defined I-VPLS instances or static-isids. The isid-policy parameters are shown below:

```
A:PE-3# configure service vpls 40 isid-policy entry ?
  - entry <range-entry-id> [create]
  - no entry <range-entry-id>

 <range-entry-id>     : [1..8191]
 <create>             : keyword


 [no] advertise-local - Configure local advertisement of the range
 [no] range           - Configure ISID range for the entry
 [no] use-def-mcast   - Use default multicast tree to propogate ISID range
```

Where:

- **advertise-local** defines whether the local ISIDs (I-VPLS ISIDs linked to the B-VPLS) or static ISIDs contained in the configured range are advertised in SPBM.
- **use-def-mcast** controls whether the ISIDs contained in the range use MFIB entries (if **no use-def-mcast** is used) or just the default tree which does not use any MFIB entry.

The **isid-policy** becomes active as soon as it is defined, as opposed to other policies in SR OS, which require the policy itself to be applied within the configuration.

The typical use of ISID-policies is to reduce the number of ISIDs being advertised and/or to save MFIB space (in deployments where MFIB space is shared with MMRP and IP Multicast). The use of ISID-policies is recommended for I-VPLS where most of the traffic is unicast or for I-VPLS where the ISID end-points are present in all the backbone edge bridges (BEBs) of the SPB network. In both cases, advertising ISIDs or consuming MFIB entries for those I-VPLSs has little value since no multicast (first case) or the default tree (second case) are as efficient as using MFIB entries.

The following configuration example will use the example topology in Figure 318. In this case, the objective of the isid-policy will be to use the default tree for all the I-VPLS services with ISIDs between 41 and 100, excluding the range 80-90. The following example shows the policy configuration in PE-3. The same policy will be configured in the rest of the SPB nodes, that is, PE-2, PE-6 and PE-5.

```
*A:PE-3# configure service
        vpls 40
            isid-policy
                entry 10 create
                    range 80 to 90
                exit
                entry 20 create
                    use-def-mcast
                    no advertise-local
                    range 41 to 79
                exit
                entry 30 create
                    use-def-mcast
                    no advertise-local
                    range 91 to 100
                exit
            exit
```

The **no advertise-local** option can only be configured if the **use-def-mcast** option is also configured.

```
*A:PE-3# configure service vpls 40 isid-policy entry 40 create no advertise-local
MINOR: SVCMGR #7855 Cannot set AdvLocal for entry - advertise-local or use-def-mcast
option must be specified
```

Overlapping ISID values can be configured as long as the actions are consistent for the same ISID. Conflicting actions are shown in the CLI.

```
*A:PE-3# configure service vpls 40 isid-policy entry 40 create
*A:PE-3>config>service>vpls>isid-policy>entry# range 82 to 85
*A:PE-3>config>service>vpls>isid-policy>entry# use-def-mcast
MINOR: SVCMGR #7854 Cannot set UseDefMctree for entry - Conflicting Actions with Entry-
10
```

The isid-policy configured for B-VPLS 40 in all the four nodes makes the SPB network to use the default tree for ISIDs 41-79 and 91-100 and not advertise those ISIDs in SPB IS-IS even if the ISID is locally defined (as in the case for ISIDs 41-100 in PE-3). As discussed in Basic SPBM Configuration, the default tree path can be checked from each node by using the **tools dump service id 40 spb default-multicast-list** command.

Due to entry 10 in the policy, ISIDs 80-90 will be advertised by PE-3 (active MC-LAG node). However, nodes PE-2 and PE-6 will not create any MFIB entry for those ISIDs until the corresponding I-VPLS ISIDs are locally created (or configured through static-isids). The following command executed on PE-2 proves that ISIDs 80-90 are indeed being advertised by PE-3:

```
*A:PE-2# show service id 40 spb database detail
===============================================================================
Router Base ISIS Instance 1027 Database
===============================================================================
```

```
Displaying Level 1 database
------------------------------------------------------------------------------
LSP ID    : PE-2.00-00                                  Level     : L1
---snip---

------------------------------------------------------------------------------
LSP ID    : PE-3.00-00                                  Level     : L1
---snip---

TLVs :
---snip---
  MT Capability :
    TLV Len        : 56
    MT ID          : 0
    SPBM Service ID:
    Sub TLV Len    : 52
      BMac Addr              : 00:00:00:03:03:03
      Base VID               : 40
      ISIDs                  :
        80       Flags:TR
        81       Flags:TR
        82       Flags:TR
        83       Flags:TR
        84       Flags:TR
        85       Flags:TR
        86       Flags:TR
        87       Flags:TR
        88       Flags:TR
        89       Flags:TR
        90       Flags:TR

------------------------------------------------------------------------------
LSP ID    : PE-5.00-00                                  Level     : L1
---snip---
==============================================================================
```

The **mfib** parameter in the **show service id 40 sap static-isids mfib** command can help understand the state of the MFIB entries added (or not) by the configured static-isid. The following possible states can be shown:

- If the static-isid is configured and programmed in the mfib the status is shown as:
  - ok
- If the static-isid is not configured and not programmed in the mfib, the reasons can be (order of priority):
  - useDefMCTree - isid policy is applied on the service for the isid.
  - sysMFibLimit - system mfib limit has been exceeded
  - addPending - adding pending due to processing delays
- If the static-isid is not configured but present in the mfib:
  - delPending - cleanup pending due to processing delays.

The following examples show some of these possible states:

```
*A:PE-5# show service id 40 sap lag-1:40 static-isids mfib
===================================
ISID Detail
===================================
ISID           Status
-----------------------------------
---snip---
80             ok
81             ok
---snip---
41             useDefMCTree
42             useDefMCTree
---snip---
8292           sysMFibLimit
8293           sysMFibLimit
---snip---
9253          addPending
9254          addPending
---snip---
===================================
```

Conclusion

SR OS supports an efficient SPBM implementation in the context of a B-VPLS, where system BMACs, vMEP OAM BMACs and SAP-BMACs are advertised in SPB IS-IS. SPBM provides a simple solution where no other control plane protocol is required in the B-VPLS to take care of the resiliency, load-balancing and multicast optimization. The SPBM implementation in the SR OS provides scale optimization through the use of control and user B-VPLSs, allows the interworking between SPB networks and PBB networks, as well as the optimization of the MFIB resources and advertisement of ISIDs through the use of ISID-policies.

Layer 3 Services

In This Section

This section provides configuration information for the following topics:

- Carrier Supporting Carrier IP VPNs
- Layer 3 VPN: VPRN Type Spoke
- NG-MVPN Configuration with MPLS
- NG-MVPN Configuration with PIM
- NG-MVPN Sender-Only, Receiver-Only
- NG-MVPN Source Redundancy
- NG-MVPN Wildcard S-PMSI
- Rosen MVPN Core Diversity
- Rosen MVPN Inter-AS Option B
- Spoke Termination for IPv6-6VPE
- Traffic Leaking from VPRN to GRT
- VPRN Inter-AS VPN Model C

Carrier Supporting Carrier IP VPNs

This chapter provides information about carrier supporting carrier IP VPN configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was originally written for and tested on 11.0.R1. The CLI in this guide corresponds to release 14.0.R5. The configuration for labeled IPv4 routes changed in SR OS release 14.0.R4, see chapter *Separate BGP RIBs for Labeled Routes*.

Overview

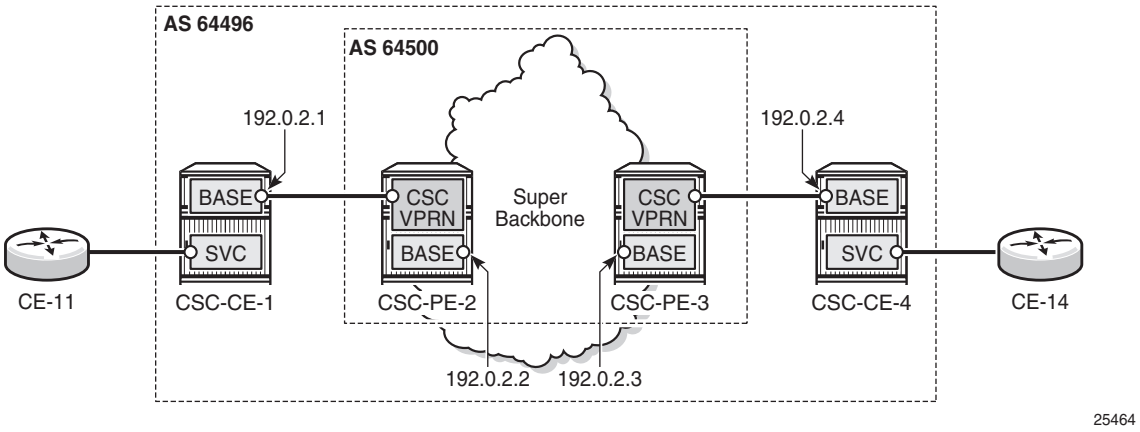
Carrier Supporting Carrier (CSC) is a solution that allows one service provider (the Customer Carrier) to use the IP VPN service of another service provider (the Super Carrier) for some or all of its backbone transport. RFC 4364 defines a Carrier Supporting Carrier solution for BGP/MPLS IP VPNs that uses MPLS at the interconnection points between the two service providers to provide a scalable and secure solution.

A simplified CSC network topology is shown in Figure 319. A CSC deployment involves the following types of devices:

- CE — Customer premises equipment dedicated to one particular business/ enterprise.
- PE — Edge router managed and operated by the Customer Carrier that connects to CEs to provide business VPN or Internet services.
- CSC-CE — Peering router managed and operated by the Customer Carrier that is connected to CSC-PEs for purposes of using the associated CSC IP VPN services for backbone transport. The CSC-CE may attach directly to CEs if it is also configured to be a PE for business VPN services.

- CSC-PE — A PE router managed and operated by the Super Carrier that supports one or more CSC IP VPN services possibly in addition to other traditional PE services.

Figure 319 **CSC Network Topology**



In the CSC solution, the CSC-CE and CSC-PE are directly connected by a link that supports MPLS. The CSC-CE distributes an MPLS label for every /32 IPv4 prefix it and any downstream PE uses as a BGP next-hop in routes associated with services offered by the Customer Carrier. BGP *must be used* as the label distribution protocol between CSC-CE and CSC-PE if the latter device is a 7x50. Typically, the Customer Carrier and Super Carrier operate as two different Autonomous Systems (AS) and therefore BGP, more specifically EBGP, is the best label distribution protocol, even if other options are available. The BGP session between CSC-CE and CSC-PE must be single-hop EBGP (or IBGP) if either device is a 7x50.

In a 7x50 CSC-PE, the interface to a CSC-CE is a special type of IP/MPLS interface that belongs to a VPRN configured for CSC mode. This special type of interface is called a CSC VPRN interface throughout the remainder of this chapter. The CSC VPRN interface has many of the same characteristics as a network interface of the base router but its association with a VRF ensures that the traffic and control plane routes of the Customer Carrier are kept separate from other services.

When a 7x50 CSC-PE receives a labeled-IPv4 route (with label L1, next-hop N1) from a CSC-CE BGP peer, the following actions take place in the CSC-PE:

1. The BGP route is installed into the routing table of the CSC VPRN (assuming the BGP route is the best route to the destination).
2. If the BGP route matches the VRF export policy, it is advertised to the core MP-BGP peers as a VPN-IPv4 route. The advertised label value is changed to L2.

3. BGP programs the line cards with an MPLS forwarding entry that swaps L2 for L1 and sends the MPLS packet over the CSC VPRN interface associated with next-hop N1.

When a 7x50 CSC-PE receives a VPN-IPv4 route (with label L2, next-hop N2) the following actions take place in the CSC-PE:

1. If the VPN-IPv4 route matches the VRF import policy of a CSC VPRN, it is installed into the routing table of that CSC VPRN.
2. If the imported (BGP-VPN) route matches the BGP export policy associated with a CSC-CE BGP peer, it is advertised to that peer as a labeled-IPv4 route. The advertised label value is changed to L3.
3. BGP programs the line cards with an MPLS forwarding entry that swaps L3 for L2 and sends the packet inside the MPLS tunnel to next-hop N2.

Once a CSC-CE has learned a labeled-IPv4 route for a remote CSC-CE and vice versa, the two CSC-CEs can set up a BGP session between themselves and exchange VPN routes over this session if they are both PEs with services. Typically this BGP session will be an IBGP session because the local and remote CSC-CEs belong to the same Autonomous System (AS). The Layer 2 VPN and Layer 3 VPN routes exchanged by the CSC-CEs are resolved by the labeled-IPv4 routes they have for each other’s /32 IPv4 address.

Configuration

This section will walk through the steps to configure the CSC solution shown in Figure 319. The IPv4 addresses in Figure 319 are the system IP addresses of the routers. The steps are the following:

- Configure CSC-CE-1
- Configure CSC service on CSC-PE-2
- Verify exchange of labeled IPv4 routes between CSC-CE-1 and CSC-PE-2
- Configure core connectivity for CSC-PE-2
- Configure core connectivity for CSC-PE-3
- Configure CSC service on CSC-PE-3
- Verify exchange of VPN-IPv4 routes between CSC-PE-2 and CSC-PE-3
- Configure CSC-CE-4
- Verify exchange of labeled IPv4 routes between CSC-PE-3 and CSC-CE-4
- Configure BGP session between CSC-CE-1 and CSC-CE-4
- Verify exchange of VPN-IPv4 routes between CSC-CE-1 and CSC-CE-4

Step 1. Configure CSC-CE-1.

This example assumes that CSC-CE-1 is a PE router with Layer 2 and Layer 3 VPN services that must extend across the CSC VPN service; assume that there are no further downstream PEs in AS 64496. The configuration of one such Layer 3 VPN service in CSC-CE-1 is shown below:

```
*A:CSC-CE-1# configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:11
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64496:1
            interface "loopback-1" create
                address 10.11.30.2/24
                loopback
            exit
            ---snip---
```

For brevity, the above configuration sample omits commands related to SAP IP interfaces, spoke-SDP IP interfaces, PE-CE routing protocols, QoS, IP filters, etc. The loopback interface is used to test whether this prefix will be learned at the remote CSC-CE-4.

The base routing instance of the CSC-CE should be configured with the appropriate router-ID and autonomous-system number and the system interface should be given an IPv4 address (usually the same as the router-id). If the router-ID is not configured, by default the system IP address is used as the router-ID. The interface to CSC-PE-2 should then be created and configured. The base router configuration of CSC-CE-1 is shown below:

```
*A:CSC-CE-1>config>router# info
----------------------------------------------
#------------------------------------------------
echo "IP Configuration"
#------------------------------------------------
        interface "int-CSC-CE-1-CSC-PE-2"
            address 192.168.12.1/30
            port 1/1/1
            no shutdown
        exit
        interface "system"
            address 192.0.2.1/32
            no shutdown
        exit
        autonomous-system 64496
----------------------------------------------
*A:CSC-CE-1>config>router#
```

BGP should be configured as the control plane protocol running on the interface to CSC-PE-2, as follows:

```
*A:CSC-CE-1# configure
    router
        bgp
            group "CSC-PE"
                peer-as 64500
                neighbor 192.168.12.2
                    family label-ipv4
                    export "static-to-BGP"
                    split-horizon
                exit
            exit
            no shutdown
        exit
    exit
*A:CSC-CE-1#
```

The BGP configuration of CSC-CE-1 is as follows:

– The peer type is EBGP (**peer-as** is different from the locally configured **autonomous-system)**

– The address family for the EBGP session is **label-ipv4** (the **neighbor** address is an IPv4 address). Family label-IPv4 causes MP-BGP negotiation of the address family for AFI=1 and SAFI=4 (IPv4 NLRI with MPLS labels), as can be observed from the following debug trace (using the command **debug router bgp open**) of the OPEN message from CSC-CE-1.This message can obviously only be seen when the BGP peer is up. The configuration for CSC-PE-2 will be shown later, but in order to have the trace message, it must be configured already.

```
1 2016/10/17 12:45:37.46 UTC MINOR: DEBUG #2001 Base BGP
"BGP: OPEN
Peer 1: 192.168.12.2 - Send (Active) BGP OPEN: Version 4
   AS Num 64496: Holdtime 90: BGP_ID 192.0.2.1: Opt Length 16
   Opt Para: Type CAPABILITY: Length = 14: Data:
     Cap_Code MP-BGP: Length 4
       Bytes: 0x0 0x1 0x0 0x4
     Cap_Code ROUTE-REFRESH: Length 0
     Cap_Code 4-OCTET-ASN: Length 4
       Bytes: 0x0 0x0 0xfb 0xf0
"
```

– The **split-horizon** command is optional. It prevents a best BGP route from the CSC-PE peer from being re-advertised back to that peer.

– The **export** command applies a BGP export policy to the session.

The configuration of the policy is as follows:

```
*A:CSC-CE-1# configure
    router
        policy-options
            begin
            prefix-list "system-IP"
                prefix 192.0.2.1/32 exact
            exit
            policy-statement "static-to-BGP"
                entry 10
                    from
                        protocol direct
                        prefix-list "system-IP"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
        exit
```

The purpose of the BGP export policy is to advertise the system IP address of CSC-CE-1 as a labeled-IPv4 BGP route toward the CSC-PE(s).

Step 2. Configure CSC service on CSC-PE-2.

CSC-PE-2 must be configured with a VPRN in **carrier-carrier-vpn** mode in order to provide CSC service to CSC-CE-1. The entire configuration of the VPRN is shown below:

```
*A:CSC-PE-2# configure
    service
        vprn 1 customer 1 create
            carrier-carrier-vpn
            router-id 192.0.2.2
            autonomous-system 64500
            route-distinguisher 64500:12
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64500:1
            network-interface "int-CSC-PE-2-CSC-CE-1" create
                address 192.168.12.2/30
                port 1/1/2
                no shutdown
            exit
            bgp
                group "CSC-CE"
                    as-override
                    export "BGP-VPN-routes"
                    peer-as 64496
                    neighbor 192.168.12.1
                        family label-ipv4
                        split-horizon
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
```

Note the following about the VPRN configuration of CSC-PE-2:

- The **carrier-carrier-vpn** command is mandatory. It cannot be configured if the VPRN currently has any SAP or spoke-SDP access interfaces configured; they must first be shutdown if necessary and then deleted.

```
*A:CSC-PE-2# configure service vprn 1 carrier-carrier-vpn
INFO: PIP #1195 Cannot toggle carrier-carrier-vpn - service interfaces present
*A:CSC-PE-2#
```

- The **auto-bind**-tunnel command should be set appropriately for the type of transport desired to other CSC-PEs, but note that GRE is not supported.

```
*A:CSC-PE-2# configure service vprn 1 auto-bind-tunnel resolution-filter gre
MINOR: SVCMGR #1538 auto-bind config not supported -
 Autobind gre not supported for carrier-carrier vprn
*A:CSC-PE-2#
```

– The interface to CSC-CE-1 must be a **network-interface**. A **network-interface** can be associated with an entire Ethernet port (as shown in the example above), a VLAN sub-interface of an Ethernet port, an entire LAG or a VLAN sub-interface of a LAG. In all cases, the associated Ethernet ports must be configured in network or hybrid mode and must reside on FP2 or higher based cards/systems.

Note the following about the BGP configuration of the CSC VPRN service in CSC-PE-2:

– The peer type is EBGP (**peer-as** is different from the locally configured **autonomous-system**).

– The address family for the EBGP session is **label-ipv4** (the **neighbor** address is an IPv4 address). Address family label-ipv4 causes MP-BGP negotiation of the address family for AFI=1 and SAFI=4 (IPv4 NLRI with MPLS labels).

– The **split-horizon** command is optional. It prevents a best BGP route from the CSC-CE peer from being re-advertised back to that peer.

– The **as-override** command replaces CSC-CE-1's AS number (64496) with CSC-PE-2's AS number (64500) in the AS_PATH attribute of routes advertised to CSC-CE-1. Without this configuration, CSC-CE-1 may reject routes originated by CSC-CE-4 as invalid due to an AS-path loop.

– The **export** command applies a BGP export policy to the session. The configuration of the policy is as follows:

```
*A:CSC-PE-2# configure
    router
        policy-options
            begin
            policy-statement "BGP-VPN-routes"
                entry 10
                    from
                        protocol bgp-vpn
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
        exit
    exit
```

The effect of the BGP export policy is to re-advertise VPN-IPv4 routes imported into the CSC VPRN (and used for forwarding) to CSC-CE-4.

Step 3. Verify exchange of labeled IPv4 routes between CSC-CE-1 and CSC-PE-2.

When Steps 1 and 2 have been completed properly, CSC-CE-1 should now be advertising the labeled-IPv4 route for its system IP address to CSC-PE-2. This can be checked from the perspective of CSC-CE-1 as follows:

```
*A:CSC-CE-1# show router bgp routes 192.0.2.1/32 label-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.168.12.1
Path Id        : None
To             : 192.168.12.2
Res. Nexthop   : n/a
Local Pref.    : n/a                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
IPv4 Label     : 262142
Origin         : IGP
AS-Path        : 64496
Route Tag      : 0
Neighbor-AS    : 64496
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CSC-CE-1#
```

CSC-CE-1 has advertised a label value of 262142 with the prefix.

The following output shows the received route from the perspective of CSC-PE-2:

```
*A:CSC-PE-2# show router 1 bgp routes 192.0.2.1/32 label-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.168.12.1
Path Id        : None
From           : 192.168.12.1
Res. Nexthop   : 192.168.12.1
Local Pref.    : None                    Interface Name : int-CSC-PE-2-CSC-CE-1
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.1
Fwd Class      : None                    Priority       : None
IPv4 Label     : 262142
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64496
Route Tag      : 0
Neighbor-AS    : 64496
Orig Validation: NotFound
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h12m50s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CSC-PE-2#
```

Step 4. Configure core connectivity for CSC-PE-2.

The next step is to configure the base router instance of CSC-PE-2 so that it can exchange VPN-IPv4 routes with CSC-PE-3 (and potentially other CSC-PEs). At a minimum this requires:

- Router-id and autonomous-system configuration.
- Network interface creation and configuration, including assignment of an IPv4 address to the system interface.
- Configuration of the IGP protocol. In this example, IS-IS is used.
- Configuration of the LDP protocol (optional).

– Configuration of RSVP LSPs used to reach remote CSC-PE devices (optional).
– Configuration of the BGP protocol.

These elements of the base router configuration of CSC-PE-2 are as follows:

```
*A:CSC-PE-2>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "IP Configuration"
#--------------------------------------------------
        interface "int-CSC-PE-2-CSC-PE-3"
            address 192.168.23.1/30
            port 1/1/1
            no shutdown
        exit
        interface "system"
            address 192.0.2.2/32
            no shutdown
        exit
        autonomous-system 64500
#--------------------------------------------------
echo "ISIS Configuration"
#--------------------------------------------------
        isis 0
            level-capability level-2
            area-id 49.01
            level 2
                wide-metrics-only
            exit
            interface "system"
                passive
                no shutdown
            exit
            interface "int-CSC-PE-2-CSC-PE-3"
                interface-type point-to-point
                no shutdown
            exit
            no shutdown
        exit
#--------------------------------------------------
echo "LDP Configuration"
#--------------------------------------------------
        ldp
            interface-parameters
                interface "int-CSC-PE-2-CSC-PE-3" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
            targeted-session
            exit
            no shutdown
        exit
#--------------------------------------------------
```

```
echo "BGP Configuration"
#------------------------------------------------
        bgp
            group "core"
                type internal
                neighbor 192.0.2.3
                    family vpn-ipv4
                exit
            exit
            no shutdown
        exit
------------------------------------------------
*A:CSC-PE-2>config>router#
```

Note the following about the BGP configuration of the base router in CSC-PE-2:

- The peer type is IBGP (**type internal**. It is also possible to configure this in a similar way as for eBGP, with the same value for **peer-as** as the locally configured **autonomous-system**).
- The transport for the IBGP session is IPv4 (the **neighbor** address is an IPv4 address).
- The **family vpn-ipv4** command causes MP-BGP negotiation of the address family for AFI=1 and SAFI=128 (=0x80), as can be observed from the following debug trace of the OPEN message from CSC-PE-2.

```
1 2016/10/17 13:36:43.04 UTC MINOR: DEBUG #2001 Base BGP
"BGP: OPEN
Peer 1: 192.0.2.3 - Send (Passive) BGP OPEN: Version 4
   AS Num 64500: Holdtime 90: BGP_ID 192.0.2.2: Opt Length 16
   Opt Para: Type CAPABILITY: Length = 14: Data:
     Cap_Code MP-BGP: Length 4
       Bytes: 0x0 0x1 0x0 0x80
     Cap_Code ROUTE-REFRESH: Length 0
     Cap_Code 4-OCTET-ASN: Length 4
       Bytes: 0x0 0x0 0xfb 0xf4
"
```

Step 5. Configure core connectivity for CSC-PE-3

The next step is to configure the base router instance of CSC-PE-3 so that it can exchange VPN-IPv4 routes with CSC-PE-2 (and potentially other CSC-PEs). At a minimum, this requires:

- Router-id and autonomous-system configuration.
- Network interface creation and configuration, including assignment of an IPv4 address to the system interface.
- Configuration of the IGP protocol. In this example IS-IS is used.
- Configuration of the LDP protocol (optional).
- Configuration of RSVP LSPs used to reach remote CSC-PE devices (optional).

– Configuration of the BGP protocol.

These elements of the base router configuration of CSC-PE-3 are as follows:

```
*A:CSC-PE-3>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "IP Configuration"
#--------------------------------------------------
        interface "int-CSC-PE-3-CSC-PE-2"
            address 192.168.23.2/30
            port 1/1/2
            no shutdown
        exit
        interface "system"
            address 192.0.2.3/32
            no shutdown
        exit
        autonomous-system 64500
#--------------------------------------------------
echo "ISIS Configuration"
#--------------------------------------------------
        isis 0
            level-capability level-2
            area-id 49.01
            level 2
                wide-metrics-only
            exit
            interface "system"
                passive
                no shutdown
            exit
            interface "int-CSC-PE-3-CSC-PE-2"
                interface-type point-to-point
                no shutdown
            exit
            no shutdown
        exit
#--------------------------------------------------
echo "LDP Configuration"
#--------------------------------------------------
        ldp
            interface-parameters
                interface "int-CSC-PE-3-CSC-PE-2" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
            targeted-session
            exit
            no shutdown
        exit
#--------------------------------------------------
echo "BGP Configuration"
#--------------------------------------------------
        bgp
```

```
            group "core"
                type internal
                cluster 192.0.2.3
                neighbor 192.0.2.2
                    family vpn-ipv4
                    split-horizon
                exit
            exit
            no shutdown
        exit----------------------------------------------
*A:CSC-PE-3>config>router#
```

Note the following about the BGP configuration of the base router in CSC-PE-3:

- The peer type is IBGP (**type internal**. If the configuration would include **peer-as** instead, the value would be the same as the locally configured **autonomous-system**).
- The transport for the IBGP session is IPv4 (the **neighbor** address is an IPv4 address).
- The **family vpn-ipv4** command causes MP-BGP negotiation of the address family for AFI=1 and SAFI=128.
- The **cluster** command configures CSC-PE-2 as a route reflector for clients in the BGP group called “core”. This is not required and in a more typical deployment, the route reflector would be a separate router from any CSC-PE.

Step 6. Configure CSC service on CSC-PE-3.

CSC-PE-3 must be configured with a VPRN in **carrier-carrier-vpn** mode in order to provide CSC service to CSC-CE-4. The entire configuration of the VPRN is shown below:

```
*A:CSC-PE-3# configure
    service
        vprn 1 customer 1 create
            carrier-carrier-vpn
            router-id 192.0.2.3
            autonomous-system 64500
            route-distinguisher 64500:13
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64500:1
            network-interface "int-CSC-PE-3-CSC-CE-4" create
                address 192.168.34.1/30
                port 1/1/1
                no shutdown
            exit
            bgp
                group "CSC-CE"
                    as-override
                    export "BGP-VPN-routes"
                    peer-as 64496
                    neighbor 192.168.34.2
```

```
                        family label-ipv4
                        split-horizon
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
```

Note the following about the VPRN configuration of CSC-PE-3:

- The **carrier-carrier-vpn** command is mandatory. It cannot be configured if the VPRN currently has any SAP or spoke-SDP “access” interfaces configured; they must first be shutdown if necessary and then deleted.
- The **auto-bind-tunnel** command should be set appropriately for the type of transport desired to other CSC-PEs, but GRE is not supported.
- The interface to CSC-CE-4 must be a **network-interface**. A **network-interface** can be associated with an entire Ethernet port (as shown in the example above), a VLAN sub-interface of an Ethernet port, an entire LAG or a VLAN sub-interface of a LAG. In all cases, the associated Ethernet ports must be configured in network or hybrid mode and must reside on FP2 or higher based cards/systems.

Note the following about the BGP configuration of the CSC VPRN service in CSC-PE-2:

- The peer type is EBGP (**peer-as** is different from the locally configured **autonomous-system**).
- The address family for the EBGP session is **label-ipv4** (the **neighbor** address is an IPv4 address). Address family label-ipv4 causes MP-BGP negotiation of the address family for AFI=1 and SAFI=4 (IPv4 NLRI with MPLS labels).
- The **split-horizon** command is optional. It prevents a best BGP route from the CSC-CE peer from being re-advertised back to that peer.
- The **as-override** command replaces CSC-CE-4’s AS number (64496) with CSC-PE-3’s AS number (64500) in the AS_PATH attribute of routes advertised to CSC-CE-4. Without this configuration, CSC-CE-4 may reject routes originated by CSC-CE-1 as invalid due to an AS-path loop.
- The **export** command applies a BGP export policy to the session. The configuration of the policy is as follows:

```
*A:CSC-PE-3# configure router
        policy-options
            begin
            policy-statement "BGP-VPN-routes"
                entry 10
                    from
```

```
                    protocol bgp-vpn
                exit
                action accept
                exit
            exit
            default-action drop
            exit
        exit
        commit
    exit
exit
```

The effect of the BGP export policy is to re-advertise VPN-IPv4 routes imported into the CSC VPRN (and used for forwarding) to CSC-CE-4.

Step 7. Verify exchange of routes between CSC-PE-2 and CSC-PE-3.

When the preceding steps have been completed properly, CSC-PE-2 should now be advertising the labeled-IPv4 route for 192.0.2.1/32 (the system IP address of CSC-CE-1) to CSC-PE-3. This can be checked from the perspective of CSC-PE-2 as follows:

```
*A:CSC-PE-2# show router bgp routes 192.0.2.1/32 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.0.2.2
Route Dist.    : 64500:12               VPN Label      : 262140
Path Id        : None
To             : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Origin         : IGP
AS-Path        : 64496
Route Tag      : 0
```

```
Neighbor-AS    : 64496
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CSC-PE-2#
```

CSC-PE-2 has advertised a label value of 262140 with the prefix.

The following output shows the received route from the perspective of CSC-PE-3:

```
*A:CSC-PE-3# show router bgp routes 192.0.2.1/32 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.0.2.2
Route Dist.    : 64500:12               VPN Label      : 262140
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-CSC-PE-3-CSC-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : 64496
Route Tag      : 0
Neighbor-AS    : 64496
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h22m32s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

```
Routes : 1
===============================================================================
*A:CSC-PE-3#
```

Also note the label swap entries that BGP programmed in the line cards of CSC-PE-2 based on the received labeled-IPv4 route from CSC-CE-1 (Label Origin = ExtCarCarVpn) and the advertised VPN-IPv4 route to CSC-PE-3:

```
*A:CSC-PE-2# show router bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.168.12.1                  262142         262140         ExtCarCarVpn
-------------------------------------------------------------------------------
Total Labels allocated:   1
===============================================================================
*A:CSC-PE-2#
```

Step 8. Configure CSC-CE-4.

This example assumes that CSC-CE-4 is a PE router with Layer 2 and Layer 3 VPN services that must extend across the CSC VPN service. The configuration of one such Layer 3 VPN service in CSC-CE-4 is as follows:

```
*A:CSC-CE-4# configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:14
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:64496:1
            interface "loopback-1" create
                address 10.14.30.2/24
                loopback
            exit
            ---snip---
```

For brevity, the preceding configuration sample omits commands related to SAP IP interfaces, spoke-SDP IP interfaces, PE-CE routing protocols, QoS, IP filters, and so on.

The base routing instance of CSC-CE-4 should be configured with the appropriate router ID and autonomous system number and the system interface should be given an IPv4 address (usually the same as the router-id). The interface to CSC-PE-3 should then be created and configured. The base router configuration of CSC-CE-4 is as follows:

```
*A:CSC-CE-4>config>router# info
----------------------------------------------
#----------------------------------------------
```

```
echo "IP Configuration"
#------------------------------------------------
        interface "int-CSC-CE-4-CSC-PE-3"
            address 192.168.34.2/30
            port 1/1/2
            no shutdown
        exit
        interface "system"
            address 192.0.2.4/32
            no shutdown
        exit
        autonomous-system 64496
------------------------------------------------
*A:CSC-CE-4>config>router#
```

BGP should be configured as the control plane protocol running on the interface to CSC-PE-4 as follows:

```
*A:CSC-CE-4# configure router
        bgp
            group "CSC-PE"
                peer-as 64500
                neighbor 192.168.34.1
                    family label-ipv4
                    export "static-to-BGP"
                    split-horizon
                exit
            exit
            no shutdown
        exit
```

Note the following about the BGP configuration of CSC-CE-4:

- The peer type is EBGP (**peer-as** is different from the locally configured **autonomous-system**).
- The address family for the EBGP session is **label-ipv4** (the **neighbor** address is an IPv4 address). Address family label-ipv4 causes MP-BGP negotiation of the address family for AFI=1 and SAFI=4 (IPv4 NLRI with MPLS labels)
- The **split-horizon** command is optional. It prevents a best BGP route from the CSC-PE peer from being re-advertised back to that peer.
- The **export** command applies a BGP export policy to the session. The configuration of the policy is shown below:

```
*A:CSC-CE-4# configure
    router
        policy-options
            begin
            prefix-list "system-IP"
                prefix 192.0.2.4/32 exact
            exit
            policy-statement "static-to-BGP"
                entry 10
                    from
                        protocol direct
```

```
                    prefix-list "system-IP"
                exit
                action accept
                exit
            exit
            default-action drop
            exit
        exit
        commit
    exit
```

The purpose of the BGP export policy is to advertise the system IP address of CSC-CE-4 as a labeled-IPv4 BGP route toward CSC-PE-3.

Step 9. Verify exchange of routes between CSC-PE-3 and CSC-CE-4.

When the preceding steps have been completed, CSC-PE-3 should be advertising the labeled-IPv4 route for 192.0.2.1/32 to CSC-CE-4. This can be checked from the perspective of CSC-PE-3 as shown below:

```
*A:CSC-PE-3# show router 1 bgp routes 192.0.2.1/32 label-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.168.34.1
Path Id        : None
To             : 192.168.34.2
Res. Nexthop   : n/a
Local Pref.    : n/a                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
IPv4 Label     : 262139
Origin         : IGP
AS-Path        : 64500 64500
Route Tag      : 0
Neighbor-AS    : 64500
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
```

```
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CSC-PE-3#
```

CSC-PE-3 has advertised a label value of 262139 with the prefix.

The following output shows the received route from the perspective of CSC-CE-4:

```
*A:CSC-CE-4# show router bgp routes 192.0.2.1/32 label-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.1/32
Nexthop        : 192.168.34.1
Path Id        : None
From           : 192.168.34.1
Res. Nexthop   : 192.168.34.1
Local Pref.    : None                   Interface Name : int-CSC-CE-4-CSC-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
IPv4 Label     : 262139
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64500 64500
Route Tag      : 0
Neighbor-AS    : 64500
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h11m34s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CSC-CE-4#
```

Note the label swap entries that BGP programmed in the line cards of CSC-PE-3 based on the received VPN-IPv4 routes from CSC-PE-2 (Label Origin = Internal) and the advertised labeled-IPv4 routes to CSC-CE-4:

```
*A:CSC-PE-3# show router 1 bgp inter-as-label
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                         Received       Advertised     Label
                                Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.2                       262140         262139         Internal
192.0.2.2                       262141         262140         Internal
-------------------------------------------------------------------------------
Total Labels allocated:   2
===============================================================================
*A:CSC-PE-3#
```

In the preceding output, the first entry for NextHop 192.0.2.2 corresponds to the prefix 192.0.2.1/32; recall from Step 7 that CSC-PE-3 received the VPN-IPv4 route with label value 262140 and it can be seen from this step that it re-advertised the route to CSC-CE-4 with label value 262139.

Step 10. Set up BGP session between CSC-CE-1 and CSC-CE-4.

The final step in the setup of the CSC solution shown in Figure 319 is the creation of a BGP session between CSC-CE-1 and CSC-CE-4 so that they can exchange routes belonging to VPN services they support. The configuration of this BGP session from the perspective of CSC-CE-1 is as follows:

```
*A:CSC-CE-1# configure router
        bgp
            group "CSC-CE"
                type internal
                neighbor 192.0.2.4
                    family vpn-ipv4
                exit
            exit
            no shutdown
```

The configuration of the BGP session from the perspective of CSC-CE-4 is very similar, as shown below.

```
*A:CSC-CE-4# configure router
        bgp
            group "CSC-CE"
                type internal
                neighbor 192.0.2.1
                    family vpn-ipv4
                exit
            exit
            no shutdown
```

Note the following about the configuration of the BGP session between CSC-CE-1 and CSC-CE-4:

- The peer type is IBGP (**type internal**. Alternatively, **peer-as** can be configured with the same value as the locally configured **autonomous-system**).
- The transport for the IBGP session is IPv4 (the **neighbor** address is an IPv4 address).
- The **family vpn-ipv4** command causes MP-BGP negotiation of the address family for AFI=1 and SAFI=128.

Step 11. Verify exchange of routes between CSC-CE-1 and CSC-CE-4.

When the preceding steps have been completed properly, CSC-PE-3 should now be able to advertise a VPN-IPv4 route for some IP prefix (for example, 10.11.30.0/24) to CSC-CE-4. This can be checked from the perspective of CSC-CE-4 as follows:

```
*A:CSC-CE-4# show router bgp routes 10.11.30.0/24 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.11.30.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:11               VPN Label      : 262143
Path Id        : None
From           : 192.0.2.1
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64496:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h08m57s
```

```
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CSC-CE-4#
```

It is also possible to check that CSC-CE-4 has properly installed the preceding VPN-IPv4 route into the routing table of the importing VPRN service, as follows.

```
*A:CSC-CE-4# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
10.11.30.0/24                                   Remote  BGP VPN   00h12m05s   170
       192.0.2.1 (tunneled:BGP)                                       0
10.14.30.0/24                                   Local   Local     00h46m42s   0
       loopback-1                                                     0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:CSC-CE-4#
```

Conclusion

Carrier Supporting Carrier is a scalable and secure solution for using an infrastructure IP VPN to transport traffic between dispersed CSC-CE devices belonging to an ISP or other service provider. Many different topology models are supported by SR OS. This chapter has explored one simplified configuration that can serve as the basis for more complicated setups.

Layer 3 VPN: VPRN Type Spoke

This chapter provides information about Layer 3 VPRN CE hub and spoke architecture.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 12.0. However, the CLI in the current edition is based on release 14.0.R5.

Summary

This chapter provides a basic technology overview and configuration examples of a network topology used for a Layer 3 VPRN CE hub and spoke architecture.

Knowledge of Nokia's Layer 3 VPN concepts is assumed throughout this document.

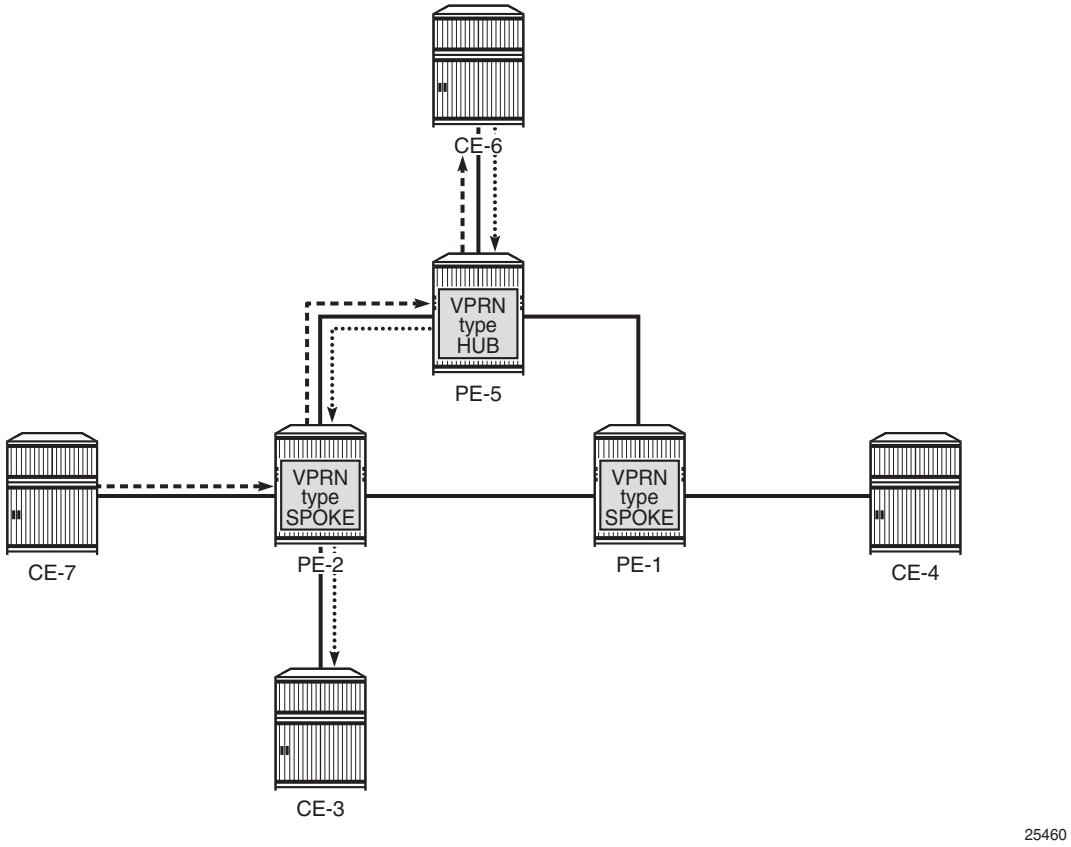
Overview

In SR OS releases earlier than 12.0, a **CE hub and spoke** architecture was partially supported. Internal optimization was available for the hub sites connected to the same PE router only. This feature is known as VPRN **type hub**. If, on the other hand, multiple spoke sites were connected to the same PE router, separate VPRN instances had to be created to maintain the split horizon forwarding behavior. This approach was complex, hard to maintain and consumed extra VPRN instances.

Release 12.0.R1 added new functionality to overcome these limitations. Introducing the VPRN **type spoke** feature allows multiple spoke sites to be kept within the same VPRN instance while at the same time maintaining the split horizon approach such that spoke sites cannot send traffic directly to each other.

The primary goal of the feature is to allow multiple spoke sites to be part of a single VPRN instance without allowing direct communication between the spoke CE sites which are part of that VPRN (of type spoke). The packet flow is demonstrated in Figure 320.

Figure 320 **CE Hub and Spoke Data Path**



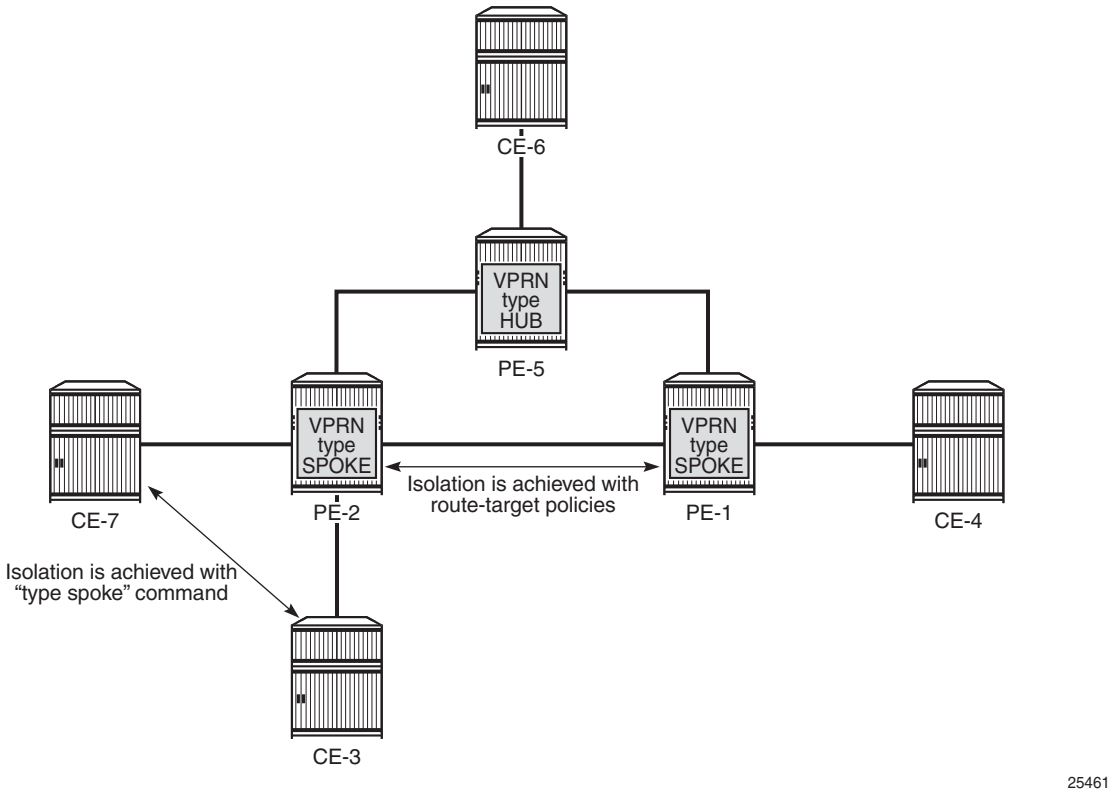
The only way for CE-7 to communicate with CE-3 is via hub site CE-6. The same applies to the communication between CE-7 and CE-4. The VPRN on PE-2 is configured as **type spoke** and has IP interfaces using SAPs or spoke SDPs that are considered spoke sites only. No direct communication between any of the spoke CE sites in the network is allowed.

This is achieved using two techniques (Figure 321).

- Use the **type spoke** command under the VPRN context as explained later.

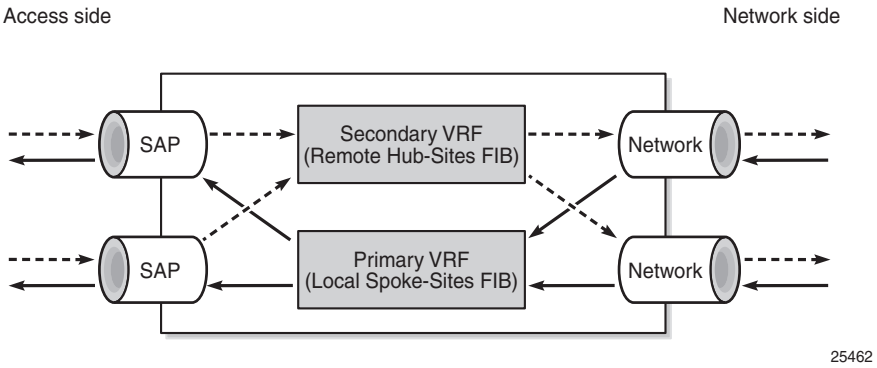
- The extended community configuration using route-target policies (this is not covered in detail in this chapter).

Figure 321 **CE Hub and Spoke Control Plane Isolation**



When a VPRN on a PE router is configured as **type spoke,** then the internal forwarding logic changes as demonstrated in Figure 322.

Figure 322 **Internal VPRN Logic on a PE Router**

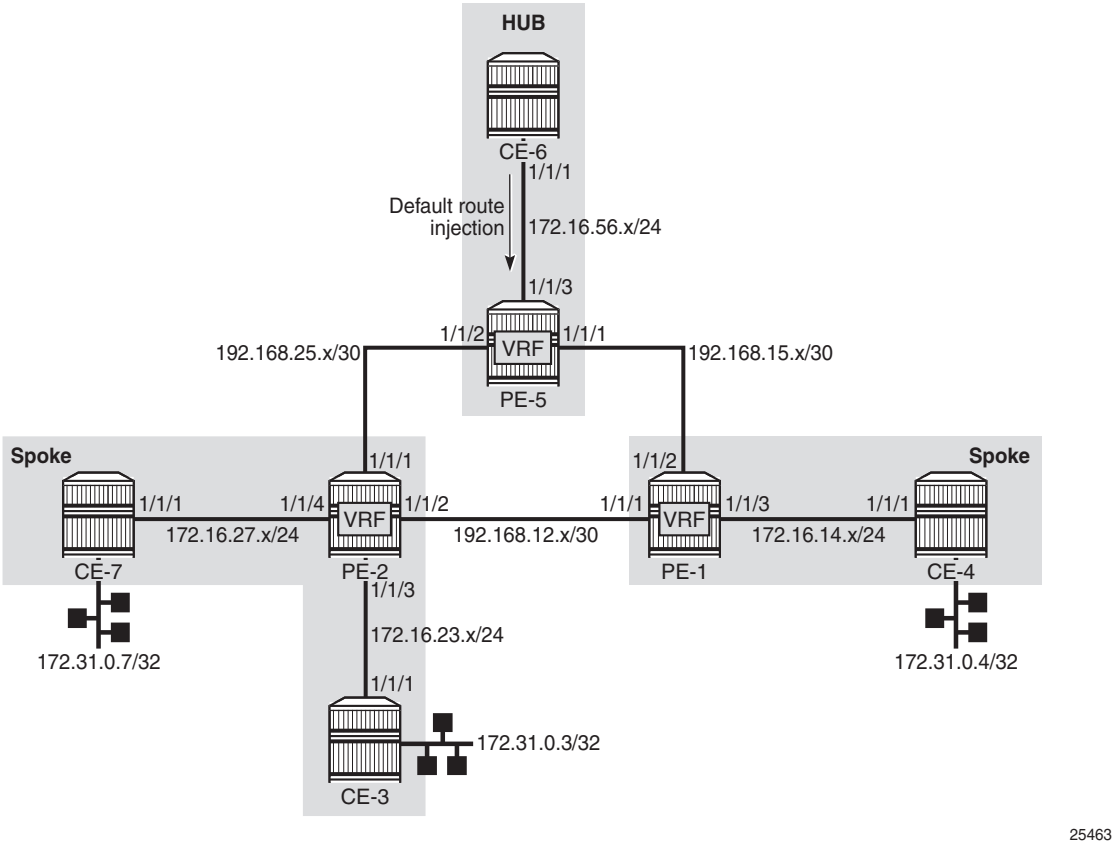


- VPRNs of type spoke create a primary and a secondary VRF internally to the VPRN:
  - The primary VRF is used for forwarding traffic from the network interfaces toward the IP interfaces using SAPs or spoke SDPs. This VRF is populated with routes learned from the spoke CE sites connected to the local PE through IP interfaces using SAPs or spoke SDPs.
  - The secondary VRF is used for forwarding traffic from the IP interfaces using SAPs or spoke SDPs toward the network interfaces or other VPRN instances. This VRF is populated with routes learned via MP-BGP from Hub sites.
- VPRNs of type spoke export routes using a specific extended community (for instance spoke-ext-comm) via an export policy to identify them as spoke site originated routes.
  - This community is not hard-coded and has to be configured manually (see configuration example later).
- VPRNs of type spoke import routes (using an import policy) received from other PEs or VPRN instances with a hub specific community only (for example hub-ext-comm). Routes with spoke-ext-comm community are ignored.
  - This community is not hard-coded and has to be configured manually (see configuration example later).
- Multiple VPRNs of type spoke and hub can coexist on the same PE if they use different VPRN instances.
- The configuration of type hub and type spoke is mutually exclusive within one VPRN instance.

Configuration

The physical topology and addressing scheme are presented in Figure 323.

Figure 323 **CE Hub and Spoke Topology and Addressing Scheme**



The configuration of PE-2 and PE-5 are the main focus of this example. The configuration of PE-1 is similar to that of PE-2.

Hub Site Configuration

Only the essential part of the configuration is provided for the hub site.

Vrf-import and export policies are used to manipulate the vrf-target in order to achieve logical isolation between the spoke sites in the network.

```
*A:PE-5# configure
    router
        policy-options
            begin
            community "hub-ext-comm" members "target:64500:11"
            community "spoke-ext-comm" members "target:64500:12"
            policy-statement "vrf-export"
```

```
                default-action accept
                    community add "hub-ext-comm"
                exit
            exit
            policy-statement "vrf-import"
                entry 10
                    from
                        community "spoke-ext-comm"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            policy-statement "export-ospf"
                entry 10
                    from
                        protocol direct
                    exit
                    action accept
                    exit
                exit
                default-action accept
                exit
            exit
            commit
        exit
```

PE-5 is configured with VPRN 1 providing OSPF connectivity to customer CE-6.

```
*A:PE-5# configure
    service
        vprn 1 customer 1 create
            vrf-import "vrf-import"
            vrf-export "vrf-export"
            route-distinguisher 64500:15
            type hub
            auto-bind-tunnel
                resolution any
            exit
            interface "int-PE-5-CE-6" create
                address 172.16.56.1/24
                sap 1/1/3:1 create
                exit
            exit
            ospf
                export "export-ospf"
                area 0.0.0.0
                    interface "int-PE-5-CE-6"
                        interface-type point-to-point
                        mtu 1500
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

```
        exit
```

At the same time, CE-6 is configured to advertise a default route which is used by all remote spoke CE sites to forward traffic via CE-6.

```
*A:CE-6# configure
    router
        policy-options
            begin
            policy-statement "export-ospf-default"
                entry 10
                    from
                        protocol static
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit

*A:CE-6# configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64500:16
            interface "int-CE-6-PE-5" create
                address 172.16.56.2/24
                sap 1/1/1:1 create
                exit
            exit
            interface "lo0" create
                shutdown
                address 172.31.0.6/32
                loopback
            exit
            static-route-entry 0.0.0.0/0
                black-hole
                    no shutdown
                exit
            exit
            ospf 192.0.2.6
                export "export-ospf-default"
                ignore-dn-bit
                suppress-dn-bit
                area 0.0.0.0
                    interface "int-CE-6-PE-5"
                        interface-type point-to-point
                        mtu 1500
                        no shutdown
                    exit
                    interface "lo0"
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

```
        exit
```

Spoke Site Configuration

According to the example topology, two spoke VPRNs are present: one VPRN with two CE spoke sites connected is located on PE-2, and another VPRN with one spoke CE site is located on PE-1. The service configuration for PE-2 is as follows with the one for PE-1 being similar.

Vrf-import and export policies are used to build a hub-and-spoke topology in order to achieve a logical isolation between spoke sites connected to different PE routers.

```
*A:PE-2# configure
    router
        policy-options
            begin
            community "hub-ext-comm" members "target:64500:11"
            community "spoke-ext-comm" members "target:64500:12"
            policy-statement "vrf-export"
                default-action accept
                    community add "spoke-ext-comm"
                exit
            exit
            policy-statement "vrf-import"
                entry 10
                    from
                        community "hub-ext-comm"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            policy-statement "export-ospf"
                default-action accept
                exit
            exit
            commit
        exit
```

PE-2 is configured with VPRN 1, which has OSPF connectivity to the customer CE-3 and CE-7. The dedicated command **type spoke** is used to prevent direct CE spoke to CE spoke communications for this VPRN.

```
*A:PE-2# configure
    service
        vprn 1
            vrf-import "vrf-import"
            vrf-export "vrf-export"
            route-distinguisher 64500:12
            type spoke
```

```
            auto-bind-tunnel
                resolution any
            exit
            interface "int-PE-2-CE-3" create
                address 172.16.23.1/24
                sap 1/1/3:1 create
                exit
            exit
            interface "int-PE-2-CE-7" create
                address 172.16.27.1/24
                sap 1/1/4:1 create
                exit
            exit
            ospf
                export "export-ospf"
                area 0.0.0.0
                    interface "int-PE-2-CE-3"
                        interface-type point-to-point
                        mtu 1500
                        no shutdown
                    exit
                    interface "int-PE-2-CE-7"
                        interface-type point-to-point
                        mtu 1500
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

For connectivity verification purposes, CE-3, CE-4, and CE-7 are configured to advertise their internal loopback interfaces via OSPF:

- CE-3 advertises 172.31.0.3/32
- CE-4 advertises 172.31.0.4/32
- CE-7 advertises 172.31.0.7/32

```
*A:CE-3# configure service
        vprn 1 customer 1 create
            route-distinguisher 64500:13
            interface "int-CE-3-PE-2" create
                address 172.16.23.2/24
                sap 1/1/1:1 create
                exit
            exit
            interface "lo0" create
                address 172.31.0.3/32
                loopback
            exit
            ospf 192.0.2.3
                ignore-dn-bit
                suppress-dn-bit
                area 0.0.0.0
                    interface "int-CE-3-PE-2"
                        interface-type point-to-point
                        mtu 1500
```

```
                exit
                interface "lo0"
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

Hub Site Verification

The routing table (RIB) for VPRN 1 on PE-5 (hub site) lists all reachable networks.

```
*A:PE-5# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                       Remote  OSPF      02d17h02m   150
       172.16.56.2                                                  1
172.16.14.0/24                                  Remote  BGP VPN   00h32m41s   170
       192.0.2.1 (tunneled)                                         0
172.16.23.0/24                                  Remote  BGP VPN   00h32m37s   170
       192.0.2.2 (tunneled)                                         0
172.16.27.0/24                                  Remote  BGP VPN   00h32m37s   170
       192.0.2.2 (tunneled)                                         0
172.16.56.0/24                                  Local   Local     02d17h31m   0
       int-PE-5-CE-6                                                0
172.31.0.3/32                                   Remote  BGP VPN   00h03m30s   170
       192.0.2.2 (tunneled)                                         0
172.31.0.4/32                                   Remote  BGP VPN   00h02m21s   170
       192.0.2.1 (tunneled)                                         0
172.31.0.7/32                                   Remote  BGP VPN   00h01m31s   170
       192.0.2.2 (tunneled)                                         0
-------------------------------------------------------------------------------
No. of Routes: 8
```

The forwarding table (FIB) for the primary VRF of VPRN 1 is displayed using following command. All remote spoke and hub sites are reachable via this VRF.

```
*A:PE-5# show router 1 fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
  NextHop
-------------------------------------------------------------------------------
0.0.0.0/0                                                       OSPF
  172.16.56.2 (int-PE-5-CE-6)
172.16.14.0/24                                                  BGP_VPN
  192.0.2.1 (VPRN Label:262142 Transport:LDP)
172.16.23.0/24                                                  BGP_VPN
```

```
  192.0.2.2 (VPRN Label:262142 Transport:LDP)
172.16.27.0/24                                                  BGP_VPN
  192.0.2.2 (VPRN Label:262142 Transport:LDP)
172.16.56.0/24                                                  LOCAL
  172.16.56.0 (int-PE-5-CE-6)
172.31.0.3/32                                                   BGP_VPN
  192.0.2.2 (VPRN Label:262142 Transport:LDP)
172.31.0.4/32                                                   BGP_VPN
  192.0.2.1 (VPRN Label:262142 Transport:LDP)
172.31.0.7/32                                                   BGP_VPN
  192.0.2.2 (VPRN Label:262142 Transport:LDP)
-------------------------------------------------------------------------------
Total Entries : 8
-------------------------------------------------------------------------------
===============================================================================
*A:PE-5#
```

The forwarding table for the secondary VRF of VPRN 1 is displayed using following command, including the **secondary** keyword. All local hub CE sites are reachable via this VRF.

```
*A:PE-5# show router 1 fib 1 secondary
===============================================================================
FIB Display
===============================================================================
Prefix                                                          Protocol
    NextHop
-------------------------------------------------------------------------------
0.0.0.0/0                                                       OSPF
    172.16.56.2 (int-PE-5-CE-6)
172.16.56.0/24                                                  LOCAL
    172.16.56.0 (int-PE-5-CE-6)
-------------------------------------------------------------------------------
Total Entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

Spoke Site Verification

The RIB for VPRN 1 on PE-2 (spoke VPRN) lists all reachable networks.

The other spoke sites connected to the remote PEs (only CE-4 here, for example: 172.31.0.4/32)) are not present in the routing table.

The local interface addresses of PE-2 (172.16.23.1/32 and 172.16.27.1/32) are present in the routing table of VPRN 1, as follows. From a FIB point of view, these are reachable from any spoke VPRN, but the spoke CE's router host addresses are not. This fact does not influence the data plane isolation for the customer networks.

```
*A:PE-2# show router 1 route-table
===============================================================================
Route Table (Service: 1)
```

```
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                       Remote  BGP VPN   00h04m33s  170
       192.0.2.5 (tunneled)                                            0
172.16.23.0/24                                  Local   Local     02d17h47m  0
       int-PE-2-CE-3                                                   0
172.16.23.1/32                                  Local   Host      02d17h47m  0
       int-PE-2-CE-3                                                   0
172.16.27.0/24                                  Local   Local     02d17h47m  0
       int-PE-2-CE-7                                                   0
172.16.27.1/32                                  Local   Host      02d17h47m  0
       int-PE-2-CE-7                                                   0
172.16.56.0/24                                  Remote  BGP VPN   00h04m33s  170
       192.0.2.5 (tunneled)                                            0
172.31.0.3/32                                   Remote  OSPF      01h47m42s  10
       172.16.23.2                                                     10
172.31.0.7/32                                   Remote  OSPF      01h45m45s  10
       172.16.27.2                                                     10
-------------------------------------------------------------------------------
No. of Routes: 8
```

The FIB for the primary VRF of VPRN 1 shows all local spoke sites are reachable via this VRF, as follows:

```
*A:PE-2# show router 1 fib 1
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
  NextHop
-------------------------------------------------------------------------------
172.16.23.0/24                                                  LOCAL
  172.16.23.0 (int-PE-2-CE-3)
172.16.23.1/32                                                  HOST
  Blackhole
172.16.27.0/24                                                  LOCAL
  172.16.27.0 (int-PE-2-CE-7)
172.16.27.1/32                                                  HOST
  Blackhole
172.31.0.3/32                                                   OSPF
  172.16.23.2 (int-PE-2-CE-3)
172.31.0.7/32                                                   OSPF
  172.16.27.2 (int-PE-2-CE-7)
-------------------------------------------------------------------------------
Total Entries : 6
```

The FIB for the secondary VRF of VPRN 1 shows the remote hub site (address 172.16.56.0/24) is reachable via this VRF, as follows:

```
*A:PE-2# show router 1 fib 1 secondary
===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                                  Protocol
  NextHop
```

```
-------------------------------------------------------------------------------
0.0.0.0/0                                                        BGP_VPN
  192.0.2.5 (VPRN Label:262141 Transport:LDP)
172.16.23.1/32                                                   HOST
  Blackhole
172.16.27.1/32                                                   HOST
  Blackhole
172.16.56.0/24                                                   BGP_VPN
  192.0.2.5 (VPRN Label:262141 Transport:LDP)
-------------------------------------------------------------------------------
Total Entries : 4
```

Spoke Sites Connectivity Verification

Without the VPRN spoke type configuration in VPRN 1 on PE-2, CE-3 takes the shortest path to CE-7, which violates the “hub and spoke” design approach explained earlier.

The VPRN has to be shut down in order to modify the type.

```
*A:PE-2# configure service vprn 1 no type
INFO: PIP #1162 Instance must be 'shutdown'
*A:PE-2# configure service vprn 1 shutdown
*A:PE-2# configure service vprn 1 no type
*A:PE-2# configure service vprn 1 no shutdown
```

Note: In this setup, a VPRN is configured on CE-3, but that is not necessary.

```
*A:CE-3# traceroute router 1 172.31.0.7
traceroute to 172.31.0.7, 30 hops max, 40 byte packets
  1  172.16.23.1 (172.16.23.1)    1.46 ms  0.642 ms  0.547 ms
  2  172.31.0.7 (172.31.0.7)    2.55 ms  1.70 ms  1.79 ms
```

After enabling the **type spoke** feature on PE-2, CE-3 takes the longest path via hub CE-6 to reach CE-7, as it should.

```
*A:PE-2# configure service vprn 1 shutdown
*A:PE-2# configure service vprn 1 type spoke
*A:PE-2# configure service vprn 1 no shutdown
```

```
*A:CE-3# traceroute router 1 172.31.0.7 no-dns
traceroute to 172.31.0.7, 30 hops max, 40 byte packets
  1  172.16.23.1    0.691 ms  0.642 ms  0.638 ms
  2  0.0.0.0  * * *
  3  172.16.56.2    2.32 ms  2.56 ms  2.25 ms
  4  172.16.56.1    2.46 ms  1.86 ms  2.27 ms
  5  172.16.27.1    1.89 ms  1.95 ms  1.99 ms
  6  172.31.0.7    4.44 ms  4.43 ms  4.90 ms
```

```
*A:CE-3#
```

Similarly, the long path is taken by CE-3 to reach CE-4, as follows. This is unrelated to the VPRN type. It is achieved by policies.

```
*A:CE-3# traceroute router 1 172.31.0.4 no-dns
traceroute to 172.31.0.4, 30 hops max, 40 byte packets
  1  172.16.23.1    0.677 ms  0.545 ms  0.546 ms
  2  0.0.0.0  * * *
  3  172.16.56.2    2.18 ms  2.41 ms  2.30 ms
  4  172.16.56.1    2.05 ms  1.94 ms  2.02 ms
  5  172.16.14.1    2.70 ms  3.13 ms  3.36 ms
  6  172.31.0.4    3.78 ms  3.76 ms  4.57 ms
*A:CE-3#
```

Conclusion

The VPRN type spoke feature completes the **CE hub and spoke** solution. It brings a new level of simplicity, scalability, and flexibility to operators using this VPRN architecture for their customers.

NG-MVPN Configuration with PIM

This chapter provides information about multicast in a VPRN service.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

Initially, this chapter was written for SR OS release 7.0.R5. The configuration in the current edition is based on SR OS release 14.0.R5. There are no prerequisites for this configuration.

Summary

Multicast VPN (MVPN) architectures describe a set of VRFs that support the transport of multicast traffic across a provider network.

RFC 6037 (herein referred to as Rosen MVPN) describes the use of multicast distribution trees (MDT) established between PEs within a given VRF. Each VRF required its own tree. Customer edge routers form protocol independent multicast (PIM) adjacencies with the PE, and PE-PE PIM adjacencies are formed across the multicast tree. PIM signaling and data streams are transported across the MDT. There were a number of limitations with the Rosen MVPN implementation including, but not limited to:

- Rosen MVPN requires a set of MDTs per VPN, which requires a PIM state per MDT. There is no option to aggregate MDT across multiple VPNs
- Customer signaling, PE discovery and Data MDT signaling are all PIM-based. There is no mechanism available to decouple these. Thus there is an incongruency between unicast and multicast VPNs using Rosen MVPN.

– There is no mechanism for using MPLS to encapsulate multicast traffic in the VPN. GRE is the only encapsulation method available in Rosen MVPN.
– Rosen MVPN multicast trees are signaled using PIM only. NG MVPN allows the use of mLDP, RSVP P2MP LSPs.
– PE to PE protocol exchanges for Rosen MVPN is achieved using PIM only. NG MVPN allows for the use of BGP signaling as per unicast Layer 3 VPNs.

Next Generation MVPN addresses these limitations by extending the idea of the per-VRF tree, by introducing the idea of Provider Multicast Service Interfaces (PMSI). These are equivalent to the default MDTs of Rosen MVPN in that they support control plane traffic (customer multicast signaling), and the data MDTs which carry multicast data traffic streams between PEs within a multicast VRF.

Next Generation MVPN allows the decoupling of the mechanism required to create a multicast VPN, such as PE auto-discovery (which PEs are members of which VPN), PMSI signaling (creation of tunnels between PEs) and customer multicast signaling (multicast signaling —IGMP/PIM — received from customer edge routers). Two types of PMSI exist:

- Inclusive (I-PMSI): contains all the PEs for an MVPN.
- Selective (S-PMSI): contains only a subset of PEs of an MVPN.

Knowledge of MPLS-VPN RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, architecture and functionality, as well as an understanding of multicast protocols, is assumed throughout.

This chapter provides configuration details required to implement the parts of Next Generation MVPN shown in Table 24.

Table 24 **Next Generation MVPN Components**

Provider Multicast Domain				**Customer Multicast Domain**		
I-PMSI	**Auto-discovery**	**C-Mcast**	**S-PMSI Creation**	**PE-based RP**	**Anycast RP on PE**	**PIM SSM**
PIM ASM	PIM	PIM join/ leave	PIM SSM with S-PMSI join TLV	X	X	X
PIM ASM	BGP A/D	PIM join/ leave	PIM SSM with S-PMSI join TLV			X

The first section of this chapter describes the common configuration required for each PE within the provider multicast domain regardless of the MVPN PE auto-discovery or customer signaling methods. This includes IGP and VPRN service configuration.

Following the common configuration, specific MVPN configuration required for the configuration for the provider multicast domain using PIM Any Source Multicast (ASM) with auto-discovery based on PIM or BGP auto-discovery (A/D), PIM used for the customer multicast signaling and PIM Source Specific Multicast (SSM) used for the S-PMSI creation are described. The customer domain configuration covers the following three cases:

1. PIM ASM with the Rendezvous Point (RP) in the provider PE
2. PIM ASM using anycast RP on the provider RPs
3. PIM SSM

Other possible options, not covered in this section but are described in the 7450 ESS/7750 SR/7950 XRS Multicast Routing Protocols Guide:

- The use of PIM SSM for the provider multicast I-PMSI.
- The use of BGP for the customer multicast signaling in the provider multicast domain.
- The provider S-PMSI creation through BGP S-PMSI A/D.
- The use of the customer RP based in the customer CE.

The use of mLDP and RSVP p2mp LSPs for the I/S-PMSI was not available in release 7.0.

The Multicast in a VPRN II example in NG-MVPN Configuration with MPLS introduces features that were not supported in Release 7.0.R5. It provides configuration details to implement:

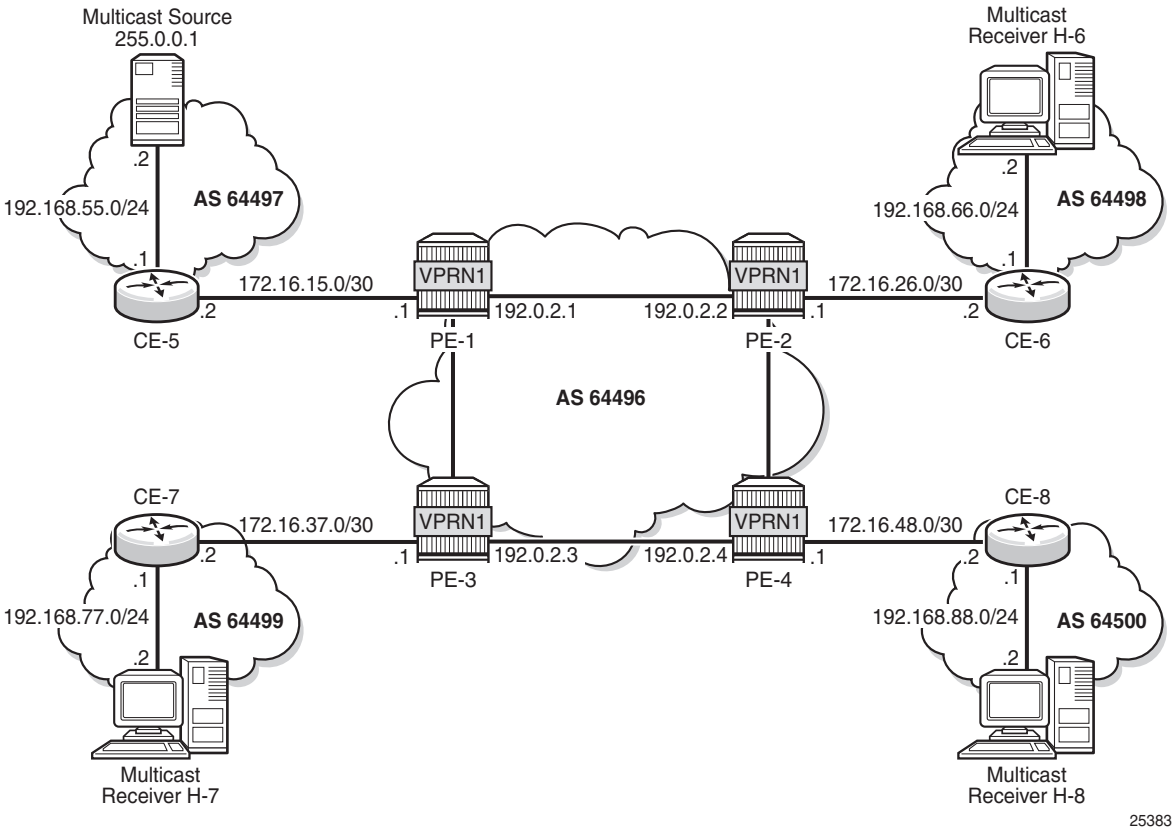
- Multicast LDP (mLDP) and RSVP-TE Point to Multi-point (P2MP) for building customer trees (C-trees) which are using MPLS instead of PIM techniques.
- MVPN source redundancy
- MDT AFI/SAFI (to fully interoperate with Cisco networks).

References

- IETF
  - RFC 6513, Multicast in MPLS/BGP IP VPNs.
  - RFC 6514, BGP Encodings and Procedures for Multicast in MPLS/ BGP IP VPNs.
- 7450 ESS/7750 SR/7950 XRS Layer 3 Services Guide

Overview

Figure 324 **Network Topology**



The network topology is displayed in Figure 324. The setup consists of four SR 7750s acting as Provider Edge (PE) routers within a single Autonomous System (AS).

- Full mesh ISIS or OSPF in each AS
- LDP on all interfaces in each AS (RSVP could also be used)
- MP-iBGP sessions between the PE routers in each AS (Route Reflectors (RRs) could also be used).
- Layer 3-VPN on all PEs with identical route targets, in the form AS-number: *vprn-service-id*

Connected to each PE is a single 7750 acting as a Customer Edge (CE) router. CE-5 has a multicast source connected, and PE-6, PE-7, and PE-8 each have a single receiver connected which will receive the multicast streams from the source. In this document, each receiver is both IGMPv2 and IGMPv3 capable. If the customer domain multicast signaling plane uses Source Specific Multicasting (SSM), then an IGMPv3 receiver is configured; if Any Source Multicasting (ASM) is used, the receiver is IGMPv2 capable.

If the receiver is IGMPv3 capable, it will issue IGMPv3 reports that will include a list of required source addresses. The receiver will join the 232.0.0.1 multicast group.

If the receiver is only IGMPv2 capable, then it will issue IGMPv2 reports which do not specify a source of the group. In this case, a Rendezvous Point is required within the PIM control plane of the multicast VRF which is source-aware. In this case, the receiver will join the 225.0.0.1 multicast group.

When the receiver wishes to become a member of any group, the source address of the group must be known to the CE. As a result, the source address must be IP reachable by each CE, so it is advertised by CE-5 to the PEs with attachment circuits in VPRN1 using BGP.

Static routes are then configured on the receiver CEs to achieve IP reachability to the source address of multicast groups. In the case of PIM ASM, any RP that is configured must also be reachable from the CE.

Multicast VPN Overview

Multicast traffic from the source is streamed toward router CE-5. Receivers connected to PE-2, PE-3 and PE-4 are interested in joining this multicast group.

All CEs are PIM enabled routers, which form a PIM adjacency with their nearest PE. The PIM adjacencies between PEs across the Provider network are achieved using I-PMSIs. I-PMSIs carry PIM control messages between PEs. Data plane traffic is transported across the I-PMSI until a configured bandwidth threshold is reached. A Selective PMSI is then signaled that carries data plane traffic. This threshold can be as low as 1kb/second and must be explicitly configured along with the S-PMSI multicast group. An S-PMSI per customer group per VPRN is configured. If no S-PMSI and threshold is configured, data traffic will continue to be forwarded across the provider network within the I-PMSI.

Configuration

The configuration is divided into the following sections:

- Provider Common Configuration
  - PE Global Configuration
  - PE VPRN Configuration
- PE VPRN Multicast Configuration
  - Auto-Discovery within Provider Domain using PIM
  - PIM Autodiscovery: Customer Signaling using PIM
- PIM Any Source Multicasting with RP at the provider PE
- PIM Any Source Multicasting with Anycast RP at the provider PE
- PIM Source Specific Multicasting
  - BGP Autodiscovery: PE VPRN Multicast Configuration
  - Data Path Using Selective-PMSIs

Provider Common Configuration

This section describes the common configuration required for each PE within the Provider multicast domain, regardless of the MVPN PE auto-discovery or customer signaling methods. This includes IGP and VPRN service configuration.

The configuration tasks can be summarized as follows:

- PE global configuration. This includes configuration of the Interior Gateway Protocol (IGP) (ISIS or OSPF); configuration of link layer LDP between PEs; configuration of iBGP between PEs, to facilitate VPRN route learning; configuration of PIM.
- VPRN configuration on PEs. This includes configuration of basic VPRN parameters (route-distinguisher, route target communities); configuration of attachment circuits toward CEs; configuration of VRF routing protocol and any policies toward CE.
- VRF PIM and MVPN parameters — I-PMSI
- CE configuration.

PE Global Configuration

Step 1. On each of the PE routers, configure the appropriate router interfaces, OSPF (or ISIS) and link layer LDP. For clarity in the following configuration steps, only the configuration for PE-1 is shown. PE-2, PE-3 and PE-4 are similar.

```
*A:PE-1# configure router
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/1
        exit
        interface "int-PE-1-PE-3"
            address 192.168.13.1/30
            port 1/1/2
        exit
        interface "system"
            address 192.0.2.1/32
        exit
        autonomous-system 64496
        ospf
            area 0.0.0.0
                interface "system"
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                exit
                interface "int-PE-1-PE-3"
                    interface-type point-to-point
                exit
            exit
        exit
        ldp
            interface-parameters
                interface "int-PE-1-PE-2"
                exit
                interface "int-PE-1-PE-3"
                exit
            exit
        exit
```

Step 2. Verify that OSPF adjacencies are formed and that LDP peer sessions are formed.

```
*A:PE-1# show router ospf neighbor

===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-PE-2                    192.0.2.2       Full       1    1       36
   0.0.0.0
int-PE-1-PE-3                    192.0.2.3       Full       1    1       34
   0.0.0.0
-------------------------------------------------------------------------------
```

```
No. of Neighbors: 2
===============================================================================
*A:PE-1#

*A:PE-1# show router ldp session ipv4
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   176       178       0d 00:07:36
192.0.2.3:0         Link      Established   175       176       0d 00:07:30
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
*A:PE-1#
```

Step 3. Configure BGP between the PEs for VPRN routing.

```
*A:PE-1# configure router
        bgp
            group "INTERNAL"
                family vpn-ipv4
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
            no shutdown
        exit
```

Step 4. Verify that BGP peer relationship is established.

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 15          Total Path Memory           : 3000
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0
```

```
Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
                64496       4    0 00h00m36s 0/0/0 (VpnIPv4)
                            4    0
192.0.2.3
                64496       3    0 00h00m29s 0/0/0 (VpnIPv4)
                            3    0
192.0.2.4
                64496       3    0 00h00m21s 0/0/0 (VpnIPv4)
                            3    0
-------------------------------------------------------------------------------
*A:PE-1#
```

Step 5. Enable PIM on all network interfaces, including the system interface. This allows the signaling of PMSIs that transport PIM signaling within each VRF.

Step 6. As each I-PMSI will be signaled using PIM ASM, a rendezvous point (RP) is required within the global PIM configuration. A static RP is used and PE-1 is selected. All PEs must be configured with this RP address.

```
*A:PE-1# configure router
        pim
            interface "system"
            exit
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-3"
            exit
            rp
                static
                    address 192.0.2.1
                        group-prefix 239.255.0.0/16
                    exit
```

```
            exit
        exit
```

Step 7. Verify PIM neighbor relationship.

```
*A:PE-1# show router pim neighbor
===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-PE-1-PE-2           1               0d 00:00:32   0d 00:01:44     105
   192.168.12.2
int-PE-1-PE-3           1               0d 00:00:25   0d 00:01:21     105
   192.168.13.2
-------------------------------------------------------------------------------
Neighbors : 2
===============================================================================
*A:PE-1#
```

PE VPRN Configuration

A VPRN (VPRN 1) is created on each PE. This will be the multicast VPRN. PE-1 is the PE containing the attachment circuit toward CE-5. CE-5 is the CE nearest the source. PE-2, PE-3, and PE-4 contain attachment circuits toward CE-6, CE-7, and CE-8 respectively. Each CE has a receiving host attached.

Step 1. Create VPRN 1 on each PE, containing a route-distinguisher and vrf-target of 64496:1. The autonomous system number is 64496. Use **auto-bind-tunnel resolution-filter ldp** for next hop tunnel route resolution.

```
*A:PE-1# configure service vprn 1 customer 1 create
            autonomous-system 64496
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
```

Step 2. Create an attachment circuit interface toward the CE.

```
*A:PE-1# configure service vprn 1
            interface "int-PE-1-CE-5" create
                address 172.16.15.1/30
                sap 1/1/3 create
                exit
            exit
```

Step 3. The source address of the multicast stream will need to be reachable by all routers (PEs and CEs) within the VPN. This will be advertised within BGP from the CE to the PE. Create a BGP peering relationship within VPRN 1 on PE-1 with CE-5.

```
*A:PE-1# configure service vprn 1
            bgp
                group "EXTERNAL"
                    type external
                    peer-as 64497
                    neighbor 172.16.15.2
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

Step 4. On CE-5, create a VPRN to support the connection of the source to the CE and to connect the CE to the PE. Two attachment circuits are required, as well as a BGP peering relationship with the PE. This uses a default address family of **ipv4**.

(A pair of IES services could also be used to provide the attachment circuits.)

```
*A:CE-5# configure service vprn 1 customer 1 create
            autonomous-system 64497
            route-distinguisher 64497:1
            interface "int-CE-5-PE-1" create
                address 172.16.15.2/30
                sap 1/1/1 create
                exit
            exit
            interface "int-CE-5-S-5" create
                address 192.168.55.1/24
                sap 1/1/3 create
                exit
            exit
            bgp
                group "EXTERNAL"
                    type external
                    peer-as 64496
                    neighbor 172.16.15.1
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

Step 5. Verify PE-CE BGP peer relationship on CE-5 and PE-1:

```
*A:CE-5# show router 1 bgp summary
===============================================================================
 BGP Router ID:192.0.2.5        AS:64497       Local AS:64497
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
```

```
Total BGP Paths         : 2           Total Path Memory           : 368
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
172.16.15.1
                64496       4    0 00h00m56s 0/0/0 (IPv4)
                            4    0
-------------------------------------------------------------------------------
*A:CE-5#


*A:PE-1# show router 1 bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
Total BGP Paths         : 5           Total Path Memory           : 944
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
```

```
                  AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                     PktSent OutQ
-------------------------------------------------------------------------------
172.16.15.2
               64497       4    0 00h00m53s 0/0/0 (IPv4)
                           4    0
-------------------------------------------------------------------------------
*A:PE-1#
```

Step 6. In order for the CE connecting to the source to be advertised within BGP, a route policy is required. The subnet containing the multicast source is 192.168.55.0/24, so a prefix-list can be used to define a match, and then used within a route policy to inject into BGP.

```
*A:CE-5# configure router policy-options
           begin
           prefix-list "SOURCE-PREFIX"
               prefix 192.168.55.0/24 exact
           exit
           policy-statement "EXPORT-SOURCE-PREFIX-TO-BGP"
               entry 10
                   from
                       prefix-list "SOURCE-PREFIX"
                   exit
                   to
                       protocol bgp
                   exit
                   action accept
                   exit
               exit
           exit
           commit
```

Step 7. Apply this policy as an export policy within the BGP context.

```
*A:CE-5# configure service vprn 1
           bgp
               export "EXPORT-SOURCE-PREFIX-TO-BGP"
           exit
```

This results in the 192.168.55.0/24 subnet being seen in the BGP RIB_OUT on CE-5.

```
*A:CE-5# show router 1 bgp routes 192.168.55.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.5        AS:64497       Local AS:64497
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
---snip---
```

```
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : 192.168.55.0/24
Nexthop        : 172.16.15.2
Path Id        : None
To             : 172.16.15.1
Res. Nexthop   : n/a
Local Pref.    : n/a                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Origin         : IGP
AS-Path        : 64497
Route Tag      : 0
Neighbor-AS    : 64497
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:CE-5#
```

It is also seen in the PE-1 VRF 1 FIB:

```
*A:PE-1# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.15.0/30                                  Local   Local     00h05m11s  0
       int-PE-1-CE-5                                                0
172.16.26.0/30                                  Remote  BGP VPN   00h04m57s  170
       192.0.2.2 (tunneled)                                         0
172.16.37.0/30                                  Remote  BGP VPN   00h04m30s  170
       192.0.2.3 (tunneled)                                         0
172.16.48.0/30                                  Remote  BGP VPN   00h04m22s  170
       192.0.2.4 (tunneled)                                         0
192.168.55.0/24                                 Remote  BGP       00h02m08s  170
       172.16.15.2                                                  0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

This prefix will also be automatically advertised within the BGP VPRN to all other PEs, and will be installed in VRF 1.

For example, on PE-2:

```
*A:PE-2# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.15.0/30                                  Remote  BGP VPN   00h05m05s  170
       192.0.2.1 (tunneled)                                         0
172.16.26.0/30                                  Local   Local     00h05m08s  0
       int-PE-2-CE-6                                                0
172.16.37.0/30                                  Remote  BGP VPN   00h04m56s  170
       192.0.2.3 (tunneled)                                         0
172.16.48.0/30                                  Remote  BGP VPN   00h04m51s  170
       192.0.2.4 (tunneled)                                         0
192.168.55.0/24                                 Remote  BGP VPN   00h02m07s  170
       192.0.2.1 (tunneled)                                         0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

Each CE containing the multicast receivers must be able to reach the source. The following output shows the VPRN configuration of CE-6 containing an interface toward PE-2 and an interface toward Receiver 6 (H-6). A static route will suffice and is configured with next hop of the PE-2 PE-CE interface.

```
*A:CE-6# configure service vprn 1 customer 1 create
            autonomous-system 64498
            route-distinguisher 64498:1
            interface "int-CE-6-H-6" create
                address 192.168.66.1/24
                sap 1/1/2 create
                exit
            exit
            interface "int-CE-6-PE-2" create
                address 172.16.26.2/30
                sap 1/1/1 create
                exit
            exit
            static-route-entry 192.168.55.0/24
                next-hop 172.16.26.1
                    no shutdown
                exit
            exit
            no shutdown
```

PE VPRN Multicast Configuration

This section gives details of the VPRN configuration that allows the support of multicasting.

Sub-sections include:

1. Auto-discovery — This is the mechanism by which each PE advertises the presence of an MVPN to other PEs. This can be achieved using PIM or using BGP. This section covers PIM auto-discovery (auto-discovery using BGP is shown later).
2. Customer domain signaling — This discusses the mechanism of transporting customer signaling.
3. Data plane connectivity — This is the signaling of S-PMSIs within the provider domain to carry each individual customer multicast stream.

This chapter describes the PIM and BGP auto-discovery mechanisms in detail. For each of these, there is an example of customer domain signaling. For completion, a single example of S-PMSI creation is also shown.

Auto-Discovery within Provider Domain Using PIM

Each PE advertises its membership of a multicast VPN using PIM through the configuration of an Inclusive PMSI (I-PMSI). This is a multicast group that is common to each VPRN. The configuration for PE 1 and 2 is as follows:

```
*A:PE-1# configure service vprn 1
            mvpn
                provider-tunnel
                    inclusive
                        pim asm 239.255.255.1
                        exit
                    exit
                exit
            exit

*A:PE-2# configure service vprn 1
            mvpn
                provider-tunnel
                    inclusive
                        pim asm 239.255.255.1
                        exit
                    exit
                exit
            exit
```

The multicast group address used for the PMSI must be the same on all PEs for this VPRN instance.

Verify that PIM in the Global routing table (GRT) has signaled the I-PMSIs.

For the PE acting as the RP for global PIM:

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                                      3
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         3
   192.0.2.1                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  3
   192.0.2.2                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-3  3
   192.0.2.3                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  2
   192.0.2.4                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 5
===============================================================================
*A:PE-1#
```

This shows an incoming (S,G) join from all other PEs within the multicast VRF, plus an outgoing (*,G) join to the same PEs.

PE-3 will have the following PIM groups.

```
*A:PE-3# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                          int-PE-3-PE-1  1
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         2
   192.0.2.3                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-3#
```

This shows an (S,G) join toward the RP at 192.0.2.1, plus a (*,G) join from the RP. These represent the outgoing and incoming PIM interfaces for the VRF.

This results in a series of PIM neighbors through the I-PMSIs within the VRF, which are maintained using PIM hellos.

```
*A:PE-1# show router 1 pim neighbor
===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface                Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-PE-1-CE-5            1               0d 00:01:03   0d 00:01:43     105
   172.16.15.2
1-mt-239.255.255.1       1               0d 00:01:24   0d 00:01:25     105
   192.0.2.2
1-mt-239.255.255.1       1               0d 00:01:18   0d 00:01:31     105
   192.0.2.3
1-mt-239.255.255.1       1               0d 00:01:11   0d 00:01:37     105
   192.0.2.4
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
*A:PE-1#
```

PIM Auto-Discovery: Customer Signaling using PIM

Consider now how the signaling plane of the customer domain is dealt with at the provider domain.

The customer domain configuration covers the following three cases:

1. PIM ASM with the RP in the provider PE.
2. PIM ASM using anycast RP on the provider RPs.
3. PIM SSM.

PIM Any Source Multicasting with RP at the Provider PE

As each PE connects to a CE which will be part of the multicast VRF, it is necessary to enable PIM on each interface containing an attachment circuit toward a CE, and to configure the I-PMSI multicast tunnel for the VRF.

There is a requirement for an RP, as customer multicast signaling will be PIM-ASM.

The RP for the customer multicast will be on PE-2. In order to facilitate this, a loopback interface is created (called RP within the VPRN 1 context of PE-2, and will be advertised to all PEs. It must also be a PIM enabled interface.

The additional configuration for the RP on PE-2 is the following:

```
*A:PE-2# configure service vprn 1
            interface "RP" create
                address 10.2.3.5/32
                loopback
            exit
            pim
                interface "RP"
                exit
                rp
                    static
                        address 10.2.3.5
                            group-prefix 225.0.0.0/8
                        exit
                    exit
                exit
                no shutdown
            exit
```

The RP must also be configured on each of the PEs and CEs.

On PE-3, the configuration is as follows:

```
*A:PE-3# configure service vprn 1
            pim
                interface "int-PE-3-CE-7"
                exit
                rp
                    static
                        address 10.2.3.5
                            group-prefix 225.0.0.0/8
                        exit
                    exit
                exit
            exit
```

The configuration on the other nodes is similar; only the interfaces are different.

Customer Edge Router Multicast Configuration

Each CE router will have a PIM neighbor peer relationship with its nearest PE.

The CE router (CE-5) containing the source will have PIM enabled on the interface connected to the source. It will also have a static RP entry, as the incoming sources need to be registered with the RP.

```
*A:CE-5# configure service vprn 1
            pim
                interface "int-CE-5-PE-1"
                exit
                interface "int-CE-5-S-5"
```

```
            exit
            rp
                static
                    address 10.2.3.5
                        group-prefix 225.0.0.0/8
                    exit
                exit
            exit
        exit
```

The CE containing the receivers will have IGMP enabled on the interface connected to the receivers. Once again, there needs to be an RP configured, as the router needs to issue PIM joins to the RP.

```
*A:CE-6# configure service vprn 1
        static-route-entry 10.0.0.0/8
            next-hop 172.16.26.1
                no shutdown
            exit
        exit
        static-route-entry 192.168.55.0/24
            next-hop 172.16.26.1
                no shutdown
            exit
        exit
        igmp
            interface "int-CE-6-H-6"
            exit
        exit
        pim
            interface "int-CE-6-PE-2"
            exit
            rp
                static
                    address 10.2.3.5
                        group-prefix 225.0.0.0/8
                    exit
                exit
            exit
        exit
```

Traffic Flow

The source sends a multicast stream using group address 225.0.0.1 toward CE-5. As the group matches the group address in the static RP configuration, the router sends a register join toward the RP. At this time, no receivers are interested in the group, so there are no entries in the Outgoing Interface List (OIL), and the number of outgoing interfaces (OIFs) is zero.

The PIM status of CE-5 within VPN 1 is as follows:

```
*A:CE-5# show router 1 pim group
===============================================================================
```

```
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
225.0.0.1                   (S,G)                        int-CE-5-S-5   0
   192.168.55.2                10.2.3.5
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-5#
```

The receiver H-6 connected to CE-6, wishes to join the group 225.0.0.1, and sends an IGMPv2 report toward CE-6. CE-6 recognizes the report, which contains no source.

```
*A:CE-6# show router 1 igmp group
===============================================================================
IGMP Interface Groups
===============================================================================

(*,225.0.0.1)                                                 UpTime: 0d 00:00:05
    Fwd List  : int-CE-6-H-6
-------------------------------------------------------------------------------
Entries : 1
===============================================================================
IGMP Host Groups
===============================================================================
No Matching Entries
===============================================================================
IGMP SAP Groups
===============================================================================
No Matching Entries
===============================================================================
*A:CE-6#
```

CE-6 is not aware of the source of the group so initiates a (*,G) PIM join toward the RP.

At the RP, a (*,G) join is received.

```
*A:PE-2# show router 1 pim group 225.0.0.1 type starg detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.0.0.1
Source Address     : *
RP Address         : 10.2.3.5
Advt Router        : 192.0.2.2
Flags              :                       Type               : (*,G)
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : self
Keepalive Timer    : Not Running
```

```
Up Time            : 0d 00:00:20        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:39
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor       :
Incoming Intf      :
Outgoing Intf List : int-PE-2-CE-6

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 0                  Discarded Packets  : 0
Forwarded Octets   : 0                  RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-2#
```

The RP can now forward traffic from itself toward CE-6, as the outgoing interface is seen as int-PE-2-CE-6.

CE-6 is now able to determine the source from the traffic stream, so it initiates a Reverse Path Forwarding (RPF) lookup of the source address in the route table, and issues an (S,G) PIM join toward the source.

The join is propagated across the provider network, from PE-2 toward PE-1 which is the resolved RPF next hop for the source.

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.0.0.1
Source Address     : 192.168.55.2
RP Address         : 10.2.3.5
Advt Router        : 172.16.15.2
Flags              : spt                   Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.15.2
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:01:15        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:44
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.15.2
Incoming Intf      : int-PE-1-CE-5
Outgoing Intf List : 1-mt-239.255.255.1

Curr Fwding Rate   : 7524.9 kbps
Forwarded Packets  : 1531063            Discarded Packets  : 0
```

```
Forwarded Octets   : 70428898           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The outgoing interface is the I-PMSI: 1-mt-239.255.255.1.

The join is received by CE-5, which contains the subnet of the source.

CE-5 now recognizes the multicast group as a valid stream. This becomes the root of the shortest path tree for the group.

```
*A:CE-5# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
225.0.0.1                   (S,G)               spt      int-CE-5-S-5   1
   192.168.55.2                10.2.3.5
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-5#
```

For completion, consider a second receiver H-7 interested in group 225.0.0.1. The IGMPv2 report is translated into a (*,G) PIM join at CE-7 toward the RP.

```
*A:CE-7# show router 1 pim group type starg
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
225.0.0.1                   (*,G)                        int-CE-7-PE-3  1
   *                           10.2.3.5
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-7#
```

At the RP (PE-2), there is now a second interface in the OIL.

```
*A:PE-2# show router 1 pim group 225.0.0.1 type starg detail
===============================================================================
PIM Source Group ipv4
```

```
===============================================================================
Group Address      : 225.0.0.1
Source Address     : *
RP Address         : 10.2.3.5
Advt Router        : 192.0.2.2
Flags              :                     Type               : (*,G)
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : self
Keepalive Timer    : Not Running
Up Time            : 0d 00:02:05         Resolved By        : rtable-u

Up JP State        : Joined              Up JP Expiry       : 0d 00:00:55
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor       :
Incoming Intf      :
Outgoing Intf List : int-PE-2-CE-6, 1-mt-239.255.255.1

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 0                   Discarded Packets  : 0
Forwarded Octets   : 0                   RPF Mismatches     : 0
Spt threshold      : 0 kbps              ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-2#
```

The second interface is the I-PMSI, which is the multicast tunnel toward all other PEs. At PE-3, the (*,G) join has the I-PMSI as an incoming interface, and the PE-CE interface as the outgoing interface.

```
*A:PE-3# show router 1 pim group type starg detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.0.0.1
Source Address     : *
RP Address         : 10.2.3.5
Advt Router        : 192.0.2.2
Flags              :                     Type               : (*,G)
MRIB Next Hop      : 192.0.2.2
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:00:32         Resolved By        : rtable-u

Up JP State        : Joined              Up JP Expiry       : 0d 00:00:28
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor       : 192.0.2.2
Incoming Intf      : 1-mt-239.255.255.1
Outgoing Intf List : int-PE-3-CE-7

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 168                 Discarded Packets  : 0
Forwarded Octets   : 7728                RPF Mismatches     : 0
Spt threshold      : 0 kbps              ECMP opt threshold : 7
```

```
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-3#
```

Once again, as the CE receives traffic from the group, it can use the source address in the packet to initiate an (S,G) join toward the source it joins the shortest path tree (SPT).

```
*A:CE-7# show router 1 pim group type sg detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.0.0.1
Source Address     : 192.168.55.2
RP Address         : 10.2.3.5
Advt Router        :
Flags              : spt               Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.37.1
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:02:40
Up Time            : 0d 00:00:52       Resolved By        : rtable-u

Up JP State        : Joined            Up JP Expiry       : 0d 00:00:08
Up JP Rpt          : Not Pruned        Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

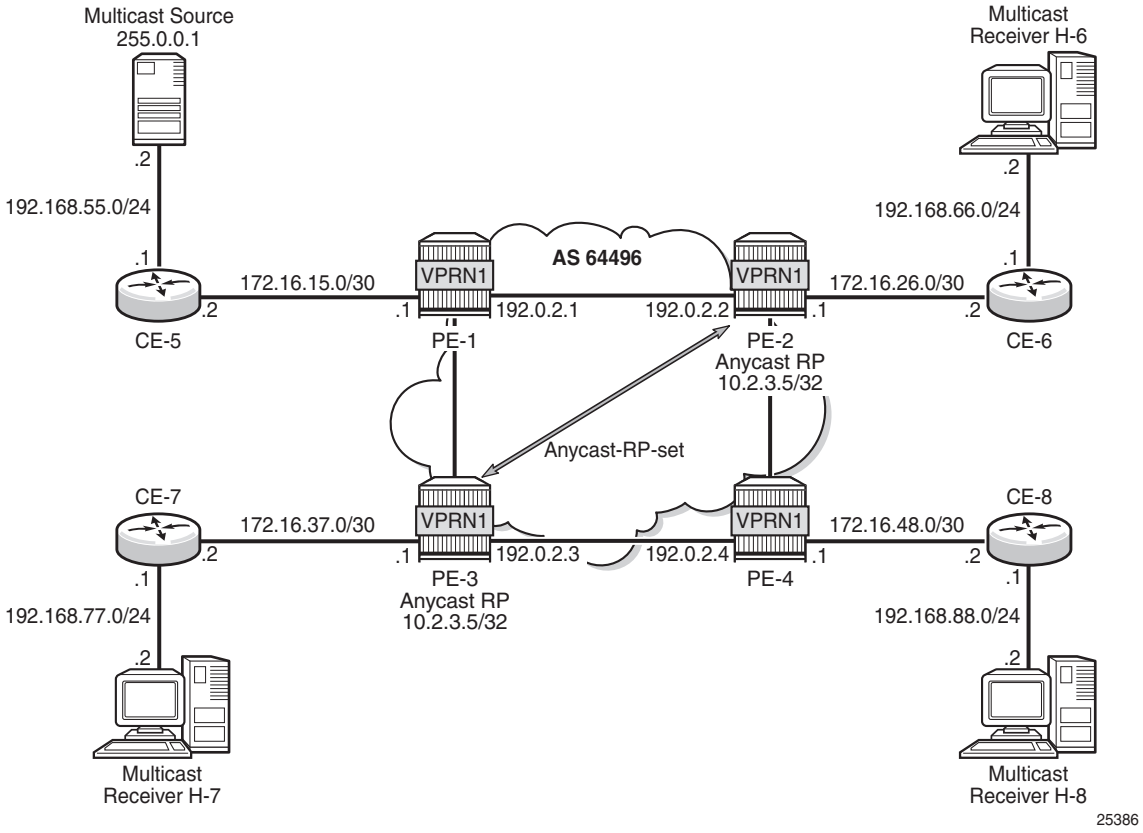
Rpf Neighbor       : 172.16.37.1
Incoming Intf      : int-CE-7-PE-3
Outgoing Intf List : int-CE-7-H-7

Curr Fwding Rate   : 5481.9 kbps
Forwarded Packets  : 791128            Discarded Packets  : 0
Forwarded Octets   : 36391888          RPF Mismatches     : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-7#
```

PIM Any Source Multicasting with Anycast RP at the Provider PE

The network topology is displayed in Figure 325. The setup consists of four 7750 SRs acting as Provider Edge (PE) routers within a single Autonomous System (AS).

Figure 325 **Network Topology for Anycast RP**



Connected to each PE is a single 7750 acting as a Customer Edge (CE) router. CE-5 has a single multicast source connected, and PEs 2 to 4 each have a single receiver connected which will receive the multicast stream from the source. In this section, each receiver is IGMPv2 capable, and will issue IGMPv2 reports. A Rendezvous Point is required by the C-signaling plane to resolve each (*,G) group state into an (S,G) state. In this case, two RPs are chosen to form an Anycast set to resolve each (*,G) group into an (S,G) state.

Multicast traffic from the source group 225.0.0.1 is streamed toward router CE-5. Receivers connected to PE-2, PE-3 and PE-4 are interested in joining this multicast group.

Anycast RP - PE VPRN Configuration

As previously stated, there is a requirement for an RP, as Customer Multicast signaling will be PIM-ASM and IGMPv2.

In this case, an Anycast RP will be used. This is configured on PE-2 and PE-3, and an anycast set is created.

As each PE contains a CE which will be part of the multicast VRF, it is necessary to enable PIM on each interface containing an attachment circuit toward a CE, and to configure the I-PMSI multicast tunnel for the VRF.

The following output shows the VPRN configuration for PE-2 containing the RP and anycast RP configuration. The loopback interface lo1 is used for inter-RP communication:

```
*A:PE-2# configure service vprn 1
            interface "RP" create
                address 10.2.3.5/32
                loopback
            exit
            interface "lo1" create
                address 10.0.0.2/32
                loopback
            exit
            pim
                interface "int-PE-2-CE-6"
                exit
                interface "RP"
                exit
                interface "lo1"
                exit
                rp
                    static
                        address 10.2.3.5
                            group-prefix 225.0.0.0/8
                        exit
                    exit
                    anycast 10.2.3.5
                        rp-set-peer 10.0.0.2
                        rp-set-peer 10.0.0.3
                    exit
                exit
                no shutdown
            exit
```

Similarly, the VPRN configuration for PE-3 is:

```
*A:PE-3# configure service vprn 1
            interface "RP" create
                address 10.2.3.5/32
                loopback
            exit
            interface "lo1" create
                address 10.0.0.3/32
                loopback
            exit
            pim
                interface "int-PE-3-CE-7"
                exit
                interface "RP"
```

```
            exit
            interface "lo1"
            exit
            rp
                static
                    address 10.2.3.5
                        group-prefix 225.0.0.0/8
                    exit
                exit
                anycast 10.2.3.5
                    rp-set-peer 10.0.0.2
                    rp-set-peer 10.0.0.3
                exit
            exit
            no shutdown
        exit
```

As previously stated, there is a requirement for an RP, as customer multicast signaling will be PIM-ASM and IGMPv2.

In this case, an anycast RP will be used. This is configured on PE-2 and PE-3, and an anycast set is created.

The anycast address will be 10.2.3.5/32 and is created as an interface called **RP** on both PE-2 and PE-3.

An additional loopback interface, called **lo1** is created on each VPRN on PEs containing the anycast address. These are used as source addresses for communication between the routers within the RP set. These addresses will be automatically advertised to all PEs as vpn-ipv4 addresses, and will be installed in the VRF 1 forwarding table of all PEs containing VPRN 1.

Note: All routers containing RP must have their own loopback address included in the RP set as well as all peer routers.

The multicast group address used for the Inclusive PMSI is chosen to be 239.255.255.1 and must be the same on all PEs for this VPRN instance. This is analogous to the MDT within the Rosen MVPN implementation.

```
*A:PE-2# configure service vprn 1
            mvpn
                provider-tunnel
                    inclusive
                        pim asm 239.255.255.1
                        exit
                    exit
                exit
            exit
```

Verify that PIM in the global routing table (GRT) has signaled the I-PMSIs.

For the PE acting as the RP for global PIM:

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                                       3
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         3
   192.0.2.1                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  3
   192.0.2.2                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-3  3
   192.0.2.3                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  2
   192.0.2.4                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 5
===============================================================================
*A:PE-1#
```

PE-3 will have:

```
*A:PE-3# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                           int-PE-3-PE-1  1
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         2
   192.0.2.3                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-3#
```

This shows a (S,G) join toward the RP at 192.0.2.1, plus a (*,G) join from RP. These represent the outgoing and incoming PIM interfaces for the VRF.

This results in a series of PIM neighbors through the I-PMSIs within the VRF, which are maintained using PIM hellos.

```
*A:PE-1# show router 1 pim neighbor
===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
```

```
-------------------------------------------------------------------------------
int-PE-1-CE-5           1               0d 00:08:01   0d 00:01:15     105
   172.16.15.2
1-mt-239.255.255.1      1               0d 00:08:22   0d 00:01:27     105
   192.0.2.2
1-mt-239.255.255.1      1               0d 00:08:15   0d 00:01:33     105
   192.0.2.3
1-mt-239.255.255.1      1               0d 00:08:09   0d 00:01:39     105
   192.0.2.4
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
*A:PE-1#
```

Verify PIM RP set on PE-2 (similar for PE-3):

```
*A:PE-2# show router 1 pim anycast
===============================================================================
PIM Anycast RP Entries ipv4
===============================================================================
Anycast RP                              Anycast RP Peer
-------------------------------------------------------------------------------
10.2.3.5                                10.0.0.2
                                        10.0.0.3
-------------------------------------------------------------------------------
PIM Anycast RP Entries : 2
===============================================================================
*A:PE-2#
```

Anycast RP — Customer Edge Router Multicast Configuration

Each CE router will have a PIM neighbor peer relationship with its nearest PE.

The CE router (CE-5) containing the source will have PIM enabled on the interface connected to the source.

```
*A:CE-5# configure service vprn 1
            pim
                interface "int-CE-5-PE-1"
                exit
                interface "int-CE-5-S-5"
                exit
                rp
                    static
                        address 10.2.3.5
                            group-prefix 225.0.0.0/8
                        exit
                    exit
                exit
            exit
```

The CE containing the receivers will have IGMP enabled on the interface connected to the receivers.

```
*A:CE-6# configure service vprn 1
            igmp
                interface "int-CE-6-H-6"
                exit
            exit
```

Traffic Flow

Figure 326 **IGMP and PIM Control Messaging Schematic**

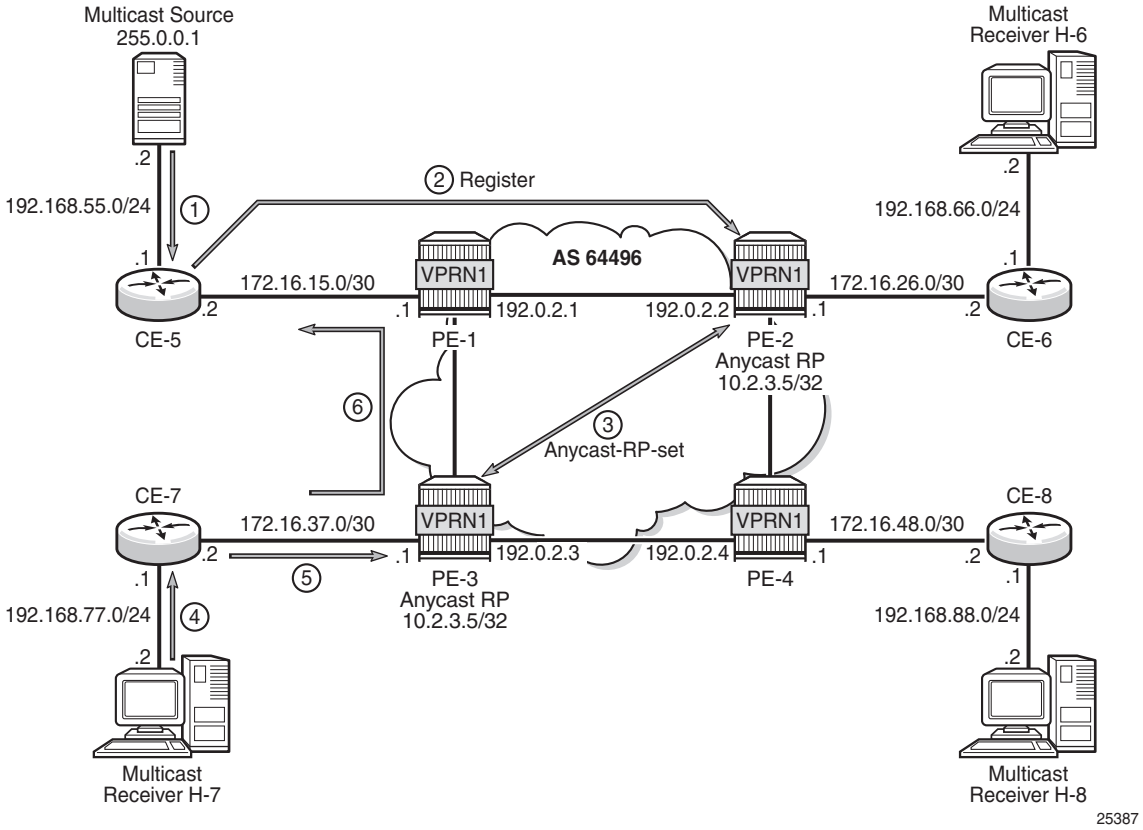


Figure 326 shows the sequence of IGMP and PIM control messaging.

1. The source multicasts a stream with group address 225.0.0.1 toward CE-5.
2. CE-5 matches the group with the group address prefix in the static RP configuration and sends a register message toward the RP.

```
*A:CE-5# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.0.0.1
Source Address     : 192.168.55.2
```

```
RP Address          : 10.2.3.5
Advt Router         : 192.0.2.5
Flags               :                       Type               : (S,G)
Mode                : sparse
MRIB Next Hop       : 192.168.55.2
MRIB Src Flags      : direct
Keepalive Timer Exp: 0d 00:03:07
Up Time             : 0d 00:00:23           Resolved By        : rtable-u

Up JP State         : Not Joined            Up JP Expiry       : 0d 00:00:00
Up JP Rpt           : Not Joined StarG      Up JP Rpt Override : 0d 00:00:00

Register State      : Pruned                Register Stop Exp  : 0d 00:00:32
Reg From Anycast RP: No

Rpf Neighbor        : 192.168.55.2
Incoming Intf       : int-CE-5-S-5
Outgoing Intf List :
Outgoing Sap List  :
Outgoing Host List :

Curr Fwding Rate    : 19716.7 kbps
Forwarded Packets   : 1221836               Discarded Packets  : 0
Forwarded Octets    : 56204456              RPF Mismatches     : 0
Spt threshold       : 0 kbps                ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-5#
```

The register message is sent to the nearest RP, the RP with the lowest IGP cost.

As the register is sent through PE-1, it is PE-1 that determines which RP will receive the message.

```
*A:PE-1# show router 1 route-table 10.2.3.5/32
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.2.3.5/32                                     Remote  BGP VPN   00h00m59s  170
       192.0.2.2 (tunneled)                                          0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-1#
```

The PE which will receive the register is 192.0.2.2 (PE-2). The PIM group status on PE-2 is:

```
*A:PE-2# show router 1 pim group
```

```
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
225.0.0.1                   (S,G)                          1-mt-239.255.* 0
   192.168.55.2                10.2.3.5
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-2#
```

This shows that RP is aware of the (S,G) status of the group 225.0.0.1, and becomes a root of a shared tree for this group. The Outgoing Interface List (OIL) is empty.

3. PE-2 will now send a register message to all other RPs within the anycast set, in this case to PE-3 (which has VPRN 1 containing address 10.0.0.3).

   The PIM status of the group 225.0.0.1 on PE-3 is:

```
*A:PE-3# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
225.0.0.1                   (S,G)                          1-mt-239.255.* 0
   192.168.55.2                10.2.3.5
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

Now both PEs within the RP set for VPRN have an (S,G) state for 225.0.0.1.

4. The receiver H-7, wishes to join the group 225.0.0.1, and sends in an IGMPv2 report toward CE-7. CE-7 recognizes the report, but has no PIM state for this group.
5. CE-7 sends a PIM join toward the RP, in this case the nearest RP will be PE-7.

   PE-3 already has (S,G) state for this group, so will forward traffic toward receiver H-7.
6. CE-7 does a Reverse Path Forwarding (RPF) lookup of the source address in the route table, and issues a PIM join toward the source.

   The join is propagated across the provider network toward PE-1, which is the resolved RPF next hop for the source.

```
*A:CE-7# show router 1 pim group type sg detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.0.0.1
Source Address     : 192.168.55.2
RP Address         : 10.2.3.5
Advt Router        :
Flags              : spt               Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.37.1
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:03:10
Up Time            : 0d 00:00:22       Resolved By        : rtable-u

Up JP State        : Joined            Up JP Expiry       : 0d 00:00:38
Up JP Rpt          : Not Pruned        Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.37.1
Incoming Intf      : int-CE-7-PE-3
Outgoing Intf List : int-CE-7-H-7

Curr Fwding Rate   : 6803.4 kbps
Forwarded Packets  : 406012            Discarded Packets  : 0
Forwarded Octets   : 18676552          RPF Mismatches     : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-7#
```

The join is received by CE-5, which contains the subnet of the source.

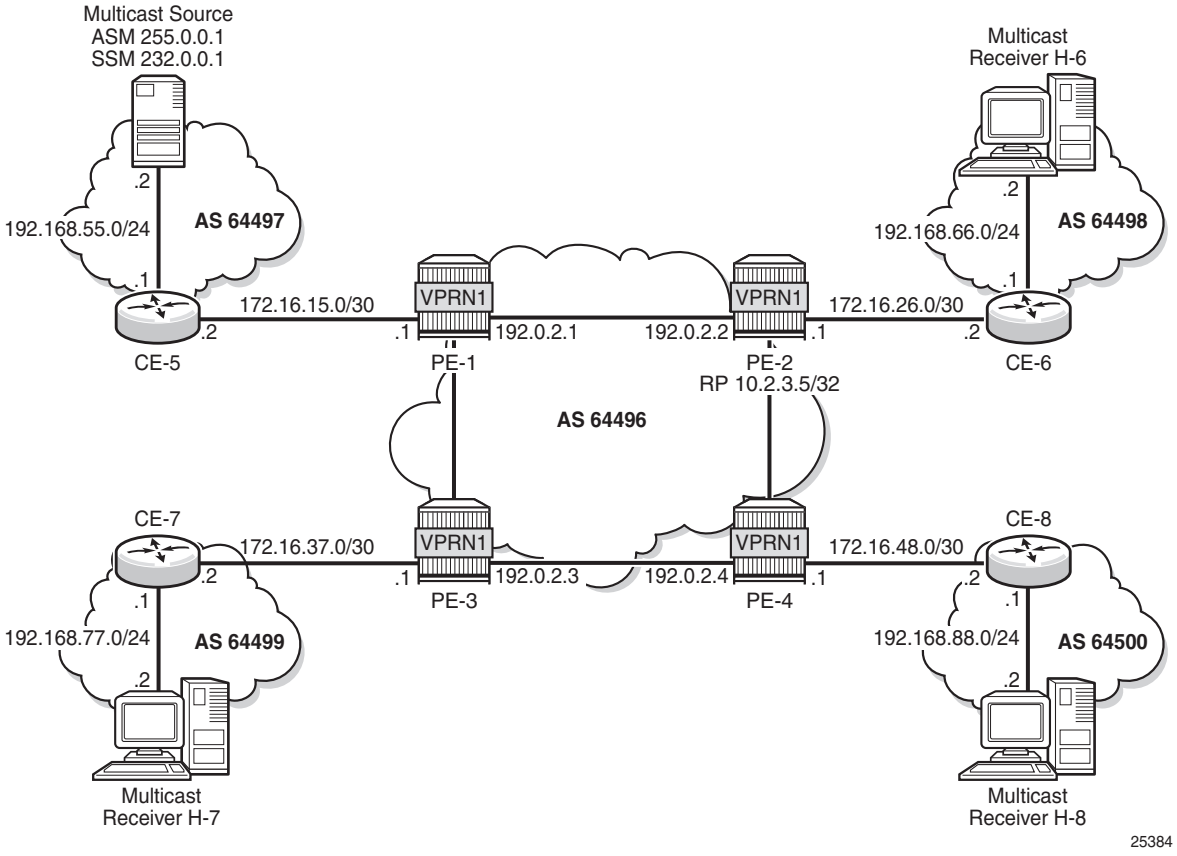
CE-5 now recognizes the multicast group as a valid stream. CE-5 becomes the root of the shortest path tree for the group.

```
*A:CE-5# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type               Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
225.0.0.1                   (S,G)              spt      int-CE-5-S-5   1
   192.168.55.2                10.2.3.5
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-5#
```

PIM Source-Specific Multicasting

There is no requirement for an RP, as customer multicast signaling will be PIM-SSM. The Multicast group address used for the PMSI must be the same on all PEs for this VPRN instance.

Figure 327 **PIM SSM in Customer Signaling Plane**



The following output shows the VPRN configuration for PIM and MVPN for PE-1.

```
*A:PE-1# configure service vprn 1
            pim
                interface "int-PE-1-CE-5"
            exit
            mvpn
                provider-tunnel
                    inclusive
                        pim asm 239.255.255.1
                        exit
                    exit
                exit
            exit
```

There is a similar configuration required for each of the other PEs. Verify that PIM in the Global routing table (GRT) has signaled the I-PMSIs.

For the PE acting as the RP for global PIM:

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                                       3
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         3
   192.0.2.1                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  3
   192.0.2.2                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-3  3
   192.0.2.3                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  2
   192.0.2.4                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 5
===============================================================================
*A:PE-1#
```

PE-3 will have:

```
*A:PE-3# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                           int-PE-3-PE-1  1
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         2
   192.0.2.3                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-3#
```

This shows a (S,G) join toward the RP at 192.0.2.1, plus a (*,G) join from RP. These represent the outgoing and incoming PIM interfaces for the VRF.

This results in a series of PIM neighbors through the I-PMSIs within the VRF, which are maintained using PIM hellos.

```
*A:PE-1# show router 1 pim neighbor
===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-PE-1-CE-5           1               0d 00:08:01   0d 00:01:15     105
   172.16.15.2
1-mt-239.255.255.1      1               0d 00:08:22   0d 00:01:27     105
   192.0.2.2
1-mt-239.255.255.1      1               0d 00:08:15   0d 00:01:33     105
   192.0.2.3
1-mt-239.255.255.1      1               0d 00:08:09   0d 00:01:39     105
   192.0.2.4
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
*A:PE-1#
```

PIM SSM — Customer Edge Router Multicast Configuration

Each CE router will have a PIM neighbor peer relationship with its nearest PE.

The CE router (CE-5) containing the source will have PIM enabled on the interface connected to the source.

```
*A:CE-5# configure service vprn 1
            pim
                interface "int-CE-5-PE-1"
                exit
                interface "int-CE-5-S-5"
                exit
            exit
```

The CE containing the receivers will have IGMP enabled on the interface connected to the receivers and PIM on the interface facing the PE.

```
*A:CE-6# configure service vprn 1
            static-route 192.168.55.0/24 next-hop 172.16.26.1
            igmp
                interface "int-CE-6-H-6"
                exit
            exit
            pim
                interface "int-CE-6-PE-2"
                exit
            exit
```

Traffic Flow

The source multicasts a stream with group address 232.0.0.1 toward CE-5. As there is no receiver interested in the group at this time, there are no outgoing interfaces, so the Outgoing Interface List (OIL) is empty.

```
*A:CE-5# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
232.0.0.1                   (S,G)                        int-CE-5-S-5   0
   192.168.55.2
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-5#
```

The receiver H-6, wishes to join the group 232.0.0.1, and so sends in an IGMPv3 report toward CE-6. CE-6 recognizes the report, which contains the source 192.168.55.2 in the include filter list.

```
*A:CE-6# show router 1 igmp group
===============================================================================
IGMP Interface Groups
===============================================================================

(192.168.55.2,232.0.0.1)                                 UpTime: 0d 00:00:05
    Fwd List  : int-CE-6-H-6
-------------------------------------------------------------------------------
Entries : 1
===============================================================================
IGMP Host Groups
===============================================================================
No Matching Entries
===============================================================================
IGMP SAP Groups
===============================================================================
No Matching Entries
===============================================================================
*A:CE-6#
```

CE-6 does a RPF lookup of the source address in the route table, and issues a PIM join toward the source.

The join is propagated across the provider network, toward PE-1 which is the resolved RPF next hop for the source.

```
*A:PE-1# show router 1 pim group detail
===============================================================================
```

```
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 192.168.55.2
RP Address         : 0
Advt Router        : 172.16.15.2
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.15.2
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:00:12        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:47
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.15.2
Incoming Intf      : int-PE-1-CE-5
Outgoing Intf List : 1-mt-239.255.255.1

Curr Fwding Rate   : 7854.4 kbps
Forwarded Packets  : 255250             Discarded Packets  : 0
Forwarded Octets   : 11741500           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The outgoing interface is the I-PMSI: 1-mt-239.255.255.1.

The join is received by CE-5, which contains the subnet of the source.

CE-5 now recognizes the multicast group as a valid stream. CE-5 becomes the root of the shortest path tree for the group.

```
*A:CE-5# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
232.0.0.1                   (S,G)                       int-CE-5-S-5   1
   192.168.55.2
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:CE-5#
```

PE BGP Auto-Discovery

Discovery of multicast-enabled virtual private networks (MVPNs) can also be achieved using BGP. To this end, any PE that is a member of a multicast VPN will advertise this using a BGP multi-protocol network layer reachability information (NLRI) update that is sent to all PEs within the AS. This update will contain an intra-AS I-PMSI auto-discovery route type, also known as an Intra-AD. These use an address family — mvpn-ipv4 — so each PE must be configured to originate and accept such updates. The following needs to be modified in the BGP context for all PE nodes:

```
configure router bgp
            group "INTERNAL"
                family vpn-ipv4 mvpn-ipv4
                exit all
```

This is achieved in the global routing table within the BGP context.

This allows each BGP speaker to advertise its capabilities within a BGP Open message.

When the peers become established, the address family capabilities should look like the following:

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 18          Total Path Memory           : 3632
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 1           Total VPN Peers             : 1
Total VPN Local Rts     : 3
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
```

```
Total RouteTgt Rem Rts  : 0          Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0          Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0          Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0          Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0          Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0          Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0          Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0          Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0          Total LblIpv6 Bkp Rts       : 0

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
                64496      60    0 00h00m22s 0/0/0 (VpnIPv4)
                           17    0           0/0/0 (MvpnIPv4)
192.0.2.3
                64496      60    0 00h00m20s 0/0/0 (VpnIPv4)
                           16    0           0/0/0 (MvpnIPv4)
192.0.2.4
                64496      60    0 00h00m16s 0/0/0 (VpnIPv4)
                           15    0           0/0/0 (MvpnIPv4)
-------------------------------------------------------------------------------
*A:PE-1#
```

BGP Auto-Discovery — PE VPRN Multicast Configuration

As each PE contains a CE which will be part of the multicast VRF, it is necessary to enable PIM on each interface containing an attachment circuit toward a CE, and to configure the I-PMSI multicast tunnel for the VRF.

In order for the BGP routes to be accepted into the VRF, a route-target community is required (vrf-target). This is configured in the **configure service vprn 1 mvpn** context and, in this case, is set to the same value as the unicast vrf-target, the vrf-target community as the **configure service vprn 1 vrf-target** context.

On each PE, the MVPN context of the VPRN instance is configured as follows:

```
*A:PE-2# configure service vprn 1
           mvpn
               auto-discovery default
               provider-tunnel
                   inclusive
                       pim asm 239.255.255.1
                       exit
                   exit
```

```
            exit
            vrf-target unicast
            exit
        exit
```

The multicast group address used for the PMSI must be the same on all PEs for this VPRN instance.

The presence of auto-discovery will initiate BGP updates between the PEs that contain an MVPN, such as Intra-AD MVPN routes, are generated and advertised to each peer

```
*A:PE-1# show router bgp routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                     192.0.2.2              100         0
      64496:1                      -                                  -
      192.0.2.2                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.3              100         0
      64496:1                      -                                  -
      192.0.2.3                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.4              100         0
      64496:1                      -                                  -
      192.0.2.4                    -
      No As-Path                   -
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

This shows that PE-1 has received an Intra-AD route from each of the other PEs, each of which has multicast VPRN 1 configured.

Examining one of the Intra-AD routes from PE-2 shows that the route-target community matches the unicast VRF-target (64496:1), and also that the PMSI tree has a multicast group address of 239.255.255.1, which matches the I-PMSI group configuration on PE-1.

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad originator-ip 192.0.2.2
```

```
        detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Original Attributes

Route Type     : Intra-Ad
Route Dist.    : 64496:1
Originator IP  : 192.0.2.2
Nexthop        : 192.0.2.2
From           : 192.0.2.2
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : no-export target:64496:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h00m40s
VPRN Imported  :  1
-------------------------------------------------------------------------------
PMSI Tunnel Attribute :
Tunnel-type    : PIM-SM Tree
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 0
Sender         : 192.0.2.2              P-Group        : 239.255.255.1
-------------------------------------------------------------------------------
---snip---

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

Verify that PIM in the global routing table (GRT) has signaled the I-PMSIs.

For the PE acting as the RP for global PIM:

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
```

```
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                                       3
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         3
   192.0.2.1                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  3
   192.0.2.2                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-3  3
   192.0.2.3                   192.0.2.1
239.255.255.1               (S,G)               spt      int-PE-1-PE-2  2
   192.0.2.4                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 5
===============================================================================
*A:PE-1#
```

This shows an incoming (S,G) join from all other PEs within the multicast VRF, plus an outgoing (*,G) join to the same PEs.

PE-3 will have the following PIM groups:

```
*A:PE-3# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.255.1               (*,G)                          int-PE-3-PE-1  1
   *                           192.0.2.1
239.255.255.1               (S,G)               spt      system         2
   192.0.2.3                   192.0.2.1
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-3#
```

This shows a (S,G) join toward the RP at 192.0.2.1, plus a (*,G) join from RP. These represent the outgoing and incoming PIM interfaces for the VRF.

This results in a series of PIM neighbors through the I-PMSIs within the VRF. As the neighbors were discovered using BGP (rather than with PIM as per Rosen MVPN), there are no PIM hellos exchanged.

```
*A:PE-1# show router 1 pim neighbor
===============================================================================
PIM Neighbor ipv4
===============================================================================
```

```
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-PE-1-CE-5           1               0d 00:16:24   0d 00:01:22     105
   172.16.15.2
1-mt-239.255.255.1      1               0d 00:01:04   never           65535
   192.0.2.2
1-mt-239.255.255.1      1               0d 00:01:03   never           65535
   192.0.2.3
1-mt-239.255.255.1      1               0d 00:00:53   never           65535
   192.0.2.4
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
*A:PE-1#
```

BGP Auto-Discovery — Customer Signaling Domain

As the customer signaling is independent from the provider PE discovery mechanism, all of the customer signaling techniques described when using PIM for auto-discovery within provider domain are also applicable when using BGP for auto-discovery, namely

- PIM Any Source Multicasting with RP at the provider PE
- PIM Any Source Multicasting with Anycast RP at the provider PE
- PIM Source Specific Multicasting

Data Path Using Selective PMSI

When a configurable data threshold for a multicast group has been exceeded, multicast traffic across the provider network can be switched to a selective PMSI (S-PMSI).

This has to be configured as a separate group and must contain a threshold which, if exceeded, will see a new PMSI signaled by the PE nearest the source, and traffic switched onto the S-PMSI.

```
*A:PE-1# configure service vprn
            mvpn
                provider-tunnel
                    inclusive
                        pim asm 239.255.255.1
                        exit
                    exit
                    selective
                        data-threshold 232.0.0.0/8 1
                        pim-ssm 232.255.1.0/24
```

```
                exit
            exit
        exit
```

This shows that when the traffic threshold for multicast groups covered by the range 232.0.0.0/8 exceeds 1 kbps between a pair of PEs, then an S-PMSI is signaled between the PEs. This is a separate multicast tunnel over which traffic in the given group now flows.

```
*A:PE-1# show router 1 pim s-pmsi detail
===============================================================================
PIM Selective provider tunnels
===============================================================================
Md Source Address  : 192.0.2.1          Md Group Address   : 232.255.1.0
Number of VPN SGs  : 1                  Uptime             : 0d 00:00:16
MT IfIndex         : 16389

VPN Group Address  : 232.0.0.1
VPN Source Address : 192.168.55.2
State              : TX Joined          Mdt Threshold      : 1
Join Timer         : 0d 00:01:02        Holddown Timer     : 0d 00:00:44
===============================================================================
PIM Selective provider tunnels Interfaces : 1
===============================================================================
*A:PE-1#
```

In this example, the (S,G) group is (192.168.55.2, 232.0.0.1). As the data rate has exceeded the configured MDT threshold of 1kbps, a new provider tunnel with a group address of 232.255.1.0 has been signaled and now carries the multicast stream.

The TX Joined state indicates that the S-PMSI has been sourced at this PE — PE-1.

Comparing this to PE-3, where a receiver is connected through a CE indicates that it has received a join to connect the S-PMSI.

```
*A:PE-3# show router 1 pim s-pmsi detail
===============================================================================
PIM Selective provider tunnels
===============================================================================
Md Source Address  : 192.0.2.1          Md Group Address   : 232.255.1.0
Number of VPN SGs  : 1                  Uptime             : 0d 00:00:24
MT IfIndex         : 24576              Egress Fwding Rate : 7790.4 kbps

VPN Group Address  : 232.0.0.1
VPN Source Address : 192.168.55.2
State              : RX Joined
Expiry Timer       : 0d 00:02:36
===============================================================================
PIM Selective provider tunnels Interfaces : 1
===============================================================================
*A:PE-3#
```

Conclusion

This chapter provides configuration on how to configure multicast within a VPRN with next generation multicast VPN techniques. Specifically, discovery of multicast VPNs using PIM and BGP auto-discovery mechanisms are described with a number of ASM and SSM signaling techniques within the customer domain.

NG-MVPN Configuration with MPLS

This chapter provides information about NG-MVPN configuration with MPLS.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 9.0.R5, but the CLI in this edition corresponds to release 14.0.R5. There are no prerequisites for this configuration.

Summary

Multicast VPN (MVPN) or Next Generation IP Multicast in an IP-VPN (NG-MVPNs) architectures describe a set of virtual routing and forwarding (VRFs) or virtual private routed networks (VPRNs) that support the transport of multicast traffic across a provider network. MVPNs are defined in RFC 6513, *Multicast in MPLS/BGP IP VPNs*, and RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/IP VPNs*.

Initial MVPN deployments were originally based on Rosen MVPN (RFC 6037) which described the protocols and procedures required to support an IP Multicast VPN. There were a number of limitations with the Rosen MVPN implementation including, but not limited to:

- Rosen MVPN requires a set of multicast distribution trees (MDTs) per VPN, which requires a PIM state per MDT. There is no option to aggregate MDTs across multiple VPNs.

- Customer signaling. Initially, PE discovery and Data MDT signaling were all PIM-based because there was no mechanism available to decouple these. Now, PE discovery is supported using a BGP MDT address family identifier/ subsequent address family identifier (AFI/SAFI), however, the data MDT still needs PIM.
- There is no mechanism for using MPLS to encapsulate multicast traffic in the VPN. GRE is the only encapsulation method available in Rosen MVPN.
- Rosen MPVN multicast trees are signaled using PIM only. MVPN allows the use of mLDP and RSVP P2MP LSPs.
- PE to PE protocol exchanges for Rosen MVPN is achieved using PIM only. MVPN allows for the use of BGP signaling as per unicast Layer 3 VPNs.

NG-MVPN addresses these limitations by extending the idea of the per-VRF tree by introducing the idea of provider multicast service interfaces (PMSIs). These are equivalent to the default MDTs of Rosen MVPN. NG-MVPN allows the decoupling of the mechanisms required to create a multicast VPN, such as PE auto-discovery (which PEs are members of which VPN), PMSI signaling (creation of tunnels between PEs), and customer multicast signaling (multicast signaling —IGMP/PIM— received from customer edge routers). Two types of PMSI exist:

- Inclusive (I-PMSI) — Contains all the PEs for a given MVPN, I-PMSI is the default multicast data path between all PEs of the same VPN.
- Selective (S-PMSI) — Contains only a subset of PEs of a given MVPN, used to optimize multicast stream distribution to only the PEs with active receivers for those streams.

The NG-MVPN Configuration with PIM chapter contains the VPN configuration required for the provider multicast domain using PIM Any Source Multicast (ASM) with auto-discovery based on PIM or BGP auto-discovery (AD), PIM used for the customer multicast signaling and PIM Source Specific Multicast (SSM) used for the S-PMSI creation. The customer domain configuration covers the following cases:

- PIM ASM with the Rendezvous Point (RP) in the provider PE
- PIM ASM using anycast RP on the provider RPs
- PIM SSM

This chapter introduces some of the features that were not supported at the time of writing of chapter NG-MVPN Configuration with PIM (Release 7.0). It provides configuration details to implement:

- Multicast LDP (mLDP) and RSVP-TE Point to Multipoint (P2MP) for building customer trees (C-trees) which are using MPLS instead of PIM techniques.
- MVPN source redundancy.
- MDT AFI/SAFI (to fully interoperate with Cisco networks).

PIM SSM is the only case addressed in this example, other PIM customer domain configurations are out of the scope, for more information refer to NG-MVPN Configuration with PIM.

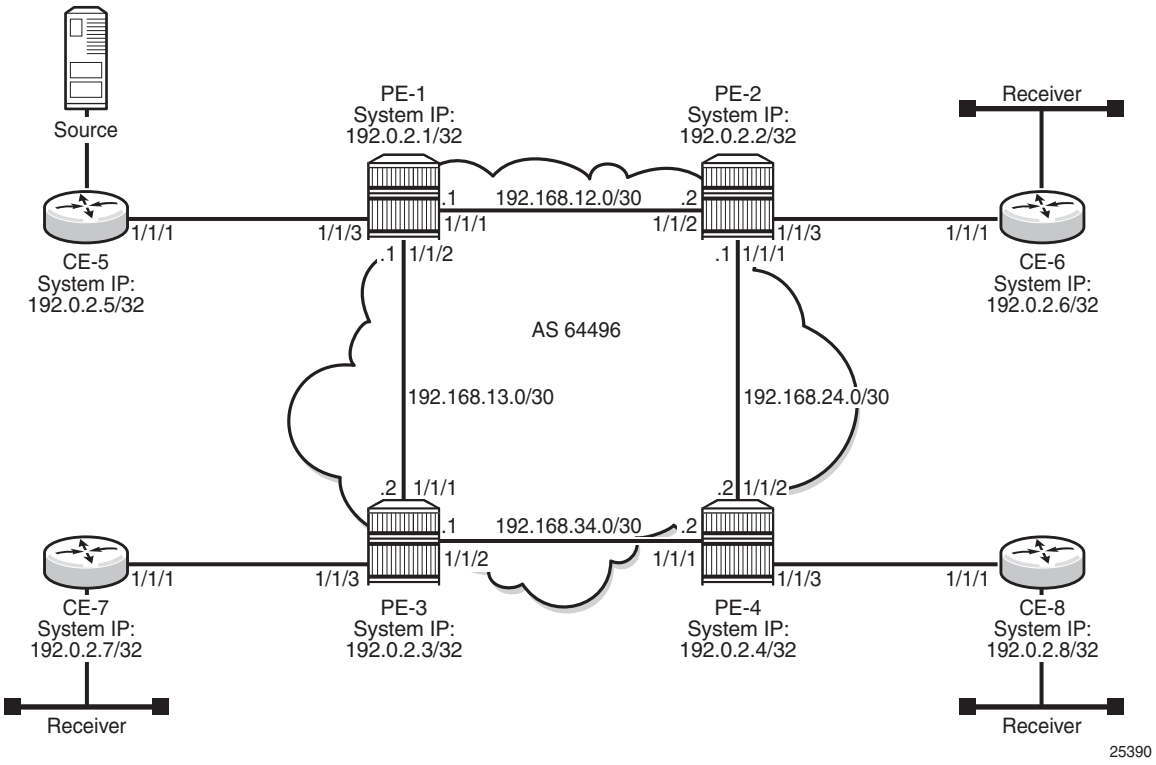
Overview

The network topology is shown in Figure 328. The setup consists of four 7750 SRs acting as provider edge (PE) routers within a single autonomous system (AS).

- Full mesh IS-IS in the AS (OSPF could be used instead)
- LDP on all interfaces in each AS (RSVP could be used instead)
- MP-iBGP sessions between the PE routers in the AS (route reflectors (RRs) could also be used).
- Layer 3 VPN on all PEs with identical route targets.

Connected to each PE is a single 7750 SR acting as a customer edge (CE) router. CE-5 has a multicast source connected, and PE-2, PE-3, and PE-4 each have a single receiver connected which will receive the multicast streams from the source. In this setup, each receiver is IGMPv3 capable. If the receiver is IGMPv3 capable, it will issue IGMPv3 reports that may include a list of required source addresses.

Figure 328 **Network Topology**



When the receiver wishes to become a member of any group, the source address of the group must be known to the CE. As a result, the source address must be IP reachable by each CE, so it is advertised using BGP by CE-5 to the PEs with attachment circuits in the VPRN. Static routes are then configured on the receiver CEs to achieve IP reachability to the source address of the multicast group.

Multicast traffic from the source is streamed toward router CE-5. Receivers connected to PE-2, PE-3 and PE-4 are interested in joining this multicast group.

CEs 5 to 8 are PIM enabled routers, which form a PIM adjacency with their nearest PE. Between the PEs across the provider network, there are no PIM adjacencies, because BGP auto-discovery and BGP signaling are used. Selective PMSI using mLDP or RSVP P2MP are out of the scope of this chapter. Selective PMSI using PIM SSM is supported too. I-PMSI and S-PMSI must use the same tunneling technology, either PIM/GRE, or mLDP, or RSVP P2MP.

Configuration

The configuration is divided into the following sections:

- Provider common configuration
  - PE global configuration
- PE VPRN configuration and PE VPRN multicast configuration for NG-MVPN
  - PMSI using mLDP
  - PMSI using RSVP-TE
  - UMH (upstream multicast hop)
- PE VPRN configuration and PE VPRN multicast configuration for Rosen MVPN using MDT AFI SAFI
  - Auto discovery using BGP MDT AFI SAFI as per Rosen MVPN version 9 with MDT using PIM SSM

Provider Common Configuration

PE Global Configuration

This section describes the common configuration required for each PE within the provider multicast domain, regardless of the MVPN PE auto-discovery or customer signaling methods. This includes interior gateway protocol (IGP) and VPRN service configuration.

The configuration tasks can be summarized as follows:

- PE global configuration.

  This includes configuration of the IGP (IS-IS will be used); configuration of link layer LDP between PEs (LDP will be used here as the method to interconnect VPRNs); configuration of iBGP between PEs to facilitate VPRN route learning.
- VPRN configuration on the PEs.

  This includes configuration of basic VPRN parameters (route-distinguisher, route target communities), configuration of attachment circuits toward CEs, configuration of VRF routing protocol and any routing policies.
- PIM within the VRF and MVPN parameters — I-PMSI
- CE configuration.

Step 1. Configure the interfaces, the IGP (IS-IS) in all PE nodes (where IS-IS redistributes route reachability to all routers) and LDP in the interfaces (link layer LDP). To facilitate the IS-IS configuration, all routers are Level2-Level1 capable within the same ISIS area-id, so there is only a single topology area in the network (all routers share the same topology). The configuration for PE-2 is displayed below.

```
*A:PE-2#  configure router
        interface "int-PE-2-PE-1"
            address 192.168.12.2/30
            port 1/1/2
        exit
        interface "int-PE-2-PE-4"
            address 192.168.24.1/30
            port 1/1/1
        exit
        interface "system"
            address 192.0.2.2/32
        exit
        autonomous-system 64496
        isis
            area-id 49.0001
            traffic-engineering
            interface "system"
                passive
            exit
            interface "int-PE-2-PE-1"
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-4"
                interface-type point-to-point
            exit
        exit
        ldp
            interface-parameters
                interface "int-PE-2-PE-1"
                exit
                interface "int-PE-2-PE-4"
                exit
            exit
        exit
```

The configuration for the rest of nodes is similar. The IP addresses can be derived from Figure 328.

Step 2. Verify that IS-IS adjacencies and LDP peer sessions are formed.

```
*A:PE-1# show router isis adjacency
===============================================================================
Router Base ISIS Instance 0 Adjacency
===============================================================================
System ID               Usage State Hold Interface                    MT-ID
-------------------------------------------------------------------------------
PE-2                    L1L2  Up    20   int-PE-1-PE-2                0
PE-3                    L1L2  Up    21   int-PE-1-PE-3                0
```

```
-------------------------------------------------------------------------------
Adjacencies : 2
===============================================================================
*A:PE-1#

*A:PE-1# show router ldp session ipv4
===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   17        18        0d 00:00:33
192.0.2.3:0         Link      Established   15        16        0d 00:00:24
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
*A:PE-1#
```

Step 3. Configure iBGP full mesh between the PEs for VPRN routing (Route Reflectors could also be an option).

```
*A:PE-1# configure
    router
        bgp
            min-route-advertisement 1
            rapid-withdrawal
            rapid-update mvpn-ipv4 mdt-safi
            group "INTERNAL"
                family vpn-ipv4 mvpn-ipv4 mdt-safi
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
            no shutdown
        exit
```

The families configured under the group “INTERNAL” are vpn-ipv4, mvpn-ipv4, and mdt-safi, since the three families are referenced in this chapter.

The mdt-safi parameter is not needed for NG-MVPN (mLDP/RSVP scenarios) and is only required for Rosen MVPN with MDT AFI SAFI.

Rapid withdrawal (configured on all PEs) disables the minimum route advertisement interval (MRAI) interval on sending BGP withdrawals. Rapid update (configured for MVPN-IPv4 and MDT AFI/SAFI address families) disables the MRAI interval on sending BGP update messages for the address family MVPN-IPv4 and MDT-SAFI).

Step 4. Verify that BGP peer relationships are established.

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
```

```
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 15          Total Path Memory           : 3000
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0
---snip---
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.2
                64496       3    0 00h00m23s 0/0/0 (VpnIPv4)
                            3    0           0/0/0 (MvpnIPv4)
                                             0/0/0 (MdtSafi)
192.0.2.3
                64496       3    0 00h00m17s 0/0/0 (VpnIPv4)
                            3    0           0/0/0 (MvpnIPv4)
                                             0/0/0 (MdtSafi)
192.0.2.4
                64496       3    0 00h00m09s 0/0/0 (VpnIPv4)
                            3    0           0/0/0 (MvpnIPv4)
                                             0/0/0 (MdtSafi)
-------------------------------------------------------------------------------
*A:PE-1#
```

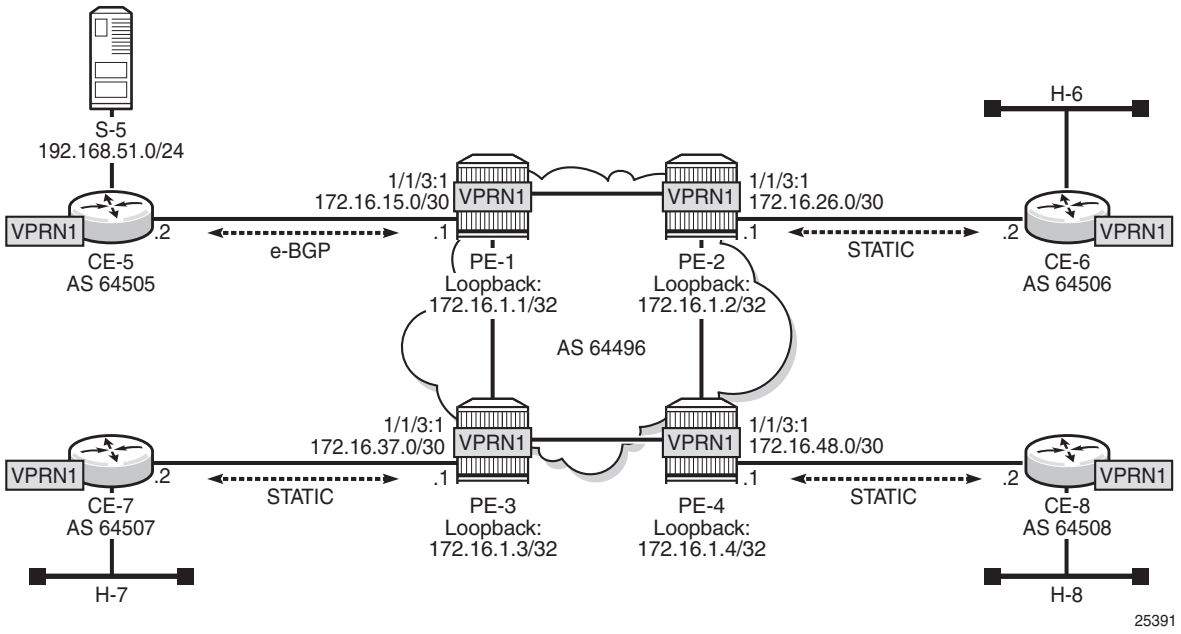
PE VPRN Configuration and PE VPRN Multicast Configuration

A VPRN is created on each PE per service (the different services using mLDP, RSVP-TE, and AFI/SAFI with PIM); these are the multicast VPRNs. PE-1 is the PE containing the attachment circuit toward CE-5. CE-5 is the CE nearest to the source. PE-2, PE-3, and PE-4 contain attachment circuits toward CE-6, CE-7, and CE-8 respectively. Each CE has a receiving host attached.

PMSI using mLDP

Figure 329 shows the details of the topology for VPRN 1.

Figure 329 **VPRN 1 Topology used for mLDP**



Unicast

Step 1. Create VPRN 1 on each PE, containing a route-distinguisher of 64496:10X (where X= number of PE) and vrf-target of 64496:100. The autonomous system number is 64496. For the next hop tunnel route resolution to connect the VPRNs between the PEs, manually configured spoke SDPs are created (other methods such as auto-bind-tunnel resolution-filter LDP resolution filter could also be used). LDP was already enabled.

```
*A:PE-1# configure
    service
        sdp 12 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 13 mpls create
            far-end 192.0.2.3
            ldp
            no shutdown
        exit
        sdp 14 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        vprn 1 customer 1 create
            description "mLDP"
```

```
            autonomous-system 64496
            route-distinguisher 64496:101
            vrf-target target:64496:100
            spoke-sdp 12 create
            exit
            spoke-sdp 13 create
            exit
            spoke-sdp 14 create
            exit
```

Step 2. Create an attachment circuit interface toward the CE and a loopback (the loopback is not mandatory, but it is configured to aid troubleshooting the routers).

```
*A:PE-1# configure
    service
        vprn 1
            interface "loopback" create
                address 172.16.1.1/32
                loopback
            exit
            interface "int-PE-1-CE-5" create
                address 172.16.15.1/30
                sap 1/1/3:1 create
                exit
            exit
```

Step 3. The source address of the multicast stream will need to be reachable by all routers (PEs and CEs) within the VPN. This will be advertised within BGP from CE-5 to PE-1. Create a BGP peering relationship with the CE as follows:

```
*A:PE-1# configure
    service
        vprn 1
            bgp
                group "EXTERNAL"
                    type external
                    peer-as 64505
                    neighbor 172.16.15.2
                    exit
                exit
                no shutdown
            exit
```

Step 4. On CE-5, create a VPRN to support the connection of the source to CE-5 and the connection from CE-5 to PE-1. Two attachment circuits are required as well as a BGP peering relationship with the PE. This uses a default BGP address family of ipv4.

```
*A:CE-5# configure
    service
        vprn 1 customer 1 create
            autonomous-system 64505
            route-distinguisher 64505:1
            interface "int-CE-5-PE-1" create
```

```
                address 172.16.15.2/30
                sap 1/1/1:1 create
                exit
            exit
            interface "int-CE-5-S-5" create
                address 192.168.51.1/24
                sap 1/1/3 create
                exit
            exit
            bgp
                group "EXTERNAL"
                    type external
                    peer-as 64496
                    neighbor 172.16.15.1
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
```

Step 5. In order for the subnet on the CE connecting to the source to be advertised within BGP, a route policy is required. The subnet containing the multicast source is 192.168.51.0/24, so a prefix-list can be used to define a match, and then used within a route policy to inject into BGP.

```
*A:CE-5# configure
    router
        policy-options
            begin
            prefix-list "SOURCE-PREFIX"
                prefix 192.168.51.0/24 exact
            exit
            policy-statement "EXPORT-SOURCE-PREFIX-TO-BGP"
                entry 10
                    from
                        prefix-list "SOURCE-PREFIX"
                    exit
                    to
                        protocol bgp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit


*A:CE-5# configure
    service
        vprn 1
            bgp
                export "EXPORT-SOURCE-PREFIX-TO-BGP"
            exit
        exit
```

Step 6. Check that the route is seen in PE-1:

```
*A:PE-1# show router 1 route-table 192.168.51.0/24

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.168.51.0/24                                 Remote  BGP       00h23m04s  170
       172.16.15.2                                                     0
-------------------------------------------------------------------------------
No. of Routes: 9
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

This prefix will also be automatically advertised within the BGP VPRN to all other PEs, and will be installed in VRF 1.

For example, on PE-4, the source subnet 192.168.51.0/24 is received via BGP VPN with a next-hop of PE-1 (192.0.2.1):

```
*A:PE-4# show router 1 route-table 192.168.51.0/24

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.168.51.0/24                                 Remote  BGP VPN   00h24m20s  170
       192.0.2.1 (tunneled)                                            0
-------------------------------------------------------------------------------
No. of Routes: 9
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-4#
```

Each CE containing a multicast receiver must be able to reach the source. As an example on CE-6, a static route is configured with next hop 172.16.26.1 of interface int-PE-2-CE-6.

```
*A:CE-6# configure
    service
        vprn 1 customer 1 create
            autonomous-system 64506
            route-distinguisher 64506:1
            interface int-CE-6-H-6 create
                address 192.168.61.1/24
                sap 1/1/2:1 create
                exit
            exit
            interface int-CE-6-PE-2 create
```

```
                address 172.16.26.2/30
                sap 1/1/1:1 create
                exit
            exit
            static-route-entry 192.168.51.0/24
                next-hop 172.16.26.1
                    no shutdown
                exit
            exit
            no shutdown
---snip---
```

After **Steps 1** to **6**, all required unicast routing is provisioned. The following sections show the configuration of the multicast in the VPRN.

Auto-Discovery and mLDP PMSI Establishment

The MP-BGP based auto-discovery is implemented with a dedicated address family defined in RFC 4760 MP_REACH_NLRI/MP_UNREACH_NLRI attributes, with AFI 1 (IPv4) or 2 (IPv6) SAFI 5 (temporary value assigned by IANA). This is the mechanism by which each PE advertises the presence of an MVPN to other PEs. This can be achieved using PIM (like in Rosen MVPN) or using BGP. With the default parameter, BGP is automatically chosen because the PMSIs are mLDP and PIM is not an option in this case. Any PE that is a member of an MVPN will advertise to the other PEs using a BGP multi-protocol network layer reachability information (NLRI) update that is sent to all PEs within the AS. This update will contain an Intra-AS I-PMSI auto-discovery route type, also known as an Intra-AD. These use an address family mvpn-ipv4, so each PE must be configured to originate and accept such updates (this was done earlier when configuring the families).

At this step (auto-discovery), the information about the PMSI is exchanged, but the PMSI is not instantiated.

As each PE contains a CE which will be part of the multicast VRF, it is necessary to enable PIM on each interface containing the attachment circuit toward a CE, and to configure the I-PMSI multicast tunnel for the VRF. In order for the BGP routes to be accepted into the VRF, a route-target community is required (vrf-target). This is configured in the **configure service vprn 1 mvpn** context and, in this case is set to the same value as the unicast vrf-target (the vrf-target community as the **configure service vprn 1 vrf-target** context).

On each PE, the PIM and MVPN context within the VPRN instance are configured as follows:

```
*A:PE-4# configure
    service
        vprn 1
            pim
```

```
            interface "loopback"
            exit
            interface "int-PE-4-CE-8"
            exit
        exit
        mvpn
            auto-discovery default
            c-mcast-signaling bgp
            provider-tunnel
                inclusive
                    mldp
                        no shutdown
                    exit
                exit
            exit
            vrf-target unicast
            exit
        exit
```

When PIM SSM is used, the configuration always shows RP static with no RP entries (this is enabled by default when PIM is provisioned). In order for the BGP routes to be accepted into the VRF, a route-target community is required (vrf-target). Although it is not mandatory for the mvpn target to be equal to the unicast target, Nokia recommends to use **vrf-target unicast** to avoid configuration mistakes and extra complexity.

The status of VPRN 1 on PE-1 is shown with the following output:

```
*A:PE-1# show router 1 mvpn

===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                   auto-discovery     : Default
UMH Selection      : Highest-Ip            SA withdrawn       : Disabled
intersite-shared   : Enabled               Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.1:2

ipmsi              : ldp
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-if-73728
Mdt-type           : sender-receiver

BSR signalling     : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
```

The following shows a debug of an Intra-AD BGP update message received by PE-1 that was sent by PE-2. The message contains the PMSI tunnel type to be used (LDP P2MP LSP), LSP identification (root node, opaque value) and the type of BGP update (Type: Intra-AD Len: 12 RD: 64496:102 Orig: 192.0.2.2):

```
3 2016/10/10 09:28:47.22 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Intra-AD Len: 12 RD: 64496:102 Orig: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:100
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 0
        Root-Node 192.0.2.2, LSP-ID 0x2001
"
```

The setup has four PEs, so every PE should see the Intra-AD routes from its peers; the following output shows the routes received in PE-1:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                     192.0.2.2              100         0
      64496:102                    -                                  -
      192.0.2.2                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.3              100         0
      64496:103                    -                                  -
      192.0.2.3                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.4              100         0
```

```
      64496:104                 -                                   -
      192.0.2.4                 -
      No As-Path                -
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The detailed output of the Intra-AD received from PE-2 shows the Tunnel-Type LDP P2MP LSP (LSP-ID is 8193), the originator id (192.0.2.2), and the route-distinguisher (64496:102):

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Original Attributes

Route Type     : Intra-Ad
Route Dist.    : 64496:102
Originator IP  : 192.0.2.2
Nexthop        : 192.0.2.2
From           : 192.0.2.2
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : no-export target:64496:100
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h15m18s
VPRN Imported  :  1
-------------------------------------------------------------------------------
PMSI Tunnel Attribute :
Tunnel-type    : LDP P2MP LSP
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 0
Root-Node      : 192.0.2.2              LSP-ID         : 8193
-------------------------------------------------------------------------------
---snip---
```

Because of the receiver-driven nature of mLDP, mLDP P2MP LSPs are set up unsolicited from the leaf PEs toward the head-end PE. The leaf PEs discover the head-end PE via I-PMSI/S-PMSI AD routes. The tunnel identifier carried in the PMSI attribute is used as the P2MP forwarding equivalence class (FEC) Element. The tunnel identifier consists of the address of the head-end PE, along with a generic LSP identifier value. The generic LSP identifier value is automatically generated by the head-end PE. The preceding show command displays the PMSI information with the detail of the root node (192.0.2.2) and the LSP-ID (8193). The PMSI was created after receiving the AD message from PE-2, where the following excerpt from the previous debug shows the same information (0x2001 in HEX is equal to 8193 in decimal).

```
Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 0
        Root-Node 192.0.2.2, LSP-ID 0x2001
```

Once the mLDP P2MP LSPs are created, the I-PMSI is instantiated in the core:

```
*A:PE-1# show router 1 pim neighbor

===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-PE-1-CE-5           1               0d 00:00:43   0d 00:01:34     105
   172.16.15.2
mpls-if-73729           1               0d 01:01:54   never           65535
   192.0.2.2
mpls-if-73730           1               0d 01:01:47   never           65535
   192.0.2.3
mpls-if-73731           1               0d 01:01:42   never           65535
   192.0.2.4
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
*A:PE-1#

*A:PE-1# show router 1 pim tunnel-interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                       Originator Address   Adm  Opr  Transport Type
-------------------------------------------------------------------------------
mpls-if-73728                   192.0.2.1            Up   Up   Tx-IPMSI
mpls-if-73729                   192.0.2.2            Up   Up   Rx-IPMSI
mpls-if-73730                   192.0.2.3            Up   Up   Rx-IPMSI
mpls-if-73731                   192.0.2.4            Up   Up   Rx-IPMSI
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
```

```
*A:PE-1#
```

Every PE has created an I-PMSI to the other PEs. Checking the mLDP P2MP LSPs that are originated, transit, or destination to PE-1:

```
 *A:PE-1# show router ldp bindings active p2mp ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73728
192.0.2.1                                   Push              --      262138
192.168.12.2                                1/1/1

8193                                        73728
192.0.2.1                                   Push              --      262138
192.168.13.2                                1/1/2

8193                                        73729
192.0.2.2                                   Pop             262138      --
  --                                          --

8193                                        73729
192.0.2.2                                   Swap            262138    262137
192.168.13.2                                1/1/2

8193                                        73730
192.0.2.3                                   Pop             262137      --
  --                                          --

8193                                        73730
192.0.2.3                                   Swap            262137    262137
192.168.12.2                                1/1/1

8193                                        73731
192.0.2.4                                   Pop             262136      --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 7
===============================================================================
---snip---
```

The two first entries in the output show the P2MP LSP where PE-1 is the root head-end (Push). The other two entries (Swap and Pop) correspond with transit and leaf for the P2MP LSPs originated by the other PEs. The command shows a P2MP-ID (8193) with an interface 73728 (matches with the **show router 1 pim tunnel interface** being the PIM interface created from PE-1) with two egress interfaces pointing to PE-2 and PE-3.

A similar command executed on PE-2 shows:

```
*A:PE-2# show router ldp bindings active p2mp ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73729
192.0.2.1                                   Pop             262138      --
  --                                          --

8193                                        73729
192.0.2.1                                   Swap            262138    262136
192.168.24.2                                1/1/1

---snip---

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 7
===============================================================================
```

On PE-2, the first entry shows that PE-2 is a leaf of the P2MP LSP tree created by PE-1 (ingress label is 262138 which was the egress label to reach PE-2 and is popped). However, the second entry shows that PE-2 is transit for the P2MP LSP going to PE-4 (ingress label 262138, egress label 262136 next hop PE-4).

The same command on PE-4 shows:

```
*A:PE-4# show router ldp bindings active p2mp ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
```

```
===============================================================================
P2MP-Id                                       Interface
RootAddr                                      Op              IngLbl    EgrLbl
EgrNH                                         EgrIf/LspId
-------------------------------------------------------------------------------
8193                                          73731
192.0.2.1                                     Pop             262136      --
  --                                            --

---snip---
-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 5
===============================================================================
```

In the first entry, the root is PE-1 and the action is Pop, being the ingress label 262136, showing that this is another leaf for the P2MP LSP started on PE-1.

To complete the information, checking on PE-3, the first entry there is a Pop where the root is PE-1, and the ingress label is 262138:

```
*A:PE-3# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                       Interface
RootAddr                                      Op              IngLbl    EgrLbl
EgrNH                                         EgrIf/LspId
-------------------------------------------------------------------------------
8193                                          73729
192.0.2.1                                     Pop             262138      --
  --                                            --

---snip---

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 5
===============================================================================
```

As a summary, each root PE has a P2MP LSP with three leaves (the other PEs) and they are also transit points to the P2MP LSPs created in the other PEs. As an additional check, an OAM ping can show the different leaves that a P2MP LSP has:

```
*A:PE-1# oam p2mp-lsp-ping ldp 8193 sender-addr 192.0.2.1 detail
P2MP identifier 8193: 88 bytes MPLS payload
===============================================================================
Leaf Information
===============================================================================
From             RTT                  Return Code
-------------------------------------------------------------------------------
192.0.2.2        =0.563ms             EgressRtr(3)
```

```
192.0.2.3        =0.600ms             EgressRtr(3)
192.0.2.4        =1.08ms              EgressRtr(3)
------------------------------------------------------------------------------
Total Leafs responded = 3
          round-trip min/avg/max  = 0.643 / 0.835 / 1.18 ms
Responses based on return code:
        EgressRtr(3)=3
*A:PE-1#
```

An easy way to see the path that the LDP P2MP LSP follows for a specific leaf is the following command (such as LDP trace from PE-1 to PE-4):

```
 *A:PE-1# oam ldp-treetrace prefix 192.0.2.4/32

ldp-treetrace for Prefix 192.0.2.4/32:

      192.168.24.2, ttl =   2 dst =      127.1.0.255 rc = EgressRtr status = Done
    Hops:      192.168.12.2

ldp-treetrace discovery state: Done
ldp-treetrace discovery status: ' OK '
Total number of discovered paths: 1
Total number of failed traces: 0

*A:PE-1#
```

The command shows that on PE-4, there is an active leaf of the P2MP LSP, and that there is an intermediate hop on PE-2.

Traffic Flow

The receiver H-8, connected to CE-8, wishes to join the group 232.1.1.1 with source 192.168.51.1 and sends an IGMPv3 report toward CE-8. CE-8 recognizes the report and sends a PIM join toward the source, therefore, it reaches PE-1 where the source is connected to through CE-5. The following output shows the debug seen on PE-4, where the PIM join is received from CE-8 and a BGP update Source Join is sent to all PEs (only the update sent to PE-1 is shown).

```
17 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPProcessSG
pimJPProcessSG: (S,G)-> (192.168.51.1,232.1.1.1) type <S,G>, i/f int-PE-4-CE-8,
 upNbr 172.16.48.1 isJoin 1 isRpt 0 holdTime 210"

18 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmFindRpfNexthop
Track (192.168.51.1,232.1.1.1) type <S,G> using 192.168.51.1"

19 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmAddSrcEntry
Added src entry for src 192.168.51.1"

20 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
```

```
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Downstream: State NoInfo Event RxJoin StandbyEvent F, (S,G) (192.168.51.1,232
.1.1.1) groupType <S,G>"

21 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Upstream: State NotJoined Event JoinDesiredTrue StandbyEvent F, (S,G) (192.16
8.51.1,232.1.1.1) groupType <S,G>"

22 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSGUpJoinDesiredTrue
No upstream interface. pSG (192.168.51.1,232.1.1.1) rpfType 3"

23 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSGUpJoinDesiredTrue
No upstream interface SG (192.168.51.1,232.1.1.1) rpfType 3"

24 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmProcessNhresEvent
RTM-Nhres Event U-RTM NEW Src 192.168.51.1 SrcRtmUse UCAST"

25 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmProcessNhresEvent
Prefix 192.168.51.0/24 numNextHops 1 owner BGP_VPN metric 20 pref 170"

26 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmSrcResolveSGsInt
Trying to resolve SG (192.168.51.1,232.1.1.1)"

27 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmNotifyRpfChange
RPF Change to Source/RP 192.168.51.1  for SG (192.168.51.1,232.1.1.1) dynMLDP F via
NH 192.0.2.1 IfIdx: 73731 RpfType: REMOTE Reason:  RTE_ADD old NH 0.0.0.0 IfIdx: 0 R
pfType: NONE mplsRpf F NextHops 1 reg 1/1 lfa 0/0"

28 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmNotifyRpfChange
SG (192.168.51.1,232.1.1.1) Source/RP 192.168.51.1 Ipmsi 73728 NhIf 0
new NhIf 73731"

29 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Upstream: State Joined Event MribChange StandbyEvent F,
(S,G) (192.168.51.1,232.1.1.1) groupType <S,G>

30 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSGUpStateJMribChange
SG (192.168.51.1,232.1.1.1), type <S,G> oldMribNhopIp 0.0.0.0 oldRpfNbrIp 0.0.0.0,
oldRpfType NONE oldRpfIf 0 rptMribNhopIp 0.0.0.0, rptRpfNbrIp 0.0.0.0 rtmReason 48
isSGExtNet : no"

31 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSGUpStateJMribChange
SG (192.168.51.1,232.1.1.1), type <S,G> newMribNhopIp 192.0.2.1
newRpfNbrIp 192.0.2.1 newRpfType REMOTE newRpfIf 73731"

32 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmUpdateSGMetric
SG metric 4294967295 pref 2147483647, new metric 20 pref 170"
```

```
35 2016/10/10 10:48:26.67 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.4
        Type: Source-
Join Len:22 RD: 64496:101 SrcAS: 64496 Src: 192.168.51.1 Grp: 232.1.1.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The following debug shows that PE-1 receives the BGP update Source Join with source 192.168.1.1 and group 232.1.1.1 and sends a PIM join toward CE-5:

```
19 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.4
        Type: Source-
Join Len:22 RD: 64496:101 SrcAS: 64496 Src: 192.168.51.1 Grp: 232.1.1.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"


20 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimProcessMvpnRouteMsg
originator 0.0.0.0: add rtType 7 nextHop 192.0.2.4 source 192.168.51.1
group 232.1.1.1"

21 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPProcessSG
pimJPProcessSG: (S,G)-> (192.168.51.1,232.1.1.1) type <S,G>, i/f mpls-if-
73728, upNbr 192.0.2.1 isJoin 1 isRpt 0 holdTime 65535"

22 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmFindRpfNexthop
Track (192.168.51.1,232.1.1.1) type <S,G> using 192.168.51.1"
```

```
23 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmAddSrcEntry
Added src entry for src 192.168.51.1"

24 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Downstream: State NoInfo Event RxJoin StandbyEvent F,
(S,G) (192.168.51.1,232.1.1.1) groupType <S,G>"

25 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Upstream: State NotJoined Event JoinDesiredTrue StandbyEvent F,
(S,G) (192.168.51.1,232.1.1.1) groupType <S,G>"

26 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimAddToJPTxPdu
pimAddToJPTxPdu: (S,G)-> (192.168.51.1,232.1.1.1), type <S,G>,
txPendFlag J isStandby F"

27 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmProcessNhresEvent
RTM-Nhres Event U-RTM NEW Src 192.168.51.1 SrcRtmUse UCAST"

28 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmProcessNhresEvent
Prefix 192.168.51.0/24 numNextHops 1 owner BGP metric 0 pref 170"

37 2016/10/10 10:51:30.14 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSGEncodeGroupSet
Encoding  Join for source 192.168.51.1"

38 2016/10/10 10:51:30.14 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSGEncodeGroupSet
num joined srcs 1, num pruned srcs 0"

39 2016/10/10 10:51:30.14 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimSendJoinPrunePdu
sending JP PDU with 1 groups, if 5 adj 172.16.15.2"

41 2016/10/10 10:51:30.15 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Upstream: State Joined Event JTTimerExp StandbyEvent F, (S,G) (192.168.51.1,2
32.1.1.1) groupType <S,G>"
```

The BGP update source join received by PE-1 is displayed with the command:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type source-join
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
```

```
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
u*>i  Source-Join                 -                      100         0
      64496:101                   64496                              -
      192.0.2.4                   192.168.51.1
      No As-Path                  232.1.1.1
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

To verify the traffic: on PE-1 there is a group 232.1.1.1 with source 192.168.51.1, the Reverse Path Forwarding (RPF) is CE-5, the multicast traffic is flowing from CE-5 to PE-1 using int-PE-1-CE-5 and the outgoing interface is using the PMSI mLDP mpls-if-73728.

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.51.1
RP Address         : 0
Advt Router        : 172.16.15.2
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.15.2
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:05:37        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:22
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.15.2
Incoming Intf      : int-PE-1-CE-5
Outgoing Intf List : mpls-if-73728

Curr Fwding Rate   : 6747.1 kbps
Forwarded Packets  : 6270331            Discarded Packets  : 0
Forwarded Octets   : 288435226          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

On PE-4, the same (S,G) arrives in the incoming interface mpls-if-73731 and the outgoing interface is int-PE-4-CE-8.

```
*A:PE-4# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.51.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 01:22:42        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:41
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73731
Outgoing Intf List : int-PE-4-CE-8

Curr Fwding Rate   : 7142.1 kbps
Forwarded Packets  : 2527624            Discarded Packets  : 0
Forwarded Octets   : 116270704          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-4#
```

When the receiver is not interested in the channel group any more, the receiver H-8 sends an IGMPv3 leave, PE-4 sends a PIM prune translated to a BGP MP_UNREACH NLRI to all PEs. As mentioned before, rapid withdrawals are sent without waiting for the MRAI (for simplicity, only one BGP update is shown in the output debug).

```
45 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPProcessSG
pimJPProcessSG: (S,G)-> (192.168.51.1,232.1.1.1) type <S,G>, i/f int-PE-4-CE-8,
upNbr 172.16.48.1 isJoin 0 isRpt 0 holdTime 210"

46 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Downstream: State Joined Event RxPrune StandbyEvent F,
(S,G) (192.168.51.1,232.1.1.1) groupType <S,G>"

47 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Downstream: State PrunePending Event PrunePendTimerExp StandbyEvent F,
(S,G) (192.168.51.1,232.1.1.1) groupType <S,G>"

48 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
```

```
"PIM[vprn 2 vprn1]: pimJPPrintFsmEvent
PIM JP Upstream: State Joined Event JoinDesiredFalse StandbyEvent F,
(S,G) (192.168.51.1,232.1.1.1) groupType <S,G>"

49 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimAddToJPTxPdu
pimAddToJPTxPdu: (S,G)-> (192.168.51.1,232.1.1.1), type <S,G>, txPendFlag P
isStandby F"

50 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmStopRpfNexthop
Stop tracking (192.168.51.1,232.1.1.1) type <S,G> with 192.168.51.1
pRtmNhop 0xa2f2c590"

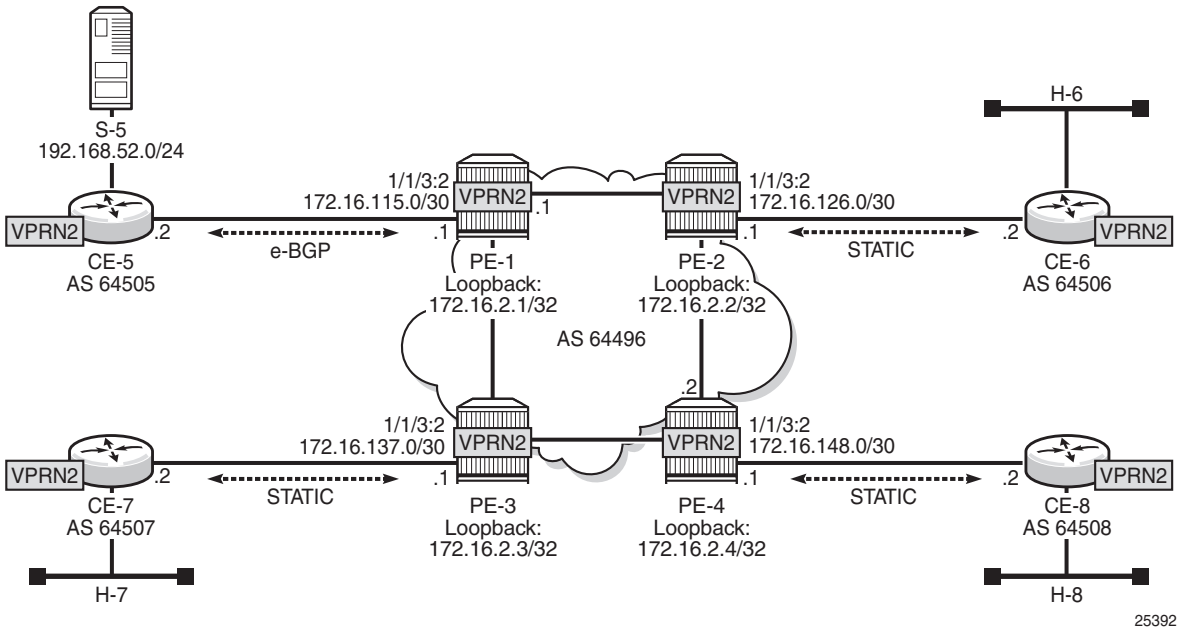
51 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 vprn1 PIM[vprn 2 vprn1]
"PIM[vprn 2 vprn1]: pimRtmDelSrcEntry
Deleted src entry for src 192.168.51.1"

52 2016/10/10 11:17:41.65 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 31
    Flag: 0x90 Type: 15 Len: 27 Multiprotocol Unreachable NLRI:
        Address Family MVPN_IPV4
        Type: Source-Join Len:22 RD: 64496:101 SrcAS: 64496
                         Src: 192.168.51.1 Grp: 232.1.1.1
"
```

PMSI using RSVP-TE

Figure 330 shows the details of the topology for VPRN 2.

Figure 330 **VPRN 2 Topology used for RSVP-TE P2MP**



Unicast

For the sake of simplicity, check **Steps 1** to **6** in PMSI using mLDP for VPRN 2 creation information. The same steps are repeated for RSVP, check Figure 330 for details. The result is the configuration in all the PEs, taking as an example PE-1:

```
*A:PE-1# configure
    service
        vprn 2 customer 1 create
            description "P2MP RSVP"
            autonomous-system 64496
            route-distinguisher 64496:201
            vrf-target target:64496:200
            interface "loopback" create
                address 172.16.2.1/30
                loopback
            exit
            interface "int-PE-1-CE-5" create
                address 172.16.115.1/30
                sap 1/1/3:2 create
                exit
            exit
            bgp
                group "EXTERNAL"
                    type external
                    peer-as 64505
                    neighbor 172.16.115.2
```

```
                exit
            exit
            no shutdown
        exit
        spoke-sdp 12 create
        exit
        spoke-sdp 13 create
        exit
        spoke-sdp 14 create
        exit
        no shutdown
    exit
```

Because RSVP is the signaling protocol to establish the P2MP LSPs, RSVP is configured on the interfaces. In addition, to use P2MP RSVP, an LSP template is needed. The template defines the characteristics of the LSP to be created, for example, make-before-break, bandwidth, administrative groups, CSPF, specific paths, etc. A basic template is used here. TE parameters specified in the template are commonly used in each RSVP PATH message for each of the branches of the P2MP RSVP LSP. The template is used in the mvpn context within the VPRN configuration (see Auto-Discovery and RSVP PMSI Establishment). The resignal timer for P2MP is configured to the minimum value of sixty minutes (60 — 10080 minutes):

```
*A:PE-1# configure
    router
        mpls
            p2mp-resignal-timer 60
            interface system
            exit
            interface int-PE-1-PE-2
            exit
            interface int-PE-1-PE-3
            exit
            path EMPTY
                no shutdown
            exit
            lsp-template VRF2 p2mp
                default-path EMPTY
                cspf
                fast-reroute facility
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit

*A:PE-1# configure router rsvp no shutdown
```

Auto-Discovery and RSVP PMSI Establishment

The MP-BGP based auto-discovery is implemented with a new address family defined in RFC 4760 MP_REACH_NLRI/MP_UNREACH_NLRI attributes, with AFI 1 (IPv4) or 2 (IPv6) SAFI 5 (temporary value assigned by IANA). This is the mechanism by which each PE advertises the presence of an MVPN to other PEs. This can be achieved using PIM (like in Rosen MVPN) or using BGP. With the default parameter, BGP is automatically chosen because the PMSIs are RSVP and PIM is not an option in this case. Any PE that is a member of an MVPN will advertise to the other PEs using a BGP multi-protocol network layer reachability information (NLRI) update that is sent to all PEs within the AS. This update will contain an Intra-AS I-PMSI auto-discovery route type, also known as an Intra-AD. These use an address family mvpn-ipv4, so each PE must be configured to originate and accept such updates (this was done earlier when configuring the families).

At this step (auto-discovery), the information about the PMSI is exchanged, but the PMSI is not instantiated.

As each PE contains a CE which will be part of the multicast VRF, it is necessary to enable PIM on each interface containing the attachment circuit toward a CE, and to configure the I-PMSI multicast tunnel for the VRF. In order for the BGP routes to be accepted into the VRF a route-target community is required (vrf-target). Although it is not mandatory for the MVPN vrf-target to be equal to the unicast target, Nokia recommends to use vrf-target unicast to avoid configuration mistakes and extra complexity.

On each PE, the multicast configuration in the VPRN instance is as follows:

```
*A:PE-1# configure service
        vprn 2
            pim
                interface "loopback"
                exit
                interface "int-PE-1-CE-5"
                exit
            exit
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                provider-tunnel
                    inclusive
                        rsvp
                            lsp-template "VRF2"
                            no shutdown
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
```

The status of VPRN 2 on PE-1 is shown with the following output:

```
*A:PE-1# show router 2 mvpn
===============================================================================
MVPN 2 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.1:3

ipmsi              : rsvp VRF2
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : VRF2-2-73732
enable-bfd-root    : false                enable-bfd-leaf    : false
Mdt-type           : sender-receiver

BSR signalling     : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
*A:PE-1#
```

The following shows a debug of an Intra-AD BGP update message received by PE-1 that was sent by PE-4. The message contains the PMSI tunnel-type to be used (RSVP P2MP LSP), the P2MP LSP ID (encoded as extended tunnel ID and P2MP-ID carried in the RSVP Session object), and the type of BGP update (Type: Intra-AD Len: 12 RD: 64496:204 Orig: 192.0.2.4):

```
29 2016/10/10 11:41:43.81 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.4
        Type: Intra-AD Len: 12 RD: 64496:204 Orig: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:200
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 0
```

```
        P2MP-ID 0x2, Tunnel-ID: 61442, Extended-Tunnel-ID 192.0.2.4
"
```

The setup has four PEs, so every PE should see the others peer Intra-AD route; the following output shows the routes received in PE-1:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                     192.0.2.2              100         0
      64496:202                    -                                  -
      192.0.2.2                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.3              100         0
      64496:203                    -                                  -
      192.0.2.3                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.4              100         0
      64496:204                    -                                  -
      192.0.2.4                    -
      No As-Path                   -
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1#
```

The detailed output of the Intra-AD received from PE-4 shows the tunnel-type RSVP-TE P2MP LSP (P2MP-ID is 2), the originator id (192.0.2.4), and the route-distinguisher (64496:204):

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad originator-ip 192.0.2.4
 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Original Attributes
```

```
Route Type     : Intra-Ad
Route Dist.    : 64496:204
Originator IP  : 192.0.2.4
Nexthop        : 192.0.2.4
From           : 192.0.2.4
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : no-export target:64496:200
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h37m08s
VPRN Imported  :  2
-------------------------------------------------------------------------------
PMSI Tunnel Attribute :
Tunnel-type    : RSVP-TE P2MP LSP
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 0
P2MP-ID        : 2                      Tunnel-ID      : 61442
Extended-Tunne*: 192.0.2.4
-------------------------------------------------------------------------------
```

For the I-PMSI, the head-end PE firstly discovers all the leaf PEs via I-PMSI A-D routes, it then signals the P2MP LSP to all the leaf PEs using RSVP-TE (subsequently adding or removing S2L (source to leaf) paths as PEs are added or removed from the MVPN).

As in the mLDP case, the demarcation of the domains is in the PE. The PE router participates in both the customer multicast domain and the provider's multicast domain. The customer's CEs are limited to a multicast adjacency with the multicast instance on the PE created to support that specific customer's IP-VPN. This way, customers are isolated from the provider's core multicast domain and other customer multicast domains while the provider's core P routers only participate in the provider's multicast domain and are isolated from all customers's multicast domains. C-trees to P-tunnels bindings are also discovered using BGP routes, instead of PIM join TLVs. MVPN c-multicast routing information is exchanged between PEs by using c-multicast routes that are carried using MCAST-VPN NLRIs.

Once the RSVP-TE P2MP LSPs are created, the I-PMSI is instantiated in the core:

```
*A:PE-1# show router 2 pim neighbor
```

```
===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-PE-1-CE-5           1               0d 00:01:20   0d 00:01:26     105
   172.16.115.2
mpls-if-73733           1               0d 00:01:51   never           65535
   192.0.2.2
mpls-if-73734           1               0d 00:01:43   never           65535
   192.0.2.3
mpls-if-73735           1               0d 00:01:36   never           65535
   192.0.2.4
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
*A:PE-1#

*A:PE-1# show router 2 pim tunnel-interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                       Originator Address   Adm  Opr  Transport Type
-------------------------------------------------------------------------------
mpls-if-73732                   192.0.2.1            Up   Up   Tx-IPMSI
mpls-if-73733                   192.0.2.2            Up   Up   Rx-IPMSI
mpls-if-73734                   192.0.2.3            Up   Up   Rx-IPMSI
mpls-if-73735                   192.0.2.4            Up   Up   Rx-IPMSI
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
*A:PE-1#
```

The following command displays the PMSIs created on a PE, taking PE-3 as an example:

```
*A:PE-3# show router 2 pim tunnel-interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                       Originator Address   Adm  Opr  Transport Type
-------------------------------------------------------------------------------
mpls-if-73732                   192.0.2.3            Up   Up   Tx-IPMSI
mpls-if-73733                   192.0.2.1            Up   Up   Rx-IPMSI
mpls-if-73734                   192.0.2.2            Up   Up   Rx-IPMSI
mpls-if-73735                   192.0.2.4            Up   Up   Rx-IPMSI
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
*A:PE-3#

*A:PE-3# tools dump router 2 mvpn provider-tunnels
===============================================================================
MVPN 2 Inclusive Provider Tunnels Originating
===============================================================================
ipmsi (RSVP)                            P2MP-ID     Tunl-ID     Ext-Tunl-ID
```

```
-------------------------------------------------------------------------------
VRF2-2-73732                          2           61441       192.0.2.3
===============================================================================
MVPN 2 Selective Provider Tunnels Originating
===============================================================================
spmsi (RSVP)                          P2MP-ID     Tunl-ID     Ext-Tunl-ID
-------------------------------------------------------------------------------
No Tunnels Found
-------------------------------------------------------------------------------
===============================================================================
MVPN 2 Inclusive Provider Tunnels Terminating
===============================================================================
ipmsi (RSVP)                          P2MP-ID     Tunl-ID     Ext-Tunl-ID
-------------------------------------------------------------------------------
mpls-if-73733                         2           61441       192.0.2.1
mpls-if-73734                         2           61441       192.0.2.2
mpls-if-73735                         2           61442       192.0.2.4
===============================================================================
MVPN 2 Selective Provider Tunnels Terminating
===============================================================================
spmsi (RSVP)                          P2MP-ID     Tunl-ID     Ext-Tunl-ID
-------------------------------------------------------------------------------
No Tunnels Found
-------------------------------------------------------------------------------
*A:PE-3#
```

Every PE has created an I-PMSI to the other PEs. As an example, PE-1 has established an LSP with name VRF2-2-73732 with PE-2, PE-3 and PE-4 as leaves. The S2L path is empty because the template did not have any S2L path configured for simplicity.

```
*A:PE-1# show router mpls p2mp-lsp detail
===============================================================================
MPLS P2MP LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : VRF2-2-73732
LSP Type         : P2mpAutoLsp              LSP Tunnel ID        : 61441
LSP Index        : 61441                    TTM Tunnel Id        : 61441
From             : 192.0.2.1
Adm State        : Up                       Oper State           : Up
LSP Up Time      : 0d 00:11:30              LSP Down Time        : 0d 00:00:00
Transitions      : 1                        Path Changes         : 1
Retry Limit      : 0                        Retry Timer          : 30 sec
Signaling        : RSVP                     Resv. Style          : SE
Hop Limit        : 255                      Negotiated MTU       : n/a
Adaptive         : Enabled                  ClassType            : 0
FastReroute      : Enabled                  Oper FR              : Enabled
FR Method        : Facility                 FR Hop Limit         : 16
FR Node Protect  : Disabled                 FR Prop Adm Grp      : Disabled
FR Object        : Enabled
CSPF             : Enabled                  ADSPEC               : Disabled
Metric           : Disabled                 Use TE metric        : Disabled
Load Bal Wt      : N/A                      ClassForwarding      : Disabled
Include Grps     :                          Exclude Grps         :
```

```
None                                              None
Least Fill      : Disabled

Revert Timer    : Disabled                  Next Revert In       : N/A
Auto BW         : Disabled
LdpOverRsvp     : Enabled
VprnAutoBind    : Enabled
IGP Shortcut    : Enabled                   BGP Shortcut         : Enabled
IGP LFA         : Disabled                  IGP Rel Metric       : Disabled
BGPTransTun     : Enabled
Oper Metric     : Disabled
Prop Adm Grp    : Disabled

P2MPInstance    : 2                         P2MP-Inst-type       : Primary
S2L Cfg Counter : 3                         S2L Oper Counter     : 3
S2l-Name        : EMPTY                     To                   : 192.0.2.2
S2l-Name        : EMPTY                     To                   : 192.0.2.3
S2l-Name        : EMPTY                     To                   : 192.0.2.4
===============================================================================
*A:PE-1#
```

Checking the RSVP-TE P2MP LSPs that are originated, transit, or destination to PE-1, the show command allows filtering by type, in this case showing the originated LSPs only:

```
*A:PE-1# show router mpls p2mp-info type originate

===============================================================================
MPLS P2MP LSPs (Originate)
===============================================================================
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 37376
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.2
Out Interface        : 1/1/1                 Out Label     : 262138
Num. of S2ls         : 2
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 37376
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.3
Out Interface        : 1/1/2                 Out Label     : 262136
Num. of S2ls         : 1
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 37376
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.4
Out Interface        : 1/1/1                 Out Label     : 262138
Num. of S2ls         : 2
-------------------------------------------------------------------------------
P2MP Cross-connect instances : 3
===============================================================================
*A:PE-1#
```

Following the path of the S2L from PE-1 to PE-4 (third entry S2L VRF2-2-73732), the outgoing interface is 1/1/1 that connects PE-1 to PE-2, so the LSP goes to PE-4 via PE-2. The return path need not be via PE-2; it may be via PE-3.

```
*A:PE-2# show router mpls p2mp-info type transit

===============================================================================
MPLS P2MP LSPs (Transit)
===============================================================================
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 37376
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.4
Out Interface        : 1/1/1                 Out Label     : 262131
Num. of S2ls         : 1
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.4             Tunnel ID     : 61442
P2MP ID              : 2                     Lsp ID        : 53248
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.1
Out Interface        : 1/1/2                 Out Label     : 262131
Num. of S2ls         : 1
-------------------------------------------------------------------------------
P2MP Cross-connect instances : 2
===============================================================================
*A:PE-2#
```

As transit, PE-2 shows that there is an LSP coming from PE-1 (VRF2-2-73732) and the outgoing interface is 1/1/1 that connects PE-2 with PE-4.

Using the same command with a different filter on PE-4, 3 P2MP LSPs are terminated, one from each remote PE (PE-1, PE-2 and PE-3). On PE-4, an S2L VRF2-2-73732 from 192.0.2.1 and P2MP ID = 2 is traced.

```
*A:PE-4# show router mpls p2mp-info type terminate
===============================================================================
MPLS P2MP LSPs (Terminate)
===============================================================================
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.1             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 37376
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.4
In Interface         : 1/1/2                 In Label      : 262131
Num. of S2ls         : 1
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.2             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 6656
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.4
In Interface         : 1/1/2                 In Label      : 262136
```

```
Num. of S2ls         : 2
-------------------------------------------------------------------------------
S2L VRF2-2-73732::EMPTY
-------------------------------------------------------------------------------
Source IP Address    : 192.0.2.3             Tunnel ID     : 61441
P2MP ID              : 2                     Lsp ID        : 8704
S2L Name             : VRF2-2-73732::EMPTY   To            : 192.0.2.4
In Interface         : 1/1/1                 In Label      : 262132
Num. of S2ls         : 2
-------------------------------------------------------------------------------
P2MP Cross-connect instances : 3
===============================================================================
*A:PE-4#
```

The following output shows the P2MP LSP on PE-1 with more detail:

```
*A:PE-1# show router mpls p2mp-lsp "VRF2-2-73732" p2mp-instance "2" s2l EMPTY
detail

===============================================================================
MPLS LSP VRF2-2-73732 S2L EMPTY (Detail)
===============================================================================
Legend :
    @ - Detour Available                        # - Detour In Use
    b - Bandwidth Protected                     n - Node Protected
    S - Strict                                  L - Loose
    A - ABR
    s - Soft Preemption
===============================================================================
-------------------------------------------------------------------------------
LSP VRF2-2-73732 S2L EMPTY
-------------------------------------------------------------------------------
LSP Name         : VRF2-2-73732         S2l LSP ID           : 37376
P2MP ID          : 2                    S2l Grp Id           : 1
Adm State        : Up                   Oper State           : Up
S2l State:       : Active                                    :
S2L Name         : EMPTY                To                   : 192.0.2.2
S2l Admin        : Up                   S2l Oper             : Up
OutInterface     : 1/1/1                Out Label            : 262138
S2L Up Time      : 0d 00:03:41          S2L Dn Time          : 0d 00:00:00
RetryAttempt     : 0                    NextRetryIn          : 0 sec
S2L Trans        : 1                    CSPF Queries         : 1
Failure Code     : noError              Failure Node         : n/a
Inter-area       : False
ExplicitHops     :
    No Hops Specified
Actual Hops      :
    192.168.12.1 (192.0.2.1) @                  Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)                    Record Label        : 262138
ComputedHops     :
    192.168.12.1(S)
 -> 192.168.12.2(S)
LastResignal     : n/a
-------------------------------------------------------------------------------
LSP VRF2-2-73732 S2L EMPTY
-------------------------------------------------------------------------------
LSP Name         : VRF2-2-73732         S2l LSP ID           : 37376
P2MP ID          : 2                    S2l Grp Id           : 2
```

```
Adm State        : Up                    Oper State           : Up
S2l State:       : Active                                     :
S2L Name         : EMPTY                 To                   : 192.0.2.3
S2l Admin        : Up                    S2l Oper             : Up
OutInterface     : 1/1/2                 Out Label            : 262136
S2L Up Time      : 0d 00:03:33           S2L Dn Time          : 0d 00:00:00
RetryAttempt     : 0                     NextRetryIn          : 0 sec
S2L Trans        : 1                     CSPF Queries         : 1
Failure Code     : noError               Failure Node         : n/a
Inter-area       : False
ExplicitHops     :
    No Hops Specified
Actual Hops      :
    192.168.13.1 (192.0.2.1) @                  Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3)                    Record Label        : 262136
ComputedHops     :
    192.168.13.1(S)
 -> 192.168.13.2(S)
LastResignal     : n/a
-------------------------------------------------------------------------------
LSP VRF2-2-73732 S2L EMPTY
-------------------------------------------------------------------------------
LSP Name         : VRF2-2-73732          S2l LSP ID           : 37376
P2MP ID          : 2                     S2l Grp Id           : 3
Adm State        : Up                    Oper State           : Up
S2l State:       : Active                                     :
S2L Name         : EMPTY                 To                   : 192.0.2.4
S2l Admin        : Up                    S2l Oper             : Up
OutInterface     : 1/1/1                 Out Label            : 262138
S2L Up Time      : 0d 00:03:26           S2L Dn Time          : 0d 00:00:00
RetryAttempt     : 0                     NextRetryIn          : 0 sec
S2L Trans        : 1                     CSPF Queries         : 1
Failure Code     : noError               Failure Node         : n/a
Inter-area       : False
ExplicitHops     :
    No Hops Specified
Actual Hops      :
    192.168.12.1 (192.0.2.1) @                  Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2) @                  Record Label        : 262138
 -> 192.168.24.2 (192.0.2.4)                    Record Label        : 262131
ComputedHops     :
    192.168.12.1(S)
 -> 192.168.12.2(S)
 -> 192.168.24.2(S)
LastResignal     : n/a
===============================================================================
*A:PE-1#
```

The last entry, VRF2-2-73732, provides the details of the S2L traced earlier, displaying the different hops (PE-1, PE-2, and PE-4), the fast reroute protection (link protection is supported only) and the labels used (262138 from PE-1 to PE-2, 262131 from PE-2 to PE-4). On PE-1, although only one has been shown, both links PE-1 to PE-3 and PE-1 to PE-2 are fast reroute protected.

If any of the protected links between PE-1 and PE-2 or PE-3 are broken, fast reroute will be initiated. The protected bypass hops are displayed with the following command:

```
*A:PE-1# show router mpls bypass-tunnel protected-lsp p2mp detail
===============================================================================
MPLS Bypass Tunnels (Detail)
===============================================================================
-------------------------------------------------------------------------------
bypass-link192.168.12.2-61442
-------------------------------------------------------------------------------
To             : 192.168.24.1        State               : Up
Out I/F        : 1/1/2               Out Label           : 262138
Up Time        : 0d 00:03:50         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 2
Type           : P2mp                Bypass Path Cost    : 30
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
    192.168.13.1(S)                  Egress Admin Groups : None
 -> 192.168.13.2(S)                  Egress Admin Groups : None
 -> 192.168.34.2(S)                  Egress Admin Groups : None
 -> 192.168.24.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.13.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.13.2 (192.0.2.3)         Record Label        : 262138
 -> 192.168.34.2 (192.0.2.4)         Record Label        : 262138
 -> 192.168.24.1 (192.0.2.2)         Record Label        : 262137

Protected LSPs -
LSP Name       : VRF2-2-73732::EMPTY
From           : 192.0.2.1           To                  : 192.0.2.2
Avoid Node/Hop : 192.168.12.2        Downstream Label    : 262138
Bandwidth      : 0 Kbps

LSP Name       : VRF2-2-73732::EMPTY
From           : 192.0.2.3           To                  : 192.0.2.2
Avoid Node/Hop : 192.168.12.2        Downstream Label    : 262134
Bandwidth      : 0 Kbps

LSP Name       : VRF2-2-73732::EMPTY
From           : 192.0.2.1           To                  : 192.0.2.4
Avoid Node/Hop : 192.168.12.2        Downstream Label    : 262138
Bandwidth      : 0 Kbps

-------------------------------------------------------------------------------
bypass-link192.168.13.2-61443
-------------------------------------------------------------------------------
To             : 192.168.34.1        State               : Up
Out I/F        : 1/1/1               Out Label           : 262136
Up Time        : 0d 00:03:42         Active Time         : n/a
Reserved BW    : 0 Kbps              Protected LSP Count : 1
Type           : P2mp                Bypass Path Cost    : 30
Setup Priority : 7                   Hold Priority       : 0
Class Type     : 0
Exclude Node   : None                Inter-Area          : False
Computed Hops  :
```

```
    192.168.12.1(S)                  Egress Admin Groups : None
 -> 192.168.12.2(S)                  Egress Admin Groups : None
 -> 192.168.24.2(S)                  Egress Admin Groups : None
 -> 192.168.34.1(S)                  Egress Admin Groups : None
Actual Hops    :
    192.168.12.1 (192.0.2.1)         Record Label        : N/A
 -> 192.168.12.2 (192.0.2.2)         Record Label        : 262136
 -> 192.168.24.2 (192.0.2.4)         Record Label        : 262135
 -> 192.168.34.1 (192.0.2.3)         Record Label        : 262134

Protected LSPs -
LSP Name       : VRF2-2-73732::EMPTY
From           : 192.0.2.1           To                  : 192.0.2.3
Avoid Node/Hop : 192.168.13.2        Downstream Label    : 262136
Bandwidth      : 0 Kbps

===============================================================================
*A:PE-1#
```

Traffic Flow

The receiver H-8, connected to CE-8, wishes to join the group 232.2.2.2 with source 192.168.52.1 and so sends an IGMPv3 report toward CE-8. CE-8 recognizes the report and sends a PIM join toward the source, therefore, it reaches PE-1 where the source is connected to through CE-5. The following output shows the debug seen on PE-4, where the PIM join is received from CE-8 and a BGP update Source Join is sent to all PEs (only the update sent to PE-1 is shown).

```
1 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPProcessSG
pimJPProcessSG: (S,G)-> (192.168.52.1,232.2.2.2) type <S,G>,
i/f int-PE-4-CE-8, upNbr 172.16.148.1 isJoin 1 isRpt 0 holdTime 210"

2 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmFindRpfNexthop
Track (192.168.52.1,232.2.2.2) type <S,G> using 192.168.52.1"

3 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmAddSrcEntry
Added src entry for src 192.168.52.1"

4 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Downstream: State NoInfo Event RxJoin StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

5 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Upstream: State NotJoined Event JoinDesiredTrue StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

6 22016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGUpJoinDesiredTrue
No upstream interface. pSG (192.168.52.1,232.2.2.2) rpfType 3"
```

```
7 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGUpJoinDesiredTrue
No upstream interface SG (192.168.52.1,232.2.2.2) rpfType 3"

8 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmProcessNhresEvent
RTM-Nhres Event U-RTM NEW Src 192.168.52.1 SrcRtmUse UCAST"

9 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmProcessNhresEvent
Prefix 192.168.52.0/24 numNextHops 1 owner BGP_VPN metric 20 pref 170"

10 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmSrcResolveSGsInt
Trying to resolve SG (192.168.52.1,232.2.2.2)"

11 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmNotifyRpfChange
RPF Change to Source/
RP 192.168.52.1  for SG (192.168.52.1,232.2.2.2) dynMLDP F via NH 192.0.2.1
IfIdx: 73734 RpfType: REMOTE Reason:  RTE_ADD old NH 0.0.0.0 IfIdx: 0 RpfType: NONE
mplsRpf F NextHops 1 reg 1/1 lfa 0/0"

12 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmNotifyRpfChange
SG (192.168.52.1,232.2.2.2) Source/RP 192.168.52.1 Ipmsi 73732 NhIf 0
new NhIf 73734"

13 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Upstream: State Joined Event MribChange StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

14 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGUpStateJMribChange
SG (192.168.52.1,232.2.2.2), type <S,G> oldMribNhopIp 0.0.0.0 oldRpfNbrIp 0.0.0.0,
oldRpfType NONE oldRpfIf 0 rptMribNhopIp 0.0.0.0, rptRpfNbrIp 0.0.0.0 rtmReason 48
isSGExtNet : no"

15 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGUpStateJMribChange
SG (192.168.52.1,232.2.2.2), type <S,G> newMribNhopIp 192.0.2.1
newRpfNbrIp 192.0.2.1 newRpfType REMOTE newRpfIf 73734"

16 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimAddToJPTxPdu
pimAddToJPTxPdu: (S,G)-> (192.168.52.1,232.2.2.2), type <S,G>, txPendFlag J
isStandby F"

17 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmUpdateSGMetric
SG metric 4294967295 pref 2147483647, new metric 20 pref 170"

20 2016/10/10 12:06:44.78 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
```

```
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.4
        Type: Source-Join Len:22 RD: 64496:201 SrcAS: 64496
                          Src: 192.168.52.1 Grp: 232.2.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:3
"
```

The following debug shows that PE-1 receives the BGP update Source Join with source 192.168.52.1 and group 232.2.2.2 and sends a PIM join toward CE-5:

```
1 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.4
        Type: Source-Join Len:22 RD: 64496:201 SrcAS: 64496
                          Src: 192.168.52.1 Grp: 232.2.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:3
"

2 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimProcessMvpnRouteMsg
originator 0.0.0.0: add rtType 7 nextHop 192.0.2.4 source 192.168.52.1
group 232.2.2.2"

3 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPProcessSG
pimJPProcessSG: (S,G)-> (192.168.52.1,232.2.2.2) type <S,G>,
i/f mpls-if-73732, upNbr 192.0.2.1 isJoin 1 isRpt 0 holdTime 65535"

4 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmFindRpfNexthop
Track (192.168.52.1,232.2.2.2) type <S,G> using 192.168.52.1"

5 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmAddSrcEntry
Added src entry for src 192.168.52.1"

6 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
```

```
PIM JP Downstream: State NoInfo Event RxJoin StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

7 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Upstream: State NotJoined Event JoinDesiredTrue StandbyEvent F, (S,G) (192.16
8.52.1,232.2.2.2) groupType <S,G>"

8 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmProcessNhresEvent
RTM-Nhres Event U-RTM NEW Src 192.168.52.1 SrcRtmUse UCAST"

9 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmProcessNhresEvent
Prefix 192.168.52.0/24 numNextHops 1 owner BGP metric 0 pref 170"

20 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGEncodeGroupSet
Encoding  Join for source 192.168.52.1"

21 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGEncodeGroupSet
num joined srcs 1, num pruned srcs 0"

22 2016/10/10 12:09:48.22 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSendJoinPrunePdu
sending JP PDU with 1 groups, if 3 adj 172.16.115.2"
```

The BGP update source join received by PE-1 is displayed with the following command:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type source-join
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Source-Join                  -                      100         0
      64496:201                    64496                              -
      192.0.2.4                    192.168.52.1
      No As-Path                   232.2.2.2
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

To verify the traffic: on PE-1, there is a group 232.2.2.2 with source 192.168.52.1, the RPF is CE-5, and the multicast traffic is flowing from CE-5 to PE-1 using int-PE-1-CE-5 and the outgoing interface is using the PMSI RSVP mpls-if-73732.

```
*A:PE-1# show router 2 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.2.2.2
Source Address     : 192.168.52.1
RP Address         : 0
Advt Router        : 172.16.115.2
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.115.2
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:26:05        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:55
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.115.2
Incoming Intf      : int-PE-1-CE-5
Outgoing Intf List : mpls-if-73732

Curr Fwding Rate   : 6690.1 kbps
Forwarded Packets  : 29636630           Discarded Packets  : 0
Forwarded Octets   : 1363284980         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

On PE-4, the same (S,G) arrives in the incoming interface mpls-if-73734 and the outgoing interface is int-PE-4-CE-8.

```
*A:PE-4# show router 2 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.2.2.2
Source Address     : 192.168.52.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:23:40        Resolved By        : rtable-u
```

```
Up JP State        : Joined             Up JP Expiry       : 0d 00:00:20
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73734
Outgoing Intf List : int-PE-4-CE-8

Curr Fwding Rate   : 6798.4 kbps
Forwarded Packets  : 27013244           Discarded Packets  : 0
Forwarded Octets   : 1242609224         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-4#
```

When the receiver is not interested in the channel group anymore, the receiver H-8 sends an IGMPv3 leave, PE-4 sends a PIM prune translated to a BGP MP_UNREACH NLRI to all PEs. As mentioned before, rapid withdrawals are sent without waiting for the MRAI (for simplicity, only one BGP update is shown in the output debug).

```
23 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPProcessSG
pimJPProcessSG: (S,G)-> (192.168.52.1,232.2.2.2) type <S,G>,
i/f int-PE-4-CE-8, upNbr 172.16.148.1 isJoin 0 isRpt 0 holdTime 210"

24 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Downstream: State Joined Event RxPrune StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

25 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Downstream: State PrunePending Event PrunePendTimerExp StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

26 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimJPPrintFsmEvent
PIM JP Upstream: State Joined Event JoinDesiredFalse StandbyEvent F,
(S,G) (192.168.52.1,232.2.2.2) groupType <S,G>"

28 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmStopRpfNexthop
Stop tracking (192.168.52.1,232.2.2.2) type <S,G> with 192.168.52.1
pRtmNhop 0xa2f2c590"

29 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimRtmDelSrcEntry
Deleted src entry for src 192.168.52.1"

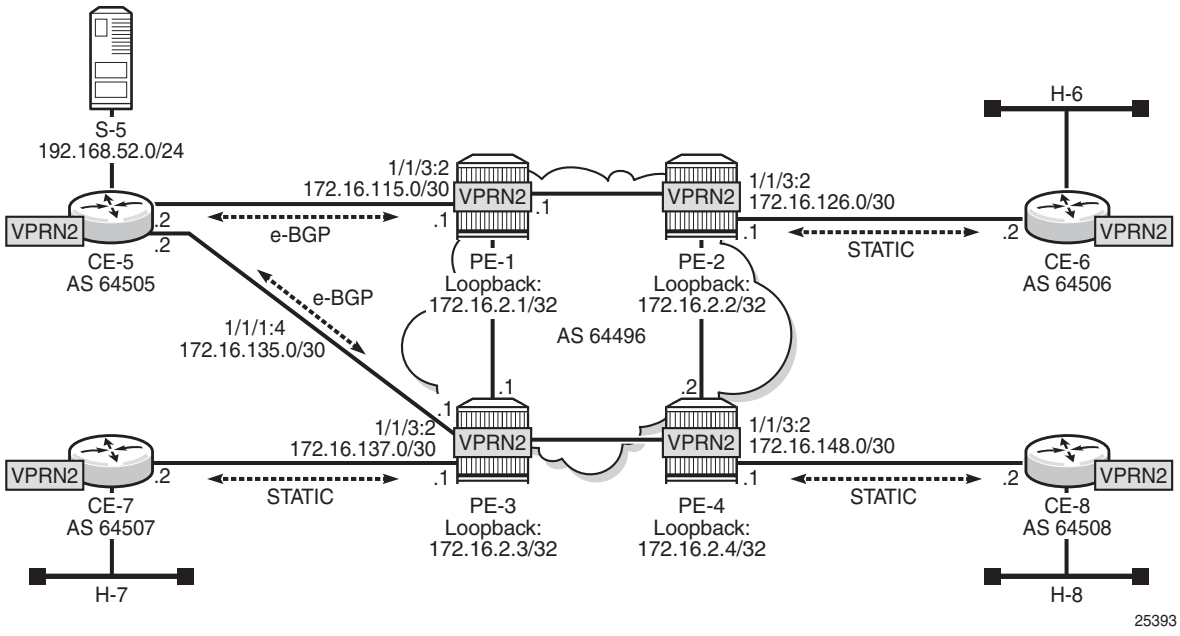
31 2016/10/10 12:34:14.34 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
```

```
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 31
    Flag: 0x90 Type: 15 Len: 27 Multiprotocol Unreachable NLRI:
        Address Family MVPN_IPV4
        Type: Source-Join Len:22 RD: 64496:201 SrcAS: 64496
                             Src: 192.168.52.1 Grp: 232.2.2.2
"
```

MVPN Source Redundancy

So far, the multicast traffic has been streamed toward router CE-5 from a single source. For redundancy purposes, the source can be redundant (two sources attached to different CEs that connect to a pair of PEs). To simulate the redundancy, CE-5 has been connected to both PE-1 and PE-3, using VPRN 2, and equal cost multi-path (ECMP) is configured with the value of 2 in all PEs. With this configuration, any PE is able to reach the source through PE-1 and PE-2. The (S,G) is the same as the one used in P2MP RSVP TE (192.168.52.1, 232.2.2.2). Figure 331 shows the VPRN 2 topology with the source redundancy.

Figure 331 **VPRN 2 Topology used for MVPN Source Redundancy**



The configuration change with respect to the previous section (P2MP RSVP-TE PMSIs) is an additional interface created in both CE-5 and PE-3 (int-CE-5-PE-3 on CE-5 and int-PE-3-CE-5 on PE-3), the addition of these interfaces to PIM and also the creation an e-BGP session between the two routers. The following is the additional configuration on PE-3 (CE-5 configuration changes are not displayed for brevity).

```
*A:PE-3# configure service
        vprn 2
            interface "int-PE-3-CE-5" create
                address 172.16.35.1/30
                sap 1/1/4:2 create
                exit
            exit
            bgp
                group "EXTERNAL"
                    type external
                    peer-as 64505
                    neighbor 172.16.35.2
                    exit
                exit
                no shutdown
            exit
            pim
                interface "int-PE-3-CE-5"
                exit
            exit
        exit
```

Checking the routes on PE-4, the source is reachable through PE-1 and PE-2 as ECMP is set to 2. If the configuration of the VPRN is provisioned with **auto-bind-tunnel resolution-filter rsvp resolution filter,** instead of static spoke-SDPs, the command **ignore-nh-metric** is also needed.

```
*A:PE-4# configure service vprn 2 ecmp 2


*A:PE-4# show router 2 route-table
===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
---snip---
192.168.52.0/24                                 Remote  BGP VPN   00h03m08s  170
       192.0.2.1 (tunneled)                                          0
192.168.52.0/24                                 Remote  BGP VPN   00h03m08s  170
       192.0.2.3 (tunneled)                                          0
-------------------------------------------------------------------------------
No. of Routes: 9
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

```
*A:PE-4#
```

When PE-4 receives a c-join/prune, PE-4 needs to find the **upstream multicast hop** (UMH) for the (S,G). This is the upstream multihop selection and is configurable. The values are highest-ip, hash-based, tunnel-status, and unicast-rt-pref

```
*A:PE-4# configure service vprn 2 mvpn umh-selection
  - no umh-selection
  - umh-selection {highest-ip|hash-based|tunnel-status|unicast-rt-pref}
```

The default is highest-ip, which is the selection of the highest /32 IP addresses (in this setup, PE-3 is preferred versus PE-1). A BGP c-join is sent with the route target equal to the VRF import extended community distributed by PE-3 for the subnet of the source (see following PE-4 debug).

```
15 2016/10/10 12:51:16.84 UTC MINOR: DEBUG #2001 vprn2 PIM[vprn 3 vprn2]
"PIM[vprn 3 vprn2]: pimSGUpStateJMribChange
SG (192.168.52.1,232.2.2.2), type <S,G> newMribNhopIp 192.0.2.3
newRpfNbrIp 192.0.2.3 newRpfType REMOTE newRpfIf 73735"

18 2016/10/10 12:51:16.84 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.4
        Type: Source-Join Len:22 RD: 64496:203 SrcAS: 64496
                          Src: 192.168.52.1 Grp: 232.2.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.3:3
"
```

The second option is hash-based, where the UMH is selected (both PEs are potentially possible UMHs) after hashing the source and group addresses of the stream. For this example, PE-3 is also preferred.

The third option, tunnel-status, is based on the status of the P2MP RSVP tunnel (not available in mLDP or PIM). The roots PE-1 and PE-3 are sending BFD messages to the leaf PE-4 (in fact this is UFD, unidirectional forwarding detection). The c-join from PE-4 for the (S,G) is sent to both PE-1 and PE-3, and in return the traffic is forwarded from both PE-1 and PE-3 for the c-group onto the I-PMSI; therefore PE-4 receives two copies of the c-(S,G) stream. By configuration, the stream from the primary PE-1 is selected by PE-4 to be forwarded to receiver H-8. If BFD messages are no longer received over the primary P2MP LSP, then the stream from the standby PE-3 is selected and forwarded to the receiver.

The configuration on PE-1 and PE-3 is similar and is as follows (only PE-3 is shown):

```
*A:PE-3# configure service vprn 2
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                umh-selection tunnel-status
                provider-tunnel
                    inclusive
                        rsvp
                            lsp-template "VRF2"
                            enable-bfd-root 100
                            no shutdown
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
```

PE-1 and PE-3 are root. On PE-4, BFD is configured as leaf and the primary PE (PE-1) and backup PE (PE-3) are also provisioned:

```
*A:PE-4# configure service vprn 2
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                umh-selection tunnel-status
                umh-pe-backup
                    umh-pe 192.0.2.1 standby 192.0.2.3
                exit
                provider-tunnel
                    inclusive
                        rsvp
                            lsp-template "VRF2"
                            enable-bfd-leaf
                            no shutdown
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
```

This BFD (UFD) configuration on the root establishes a session with the leaf. The root only transmits BFD packets; it doesn't receive any.

```
*A:PE-1# show router 2 bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
mpls-if-73736                 Up                   100       0         3
   127.0.0.0                  pim                  4305      0         central
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
*A:PE-1#
```

On PE-4, two BFD sessions are received, one from each root (note that BFD packets are only received):

```
*A:PE-4# show router 2 bfd session
===============================================================================
Legend:  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
If/Lsp Name/Svc-Id/RSVP-sess  State                Tx Intvl  Rx Intvl  Multipl
  Rem Addr/Info/SdpId:VcId    Protocols            Tx Pkts   Rx Pkts   Type
   LAG port                      LAG ID
-------------------------------------------------------------------------------
mpls-if-73739                 Up                   1000      100       3
   192.0.2.3                  pim                  0         793       central
mpls-if-73740                 Up                   1000      100       3
   192.0.2.1                  pim                  0         792       central
-------------------------------------------------------------------------------
No. of BFD sessions: 2
===============================================================================
*A:PE-4#
```

PE-4 delivers the multicast traffic from the primary configured UMH, PE-1. If, as an example of a failure condition, PE-1 goes down (reboot), PE-4 will switch to the PE-3 P2MP LSP.

MDT AFI SAFI for Rosen MVPN

In Rosen MVPN up to version 6, the default MDT is PIM sparse mode only, and there is no auto-discovery mechanism available. In SR-OS Release 7.0, and later, it is possible to configure PIM SSM with auto-discovery, using AFI 1 and SAFI 5. Rosen MVPN version 7 allows use of MDT PIM SM or SSM, and auto-discovery based on AFI 1 and SAFI 66 to distribute the default MDT. Rosen MVPN version 9 adds a new MDT NLRI. SR OS has added the capability and support of MDT SAFI as specified in RFC 6037.

MDT SAFI is used to discover PEs in a specific MVPN, so that PIM SSM can be used for default MDT. The basic idea is the same as MVPN BGP auto-discovery, but it uses a different BGP SAFI. BGP messages in which AFI=1 and SAFI=66 are "MDT-SAFI" messages. The NLRI format is 8-byte-RD:IPv4-address followed by the MDT group address. The IPv4 address identifies the PE that originated this route and the RD identifies a VRF in that PE. The group address must be an IPv4 multicast group address and is used to build the P-tunnels.

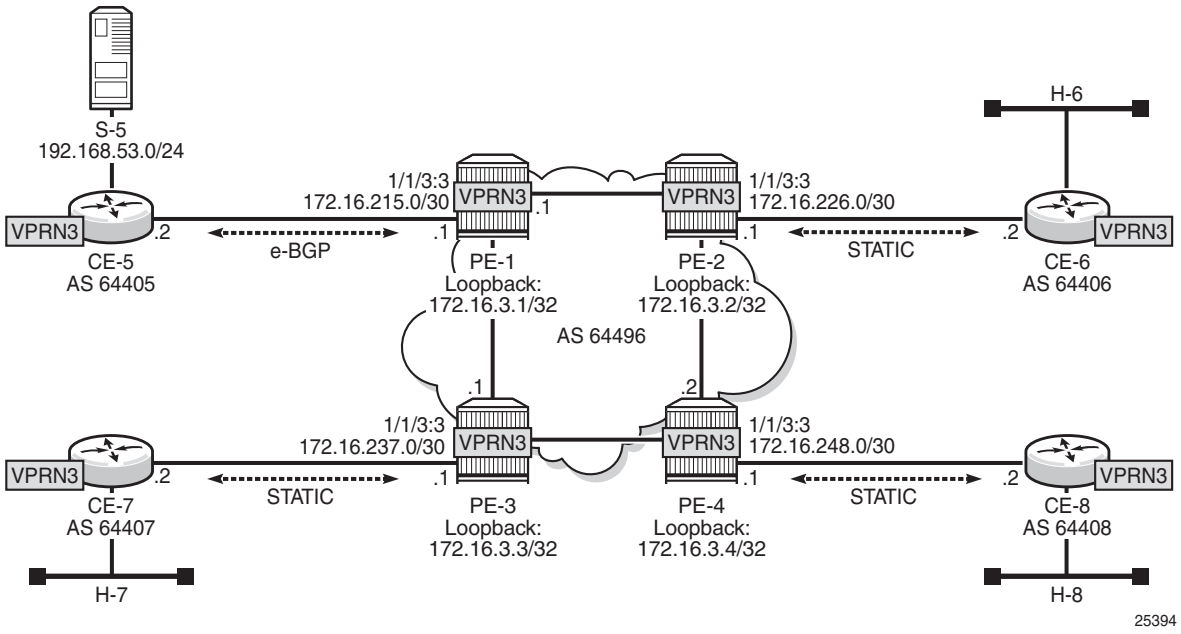
All PEs attached to a given MVPN must specify the same group-address. MDT-SAFI routes do not carry RTs and the group address is used to associate a received MDT-SAFI route with a VRF.

MDT SAFI can only be used when the implicit provider tunnel is PIM GRE based with a specific IPv4 group address.

For additional information on the use of PIM PMSIs, see NG-MVPN Configuration with PIM.

Figure 332 shows the topology of VPRN 3.

Figure 332 **VPRN 3 Topology used for MDT SAFI**



In this scenario, there is no MPLS based PMSI, there is PIM in the core for the control plane and the data traffic is GRE encapsulated. PIM needs to be configured in the base router on interface system and on the other interfaces pointing to other PEs. PIM is used for c-signaling. In addition, auto-discovery is provisioned to use mdt-safi and a PIM SSM inclusive PMSI with address 239.1.1.1 as the default MDT. The configuration is as follows on PE-4:

```
*A:PE-4# configure router
        pim
            interface "system"
            exit
            interface "int-PE-4-PE-2"
            exit
            interface "int-PE-4-PE-3"
            exit
        exit
    exit

*A:PE-4# configure service
        vprn 3 customer 1 create
            description "PIM SSM / MDT SAFI"
            autonomous-system 64496
            route-distinguisher 64496:304
            vrf-target target:64496:300
            interface "loopback" create
                address 172.16.3.4/32
                loopback
            exit
            interface "int-PE-4-CE-8" create
```

```
            address 172.16.248.1/30
            sap 1/1/3:3 create
            exit
        exit
        pim
            interface "loopback"
            exit
            interface "int-PE-4-CE-8"
            exit
        exit
        mvpn
            auto-discovery mdt-safi
            provider-tunnel
                inclusive
                    pim ssm 239.1.1.1
                    exit
                exit
            exit
            vrf-target unicast
            exit
        exit
        spoke-sdp 341 create
        exit
        spoke-sdp 342 create
        exit
        spoke-sdp 343 create
        exit
        no shutdown
```

The following debug output shows a BGP update with MDT AFI SAFI on PE-4:

```
12 2016/10/10 13:00:48.22 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 62
    Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
        Address Family MDT-SAFI
        NextHop len 4 NextHop 192.0.2.4
        [MDT-SAFI] Addr 192.0.2.4, Group 239.1.1.1, RD 64496:304
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:300
"
```

The following output shows the MDT-SAFI routes that have been learned at PE-4:

```
*A:PE-4# show router bgp routes mdt-safi
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
```

```
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MDT-SAFI Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop                    Group-Addr                        Label
      As-Path
-------------------------------------------------------------------------------
u*>i  64496:301:192.0.2.1                              100         0
      192.0.2.1                  239.1.1.1                         -
      No As-Path
u*>i  64496:302:192.0.2.2                              100         0
      192.0.2.2                  239.1.1.1                         -
      No As-Path
u*>i  64496:303:192.0.2.3                              100         0
      192.0.2.3                  239.1.1.1                         -
      No As-Path
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-4#
```

Conclusion

This chapter provides information to configure multicast within a VPRN with next generation multicast VPN techniques. Specifically, the use of MPLS I-PMSIs (mLDP and P2MP RSVP-TE), MVPN source redundancy, and the complete set of features needed to interoperate with Rosen MVPN in live deployments are covered.

Rosen MVPN Core Diversity

This chapter provides information about Rosen multicast virtual private network (MVPN) core diversity.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS release 13.0.R4, using Rosen MVPN. Default multicast distribution trees (MDTs) for each virtual private routed network (VPRN) are signaled using protocol independent multicasting (PIM) and auto-discovery uses border gateway protocol multicast distribution tree sub-address family indicator (BGP MDT-SAFI) network layer routing. Chassis mode D is required.

Overview

This chapter describes a service provider core network used by multiple content providers to deliver multicast services to multiple customers using Rosen MVPN. If the same set of PEs is used to deliver the MVPN, the MDTs will all be routed across the same paths between the set of PEs. Because each MDT is signaled using PIM, and the source of all MDTs is the system address of the PE, the path to this source is the same.

Each remote PE then sends a PIM join toward this PE with its source address set to the system address. For multiple VPNs between the same set of PEs, the MDT will follow the same path.

If there is a requirement to deliver content from each content provider across different MVPNs that use diversely routed MDTs, multiple IGP instances can be used: up to three different instances of IGP, OSPF or ISIS, can exist. In this chapter, two instances of OSPF are used to create incongruent topologies providing isolation between the MDTs of two different MVPNs: a default OSPF instance and OSPF instance 1. A separate /32 loopback address can be used as the MDT source address that is advertised in the non-default IGP, which can also be used as the BGP next hop for labeled IPv4 routes representing the customer source addresses.

Knowledge of multi-protocol BGP (MP-BGP) and RFC 4364 (*BGP/MPLS IP VPNs*) is assumed throughout this chapter, as well as the original RFC 6037.

All PEs within an MVPN create a default MDT with their own system address as the source. Auto-discovery of PEs within a Rosen MVPN is achieved using the BGP route type of multicast data tree subsequent address family identifier (MDT-SAFI). Each PE originates an MDT-SAFI route update per MVPN. This route advertises the presence of the MVPN on a specific PE.

Each MDT-SAFI update contains attributes, including the following:

1. Route distinguisher
2. Route target extended community
3. MDT source address (usually the system address)
4. Group address of MDT

Upon receipt of an MDT-SAFI route update, each remote PE accepts or rejects the route based on the route target extended community value. If the route is accepted, a remote PE sends a PIM (S,G) join to this local PE. The (S,G) values are taken from the MDT-SAFI. The set of MDTs extend the c-multicast data tree across the MVPN and form PIM adjacencies between PEs within the MVPN. The neighbor address across the set of PIM-enabled tunnels is the default MDT source address, usually the system address.

When established, the default MDT is used to transport c-multicast PIM signaling between PEs.

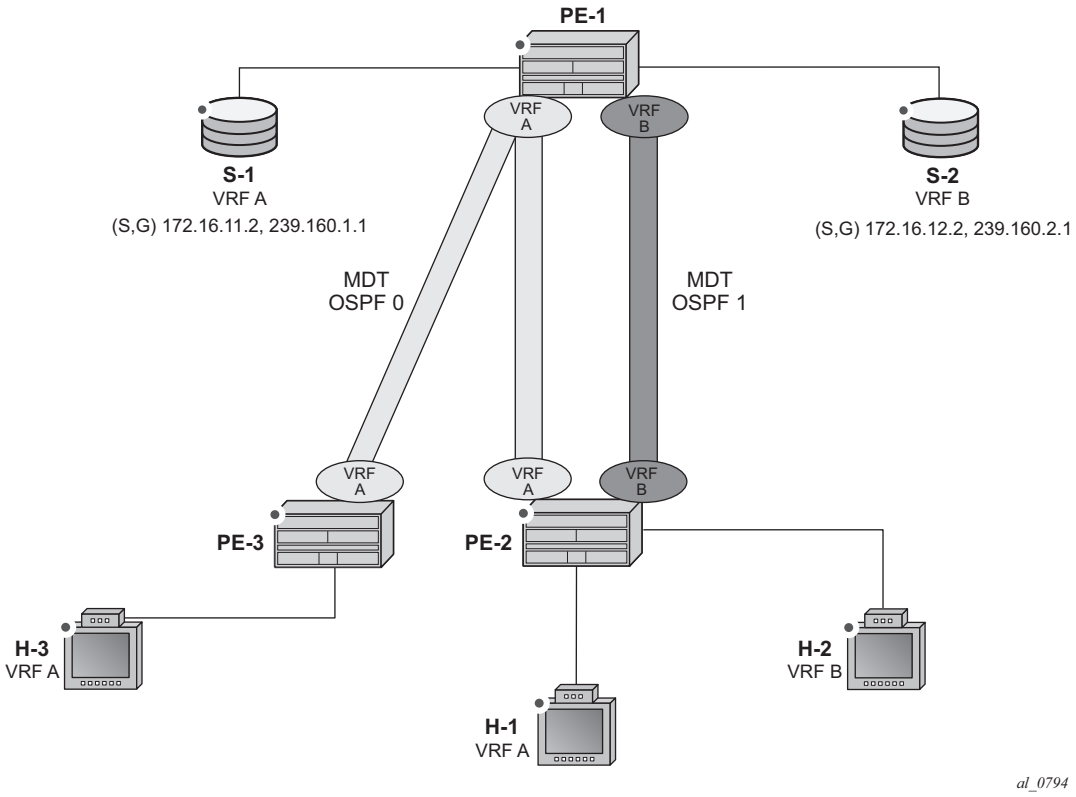
If a source S, of a c-multicast group G, is connected to a sender PE, the route to the source is advertised to remote PEs as a BGP-labeled VPN-IPv4 route.

Therefore, an (S,G) join toward this source at a remote PE will perform a reverse path forwarding (RPF) look-up of the unicast VRF table to find a suitable PIM-enabled interface. The next hop needs be resolved to the MDT source address of the sender PE. A PIM join must now be forwarded toward the sender PE that has a PIM neighbor that matches the next hop for this route, the system address of the sender PE. This is the default MDT.

The system address is a significant address in this process. Any other VPRN that uses the same set of PEs will also signal a set of default MDTs using a different group address, so they will follow the same path between PEs across the provider network.

Figure 333 shows an example of core diversity; multicast sources provided by two separate content providers are connected to a provider network. There is a requirement to provide topology diversity so that the default MDTs between the same PEs are routed across different paths within the core.

Figure 333 **Core Diversity Schematic**



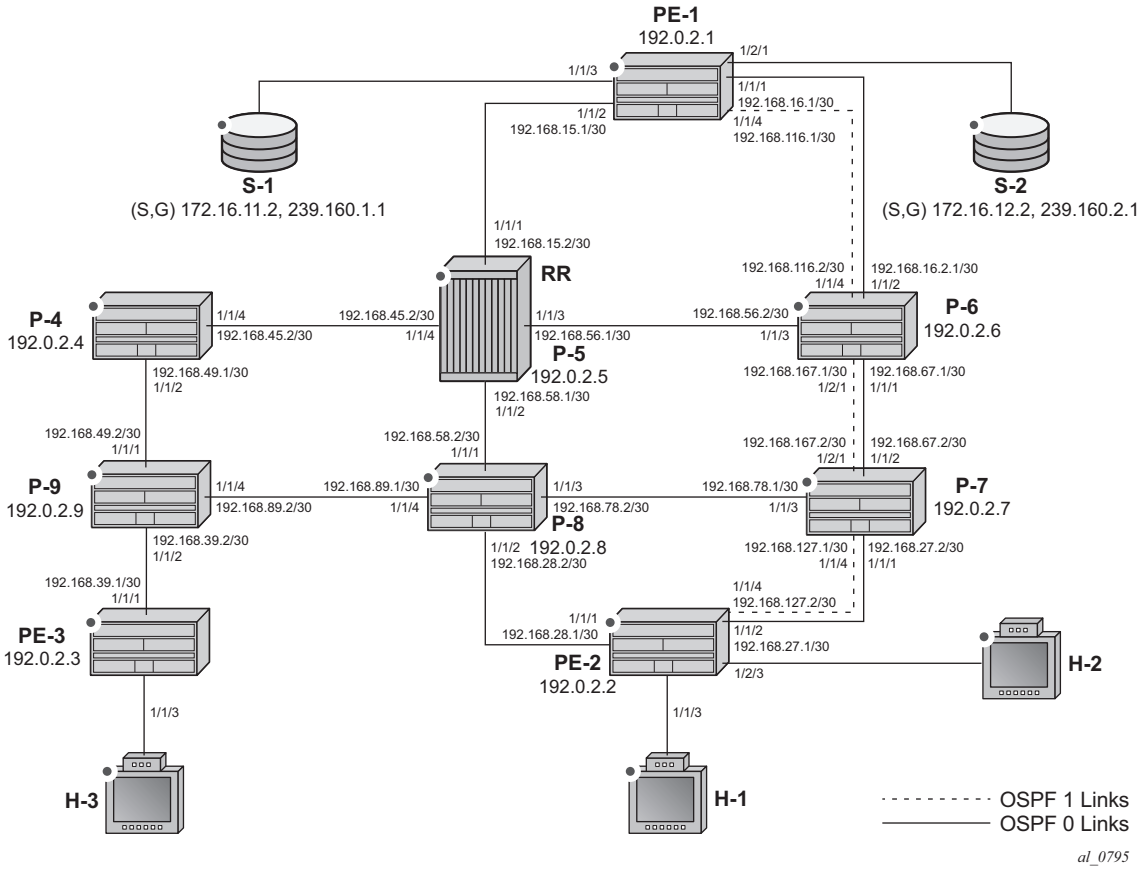
Content servers from two separate content providers are connected to PE-1 with directly connected multicast sources. For simplicity, this example uses only a single multicast group for provider S-1 and S-2.

Source S-1 is reachable via VRF A and source S-2 is reachable via VRF B.

Topology isolation for the multicast data trees of each VPRN can be provided by using two separate IGP instances; in this case, OSPF instances. Multi-instance IS-IS could also be used.

Figure 334 shows a schematic of the full network, including the c-multicast groups.

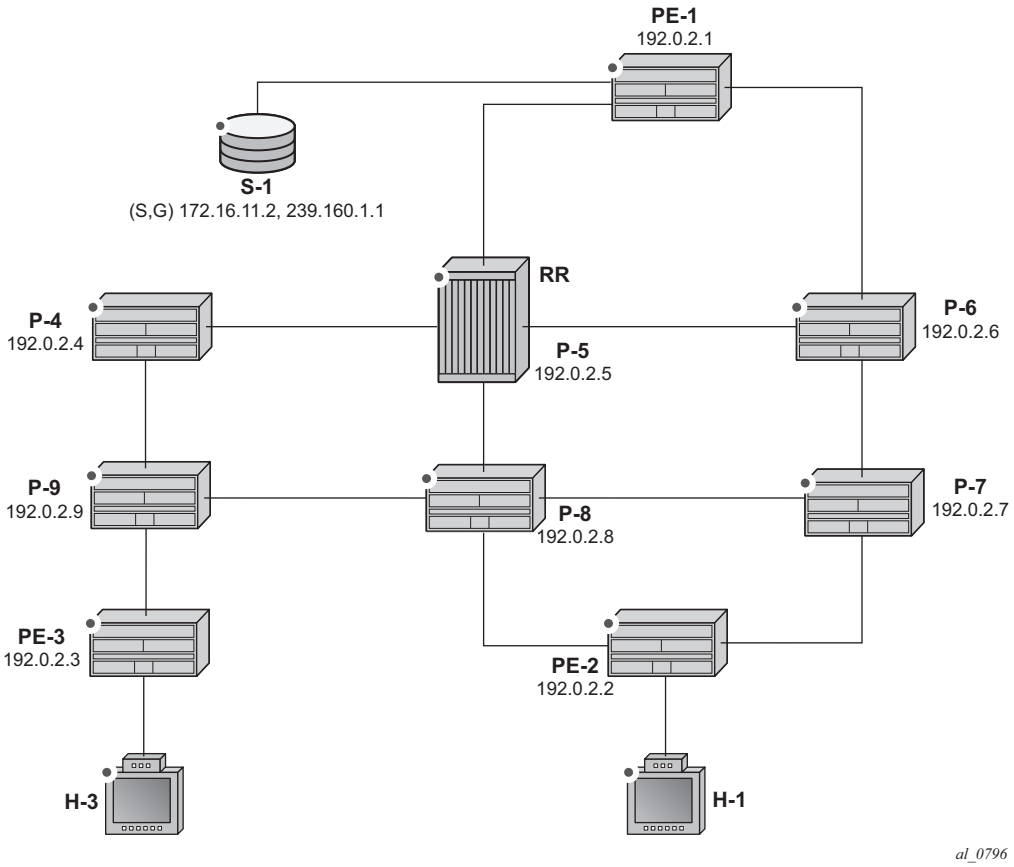
Figure 334 **Core Diversity Network**



All routers have interfaces in the OSPF base instance (instance 0) and routers interconnected by the dotted lines have interfaces in both the base instance and OSPF 1.

Figure 334 shows the extent of the OSFP base instance within the core network.

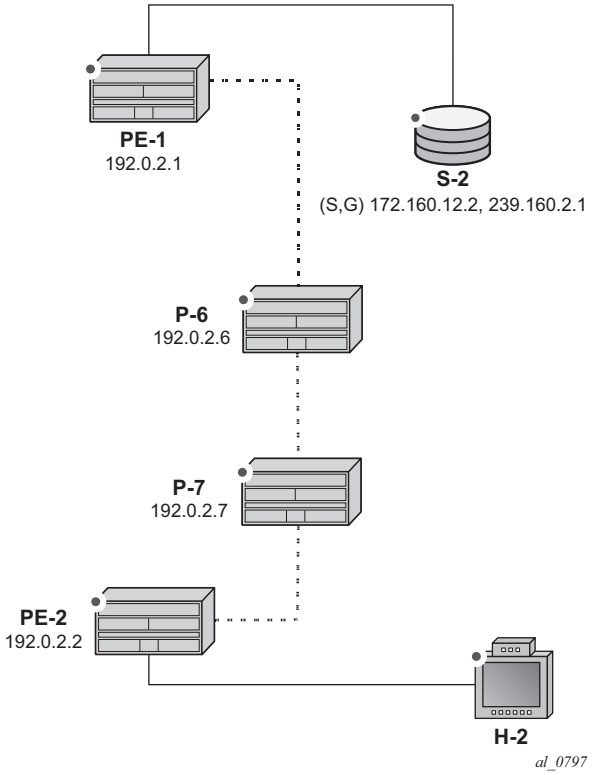
Figure 335 **Core Diversity Network — Base OSPF**



In this case, assume that the shortest path between PE-1 and PE-2 is the path via P-5 and P-8.

Similarly, Figure 336 shows the extent of OSPF instance 1.

Figure 336 **Core Diversity Network - OSPF Instance 1**



The only path available between PE-1 and PE-2 is now completely diverse from the shortest path advertised between the same pair of PEs in the base OSPF instance.

Therefore, for any MDT to be signaled across the OSFP 1 topology, only addresses advertised within OSPF 1 must be used. As previously stated, the system address is used as the default MDT source address. This system address is not advertised within the OSPF 1 topology, so a replacement /32 loopback address is used as the default MDT source address within OSPF 1.

VPN-IPv4 routes that may represent a customer multicast source address should be reachable via the default MDT. In the non-default topology, the c-multicast signaling across the MVPN must resolve the c-multicast route via the MDT, which has its root at the non-default /32 loopback. Therefore, the VPN-IPv4 prefix representing the possible source routes needs to be advertised containing the non-default /32 loopback.

This can be achieved in one of two ways:

1. Use a route policy at the advertising PE that changes the BGP next hop to match the MDT source address for non-default topology MDTs.

2. Use the BGP connector attribute for all VPN-IPv4 route prefixes within a multicast VPRN that has auto-discovery set to MDT-SAFI. The connector attribute contains the MDT source address within the originator field.

This chapter describes the use of the connector attribute.

If the default IGP instance is used, the BGP next hop of the VPN-IPv4 route matches the source address of the default MDT.

Therefore, if a second /32 loopback is used that replaces the system address as MDT source address and also as the next hop for source address RPF look-up, the loopback could be advertised within the non-default IGP instance, and the paths between the PEs would follow this topology.

Core diversity is achieved by configuring the following steps:

1. Configuring multiple OSPF instances, as shown in Figure 335 and Figure 336, and including the appropriate interfaces. This includes a separate loopback address per instance.
2. Configuring separate VPRNs with their own MDTs using BGP MDT-SAFI auto-discovery and PIM signaling across the appropriate PEs.
3. Configuring the VPRN that uses the base OSPF instance to use the system address as the source addresses for the MDTs (this is default behavior).
4. Configuring a separate loopback (/32) address that is advertised within OSFP instance 1 only.
5. Configuring the VPRN that uses the OSPF instance 1 to use the separate loopback system address as the source addresses for the MDTs.
6. Ensuring the unicast route that represents the c-source address is advertised as a VPN-IPv4 route and has a BGP connector attribute that contains an address that matches the MDT source address of the originating PE.

Configuration

The following configuration tasks must be completed as a prerequisite:

- Full mesh OSPF base instance between each of the nodes. However, IS-IS could also be used for any or all of the IGP instances.
- Link-layer LDP between each P and PE router.
- PIM enabled on each router network interface.

Global BGP Configuration

The first step is to configure an iBGP session between each of the PEs and the route reflector (P-5) shown in Figure 334. The address families negotiated between the iBGP peers are **vpn-ipv4**, for unicast routing, and **mdt-safi** for multicast routing.

The iBGP configuration for PE-1 is the following:

```
*A:PE-1# configure
    router
        bgp
            group "INTERNAL"
                family vpn-ipv4 mdt-safi
                type internal
                neighbor 192.0.2.5
                exit
            exit
```

The configuration for the other PE nodes is the same.

P-5 is the route reflector for PE-1, PE-2, and PE-3, as follows:

```
*A:P-5# configure
    router
        bgp
            cluster 0.0.0.1
            group "RR_CLIENTS"
                family vpn-ipv4 mdt-safi
                type internal
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

On PE-1, verify that the BGP session with the route reflector at P-5 is established with address families **mdt-safi** and **vpn-ipv4** capabilities negotiated:

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
Total BGP Paths         : 20          Total Path Memory           : 4000
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
```

```
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.5
                64496       5    0 00h01m00s 0/0/0 (VpnIPv4)
                            5    0           0/0/0 (MdtSafi)
-------------------------------------------------------------------------------
*A:PE-1#
```

Configuring VPRN on PEs

There are two VPRNs:

- VPRN 1 using the base instance OSPF topology. This is present on PE-1, PE-2, and PE-3.
- VPRN 2 using OSPF instance 1. This is present on PE-1 and PE-2.

The following output displays the configuration for VPRN 1 for the sender PE-1.

```
*A:PE-1# configure
    service
```

```
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-1-S-1" create
                address 172.16.11.1/24
                sap 1/1/3 create
                exit
            exit
            pim
                apply-to all
                no shutdown
            exit
            mvpn
                auto-discovery mdt-safi
                provider-tunnel
                    inclusive
                        pim ssm 239.160.1.1
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
            no shutdown
```

There is a single interface toward S-1, from which the multicast group is received.

PIM is enabled and applied to all interfaces.

The MVPN configuration enables BGP MDT-SAFI as the auto-discovery mechanism. The provider tunnels between the PEs within the MVPN are PIM SSM multicast data trees with a group address of 239.160.1.1.

The configuration for VPRN 1 for the receiver PE-2 is the following.

```
*A:PE-2# configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-2-H-1" create
                address 172.16.21.1/24
                sap 1/1/3 create
                exit
```

```
            exit
            igmp
                interface "int-PE-2-H-1
                    no shutdown
                exit
                no shutdown
            exit
            pim
                apply-to all
                no shutdown
            exit
            mvpn
                auto-discovery mdt-safi
                provider-tunnel
                    inclusive
                        pim ssm 239.160.1.1
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
            no shutdown
```

The configuration for VPRN 1 for receiver PE-3 is as follows.

```
*A:PE-3# configure service
         vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-3-H-3" create
                address 172.16.33.1/24
                sap 1/1/3 create
                exit
            exit
            igmp
                interface "int-PE-3-H-3"
                    no shutdown
                exit
                no shutdown
            exit
            pim
                apply-to all
                no shutdown
            exit
            mvpn
                auto-discovery mdt-safi
                provider-tunnel
                    inclusive
                        pim ssm 239.160.1.1
                        exit
                    exit
```

```
            exit
            vrf-target unicast
            exit
        exit
        no shutdown
```

At PE-2, the MDT SAFI NLRI advertised by PE-1 is as follows:

```
A:PE-2# show router bgp routes mdt-safi grp-address 239.160.1.1 source-ip 192.0.2.1
detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MDT-SAFI Routes
===============================================================================
Original Attributes

Route Dist.    : 64496:1
Source Addr    : 192.0.2.1
Group Addr     : 239.160.1.1
Nexthop        : 192.0.2.1
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:64496:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h17m09s

---snip---
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

The source and group address is used by PE-2 (and PE-3) to join the MDT that has its root at PE-1. The source address used is the system address of PE-1.

Examining the MDTs for this VPRN at PE-1 shows the state, as follows:

```
*A:PE-1# show router pim group 239.160.1.1
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type               Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.160.1.1                 (S,G)              spt      system         2
   192.0.2.1
239.160.1.1                 (S,G)              spt      int-PE-1-P-5   1
   192.0.2.2
239.160.1.1                 (S,G)              spt      int-PE-1-P-5   1
   192.0.2.3
-------------------------------------------------------------------------------
Groups : 3
===============================================================================
*A:PE-1#
```

The MDT with the root of its tree at PE-1 is as follows:

```
*A:PE-1# show router pim group 239.160.1.1 detail source 192.0.2.1

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.1.1
Source Address     : 192.0.2.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : self
Keepalive Timer Exp: 0d 00:03:05
Up Time            : 0d 00:10:28        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:31
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : system
Outgoing Intf List : system, int-PE-1-P-5

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 6                  Discarded Packets  : 0
Forwarded Octets   : 468                RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

```
*A:PE-1#
```

The source address of the tree is the system address of the router, which is determined from the MDT SAFI NLRI that is advertised to all other PEs via the route reflector. Also, the outgoing interface list contains an interface (int-PE-1-P-5) that is OSPF enabled, and advertised within the base OSPF instance.

From the MDT on PE-2, which has its root on PE-1, the incoming interface is an OSPF interface advertised in the base OSPF instance, as shown.

```
*A:PE-2# show router pim group 239.160.1.1 detail source 192.0.2.1

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.1.1
Source Address     : 192.0.2.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              : spt               Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.168.28.2
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:03:01
Up Time            : 0d 00:09:53       Resolved By        : rtable-u

Up JP State        : Joined            Up JP Expiry       : 0d 00:00:06
Up JP Rpt          : Not Joined StarG  Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.168.28.2
Incoming Intf      : int-PE-2-P-8
Outgoing Intf List : system

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 2                 Discarded Packets  : 0
Forwarded Octets   : 156               RPF Mismatches     : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-2#
```

The incoming interface shown is “int-PE-2-P-8”. Similarly for PE-3, the incoming interface is “int-PE-3-P-9”.

```
*A:PE-3# show router pim group 239.160.1.1 detail source 192.0.2.1

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.1.1
```

```
Source Address     : 192.0.2.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.168.39.2
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:03:27
Up Time            : 0d 00:09:56        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:03
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.168.39.2
Incoming Intf      : int-PE-3-P-9
Outgoing Intf List : system

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 6                  Discarded Packets  : 0
Forwarded Octets   : 468                RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-3#
```

VPRN Using Non-Default IGP Instance

A VPRN instance is configured on each of PE-1 and PE-2 that uses an MDT topology governed by the non-default OSPF instance.

Additional interfaces need to be configured.

```
*A:PE-1# configure
    router
        interface "int-PE-1-P-6a"
            address 192.168.116.1/30
            port 1/1/4
        exit
        interface "loop-1"
            address 192.0.3.1/32
            loopback
        exit
```

There are parallel links between PE-1 and P-6. The interface name of the second link contains the suffix **a.**

In a Rosen MVPN, each PE constructs a default MDT to all other PEs in the multicast VPN domain, as defined by the MDT SAFI BGP update received. The MDT update contains the source address of the MDT to which each PE should join.

When each of the other PEs receives the MDT SAFI network layer reachability information (NLRI), a PIM join is sent to the source address within the global PIM routing instance to create the MDT.

The MDT source address is usually the system address. Because the system address is advertised in the base instance of OSPF, another /32 address must be used as the source address for the default MDT. Therefore, a second loopback address is configured and used as the source address for the default MDT. For PE-1, this interface is called loop-1, and it is advertised in OSPF 1.

The interface loop-1 will be used as the source address for the MDTs and the next hop for the unicast route representing the source address of the c-multicast group.

The non-default OSPF instance for PE-1 is configured as follows, where 192.0.3.1 is the OSPF 1 router-ID. The router ID need not be equal to the IP address for loop-1, but in this case it is.

```
*A:PE-1# configure
    router
        ospf 1 192.0.3.1
            area 0.0.0.0
                interface "int-PE-1-P-6a"
                    interface-type point-to-point
                exit
                interface "loop-1"
                exit
            exit
```

LDP is also required for BGP next-hop resolution and is configured as follows for PE-1.

```
*A:PE-1# configure
    router
        ldp
            interface-parameters
                interface int-PE-1-P-6a
                    ipv4
                        transport-address interface
                    exit
                exit
            exit
```

The transport address is set to interface, rather than the default of system address; this is because the system address is not reachable within OSPF instance 1.

For completeness, the configuration of the additional interfaces, OSPF instance 1 and LDP of PE-2 is shown in the following three outputs.

```
*A:PE-2# configure
    router
        interface "int-PE-2-P-7a"
            address 192.168.127.1/30
            port 1/1/4
        exit
        interface "loop-1"
            address 192.0.3.2/32
            loopback
        exit
```

The following output displays the OSPF 1 instance configuration:

```
*A:PE-2# configure
    router
        ospf 1 192.0.3.2
            area 0.0.0.0
                interface "int-PE-2-P-7a"
                    interface-type point-to-point
                exit
                interface "loop-1"
                exit
            exit
            no shutdown
        exit
```

The following output displays the LDP configuration:

```
*A:PE-2# configure
    router
        ldp
            interface-parameters
                interface "int-PE-2-P-7a"
                    ipv4
                        transport-address interface
                    exit
                exit
            exit
```

PIM needs to be enabled on all interfaces.

The MDT source address for VPRN 2 is the loop-1 address. Each PE within this VPRN has to join the MDT sourced at PE-1, so the MDT SAFI NLRI must advertise the source address of the MDT group as loop-1. This is achieved by specifying the MDT SAFI source address within the MVPN context. The following output displays the VPRN configuration for PE-1.

```
*A:PE-1# configure
    service
         vprn 2 customer 1 create
            route-distinguisher 64496:2
```

```
auto-bind-tunnel
    resolution-filter
        ldp
    exit
    resolution filter
exit
vrf-target target:64496:2
interface "int-PE-1-S-2" create
    address 172.16.12.1/24
    sap 1/2/1 create
    exit
exit
pim
    apply-to all
    no shutdown
exit
mvpn
    auto-discovery mdt-safi source-address 192.0.3.1
    provider-tunnel
        inclusive
            pim ssm 239.160.2.1
            exit
        exit
    exit
    vrf-target target:64496:2
    exit
exit
no shutdown
```

The MDT SAFI source address modification is only required on PEs that use the non-default /32 addresses. The system address must not be explicitly configured as the MDT source address for MVPNs that use the default IGP instance. As previously stated, only three MVPNs can be used to create core diversity, one of which must be the default instance. Configuring the system address as a source address prevents the creation of a third MVPN because only two MVPNs are allowed to use explicitly configured MDT source addresses.

Verification of Core Diversity

The MDT SAFI NLRI advertised by the PE-1 sender router contains the following information.

```
*A:PE-1# show router bgp routes mdt-safi hunt rd 64496:2 | match "RIB Out" post-lines
25 pre-lines 1
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Dist.    : 64496:2
Source Addr    : 192.0.3.1
Group Addr     : 239.160.2.1
Nexthop        : 192.0.2.1
To             : 192.0.2.5
```

```
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:64496:2
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.5
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
*A:PE-1
```

The source address is set to 192.0.3.1, which is the address of the loopback address used in the non-default OSPF instance 1 of PE-1.

The following output shows the MDT that has its root at PE-1, and that the source address is set to 192.0.3.1. The outgoing interface list includes the router interface contained within the OSPF 1 instance, proving that the non-default OSPF instance is used.

```
*A:PE-1# show router pim group 239.160.2.1 source 192.0.3.1 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.2.1
Source Address     : 192.0.3.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : self
Keepalive Timer Exp: 0d 00:03:06
Up Time            : 0d 00:06:27        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:32
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : loop-1
Outgoing Intf List : system, int-PE-1-P-6a

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 16                 Discarded Packets  : 0
Forwarded Octets   : 1248               RPF Mismatches     : 0
```

```
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The PIM interfaces within VPRN 2 are now present on PE-1, as follows:

```
*A:PE-1# show router 2 pim interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                   Adm  Opr  DR Prty         Hello Intvl  Mcast Send
   DR
-------------------------------------------------------------------------------
int-PE-1-S-2                Up   Up   1               30           auto
   172.16.12.1
2-mt-239.160.2.1            Up   Up   1               N/A          auto
   192.0.3.2
-------------------------------------------------------------------------------
Interfaces : 2 Tunnel-Interfaces : 0
===============================================================================
*A:PE-1#
```

Likewise, for PE-2, the PIM interfaces within VPRN 2 are displayed, as follows:

```
*A:PE-2# show router 2 pim interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                   Adm  Opr  DR Prty         Hello Intvl  Mcast Send
   DR
-------------------------------------------------------------------------------
int-PE-2-H-2                Up   Up   1               30           auto
   172.16.22.1
2-mt-239.160.2.1            Up   Up   1               N/A          auto
   192.0.3.2
-------------------------------------------------------------------------------
Interfaces : 2 Tunnel-Interfaces : 0
===============================================================================
*A:PE-2#
```

Within the VPRN, there are PIM neighbors shown via the MDT. On PE-2, the PIM neighbor is 192.0.3.1, as follows:

```
*A:PE-2# show router 2 pim neighbor
===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
2-mt-239.160.2.1        1               0d 00:06:29   0d 00:01:36     105
   192.0.3.1
-------------------------------------------------------------------------------
```

```
Neighbors : 1
==============================================================================
*A:PE-2#
```

The PIM interface on PE-2 designated as 2-mt-239.160.2.1 with a neighbor address of 192.0.3.1 is the MDT interface toward PE-1.

The prefix that represents the source address on PE-1 is advertised as a VPN-IPv4 route, which contains a BGP connector attribute.

This can be shown when the VPN-IPv4 route is examined on PE-2, as follows:

```
*A:PE-2# show router bgp routes 172.16.12.0/24 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 172.16.12.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:2                VPN Label      : 262141
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-2-P-8
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : RD 64496:2, Originator 192.0.3.1
Community      : target:64496:2
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h00m26s
VPRN Imported  :  2

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

```
Routes : 1
```

The originator value within the connector attribute is shown to be 192.0.3.1, which is the same as the MDT source address of PE-1. The BGP next hop is still set to the system address of PE-1, so the unicast route can still be resolved via an LDP tunnel.

PIM will now resolve the c-source address RPF using the originator value within the connector attribute.

Similarly, for VPRN 1, the route on PE-1 representing the source address is also advertised as a VPN-IPv4 address that contains a BGP connector attribute.

```
*A:PE-2# show router bgp routes vpn-ipv4 172.16.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 172.16.11.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262134
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-2-P-8
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : RD 64496:1, Originator 192.0.2.1
Community      : target:64496:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h15m52s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
```

```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

Verification of Multicast Traffic

An IGMPv3 query is initiated from all 3 hosts: H-1, H-2, and H-3 in Figure 1, and the multicast streams from S-1 and S-2 into interfaces on the two VPRNs are enabled.

Consider VPRN 1, which uses the default topology. On PE-1, the group 239.160.1.123 can be shown. The outgoing and incoming interface lists are populated, with the outgoing interface being the MDT interface for the VPRN:

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.1.123
Source Address     : 172.16.11.2
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.11.2
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:02:15        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.11.2
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : 1-mt-239.160.1.1

Curr Fwding Rate   : 7134.2 kbps
Forwarded Packets  : 16467196           Discarded Packets  : 0
Forwarded Octets   : 757491016          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The same groups can be shown within VPRN 1 on PE-2.

```
*A:PE-2# show router 1 pim group detail
```

```
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.1.123
Source Address     : 172.16.11.2
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                     Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:14:49         Resolved By        : rtable-u

Up JP State        : Joined              Up JP Expiry       : 0d 00:00:58
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : 1-mt-239.160.1.1
Outgoing Intf List : int-PE-2-H-1

Curr Fwding Rate   : 5764.4 kbps
Forwarded Packets  : 14046700            Discarded Packets  : 0
Forwarded Octets   : 646148200           RPF Mismatches     : 0
Spt threshold      : 0 kbps              ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-2#
```

The MDT is now the incoming interface with an upstream RPF neighbor of 192.0.2.1, the system address of PE-1. Similarly for PE-3:

```
*A:PE-3# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.1.123
Source Address     : 172.16.11.2
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                     Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:14:18         Resolved By        : rtable-u

Up JP State        : Joined              Up JP Expiry       : 0d 00:00:36
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No
```

```
Rpf Neighbor       : 192.0.2.1
Incoming Intf      : 1-mt-239.160.1.1
Outgoing Intf List : int-PE-3-H-3

Curr Fwding Rate   : 7233.6 kbps
Forwarded Packets  : 4188552            Discarded Packets  : 0
Forwarded Octets   : 192673392          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-3#
```

Consider VPRN 2, which uses the non-default topology. On PE-1, the group 239.160.2.123 can be shown. The outgoing and incoming interface lists are populated, with the outgoing interface being the MDT interface for the VPRN.

```
*A:PE-1# show router 2 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.2.123
Source Address     : 172.16.12.2
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
MRIB Next Hop      : 172.16.12.2
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:00:39        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.12.2
Incoming Intf      : int-PE-1-S-2
Outgoing Intf List : 2-mt-239.160.2.1

Curr Fwding Rate   : 4571.7 kbps
Forwarded Packets  : 9960897            Discarded Packets  : 0
Forwarded Octets   : 458201262          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

The outgoing interface list is again populated with the MDT being the interface. This MDT is encapsulated in the multicast tree shown in the global PIM context as multicast group 239.160.2.1 with source address 192.0.3.1. This can be shown to have an outgoing interface list containing the interface int-PE-1-P-6a, which is an OSPF 1 interface and was shown in a previous output.

```
*A:PE-1# show router pim group detail 239.160.2.1
---snip---

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.160.2.1
Source Address     : 192.0.3.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : self
Keepalive Timer Exp: 0d 00:03:18
Up Time            : 0d 03:19:45        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:14
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : loop-1
Outgoing Intf List : system, int-PE-1-P-6a

Curr Fwding Rate   : 4713.5 kbps
Forwarded Packets  : 9842087            Discarded Packets  : 0
Forwarded Octets   : 452737186          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-1#
```

Conclusion

MVPN Core Diversity allows service providers to provide separation in terms of topology between content providers that use a core network to provide transport between source and receivers in a multicast VPN. This chapter provides the configuration for multiple instances of OSPF which, together with the associated commands and outputs, can be used for verifying and troubleshooting.

Rosen MVPN Inter-AS Option B

This chapter provides information about Rosen MVPN: Inter-AS Option B configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS 11.0.R3. The CLI in the current edition is based on SR OS release 14.0.R5.

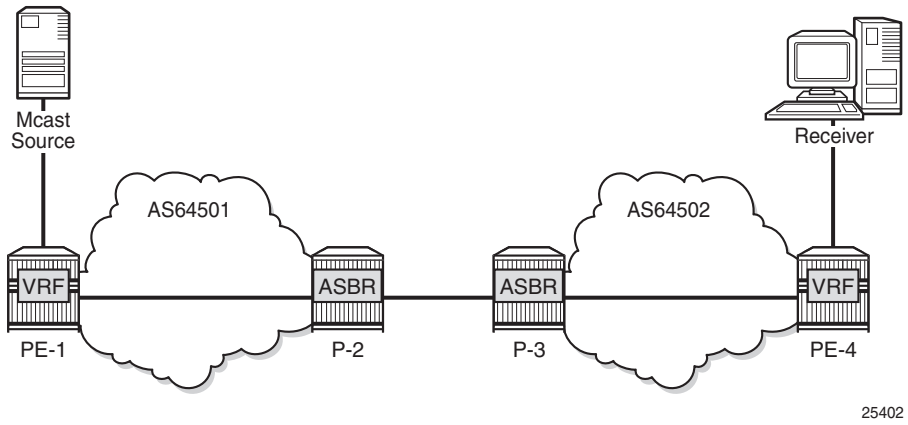
Knowledge of the Nokia multicast and Layer 3 VPNs concepts are assumed throughout this document.

Overview

This chapter covers a basic technology overview, the network topology and configuration examples which are used for multicast virtual private network (MVPN) inter-autonomous system (AS) option B.

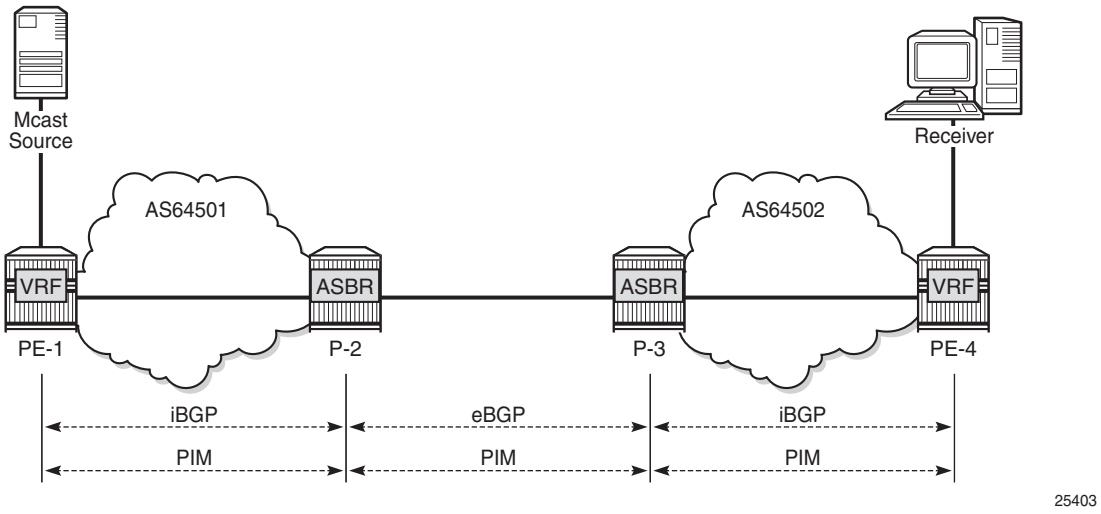
The Inter-AS MVPN feature allows the setup of multicast distribution trees (MDTs) that span multiple autonomous systems.

Figure 337 **General Topology for Inter-AS MVPN**



This example covers Rosen MVPN Inter-AS support (Option-B). Inter-AS Option B is supported for protocol independent multicast (PIM) source-specific multicast (SSM) with Rosen MVPN using multicast distribution tree (MDT) subsequent address family indicator (SAFI), using the border gateway protocol (BGP) connector attribute and PIM reverse path forwarding (RPF) vector.

Figure 338 **Protocols Used for Inter-AS MVPN**



The following assumptions are made:

- PE-1 is the sender PE because the multicast source is directly connected to this router.

- PE-4 is the receiver PE because the multicast receiver is directly connected to this router.
- P-2 and P-3 are ASBR routers according to Inter-AS model.

The multicast receiver and source can be indirectly connected to PE routers via CE routers, but for the core multicast distribution, these variations are conceptually the same. For simplicity, the PE and P router configurations will be provided.

There are several challenges which have to be solved in order to make the complete inter-AS solution operational:

Challenge 1:

In case of Inter-AS MVPN Option B, routing information toward the source PE is not available in a remote AS domain because IGP routes are not exchanged between ASs.

As a result, a PIM-P join would never be sent upstream (from the receiver PE to the sender PE in a different AS). However, the PIM-P join has to be propagated from PE-4 to PE-1. Therefore, a solution is required to issue PIM-P join and perform RPF.

Solution:

Use a PIM reverse path forwarding (RPF) vector (RPFV) to segment the PIM-P propagation. In this example there are three segments:

- PE-4 -> ASBR P-3
- ASBR P-3 -> ASBR P-2
- ASBR P-2 -> PE-1

The RPF vector is added to a PIM join at the PE router when the following option is enabled:

```
*A:PE-4# configure router pim rpfv
  - no rpfv [core] [mvpn]
  - rpfv core mvpn
  - rpfv core
  - rpfv mvpn
 <core>               : Proxy RPF vector for core
 <mvpn>               : Proxy RPF vector for inter-AS rosen mvpn
```

The keyword **mvpn** enables “mvpn RPF vector” processing for Inter-AS Option B MVPN based on RFC 5496 and RFC 6513. If a core RPF vector is received, it will be dropped before a message is processed.

All routers which are used for multicast traffic transportation must have this option enabled to allow RPF vector processing. If the option is not enabled, the RPF vector is dropped and the PIM join is processed as if the PIM vector is not present.

Details about RPF Vector can be found in the following RFCs: 5496, 5384, 6513.

Challenge 2:

With Inter-AS MVPN Option B, the BGP next-hop is modified by the local and remote ASBRs during re-advertisement of VPN IPv4 routes. When the BGP next-hop is changed, information regarding the originator of the prefix is lost when the advertisement reaches the receiver PE node. Therefore, a solution is required to do a successful RPF check for the VPN source at receiver VPRN.



Note: This challenge does not apply to Model C because in Model C the BGP next-hop for VPN routes is not updated.

Solution:

A transitive BGP attribute - connector - is used to advertise an address of a sender PE node which is carried inside a VPN IPv4 update. The BGP connector attribute allows the sender PE address information to be available to the receiver PE so that a receiver PE is able to associate VPN IPv4 advertisement to the corresponding source PE.

Inter-AS Option B will work when the following criteria are met:

- Rosen MVPN is used with PIM SSM
- BGP MDT-SAFI address family is used
- PIM RPF vector is configured
- BGP connector attribute is used for VPN-IPv4 updates

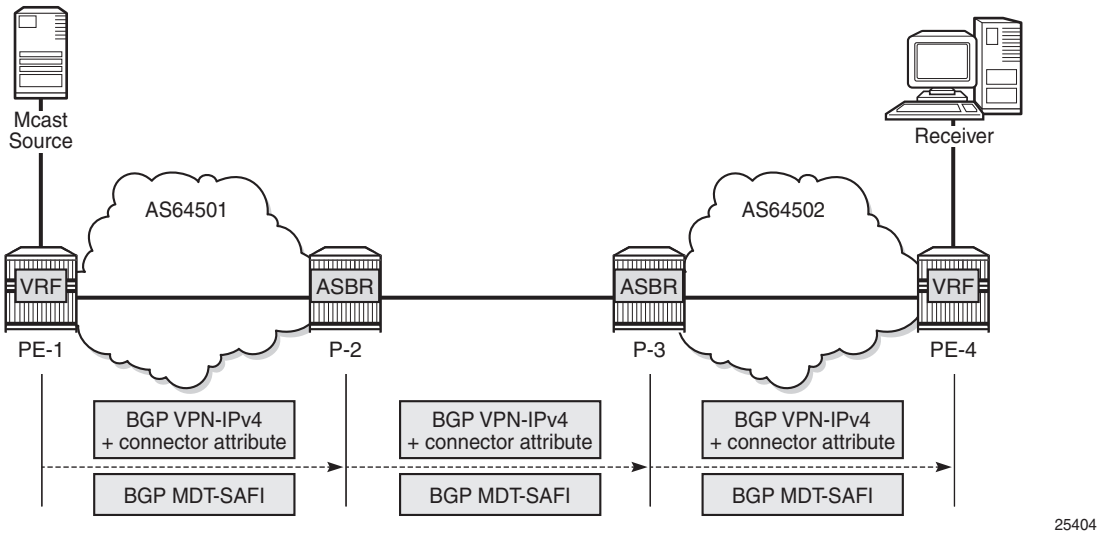
SR OS inter-AS Option B is designed to be standard compliant based on the following RFCs:

- RFC 5384, *The Protocol Independent Multicast (PIM) Join Attribute Format*
- RFC 5496, *The Reverse Path Forwarding (RPF) Vector TLV*
- RFC 6513, *Multicast in MPLS/BGP IP VPNs*

Three global signaling stages can be identified when Inter-AS MVPN is configured:

Stage 1: BGP core signaling

Figure 339 **BGP Signaling Steps**



The sender PE sends VPN-IPv4 and MDT-SAFI BGP updates for this particular MVPN:

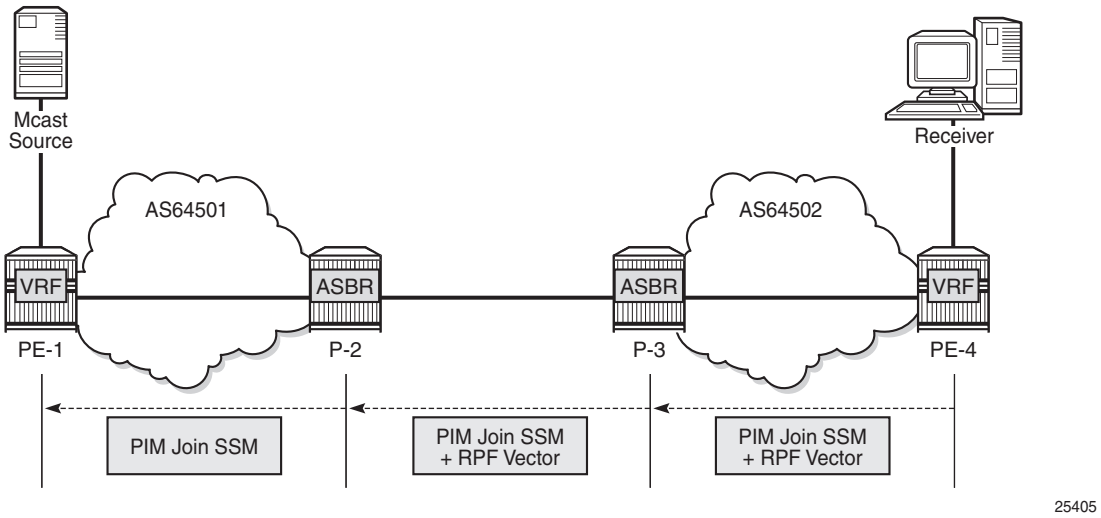
- Every ASBR propagates VPN-IPv4 and MDT-SAFI BGP updates:
  - Next hop (NH) attribute is modified every time
  - Connector attribute stays untouched

When this stage is completed, all routers have necessary information:

- to start PIM signaling in the core network (PIM-P) to prepare the default MDT
- to start PIM signaling of customer multicast streams (PIM-C) inside the VPN

Stage 2: Core PIM signaling

Figure 340 **PIM-P Signaling Steps for Default MDT**



PE-4 determines the reverse path to the source based on the RPF vector (ASBR P-3 IP address) and not based on the IP address of the multicast source (PE-1) which is unknown to PE-4.

PE-4 inserts an RPF vector and sends a PIM-P join to the immediate next-hop to reach ASBR P-3. Intermediate P-routers (if present) do not change the RPF vector.

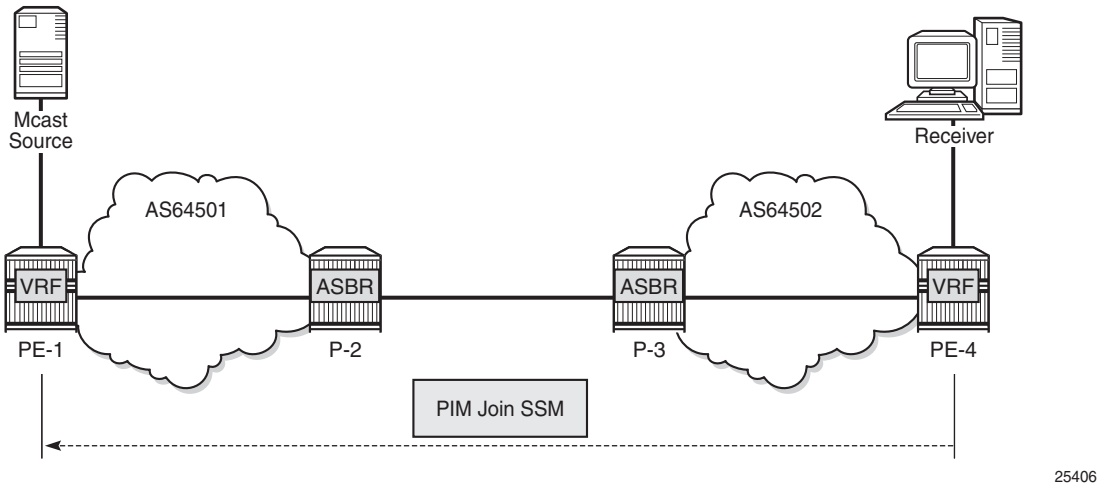
P-3 finds itself in the RPF vector and has to make a decision based on MDT-SAFI BGP table:

- P-3 determines the reverse path to the multicast source based on the RPF vector (ASBR P-2 IP address).
- If the multicast source and the NH do not match, P-3 has to use the RPFV.
- P-3 modifies the PIM-P join received from PE-4 with ASBR P-2's IP address as the upstream (taken from next hop MDT-SAFI network layer reachability information (NLRI)).
- P-2 can match the source IP with the NH in BGP MDT-SAFI. Therefore, there is no need for the RPF vector to be used.
- P-2 removes the RPF vector and sends a normal PIM-P join toward PE-1.

When this stage is completed, the default MDT is established for this MVPN and PE routers have the necessary information to start PIM signaling inside the VPRN (PIM-C).

Stage 3: Customer PIM signaling

Figure 341 **PIM-C Signaling**



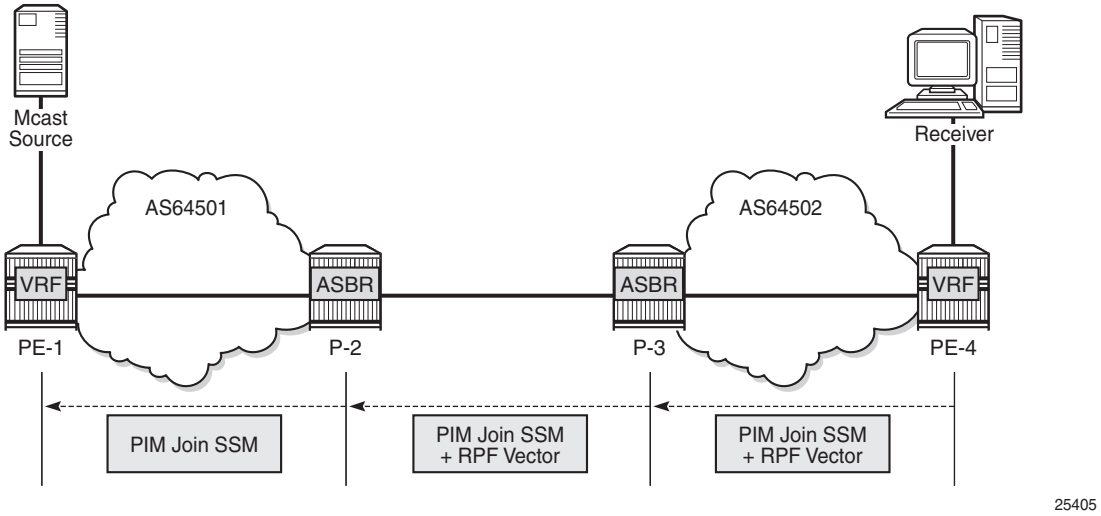
A PIM-C join is sent to the source PE using the existing tunnel infrastructure to the RPF neighbor PE-1 provided by the BGP connector attribute of the vpn-ipv4 route of the multicast source.

When this stage is completed, the customer multicast flows throughout the network in a default MDT.

Stage 4: The multicast stream threshold is reached.

Note: This stage is optional and applicable when S-PMSI instance and S-PMSI threshold are configured.

Figure 342 **PIM-P Signaling Steps for Data MDT**



The process is similar to the default MDT setup:

- PE-4 determines reverse path to the source based on the RPF vector (ASBR P-3's IP address) and not based on the IP address of the multicast source (PE-1) which is unknown to PE-4.
- PE-4 inserts an RPF vector and sends a PIM-P Join to the immediate next hop to reach ASBR P-3.
- Intermediate P-routers (if present) do not change the RPF vector.
- P-3 finds itself in the RPF vector and has to make a decision based on the MDT-SAFI BGP table:
  - P-3 determines the reverse path to the multicast source based on the RPF Vector (ASBR P-2's IP address).
  - If the multicast source and the NH do not match, P-3 has to use the RPFV.
  - P-3 modifies the PIM-P join received from PE-4 with ASBR P-2's IP address as upstream (taken from next hop MDT-SAFI NLRI).
- P-2 can match the source IP with the NH in the BGP MDT-SAFI. Therefore, there is no need for the RPF vector to be used.
- P-2 removes the RPF vector and sends a normal PIM-P join toward PE-1.

When this optional stage is completed, the customer multicast flows in a dedicated Data MDT.

Known interoperability issues:

The SR OS implementation was also designed to interoperate with Cisco routers' Inter-AS implementations that do not fully comply with the RFC 5384 and RFC 5496.

When the following option is enabled:

```
configure router pim rpfv mvpn
```

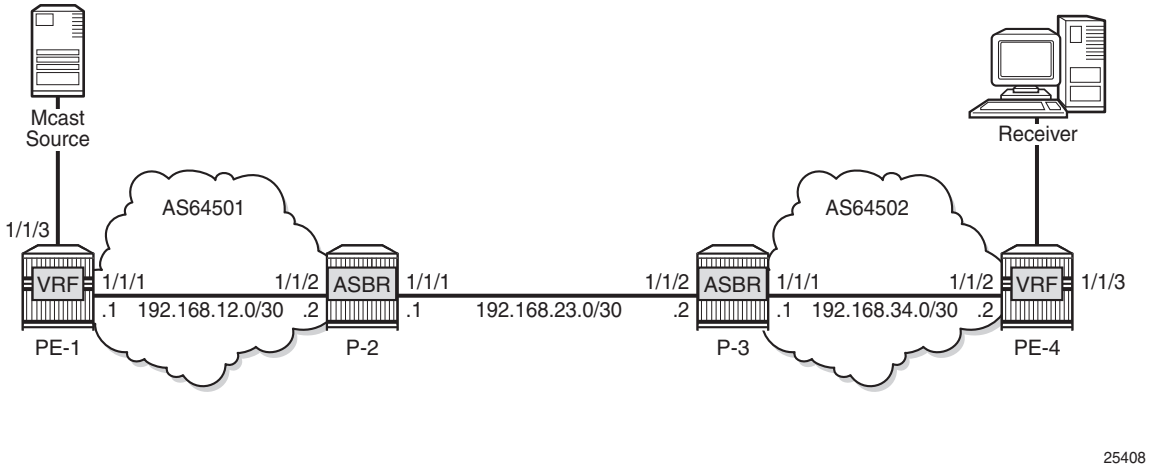
Cisco routers need to be configured to include **RD** in an RPF vector using the following command for interoperability:

```
ip multicast vrf <name> rpf proxy rd vector
```

Configuration

The example topology is shown in Figure 343.

Figure 343 **Example Topology Details**



The following parameters are used in the example scenario:

- VPRN 1 is used
- Customer multicast group is 232.0.0.0/8
- Default MDT multicast group is 239.255.0.1
- Data MDT multicast group is 239.255.1.0/24
- Multicast source is 172.16.1.1
- PE-x routers have system IP addresses 192.0.2.x

- P-x routers have system IP addresses 192.0.2.x
- Interface between Router A and B has IP address 192.168.AB.x

Global BGP configuration for PE-1 router using the mdt-safi family with an iBGP neighbor to P-2. The system interface IP address is used for the iBGP session.

```
*A:PE-1# configure router
        bgp
            group "iBGP"
                family vpn-ipv4 mdt-safi
                type internal
                neighbor 192.0.2.2
                    next-hop-self
                exit
            exit
```

The global BGP configuration for P-2 router is using the mdt-safi family with an iBGP neighbor to PE-1 and an eBGP neighbor to P-3. The system interface IP address is used for the iBGP session and the network interface IP address is used for the eBGP session.

```
*A:P-2# configure router
        bgp
            enable-inter-as-vpn
            group "eBGP"
                family vpn-ipv4 mdt-safi
                neighbor 192.168.23.2
                    type external
                    peer-as 64502
                exit
            exit
            group "iBGP"
                family vpn-ipv4 mdt-safi
                neighbor 192.0.2.1
                    next-hop-self
                    type internal
                exit
            exit
```

The global BGP configuration for the router P-3 is using the mdt-safi family with an iBGP neighbor to PE-4 and an eBGP neighbor to P-2. The system interface IP address is used for the iBGP session and the network interface IP address is used for the eBGP session.

```
*A:P-3# configure router
        bgp
            enable-inter-as-vpn
            group "eBGP"
                family vpn-ipv4 mdt-safi
                neighbor 192.168.23.1
                    type external
                    peer-as 64501
                exit
            exit
```

```
            group "iBGP"
                family vpn-ipv4 mdt-safi
                neighbor 192.0.2.4
                    next-hop-self
                    type internal
                exit
            exit
```

The global BGP configuration for router PE-4 is using the mdt-safi family with an iBGP neighbor to P-3 is as follows. The system interface IP address is used for the iBGP session.

```
*A:PE-4# configure router
        bgp
            group "iBGP"
                family vpn-ipv4 mdt-safi
                type internal
                neighbor 192.0.2.3
                    next-hop-self
                exit
            exit
```

The global PIM configuration for all routers is as follows:.

```
configure router
        pim
            rpf-table both
            apply-to non-ies
            rpfv mvpn
        exit
```

The VPRN configuration for the PE routers is as follows:

```
*A:PE-1# configure service
        vprn 1 customer 1 create
            route-distinguisher 1:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:1:1
            interface "int-PE-1-S-1" create
                address 172.16.1.2/30
                sap 1/1/3 create
                exit
            exit
            pim
                apply-to all
            exit
            mvpn
                auto-discovery mdt-safi
                provider-tunnel
                    inclusive
```

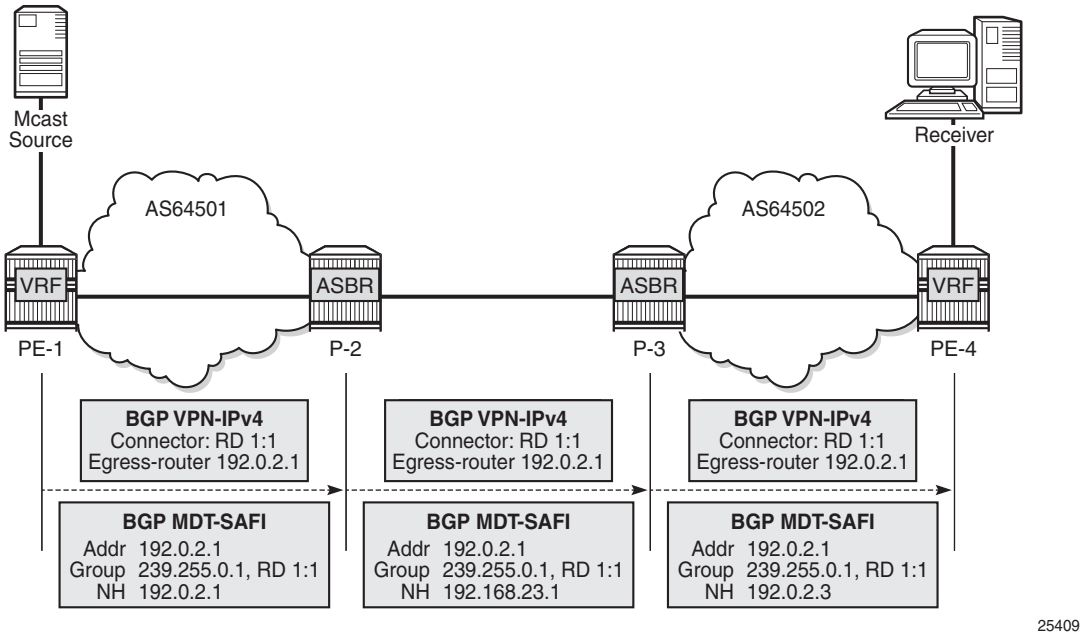
```
                        pim ssm 239.255.0.1
                        exit
                    exit
                    selective
                        data-threshold 232.0.0.0/8 1
                        pim-ssm 239.255.1.0/24
                    exit
                exit
                vrf-target unicast
                exit
            exit
            no shutdown
        exit

*A:PE-4# configure service
        vprn 1 customer 1 create
            route-distinguisher 4:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:1:1
            interface "int-PE-4-H-4" create
                address 172.16.4.1/30
                sap 1/1/3 create
                exit
            exit
            igmp
                interface "int-PE-4-H-4"
                exit
            exit
            pim
                apply-to all
            exit
            mvpn
                auto-discovery mdt-safi
                provider-tunnel
                    inclusive
                        pim ssm 239.255.0.1
                        exit
                    exit
                    selective
                        data-threshold 232.0.0.0/8 1
                        pim-ssm 239.255.1.0/24
                    exit
                exit
                vrf-target unicast
                exit
            exit
            no shutdown
        exit
```

MVPN Verification and Debugging

BGP Core Signaling

Figure 344 **BGP Signaling Steps**



On PE-1, the **debug router bgp update** output shows the BGP update messages which are sent to P-2. The VPN-IPv4 update contains a connector attribute and the MDT-SAFI update is used for signaling multicast group 239.255.0.1.

```
1 2016/10/23 08:31:51.67 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 79
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family VPN_IPV4
        NextHop len 12 NextHop 192.0.2.1
        172.16.1.0/30 RD 1:1 Label 262143
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:1:1
    Flag: 0xc0 Type: 20 Len: 14 Connector:
        RD 1:1, Egress-router 192.0.2.1
"
```

```
2 2016/10/23 08:31:51.67 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 62
    Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
        Address Family MDT-SAFI
        NextHop len 4 NextHop 192.0.2.1
        [MDT-SAFI] Addr 192.0.2.1, Group 239.255.0.1, RD 1:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:1:1
"
```

On P-2, the **debug router bgp update** output shows the BGP update messages which are sent to P-3. The VPN-IPv4 update contains an unmodified connector attribute and the MDT-SAFI update is used for signaling multicast group 239.255.0.1.

```
4 2016/10/23 08:28:16.14  UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.23.2
"Peer 1: 192.168.23.2: UPDATE
Peer 1: 192.168.23.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 78
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family VPN_IPV4
        NextHop len 12 NextHop 192.168.23.1
        172.16.1.0/30 RD 1:1 Label 262140
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:1:1
    Flag: 0xc0 Type: 20 Len: 14 Connector:
        RD 1:1, Egress-router 192.0.2.1
"


3 2016/10/23 08:28:16.14 UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.23.2
"Peer 1: 192.168.23.2: UPDATE
Peer 1: 192.168.23.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 54
    Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
        Address Family MDT-SAFI
        NextHop len 4 NextHop 192.168.23.1
        [MDT-SAFI] Addr 192.0.2.1, Group 239.255.0.1, RD 1:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:1:1
"
```

On P-3, the **debug router bgp update** output shows the BGP update messages which are sent to PE-4. The VPN-IPv4 update contains an unmodified connector attribute and the MDT-SAFI update is used for signaling multicast group 239.255.0.1.

```
9 2016/10/23 08:29:54.70 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family VPN_IPV4
        NextHop len 12 NextHop 192.0.2.3
        172.16.1.0/30 RD 1:1 Label 262140
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:1:1
    Flag: 0xc0 Type: 20 Len: 14 Connector:
        RD 1:1, Egress-router 192.0.2.1
"


11 2016/10/23 08:29:54.70 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 61
    Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
        Address Family MDT-SAFI
        NextHop len 4 NextHop 192.0.2.3
        [MDT-SAFI] Addr 192.0.2.1, Group 239.255.0.1, RD 1:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:1:1
"
```

The BGP tables on PE-1 and PE-4 are updated accordingly. The most interesting aspect here is the MDT-SAFI routes received.

PE-4 has one MDT-SAFI update received from PE-1. The next-hop was modified according to the Option-B model.

```
*A:PE-4# show router bgp neighbor 192.0.2.3 received-routes mdt-safi
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MDT-SAFI Routes
```

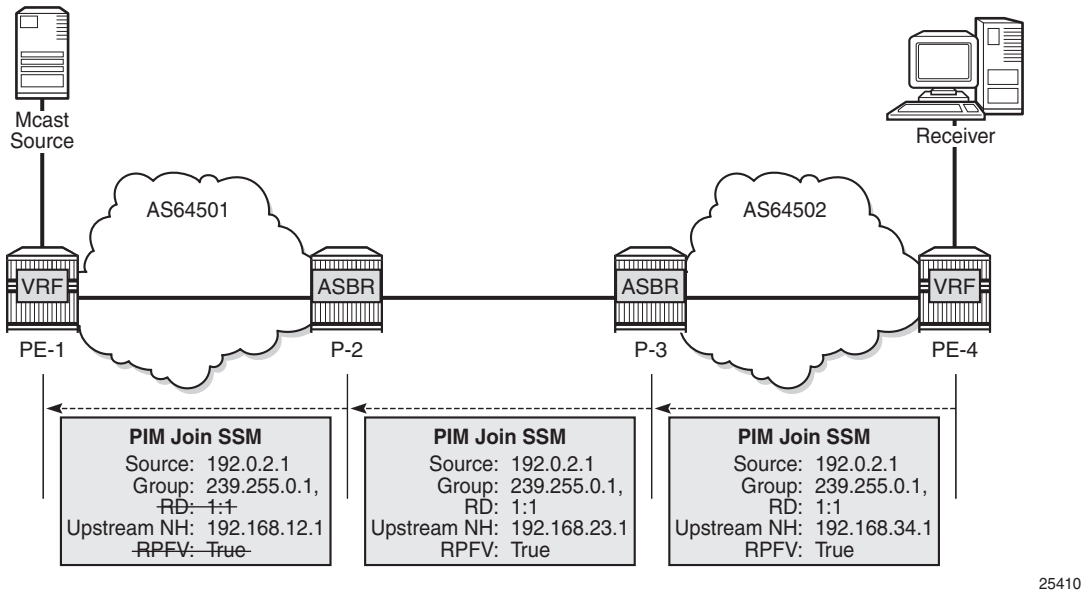
```
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop                    Group-Addr                        Label
      As-Path
-------------------------------------------------------------------------------
u*>i  1:1:192.0.2.1                                    100         None
      192.0.2.3                  239.255.0.1                       -
      64501
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-4#
```

PE-1 has one MDT-SAFI update received from PE-4. The next-hop was modified according to the Option B model.

```
*A:PE-1# show router bgp neighbor 192.0.2.2 received-routes mdt-safi
===============================================================================
 BGP Router ID:192.0.2.1        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MDT-SAFI Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop                    Group-Addr                        Label
      As-Path
-------------------------------------------------------------------------------
u*>i  4:1:192.0.2.4                                    100         None
      192.0.2.2                  239.255.0.1                       -
      64502
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

Core PIM Signaling

Figure 345 **PIM-P Signaling Steps for Default MDT**



On PE-4, the **debug router pim packet jp** output shows the PIM join/prune message which is sent to P-3. This message contains the original source of the multicast traffic (PE-1: 192.0.2.1) and the RPF Vector (P-3: 192.0.2.3).

```
10 2016/10/23 11:27:20.39 UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
[000 00:20:52.320] PIM-TX ifId 2 ifName int-PE-4-P-3 0.0.0.0 -> 224.0.0.13 Length: 48
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x3d2c
Upstream Nbr IP : 192.168.34.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 239.255.0.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Joined Srcs:
                192.0.2.1/32 Flag S  <S,G> JA={rpfvMvpn 192.0.2.3 1:1}
"
```

On P-3, **the debug router pim packet jp** output shows the PIM join/prune message which is propagated to P-2. The source of multicast traffic is untouched while the RPF Vector is modified for Inter-AS propagation.

```
24 2016/10/23 11:28:09.70  UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
[000 00:20:51.420] PIM-TX ifId 2 ifName int-P-3-P-2 0.0.0.0 -> 224.0.0.13 Length: 48
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x3286
Upstream Nbr IP : 192.168.23.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 239.255.0.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Joined Srcs:
                192.0.2.1/32 Flag S  <S,G> JA={rpfvMvpn 192.168.23.1 1:1}
"
```

On P-2, the **debug router pim packet jp** output shows the PIM join/prune message which is propagated to P-1. The source of multicast traffic is untouched while the RPF Vector is not present anymore.

```
23 2016/10/23 11:26:34.01 UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
[000 00:20:51.490] PIM-TX ifId 3 ifName int-P-2-PE-1 0.0.0.0 -> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x563f
Upstream Nbr IP : 192.168.12.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 239.255.0.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Joined Srcs:
                192.0.2.1/32 Flag S  <S,G>
"
```

As a result of this signaling, the default MDT is established between the two ASs. This can be checked with the **show router pim group** command.

The following PE-1 output shows the active multicast groups which are used as default MDT.

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (S,G)               spt      system         2
   192.0.2.1
239.255.0.1                 (S,G)               spt      int-PE-1-P-2   1
   192.0.2.4
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-1#
```

The following PE-4 output shows the active multicast groups which are used as default MDT:

```
*A:PE-4# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (S,G)               spt      int-PE-4-P-3   1
   192.0.2.1
239.255.0.1                 (S,G)               spt      system         2
   192.0.2.4
-------------------------------------------------------------------------------
```

```
Groups : 2
===============================================================================
*A:PE-4#
```

The detailed information about the PIM-P group shows that the default MDT is used to deliver traffic. Key parameters such as the incoming/outgoing interfaces and non-zero traffic counters allow this conclusion to be made.

PE-4 has the incoming interface “int-PE-4-P-3”, and outgoing interface “system”, as follows:

```
*A:PE-4# show router pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 192.0.2.1
RP Address         : 0
Advt Router        : 192.0.2.3

Upstream RPFV Nbr  : 192.168.34.1
RPFV Type          : Mvpn 1:1           RPFV Proxy         : 192.0.2.3

Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.168.34.1
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:03:13
Up Time            : 0d 00:42:26        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:33
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.168.34.1
Incoming Intf      : int-PE-4-P-3
Outgoing Intf List : system

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 86                 Discarded Packets  : 0
Forwarded Octets   : 6708               RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
===============================================================================
```

PE-1 has incoming the interface “system”, and outgoing interfaces “system, int-PE-1-P-2”, as follows:

```
*A:PE-1# show router pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 192.0.2.1
```

```
RP Address          : 0
Advt Router         : 192.0.2.1
Flags               : spt                Type               : (S,G)
Mode                : sparse
MRIB Next Hop       :
MRIB Src Flags      : self
Keepalive Timer Exp: 0d 00:03:13
Up Time             : 0d 00:42:50        Resolved By        : rtable-m

Up JP State         : Joined             Up JP Expiry       : 0d 00:00:09
Up JP Rpt           : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

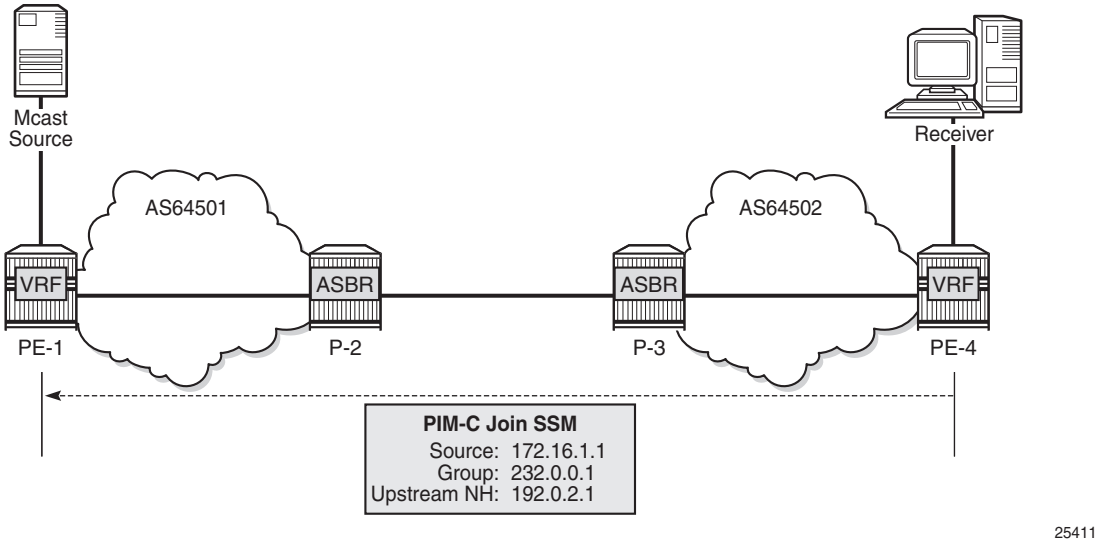
Register State      : No Info
Reg From Anycast RP: No

Rpf Neighbor        :
Incoming Intf       : system
Outgoing Intf List  : system, int-PE-1-P-2

Curr Fwding Rate    : 0.0 kbps
Forwarded Packets   : 89                 Discarded Packets  : 0
Forwarded Octets    : 6942               RPF Mismatches     : 0
Spt threshold       : 0 kbps             ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
```

Customer PIM Signaling

Figure 346 **PIM-C Signaling**



The PIM-C Join is sent to the sender PE using the existing tunnel infrastructure.

On PE-4, the **debug router 1 pim packet jp** output shows the PIM join/prune message which is sent to PE-1 using PMSI interface “1-mt-239.255.0.1” inside VPRN 1. All of this information and more can be found in the output of the **debug** command.

```
10 2016/10/23 13:27:26.46 UTC MINOR: DEBUG #2001 vprn1 PIM[Instance 2 vprn1]
"PIM[Instance 2 vprn1]: Join/Prune
[000 00:36:54.980] PIM-TX ifId 16385 ifName 1-mt-239.255.0.1 0.0.0.0 -
> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x7dd6
Upstream Nbr IP : 192.0.2.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 232.0.0.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Join Srcs:
                172.16.1.1/32 Flag S <S,G>
"
```

The detailed information about the PIM-C group for a particular VPRN shows that the default MDT is used to deliver traffic. For this purpose, the **show router 1 pim group detail** command is used. Key parameters such as the correct multicast group, correct incoming/outgoing interfaces and non-zero flow rate allow this conclusion to be made.

PE-1 has the incoming interface “int-PE-1-S-1”, and outgoing interface “1-mt-239.255.0.1”. If the threshold hasn’t been reached to set up a selective provider tunnel, only one outgoing interface is listed.

Note: In order to generate this output, the data threshold for the selective provider tunnel was temporarily raised to 100000 kbps in VPRN 1.

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.1.1
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:00:11        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.1.1
```

```
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : 1-mt-239.255.0.1

Curr Fwding Rate   : 9771.0 kbps
Forwarded Packets  : 274840            Discarded Packets  : 0
Forwarded Octets   : 12642640          RPF Mismatches     : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

PE-4 has the incoming interface “1-mt-239.255.0.1”, and outgoing interface “int-PE-4-H-4” to the receiving host. As long as there is no S-PMSI, the following output can be seen.

```
*A:PE-4# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.3
Flags              :                   Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:07:44       Resolved By        : rtable-u

Up JP State        : Joined            Up JP Expiry       : 0d 00:00:16
Up JP Rpt          : Not Joined StarG  Up JP Rpt Override : 0d 00:00:00

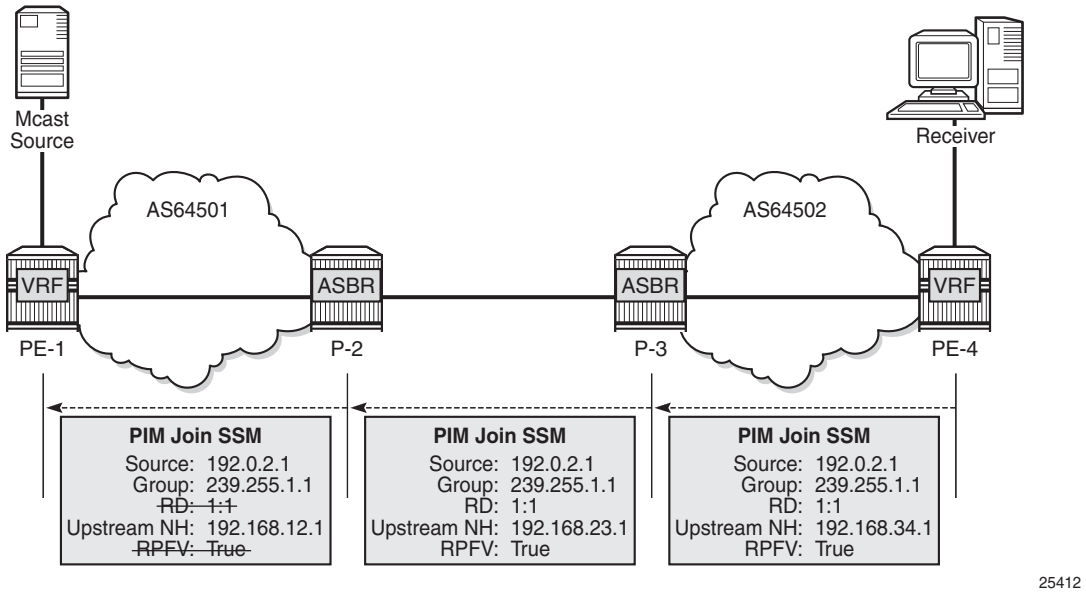
Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : 1-mt-239.255.0.1
Outgoing Intf List : int-PE-4-H-4

Curr Fwding Rate   : 8717.0 kbps
Forwarded Packets  : 11811923          Discarded Packets  : 0
Forwarded Octets   : 543348458         RPF Mismatches     : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-4#
```

When Multicast Stream Threshold is Reached

Figure 347 **PIM-P Signaling Steps for Data MDT**



On PE-4, the **debug router pim packet jp** output shows the PIM join/prune message which is sent to P-3. This message contains the original source of multicast traffic (PE-1: 192.0.2.1) and the RPF Vector (P-3: 192.0.2.3).

Note: A new multicast group (239.255.1.1) is signaled for purposes of establishing the data MDT.

```
16 2016/10/23 14:39:08.45 UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
[000 00:44:09.130] PIM-TX ifId 2 ifName int-PE-4-P-3 0.0.0.0 -> 224.0.0.13 Length: 48
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x3c2c
Upstream Nbr IP : 192.168.34.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 239.255.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Joined Srcs:
                192.0.2.1/32 Flag S  <S,G> JA={rpfvMvpn 192.0.2.3 1:1}

"
```

On P-3, the **debug router pim packet jp** output shows the PIM join/prune message which is propagated to P-2. The source of multicast traffic is untouched while the RPF Vector is modified for Inter-AS propagation.

```
71 2016/10/23 14:41:00.68 UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
```

```
[000 00:44:08.240] PIM-TX ifId 2 ifName int-P-3-P-2 0.0.0.0 -> 224.0.0.13 Length: 48
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x3186
Upstream Nbr IP : 192.168.23.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 239.255.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Joined Srcs:
                192.0.2.1/32 Flag S  <S,G> JA={rpfvMvpn 192.168.23.1 1:1}

"
```

On P-2, the **debug router pim packet jp** output shows the PIM join/prune message which is propagated to PE-1. The source of multicast traffic is untouched while the RPF Vector is not present anymore.

```
44 2016/10/23 14:43:25.13 UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
[000 00:44:08.290] PIM-TX ifId 3 ifName int-P-2-PE-1 0.0.0.0 -> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x553f
Upstream Nbr IP : 192.168.12.1 Resvd: 0x0, Num Groups 1, HoldTime 210
        Group: 239.255.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
        Joined Srcs:
                192.0.2.1/32 Flag S  <S,G>

"
```

As a result of this signaling, the Data MDT is established between the two ASs. This can be checked with **show router pim group** command.

The PE-1 output shows an additional multicast group (239.255.1.1), which was created in the global routing table (GRT).

```
*A:PE-1# show router pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (S,G)               spt      system         2
   192.0.2.1
239.255.0.1                 (S,G)               spt      int-PE-1-P-2   1
   192.0.2.4
239.255.1.1                 (S,G)                        system         1
   192.0.2.1
-------------------------------------------------------------------------------
Groups : 3
===============================================================================
*A:PE-1#
```

The PE-4 output shows an additional multicast group (239.255.1.1), which was created in the GRT.

```
*A:PE-4# show router pim group
===============================================================================
```

```
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (S,G)               spt      int-PE-4-P-3   1
   192.0.2.1
239.255.0.1                 (S,G)               spt      system         2
   192.0.2.4
239.255.1.1                 (S,G)                        int-PE-4-P-3   1
   192.0.2.1
-------------------------------------------------------------------------------
Groups : 3
===============================================================================
*A:PE-4#
```

The detailed information about the PIM group in a VPRN shows that the data MDT is used to receive traffic instead of the default MDT.

The PE-4 output for multicast groups in a VPRN 1 has slightly changed: a new line “Incoming SPMSI Intf” was added. This indicates that the S-PMSI instance and dedicated Data MDT are used for this particular multicast group. The non-zero rate for the multicast flow is also an indication that multicast traffic is forwarded.

```
*A:PE-4# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.3
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:08:37        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:22
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : 1-mt-239.255.0.1
Incoming SPMSI Intf: 1-mt-239.255.0.1*
Outgoing Intf List : int-PE-4-H-4

Curr Fwding Rate   : 9811.2 kbps
Forwarded Packets  : 13260102           Discarded Packets  : 0
Forwarded Octets   : 609964692          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
```

```
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-4#
```

The **show router 1 pim s-pmsi detail** command can also be used to verify existence of the S-PMSI instance for the VPRN 1. The output includes the multicast group inside the VPRN, the multicast source IP, the multicast group which is used for S-PMSI tunneling and the current forwarding rate.

```
*A:PE-4# show router 1 pim s-pmsi detail
===============================================================================
PIM Selective provider tunnels
===============================================================================
Md Source Address  : 192.0.2.1          Md Group Address   : 239.255.1.1
Number of VPN SGs  : 1                  Uptime             : 0d 00:10:58
MT IfIndex         : 24589              Egress Fwding Rate : 8629.2 kbps

VPN Group Address  : 232.0.0.1
VPN Source Address : 172.16.1.1
State              : RX Joined
Expiry Timer       : 0d 00:01:55
===============================================================================
PIM Selective provider tunnels Interfaces : 1
===============================================================================
*A:PE-4#
```

Conclusion

Inter-AS MVPN offers flexibility for the operators who can use it to provide additional value added services to their customers. Before implementing this feature in the network the following are required:

- The RPF vector must be enabled on every router for inter-AS MVPN.
- Can be used only with Rosen MVPN with PIM SSM and MDT SAFI.

NG-MVPN Sender-Only, Receiver-Only

This chapter provides information about next generation multicast virtual private network (NG-MVPN) sender-only and receiver-only configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The sender-only/receiver-only feature as described in this chapter is supported in SR OS release 11.0.R1, and later. The CLI in this edition is based on SR OS release 14.0.R5.

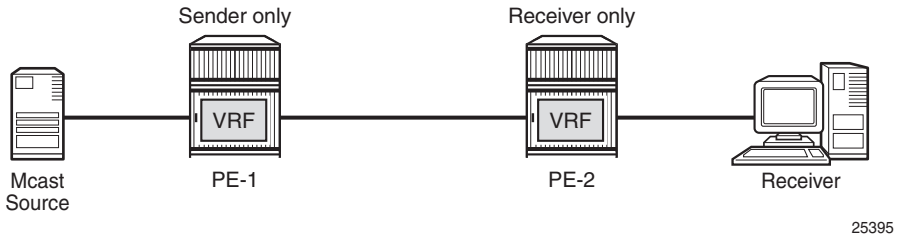
Knowledge of the Nokia multicast and Layer 3 VPNs concepts are assumed throughout this document.

Overview

This example covers a basic technology overview, the network topology, and configuration examples which are used for the Multicast VPN (MVPN) sender-only, receiver-only feature.

By default, if multiple PE nodes form a peering relationship within a common MVPN instance, then each PE node originates a multicast tree locally towards the remaining PE nodes that are members of this MVPN instance. This behavior creates a full mesh of Inclusive-Provider Multicast Service Interfaces (I-PMSIs) across all PE nodes in the MVPN.

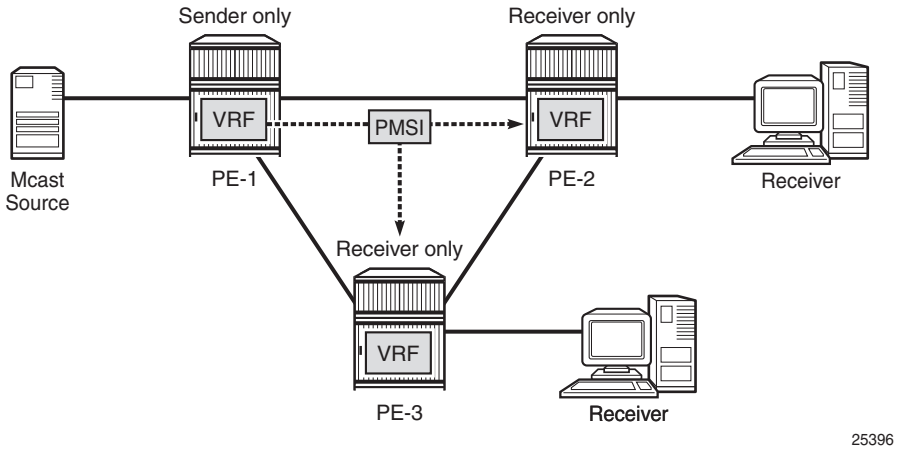
Figure 348 **Default PMSI Hierarchy**



It is often the case that an MVPN has many sites with multicast receivers, but only a few sites that host either both receivers and sources, or sources only.

The MVPN sender-only/receiver-only feature optimizes control and data plane resources by preventing unnecessary I-PMSI meshing when a PE hosts multicast sources only, or multicast receivers only, for an MVPN. An example of such an optimization is presented in Figure 349.

Figure 349 **Optimized PMSI Structure**



The general rules to follow are:

- For PE nodes that host only multicast sources for a given MVPN, operators can now block these PEs, through configuration, from joining I-PMSIs from the other PEs in this MVPN.
- For PE nodes that host only multicast receivers for a given MVPN, operators can now block these PEs, through configuration, to set-up a local I-PMSI to the other PEs in this MVPN.

MVPN sender-only/receiver-only is supported with next generation-MVPN for both IPv4 and IPv6 customer multicast using:

- IPv4 RSVP-TE provider tunnels
- IPv4 LDP provider tunnels

Extra attention should be given to the Bootstrap Router/Rendezvous Point (BSR/RP) placement when sender-only/receiver-only is enabled:

- The RP should be at the sender-receiver or sender-only site so that (*,G) traffic can be sent over the tunnel

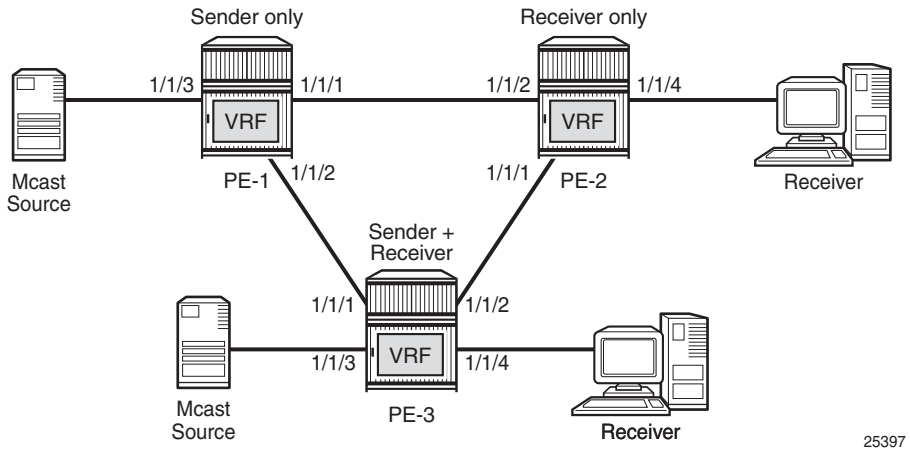
Note: (*,G) refers to an individual multicast stream indicating any source (*) and the multicast group (G) used by the stream.

- The BSR should be deployed at the sender-receiver site.
- The BSR can be at a sender-only site if the RPs are at the same site.

Configuration

The example topology is shown in Figure 350.

Figure 350 **Example Topology**



To configure the sender-only/receiver-only feature, the following configuration command is used:

```
*A:PE>config>service>vprn>mvpn# mdt-type
  - mdt-type {sender-only|receiver-only|sender-receiver}
  - no mdt-type
```

sender-receiver is the default option and is visible using the **info detail** command.

This command restricts the MVPN instance to a specific role and provides an option to configure either a sender-only or receiver-only mode per PE node per service.

Parameters:

sender-only — MVPN has only senders connected to PE node.

receiver-only — MVPN has only receivers connected to PE node.

sender-receiver — MVPN has both sender and receivers connected to PE node.

Considerations:

- Two general approaches for building MVPNs will be covered in detail in this example:
  - Point-to-multipoint (P2MP) RSVP MVPNs
  - Multicast LDP (mLDP) MVPNs
- IPv4 and IPv6 multicast streaming are used for every MVPN at the same time.
- Basic principles of an MVPN including I-PMSI, S-PMSI, mLDP and P2MP RSVP are covered in the NG-MVPN Configuration with PIM and chapters of this guide.

PIM SSM is used for IPv4/IPv6 Customer (C)-multicast groups.

Initial Configuration

Step 1. The PE routers already have the following configuration:
- Interfaces (IPv4/IPv6)
- IGP (IS-IS or OSPF/OSPFv3)
- LDP (IPv4 only suffices)
- MPLS/RSVP
- BGP

Step 2. The MPLS/RSVP configuration on PE-1 is as follows. An P2MP LSP template is created with an empty path, without explicit hops.

```
*A:PE-1# configure router
        mpls
            interface system
```

```
            exit
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-3"
            exit
            no shutdown
        exit
        rsvp no shutdown
        mpls
            path EMPTY
                no shutdown
            exit
            lsp-template MVPN-P2MP-LSP p2mp
                default-path EMPTY
                cspf
                no shutdown
            exit
        exit
```

Step 3. The BGP configuration on PE-1 is as follows. There is no route reflector used.

```
*A:PE-1# configure router
        bgp
            min-route-advertisement 1
            rapid-withdrawal
            rapid-update mvpn-ipv4 mvpn-ipv6
            enable-peer-tracking
            group INTERNAL
                family vpn-ipv4 vpn-ipv6 mvpn-ipv4 mvpn-ipv6
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

RSVP-Based MVPN Configuration

Step 1. Configure a basic MVPN using P2MP RSVP as a transport protocol for C-multicast groups. For this setup, PE-2 and PE-3 are configured to receive the following multicast groups:

- IPv4 group 232.0.0.1 source 172.16.1.1
- IPv6 group FF3E::8000:1 source 2001:DB8:1::1

Step 2. Configure the MDT type for the MVPN.

Based on the example topology, PE-1 is configured as **sender-only** for the MVPN.

```
*A:PE-1# configure service
        vprn 1 customer 1 create
```

```
        description "RSVP-based MVPN"
        ecmp 2
        autonomous-system 64500
        route-distinguisher 64500:101
        auto-bind-tunnel
            resolution-filter
                ldp
                rsvp
            exit
            resolution filter
        exit
        vrf-target target:64500:1
        interface "int-PE-1-S-1" create
            description "to multicast source"
            address 172.16.1.2/30
            ipv6
                address 2001:db8:1::2/126
            exit
            sap 1/1/3 create
            exit
        exit
        pim
            no ipv6-multicast-disable
            apply-to all
        exit
        mvpn
            auto-discovery default
            c-mcast-signaling bgp
            mdt-type sender-only
            provider-tunnel
                inclusive
                    rsvp
                        lsp-template "MVPN-P2MP-LSP"
                        no shutdown
                    exit
                exit
            exit
            vrf-target unicast
            exit
        exit
        service-name "RSVP-based MVPN"
        no shutdown
```

Based on the example topology, PE-2 is configured as **receiver-only** for the MVPN. PE-2 has static joins for the IPv4 and IPv6 multicast groups:

- group 232.0.0.1,source 172.16.1.1
- group FF3E::8000:1, source 2001:DB8:1::1

```
*A:PE-2# configure service
       vprn 1 customer 1 create
            description "RSVP-based MVPN"
            ecmp 2
            autonomous-system 64500
            route-distinguisher 64500:102
            ignore-nh-metric
            auto-bind-tunnel
                resolution-filter
```

```
        ldp
        rsvp
    exit
    resolution filter
exit
vrf-target target:64500:1
interface "int-PE-2-H-2" create
    description "to receiver Host-2"
    address 172.16.2.2/30
    ipv6
        address 2001:db8:2::2/126
    exit
    sap 1/1/4 create
    exit
exit
igmp
    interface "int-PE-2-H-2"
        static
            group 232.0.0.1
                source 172.16.1.1
            exit
        exit
        no shutdown
    exit
    no shutdown
exit
mld
    interface "int-PE-2-H-2"
        static
            group ff3e::8000:1
                source 2001:db8:1::1
            exit
        exit
        no shutdown
    exit
    no shutdown
exit
pim
    no ipv6-multicast-disable
exit
mvpn
    auto-discovery default
    c-mcast-signaling bgp
    mdt-type receiver-only
    provider-tunnel
        inclusive
            rsvp
                lsp-template "MVPN-P2MP-LSP"
                no shutdown
            exit
        exit
    exit
    vrf-target unicast
    exit
exit
service-name "RSVP-based MVPN"
no shutdown
```

Based on the example topology, PE-3 is configured as **sender-receiver** (default) for the MVPN. PE-1 has also static joins for the IPv4 and IPv6 multicast groups:

- group 232.0.0.1,source 172.16.1.1
- group FF3E::8000:1, source 2001:DB8:1::1

The interface to the local source for PE-3 is not configured in this example. PE-3 acts as a receiver, not as a sender. Nonetheless, it is configured as sender-receiver and that has its consequences for the I-PMSIs that will be established.

```
*A:PE-3# configure service
        vprn 1 customer 1 create
            description "RSVP-based MVPN"
            ecmp 2
            autonomous-system 64500
            route-distinguisher 64500:103
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:64500:1
            interface "int-PE-3-H-3" create
                description "to receiver Host-3"
                address 172.16.3.2/30
                ipv6
                    address 2001:db8:3::2/126
                exit
                sap 1/1/4 create
                exit
            exit
            igmp
                interface "int-PE-3-H-3"
                    static
                        group 232.0.0.1
                            source 172.16.1.1
                        exit
                    exit
                    no shutdown
                exit
                no shutdown
            exit
            mld
                interface "int-PE-3-H-3"
                    static
                        group ff3e::8000:1
                            source 2001:db8:1::1
                        exit
                    exit
                    no shutdown
                exit
                no shutdown
            exit
            pim
```

```
        no ipv6-multicast-disable
        apply-to all
    exit
    mvpn
        auto-discovery default
        c-mcast-signaling bgp
        provider-tunnel
            inclusive
                rsvp
                    lsp-template "MVPN-P2MP-LSP"
                    no shutdown
                exit
            exit
        exit
        vrf-target unicast
        exit
    exit
    service-name "RSVP-based MVPN"
    no shutdown
```

The PIM instance must be **shutdown** before the mdt-type is modified; this leads to a multicast service disruption. Trying to change the mdt-type with PIM instance active will result in the following message being displayed.

```
*A:PE-1# configure service vprn 1 mvpn mdt-type receiver-only
MINOR: PIM #1100 PIM instance must be shutdown before changing this configuration
```

RSVP-Based MVPN Verification and Debugging

MDT-Type Verification

The status of the MVPN can be checked using the **show>router <***service-number***> mvpn** command:

PE-1 output:

```
*A:PE-1# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                   auto-discovery     : Default
UMH Selection      : Highest-Ip            SA withdrawn       : Disabled
intersite-shared   : Enabled               Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.1:2

ipmsi              : rsvp MVPN-P2MP-LSP
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : MVPN-P2MP-LSP-1-73728
```

```
enable-bfd-root    : false                enable-bfd-leaf    : false
Mdt-type           : sender-only

BSR signaling      : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
```

PE-2 output:

```
*A:PE-2# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.2:2

ipmsi              : rsvp MVPN-P2MP-LSP
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-virt-if-1005857
enable-bfd-root    : false                enable-bfd-leaf    : false
Mdt-type           : receiver-only

BSR signaling      : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
```

PE-3 output:

```
*A:PE-3# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.3:2

ipmsi              : rsvp MVPN-P2MP-LSP
i-pmsi P2MP AdmSt  : Up
```

```
i-pmsi Tunnel Name : MVPN-P2MP-LSP-1-73728
enable-bfd-root    : false                 enable-bfd-leaf    : false
Mdt-type           : sender-receiver

BSR signaling      : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

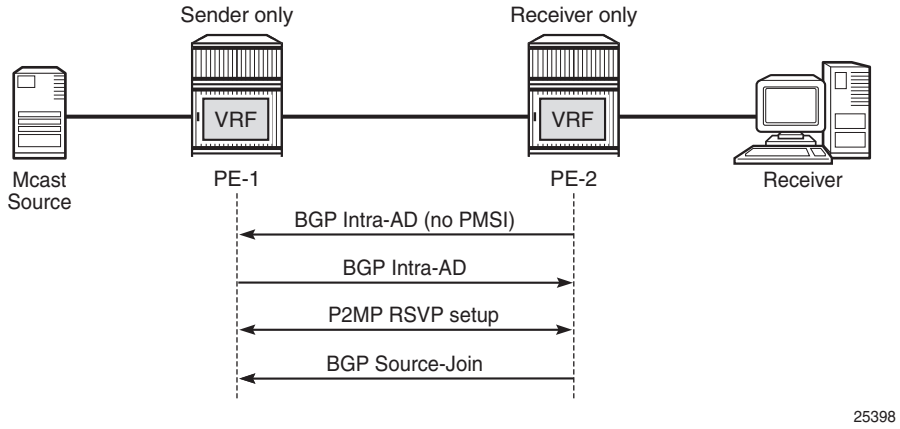
===============================================================================
```

BGP Verification and Debugging

When the MDT type is changed, the BGP signaling is slightly modified in order to achieve the signaling optimization.

The PE router does not include the PMSI part in the Intra-AD BGP messages when the MVPN is configured with mdt-type as **receiver-only**. The message flow is presented in Figure 351.

Figure 351 **RSVP-Based BGP Message Flow Between PE-1 and PE-2**



The following BGP debug output is taken from PE-2 and demonstrates the message flow between PE-1 and PE-2 for the MVPN-IPv4 address family.

There is no PMSI part in the BGP Intra-AD message sent by PE-2 (message 7), but the PMSI part is present in the BGP Intra-AD message received from **sender-only** PE-1 (message 1).

```
1 2016/10/11 11:23:52.88 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
```

```
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        P2MP-ID 0x1, Tunnel-ID: 61441, Extended-Tunnel-ID 192.0.2.1
"

7 2016/10/11 11:24:04.41 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 66
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Intra-AD Len: 12 RD: 64500:102 Orig: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
"

20 2016/10/11 11:24:06.85 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Source-Join Len:22 RD: 64500:101 SrcAS: 64500
                          Src: 172.16.1.1 Grp: 232.0.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
```

```
        target:192.0.2.1:2
"
```

Similar behavior is observed for IPv6 multicast.The following BGP debug output is taken from PE-2 and demonstrates the message flow between PE-1 and PE-2 for the MVPN-IPv6 address family.

There is no PMSI part in the Intra-AD message sent by PE-2 (message 8).

```
2 2016/10/11 11:23:52.88 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        P2MP-ID 0x1, Tunnel-ID: 61441, Extended-Tunnel-ID 192.0.2.1
"

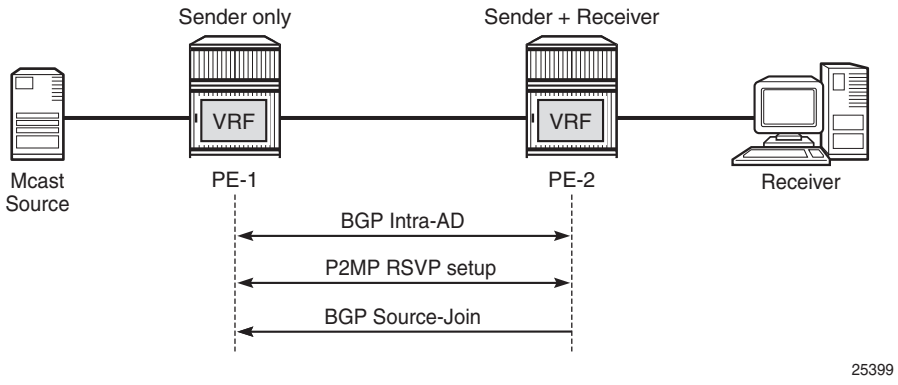

8 2016/10/11 11:24:04.41 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 66
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.2
        Type: Intra-AD Len: 12 RD: 64500:102 Orig: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
"


20 2015/10/09 09:25:15.75 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
```

```
    Total Path Attr Length = 100
    Flag: 0x90 Type: 14 Len: 57 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.2
        Type: Source-Join Len:46 RD: 64500:101 SrcAS: 64500
                         Src: 2001:db8:1::1 Grp: ff3e::8000:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The PE router does not change its BGP behavior when the MVPN is configured with mdt-type as **sender-only**. The message flow is presented in Figure 352.

Figure 352 **RSVP-Based BGP Message Flow Between PE-1 and PE-3**



The BGP following debug output is taken from PE-3 and demonstrates the message flow between PE-1 and PE-3 for the MVPN-IPv4 address family.

The PMSI part is present in debug message 1, which is sent by PE-1 (**sender-only**).

```
1 2016/10/11 11:23:52.85 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
```

```
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        P2MP-ID 0x1, Tunnel-ID: 61441, Extended-Tunnel-ID 192.0.2.1
"


13 2016/10/11 11:24:20.69 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.3
        Type: Intra-AD Len: 12 RD: 64500:103 Orig: 192.0.2.3
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        P2MP-ID 0x1, Tunnel-ID: 61441, Extended-Tunnel-ID 192.0.2.3
"


29 2016/10/11 11:24:22.61 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.3
        Type: Source-Join Len:22 RD: 64500:101 SrcAS: 64500
                          Src: 172.16.1.1 Grp: 232.0.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
```

Similar behavior is observed for IPv6 multicast.

The following BGP debug output is taken from PE-3 and demonstrates the message flow between PE-1 and PE-3 for the MVPN-IPv6 address family.

The PMSI part is present in debug message 4, which is sent by PE-1 (**sender-only**).

```
2 2016/10/11 11:23:52.85 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        P2MP-ID 0x1, Tunnel-ID: 61441, Extended-Tunnel-ID 192.0.2.1
"


14 2016/10/11 11:24:20.69 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.3
        Type: Intra-AD Len: 12 RD: 64500:103 Orig: 192.0.2.3
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 17 PMSI:
        Tunnel-type RSVP-TE P2MP LSP (1)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        P2MP-ID 0x1, Tunnel-ID: 61441, Extended-Tunnel-ID 192.0.2.3
"


30 2016/10/11 11:24:23.06 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
```

```
    Total Path Attr Length = 100
    Flag: 0x90 Type: 14 Len: 57 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.3
        Type: Source-Join Len:46 RD: 64500:101 SrcAS: 64500
                         Src: 2001:db8:1::1 Grp: ff3e::8000:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The BGP routing table of each router is populated accordingly.

PE-1 (**sender-only**) has two Intra-Ad and two Source-Join messages from PE-2 and PE-3.

```
*A:PE-1# show router bgp routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                     192.0.2.2              100         0
      64500:101                    -                                  -
      192.0.2.2                    -
      No As-Path                   -
u*>i  Source-Join                  -                      100         0
      64500:101                    64500                              -
      192.0.2.2                    172.16.1.1
      No As-Path                   232.0.0.1
*>i   Source-Join                  -                      100         0
      64500:102                    64500                              -
      192.0.2.3                    172.16.1.1
      No As-Path                   232.0.0.1
u*>i  Intra-Ad                     192.0.2.3              100         0
      64500:103                    -                                  -
      192.0.2.3                    -
      No As-Path                   -
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

PE-2 (receiver-only) has two Intra-Ad messages from PE-1 and PE-3.

```
*A:PE-2# show router bgp routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                    192.0.2.1              100         0
      64500:101                   -                                  -
      192.0.2.1                   -
      No As-Path                  -
u*>i  Intra-Ad                    192.0.2.3              100         0
      64500:103                   -                                  -
      192.0.2.3                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PE-3 (**sender-receiver**) has two Intra-Ad messages: one from PE-1 and one from PE-2.

```
*A:PE-3# show router bgp routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                    192.0.2.1              100         0
      64500:101                   -                                  -
      192.0.2.1                   -
      No As-Path                  -
u*>i  Intra-Ad                    192.0.2.2              100         0
      64500:102                   -                                  -
      192.0.2.2                   -
      No As-Path                  -
```

```
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

RSVP Verification and Debugging

When BGP intra-AD messages are exchanged, every PE starts to build multicast tunnels based on the following criteria:

- PE nodes which are configured as **sender-only** for an MVPN do not join P2MP LSPs from other PEs in this MVPN.
- PE nodes which are configured as receiver-only for an MVPN do not originate P2MP LSPs to other PEs in this MVPN.

The RSVP session can be checked with the **show>router>rsvp>session** command:

PE-1 (192.0.2.1) has two originating LSPs: one toward PE-2 (192.0.2.2) and one toward PE-3 (192.0.2.3). PE-1 also has one incoming LSP from PE-3 (**mdt-type sender-receiver**).

```
*A:PE-1# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.2       61441  60416 MVPN-P2MP-LSP-1-73728::EMPTY Up
192.0.2.1       192.0.2.3       61441  60416 MVPN-P2MP-LSP-1-73728::EMPTY Up
192.0.2.3       192.0.2.1       61441  59904 MVPN-P2MP-LSP-1-73728::EMPTY Up
-------------------------------------------------------------------------------
Sessions : 3
```

PE-2 (192.0.2.2) has two incoming LSPs from PE-1 (192.0.2.1) and PE-3 (192.0.2.3) and no originating LSPs due to the fact that PE-2 has **mdt-type receiver-only**.

```
*A:PE-2# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.2       61441  60416 MVPN-P2MP-LSP-1-73728::EMPTY Up
192.0.2.3       192.0.2.2       61441  59904 MVPN-P2MP-LSP-1-73728::EMPTY Up
-------------------------------------------------------------------------------
Sessions : 2
```

PE-3 (192.0.2.3) has two originating LSPs: one toward PE-2 (192.0.2.2) and one toward PE-1 (192.0.2.1). PE-3 also has one incoming LSP from PE-1 (**mdt-type sender-only**).

Theoretically there is no need for the LSP from PE-3 toward PE-1, because PE-1 is a sender-only; this minor limitation should be taken into account during planning phase.

```
*A:PE-3# show router rsvp session
===============================================================================
RSVP Sessions
===============================================================================
From            To              Tunnel LSP   Name                          State
                                ID     ID
-------------------------------------------------------------------------------
192.0.2.1       192.0.2.3       61441  60416 MVPN-P2MP-LSP-1-73728::EMPTY Up
192.0.2.3       192.0.2.2       61441  59904 MVPN-P2MP-LSP-1-73728::EMPTY Up
192.0.2.3       192.0.2.1       61441  59904 MVPN-P2MP-LSP-1-73728::EMPTY Up
-------------------------------------------------------------------------------
Sessions : 3
```

Additional details about originating P2MP paths can be found using the following command:

show>router>mpls p2mp-lsp <*lsp name***> p2mp-instance <***service number***> s2l**

PE-1 output:

```
*A:PE-1# show router mpls p2mp-lsp "MVPN-P2MP-LSP-1-73728" p2mp-instance "1" s2l
===============================================================================
MPLS LSP MVPN-P2MP-LSP-1-73728 S2L
===============================================================================
-------------------------------------------------------------------------------
LSP Name         : MVPN-P2MP-LSP-1-737* P2MP ID              : 1
Adm State        : Up                   Oper State           : Up
P2MPInstance     : 1                    Inst-type            : Primary
Adm State        : Up                   Oper State           : Up
-------------------------------------------------------------------------------
S2l Name                          To              Next Hop        Adm  Opr
-------------------------------------------------------------------------------
EMPTY                             192.0.2.2       192.168.12.2    Up   Up
EMPTY                             192.0.2.3       192.168.13.2    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-1#
```

PE-2 output:

```
*A:PE-2# show router mpls p2mp-lsp
===============================================================================
MPLS P2MP LSPs (Originating)
===============================================================================
LSP Name                                            Tun     Fastfail  Adm  Opr
                                                    Id      Config
```

```
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
```

PE-3 output:

```
*A:PE-3# show router mpls p2mp-lsp "MVPN-P2MP-LSP-1-73728" p2mp-instance "1" s2l
===============================================================================
MPLS LSP MVPN-P2MP-LSP-1-73728 S2L
===============================================================================
-------------------------------------------------------------------------------
LSP Name         : MVPN-P2MP-LSP-1-737* P2MP ID              : 1
Adm State        : Up                   Oper State           : Up
P2MPInstance     : 1                    Inst-type            : Primary
Adm State        : Up                   Oper State           : Up
-------------------------------------------------------------------------------
S2l Name                          To              Next Hop        Adm  Opr
-------------------------------------------------------------------------------
EMPTY                             192.0.2.1       192.168.13.1    Up   Up
EMPTY                             192.0.2.2       192.168.23.1    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

Multicast Stream Verification

The status of the multicast groups/streams can be verified using the **show>router <*sid*>>pim group detail ipv6** command:

There is an IPv4 sender connected to PE-1. The physical interface where the sender is connected is used as the incoming interface. An I-PMSI is used as the outgoing interface.

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                      Type              : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.1.1
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:13:26          Resolved By       : rtable-u

Up JP State        : Joined               Up JP Expiry      : 0d 00:00:00
Up JP Rpt          : Not Joined StarG     Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No
```

```
Rpf Neighbor       : 172.16.1.1
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73728

Curr Fwding Rate   : 4571.1 kbps
Forwarded Packets  : 1456847            Discarded Packets  : 0
Forwarded Octets   : 67014962           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
```

There is an IPv4 receiver connected to PE-2. An I-PMSI is used as the incoming interface and the physical interface where the receiver is connected is used as the outgoing interface.

```
*A:PE-2# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
---snip---

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73728
Outgoing Intf List : int-PE-2-H-2
---snip---
```

There is an IPv4 receiver connected to PE-3. An I-PMSI is used as the incoming interface and the physical interface where receiver is connected is used as the outgoing interface.

```
*A:PE-3# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
---snip---
Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73729
Outgoing Intf List : int-PE-3-H-3
---snip---
```

Similar behavior is observed for IPv6 multicast.

An IPv6 sender is connected to PE-1. The physical interface where the sender is connected is used as the incoming interface. An I-PMSI is used as the outgoing interface.

```
*A:PE-1# show router 1 pim group detail ipv6
===============================================================================
PIM Source Group ipv6
===============================================================================
Group Address      : ff3e::8000:1
Source Address     : 2001:db8:1::1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                     Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 2001:db8:1::1
---snip---
Rpf Neighbor       : 2001:db8:1::1
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73728
---snip---
```

An IPv6 receiver is connected to PE-2. An I-PMSI is used as the incoming interface and the physical interface where the receiver is connected is used as the outgoing interface.

```
*A:PE-2# show router 1 pim group detail ipv6
===============================================================================
PIM Source Group ipv6
===============================================================================
Group Address      : ff3e::8000:1
Source Address     : 2001:db8:1::1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                     Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
---snip---
Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73728
Outgoing Intf List : int-PE-2-H-2
---snip---
```

An IPv6 receiver is connected to PE-3. An I-PMSI is used as the incoming interface and the physical interface where the receiver is connected is used as the outgoing interface.

```
*A:PE-3# show router 1 pim group detail ipv6
===============================================================================
PIM Source Group ipv6
===============================================================================
Group Address      : ff3e::8000:1
Source Address     : 2001:db8:1::1
RP Address         : 0
```

```
Advt Router        : 192.0.2.1
Flags              :                    Type              : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
---snip---
Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73729
Outgoing Intf List : int-PE-3-H-3
---snip---
```

mLDP-Based MVPN Configuration

Step 1. Reconfigure VPRN 1 to make it mLDP-based. The resolution-filter should only be LDP (no RSVP anymore) for auto-bind-tunnel. The MVPN context also changes: the inclusive provider-tunnel is mLDP-based. The MDT-type remains the same: PE-1 is sender-only, PE-2 is receiver-only and PE-3 is sender-receiver (default).

PE-2 and PE-3 have static joins for the IPv4/IPv6 multicast groups:

- group 232.0.0.1,source 172.16.1.1
- group FF3E::8000:1, source 2001:DB8:1::1

Step 2. The VPRN 1 configuration on PE-1 is as follows:

```
*A:PE-1# configure service
        vprn 1 customer 1 create
            description "mLDP-based MVPN"
            ecmp 2
            autonomous-system 64500
            route-distinguisher 64500:101
            ignore-nh-metric
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:1
            interface "int-PE-1-S-1" create
                description "to multicast source"
                address 172.16.1.2/30
                ipv6
                    address 2001:db8:1::2/126
                exit
                sap 1/1/3 create
                exit
            exit
            pim
                no ipv6-multicast-disable
                apply-to all
            exit
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
```

```
                mdt-type sender-only
                provider-tunnel
                    inclusive
                        mldp
                            no shutdown
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
            service-name "mLDP-based MVPN"
            no shutdown
        exit
```

Based on the example topology, PE-2 is configured as receiver-only for the MVPN. PE-2 has also static joins for the IPv4 and IPv6 multicast groups:

- group 232.0.0.1,source 172.16.3.1
- group FF3E::8000:1, source 2001:DB8:3::1

```
*A:PE-2# configure service
        vprn 1 customer 1 create
            description "mLDP-based MVPN"
            ecmp 2
            autonomous-system 64500
            route-distinguisher 64500:102
            ignore-nh-metric
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:1
            interface "int-PE-2-H-2" create
                description "to receiver Host-2"
                address 172.16.2.2/30
                ipv6
                    address 2001:db8:2::2/126
                exit
                sap 1/1/4 create
                exit
            exit
            igmp
                interface "int-PE-2-H-2"
                    static
                        group 232.0.0.1
                            source 172.16.1.1
                        exit
                    exit
                    no shutdown
                exit
                no shutdown
            exit
            mld
                interface "int-PE-2-H-2"
                    static
```

```
                        group ff3e::8000:1 source 2001:db8:1::1
                    exit
                    no shutdown
                exit
                no shutdown
            exit
            pim
                no ipv6-multicast-disable
                apply-to all
            exit
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                mdt-type receiver-only
                provider-tunnel
                    inclusive
                        mldp
                            no shutdown
                        exit
                    exit
                exit
                vrf-target unicast
                exit
            exit
            service-name "mLDP-based MVPN"
            no shutdown
        exit
```

Based on the example topology, PE-3 is configured as **sender-receiver** (default) for the MVPN. PE-3 has also static joins for the IPv4 and IPv6 multicast groups:

- group 232.0.0.1,source 172.16.3.1
- group FF3E::8000:1, source 2001:DB8:3::1

```
*A:PE-3# configure service
        vprn 1 customer 1 create
            description "mLDP-based MVPN"
            ecmp 2
            autonomous-system 64500
            route-distinguisher 64500:103
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64500:1
            interface "int-PE-3-H-3" create
                description "to receiver Host-3"
                address 172.16.3.2/30
                ipv6
                    address 2001:db8:3::2/126
                exit
                sap 1/1/4 create
                exit
            exit
            igmp
```

```
            interface "int-PE-3-H-3"
                static
                    group 232.0.0.1
                        source 172.16.1.1
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
        mld
            interface "int-PE-3-H-3"
                static
                    group ff3e::8000:1
                        source 2001:db8:1::1
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
        pim
            no ipv6-multicast-disable
            apply-to all
        exit
        mvpn
            auto-discovery default
            c-mcast-signaling bgp
            provider-tunnel
                inclusive
                    mldp
                        no shutdown
                    exit
                exit
            exit
            vrf-target unicast
            exit
        exit
        service-name "mLDP-based MVPN"
        no shutdown
    exit
```

mLDP-Based MVPN Verification and Debugging

MDT-Type Verification

The status of the MVPN can be checked using the following command:

```
show router <service-number> mvpn
```

PE-1 output:

```
*A:PE-1# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.1:2

ipmsi              : ldp
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-if-73729
Mdt-type           : sender-only

BSR signaling      : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
```

PE-2 output:

```
*A:PE-2# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.2:2

ipmsi              : ldp
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-virt-if-1005859
Mdt-type           : receiver-only

BSR signaling      : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
```

PE-3 output:

```
*A:PE-3# show router 1 mvpn
```

```
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling           : Bgp                   auto-discovery     : Default
UMH Selection       : Highest-Ip            SA withdrawn       : Disabled
intersite-shared    : Enabled               Persist SA         : Disabled
vrf-import          : N/A
vrf-export          : N/A
vrf-target          : unicast
C-Mcast Import RT   : target:192.0.2.3:2

ipmsi               : ldp
i-pmsi P2MP AdmSt   : Up
i-pmsi Tunnel Name  : mpls-if-73730
Mdt-type            : sender-receiver

BSR signaling     : none
Wildcard s-pmsi     : Disabled
Multistream-SPMSI   : Disabled
s-pmsi              : none
data-delay-interval: 3 seconds
enable-asm-mdt      : N/A

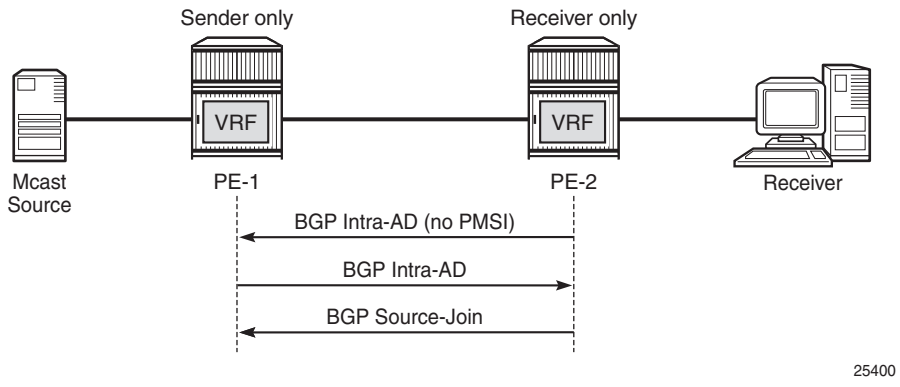
===============================================================================
*A:PE-3#
```

BGP Verification and Debugging

When the MDT type is changed, the BGP signaling is slightly modified in order to achieve the signaling optimization.The PE router does not include the PMSI part in Intra-AD BGP messages when the MVPN is configured with mdt-type as **receiver-only**.

The message flow is presented in Figure 353.

Figure 353 **mLDP-Based BGP Message Flow Between PE-1 and PE-2**



In order to demonstrate the BGP message flow sequence the following initialization steps are taken on PE-2:

1. Bring down the VPRN service, PIM protocol in a VPRN and IGMP/MLD protocol. As a result, the state of all signaling protocols is cleared.
2. Bring up the VPRN service. BGP exchanges unicast routing information.
3. Bring up the IPv4 PIM protocol. BGP exchanges IPv4 multicast routing information in order to build the PMSI infrastructure.
4. Bring up IGMP and add a static IGMP join where it is applicable. BGP exchanges IPv4 multicast routing information in order to propagate the multicast traffic to the receiver.
5. Bring up the IPv6 PIM protocol. BGP exchanges IPv6 multicast routing information in order to build the PMSI infrastructure.
6. Bring up MLD and add a static MLD join where it is applicable. BGP exchanges IPv6 multicast routing information in order to propagate the multicast traffic to the receiver.

The following BGP debug is taken from PE-2 and demonstrates the message flow between PE-2 and PE-1. VPN-IPv4 and VPN-IPv6 updates are not present in this output.

Step 1. Bring down the VPRN service and protocols to clear the state of all signaling protocols.

```
*A:PE-2# configure service
        vprn 1
            shutdown
            pim shutdown
            pim ipv6-multicast-disable
            igmp shutdown
            mld shutdown
        exit
```

Step 2. Enable the VPRN service on PE-2. PE-2 immediately receives Intra-AD messages from PE-1 because the remote VPRN service is already enabled for IPv4 and IPv6 multicast propagation.

```
*A:PE-2# configure service vprn 1 no shutdown


16 2016/10/11 12:40:09.93 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
```

```
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        Root-Node 192.0.2.1, LSP-ID 0x2001
"


13 2016/10/11 12:40:09.94 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        Root-Node 192.0.2.1, LSP-ID 0x2001
"
```

Step 3. Enable only PIM IPv4 for the service on PE-2. PE-2 immediately sends Intra-AD messages to PE-1. Note: there is no PMSI part present in the debug message sent by receiver-only PE-2.

```
*A:PE-2# configure service vprn 1 pim no shutdown


16 2016/10/11 12:40:13.49 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 66
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Intra-AD Len: 12 RD: 64500:102 Orig: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
```

```
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
"
```

Step 4. Bring up IGMP and add a static IGMP join for the service on a PE-2. PE-2 immediately sends a source-join message to PE-3 and receives a source-AD message from PE-1.

```
*A:PE-2# configure service
        vprn 1
            igmp
                interface "int-PE-2-H-2"
                    static
                        group 232.0.0.1 source 172.16.1.1
                    exit
                    no shutdown
                exit
                no shutdown
            exit


18 2016/10/11 12:40:13.49 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Source-Join Len:22 RD: 64500:101 SrcAS: 64500
                          Src: 172.16.1.1 Grp: 232.0.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

Step 5. Enable PIM IPv6 for the service on PE-2. PE-2 immediately sends Intra-AD messages to PE-3.

```
*A:PE-2# configure service vprn 1 pim no ipv6-multicast-disable


20 2016/10/11 12:41:37.99 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 66
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.2
        Type: Intra-AD Len: 12 RD: 64500:102 Orig: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
```

```
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
"
```

Step 6. Bring up MLD and add a static MLD join for the service on a PE-2. PE-2 immediately sends a source-join message to PE-3 and receives a source-AD message from PE-3.

```
*A:PE-2# configure service
        vprn 1
            mld
                interface "int-PE-2-H-2"
                    static
                        group ff3e::8000:1 source 2001:db8:1::1
                    exit
                    no shutdown
                exit
                no shutdown
            exit


22 2016/10/11 12:41:46.60 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 100
    Flag: 0x90 Type: 14 Len: 57 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.2
        Type: Source-Join Len:46 RD: 64500:101 SrcAS: 64500
                          Src: 2001:db8:1::1 Grp: ff3e::8000:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The same information can be gathered using the following show commands.

show>router>bgp>neighbor <*peer*> advertised-routes [mvpn-ipv4|mvpn-ipv6]

show>router>bgp>neighbor <*peer*> received-routes [mvpn-ipv4|mvpn-ipv6]

PE-2 output for the advertised routes for the mvpn-ipv4 address family is as follows:

```
*A:PE-2# show router bgp neighbor 192.0.2.1 advertised-routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
```

```
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
i     Source-Join                 -                      100         0
      64500:101                   64500                              -
      192.0.2.2                   172.16.1.1
      No As-Path                  232.0.0.1
i     Intra-Ad                    192.0.2.2              100         0
      64500:102                   -                                  -
      192.0.2.2                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-2#
```

PE-2 output for the advertised routers for the mvpn-ipv6 address family is as follows:

```
*A:PE-2# show router bgp neighbor 192.0.2.1 advertised-routes mvpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv6 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
i     Source-Join                 -                      100         0
      64500:101                   64500                              -
      192.0.2.2                   2001:db8:1::1
      No As-Path                  ff3e::8000:1
i     Intra-Ad                    192.0.2.2              100         0
      64500:102                   -                                  -
      192.0.2.2                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-2#
```

PE-2 output for the received routes for the mvpn-ipv4 address family is as follows:

```
*A:PE-2# show router bgp neighbor 192.0.2.1 received-routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                     192.0.2.1              100         0
      64500:101                    -                                  -
      192.0.2.1                    -
      No As-Path                   -
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

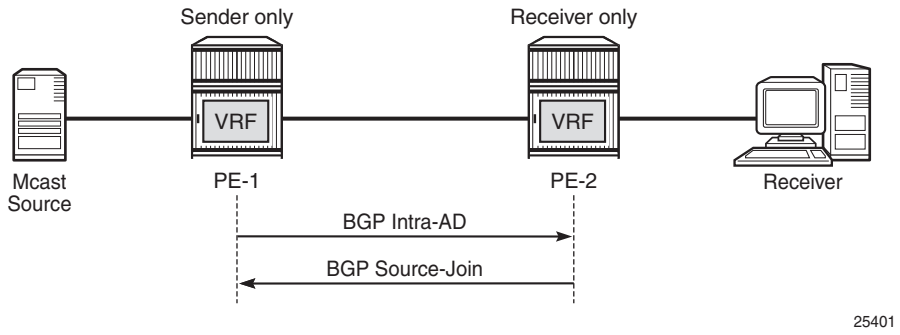
PE-2 output for the received routes for the mvpn-ipv6 address family is as follows:

```
*A:PE-2# show router bgp neighbor 192.0.2.1 received-routes mvpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv6 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                     192.0.2.1              100         0
      64500:101                    -                                  -
      192.0.2.1                    -
      No As-Path                   -
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

The PE router does not change the BGP behavior when the MVPN is configured with mdt-type as **sender-only**. A schematic of the message flow is presented in Figure 354.

Figure 354 **mLDP-Based BGP Message Flow Between PE-1 and PE-3**



In order to demonstrate the BGP message flow sequence, the following initialization steps are taken:

1. Bring down the VPRN service, PIM protocol in the VPRN and IGMP/MLD protocol. As a result, the state of all signaling protocols is cleared.
2. Bring up the VPRN service. BGP exchanges unicast routing information.
3. Bring up the IPv4 PIM protocol. BGP exchanges IPv4 multicast routing information in order to build the PMSI infrastructure.
4. Bring up IGMP and add a static IGMP join where it is applicable. BGP exchanges IPv4 multicast routing information in order to propagate the multicast traffic to the receiver.
5. Bring up the IPv6 PIM protocol. BGP exchanges IPv6 multicast routing information in order to build the PMSI infrastructure.
6. Bring up MLD and add a static MLD join where it is applicable. BGP exchanges IPv6 multicast routing information in order to propagate the multicast traffic to the receiver.

The following BGP debug output is taken from PE-3 and demonstrates the message flow between PE-1 and PE-3.

The PMSI part is present in debug messages sent by PE-1 (**sender-only**).

Step 1. Bring down the VPRN service and protocols to clear the state of all signaling protocols.

```
*A:PE-3# configure service
        vprn 1
            shutdown
```

```
            pim shutdown
            pim ipv6-multicast-disable
            igmp shutdown
            mld shutdown
        exit
```

Step 2. Enable the VPRN service on PE-3. PE-3 immediately receives Intra-AD messages from PE-1 because the remote VPRN service is already enabled for IPv4 and IPv6 multicast propagation. The PMSI attribute is present in both messages.

```
*A:PE-3# configure service vprn 1 no shutdown


6 2016/10/11 14:17:53.63 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        Root-Node 192.0.2.1, LSP-ID 0x2001
"


5 2016/10/11 14:17:53.63 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.1
        Type: Intra-AD Len: 12 RD: 64500:101 Orig: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
```

```
        MPLS Label 0
        Root-Node 192.0.2.1, LSP-ID 0x2001
"
```

Step 3. Enable PIM IPv4 only for the service on PE-3. PE-3 immediately sends Intra-AD messages to PE-1.

```
*A:PE-3# configure service vprn 1 pim no shutdown


9 2016/10/11 14:17:54.78 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.3
        Type: Intra-AD Len: 12 RD: 64500:103 Orig: 192.0.2.3
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        Root-Node 192.0.2.3, LSP-ID 0x2001
"
```

Step 4. Bring up IGMP and add a static IGMP join for the service on a PE-3. PE-3 immediately sends a source-join message to PE-1 and receives a source-AD message from PE-1.

```
*A:PE-3# configure service vprn 1 igmp no shutdown


*A:PE-3# configure service vprn 1 igmp interface "int-PE-3-H-3"
static group 232.0.0.1 source 172.16.1.1


12 2016/10/11 14:17:57.80 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.3
        Type: Source-
Join Len:22 RD: 64500:101 SrcAS: 64500 Src: 172.16.1.1 Grp: 232.0.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
```

```
     Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

Step 5. Enable PIM IPv6 for the service on PE-3. PE-3 immediately sends Intra-AD messages to PE-1.

```
*A:PE-3# configure service vprn 1 pim no ipv6-multicast-disable


14 2016/10/11 14:17:57.82 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.3
        Type: Intra-AD Len: 12 RD: 64500:103 Orig: 192.0.2.3
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 22 PMSI:
        Tunnel-type LDP P2MP LSP (2)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf not required]
        MPLS Label 0
        Root-Node 192.0.2.3, LSP-ID 0x2001
"
```

Step 6. Bring up MLD and add a static MLD join for the service on a PE-3. PE-3 immediately sends a source-join message to PE-1.

```
*A:PE-3# configure service vprn 1 mld no shutdown


*A:PE-3# configure service vprn 1 mld interface "int-PE-3-H-3" static
group ff3e::8000:1 source 2001:db8:1::1


15 2016/10/11 14:18:00.92 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 100
    Flag: 0x90 Type: 14 Len: 57 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV6
        NextHop len 4 NextHop 192.0.2.3
        Type: Source-Join Len:46 RD: 64500:101 SrcAS: 64500
                          Src: 2001:db8:1::1 Grp: ff3e::8000:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
```

```
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The same information can be gathered using the following show commands.

show router bgp neighbor <*peer*> advertised-routes [**mvpn-ipv4**|**mvpn-ipv6**]

show router bgp neighbor <*peer*> received-routes [**mvpn-ipv4**|**mvpn-ipv6**]

PE-3 output for the advertised routes for the mvpn-ipv4 address family is as follows:

```
*A:PE-3# show router bgp neighbor 192.0.2.1 advertised-routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
i     Source-Join                 -                      100         0
      64500:101                   64500                              -
      192.0.2.3                   172.16.1.1
      No As-Path                  232.0.0.1
i     Intra-Ad                    192.0.2.3              100         0
      64500:103                   -                                  -
      192.0.2.3                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 2
===============================================================================v
```

PE-3 output for the advertised routes for the mvpn-ipv6 address family is as follows:

```
*A:PE-3# show router bgp neighbor 192.0.2.1 advertised-routes mvpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv6 Routes
```

```
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
i     Source-Join                 -                      100         0
      64500:101                   64500                              -
      192.0.2.3                   2001:db8:1::1
      No As-Path                  ff3e::8000:1
i     Intra-Ad                    192.0.2.3              100         0
      64500:103                   -                                  -
      192.0.2.3                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PE-3 output for the received routes for the mvpn-ipv4 address family is as follows:

```
*A:PE-3# show router bgp neighbor 192.0.2.1 received-routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                    192.0.2.1              100         0
      64500:101                   -                                  -
      192.0.2.1                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-3 output for the received routes for the mvpn-ipv6 address family is as follows:

```
*A:PE-3# show router bgp neighbor 192.0.2.1 received-routes mvpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
```

```
BGP MVPN-IPv6 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
u*>i  Intra-Ad                    192.0.2.1              100         0
      64500:101                   -                                  -
      192.0.2.1                   -
      No As-Path                  -
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

LDP Verification and Debugging

When BGP intra-AD messages are exchanged, every PE starts to build a multicast tunnel based on the following criteria:

PE nodes which are configured as **sender-only** do not distribute mLDP forward equivalence classes (FECs) to remote PEs for this MVPN.

PE nodes which are configured as receiver-only do not include the PMSI part for intra-AD messages and remote PEs do not send mLDP FECs for this MVPN.

LDP bindings can be verified using the following command:

show router ldp bindings p2mp

PE-1 (192.0.2.1) has two egress FECs due to the fact that PE-1 has the mdt-type sender-only.

```
*A:PE-1# show router ldp bindings p2mp ipv4
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID 2001:db8::1)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
P2MP-Id
RootAddr                                  Interface       IngLbl    EgrLbl
EgrNH                                     EgrIf/LspId
Peer
-------------------------------------------------------------------------------
8193
192.0.2.1                                 73729             --      262135
192.168.12.2                              1/1/1
```

```
192.0.2.2:0

8193
192.0.2.1                                73729             --      262133
192.168.13.2                             1/1/2
192.0.2.3:0

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Bindings: 2
===============================================================================
---snip---
```

PE-2 (192.0.2.2) has two ingress FECs due to the fact that PE-2 has mdt-type receiver-only.

```
*A:PE-2# show router ldp bindings p2mp ipv4 opaque-type generic

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID 2001:db8::2)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
P2MP-Id
RootAddr                                 Interface       IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
Peer
-------------------------------------------------------------------------------
8193
192.0.2.1                                73736           262135U     --
  --                                       --
192.0.2.1:0

8193
192.0.2.3                                73735           262136U     --
  --                                       --
192.0.2.3:0

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Bindings: 2
===============================================================================
```

PE-3 (192.0.2.3) has one ingress FEC and one egress FECs due to the fact that PE-3 has the default mdt-type sender-receiver. There is only an egress FEC to PE-2 (receiver-only), but not to PE-1. PE-1 can never be a receiver, since it is configured as sender-only.

```
*A:PE-3# show router ldp bindings p2mp ipv4 opaque-type generic

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID 2001:db8::3)
```

```
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
P2MP-Id
RootAddr                                    Interface       IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
Peer
-------------------------------------------------------------------------------
8193
192.0.2.1                                   73737           262133U     --
  --                                          --
192.0.2.1:0

8193
192.0.2.3                                   73736             --      262136
192.168.23.1                                1/1/2
192.0.2.2:0

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Bindings: 2
===============================================================================
```

Multicast Stream Verification

The status of a multicast group/stream can be verified using the following command:

show router <*sid*> pim group detail [**ipv6**]

An IPv4 sender is connected to PE-1. The physical interface where the source is connected is used as incoming interface and the I-PMSI is used as outgoing interface.

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.1.1
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 02:58:29        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00
```

```
Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.1.1
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73729

Curr Fwding Rate   : 4774.2 kbps
Forwarded Packets  : 138754864          Discarded Packets  : 0
Forwarded Octets   : 6382723744         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

There is an IPv4 receiver connected to PE-2. The I-PMSI is used as incoming interface and the physical interface where the receiver is connected is used as outgoing.

```
*A:PE-2# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:40:00        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:01:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73736
Outgoing Intf List : int-PE-2-H-2

Curr Fwding Rate   : 4741.9 kbps
Forwarded Packets  : 29507636           Discarded Packets  : 0
Forwarded Octets   : 1357351256         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

There is IPv4 receiver connected to PE-3. The I-PMSI is used as incoming interface and the physical interface where the receiver is connected is used as outgoing.

```
*A:PE-3# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.0.0.1
Source Address     : 172.16.1.1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 02:32:09        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:50
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73737
Outgoing Intf List : int-PE-3-H-3

Curr Fwding Rate   : 4787.9 kbps
Forwarded Packets  : 117212483          Discarded Packets  : 0
Forwarded Octets   : 5391774218         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

Similar behavior is observed for IPv6 multicast.

There is an IPv6 sender connected to PE-1. The physical interface where the sender is connected is used as the incoming interface and the I-PMSI is used as the outgoing interface.

```
*A:PE-1# show router 1 pim group detail ipv6
===============================================================================
PIM Source Group ipv6
===============================================================================
Group Address      : ff3e::8000:1
Source Address     : 2001:db8:1::1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 2001:db8:1::1
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 03:04:47        Resolved By        : rtable6-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:12
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00
```

```
Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 2001:db8:1::1
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73729

---snip---
```

There is an IPv6 receiver connected to PE-2. An I-PMSI is used as the incoming interface and the physical interface where the receiver is connected is used as the outgoing interface.

```
*A:PE-2# show router 1 pim group detail ipv6
===============================================================================
PIM Source Group ipv6
===============================================================================
Group Address      : ff3e::8000:1
Source Address     : 2001:db8:1::1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 02:06:22        Resolved By        : rtable6-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:38
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73736
Outgoing Intf List : int-PE-2-H-2
---snip---
```

There is an IPv6 receiver connected to PE-3. An I-PMSI is used as the incoming interface and the physical interface where the receiver is connected is used as the outgoing interface.

```
*A:PE-3# show router 1 pim group detail ipv6
===============================================================================
PIM Source Group ipv6
===============================================================================
Group Address      : ff3e::8000:1
Source Address     : 2001:db8:1::1
RP Address         : 0
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
```

```
Keepalive Timer    : Not Running
Up Time            : 0d 02:31:00        Resolved By        : rtable6-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:01:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73737
Outgoing Intf List : int-PE-3-H-3

---snip---
```

Conclusion

The sender-only/receiver-only feature provides significant signaling optimization in the core network for RSVP and LDP protocols and is recommended to be used when such functionality is required. The following are required before implementing this feature in the network:

- MDT-types **sender-only**, **receiver-only** and **sender-receiver** are enabled per MVPN.
- The default mdt-type is **sender-receiver** mode for backward compatibility.
- This is purely a control plane feature and there are no hardware dependencies, except for requiring chassis mode C or later.
- Rosen MPVN or MDT-SAFI based MVPNs are not supported.
- IPv4 and IPv6 C-signaling are supported.
- mLDP and RSVP demonstrate slightly different behavior due to the nature of each protocol.
- mLDP provides a better optimization than RSVP in all cases, as mLDP does not initiate LSPs to sender-only routers.

NG-MVPN Wildcard S-PMSI

This chapter provides information about next generation multicast virtual private networks (NG-MVPNs): use of wildcard selective PMSI.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The chapter was initially written based on SR OS release 13.0.R4, but the CLI in the current edition is based on release 14.0.R5.

Chassis mode C or D is required for MPLS provider tunnels using multicast label distribution protocol (mLDP) or point-to-multipoint (P2MP) resource reservation protocol with traffic engineering (RSVP-TE) label switched paths (LSPs). SR OS release 12.0.R4 or later is required for route reflectors (RRs) peering with multicast virtual private network (MVPN) PEs.

Provider multicast service interfaces (PMSIs) are signaled using mLDP. PE MVPN auto-discovery uses BGP MVPN IPv4 network layer routing.

Knowledge of multi-protocol BGP (MP-BGP), RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, RFC 6513, *Multicast in MPLS/BGP IP VPNs*/RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs*, and RFC 6625, *Wildcards in Multicast VPN Auto-Discovery Routes*, is assumed throughout this chapter.

Overview

Consider a service provider core network used to deliver multicast services to a number of receiver PEs using Next Generation MVPN techniques, as defined in RFC 6513/6514, where multicast traffic is forwarded between PEs across a mesh of provider tunnels.

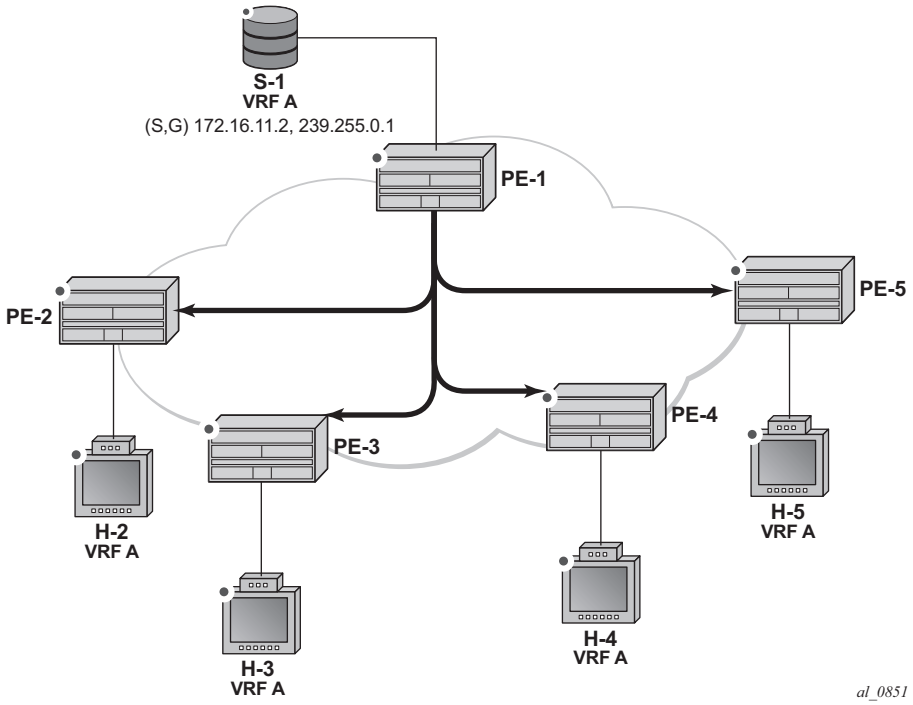
The provider tunnel used is the default Inclusive PMSI (I-PMSI) that is instantiated between all source and receiver PEs. This results in a full mesh of provider tunnels between all PEs in the MVPN. In a large network, this can result in an inefficient use of bandwidth because multicast traffic is forwarded to all PEs regardless of whether the PE has an interested receiver. Some of the mesh scaling issues can be mitigated by using source-only/destination-only configuration on PEs. However, this technique requires additional configuration and is not fully optimal when mLDP is used in the core.

To address the preceding limitation, wildcard Selective PMSI (S-PMSI) has been developed, as per RFC 6625. In the standard customer signaling notation of (C-S,C-G), this becomes (C-*,C-*). Using methods defined in RFC 6625, it is possible to use a (C-*,C-*) S-PMSI as the default tunnel, where the receiver PE can join the S-PMSI by mapping the channel join to a wildcard channel group. Multiple channels can be transported by the wildcard (C-*,C-*) S-PMSI, where an S-PMSI auto-discovery route is advertised with an empty channel group and source address:

1. Bandwidth savings can be achieved by the delivery of multicast channels on S-PMSIs, because traffic is not forwarded to PEs that have no interested receivers.
2. Control plane savings can be achieved by eliminating the need for the full tunnel mesh between all PES. The wildcard S-PMSI is only signaled on PEs containing attached upstream multicast sources, for which the PE is resolved as an upstream multicast hop (UMH) within the MVPN.

Figure 355 shows the concept of an MVPN.

Figure 355 **Multicast VPN**



In Figure 355, PE-1 has a directly connected multicast source (S-1). For clarity, consider this MVPN as a single multicast group. PE-1 is configured as a sender PE because it is the PE closest to the source. PE-2, PE-3, PE-4, and PE-5 are configured as receiver-only PEs because they have connected receiver hosts H-2, H-3, H-4, and H-5, respectively. Hosts H-2 to H-5 connected to receiver PEs can receive multicast channels from the source, S-1, connected to the source PE, PE-1, within the same virtual private routed network (VPRN).

Within the provider network, multicast traffic is delivered from the source PE to the receiver PE across a PMSI tunnel. This tunnel is, in this case, a P2MP LSP, with its root at PE-1 and with a leaf at each of the receiver PEs. Any traffic that is forwarded into the tunnel interface is replicated so that a single copy of a multicast stream is received at all PEs.

The PMSI tunnel is created after each PE has declared themselves as a member of the MVPN using BGP MVPN auto-discovery techniques.

There are two choices of PMSI:

- An I-PMSI, which is created on each PE within the MVPN, with a root at each PE and a leaf at all other PEs that are members of the MVPN. The I-PMSI is the default tunnel for all multicast traffic carried between sender PE and receiver PEs. When at least one receiver PE has a host interested in becoming a member of a multicast group, traffic for that group is delivered to all PEs via the I-PMSI, regardless of whether they have an interested host. Receiver PEs with no such interested host then drop the traffic.
- An S-PMSI, which is created to carry multicast traffic to the subset of receiver PEs that have connected hosts interested in receiving multicast traffic. This can be for a specific group, so that one S-PMSI carries traffic for one multicast group, denoted as (C-S,C-G) or (C-*,C-G). The S-PMSI can also be signaled to carry traffic for multiple multicast groups, denoted using a wildcard: (C-*,C-*) or (C-S,C-*). The wildcard S-PMSI can be signaled in place of the I-PMSI, so that all traffic can be carried on the S-PMSI by default. In this case, no I-PMSI is signaled.

In the case of an I-PMSI, the tunnel type is included in the BGP auto-discovery intra-AD route originated and advertised to all other PEs within the VPRN.

If a wildcard S-PMSI is to be used and no I-PMSI tunnel is to be signaled, then an intra-AD route update for I-PMSI is advertised with no tunnel type attribute included. In addition, the source PE will originate an additional S-PMSI auto-discovery route containing no source-encoding wildcard information.

Table 25 shows the S-PMSI MVPN BGP network layer reachability information (NLRI) advertisement.

Table 25 **S-PMSI Auto-Discovery BGP NLRI**

Route Distinguisher (8 octets)
Multicast Source Length (1 octet)
Multicast Source (variable)
Multicast Group Length (1 octet)
Multicast Group (variable)
Originating Router IP Address

To signal the S-PMSI as wildcard (C-*,C-*) S-PMSI, the multicast source length and multicast group length fields are encoded with the value of zero (0), representing (C-*,C-*) wildcard.

The objectives are to:

- Configure multicast in a VPRN on PE-1 to PE-5 using mLDP as the tunnel signaling method.
- Connect multicast sources to the sender PE-1.
- Create receiver hosts that can receive multicast traffic from the source, and to observe the effect on the provider network.

The following configuration tasks should be completed as a prerequisite:

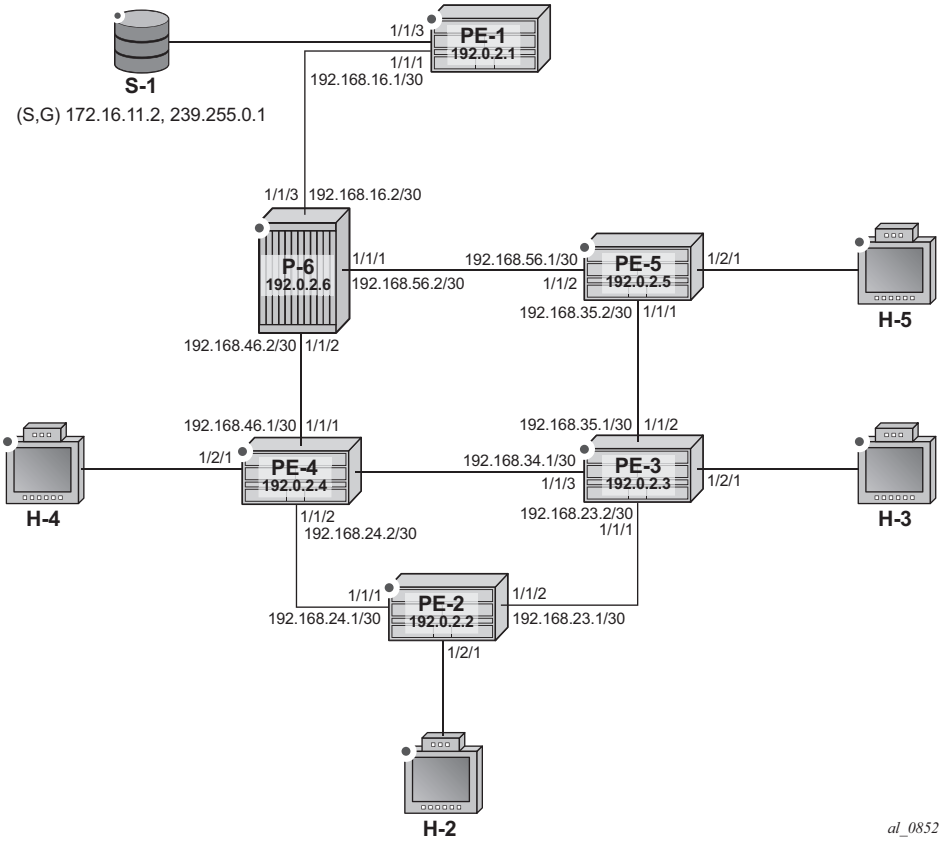
- Full mesh IS-IS or OSPF between each of the PE routers and the RR.
- Link-layer LDP between all PEs.
- mLDP used as the provider tunnel signaling protocol. This is enabled by default when link-layer LDP is enabled.

RSVP-TE is also supported as a provider tunnel signaling mechanism and could be used instead of mLDP.

Configuration

The example topology is shown in Figure 356, containing five PE routers. P-6 acts as an RR.

Figure 356 **Schematic Topology**



Global BGP Configuration

The first step is to configure an IBGP session between each of the PEs and the RR (PE-6) shown in Figure 356. The address families negotiated between the IBGP peers are vpn-ipv4 (unicast routing) and mvpn-ipv4 (multicast routing).

The configuration for PE1 is:

```
configure
    router
        bgp
            group INTERNAL
                family vpn-ipv4 mvpn-ipv4
                type internal
                neighbor 192.0.2.6
            exit
        exit
```

The configuration for the other PE nodes is exactly the same.

The configuration for the RR at P-6 is:

```
configure
    router
        bgp
            cluster 0.0.0.1
            group "RR_CLIENTS"
                family vpn-ipv4 mvpn-ipv4
                type internal
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

On PE-1, verify that the BGP session with RR at P-6 is established with address families vpn-ipv4 and mvpn-ipv4 capabilities negotiated:

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:65545       Local AS:65545
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
Total BGP Paths         : 17          Total Path Memory           : 3400
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 6
Total VPN-IPv4 Rem. Rts : 6           Total VPN-IPv4 Rem. Act. Rts: 4
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 6           Total MVPN-IPv4 Rem Act Rts : 4
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
```

```
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                    AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                       PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.6
               65545        3    0 00h00m08s 0/0/0 (VpnIPv4)
                            3    0           0/0/0 (MvpnIPv4)
-------------------------------------------------------------------------------
*A:PE-1#
```

The same command can be used on the other PEs to verify their BGP sessions to the RR.

Configuring VPRN on PEs

The following outputs show the VPRN configurations for each PE. The specific MVPN configuration is shown later.

The VPRN configuration for PE-1 is:

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 65545:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:65545:1
            interface "int-PE-1-S-1" create
                address 172.16.11.1/24
                sap 1/1/3 create
                exit
            exit
            interface "rp" create
                address 10.0.0.1/32
                loopback
```

```
            exit
            pim
                apply-to all
                rp
                    static
                        address 10.0.0.1
                            group-prefix 239.255.0.0/16
                        exit
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

There is a single interface toward S-1 from which the multicast group is transmitted.

If the customer signaling uses PIM ASM, a customer Rendezvous Point (RP) is required.

A loopback interface called “rp” acts as the RP for all group prefixes in the 239.255.0.0/16 range. This will be the RP for all groups.

MVPN configuration enables BGP as both the auto-discovery mechanism and the customer multicast signaling protocol across the VPRN. The provider tunnel between PEs within the MVPN is signaled using mLDP.

PE-2 contains an attached receiver, therefore a single interface is configured to accommodate this, as follows. The RP is configured as a static RP:

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 65545:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:65545:1
            interface "int-PE-2-H-2" create
                address 172.16.22.1/24
                sap 1/2/1 create
                exit
            exit
            igmp
                interface "int-PE-2-H-2”
                    no shutdown
                exit
                no shutdown
            exit
            pim
                apply-to all
                rp
                    static
```

```
                        address 10.0.0.1
                            group-prefix 239.255.0.0/16
                        exit
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

PE-3 also has an attached receiver, as follows:

```
configure
    service
        vprn 1 customer 1 create
            route-distinguisher 65545:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:65545:1
            interface "int-PE-3-H-3" create
                address 172.16.33.1/24
                sap 1/2/1 create
                exit
            exit
            igmp
                interface "int-PE-3-H-3"
                    no shutdown
                exit
                no shutdown
            exit
            pim
                apply-to all
                rp
                    static
                        address 10.0.0.1
                            group-prefix 239.255.0.0/16
                        exit
                    exit
                exit
                no shutdown
            exit
            no shutdown
```

The configuration for PE-4 and PE-5 is similar.

MVPN Configuration for PEs

The provider-tunnel inclusive configuration specifies that a wildcard S-PMSI will be used instead of an I-PMSI as the default PMSI. The MVPN configuration for all PEs is:

```
configure
    service
        vprn 1
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                provider-tunnel
                    inclusive
                        mldp
                            no shutdown
                        exit
                        wildcard-spmsi
                    exit
                exit
                vrf-target unicast
                exit
            exit
```

The keyword **wildcard-spmsi** reduces the number of PMSIs signaled. If there are no group sources on the receiver PEs, there will be no S-PMSI signaled. This has an effect similar to configuring the receiver PEs as MDT-type receiver-only.

Provider Tunnel Signaling

Each PE originates BGP MVPN intra-AD routes to determine membership of an MVPN.

The provider tunnels constructed between the PEs within the VPRN are signaled on receipt of an intra-AD route update from other PEs. The intra-AD update message contains details of the originator, along with the VRF route-target extended community. If an I-PMSI is to be signaled, a PMSI tunnel attribute is included that determines the tunnel type, such as LDP P2MP LSP. PEs that receive this intra-AD update will import the route into the MVPN, then signal a P2MP LDP label map toward the originator, which is the root of the LDP P2MP LSP.

However, if a wildcard S-PMSI is to be used as the default PMSI, no PMSI tunnel attribute is included within the intra-AD update.

The following output shows a BGP update originated by PE-1, and received by PE-2:

```
*A:PE-2# show router bgp routes mvpn-ipv4 type intra-ad rd 65545:1 detail originator-
ip 192.0.2.1
===============================================================================
 BGP Router ID:192.0.2.2        AS:65545       Local AS:65545
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
```

```
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Original Attributes

Route Type     : Intra-Ad
Route Dist.    : 65545:1
Originator IP  : 192.0.2.1
Nexthop        : 192.0.2.1
From           : 192.0.2.6
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : no-export target:65545:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.6
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h22m08s
VPRN Imported  :  1
---snip---
```

There is no PMSI tunnel attribute included, and the route is imported into the correct VPRN (VPRN 1).

The intra-AD originated by PE-2 is:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad rd 65545:1 originator-ip
192.0.2.2 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65545       Local AS:65545
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Route Type     : Intra-Ad
Route Dist.    : 65545:1
Originator IP  : 192.0.2.2
Nexthop        : 192.0.2.2
From           : 192.0.2.6
```

```
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : no-export target:65545:1
Cluster        : 0.0.0.1
Originator Id  : 192.0.2.2              Peer Router Id : 192.0.2.6
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h23m49s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

This output also contains no PMSI tunnel attribute: PE-2 has no group source and there is no S-PMSI signaled. All other receiver PEs will originate a similar intra-AD route.

The following output shows all intra-AD routes originated by the PEs within the VPRN, as received by PE-1:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type intra-ad rd 65545:1
===============================================================================
 BGP Router ID:192.0.2.1        AS:65545       Local AS:65545
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
i     Intra-Ad                    192.0.2.1              100         0
      65545:1                     -                                  -
      192.0.2.1                   -
      No As-Path                  -
u*>i  Intra-Ad                    192.0.2.2              100         0
```

```
      65545:1                   -                                   -
      192.0.2.2                 -
      No As-Path                -
u*>i  Intra-Ad                  192.0.2.3           100         0
      65545:1                   -                                   -
      192.0.2.3                 -
      No As-Path                -
u*>i  Intra-Ad                  192.0.2.4           100         0
      65545:1                   -                                   -
      192.0.2.4                 -
      No As-Path                -
u*>i  Intra-Ad                  192.0.2.5           100         0
      65545:1                   -                                   -
      192.0.2.5                 -
      No As-Path                -
------------------------------------------------------------------------------
Routes : 5
==============================================================================
*A:PE-1#
```

Instead of an I-PMSI being signaled, an S-PMSI AD route update is advertised by PE-1 to all receiver PEs within the MVPN. The NLRI encoding has a zero length field for both source and group addresses, so is seen to represent multicast group (C-*,C-*). This is wildcard nomenclature for both source and group addresses.

The BGP route as advertised by PE-1:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type spmsi-ad rd 65545:1 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65545       Local AS:65545
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : Spmsi-Ad
Route Dist.    : 65545:1
Originator IP  : 192.0.2.1
Source IP      : 0.0.0.0
Group IP       : 0.0.0.0
Nexthop        : 192.0.2.1
To             : 192.0.2.6
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
```

```
Connector      : None
Community      : target:65545:1
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 192.0.2.6
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                     Dest Class     : 0
-------------------------------------------------------------------------------
PMSI Tunnel Attribute :
Tunnel-type    : LDP P2MP LSP
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 0
Root-Node      : 192.0.2.1             LSP-ID         : 8193
-------------------------------------------------------------------------------
---snip---
```

The source IP and group IP address fields are advertised as 0.0.0.0, and the tunnel type attribute is now included as an LDP P2MP LSP.

The following output shows the MVPN status at PE-1, with the I-PMSI tunnel name containing a wildcard S-PMSI denoted by (W):

```
*A:PE-1# show router 1 mvpn

===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Disabled
vrf-import         : N/A
vrf-export         : N/A
vrf-target         : unicast
C-Mcast Import RT  : target:192.0.2.1:2

ipmsi              : ldp
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-if-73728(W)
Mdt-type           : sender-receiver

BSR signalling     : none
Wildcard s-pmsi    : Enabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
*A:PE-1#
```

At this point, there is no interested host and no customer multicast flow (c-flow), so there is no S-PMSI LDP P2MP LSP signaled.

Data Transmission at Source PE

Multicast traffic for a particular group will be forwarded between the sender and receiver PE over a provider tunnel (PMSI). Because there is no default I-PMSI present, the receiver PE has to choose an S-PMSI to be used for forwarding, based on the NLRI contained within the S-PMSI AD routes.

The provider tunnel is signaled using a P2MP LDP label mapping message toward the root address signaled in the wildcard S-PMSI AD BGP update message. As previously shown, the update message is based on whether traffic is being forwarded on the shared or source-based shortest path tree.

When joining the shared tree, a c-multicast shared-join is sent toward the appropriate PE, which represents the UMH toward the RP. The UMH is chosen from the unicast route that represents the RP address. When joining the shortest path tree, a source-join c-multicast route is sent toward the UMH chosen from the unicast route that represents the actual source address. In both cases, the VPN-IPv4 unicast route advertises a VRF route import community that contains the system address as a next hop. Upon receipt of these updates, the UMH PE will forward traffic along the signaled wildcard S-PMSI.

Each S-PMSI is bound to one or more flows, as determined by the NLRI contained within the S-PMSI BGP update. The use of the wildcard within the BGP update to replace the c-source and c-group allows multiple flows to be bound to a single provider tunnel.



Note: Traffic will only be forwarded upon reception of a BGP MVPN source-join or shared-join BGP route at the sender PE.

Data Reception at Receiver PE

When the sender PE has originated an S-PMSI AD route update, each receiver PE will install the route into its VRF. The S-PMSIs installed are used to select an appropriate upstream multicast router for a c-flow when an attached receiver is interested in receiving traffic from that c-flow.

The receiver PE will receive a flow based on the best match of the incoming (C-S,C-G) or (C-*,C-G) IGMP/MLD or PIM join.

If an IGMP/MLD group membership query or PIM join is received by the receiver PE over an attachment circuit for a group, the contained (C-S,C-G) or (C-*,C-G) must match the (C-S,C-G) contained within any installed S-PMSI AD route. In the case of the wildcard S-PMSI being the only installed NLRI, this will be a match; that is, incoming (C-*,C-G) or (C-S,C-G) will match the S-PMSI (C-*,C-*).

In this example, the c-group flow is 239.255.0.1.

Traffic Flow

A traffic stream representing a c-flow is enabled on PE-1: group address 239.255.0.1 with source address of 172.16.11.2. The RP for this group is found locally on PE-1. The outgoing interface list is empty:

```
*A:PE-1# show router 1 pim group detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 172.16.11.2
RP Address         : 10.0.0.1
Advt Router        : 192.0.2.1
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.11.2
MRIB Src Flags     : direct
Keepalive Timer Exp: 0d 00:03:03
Up Time            : 0d 00:49:27        Resolved By        : rtable-u

Up JP State        : Not Joined         Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : Pruned             Register Stop Exp  : 0d 00:00:44
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.11.2
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List :
Outgoing Sap List  :
Outgoing Host List :

Curr Fwding Rate   : 19717.1 kbps
Forwarded Packets  : 158271505          Discarded Packets  : 0
Forwarded Octets   : 7280489230         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

A host connected to PE-2 sends a (C-*,C-G) IGMP v2 group membership query for group 239.255.0.1.

PE-2 sends a BGP MVPN shared-join route update toward PE-1, where the RP address of the group 10.0.0.1 is found.

The following debug output from PE-2 shows the shared-join BGP route update transmitted by PE-2:

```
1 2016/10/10 07:44:16.19 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.6
"Peer 1: 192.0.2.6: UPDATE
Peer 1: 192.0.2.6 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
      Type: Shared-Join Len:22 RD: 65545:1 SrcAS: 65545 Src: 10.0.0.1 Grp: 239.255.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

Upon receipt of the shared-join, traffic flows on the shared tree toward the receiver PE. This will flow on the default wildcard S-PMSI, as shown in the outgoing interface list:

```
*A:PE-1# show router 1 pim group 239.255.0.1 type starg detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : *
RP Address         : 10.0.0.1
Advt Router        : 192.0.2.1
Flags              :                    Type               : (*,G)
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : self
Keepalive Timer    : Not Running
Up Time            : 0d 00:03:49        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:11
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor       :
Incoming Intf      :
Outgoing Intf List : mpls-if-73728(W)

Curr Fwding Rate   : 0.0 kbps
```

```
Forwarded Packets  : 0                  Discarded Packets  : 0
Forwarded Octets   : 0                  RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

When traffic is received on the shared tree by PE-2, the source address is learned, so a source-join BGP route update is sent toward the UMH PE, which contains the source address of 172.16.11.2. The UMH is chosen from the unicast route-table using a lookup for the best route matching the source address.

The following debug output from PE-2 shows the BGP source-join route update toward the source of group 239.255.0.1, as transmitted by PE-2:

```
3 2016/10/10 07:44:45.19 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.6
"Peer 1: 192.0.2.6: UPDATE
Peer 1: 192.0.2.6 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Source-Join Len:22 RD: 65545:1 SrcAS: 65545 Src: 172.16.11.2
                          Grp: 239.255.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The c-flow toward host H-2 flows along the shortest path tree, and on PE-1 the outgoing interface list is populated with the wildcard S-PMSI:

```
*A:PE-1# show router 1 pim group detail 239.255.0.1 source 172.16.11.2

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 172.16.11.2
RP Address         : 10.0.0.1
Advt Router        : 192.0.2.1
Flags              : spt, rpt-prn-des   Type              : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.11.2
MRIB Src Flags     : direct
Keepalive Timer Exp: 0d 00:02:01
Up Time            : 0d 01:10:28        Resolved By       : rtable-u
```

```
Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Pruned             Up JP Rpt Override : 0d 00:00:00

Register State     : Pruned             Register Stop Exp  : 0d 00:00:13
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.11.2
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73728(W)

Curr Fwding Rate   : 6915.5 kbps
Forwarded Packets  : 198813564          Discarded Packets  : 0
Forwarded Octets   : 9145423944         RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The outgoing interface is the MPLS interface mpls-if-73728. This maps to a P2MP LDP label binding from which the p2mp-id can be derived:

```
*A:PE-1# show router ldp bindings active p2mp
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                 Interface
RootAddr                                Op              IngLbl    EgrLbl
EgrNH                                   EgrIf/LspId
-------------------------------------------------------------------------------
8193                                    73728
192.0.2.1                               Push              --      262137
192.168.16.2                            1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
---snip---
```

After the source-join is received, the sender PE will advertise a source-active AD BGP route to all PEs within the MVPN, to announce the presence of a (C-S,C-G) group. The following debug output shows the source-active AD route received on PE-2:

```
4 2016/10/10 07:45:06.29 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.6
"Peer 1: 192.0.2.6: UPDATE
Peer 1: 192.0.2.6 - Received BGP UPDATE:
```

```
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 29 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.1
        Type: Source-AD Len: 18 RD: 65545:1 Src: 172.16.11.2 Grp: 239.255.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.1
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        0.0.0.1
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:65545:1
"
```

The PIM status of the group on receiver PE-2 shows that the incoming interface is the wildcard S-PMSI originated on PE-1, as denoted by the (W):

```
*A:PE-2# show router 1 pim group 239.255.0.1 source 172.16.11.2 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 172.16.11.2
RP Address         : 10.0.0.1
Advt Router        : 192.0.2.1
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.1
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:02:21
Up Time            : 0d 00:12:43        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:16
Up JP Rpt          : Not Pruned         Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.1
Incoming Intf      : mpls-if-73729(W)
Outgoing Intf List : int-PE-2-H-2

Curr Fwding Rate   : 7683.3 kbps
Forwarded Packets  : 15692528           Discarded Packets  : 0
Forwarded Octets   : 721856288          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-2#
```

The S-PMSI is an LDP P2MP LSP. The LDP label binding for P2MP LSP-Id 8193 at PE-2 shows that the label operation is a label pop:

```
*A:PE-2# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73729
192.0.2.1                                   Pop             262136      --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
*A:PE-2#
```

PE-3 has no host joined to c-flow group 239.255.0.1. However, it contains the PIM state for this group due to the presence of the source-active AD route within the VRF. This route was received when the host connected to PE-2 joined the c-flow group.

The following output shows the source-active AD route within PE-3 for group 239.255.0.1 from source 172.16.11.2:

```
*A:PE-3# show router bgp routes mvpn-ipv4 type source-ad rd 65545:1
===============================================================================
 BGP Router ID:192.0.2.3        AS:65545       Local AS:65545
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                 OriginatorIP          LocalPref   MED
      RD                        SourceAS                          Label
      Nexthop                   SourceIP
      As-Path                   GroupIP
-------------------------------------------------------------------------------
u*>i  Source-Ad                 -                     100         0
      65545:1                   -                               -
      192.0.2.1                 172.16.11.2
      No As-Path                239.255.0.1
-------------------------------------------------------------------------------
```

```
Routes : 1
===============================================================================
*A:PE-3#
```

However, traffic is not received from the S-PMSI because there is no label binding for the LDP P2MP LSP. The following output shows that there is no label binding for the LSP Id 8193, which has its root on PE-1:

```
*A:PE-3# show router ldp bindings p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings
===============================================================================
P2MP-Id
RootAddr                                    Interface       IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
Peer
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
*A:PE-3#
```

A static IGMPv2 (*,G) group is created on interface int-PE-3-H3 for group 239.255.0.1 toward PE-3. The following debug output shows the process.

```
4 2016/10/10 08:25:30.13 UTC MINOR: DEBUG #2001 vprn1 IGMP[vprn1 inst 2]
"IGMP[vprn1 inst 2]: igmpIfSrcAdd
Adding i/f source entry for interface int-PE-3-H-3 [ifIndex 2] (*,239.255.0.1) to IGMP
fwdList Database, redir if N/A"

3 2016/10/10 08:25:30.13 UTC MINOR: DEBUG #2001 vprn1 IGMP[vprn1 inst 2]
"IGMP[vprn1 inst 2]: igmpProcessGroupRec
Process group rec CHG_TO_EXCL received on interface int-PE-3-H-3 [ifIndex 2] for group
239.255.0.1 in mode INCLUDE. Num srcs 0"

2 2016/10/10 08:25:30.13 UTC MINOR: DEBUG #2001 vprn1 IGMP[vprn1 inst 2]
"IGMP[vprn1 inst 2]: igmpIfGroupAdd
Adding 239.255.0.1 to IGMP interface int-PE-3-H-3 [ifIndex 2] database"

1 2016/10/10 08:25:08.60 UTC MINOR: DEBUG #2001 vprn1 IGMP[vprn1 inst 2]
"IGMP[vprn1 inst 2]: igmpIfSGStaticAdd
Adding Static SG (0.0.0.0,239.255.0.1) to IGMP interface int-PE-3-H-3 [ifIndex 2]"
```

A similar process takes place when receiver host H-3 sends an unsolicited IGMP v2 group membership query for this group. The first message would correspond to the IGMP query instead.

After the IGMP interface source entry has been added for interface int-PE-3-H-3, an mLDP P2MP label mapping message is sent from PE-3 toward the root node PE-1, as follows:

```
8 2016/10/11 12:39:26.34 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 185) to 192.0.2.4:0
Protocol version = 1
Label 262136 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 8193
"

7 2016/10/11 12:39:26.34 UTC MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 262136 for P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId:
8193 to peer 192.0.2.4:0.
```

BGP shared-join and source-join BGP route updates are sent via the RR toward the RP (source = 10.0.0.1) and the actual source (172.16.11.2), respectively:

```
9 2016/10/11 12:39:25.51 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.6
"Peer 1: 192.0.2.6: UPDATE
Peer 1: 192.0.2.6 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 100
    Flag: 0x90 Type: 14 Len: 57 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.3
        Type: Shared-Join Len:22 RD: 65545:1 SrcAS: 65545 Src: 10.0.0.1
                          Grp: 239.255.0.1
        Type: Source-Join Len:22 RD: 65545:1 SrcAS: 65545 Src: 172.16.11.2
                          Grp: 239.255.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:192.0.2.1:2
"
```

The PIM status for the c-group 239.255.0.1 on PE-3 is now:

```
*A:PE-3# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (*,G)                          mpls-if-73729* 1
   *                           10.0.0.1
239.255.0.1                 (S,G)               spt        mpls-if-73729* 1
```

```
   172.16.11.2                 10.0.0.1
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

Assume that a receiver on each of PE-4 and PE-5 needs to join group 239.255.0.1, as shown in Figure 357.

Figure 357 **S-PMSI P2MP LSP Schematic**

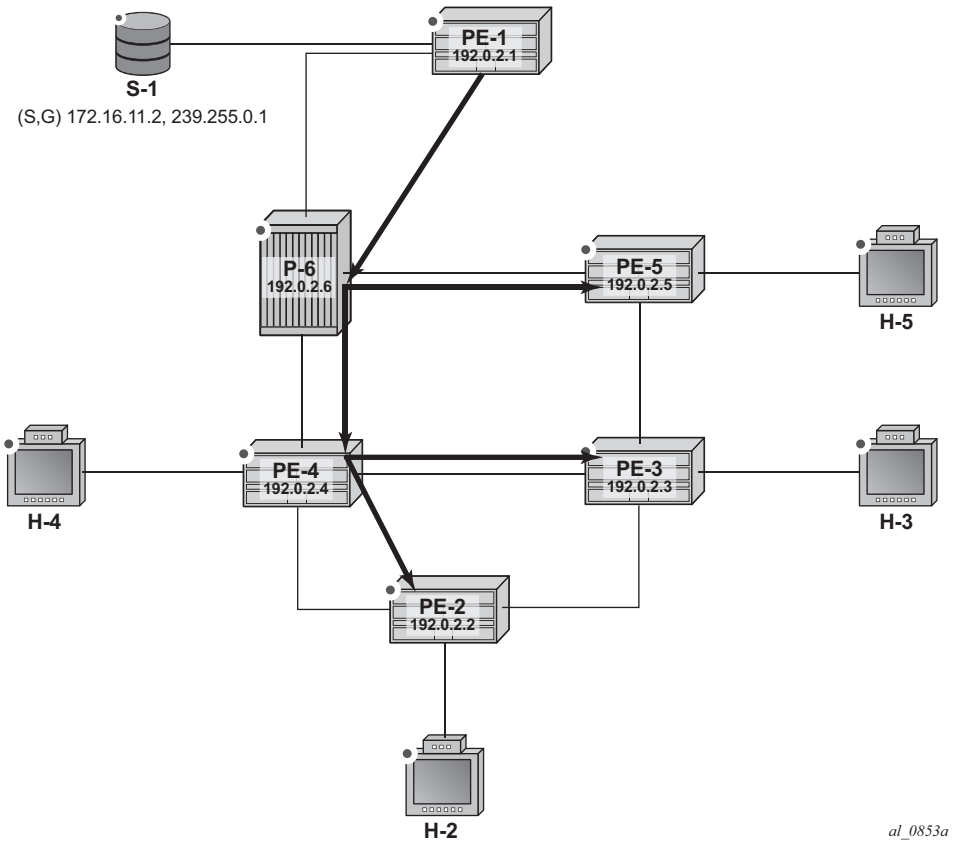


Figure 357 shows the S-PMSI P2MP LSP. The next set of outputs shows the P2MP label mapping of the LDP LSP between PE-1 and the receiver PEs.

The root of the S-PMSI is on PE-1, as follows:

```
*A:PE-1# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
```

```
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73728
192.0.2.1                                   Push              --      262137
192.168.16.2                                1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
*A:PE-1#
```

The egress label on PE-1 becomes the ingress label on P-6. P-6 has two leaves: one toward PE-4 and one toward PE-5, as follows:

```
*A:P-6# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.6:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        Unknw
192.0.2.1                                   Swap            262137    262136
192.168.46.1                                1/1/2

8193                                        Unknw
192.0.2.1                                   Swap            262137    262136
192.168.56.1                                1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
```

On PE-5, the following output shows that the LSP terminates as a leaf, as the operation (Op) is shown as “pop”:

```
*A:PE-5# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
```

```
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
-------------------------------------------------------------------------------
8193                                     73728
192.0.2.1                                Pop             262136      --
  --                                       --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
*A:PE-5#
```

On PE-4, the P2MP LSP has 3 entries: a pop operation to receiver H-4, and two label swaps toward PE-3 and PE-2:

```
*A:PE-4# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                  Interface
RootAddr                                 Op              IngLbl    EgrLbl
EgrNH                                    EgrIf/LspId
-------------------------------------------------------------------------------
8193                                     73728
192.0.2.1                                Pop             262136      --
  --                                       --

8193                                     73728
192.0.2.1                                Swap            262136    262136
192.168.24.1                             1/1/2

8193                                     73728
192.0.2.1                                Swap            262136    262136
192.168.34.1                             1/1/3

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 3
===============================================================================
*A:PE-4#
```

PE-2 and PE-3 are termination PEs for P2MP leaf. On PE-2, the pop operation is shown:

```
*A:PE-2# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
```

```
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73729
192.0.2.1                                   Pop             262136      --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
*A:PE-2#
```

On PE-3, the P2MP pop operation is shown:

```
*A:PE-3# show router ldp bindings active p2mp p2mp-id 8193 root 192.0.2.1
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3:0)
             (IPv6 LSR ID ::[0])
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op              IngLbl    EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73729
192.0.2.1                                   Pop             262136      --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
*A:PE-3#
```

Conclusion

MVPN wildcard Selective PMSI (S-PMSI), developed as per RFC 6625, provides an optimal solution for multicast routing in a VPRN. This protocol provides simple configuration, operation, and fast protection time in conjunction with MPLS and LDP fast-failover schemes. Wildcard S-PMSI can be used in a multicast network to avoid a large full mesh of an I-PMSI.

NG-MVPN Source Redundancy

This chapter provides information about MVPN source redundancy.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

The chapter was initially written for release 12.0.R1, using multicast LDP as the provider tunnel signaling mechanism for IPv4 multi-casting. The customer multicast signaling protocol within the VPN must be BGP. The CLI in the current edition corresponds to 14.0.R5.

Summary

Multicast source redundancy allows operators to provide multiple geo-redundant sources for the same multicast group in a multicast virtual private network (MVPN). For instance, in an IPTV environment where a TV channel maps to a multicast group, the same TV channel can be provided from sources in a geographically diverse manner where a national broadcaster can have multiple sources from two or more regional distribution centers.

Knowledge of Multi-Protocol BGP (MP-BGP) and RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, is assumed throughout this chapter, as well as Protocol Independent Multicast (PIM), RFC 6513, *Multicast in MPLS/BGP IP VPNs*, and RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs*.

Overview

Hosts connected to receiver PEs can receive TV channels from a specific source, with a regional backup source available in case of a failure.

Figure 358 **Source Redundancy Example.**

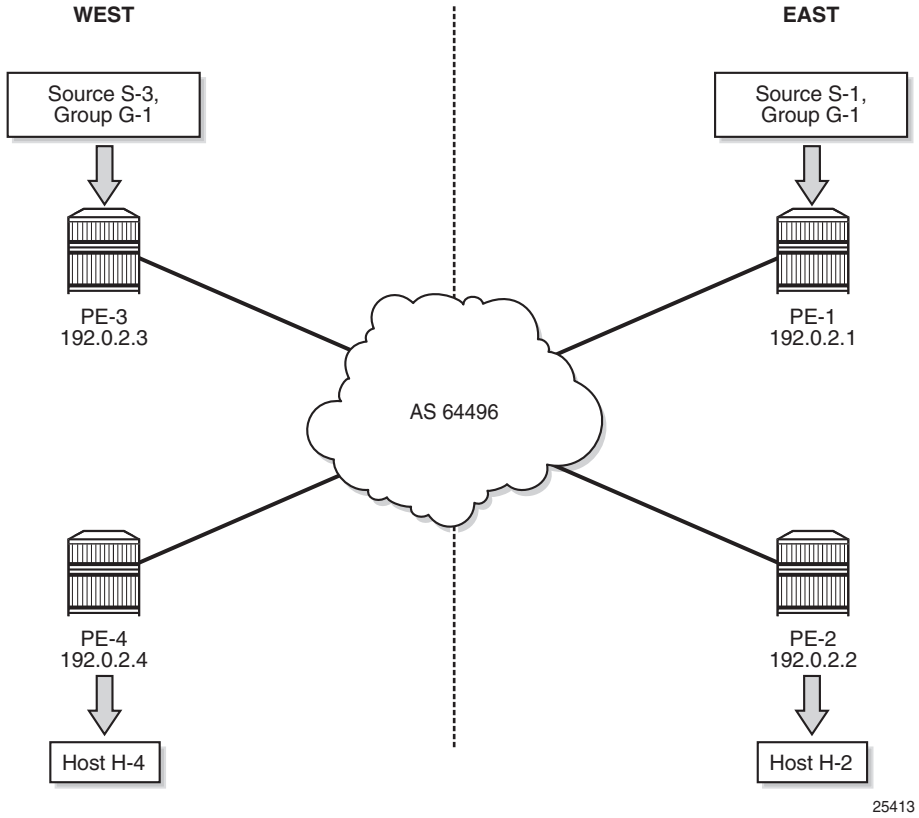


Figure 358 shows the concept of source redundancy. PE-1 and PE-3 have directly connected multicast sources. For clarity, consider a single multicast group with two separate sources connected at different sites. The content of each group is identical at a given time (allowing for transmission delay), as is expected for an IPTV channel. PE-1 and PE-3 are referred to as sender PEs because they are closer to the source; PE-2 and PE-4 are referred to as receiver PEs because they are closer to hosts H-2 and H-4.

A multicast group, group 1 (G-1) has two sources: Source 1 (S-1) in the east region and Source 3 (S-3) in the west region which are connected to PE-1 and PE-3 respectively. Receivers connected to PE-2 in the east region will join group (S-1,G-1) and receivers connected to PE-4 in the west region will join group (S-3,G-1). The presence of each source is declared within the multicast VPN by the sender PE. When a multicast group becomes active, a BGP Source Active auto-discovery (SA) route is advertised to all PEs within the multicast VPN. This must occur even if no receiver indicates that it wishes to become a member of this group. In other words, the SA must be persistently present in the receiver PEs when the source is available.

Should either source fail or become unavailable, then the sender PE will notify the receiver PEs by sending an NLRI unreachable BGP SA Route that declares the absence of the source. All hosts that are members of this group will then switch to receive traffic from the remaining active source. Only customer multicast joins received as IGMP (*,G) queries or PIM (*,G) joins at the receiver PE are valid, because the source address is not specified.

Source redundancy is achieved by:

- Configuring a list of redundant sources within each receiver PE.
- Configuring the sender PEs to originate a BGP Source Active Auto Discovery for each detected active multicast source, regardless of whether a receiver is joined to the multicast group or not. As a result, a Source Active route is originated on a per (S,G) basis.

  For multiple SAs to be persistently present in the receiver PEs, one of the following two conditions must be configured within the sender PEs:

  – Either disable inter-site shared trees on the sender PEs, such that there is no c-tree with root at the RP. Any active source will announce its presence using a BGP SA to all receiver PEs so no shared joins are sent by receiver PEs to RP, or

  – Leave inter-site shared trees as enabled, but configured so that the SA AD route for each multicast group is persistently present in the receiver PEs, even in the absence of requesting hosts for each group. Shared and Source Joins are sent by the receiver PEs.

  Both of the preceding options are supported. The default behavior has inter-site shared trees enabled without persistency. In this example, inter-site shared trees at the sender PEs are enabled with Source Active routes set to be persistent.
- Ensuring that the preferred source is IP reachable within the VPRN from the receiver PE. This must be a remote source advertised from a remote PE within the VPRN.
- Receiver PEs will accept the Source Active route(s) into the appropriate Multicast VRF.

- Ensuring the preferred active source should have a higher BGP Local Preference. This is achieved using a route policy. Any other sources from the redundant list should exist as suppressed standby sources, but the (S,G) state should exist if the source is active – when a valid BGP MVPN Source Active route for that source has been received.

All of these conditions are achieved by configuration.

In order to allow each receiver PE to choose a preferred source, each SA route advertised by the sender PE will be tagged with a community value. Each receiver PE can then use the community value contained within each SA route update received to set the Local Preference BGP attribute to a value such that the receiver PE can choose the most preferred active source.

The objectives are:

- To configure multicast in a VPRN on PE-1 to PE-4 with inter-site-shared trees enabled on the receiver PEs and Source Active routes persistently present, for reasons previously described.
- To connect redundant sources to the sender PE-1 and PE-3, with each multicast source having the same group address. For ease of configuration, a single redundant source is used.
- To advertise each source to the receiver PEs (PE-2 and PE-4), using appropriate route policies for adding community strings to the BGP Source Active Auto-Discovery routes.
- To configure appropriate route policies that allow each BGP SA route to have the correct Local Preference set, based on the community strings present.
- To allow receivers to connect to the appropriate source, using (*,G) joins.

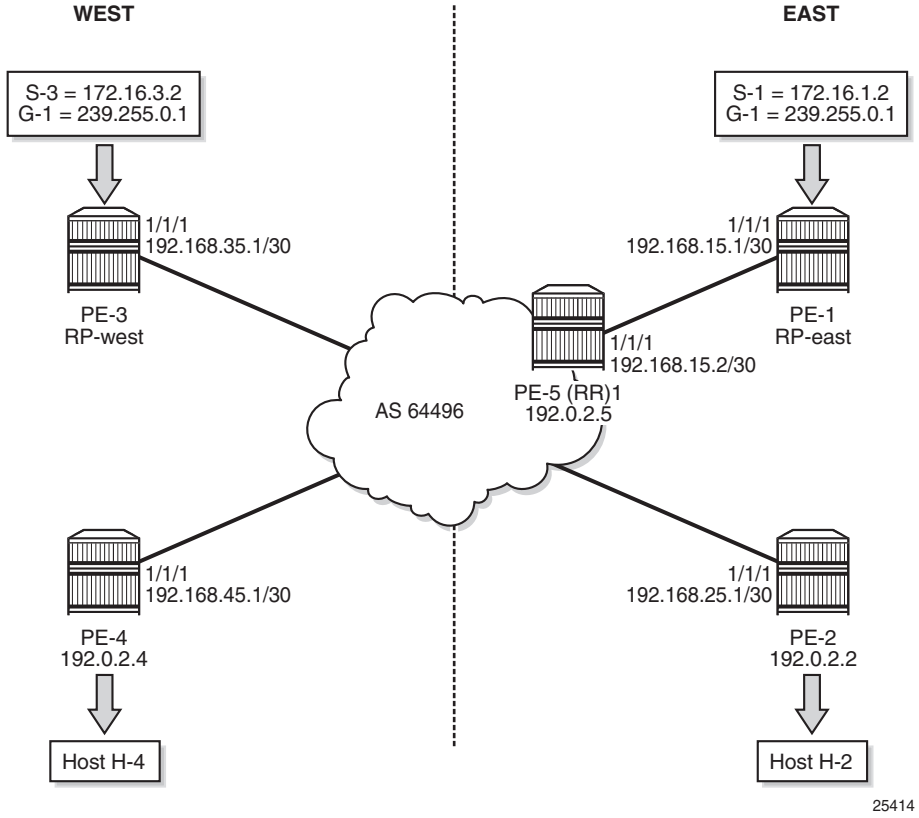
The following configuration tasks should be completed as a pre-requisite:

- Full mesh IS-IS or OSPF between each of the PE routers and the route reflector.
- Link-layer LDP between all PEs. RSVP could also be used.
- Multicast LDP is used as the provider tunnel signaling protocol. This is enabled by default when link layer LDP is enabled. RSVP and PIM SSM are also supported as provider tunnel signaling mechanisms and could be used.

Configuration

The example topology is shown in Figure 359, containing the four PEs plus the route reflector at P-5.

Figure 359 **Schematic Topology**



Global BGP Configuration

The first step is to configure an iBGP session between each of the PEs and the route reflector (RR) seen in Figure 359. The address families negotiated between the iBGP peers are vpn-ipv4 (unicast routing) and mvpn-ipv4 (multicast routing).

The BGP configuration for all PE nodes is identical:

```
*A:PE-1# configure
    router
        bgp
            group "INTERNAL"
                family vpn-ipv4 mvpn-ipv4
                type internal
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
        exit
```

The configuration for the Route Reflector at P-5 is:

```
*A:P-5# configure
    router
        bgp
            group "RRclients"
                family vpn-ipv4 mvpn-ipv4
                type internal
                cluster 1.1.1.1
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
            no shutdown
        exit
```

On PE-1, verify that the BGP session with RR at P-5 is established with address families “vpn-ipv4” and “mvpn-ipv4” capabilities negotiated:

```
*A:PE-1# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
Total BGP Paths         : 12          Total Path Memory           : 2400
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
```

```
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.5
                64496       3    0 00h00m13s 0/0/0 (VpnIPv4)
                            3    0           0/0/0 (MvpnIPv4)
-------------------------------------------------------------------------------
*A:PE-1#
```

The same command can be used on the other PEs to verify their BGP sessions to the RR.

Configuring VPRN on PEs

The following outputs show the VPRN configurations for each PE. The specific MVPN configuration is shown later.

PE-1

The VPRN configuration for PE-1 is as follows:

```
*A:PE-1# configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-1-S-1" create
                address 172.16.1.1/24
                sap 1/1/3:1.1 create
                exit
            exit
            interface "RP" create
                address 10.10.10.1/32
                loopback
            exit
```

```
            pim
                apply-to all
                rp
                    static
                        address 10.10.10.1
                            group-prefix 239.0.0.0/8
                        exit
                    exit
                exit
                no shutdown
            exit
        exit
```

There is a single interface toward S-1 from which the multicast group is generated.

If the customer signaling uses PIM ASM, then the customer Rendezvous Point (RP) must be positioned on the sender PE because registration of the source with the RP causes the SA to be sent to the remote source PEs.

A loopback interface called “**RP**” acts as the RP for all group prefixes in the 239.0.0.0/8 range. This will be the RP for the East groups.

PE-2

PE-2 has a receiver attached, and a single interface is configured to accommodate this. The RP configured is that of the East region and has a configuration as follows:

```
*A:PE-2# configure
    service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-2-H-2" create
                address 172.16.2.1/24
                sap 1/1/3:1.1 create
                exit
            exit
            igmp
                interface "int-PE-2-H-2"
                exit
            exit
            pim
                apply-to all
                rp
                    static
                        address 10.10.10.1
                            group-prefix 239.0.0.0/8
                        exit
                    exit
```

```
                exit
                no shutdown
            exit
        exit
```

PE-3

PE-3 serves as the RP for the West region and uses a different IP address for the Rendezvous Point interface.

```
*A:PE-3# configure service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64496:1
            interface "int-PE-3-S-3" create
                address 172.16.3.1/24
                sap 1/1/3:1.1 create
                exit
            exit
            interface "RP" create
                address 10.10.10.3/32
                loopback
            exit
            pim
                apply-to all
                rp
                    static
                        address 10.10.10.3
                            group-prefix 239.0.0.0/8
                        exit
                    exit
                exit
                no shutdown
            exit
        exit
```

PE-4

PE-4 also has a receiver, and uses the West sender PE (PE-3) as the Rendezvous Point.

```
*A:PE-4# configure service
        vprn 1 customer 1 create
            route-distinguisher 64496:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
```

```
            vrf-target target:64496:1
            interface "int-PE-4-H-4" create
                address 172.16.2.1/24
                sap 1/1/3:1.1 create
                exit
            exit
            igmp
                interface "int-PE-4-H-4"
                exit
            exit
            pim
                apply-to all
                rp
                    static
                        address 10.10.10.3
                            group-prefix 239.0.0.0/8
                        exit
                    exit
                exit
                no shutdown
            exit
        exit
```

MVPN Configuration for Source PEs

At the PEs closest to the sources (PE-1 and PE-3), Source Active auto-discovery BGP routes are generated when the source is active.

This applies for PIM-ASM (*,G) joins only, or IGMP (*,G) membership queries received by the provider domain. These are received by all PEs.

Inter-site trees must be disabled for this to occur. Alternatively, inter-site trees can be enabled such that when a source is discovered, a Source Active is advertised to each other PE in the MVPN. This occurs regardless of whether any receivers wish to become members of the multicast groups.

As previously stated, the presence of the SA in the receiver PEs means that no shared joins routes are generated toward the C-RPs.

The MVPN configuration enables BGP as both auto-discovery mechanism and the customer multicast signaling protocol across the VPRN. The provider tunnel between PEs within the MVPN is signaled using Multicast LDP.

The MVPN configuration for each PE should be as follows:

```
*A:PE-1# configure
    service
        vprn 1
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
```

```
            provider-tunnel
                inclusive
                    mldp
                        no shutdown
                    exit
                exit
            exit
        exit
```

The VPRN MVPN configuration for PE-2, PE-3, and PE-4 is identical.

Sender PE Route Policies

The choice of active and standby sources by the receiver PEs is determined by the “best route” policy. PE-1 and PE-3 advertise BGP Source Active Auto Discovery routes when a source is active. This is received by all PEs within the MVPN. As two different sources advertise the same group, it is necessary to differentiate between them.

Assuming that receiver PE-2 prefers the source from PE-1, and PE-4 prefers the source active on PE-3, then the export policy for MVPN routes on PE-1 requires the following steps:

1. Set a community value at PE-1 for the (S,G) multicast group – call this “blue” with value 64496:11.
2. Set the route target community for the VPRN – 64496:1.
3. Create a policy statement that becomes the export policy for MVPN routes within PE-1.
4. Create a policy statement entry (entry 10) that adds the community value “blue” along with the route target for Source Active AD BGP routes. Source Active AD routes are MVPN type 5 routes.
5. Create a policy statement entry default-action that adds the route target for all other MVPN AD BGP routes (such as Intra-AD (type 1)) that are exported to the MVPN PEs.

```
*A:PE-1# configure
    router
        policy-options
            begin
            community "blue" members "64496:11"
            community "MVPN1_RT" members "target:64496:1"
            policy-statement "MVPN1_export"
                entry 10
                    description "match MVPN routes - type 5 Source AD -
                                 add RT and 'blue' community"
                    from
                        mvpn-type 5
                        family mvpn-ipv4
```

```
                    exit
                    action accept
                        community add "blue" "MVPN1_RT"
                    exit
                exit
                default-action accept
                    community add "MVPN1_RT"
                exit
            exit
            commit
        exit
```

6. Apply as an export policy within the MVPN context.

The import policy requires that all imported MVPN BGP routes have the correct route target extended community value, specifically “target:64496:1”.

1. Create a policy statement that becomes the import policy for PE-1.
2. Create a policy statement entry (entry 10) that matches the community of the route target extended community for all MVPN BGP routes. These include the Intra-AD and Source-Join routes.

```
*A:PE-1# configure
    router
        policy-options
            begin
                policy-statement "MVPN1_import"
                    entry 10
                        from
                            community "MVPN1_RT"
                        exit
                        action accept
                        exit
                    exit
                exit
            commit
        exit
```

Enable the inter-site-shared type 5 advertisement persistency so that source ADs are advertised when multicast sources are active. Alternatively, inter-site shared trees can be disabled using the **no intersite shared** command. In this example, only inter-site shared MVPN type 5 persistency is shown.

The additional configuration in the MVPN context is the following.



Note: The PIM instance must be shut down when the intersite-shared configuration is modified.

```
*A:PE-1# configure
    service
        vprn 1
```

```
            mvpn
                intersite-shared persistent-type5-adv
                vrf-import "MVPN1_import"
                vrf-export "MVPN1_export"
            exit
```

For PE-3 (the other sender PE), similar import and export policies are required. In this case, the community will be called “red” and is added to the Source Active AD route generated when the source is active.

The requirements for the export policy for PE-3 are as follows:

```
*A:PE-3# configure
    router
        policy-options
            begin
            community "red" members "64496:33"
            community "MVPN1_RT" members "target:64496:1"
            policy-statement "MVPN1_export"
                entry 10
                    description "match MVPN routes - type 5 Source AD -
                                 add RT and 'red' community"
                    from
                        mvpn-type 5
                        family mvpn-ipv4
                    exit
                    action accept
                        community add "red" "MVPN1_RT"
                    exit
                exit
                default-action accept
                    community add "MVPN1_RT"
                exit
            exit
            commit
        exit
```

The import policy is exactly the same as for PE-1.

Apply the import and export policies to the MVPN context of the sender PE (PE-3) and enable inter-site-shared type 5 advertisement persistency with the same command as on PE-1.

Receiver PE Configuration

PE-2 and PE-4 are the receiver PEs. These will receive the Source Active AD routes and initiate Joins toward the preferred source.

When a Source-Active AD route is received, the community value is examined and the Local Preference value of the route is set using a Route Policy. The preferred source is determined by the SA AD route with the highest Local Preference value.

In the case of PE-2, the preferred source is that advertised by PE-1, the “blue” source as previously referenced. PE-2 sets the Local Preference to 200. The SA AD tagged with the “red” community has the Local Preference set to 50.

For PE-4, the reverse applies: SA AD routes tagged with the “red” community have the Local Preference set to 200, and “blue” SA AD routes have the Local Preference set to 50.

Once again, assuming that the PE-2 receiver prefers the source from PE-1 and PE-4 prefers the source active on PE-3, the import policy for MVPN routes on PE-2 requires the following steps:

1. Set a community value at PE-2 for the (S,G), call this “blue” with value 64496:11.
2. Set the route target community for the VPRN to 64496:1.
3. Create a prefix list that matches the multicast group address, in this case 239.255.0.0/24.
4. Create a policy statement that becomes the import policy for MVPN routes within PE-1.
5. Create a policy statement entry (entry 10) that matches the following attributes:
   - Source Active AD BGP routes type. Source Active AD routes are classed as MVPN type 5 routes, and
   - Community value “blue” AND Route Target extended community, and
   - Group address prefix 239.255.0.0/24

   If the BGP route matches all three conditions, then set the Local Preference to 200.
6. Create a policy statement default-action that accepts all other MVPN BGP routes, including SA routes tagged with the “red” community value.

   The import policy statement looks like:

```
*A:PE-2# configure
    router
        policy-options
            begin
            prefix-list "group_239.255.x.y"
                prefix 239.255.0.0/16 longer
            exit
            community "red" members "64496:33"
            community "blue" members "64496:11"
            community "MVPN1_RT" members "target:64496:1"
            policy-statement "MVPN1_import"
                entry 10
                    description "allow MVPN source-ad - set LP to 200 for 'blue'"
                    from
```

```
                        community expression "[blue] AND [MVPN1_RT]"
                        mvpn-type 5
                        group-address "group_239.255.x.y"
                    exit
                    action accept
                        local-preference 200
                    exit
                exit
                entry 20
                    description "allow MVPN source-ad - set LP to 50 for 'red'"
                    from
                        community expression "[red] AND [MVPN1_RT]"
                        mvpn-type 5
                        group-address "group_239.255.x.y"
                    exit
                    action accept
                        local-preference 50
                    exit
                exit
                default-action accept
                exit
            exit
            commit
```

The export policy for PE-2 MVPN routes requires each MVPN route to be tagged with the route target extended community for VPRN 1. The following policy statement is created:

```
*A:PE-2# configure router
        policy-options
            begin
            policy-statement "MVPN1_export"
                entry 10
                    from
                        family mvpn-ipv4
                    exit
                    action accept
                        community add "MVPN1_RT"
                    exit
                exit
            exit
            commit
```

7. Create a list of redundant sources. This is a list of prefixes that match the source addresses of redundant multicast groups. This is an important parameter because the receiver PEs only create active and standby (S,G) states for groups with source address prefixes that are contained in this list.
8. Before any hosts attempt to join the multicast groups, the decision must be made to enable or disable inter-site shared trees at the receiver PEs. In this example, only the Inter-site shared trees disabled option will be considered. In order to make this configuration change, it is necessary to shut the PIM protocol down before and re-enable when completed.

The additional MVPN configuration for PE-2 is shown in the following output.



Note: The redundant source prefix list is included, and inter-site shared trees are disabled.

```
*A:PE-2# configure
    service
        vprn 1
            mvpn
                no intersite-shared
                red-source-list
                    src-prefix 172.16.1.0/24
                    src-prefix 172.16.3.0/24
                exit
                vrf-import "MVPN1_import"
                vrf-export "MVPN1_export"
            exit
```

PE-4 requires a similar set of import and export policies. In this case, the “red” sources have the highest Local Preference value, based on the community string added by the export policy of PE-3.

```
*A:PE-4# configure
    router
        policy-options
            begin
            prefix-list "group_239.255.x.y"
                prefix 239.255.0.0/16 longer
            exit
            community "red" members "64496:33"
            community "blue" members "64496:11"
            community "MVPN1_RT" members "target:64496:1"
            policy-statement "MVPN1_import"
                entry 10
                    description "allow MVPN source-ad - set LP to 200 for 'red'"
                    from
                        community expression "[red] AND [MVPN1_RT]"
                        mvpn-type 5
                        group-address "group_239.255.x.y"
                    exit
                    action accept
                        local-preference 200
                    exit
                exit
                entry 20
                    description "allow MVPN source-ad - set LP to 50 for 'blue'"
                    from
                        community expression "[blue] AND [MVPN1_RT]"
                        mvpn-type 5
                        group-address "group_239.255.x.y"
                    exit
                    action accept
                        local-preference 50
                    exit
                exit
                default-action accept
```

```
            exit
        exit
        commit
```

The export policy for MVPN routes adds the route target extended community. It is exactly the same export policy as for PE-2.

The additional MVPN configuration for VPRN 1 on PE-4 is identically the same as for PE-2.

Each PE within the MVPN originates an Intra-AD BGP route. This notifies the other PEs within the VPRN. This is used to create a set of Inclusive Provider Multicast Service Interfaces (I-PMSI) between each PE. In this case, I-PMSIs are signaled using mLDP.

Using PE-1 as an example, the set of Intra-AD routes can be seen using the following command:

```
*A:PE-1# show router bgp routes mvpn-ipv4
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                    OriginatorIP           LocalPref   MED
      RD                           SourceAS                           Label
      Nexthop                      SourceIP
      As-Path                      GroupIP
-------------------------------------------------------------------------------
i     Intra-Ad                     192.0.2.1              100         0
      64496:1                      -                                  -
      192.0.2.1                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.2              100         0
      64496:1                      -                                  -
      192.0.2.2                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.3              100         0
      64496:1                      -                                  -
      192.0.2.3                    -
      No As-Path                   -
u*>i  Intra-Ad                     192.0.2.4              100         0
      64496:1                      -                                  -
      192.0.2.4                    -
      No As-Path                   -
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-1#
```

At this moment, there are no connected sources detected and no receivers wishing to join any multicast sources.

Each I-PMSI is seen as a PIM tunnel interface. As there are four routers in the MVPN, there are four I-PMSIs.

```
*A:PE-1# show router 1 pim tunnel-interface
===============================================================================
PIM Interfaces ipv4
===============================================================================
Interface                       Originator Address   Adm  Opr  Transport Type
-------------------------------------------------------------------------------
mpls-if-73730                   192.0.2.1            Up   Up   Tx-IPMSI
mpls-if-73731                   192.0.2.3            Up   Up   Rx-IPMSI
mpls-if-73732                   192.0.2.4            Up   Up   Rx-IPMSI
mpls-if-73733                   192.0.2.2            Up   Up   Rx-IPMSI
-------------------------------------------------------------------------------
Interfaces : 4
===============================================================================
*A:PE-1#
```

In order to be able to reach the source, a route for each source is included in the VRF for VPRN 1.

For PE-2, this looks as follows:

```
*A:PE-2# show router 1 route-table

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.10.10.1/32                                   Remote  BGP VPN   00h19m29s  170
       192.0.2.1 (tunneled)                                          0
10.10.10.3/32                                   Remote  BGP VPN   00h19m29s  170
       192.0.2.3 (tunneled)                                          0
172.16.1.0/24                                   Remote  BGP VPN   00h19m29s  170
       192.0.2.1 (tunneled)                                          0
172.16.2.0/24                                   Local   Local     06h01m20s  0
       int-PE-2-H-2                                                  0
172.16.3.0/24                                   Remote  BGP VPN   00h19m29s  170
       192.0.2.3 (tunneled)                                          0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-2#
```

The sources at 172.16.1.0/24 and 172.16.3.0/24 are learned as BGP VPN routes.

The following output shows the BGP routes for these prefixes, for example for prefix 172.16.1.0/24 on PE-2:

```
*A:PE-2# show router bgp routes 172.16.1.0/24 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 172.16.1.0/24
Nexthop        : 192.0.2.1
Route Dist.    : 64496:1                VPN Label      : 262138
Path Id        : None
From           : 192.0.2.5
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-2-P-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64496:1 l2-vpn/vrf-imp:192.0.2.1:3
                 source-as:64496:0
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h30m08s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-2#
```

This prefix is advertised with three communities:

- A route target extended community
- An l2-vpn/vrf-import extended community.

- A source-AS extended community (not used in Intra-AS context).

The l2-vpn/vrf-import extended community is significant as it is a unique value. It represents a specific MVPN on a specific PE and is comprised of a 32 bit value that identifies the PE plus an index identifying the VRF. The 32 bit value is the system address. The index (3) can be derived from the command:

```
*A:PE-2# admin display-config index | match vprn1
        virtual-router "vprn1" 3 0
```

Therefore, the l2-vpn/vrf-import community for VPRN 1 on PE-1 is 192.0.2.1:3

This community attribute is included within the source-join BGP route that is sent in a BGP update by a receiver PE as it tries to join a multicast group with a source address that matches the 172.16.1.0/24 prefix. This ensures that the source-join route is only accepted as a valid route and imported by the PE that originated the source address prefix. This is explained in the following section.

Enable Redundant Sources

The redundant sources are now enabled so that multicast traffic flows into both PE-1 and PE-3, using groups (S-1,G-1) and (S-3,G-1), respectively.

On each of these PEs, a source active AD route is generated. By examining each receiver PE, these can be clearly seen.

For PE-2, the source active AD routes can be seen using the following command.

```
*A:PE-2# show router bgp routes mvpn-ipv4 type source-ad
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
Flag  RouteType                   OriginatorIP           LocalPref   MED
      RD                          SourceAS                           Label
      Nexthop                     SourceIP
      As-Path                     GroupIP
-------------------------------------------------------------------------------
u*>i  Source-Ad                   -                      200         0
      64496:1                     -                                  -
      192.0.2.1                   172.16.1.2
      No As-Path                  239.255.0.1
u*>i  Source-Ad                   -                      50          0
      64496:1                     -                                  -
```

```
      192.0.2.3                   172.16.3.2
      No As-Path                  239.255.0.1
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-2#
```

There are two routes present, one from each source for the same group from PE-1 and PE-3.

The PIM groups can now be seen on PE-1 as follows:

```
*A:PE-1# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (S,G)                          int-PE-1-S-1   0
   172.16.1.2                  10.10.10.1
239.255.0.1                 (S,G)                          mpls-if-73734  0
   172.16.3.2                  10.10.10.1
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-1#
```

There are two groups at PE-1. In addition to its locally connected source, PE-1 has also received a source active from PE-3 which has an incoming interface of the I-PMSI toward PE-3. The outgoing interface list is empty as there is no host wishing to become a group member.

Similarly, on the other sender, PE-3.

```
*A:PE-3# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (S,G)                          mpls-if-73735  0
   172.16.1.2                  10.10.10.3
239.255.0.1                 (S,G)                          int-PE-3-S-3   0
   172.16.3.2                  10.10.10.3
-------------------------------------------------------------------------------
Groups : 2
===============================================================================
*A:PE-3#
```

By examining the receiver PE-2, it can be seen that the Source AD route for (S,G) (172.16.1.2, 239.255.0.1) from PE-1 has a higher local preference so it is chosen as the preferred (active) source. Examining these routes in more detail shows that each route is tagged with two communities: the route target extended community and the “red” or “blue” community, as seen in the following output.

```
*A:PE-2# show router bgp routes mvpn-ipv4 type source-ad hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Route Type     : Source-Ad
Route Dist.    : 64496:1
Source IP      : 172.16.1.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.1
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 200                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : 64496:11 no-export target:64496:1
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h05m54s
VPRN Imported  :  1

Route Type     : Source-Ad
Route Dist.    : 64496:1
Source IP      : 172.16.3.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.3
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 50                     Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
```

```
Community      : 64496:33 no-export target:64496:1
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.3              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h27m35s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The local preference is set based on these community values.

A debug of the received BGP Source AD routes is as follows for PE-2:

```
1 2016/10/13 09:32:19.25 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 29 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.1
        Type: Source-AD Len: 18 RD: 64496:1 Src: 172.16.1.2 Grp: 239.255.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 8 Community:
        64496:11
        no-export
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.1
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:1
"


2 2016/10/13 09:32:19.25 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 29 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.3
        Type: Source-AD Len: 18 RD: 64496:1 Src: 172.16.3.2 Grp: 239.255.0.1
```

```
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 8 Community:
        64496:33
        no-export
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:1
"
```

Similarly, the Source Active routes on receiver PE-4 show that the highest local preference value of 200 is set for the SA route received from PE-3 with an originator ID of 192.0.2.3, as follows:

```
*A:PE-4# show router bgp routes mvpn-ipv4 type source-ad hunt
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Route Type     : Source-Ad
Route Dist.    : 64496:1
Source IP      : 172.16.1.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.1
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 50                     Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : 64496:11 no-export target:64496:1
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h12m47s
VPRN Imported  :  1
```

```
Route Type     : Source-Ad
Route Dist.    : 64496:1
Source IP      : 172.16.3.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.3
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 200                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : 64496:33 no-export target:64496:1
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.3              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h00m14s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

Host Group Membership

If the hosts then send a (*,G) request to join the group, a source-join route is originated at each receiver PE toward the preferred source from the redundant list.

The following output shows a join originated by PE-2:

```
*A:PE-2# show debug
debug
    router "Base"
        bgp
            update neighbor 192.0.2.5
        exit
    exit
exit

30 2016/10/13 11:34:31.47 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 84
```

```
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 192.0.2.2
        Type: Source-Join Len:22 RD: 64496:1 SrcAS: 64496 Src: 172.16.1.2
                          Grp: 239.255.0.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64496:1
        target:192.0.2.1:3
"
```

When an active source AD route is present, there is no shared join sent to the RP. Because the source address is known, only a source-join needs to be sent. The source-join is trying to become a member of group 239.255.0.1 with a source address of 172.16.1.2. As this is sent as a BGP routing update, this must be accepted by the MVPN VRF at the PE that originated the unicast route that represents the c-multicast source. As previously mentioned, there are two extended community values. The second of these is the l2-vpn/vrf-import route target for 192.0.2.1 (PE-1), so only PE-1 will accept this route.

Examining the PIM state table for PE-2 shows the presence of a group with multiple sources.

```
*A:PE-2# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (*,G)                          mpls-if-73735  1
   *                           10.10.10.1
239.255.0.1                 (S,G)               spt        mpls-if-73735  1
   172.16.1.2                  10.10.10.1          A
239.255.0.1                 (S,G)                          mpls-if-73733  1
   172.16.3.2                  10.10.10.1          S
-------------------------------------------------------------------------------
Groups : 3
===============================================================================
*A:PE-2#
```

Each (S,G) has a state of either Active (A) or Standby (S), and the active group is chosen based on the Source Active AD with the highest local preference.

As a direct comparison, PE-4 also has the same two (S,G) states, but has a reversed active and standby source.

```
*A:PE-4# show router 1 pim group
===============================================================================
Legend:  A = Active   S = Standby
===============================================================================
PIM Groups ipv4
===============================================================================
Group Address               Type                Spt Bit  Inc Intf       No.Oifs
   Source Address              RP                  State    Inc Intf(S)
-------------------------------------------------------------------------------
239.255.0.1                 (*,G)                          mpls-if-73734  1
   *                           10.10.10.3
239.255.0.1                 (S,G)                          mpls-if-73733  1
   172.16.1.2                  10.10.10.3          S
239.255.0.1                 (S,G)               spt        mpls-if-73734  1
   172.16.3.2                  10.10.10.3          A
-------------------------------------------------------------------------------
Groups : 3
===============================================================================
*A:PE-4#
```

The Source Active ADs received on PE-4 have their local preference values based on the community string value.

```
*A:PE-4# show router bgp routes mvpn-ipv4 type source-ad hunt
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Route Type     : Source-Ad
Route Dist.    : 64496:1
Source IP      : 172.16.1.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.1
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 50                      Interface Name : NotAvailable
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : 0
AIGP Metric    : None
Connector      : None
Community      : 64496:11 no-export target:64496:1
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.1               Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
```

```
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h17m46s
VPRN Imported  :  1

Route Type     : Source-Ad
Route Dist.    : 64496:1
Source IP      : 172.16.3.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.3
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 200                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : 64496:33 no-export target:64496:1
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.3              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m30s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-4#
```

Sender PE MVPN Status

The MVPN status of the PE-1 sender PE is as follows:

```
*A:PE-1# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                  auto-discovery     : Default
UMH Selection      : Highest-Ip           SA withdrawn       : Disabled
intersite-shared   : Enabled              Persist SA         : Enabled
vrf-import         : MVPN1_import
vrf-export         : MVPN1_export
vrf-target         : N/A
C-Mcast Import RT  : target:192.0.2.1:3
```

```
ipmsi              : ldp
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-if-73730
Mdt-type           : sender-receiver

BSR signalling     : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
*A:PE-1#
```

The C-Mcast Import RT is set to <system-address>:<VPRN index>.

The VPRN index is derived from the following command:

```
*A:PE-1# admin display-config index | match vprn1
        virtual-router "vprn1" 3 0
```

Any Source Join received must include this attribute along with the route target extended community. As previously stated, this is advertised within the VPN-IPv4 routes as a BGP attribute.

A source join received from PE-2 to join (S,G) (172.16.1.2, 239.255.0.1) is as follows:

```
*A:PE-1# show router bgp routes mvpn-ipv4 type source-join hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Route Type     : Source-Join
Route Dist.    : 64496:1
Source AS      : 64496
Source IP      : 172.16.1.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.2
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      :  no-export target:64496:1 target:192.0.2.1:3
```

```
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.2              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h14m05s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

The PIM status for this group on sender PE-1 is as follows:

```
*A:PE-1# show router 1 pim group 239.255.0.1 source 172.16.1.2 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 172.16.1.2
RP Address         : 10.10.10.1
Advt Router        : 192.0.2.1
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.1.2
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:24:20        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.1.2
Incoming Intf      : int-PE-1-S-1
Outgoing Intf List : mpls-if-73730

Curr Fwding Rate   : 8113.3 kbps
Forwarded Packets  : 13322359           Discarded Packets  : 0
Forwarded Octets   : 612828514          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-1#
```

The outgoing interface list is the I-PMSI, and traffic is seen to be flowing because the current forwarding rate is non-zero.

Similarly for sender PE-3, the MVPN status is:

```
*A:PE-3# show router 1 mvpn
===============================================================================
MVPN 1 configuration data
===============================================================================
signaling          : Bgp                   auto-discovery     : Default
UMH Selection      : Highest-Ip            SA withdrawn       : Disabled
intersite-shared   : Enabled               Persist SA         : Enabled
vrf-import         : MVPN1_import
vrf-export         : MVPN1_export
vrf-target         : N/A
C-Mcast Import RT  : target:192.0.2.3:2

ipmsi              : ldp
i-pmsi P2MP AdmSt  : Up
i-pmsi Tunnel Name : mpls-if-73729
Mdt-type           : sender-receiver

BSR signalling     : none
Wildcard s-pmsi    : Disabled
Multistream-SPMSI  : Disabled
s-pmsi             : none
data-delay-interval: 3 seconds
enable-asm-mdt     : N/A

===============================================================================
*A:PE-3#
```

The Source-Join route on PE-3 for this multicast group is:

```
*A:PE-3# show router bgp routes mvpn-ipv4 type source-join hunt
===============================================================================
 BGP Router ID:192.0.2.3        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MVPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Route Type     : Source-Join
Route Dist.    : 64496:1
Source AS      : 64496
Source IP      : 172.16.3.2
Group IP       : 239.255.0.1
Nexthop        : 192.0.2.4
From           : 192.0.2.5
Res. Nexthop   : 0.0.0.0
```

```
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : no-export target:64496:1 target:192.0.2.3:2
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.4              Peer Router Id : 192.0.2.5
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h44m35s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-3#
```

The PIM state for this group is as follows:

```
*A:PE-3# show router 1 pim group 239.255.0.1 source 172.16.3.2 detail
===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 239.255.0.1
Source Address     : 172.16.3.2
RP Address         : 10.10.10.3
Advt Router        : 192.0.2.3
Flags              : spt                Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.3.2
MRIB Src Flags     : direct
Keepalive Timer Exp: 0d 00:00:53
Up Time            : 0d 00:15:18        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.3.2
Incoming Intf      : int-PE-3-S-3
Outgoing Intf List : mpls-if-73729

Curr Fwding Rate   : 7765.2 kbps
Forwarded Packets  : 13459708           Discarded Packets  : 0
Forwarded Octets   : 619146568          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
```

```
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
*A:PE-3#
```

The preferred source remains active unless:

- The multicast source ceases to exist, the source PE withdraws the Source Active AD route
- Or a Source Active AD is received with a higher local preference.

Conclusion

MVPN Source Redundancy provides an optimal solution for multicast routing in a VPRN. This protocol provides simple configuration, operation and guaranteed fast protection time. It could be utilized in a regionalized IPTV solution where multiple sources for the same TV channel are used.

Spoke Termination for IPv6-6VPE

This chapter provides information about spoke termination for IPv6-6VPE.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR-OS 9.0. The CLI in the current edition corresponds to 14.0.R5.

Overview

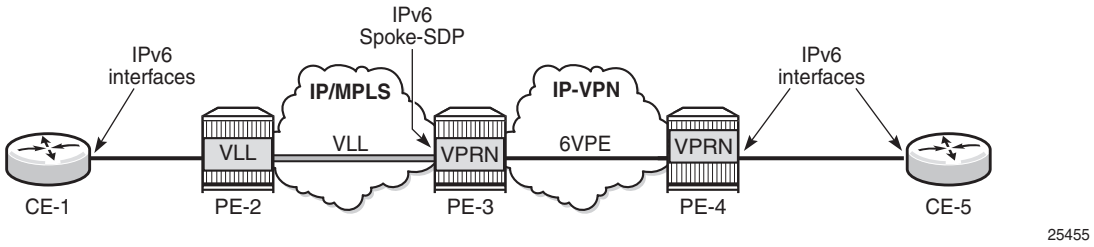
RFC 4659, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*, standardized the use of an IPv6 over IPv4 tunneling scheme. SR OS supports the standardized IPv6 over IPv4 tunneling scheme for Virtual Private Routed Network (VPRN) services using Multi-Protocol Border Gateway Protocol (MP-BGP), also known as 6VPE. The 7750 SR also supports pseudowire termination by a VPRN from an Epipe Virtual Leased Line (VLL) or VPLS spoke SDP where the pseudowire can be given IPv6 addresses and run IPv6 protocols. In the example used in this chapter, any advertisements across the Multi-Protocol Labeled Switching (MPLS) network between Virtual Private Routed Network (VPRN) Provider Edge (PE) devices will use 6VPE. The goal of this section is to list configuration guidelines for IPv6 spoke termination to a VPRN over an Epipe VLL and transporting IPv6 packets over 6VPE tunnels between PE devices.

This solution is to be used where a service provider is providing VPRN services built on a transport network whose Interior Gateway Protocol (IGP) is using IPv4 addressing on the network interfaces. The customer's CE and the service provider's PE must support IPv6 pseudowires, IPv6 interfaces and in addition, the service provider also must be able to support the advertisements of IPv6 prefixes between CE-PE peerings and between the transport PE routers using MP-BGP. The advertisement of IPv6 prefixes across the MPLS network and the transport of IPv6 traffic is tunneled using 6VPE.

The VPRN PE has the ability to support spoke termination of Epipe VLL services on access with IPv6 addressing between the CE and VPRN PE. The IPv6 spoke termination on VPRN services has the same functionality as VPRN IPv4 spoke termination.

The example in Figure 360 illustrates a CE device that connects to a VPRN PE on an IPv6 interface addressing using spoke termination.

Figure 360 **Spoke Termination for IPv6**



CE-1 is connected to the VPRN service on PE-3, using IPv6 interfaces. CE-1 reaches PE-3 by connecting to PE-2. PE-2 uses an Epipe VLL for transport to the VPRN on the PE-3. The connectivity between the VLL service on the VPRN service on PE-3 is using spoke termination with IPv6 addressing on the spoke-service distribution point (spoke SDP) interface on PE-3.

Figure 361 **IPv6 Addressing and IPv6 Prefixes**

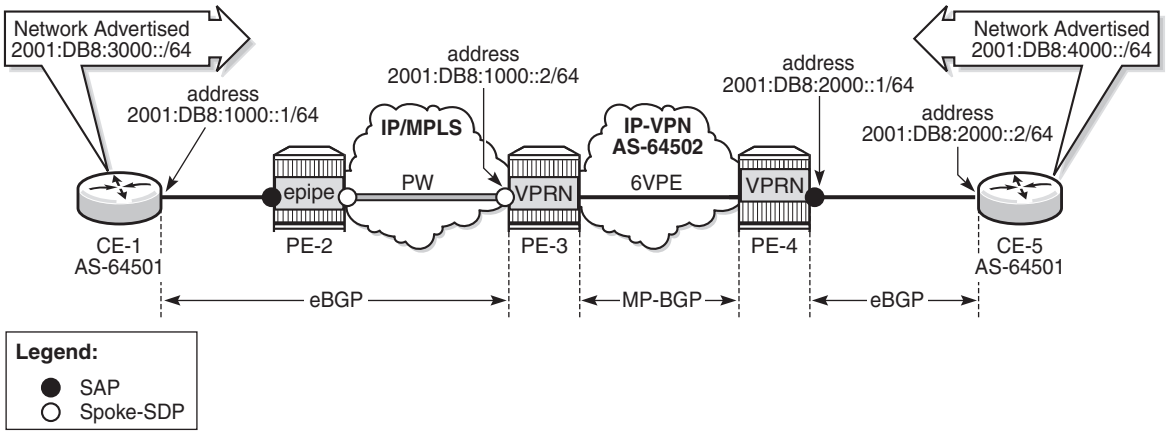


Figure 361 shows the overall IPv6 addressing from interfaces to prefixes advertised from CE-1 and CE-5 across the VPRN network.

- Link between CE-1 and PE-3: 2001:DB8:1000::/64
- Link between CE-5 and PE-4: 2001:DB8:2000::/64
- Advertised Prefix from CE-1: 2001:DB8:3000::/64
- Advertised Prefix from CE-5: 2001:DB8:4000::/64

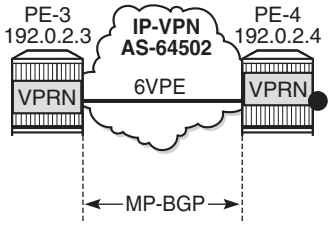
PE-3 has an MP-eBGP session with CE-1 to receive and advertise IPv6 routes. PE-3 also has an MP-iBGP peering session with PE-4 to use 6VPE to tunnel IPv6 routes and traffic to and from PE-4. PE-4 has an IPv6 SAP interface to CE-5 and uses MP-eBGP to advertise and receive routes to/from CE-5 (no spoke-termination). The configuration of PE-3 will be included to provide examples of the end-to-end VPRN service using a 6VPE model.

This network topology will illustrate the use of spoke termination using IPv6 interfaces and the tunneling of IPv6 traffic over a 6VPE MPLS network.

Configuration

First an MPLS network is established where the VPRN service can use 6VPE to tunnel traffic across the IPv4 IGP.

Figure 362 **MP-BGP VPN IPv6**



PE-3 and PE-4 in Figure 362 are edge routers running VPRN Services on access with IPv6 Interfaces. The MPLS network is configured using IPv4 link addressing. Interior Border Gateway Protocol (iBGP) peerings need to be established with MP-BGP for VPN-IPv6 address families between PE-3 and PE-4.

```
*A:PE-3# configure
    router
        autonomous-system 64502
        bgp
            group "iBGP"
                description "iBGP peering in AS 64502"
                family vpn-ipv6
                type internal
                neighbor 192.0.2.4
                    description "PE-4"
                exit
            exit
        exit
    exit
exit

*A:PE-4# configure
    router
        autonomous-system 64502
        bgp
            group "iBGP"
                description "iBGP peering in AS 64502"
                family vpn-ipv6
                type internal
                neighbor 192.0.2.3
                    description "PE-3"
                exit
            exit
        exit
    exit
exit
```

Configuring address family vpn-ipv6 between VPRN PE edge routers in BGP enables the functionality of MP-BGP for the Layer 3 VPNs supporting the customer's IPv6 Addressing (6VPE).

Verify BGP sessions for VPN-IPv6 address families between PE-3 and PE-4.

```
*A:PE-3# show router bgp neighbor 192.0.2.4
===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                 : 192.0.2.4
Description          : PE-4
Group                : iBGP
-------------------------------------------------------------------------------
Peer AS              : 64502            Peer Port            : 50222
Peer Address         : 192.0.2.4
Local AS             : 64502            Local Port           : 179
Local Address        : 192.0.2.3
Peer Type            : Internal         Dynamic Peer         : No
State                : Established      Last State           : Established
Last Event           : recvKeepAlive
Last Error           : Cease (Connection Collision Resolution)
Local Family         : VPN-IPv6
Remote Family        : VPN-IPv6
---snip---

*A:PE-4# show router bgp neighbor 192.0.2.3
===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                 : 192.0.2.3
Description          : PE-3
Group                : iBGP
-------------------------------------------------------------------------------
Peer AS              : 64502            Peer Port            : 179
Peer Address         : 192.0.2.3
Local AS             : 64502            Local Port           : 50222
Local Address        : 192.0.2.4
Peer Type            : Internal         Dynamic Peer         : No
State                : Established      Last State           : Active
Last Event           : recvKeepAlive
Last Error           : Unrecognized Error
Local Family         : VPN-IPv6
Remote Family        : VPN-IPv6
---snip---
```

After the MP-BGP sessions are established for VPN-IPv6 address-family, 6VPE tunnel support is provided between PE-3 and PE-4.

Figure 363 **Spoke Termination for IPv6 Addressing**

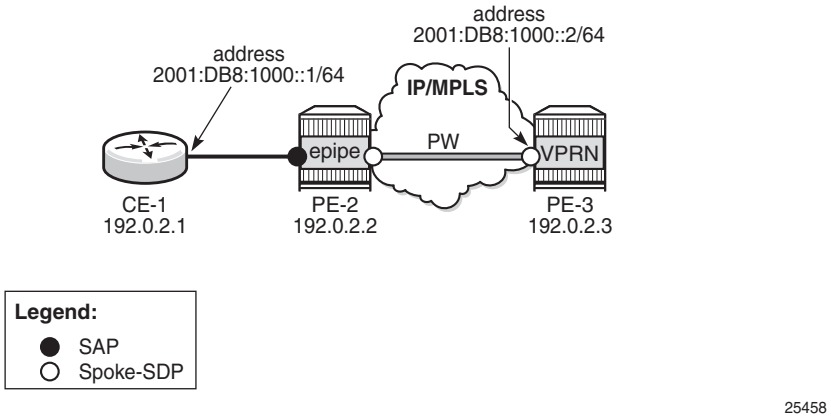


Figure 363 illustrates the model for spoke termination for IPv6 using VPRN services. CE-1 is configured with IPv6 addressing on the access interface facing the VPRN service. CE-1's access is backhauled to the VPRN service on PE-3 using Epipe VLL with spoke termination. Only Epipe VLL is supported for IPv6 spoke termination within the VPRN in R8.0. The configuration of the Epipe VLL on PE-2 is as follows:

```
*A:PE-2# configure
    service
        sdp 231 mpls create
            far-end 192.0.2.3
            ldp
            no shutdown
        exit
        epipe 1 customer 1 create
            sap 1/1/2 create
            exit
            spoke-sdp 231:1 create
            exit
            no shutdown
        exit
```

The example is taken from PE-2 which has been configured using the Epipe VLL service with a SAP interface facing the customer and a spoke SDP facing PE-3. The spoke SDP is terminated into the customer's VPRN service on PE-3.

PE-3 is now ready to be configured for the IPv6 spoke SDP. Review the possible IPv6 options for spoke SDP interfaces on the CLI for VPRN Services (compliant with RFC 4213, *Basic Transition Mechanisms for IPv6 Hosts and Routers <draft-ietf-v6ops-mech-v2-07.txt>)*:

- Interface spoke SDP (IPv6 options only)

```
*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2"
---snip---
```

```
[no] ipv6            + Enables/Configures IPv6 for a VPRN interface
---snip---


*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2" ipv6
  - ipv6
  - no ipv6

 [no] address         - Assigns an IPv6 address to the VPRN interface.
 [no] bfd             - Configure BFD parameters
 [no] dad-disable     - Disable Duplicate Address Detection
 [no] dhcp6-relay     + Configure DHCPv6 relay parameters for the VPRN interface
 [no] dhcp6-server    + Configure DHCPv6 server parameters for the VPRN interface
      icmp6           + Configure ICMPv6 parameters for the VPRN interface
 [no] link-local-add* - Configure link-local address
 [no] local-dhcp-ser* - Assign a DHCP server to the interface
 [no] local-proxy-nd  - Enable/disable local proxy Neighbor Discovery on the VPRN
                        interface
 [no] neighbor        - Configure IPv6-to-MAC address mapping on the VPRN interface
 [no] neighbor-limit  - Configures the maximum amount of IPv6 neighbor entries
 [no] proxy-nd-policy - Configure a proxy Neighbor Discovery policy for the VPRN
                        interface
 [no] qos-route-look* - Enable/Disable Qos route lookup for the interface
 [no] reachable-time  - Configure neighbor reachability detection timer
 [no] secure-nd       + Configure Secure Neighbor Discovery (SEND) parameters for
                        the interface
 [no] stale-time      - Configure the time a neighbor discovery cache entry can
                        remain stale before being removed
 [no] tcp-mss         - Configure TCP maximum segment size for the interface
 [no] urpf-check      + Enables/Configures unicast RPF check for an interface
 [no] vrrp            + Context to create and configure VRRP virtual router instance
                        on the interface
```

- IPv6 address

```
*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2" ipv6 address
  - address <ipv6-address/prefix-length> [eui-64] [preferred]
            [track-srrp <srrp-instance>] [modifier <cga-modifier>]
  - no address <ipv6-address/prefix-length>

 <ipv6-address/pref*> : ipv6-address  x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                      x:x:x:x:x:x:d.d.d.d
                                      x [0..FFFF]H
                                      d [0..255]D
                                      (no multicast address)
                        prefix-length [1..128]
 <eui-64>             : keyword
 <preferred>          : keyword
 <srrp-instance>      : [1..4294967295]
 <cga-modifier>       : [0x0..0xFFFFFFFF...(32 hex nibbles)]
```

- DHCPv6 relay parameters for the VPRN service (default settings)

```
*A:PE-3>config>service>vprn>if>ipv6>dhcp6-relay# info detail
----------------------------------------------
                        shutdown
                        no description
                        no lease-populate
                        no neighbor-resolution
```

```
                        option
                            no interface-id
                            no remote-id
                        exit
                        no source-address
                        no link-address
                        no user-db
                        no python-policy
                        no server
```

- DHCPv6 server parameters for the VPRN service (default)

```
*A:PE-3>config>service>vprn>if>ipv6>dhcp6-server# info detail
----------------------------------------------
                        prefix-delegation
                            shutdown
                        exit
                        max-nbr-of-leases 8000
```

- ICMPv6 (default)

```
*A:PE-3>config>service>vprn>if>ipv6>icmp6# info detail
----------------------------------------------
                        packet-too-big 100 10
                        param-problem 100 10
                        redirects 100 10
                        time-exceeded 100 10
                        unreachables 100 10
```

- Link-local-addressing, for the VPRN interface. By default, link-local addressing is assigned dynamically. Use this command if you would like to add a static link-local-address.

```
*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2" ipv6 link-local-address
  - link-local-address <ipv6-address> [preferred]
  - no link-local-address

 <ipv6-address>       : ipv6-address   - x:x:x:x:x:x:x:x
                                         x:x:x:x:x:x:d.d.d.d
                                         x [0..FFFF]H
                                         d [0..255]D
 <preferred>          : keyword
```

- Neighbor

  * IPv6 to MAC address mapping on the VRPN Interface

```
*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2" ipv6 neighbor
  - neighbor <ipv6-address> <mac-address>
  - no neighbor <ipv6-address>

 <ipv6-address>       : x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                    x:x:x:x:x:x:d.d.d.d
                                    x [0..FFFF]H
                                    d [0..255]D
                        prefix-length [1..128]
 <mac-address>        : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx
```

- Enabling Local Proxy Neighbor Discovery

```
*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2" ipv6 local-proxy-nd
  - local-proxy-nd
  - no local-proxy-nd
```

- VRRP

```
*A:PE-3# configure service vprn 1 interface "int-PE-3-PE-2" ipv6 vrrp
  - no vrrp <virtual-router-id>
  - vrrp <virtual-router-id> [owner]

 <virtual-router-id>  : [1..255]
 <owner>              : keyword

 [no] backup          - Configure virtual router IP addresses for the interface
 [no] bfd-enable      - Configure a BFD interface
 [no] init-delay      - Configure VRRP initialization delay timer
 [no] mac             - Configure a Virtual MAC address to use in Neighbor Discovery
 [no] master-int-inh* - Allow/disallow the master instance to dictate the master
                        down timer (non-owner context only)
 [no] message-interv* - Configure the interval for sending VRRP Advertisement
                        messages
 [no] ping-reply      - Allow/disallow non-owner master to reply to ICMP Echo
                        requests (non-owner context only)
 [no] policy          - Associate a VRRP Priority Control Policy with the virtual
                        router instance (non-owner context only)
 [no] preempt         - Allow/disallow the virtual router instance to override an
                        existing non-owner master (non-owner context only)
 [no] priority        - Configure the base priority for the virtual router
                        instance (non-owner context only)
 [no] shutdown        - Administratively enable/disable the virtual router
                        instance (non-owner context only)
 [no] standby-forwar* - Allow/disallow the forwarding of packets by a standby router
 [no] telnet-reply    - Allow/disallow non-owner master to reply to Telnet requests
                        (non-owner context only)
 [no] traceroute-rep* - Allow/disallow non-owner master to reply to traceroute
                        requests (non-owner context only)
```

The VPRN on PE-3 will export IPv6 routes (IPv6 route on CE-5) to CE-1. A route policy needs to be configured.

```
*A:PE-3# configure
    router
        policy-options
            begin
            prefix-list "PE-3-CE-1"
                prefix 2001:db8:4000::/64 exact
            exit
            policy-statement "PE-3-BGP-CE-1"
                entry 10
                    from
                        prefix-list "PE-3-CE-1"
                    exit
                    action accept
                    exit
                exit
                default-action drop
```

```
                exit
            exit
            commit
        exit
    exit
exit
```

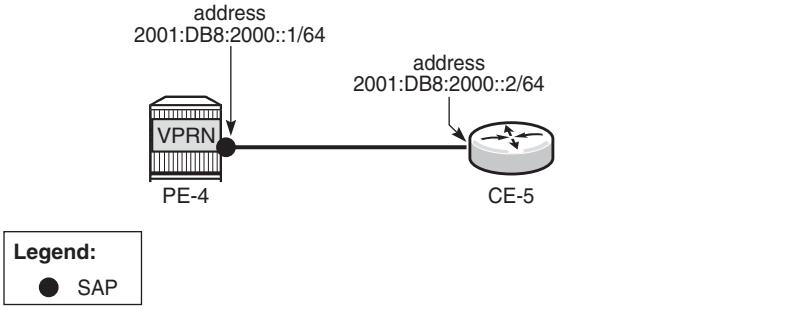
The configuration for the VPRN service on PE-3 with IPv6 interface (spoke SDP) in reference to Figure 363:

```
*A:PE-3# configure
    service
        sdp 321 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        vprn 1 customer 1 create
            router-id 192.0.2.6
            autonomous-system 64502
            route-distinguisher 64502:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64502:1
            interface "loopback" create
                address 192.0.2.6/32
                loopback
            exit
            interface "int-PE-3-PE-2" create
                description "Spoke SDP"
                ipv6
                    address 2001:db8:1000::2/64
                exit
                spoke-sdp 321:1 create
                exit
            exit
            bgp
                router-id 192.0.2.6
                group "Spoke-CE-1-PE-3"
                    family ipv6
                    peer-as 65501
                    local-address 2001:db8:1000::2
                    neighbor 2001:db8:1000::1
                        as-override
                        type external
                        export "PE-3-BGP-CE-1"
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
exit
```

In the preceding configuration example, PE-3 has been configured with an IPv6 spoke SDP (spoke termination) with spoke interface int-PE-3-PE-2. The VPRN configuration has also been set up for MP-eBGP peering to CE-1 through the IPv6 spoke interface. The MP-eBGP peering will be receiving and advertising IPv6 prefixes from/to CE-1. Route policy configuration has been included to show how IPv6 routes are advertised to CE-1 from PE-3 (policy-statement PE-3-BGP-CE-1).

The configuration on PE-4 is similar, but there is a SAP interface to CE-5 instead of a spoke-SDP.

Figure 364 **PE-4 VPRN with SAP to CE-5**



The IPv6 configuration options for the SAP interface (int-PE-4-CE-5) are similar to those in the preceding example for the spoke SDP on PE-3. The PE-4 BGP export policy (PE-4-BGP-CE-5) is also similar to the example for PE-3 in advertising the learned IPv6 route to CE-5.

```
*A:PE-4# configure
    router
        policy-options
            begin
            prefix-list "PE-4-CE-5"
                prefix 2001:db8:3000::/64 exact
            exit
            policy-statement "PE-4-BGP-CE-5"
                entry 10
                    from
                        prefix-list "PE-4-CE-5"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
        exit
    exit
exit
```

```
*A:PE-4# configure
    service
        vprn 1 customer 1 create
            router-id 192.0.2.7
            autonomous-system 64502
            route-distinguisher 64502:1
            auto-bind-tunnel
                resolution-filter
                    ldp
                exit
                resolution filter
            exit
            vrf-target target:64502:1
            interface "loopback" create
                address 192.0.2.7/32
                loopback
            exit
            interface "int-PE-4-CE-5" create
                ipv6
                    address 2001:db8:2000::1/64
                exit
                sap 1/1/1 create
                exit
            exit
            bgp
                router-id 192.0.2.7
                group "CE-5-PE-4"
                    family ipv6
                    peer-as 64501
                    local-address 2001:db8:2000::1
                    neighbor 2001:db8:2000::2
                        as-override
                        type external
                        export "PE-4-BGP-CE-5"
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
exit
```

In this setup, the configuration on CE-1 is as follows.

```
*A:CE-1# configure
    router
        static-route-entry 2001:db8:3000::/64
            black-hole
                no shutdown
            exit
        exit
        policy-options
            begin
            prefix-list "CE-1-192.0.2.1"
                prefix 2001:db8:3000::/64 exact
            exit
            policy-statement "CE-1-sys-to-eBGP"
```

```
                entry 10
                    from
                        prefix-list "CE-1-192.0.2.1"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
        exit
        bgp
            router-id 192.0.2.1
            group "eBGP_to_64502"
                description "eBGP_to_PE-3_AS64502"
                family ipv6
                type external
                peer-as 64502
                local-address 2001:db8:1000::1
                neighbor 2001:db8:1000::2
                    export "CE-1-sys-to-eBGP"
                exit
            exit
            no shutdown
        exit
    exit
exit

*A:CE-1# configure
    service
        ies 1 customer 1 create
            interface "int-CE-1-PE-2" create
                description "SAP_toward_VPRN_Service"
                ipv6
                    address 2001:db8:1000::1/64
                exit
                sap 1/1/1 create
                exit
            exit
            no shutdown
        exit
    exit
exit
```

The configuration on CE-5 is similar.

The following command on PE-2 shows that the Epipe VLL is established with the SAP facing CE-1 and spoke SDP facing VPRN on PE-3.

```
*A:PE-2# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
```

```
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 10/20/2016 12:54:38
Last Mgmt Change  : 10/20/2016 12:54:11
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 9190
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Force QTag Fwd    : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                             Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/2                              null         9212    9212    Up   Up
sdp:231:1 S(192.0.2.3)                 Spok         0       9190    Up   Up
===============================================================================
*A:PE-2#
```

The same command can be launched on PE-3 to verify that the VPRN service is up and that the spoke SDP is up (admin state up/oper state up).

```
*A:PE-3# show service id 1 base
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPRN
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 10/20/2016 12:56:13
Last Mgmt Change  : 10/20/2016 12:56:13
Admin State       : Up                  Oper State        : Up

Route Dist.       : 64502:1             VPRN Type         : regular
Oper Route Dist   : 64502:1
Oper RD Type      : configured
AS Number         : 64502               Router Id         : 192.0.2.6
ECMP              : Enabled             ECMP Max Routes   : 1
Max IPv4 Routes   : No Limit

Auto Bind Tunnel
Resolution        : filter
Filter Protocol   : ldp

Max IPv6 Routes   : No Limit
Ignore NH Metric  : Disabled
Hash Label        : Disabled
Entropy Label     : Disabled
Vrf Target        : target:64502:1
Vrf Import        : None
Vrf Export        : None
MVPN Vrf Target   : None
```

```
MVPN Vrf Import   : None
MVPN Vrf Export   : None
Car. Sup C-VPN    : Disabled
Label mode        : vrf
BGP VPN Backup    : Disabled
BGP Export Inacti*: Disabled

SAP Count         : 0                   SDP Bind Count    : 1
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:321:1 S(192.0.2.2)                  TLDP         0       9190    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

The following command shows that the IPv6 interface is established and its IPv6 address is preferred (2001:DB8:1000::2/64). The IPv6 link local address has been dynamically assigned and preferred (FE80::48C6:FFFF:FE00:0/64).

```
*A:PE-3# show service id 1 interface
===============================================================================
Interface Table
===============================================================================
Interface-Name                   Adm         Opr(v4/v6)  Type    Port/SapId
   IP-Address                                            PfxState
-------------------------------------------------------------------------------
loopback                         Up          Up/Down     VPRN    loopback
   192.0.2.6/32                                          n/a
int-PE-3-PE-2                    Up          Down/Up     VPRN    spoke-321:1
   2001:db8:1000::2/64                                   PREFERRED
   fe80::48c6:ffff:fe00:0/64                             PREFERRED
-------------------------------------------------------------------------------
Interfaces : 2
===============================================================================
*A:PE-3#
```

With the following command an extensive list of parameters is displayed, including IPv6-related fields that can be checked if configured: DHCP6-relay, DHCP6-server, and so on. It is possible to use filters to reduce the output.

```
*A:PE-3# show service id 1 all
```

After verification of the services (Epipe, VPRN), the MP-eBGP peering connectivity (through IPv6 interfaces) on the VPRN between PE-3 and CE-1 can be verified as follows:

```
*A:PE-3# show router 1 bgp neighbor
===============================================================================
BGP Neighbor
===============================================================================
```

```
-------------------------------------------------------------------------------
Peer                : 2001:db8:1000::1
Description         : (Not Specified)
Group               : Spoke-CE-1-PE-3
-------------------------------------------------------------------------------
Peer AS             : 64501            Peer Port           : 179
Peer Address        : 2001:db8:1000::1
Local AS            : 64502            Local Port          : 49218
Local Address       : 2001:db8:1000::2
Peer Type           : External         Dynamic Peer        : No
State               : Established      Last State          : Active
Last Event          : recvKeepAlive
Last Error          : Unrecognized Error
Local Family        : IPv6
Remote Family       : IPv6
---snip---
IPv6 Recd. Prefixes  : 1               IPv6 Active Prefixes : 1
---snip---
Local Capability    : RtRefresh MPBGP 4byte ASN
Remote Capability   : RtRefresh MPBGP 4byte ASN
---snip---
-------------------------------------------------------------------------------
Neighbors : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
*A:PE-3#
```

Not only is the MP-eBGP session on the VPRN established, but the MP-BGP capabilities are also supported (locally and remotely). PE-3 and its BGP peer CE-1 have advertised and received an IPv6 prefix.

The same commands can be launched on PE-4. The status of the VPRN service on PE-4 and its interface to CE-5 can be verified as follows:

```
*A:PE-4# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                  Vpn Id            : 0
Service Type      : VPRN
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                  Creation Origin   : manual
Last Status Change: 10/20/2016 12:55:47
Last Mgmt Change  : 10/20/2016 12:55:47
Admin State       : Up                 Oper State        : Up

Route Dist.       : 64502:1            VPRN Type         : regular
Oper Route Dist   : 64502:1
Oper RD Type      : configured
AS Number         : 64502              Router Id         : 192.0.2.7
ECMP              : Enabled            ECMP Max Routes   : 1
Max IPv4 Routes   : No Limit

Auto Bind Tunnel
Resolution        : filter
```

```
Filter Protocol   : ldp

Max IPv6 Routes   : No Limit
Ignore NH Metric  : Disabled
Hash Label        : Disabled
Entropy Label     : Disabled
Vrf Target        : target:64502:1
Vrf Import        : None
Vrf Export        : None
MVPN Vrf Target   : None
MVPN Vrf Import   : None
MVPN Vrf Export   : None
Car. Sup C-VPN    : Disabled
Label mode        : vrf
BGP VPN Backup    : Disabled
BGP Export Inacti*: Disabled

SAP Count         : 1                   SDP Bind Count    : 0
VSD Domain        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1                               null         9212    9212    Up   Up
===============================================================================
*A:PE-4#
```

The VPRN service is up and the SAP is up.

The following output shows that the IPv6 interface is established on PE-4 and its IPv6 address is preferred.

```
*A:PE-4# show service id 1 interface
===============================================================================
Interface Table
===============================================================================
Interface-Name                  Adm         Opr(v4/v6)  Type    Port/SapId
   IP-Address                                           PfxState
-------------------------------------------------------------------------------
loopback                        Up          Up/Down     VPRN    loopback
   192.0.2.7/32                                                  n/a
int-PE-4-CE-5                   Up          Down/Up     VPRN    1/1/1
   2001:db8:2000::1/64                                           PREFERRED
   fe80::48c7:ffff:fe00:0/64                                     PREFERRED
-------------------------------------------------------------------------------
Interfaces : 2
===============================================================================
*A:PE-4#
```

MP-iBGP, providing 6VPE, has been configured and built between PE-3 and PE-4 across the MPLS network. Now, propagate the advertisements of the IPv6 prefixes learned on PE-3 from CE-1 (2001:DB8:3000::/64) and on PE-4 from CE-5 (2001:DB8:4000::/64) across the MPLS network using MP-iBGP (6VPE).

CE-1 advertises IPv6 prefix 2001:DB8:3000::/64 and CE-5 advertises IPv6 prefix2001:DB8:4000::/64.

Verify on PE-3 whether VPN-IPv6 routes were received from and advertised to its iBGP peer PE-4, as follows:

```
*A:PE-3# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.3        AS:64502       Local AS:64502
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 1
Total BGP Paths         : 13          Total Path Memory           : 2456
---snip---
Total VPN Peer Groups   : 1           Total VPN Peers             : 1
Total VPN Local Rts     : 4
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 2           Total VPN-IPv6 Rem. Act. Rts: 2
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0

Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0
---snip---
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                    AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                       PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.4
PE-4
                 64502      52    0 00h23m09s 2/2/2 (VpnIPv6)
                            54    0
-------------------------------------------------------------------------------
*A:PE-3#
```

PE-3 has received and learned a valid and best route for IPv6 prefix 2001:DB8:3000::/64 with a BGP next hop of 2001:DB8:1000:: (CE-1), as follows:

```
*A:PE-3# show router 1 bgp routes ipv6
===============================================================================
 BGP Router ID:192.0.2.6        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     Label
```

```
      As-Path
-------------------------------------------------------------------------------
u*>?  2001:db8:3000::/64                                      None        None
      2001:db8:1000::1                                        None        -
      64501
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-3#
```

The following output shows the advertised IPv6 prefix of 2001:DB8:4000::/64 advertised by PE-3 to eBGP peer CE-1.

```
*A:PE-3# show router 1 bgp neighbor 2001:db8:1000::1 advertised-routes ipv6
===============================================================================
 BGP Router ID:192.0.2.6        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                                 LocalPref   MED
      Nexthop (Router)                                        Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
?     2001:db8:4000::/64                                      n/a         None
      2001:db8:1000::2                                        None        -
      64502 64502
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-3#
```

The IPv6 route 2001:DB8:4000::/64 originates from CE-5 and was advertised from CE-5 to its eBGP peer PE-4, then from PE-4 as VPN-IPv6 route to its iBGP peer PE-3 with next-hop PE-4, as follows:

```
*A:PE-3# show router bgp routes vpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.3        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv6 Routes
===============================================================================
Flag  Network                                                 LocalPref   MED
      Nexthop (Router)                                        Path-Id     Label
      As-Path
```

```
-------------------------------------------------------------------------------
u*>i  64502:1:2001:db8:2000::/64                        100         None
      ::ffff:192.0.2.4                                  None        262140
      No As-Path
u*>?  64502:1:2001:db8:4000::/64                        100         None
      ::ffff:192.0.2.4                                  None        262140
      64501
-------------------------------------------------------------------------------
Routes : 2
```

PE-3 is advertising the VPN-IPv6 route of 2001:DB8:3000::/64 to its MP-iBGP peer PE-4, as follows. The IPv6 prefix 2001:DB8:3000::/64 was learned from CE-1 in an MP-eBGP session.

```
*A:PE-3# show router bgp neighbor 192.0.2.4 advertised-routes vpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.3        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv6 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     64502:1:2001:db8:1000::/64                        100         None
      ::ffff:192.0.2.3                                  None        262140
      No As-Path
?     64502:1:2001:db8:3000::/64                        100         None
      ::ffff:192.0.2.3                                  None        262140
      64501
-------------------------------------------------------------------------------
Routes : 2
```

The list of VPN-IPv6 routes on PE-4 includes the VPN-IPv6 route that is being learned from PE-3: 2001:DB8:3000::/64, as follows:

```
*A:PE-4# show router bgp routes vpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP VPN-IPv6 Routes
===============================================================================
Flag  Network                                           LocalPref   MED
      Nexthop (Router)                                  Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
```

```
u*>i  64502:1:2001:db8:1000::/64                          100         None
      ::ffff:192.0.2.3                                    None        262140
      No As-Path
u*>?  64502:1:2001:db8:3000::/64                          100         None
      ::ffff:192.0.2.3                                    None        262140
      64501
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-4#
```

The following output shows the advertised VPN-IPv6 route of 2001:DB8:4000::/64 from PE-4 to PE-3.

```
*A:PE-4# show router bgp neighbor 192.0.2.3 advertised-routes vpn-ipv6
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv6 Routes
===============================================================================
Flag  Network                                             LocalPref   MED
      Nexthop (Router)                                    Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     64502:1:2001:db8:2000::/64                          100         None
      ::ffff:192.0.2.4                                    None        262140
      No As-Path
?     64502:1:2001:db8:4000::/64                          100         None
      ::ffff:192.0.2.4                                    None        262140
      64501
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
*A:PE-4#
```

The following output from PE-4 shows the IPv6 prefix 2001:DB8:4000::/64 learned from CE-5.

```
*A:PE-4# show router 1 bgp routes ipv6
===============================================================================
 BGP Router ID:192.0.2.7        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                             LocalPref   MED
```

```
      Nexthop (Router)                                        Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>?  2001:db8:4000::/64                                      None        None
      2001:db8:2000::2                                        None        -
      64501
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-4#
```

The following output verifies the advertisement of IPv6 prefix 2001:DB8:3000::/64 to CE-5.

```
*A:PE-4# show router 1 bgp neighbor 2001:db8:2000::2 advertised-routes ipv6
===============================================================================
 BGP Router ID:192.0.2.7        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                                 LocalPref   MED
      Nexthop (Router)                                        Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
?     2001:db8:3000::/64                                      n/a         None
      2001:db8:2000::1                                        None        -
      64502 64502
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-4#
```

The final verification of CE-1 and CE-5 shows that IPv6 routes for AS 64501 have been received and are valid across the VPRN service, as follows:

```
*A:CE-1# show router bgp routes ipv6
===============================================================================
 BGP Router ID:192.0.2.1        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                                 LocalPref   MED
      Nexthop (Router)                                        Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
```

```
u*>?  2001:db8:4000::/64                                        None        None
      2001:db8:1000::2                                          None        -
      64502 64502
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CE-1#

*A:CE-5# show router bgp routes ipv6
===============================================================================
 BGP Router ID:192.0.2.5        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv6 Routes
===============================================================================
Flag  Network                                                   LocalPref   MED
      Nexthop (Router)                                          Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>?  2001:db8:3000::/64                                        None        None
      2001:db8:2000::1                                          None        -
      64502 64502
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:CE-5#
```

Conclusion

Spoke termination for IPv6-6VPE extends the use of spoke terminated interfaces from an Epipe VLL into a VPRN service using IPv6 interfaces on the access. Supporting the requirement of IPv6 interfaces, routing of IPv6 prefixes and the use of 6VPE for IPv6 tunneling over an IPv4 network allows SR OS to provide capabilities to support the growth of IPv6 architectures. This chapter provides examples of this feature with **show** commands for guidance.

Traffic Leaking from VPRN to GRT

This chapter provides information about Traffic Leaking from VPRN to GRT.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter is based on SR OS release 14.0 R4.

Overview

RFC 4364 *BGP/MPLS IP Virtual Private Networks (VPNs)* describes a method of distributing routing information using BGP and MPLS forwarding data to provide a Layer 3 VPN service to end users. Each virtual private routed network (VPRN) consists of a set of customer sites connected to one or more PE routers. Each associated PE router maintains a separate IP forwarding table for each VPRN. Additionally, the PE routers exchange the routing information configured or learned from all customer sites via multi-protocol border gateway protocol (MP-BGP) peering. Each route exchanged via the MP-BGP protocol includes a route distinguisher (RD), which identifies the VPRN association and resolves any IP address overlap.

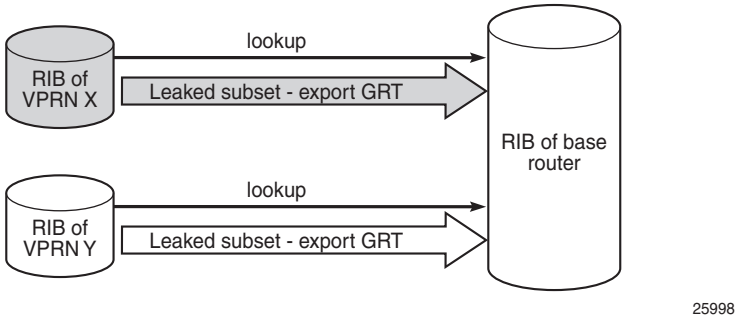
It has always been possible to route between one VPRN and another, using scenarios such as "extranet", "hub and spoke" and so on, using the vrf-import and vrf-export policies for BGP VPN-IPv4 route distribution.

Traffic leaking to the global route table (GRT) allows service providers to offer VPRN and Internet services over a single virtual routing and forwarding VRF interface. Packets entering a VRF interface can have route processing results derived from the VRF or the GRT. The leaking and preferred lookup settings are configured on a per-VPRN basis.

To allow data flowing from a VPRN to the base router, routing information from the base router must be made available for lookup by the VPRN. The GRT lookup can be general (for example, any lookup miss in the virtual routing and forwarding (VRF) table can be resolved in the GRT), or specific (for example, specific routes should only be looked up in the GRT and ignored by the VPRN).

To enable the GRT lookup from the VPRN, the command enable-grt is used. This only provides part of the solution, because packets can now be forwarded from the VPRN to the GRT, but not in the opposite direction. The GRT needs to learn specific destination prefixes from the VPRN and this is achieved by route leaking from the VPRN to the GRT, using policies (command export-grt). By default, the maximum number of leaked routes from a VPRN to the GRT is five, but this limit can be modified or even removed. Prefixes should be globally unique within the service provider network and if these are propagated outside the provider's network, they must be from the public IP space and globally unique.

Figure 365 **VPRN to GRT leak process**

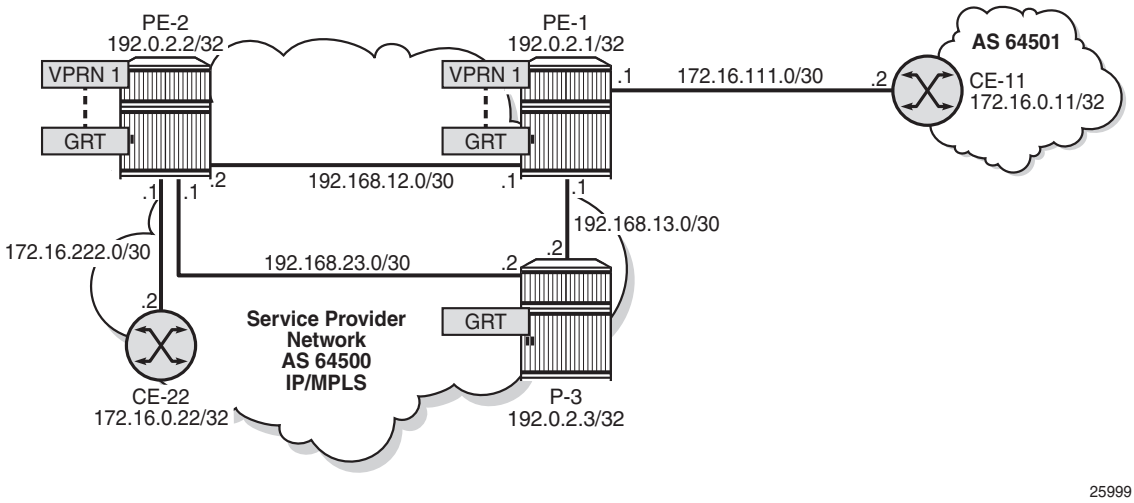


The method described in this chapter allows the network administrator to leak certain routes or all the routes that are inside a VPRN to the GRT. Route leaking from VPRN to GRT is protocol-independent and can be applied for BGP, OSPF(v3), IS-IS, static, local routes, and so on. For BGP routes, there is an improved route leaking mechanism that allows leaking routes including all their attributes; see chapter BGP Route Leaking.

Configuration

Figure 366 shows the example topology used in this chapter, including the IPv4 addresses. The interfaces also have IPv6 addresses, which will be shown in Figure 368.

Figure 366 **Example Topology with IPv4 Addresses**



Initial Configuration

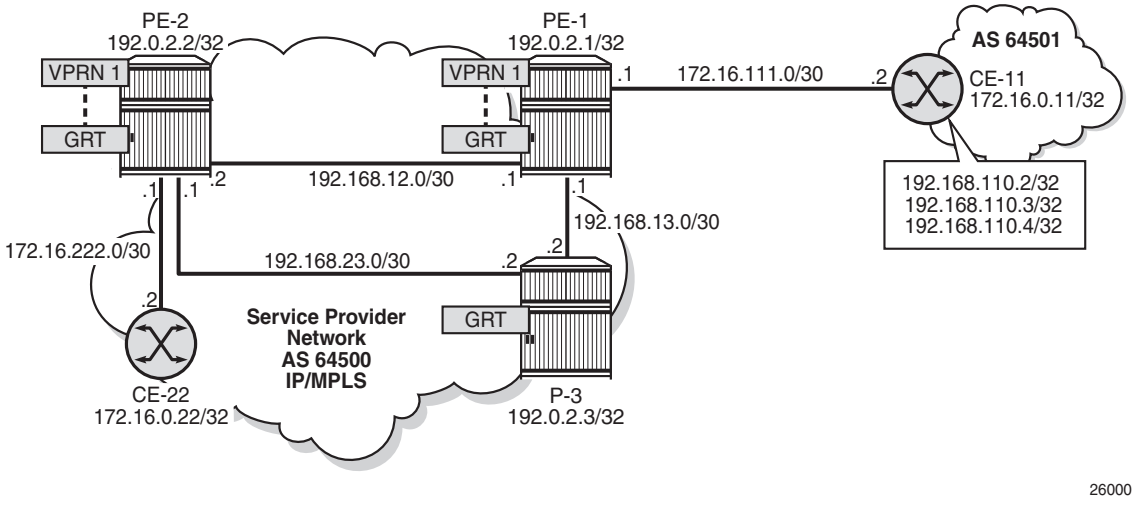
The nodes in the example topology have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces
- IGP (IS-IS or OSPF) between the PEs
- LDP between the PEs
- VPRN 1 on PE-1 and PE-2
- BGP (IBGP between the PEs and between PE-2 and CE-22; EBGP between PE-1 and CE-11)
  - On the PEs, BGP is configured in the base router and in the VPRNs.
- Loopback addresses on CE-11, such as 192.168.110.2/32.
- Export policies on CE-11 to export routes from direct with certain prefixes.

Protocol-Independent IPv4 Route Leaking from VPRN to GRT

Figure 367 shows the topology with the IP addresses for this example. Route leaking from VPRN to GRT is protocol independent and in this example, VPRN 1 on PE-1 will leak local routes, static routes, and imported BGP routes to the GRT. VPRN 1 on PE-2 will leak local routes and IS-IS routes. OSPF routes can also be leaked, but that is not shown here.

Figure 367 **IPv4 VPRN to GRT route leaking for ISIS**



GRT-leak is by default disabled. The routing table for VPRN 1 on PE-1 contains local routes, static routes, and BGP routes that are learned from CE-11, as follows:

```
*A:PE-1# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.1/32                                   Local   Local     00h06m26s   0
       system                                                         0
172.16.111.0/30                                 Local   Local     00h06m26s   0
       int-PE-1-CE-11                                                 0
192.168.110.2/32                                Remote  BGP       00h05m19s   170
       172.16.111.2                                                   0
192.168.110.3/32                                Remote  BGP       00h05m19s   170
       172.16.111.2                                                   0
192.168.110.4/32                                Remote  BGP       00h05m19s   170
       172.16.111.2                                                   0
192.168.120.0/24                                Remote  Static    00h00m12s   5
       172.16.111.2                                                   1
```

```
-------------------------------------------------------------------------------
No. of Routes: 6
```

By default, the GRT is not learning the VPRN routes, as follows:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     00h06m27s  0
       system                                                          0
192.0.2.2/32                                    Remote  OSPF      00h06m02s  10
       192.168.12.2                                                    10
192.0.2.3/32                                    Remote  OSPF      00h05m53s  10
       192.168.13.2                                                    10
192.168.12.0/30                                 Local   Local     00h06m28s  0
       int-PE-1-PE-2                                                   0
192.168.13.0/30                                 Local   Local     00h06m28s  0
       int-PE-1-P-3                                                    0
192.168.23.0/30                                 Remote  OSPF      00h06m02s  10
       192.168.12.2                                                    20
-------------------------------------------------------------------------------
No. of Routes: 6
```

On PE-2, the routing table for VPRN 1 contains local routes and IS-IS routes, as follows:

```
 *A:PE-2# show router 1 route-table
===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.2/32                                   Local   Local     00h06m52s  0
       system                                                          0
172.16.2.22/32                                  Remote  ISIS      00h06m22s  15
       172.16.222.2                                                    10
172.16.222.0/30                                 Local   Local     00h06m52s  0
       int-PE-2-CE-22                                                  0
-------------------------------------------------------------------------------
No. of Routes: 3
```

By default, none of these VPRN routes is leaked to the GRT, as follows:

```
*A:PE-2# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h06m49s  15
```

```
       192.168.12.1                                                    10
192.0.2.2/32                                  Local   Local     00h06m55s  0
       system                                                          0
192.0.2.3/32                                  Remote  ISIS      00h06m43s  15
       192.168.23.2                                                    10
192.168.12.0/30                               Local   Local     00h06m55s  0
       int-PE-2-PE-1                                                   0
192.168.13.0/30                               Remote  ISIS      00h06m49s  15
       192.168.12.1                                                    20
192.168.23.0/30                               Local   Local     00h06m55s  0
       int-PE-2-P-3                                                    0
-------------------------------------------------------------------------------
No. of Routes: 6
```

To enable VPRN to GRT leaking, the following route policy is configured in VPRN 1, on PE-1:

```
configure
    router
        policy-options
            begin
            policy-statement "LeakVPRNtoGRT_pref8"
                entry 10
                    action accept
                        preference 8
                    exit
                exit
            exit
            commit
```

This policy allows leaking all routes from a VPRN to the base router, without any match criteria. However, when routes are leaked from VPRNs to the GRT, they need to be unique and only routes that need to be known in the GRT should be leaked. By default, the preference for a leaked route is 180. The preference can be manually configured to a lower value, such as 8, to avoid network inconsistencies between the IGP and the RT on the router where the routes are leaked.

On PE-2, a similar policy is applied, but without the preference configuration, as follows. This implies that the routes leaked on PE-2 will have default preference 180.

```
configure
    router
        policy-options
            begin
            policy-statement "LeakVPRNtoGRT"
                entry 10
                    action accept
                    exit
                exit
            exit
            commit
```

The policy must be applied to VPRN 1, as follows:

```
configure
    service
        vprn 1
            grt-lookup
                enable-grt
                exit
                export-grt "Leak-VPRN-toGRT"
            exit
        exit
```

When **enable-grt** is configured, any lookup miss in the VRF table will be resolved in the GRT, if available. This only works from VPRN to GRT and does not require route leaking. However, the GRT needs to be able to route packets back to the VPRN and it cannot perform a lookup in the routing table of the VPRN. Therefore, route leaking from VPRN to GRT is required, and **export-grt** is configured. Prefixes in the VPRN must be leaked to the GRT through a policy. Prefixes leaked from any VPRN should never conflict with prefixes leaked from any other VPRN or existing prefixes in the GRT.

This configuration is protocol-independent. Route leaking from VPRN to GRT is applicable for all kinds of learned routes, such as static routes, local routes, IS-IS, OSPF, BGP, and so on.

After routes are leaked from the VPRN to the GRT, the routing table of the base router includes the leaked routes, with protocol "VPN Leak". For PE-1, the routing table contains the following routes:

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
172.16.1.1/32                                   Remote  VPN Leak  00h00m11s  8
       system                                                          0
172.16.111.0/30                                 Remote  VPN Leak  00h00m11s  8
       int-PE-1-CE-11                                                  0
192.0.2.1/32                                    Local   Local     00h11m58s  0
       system                                                          0
192.0.2.2/32                                    Remote  ISIS      00h11m35s  15
       192.168.12.2                                                    10
192.0.2.3/32                                    Remote  ISIS      00h11m32s  15
       192.168.13.2                                                    10
192.168.12.0/30                                 Local   Local     00h11m58s  0
       int-PE-1-PE-2                                                   0
192.168.13.0/30                                 Local   Local     00h11m58s  0
       int-PE-1-P-3                                                    0
192.168.23.0/30                                 Remote  ISIS      00h11m35s  15
       192.168.12.2                                                    20
192.168.110.3/32                                Remote  VPN Leak  00h00m11s  8
       172.16.111.2                                                    0
192.168.110.4/32                                Remote  VPN Leak  00h00m11s  8
       172.16.111.2                                                    0
```

```
192.168.120.0/24                                Remote  VPN Leak  00h00m11s  8
       172.16.111.2                                                 0
-------------------------------------------------------------------------------
No. of Routes: 11
```

Regardless of the preference of the original routes in VPRN 1, all the leaked routes in the GRT have preference 8, as configured. By default, only five routes are leaked. This export limit can be overruled, as follows:

```
*A:PE-1# configure service vprn 1 grt-lookup export-limit 10
```

The following command shows only the routes leaked from any VPRN to GRT on PE-1:

```
*A:PE-1# show router route-table protocol vpn-leak all
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                           Active     Metric
-------------------------------------------------------------------------------
172.16.1.1/32                                   Remote  VPN Leak  00h02m56s  8
       system                                              Y            0
172.16.111.0/30                                 Remote  VPN Leak  00h02m56s  8
       int-PE-1-CE-11                                      Y            0
192.168.110.2/32                                Remote  VPN Leak  00h02m56s  8
       172.16.111.2                                        Y            0
192.168.110.3/32                                Remote  VPN Leak  00h02m56s  8
       172.16.111.2                                        Y            0
192.168.110.4/32                                Remote  VPN Leak  00h02m56s  8
       172.16.111.2                                        Y            0
192.168.120.0/24                                Remote  VPN Leak  00h02m56s  8
       172.16.111.2                                        Y            0
-------------------------------------------------------------------------------
No. of Routes: 6
```

On PE-2, the following routes are leaked from VPRN to GRT:

```
*A:PE-2# show router route-table protocol vpn-leak all
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                           Active     Metric
-------------------------------------------------------------------------------
172.16.1.2/32                                   Remote  VPN Leak  00h00m05s  180
       system                                              Y            0
172.16.2.22/32                                  Remote  VPN Leak  00h00m05s  180
       172.16.222.2                                        Y            0
172.16.222.0/30                                 Remote  VPN Leak  00h00m05s  180
       int-PE-2-CE-22                                      Y            0
-------------------------------------------------------------------------------
No. of Routes: 3
```

Different types of routes are leaked to the GRT with protocol type "VPN Leak" and all of them get the same preference, configured or default. The detailed output for any leaked route in the preceding list for PE-1 shows protocol VPN_LEAK and preference 8, as follows:

```
*A:PE-1# show router route-table protocol vpn-leak 192.168.110.2/32 extensive
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 192.168.110.2/32
  Protocol              : VPN_LEAK
  Age                   : 00h07m57s
  Preference            : 8
  Next-Hop              : 172.16.111.2
    Interface           : int-PE-1-CE-11 (VPRN 1)
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    Metric              : 0
    ECMP-Weight         : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-1#
```

Export IPv4 VPN-Leak Routes to Routing Protocols

Until now, the VPN-leak routes are leaked locally to the GRT, but they are not advertised in IS-IS, OSPF, or BGP. Router P-3 has not learned any of the leaked routes, as follows:

```
*A:P-3# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      00h50m04s  15
       192.168.13.1                                                    10
192.0.2.2/32                                    Remote  ISIS      00h50m04s  15
       192.168.23.1                                                    10
192.0.2.3/32                                    Local   Local     00h50m13s  0
       system                                                          0
192.168.12.0/30                                 Remote  ISIS      00h50m04s  15
       192.168.13.1                                                    20
192.168.13.0/30                                 Local   Local     00h50m13s  0
       int-P-3-PE-1                                                    0
192.168.23.0/30                                 Local   Local     00h50m13s  0
       int-P-3-PE-2                                                    0
-------------------------------------------------------------------------------
No. of Routes: 6
```

To reduce the number of routes to be exported, the export will only be configured on PE-1 and a match criterion will be added for the routes to be leaked, as follows:

```
configure
    router
        policy-options
            begin
            prefix-list "192.168.110.0"
                prefix 192.168.110.0/24 longer
            exit
            policy-statement "LeakVPRNtoGRT_pref8_110"
                entry 10
                    from
                        prefix-list "192.168.110.0"
                    exit
                    action accept
                        preference 8
                    exit
                exit
            exit
            commit

configure service vprn 1 grt-lookup export-grt "LeakVPRNtoGRT_pref8_110"
```

VPN-leak routes can be exported to any routing protocol. Prefix lists can be used to filter routes, but that is not configured in this example. The following export policy is configured on PE-1 to export the VPN-leak routes:

```
configure
    router
        policy-options
            begin
            policy-statement "export-vpn-leak"
                entry 10
                    from
                        protocol vpn-leak
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

The same export policy will be used for export to IS-IS, OSPF, and BGP.

Export IPv4 VPN-Leak Routes to IS-IS

The export policy is applied in the IS-IS context on PE-1, as follows:

```
configure
    router
        isis
```

```
            export "export-vpn-leak"
        exit
    exit
exit
```

The leaked routes are now advertised via IS-IS and appear as IS-IS routes with default preference for IS-IS routes on PE-2 and P-3. The route table on P-3 looks as follows:

```
*A:P-3# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                     Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  ISIS      01h32m01s  15
       192.168.13.1                                                10
192.0.2.2/32                                    Remote  ISIS      01h32m01s  15
       192.168.23.1                                                10
192.0.2.3/32                                    Local   Local     01h32m10s  0
       system                                                       0
192.168.12.0/30                                 Remote  ISIS      01h32m01s  15
       192.168.13.1                                                20
192.168.13.0/30                                 Local   Local     01h32m10s  0
       int-P-3-PE-1                                                 0
192.168.23.0/30                                 Local   Local     01h32m10s  0
       int-P-3-PE-2                                                 0
192.168.110.2/32                                Remote  ISIS      00h00m06s  15
       192.168.13.1                                                10
192.168.110.3/32                                Remote  ISIS      00h00m06s  15
       192.168.13.1                                                10
192.168.110.4/32                                Remote  ISIS      00h00m06s  15
       192.168.13.1                                                10
-------------------------------------------------------------------------------
No. of Routes: 9
```

The export policy is removed from the IS-IS context on PE-1, as follows:

```
*A:PE-1# configure router isis no export
```

Export IPv4 VPN-Leak Routes to OSPF

When OSPF is used instead of IS-IS, the behavior is similar. The export policy is applied in the OSPF context on PE-1, as follows:

```
configure
    router
        ospf
            export "export-vpn-leak"
        exit
    exit
exit
```

To export routes into OSPF using a policy, the router must be configured as ASBR, as follows:

```
*A:PE-1# configure router ospf asbr
```

The routes with protocol VPN-leak on PE-1 are now exported in OSPF to PE-2 and P-3. The default preference for external OSPF routes is 150. On PE-3, the routing table contains the following OSPF routes:

```
*A:P-3# show router route-table protocol ospf
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                        Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                      Remote  OSPF      00h10m37s  10
       192.168.13.1                                                     10
192.0.2.2/32                                      Remote  OSPF      02h13m06s  10
       192.168.23.1                                                     10
192.168.12.0/30                                   Remote  OSPF      00h10m37s  10
       192.168.13.1                                                     20
192.168.110.2/32                                  Remote  OSPF      00h03m50s  150
       192.168.13.1                                                     1
192.168.110.3/32                                  Remote  OSPF      00h03m50s  150
       192.168.13.1                                                     1
192.168.110.4/32                                  Remote  OSPF      00h03m50s  150
       192.168.13.1                                                     1
-------------------------------------------------------------------------------
No. of Routes: 6
```

The export policy is removed from the OSPF context on PE-1 as follows:

```
*A:PE-1# configure router ospf no export
```

Export IPv4 VPN-Leak Routes to BGP

The export policy is applied in the general BGP context of PE-1, as follows:

```
configure
    router
        bgp
            export "export-vpn-leak"
        exit
    exit
exit
```

The VPN-leak routes from PE-1 will be advertised as BGP routes to BGP neighbors PE-2 and PE-3, and the routing tables will contain BGP routes with preference 170. P-3 has the following BGP routes:

```
*A:P-3# show router route-table protocol bgp
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.168.110.2/32                                Remote  BGP       00h04m07s  170
       192.168.13.1                                                    0
192.168.110.3/32                                Remote  BGP       00h04m07s  170
       192.168.13.1                                                    0
192.168.110.4/32                                Remote  BGP       00h04m07s  170
       192.168.13.1                                                    0
-------------------------------------------------------------------------------
No. of Routes: 3
```

If it is required to preserve the BGP path attributes in the leaking process, you must use the BGP Route Leaking process described in chapter BGP Route Leaking. However, with this protocol-independent route leaking mechanism, it is possible to leak non-BGP routes to the GRT that will be advertised as BGP routes.

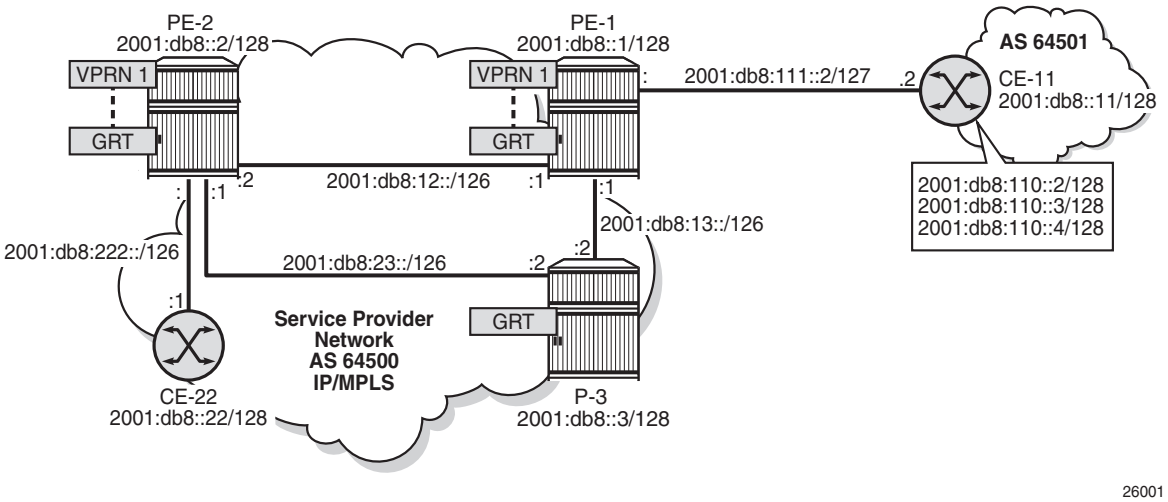
The export policy is removed from the BGP context, as follows:

```
*A:PE-1# configure router bgp no export
```

Protocol-Independent IPv6 Route Leaking from VPRN to GRT

Figure 368 shows the topology and the IP addresses used for IPv6. CE-11 exports routes such as 2001:db8:110::2/128 to VPRN 1on PE-1. On PE-1, local routes, static routes, and BGP routes will be leaked to the GRT. On PE-2, IS-IS routes and local routes are leaked to the GRT.

Figure 368 **Example Topology with IPv6 Addresses**



The IPv6 routing table for VPRN 1 on PE-1 includes local addresses, a static route, and three BGP routes exported by CE-11, as follows:

```
*A:PE-1# show router 1 route-table ipv6
===============================================================================
IPv6 Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8::1:1/128                               Local   Local     05h34m15s   0
       system                                                         0
2001:db8:110::2/128                             Remote  BGP       00h32m43s   170
       2001:db8:111::1                                                0
2001:db8:110::3/128                             Remote  BGP       00h32m43s   170
       2001:db8:111::1                                                0
2001:db8:110::4/128                             Remote  BGP       00h32m43s   170
       2001:db8:111::1                                                0
2001:db8:111::/127                              Local   Local     05h34m15s   0
       int-PE-1-CE-11                                                 0
2001:db8:120::/120                              Remote  Static    04h52m38s   5
       2001:db8:111::1                                                1
-------------------------------------------------------------------------------
No. of Routes: 6
```

By default, route leaking is disabled and the IPv6 GRT on PE-1 does not contain any of the IPv6 routes in VPRN 1, as follows:

```
*A:PE-1# show router route-table ipv6
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age         Pref
      Next Hop[Interface Name]                                      Metric
```

```
-------------------------------------------------------------------------------
2001:db8::1/128                                 Local   Local     05h34m17s  0
       system                                                         0
2001:db8::2/128                                 Remote  OSPF3     00h00m28s  10
       fe80::48c5:1ff:fe01:1-"int-PE-1-PE-2"                        10
2001:db8::3/128                                 Remote  OSPF3     00h00m20s  10
       fe80::48c6:1ff:fe01:2-"int-PE-1-P-3"                         10
2001:db8:12::/126                               Local   Local     05h34m17s  0
       int-PE-1-PE-2                                                  0
2001:db8:13::/126                               Local   Local     05h34m16s  0
       int-PE-1-P-3                                                   0
2001:db8:23::/126                               Remote  OSPF3     00h00m20s  10
       fe80::48c6:1ff:fe01:2-"int-PE-1-P-3"                         20
-------------------------------------------------------------------------------
No. of Routes: 6
```

The VPN-leak route policy is the same as for IPv4 routes, and is applied in the VPRN context in the same way as for IPv4 routes, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "LeakVPRNtoGRT_pref8"
                entry 10
                    action accept
                        preference 8
                    exit
                exit
            exit
            commit


configure
    service
        vprn 1
            grt-lookup
                enable-grt
                exit
                export-grt "LeakVPRNtoGRT_pref8"
            exit
        exit
    exit
exit
```

On PE-1, the VPN leak policy sets the preference to 8, and on PE-2, the default preference is used, as in the preceding example for IPv4.

On PE-2, the IPv6 routing table for VPRN1 contains two local routes and one IS-IS route, as follows:

```
*A:PE-2# show router 1 route-table ipv6
===============================================================================
IPv6 Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
```

```
-------------------------------------------------------------------------------
2001:db8:1::2/128                               Local   Local     05h54m22s  0
       system                                                         0
2001:db8:2::22/128                              Remote  ISIS      05h53m52s  15
       fe80::48c8:1ff:fe01:1-"int-PE-2-CE-22"                         10
2001:db8:222::/127                              Local   Local     05h54m22s  0
       int-PE-2-CE-22                                                 0
-------------------------------------------------------------------------------
No. of Routes: 3
```

All of these routes are leaked to the GRT on PE-2 with default preference 180, as follows:

```
*A:PE-2# show router route-table ipv6 protocol vpn-leak
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8:1::2/128                               Remote  VPN Leak  00h09m37s  180
       system                                                         0
2001:db8:2::22/128                              Remote  VPN Leak  00h09m37s  180
       fe80::48c8:1ff:fe01:1-"int-PE-2-CE-22"                         0
2001:db8:222::/127                              Remote  VPN Leak  00h09m37s  180
       int-PE-2-CE-22                                                 0
-------------------------------------------------------------------------------
No. of Routes: 3
```

On PE-1, the IPv6 routing table for VPRN 1 contains six routes, but by default, a maximum of five routes are leaked, as follows:

```
*A:PE-1# show router route-table ipv6 protocol vpn-leak
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8::1:1/128                               Remote  VPN Leak  00h18m41s  8
       system                                                         0
2001:db8:110::2/128                             Remote  VPN Leak  00h18m41s  8
       2001:db8:111::1                                                0
2001:db8:110::3/128                             Remote  VPN Leak  00h18m41s  8
       2001:db8:111::1                                                0
2001:db8:111::/127                              Remote  VPN Leak  00h18m41s  8
       int-PE-1-CE-11                                                 0
2001:db8:120::/120                              Remote  VPN Leak  00h18m41s  8
       2001:db8:111::1                                                0
-------------------------------------------------------------------------------
No. of Routes: 5
```

The export limit for IPv6 routes is removed, as follows:

```
*A:PE-1# configure service vprn 1 grt-lookup export-v6-limit 0
```

As a result, there is no limit to the number of leaked IPv6 routes, and all six IPv6 routes are leaked from VPRN 1 to the GRT with the configured preference 8, as follows:

```
*A:PE-1# show router route-table ipv6 protocol vpn-leak
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8::1:1/128                               Remote  VPN Leak  00h00m11s  8
       system                                                       0
2001:db8:110::2/128                             Remote  VPN Leak  00h00m11s  8
       2001:db8:111::1                                              0
2001:db8:110::3/128                             Remote  VPN Leak  00h00m11s  8
       2001:db8:111::1                                              0
2001:db8:110::4/128                             Remote  VPN Leak  00h00m11s  8
       2001:db8:111::1                                              0
2001:db8:111::/127                              Remote  VPN Leak  00h00m11s  8
       int-PE-1-CE-11                                               0
2001:db8:120::/120                              Remote  VPN Leak  00h00m11s  8
       2001:db8:111::1                                              0
-------------------------------------------------------------------------------
No. of Routes: 6
```

The details for any of the routes shows that the protocol is VPN-leak and the preference is 8, as follows:

```
*A:PE-1# show router route-table protocol vpn-leak 2001:db8:110::2/128 extensive
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8:110::2/128
  Protocol              : VPN_LEAK
  Age                   : 00h08m25s
  Preference            : 8
  Next-Hop              : 2001:db8:111::1
    Interface           : int-PE-1-CE-11 (VPRN 1)
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    Metric              : 0
    ECMP-Weight         : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
*A:PE-1#
```

Export IPv6 VPN-Leak Routes to Routing Protocols

Until now, the IPv6 VPN-leak routes are leaked locally to the GRT, but they are not advertised in IS-IS, OSPFv3, or BGP. Router P-3 has not learned any of the leaked IPv6 routes, as follows:

```
*A:P-3# show router route-table ipv6
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                        Metric
-------------------------------------------------------------------------------
2001:db8::1/128                                   Remote  OSPF3     00h41m47s  10
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                           10
2001:db8::2/128                                   Remote  OSPF3     00h41m47s  10
       fe80::48c5:1ff:fe01:2-"int-P-3-PE-2"                           10
2001:db8::3/128                                   Local   Local     06h15m21s  0
       system                                                         0
2001:db8:12::/126                                 Remote  OSPF3     00h41m47s  10
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                           20
2001:db8:13::/126                                 Local   Local     06h15m20s  0
       int-P-3-PE-1                                                   0
2001:db8:23::/126                                 Local   Local     06h15m19s  0
       int-P-3-PE-2                                                   0
-------------------------------------------------------------------------------
No. of Routes: 6
```

To reduce the number of VPN-leak routes, a match criterion is added to the route policy on PE-1, as follows:

```
configure
    router
        policy-options
            begin
            prefix-list "2001:db8:110::"
                prefix 2001:db8:110::/125 longer
            exit
            policy-statement "LeakVPRNtoGRT_pref8_110"
                entry 20
                    from
                        prefix-list "2001:db8:110::"
                    exit
                    action accept
                        preference 8
                    exit
                exit
            exit
            commit


configure
    service
        vprn 1
            grt-lookup
                enable-grt
```

```
                exit
                export-grt "LeakVPRNtoGRT_pref8_110"
            exit
        exit
    exit
exit
```

The following IPv6 routes are leaked from VPRN 1 to GRT on PE-1:

```
*A:PE-1# show router route-table ipv6 protocol vpn-leak
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8:110::2/128                             Remote  VPN Leak  00h03m20s  8
       2001:db8:111::1                                                0
2001:db8:110::3/128                             Remote  VPN Leak  00h03m20s  8
       2001:db8:111::1                                                0
2001:db8:110::4/128                             Remote  VPN Leak  00h03m20s  8
       2001:db8:111::1                                                0
-------------------------------------------------------------------------------
No. of Routes: 3
```

IPv6 VPN-leak routes can be exported to routing protocols IS-IS, OSPFv3, and BGP.

The export policy on PE-1 is the same as in all the preceding examples for IPv4, as follows:

```
configure
    router
        policy-options
            begin
            policy-statement "export-vpn-leak"
                entry 10
                    from
                        protocol vpn-leak
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

Export IPv6 VPN-Leak Routes to IS-IS

The export policy for IPv6 routes of protocol VPN-leak is applied for IS-IS, as follows:

```
configure
    router
        isis
            export "export-vpn-leak"
```

```
        exit
    exit
exit
```

The three IPv6 VPN-leak routes from PE-1 are now advertised by IS-IS to PE-2 and P-3. The routing table on P-3 contains the following IPv6 IS-IS routes:

```
*A:P-3# show router route-table ipv6 protocol isis
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8:110::2/128                             Remote  ISIS      00h00m52s  15
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         10
2001:db8:110::3/128                             Remote  ISIS      00h00m52s  15
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         10
2001:db8:110::4/128                             Remote  ISIS      00h00m52s  15
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         10
-------------------------------------------------------------------------------
No. of Routes: 3
```

The export policy is removed for IS-IS, as follows:

```
*A:PE-1# configure router isis no export
```

Export IPv6 VPN-Leak Routes to OSPFv3

The export policy for IPv6 routes of protocol VPN-leak is applied for OSPFv3, as follows:

```
configure
    router
        ospf3
            export "export-vpn-leak"
        exit
    exit
exit
```

Routes can only be exported to OSPFv3 if the router is configured as ASBR, as follows:

```
*A:PE-1# configure router ospf3 asbr
```

The IPv6 VPN-leak routes from PE-1 are now advertised by OSPFv3 to PE-2 and P-3. The routing table on P-3 contains the following IPv6 OSPFv3 routes:

```
*A:P-3# show router route-table ipv6 protocol ospf3
===============================================================================
```

```
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8::1/128                                 Remote  OSPF3     01h10m06s  10
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         10
2001:db8::2/128                                 Remote  OSPF3     01h10m06s  10
       fe80::48c5:1ff:fe01:2-"int-P-3-PE-2"                         10
2001:db8:12::/126                               Remote  OSPF3     01h10m06s  10
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         20
2001:db8:110::2/128                             Remote  OSPF3     00h01m57s  150
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         1
2001:db8:110::3/128                             Remote  OSPF3     00h01m57s  150
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         1
2001:db8:110::4/128                             Remote  OSPF3     00h01m57s  150
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         1
-------------------------------------------------------------------------------
No. of Routes: 6
```

The preference for remote OSPFv3 routes is by default 150. The export policy is removed for OSPFv3, as follows:

```
*A:PE-1# configure router ospf3 no export
```

Export IPv6 VPN-Leak Routes to BGP

The export policy for IPv6 routes of protocol VPN-leak is applied for BGP, as follows:

```
configure
    router
        bgp
            export "export-vpn-leak"
        exit
    exit
exit
```

The three IPv6 VPN-leak routes from PE-1 are now advertised by BGP to PE-2 and P-3. The routing table on P-3 contains the following IPv6 BGP routes:

```
*A:P-3# show router route-table ipv6 protocol bgp
===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
2001:db8:110::2/128                             Remote  BGP       00h03m44s  170
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         0
2001:db8:110::3/128                             Remote  BGP       00h03m44s  170
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         0
2001:db8:110::4/128                             Remote  BGP       00h03m44s  170
       fe80::48c4:1ff:fe01:1-"int-P-3-PE-1"                         0
```

```
-------------------------------------------------------------------------------
No. of Routes: 3
```

The export policy is removed for BGP, as follows:

```
*A:PE-1# configure router bgp no export
```

In this example, BGP leaked IPv6 routes are advertised by BGP. For scenarios with only BGP routes, a dedicated BGP route leaking mechanism that preserves all attributes is preferred, as described in chapter BGP Route Leaking. However, with the same configuration as in this chapter, it is possible to leak non-BGP routes and advertise them using BGP.

Conclusion

Routes learned in a VPRN can be leaked to the base router and advertised using routing protocols. The mechanism described in this chapter is protocol-independent: all kinds of routes can be leaked from a VRF to the GRT: local, static, IS-IS, OSPF, BGP routes, and so on. In some cases, it might be useful to leak the routes from a VPRN to the entire network using the routing protocol, in order to access the resources defined inside the VRF. Routes that are leaked from VPRNs to the GRT must be unique in the network where they will be advertised. For BGP routes, the protocol-independent route leaking mechanism described here does not preserve the attributes, unlike the dedicated BGP route leaking feature.

VPRN Inter-AS VPN Model C

This chapter provides information about virtual private routed network (VPRN) inter-autonomous system (AS) virtual private network (VPN) model C.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for release 7.0. The CLI in the current edition corresponds to 14.0.R5. The address family label-ipv4 is supported in SR OS release 14.0.R4, and later, see chapter Separate BGP RIBs for Labeled Routes.There are no prerequisites for this configuration.

Overview

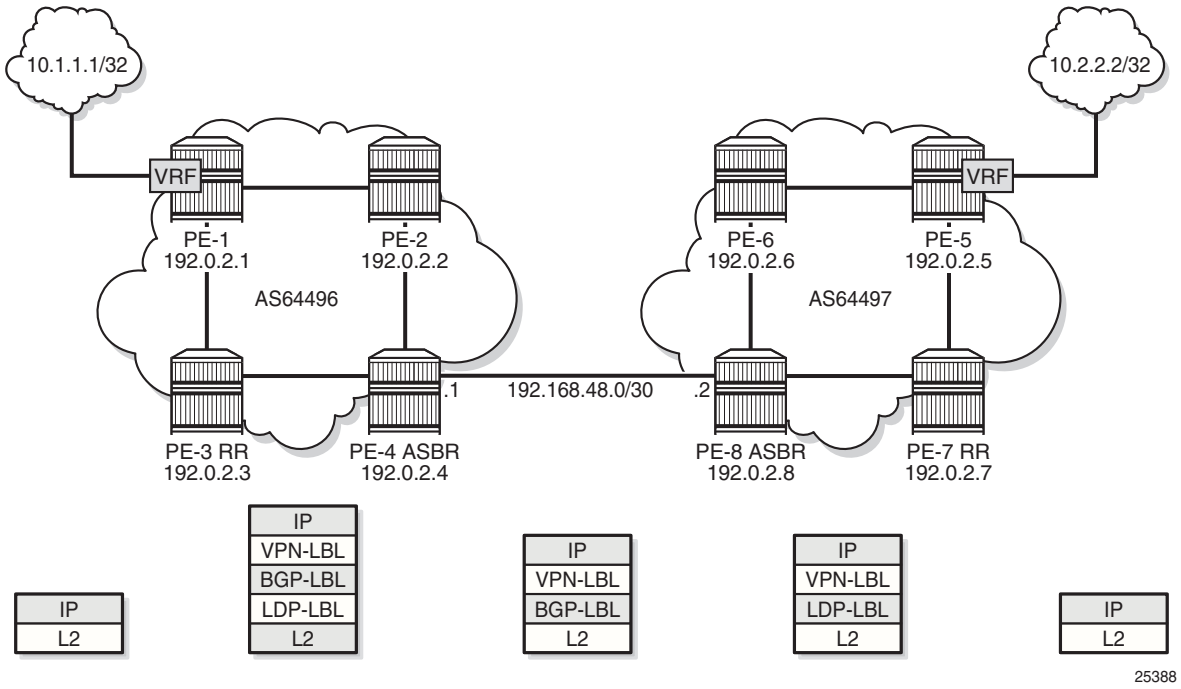
Introduction

Section 10 of RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, describes three potential methods for service providers to interconnect their IP-VPN (Internet Protocol — Virtual Private Network) backbones in order to provide an end-to-end MPLS-VPN where one or more sites of the VPN are connected to different service provider autonomous systems. The purpose of this chapter is to describe the configuration and troubleshooting for inter-AS VPN model C.

In this architecture, VPN prefixes are neither held, nor re-advertised by the Autonomous System Border Router — Provider Edge (ASBR-PE) routers. However, the ASBR-PE maintains labeled IPv4 /32 routes to other PE routers within its own AS. The ASBR then redistributes these /32 IPv4 prefixes in external Border Gateway Protocol (eBGP) to the ASBR-PE in other service providers ASs. Using this methodology, it is possible for PE routers in different ASs to establish multi-hop Multi Protocol — external Border Gateway Protocol (MP-eBGP) sessions to each other in order to exchange customer VPN prefixes over those connections.

The /32 IPv4 routes for the PE routers in the other service providers AS will need to be redistributed into the interior gateway protocol (IGP) in the local AS together with an assigned label. Because most service providers do not like redistribution of loopback addresses from another service provider into the local IGP, a potential solution can be found by imposing a three-level label stack on the ingress PE. The bottom-level label would be assigned by the egress PE (advertised in multi-hop MP-eBGP without next-hop override) and is commonly referred to as the VPN-label. The middle label would be assigned by the local ASBR-PE and would correspond to the /32 route of the egress PE (in a different AS) using BGP-LBL (RFC 3107, *Carrying Label Information in BGP-4*). The top level label would then be the label assigned by the local ASBR-PE(s) /32 loopback address, which would be assigned by the IGP next-hop of the ingress PE. This label is referred to as the LDP-LBL. Figure 369 reflects this mechanism. The VPN-LBL is assigned by PE-5, the BGP-LBL is assigned by PE-4 and the LDP-LBL is also assigned by PE-4.

Figure 369 **Inter-AS VPN Model C**



The VPN connectivity is established using Labeled VPN route exchange using MP-eBGP without next-hop override. The PE connectivity will be established as follows.

EBGP PE /32 loopback leaking routing exchange using eBGP LBL (RFC 3107) at the ASBR-PE. The /32 PE routes learned from the other AS through the ASBR-PE are further distributed into the local AS using iBGP and optionally Route Reflectors (RRs). This model uses a three label stack and is referred to as Model C. Resilience for ASBR-PE failures is dependent on BGP.

Figure 370 **Protocol Overview**

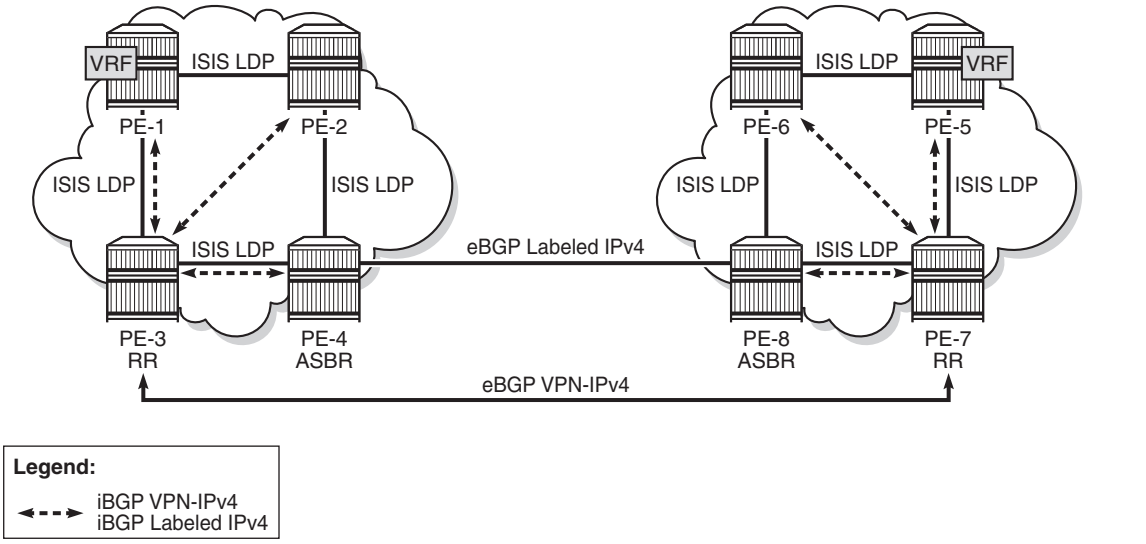


Figure 370 gives an overview of all protocols used when implementing Inter-AS Model C. Inside each AS, there is an IS-IS adjacency and a link LDP session between each pair of adjacent nodes. As an alternative, OSPF can be used as IGP. There is also an iBGP session between each PE and the RR. The address family is both VPN-IPv4 for the exchange of customer VPN prefixes and Labeled IPv4 for the exchange of labeled IPv4 prefixes. When no RR is used, a full mesh of iBGP sessions can be established in each AS.

Between the ASBRs, there is an eBGP session for the exchange of labeled IPv4 prefixes. The ASBRs will override the next-hop for those prefixes. Between the RRs in the different ASs, there is an eBGP session for the exchange of VPN customer prefixes. The RRs will not override the next-hop for those prefixes.

The main advantage of this model is that no VPN routes need to be held on the ASBR-PEs and therefore, it scales the best among all the three Inter-AS IP-VPN models. However, leaking /32 PE addresses between service providers creates some security concerns. Therefore, we see Model C typically deployed within a service provider network.

The example topology is displayed in Figure 369 and consists of two times four (2 x 4) SR OS nodes located in different autonomous systems. There is an AS interconnection from ASBR PE-4 to ASBR PE-8. PE-3 and PE-7 will act as RRs for their AS. It is assumed that an IP-VPN is already configured in each AS. Following configuration tasks should be done first:

- IS-IS or OSPF on all interfaces within each of the ASs.

- LDP on all interfaces within each of the ASs.
- MP-iBGP sessions between the PE routers and the RRs in each of the ASs, as shown in the following section.
- IP-VPN on PE-1 and on PE-5 with identical route targets.
- A loopback interface in the VRF on PE-1 and PE-5.

Configuration

The first step is to configure an MP-eBGP session between the ASBRs in both ASs. This session will be used to redistribute labeled IPv4 routes for the /32 system IP addresses between the ASs. These MP-BGP extensions are described in RFC 3107.

The configuration for ASBR PE-4 is as follows. The address family **label-ipv4** is required to enable the advertising of labeled IPv4 routes. This address family is also required on the RR neighbor in order to propagate the labeled IPv4 routes toward the other PEs in the AS.

```
configure
    router
        autonomous-system 64496
        bgp
            group EBGP
                family label-ipv4
                type external
                peer-as 64497
                neighbor 192.168.48.2
                exit
            exit
            group "IBGP"
                family vpn-ipv4 label-ipv4
                peer-as 64496
                neighbor 192.0.2.3
                exit
            exit
       exit
    exit
exit
```

Address family **label-ipv**4 is required to advertise labeled IPv4 routes toward each neighbor PE. Address family **vpn-ipv4** is required to advertise IPv4 customer routes within the AS. The configuration for RR PE-3 is as follows:

```
configure
    router
        autonomous-system 64496
        bgp
            group "IBGP"
                family vpn-ipv4 label-ipv4
```

```
                peer-as 64496
                cluster 1.1.1.1
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
    exit
exit
```

On the remaining PE nodes in AS 64496, the address families label-ipv4 and vpn-ipv4 must be enabled, as follows:

```
configure
    router
        autonomous-system 64496
        bgp
            group "IBGP"
                family vpn-ipv4 label-ipv4
                peer-as 64496
                neighbor 192.0.2.3
                exit
            exit
        exit
    exit
exit
```

The configuration for the nodes in AS 64497 is very similar. The IP addresses can be derived from Figure 369.

On ASBR PE-4, verify that the EBGP and IBGP sessions are up, as follows:

```
*A:PE-4# show router bgp summary
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 2           Total Peers                 : 2
Total BGP Paths         : 15          Total Path Memory           : 3000
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0

Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0

Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0
```

```
Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.3
               64496        8    0 00h02m11s 0/0/0 (VpnIPv4)
                            7    0           0/0/0 (Lbl-IPv4)
192.168.48.2
               64497        3    0 00h00m14s 0/0/0 (Lbl-IPv4)
                            4    0
-------------------------------------------------------------------------------
*A:PE-4#
```

Both ASBRs have MP-BGP capabilities. At this time, no prefixes have been received from the remote ASBR. To enable the advertising of labeled IPv4 routes for the system loopback interfaces, an export policy must be created and applied to the BGP session on both ASBRs. The policy configuration is as follows for ASBR PE-4. The configuration for ASBR PE-8 is very similar, the IP addresses can be derived from Figure 369.

```
configure
    router
        policy-options
            begin
            prefix-list "PE-sys"
                prefix 192.0.2.1/32 exact
                prefix 192.0.2.2/32 exact
                prefix 192.0.2.3/32 exact
                prefix 192.0.2.4/32 exact
            exit
            policy-statement "PE-sys-to-BGP"
                entry 10
                    from
```

```
                    prefix-list "PE-sys"
                exit
                to
                    protocol bgp-label
                exit
                action accept
                exit
            exit
        exit
        commit
    exit
    bgp
        group "EBGP"
            neighbor 192.168.48.2
                export "PE-sys-to-BGP"
            exit
        exit
    exit
```

After creating and applying the export policies on both ASBRs, labeled IPv4 routes will be advertised toward the remote AS for system IP addresses of the PE nodes in the local AS.

On ASBR PE-4, verify if labeled IPv4 routes have been received from ASBR PE-8:

```
*A:PE-4# show router bgp neighbor 192.168.48.2 received-routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                              LocalPref   MED
      Nexthop (Router)                                     Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
i     192.0.2.1/32                                         n/a         None
      192.168.48.2                                         None        131064
      64497 64496
i     192.0.2.2/32                                         n/a         None
      192.168.48.2                                         None        131066
      64497 64496
i     192.0.2.3/32                                         n/a         None
      192.168.48.2                                         None        131065
      64497 64496
i     192.0.2.4/32                                         n/a         None
      192.168.48.2                                         None        131067
      64497 64496
u*>i  192.0.2.5/32                                         n/a         20
      192.168.48.2                                         None        131061
      64497
u*>i  192.0.2.6/32                                         n/a         10
```

```
      192.168.48.2                                          None        131062
      64497
u*>i  192.0.2.7/32                                          n/a         10
      192.168.48.2                                          None        131060
      64497
u*>i  192.0.2.8/32                                          n/a         None
      192.168.48.2                                          None        131063
      64497
------------------------------------------------------------------------------
Routes : 8
==============================================================================
*A:PE-4#
```

The preceding output shows that four labeled IPv4 routes have been received: one route for every system IP address in the remote AS with a label attached.

More detailed information about the advertised route from PE-5 can be seen with following command on PE-4:

```
*A:PE-4# show router bgp routes 192.0.2.5/32 label-ipv4 hunt
==============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
==============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

==============================================================================
BGP Routes
==============================================================================
------------------------------------------------------------------------------
RIB In Entries
------------------------------------------------------------------------------
Network        : 192.0.2.5/32
Nexthop        : 192.168.48.2
Path Id        : None
From           : 192.168.48.2
Res. Nexthop   : 192.168.48.2
Local Pref.    : None                   Interface Name : int-PE-4-PE-8
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 20
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.8
Fwd Class      : None                   Priority       : None
IPv4 Label     : 131061
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64497
Route Tag      : 0
Neighbor-AS    : 64497
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
```

```
Last Modified  : 00h02m22s
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.5/32
Nexthop        : 192.0.2.4
Path Id        : None
To             : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 20
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
IPv4 Label     : 131060
Origin         : IGP
AS-Path        : 64497
Route Tag      : 0
Neighbor-AS    : 64497
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0

Network        : 192.0.2.5/32
Nexthop        : 192.168.48.1
Path Id        : None
To             : 192.168.48.2
Res. Nexthop   : n/a
Local Pref.    : n/a                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.8
IPv4 Label     : 131060
Origin         : IGP
AS-Path        : 64496 64497
Route Tag      : 0
Neighbor-AS    : 64496
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
*A:PE-4#
```

In the RIB In entries, the received label from PE-8 can be seen (131061). In the RIB Out entries, the locally assigned label for this prefix can be seen (131060). These labels need not match. The label mapping can also be seen with following command:

```
*A:PE-4# show router bgp inter-as-label
```

```
===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.1                     0              131064         Internal
192.0.2.2                     0              131065         Internal
192.168.48.2                  131060         131062         External
192.168.48.2                  131061         131060         External
192.168.48.2                  131062         131061         External
192.168.48.2                  131063         131063         External
192.0.2.3                     0              131066         Internal
192.0.2.4                     0              131067         Edge
-------------------------------------------------------------------------------
Total Labels allocated:   8
===============================================================================
*A:PE-4#
```

Verify that the routes have been installed in the routing table:

```
*A:PE-4# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                                Type    Proto     Age        Pref
      Next Hop[Interface Name]                                        Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                      Remote  ISIS      00h10m15s  18
       192.168.24.1                                                     20
192.0.2.2/32                                      Remote  ISIS      00h10m16s  18
       192.168.24.1                                                     10
192.0.2.3/32                                      Remote  ISIS      00h10m16s  18
       192.168.34.1                                                     10
192.0.2.4/32                                      Local   Local     00h10m17s  0
       system                                                           0
192.0.2.5/32                                      Remote  BGP_LABEL 00h03m29s  170
       192.168.48.2                                                     0
192.0.2.6/32                                      Remote  BGP_LABEL 00h03m29s  170
       192.168.48.2                                                     0
192.0.2.7/32                                      Remote  BGP_LABEL 00h03m29s  170
       192.168.48.2                                                     0
192.0.2.8/32                                      Remote  BGP_LABEL 00h03m29s  170
       192.168.48.2                                                     0
192.168.12.0/30                                   Remote  ISIS      00h10m16s  18
       192.168.24.1                                                     20
192.168.13.0/30                                   Remote  ISIS      00h10m16s  18
       192.168.34.1                                                     20
192.168.24.0/30                                   Local   Local     00h10m17s  0
       int-PE-4-PE-2                                                    0
192.168.34.0/30                                   Local   Local     00h10m17s  0
       int-PE-4-PE-3                                                    0
192.168.48.0/30                                   Local   Local     00h10m17s  0
       int-PE-4-PE-8                                                    0
-------------------------------------------------------------------------------
No. of Routes: 13
```

```
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
*A:PE-4#
```

Verify that the BGP labeled routes are further advertised toward all the PEs in the AS (through the RR) and are installed in the routing table on all PEs by using the preceding command on the other PEs.

At this point, all PEs in one AS have the /32 system IPs of the remote PEs in their routing table. All PEs in one AS have also received labels for all /32 system IPs of the remote PEs. Therefore, an MP-eBGP session can be created between the RRs in the different ASs to exchange VPN-IPv4 routes.

The configuration for RR PE-3 is as follows. The configuration for RR PE-7 is very similar. The IP addresses can be derived from Figure 370.

```
configure
    router
        bgp
            group "peer-AS-RR"
                family vpn-ipv4
                multihop 10
                peer-as 64497
                neighbor 192.0.2.7
                exit
            exit
        exit
```

On the RRs, verify that the MP-eBGP session is up:

```
*A:PE-3# show router bgp neighbor 192.0.2.7

===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                  : 192.0.2.7
Description           : (Not Specified)
Group                 : peer-AS-RR
-------------------------------------------------------------------------------
Peer AS               : 64497            Peer Port            : 49380
Peer Address          : 192.0.2.7
Local AS              : 64496            Local Port           : 179
Local Address         : 192.0.2.3
Peer Type             : External         Dynamic Peer         : No
State                 : Established      Last State           : Established
Last Event            : recvKeepAlive
Last Error            : Cease (Connection Collision Resolution)
Local Family          : VPN-IPv4
Remote Family         : VPN-IPv4
---snip---
```

The EBGP session between the two RRs is established.

The VPRNs on PE-1 in AS 64496 and PE-5 in AS 64497 are now interconnected. Packets originating in AS 64496 with a destination in AS 64497 will have 3 labels in AS 64496. Originate a VPRN ping on PE-1 toward the VPRN loopback IP address on PE-5:

```
*A:PE-1# ping router 1 10.2.2.2
PING 10.2.2.2 56 data bytes
64 bytes from 10.2.2.2: icmp_seq=1 ttl=64 time=5.76ms.
64 bytes from 10.2.2.2: icmp_seq=2 ttl=64 time=3.62ms.
64 bytes from 10.2.2.2: icmp_seq=3 ttl=64 time=5.05ms.
64 bytes from 10.2.2.2: icmp_seq=4 ttl=64 time=4.48ms.
64 bytes from 10.2.2.2: icmp_seq=5 ttl=64 time=4.40ms.
---- 10.2.2.2 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.62ms, avg = 4.66ms, max = 5.76ms, stddev = 0.713ms
*A:PE-1#
```

The top label is the LDP label to reach the exit point of the AS (PE-4). This label can be seen with following command on PE-1:

```
*A:PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1:0)
             (IPv6 LSR ID ::)
===============================================================================
Legend: U - Label In Use,  N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        LF - Lower FEC, UF - Upper FEC, e - Label ELC
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op              IngLbl    EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                            Push              --      131068
192.168.12.2                            1/1/1

192.0.2.4/32                            Swap            131068    131068
192.168.12.2                            1/1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
*A:PE-1#
```

The middle label is the label assigned by MP-BGP on the local ASBR-PE to reach the remote PE in the remote AS. This label can be seen with following command on PE-1:

```
*A:PE-1# show router bgp routes 192.0.2.5/32 label-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 192.0.2.5/32
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.12.2 (LDP)
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 20
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.4              Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
IPv4 Label     : 131060
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : 64497
Route Tag      : 0
Neighbor-AS    : 64497
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h38m04s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

The bottom label is the VPN label assigned by the remote PE in the remote AS for the destination network. This label is retrieved on PE-1, as follows:

```
*A:PE-1# show router bgp routes 10.2.2.2/32 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
```

```
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.2.2.2/32
Nexthop        : 192.0.2.5
Route Dist.    : 64497:1                VPN Label      : 131067
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : target:64496:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : 64497
Route Tag      : 0
Neighbor-AS    : 64497
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h46m52s
VPRN Imported  :  1

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
*A:PE-1#
```

Conclusion

Inter-AS option C allows the delivery of Layer 3 VPN services to customers who have sites connected in different ASs. This example shows the configuration of inter-AS option C (specific to this feature) together with the associated show output which can be used for verification and troubleshooting.

Quality of Service

In This Section

This section provides configuration information for the following topics:

- Class Fair Hierarchical Policing for SAPs
- FP and Port Queue Groups
- Pseudowire QoS
- QoS Architecture and Basic Operation

Class Fair Hierarchical Policing for SAPs

This chapter provides information to configure Class Fair Hierarchical Policing for SAPs.

Topics in this chapter include:

- Applicability
- Summary
- Overview
- Configuration
- Conclusion

Applicability

The information in this chapter is applicable to all of the Nokia 7x50 platforms and is focused on the FP2 chipset, which is used in the IOM3-XP/IMMs and in the 7750 SRc-12/4. The configuration was tested on release 9.0.R1. There are no specific pre-requisites for this configuration.

Summary

The Quality of Service (QoS) features of the 7x50 platforms provide traffic control with both shaping and policing.

Shaping is achieved using a queue; packets are placed on the queue and a scheduler removes packets from the queue at a given rate. This provides an upper bound to the traffic rate sent, thereby protecting down stream devices from bursts. However, shaping can introduce latency and jitter as packets are delayed in the queue. Packets can be dropped when the queue is full or statistically when weighted random early discard is applied. Configuration of shaping on the 7x50 is described in QoS Architecture and Basic Operation.

Policing is another mechanism for controlling traffic rates but it does not introduce latency/jitter. This is achieved using a token bucket mechanism which drops certain packets from the traffic. A common disadvantage of policing implementations is that they are usually applicable to a single level of traffic priority and have no way to fairly share capacity between multiple streams at the same priority level. Nokia's Class Fair Hierarchical Policing (CFHP) addresses these problems by implementing a four level prioritized policing hierarchy which also provides weighted fairness for traffic at a given priority.

Regardless of whether shaping or policing is being used, the preceding QoS classification and subsequent packet marking functionality is similar for both and is covered in more detail in QoS Architecture and Basic Operation.

This note describes the configuration and operation of CFHP when applied to Service Access Points (SAPs). It is also possible to use CFHP for subscribers in a Triple Play Service Delivery Architecture (TPSDA) environment but it is beyond the scope of this note.

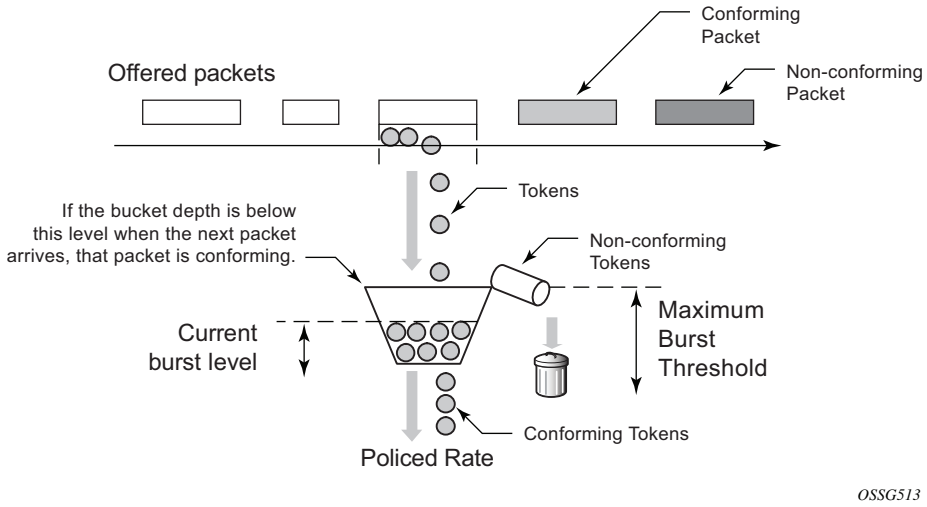
Overview

Policers

CFHP can be used both for ingress and egress QoS. The basic element is a policer which can apply both a committed information rate (CIR) and peak information rate (PIR) to a traffic flow (determined by the ingress classification). Traffic is directed to a policer by assigning a forwarding class (FC) to the policer.

To describe the operation of a policer we will use a token bucket model, this is shown in Figure 371.

Figure 371 **Policer Token Bucket Model**



The policer is modeled by a bucket being filling with tokens which represent the bytes in the packets passing through the policer. The bucket drains at a given rate (the policed rate) and if the token (byte) arrival rate exceeds the drain rate then the bucket will fill. The bucket has a maximum depth, defined by a maximum burst threshold. If tokens for a packet arrive in the bucket when the current burst level of tokens is below the maximum burst threshold then the packet is considered to be conforming and all of its tokens are accepted into the bucket. If a packet's tokens arrive when the current burst level has exceeded the maximum burst threshold then none its tokens are accepted into the bucket and the packet is considered to be non-conforming (in the representation, these tokens over-flow into a waste bin).

Table 26 shows an example of the two possibilities.

Table 26 **Burst Levels**

Maximum burst threshold = 2000 tokens (bytes) **Policed rate 2 Mbps = 250000 bytes/sec (250 tokens/ms)**				
Arrival Time	**Packet Size**	**Current Burst Level**	**Conforming Packet**	**New Burst Level**
T0	1024	1500	Yes	1500 + 1024 = 2524
T0 + 1ms	128	2524 - 250 = 2274	No	2274

When the first packet arrives the current burst level is below the maximum burst threshold so it is conforming, however, when the second packet arrives the current burst level is above the maximum burst threshold so it is non-conforming.

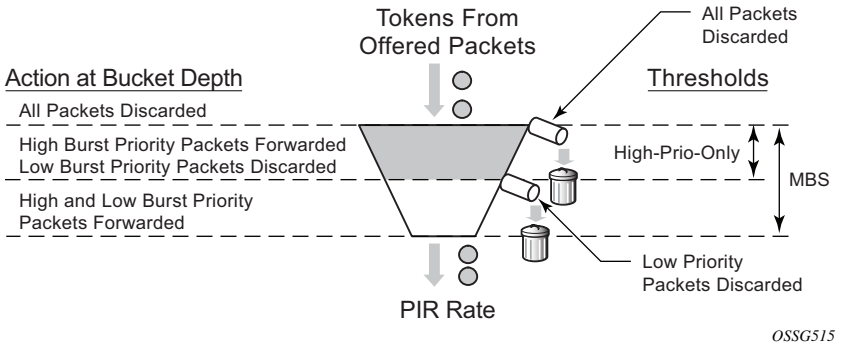
An important aspect of the implementation of hierarchical policing is the ability of a policer bucket to have multiple burst thresholds. The tokens for each arriving packet are only compared against a single threshold relating to the characteristics of packet. These burst thresholds allow specific granular QoS control.

Policer Buckets

A policer uses up to 3 buckets depending on its configuration. A PIR bucket to control the traffic rate which is always used though its rate could be max, there can be an optional CIR bucket if a CIR rate is defined for dynamically profiling (in-profile/out-of-profile) packets, finally there may be a fair information rate (FIR) bucket used to maintain traffic fairness in a hierarchical policing scenario when multiple child policers are configured at the same parent priority level.

The PIR bucket is drained at the PIR rate and has two burst thresholds, one for high burst priority traffic (defined by the maximum burst size (MBS)) and a second for low burst priority traffic (defined by the MBS minus high-prio-only), see Figure 372. The traffic burst priority is determined at ingress by the configured priority of either high or low, and at the egress by the profile state of the packets (in-profile=high, out-of-profile=low). Note that by default all FCs are low burst priority. If a packet conforms at the PIR bucket (its tokens enter the bucket) then the packet is forwarded, otherwise the packet is discarded. Discarding logically results in the packet's tokens not being placed into the CIR, FIR or parent policer buckets.

Figure 372 **Peak Information Rate (PIR) Bucket**

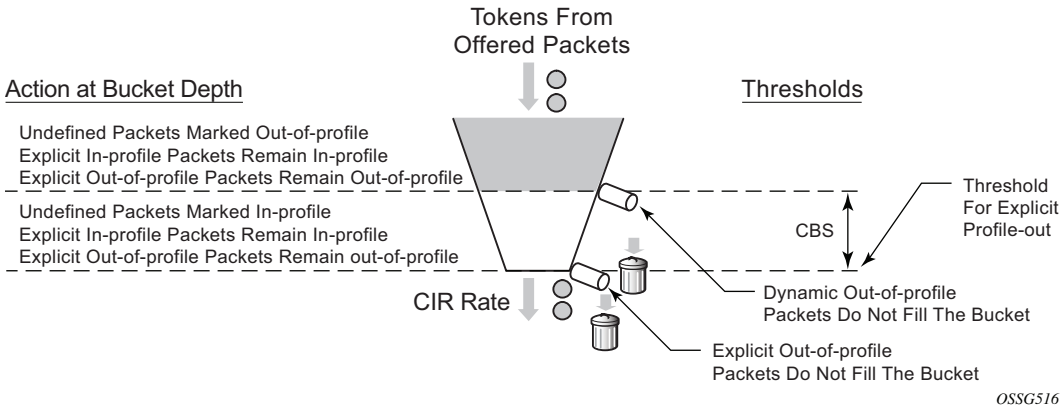


The CIR bucket is drained at the CIR rate and has one configurable burst threshold (defined by the committed burst size (CBS)). At the ingress, if the bucket level is below this threshold traffic is determined to be in-profile so the only action of the CIR bucket is to set the state of dynamically profiled packets to be either in-profile or out-of-profile. At the egress, re-profiling only affects Dot1P and DEI (Layer 2) egress marking (if the frame is double tagged, only the outer VLAN tag is remarked).

The CBS threshold is used when operating in color-blind mode, the profile of incoming packets is undefined and dynamically set based on the current burst level in the CIR bucket compared to the CBS threshold. It is also possible to operate (simultaneously) in color-aware mode, where the classification of incoming packets is used to explicitly determine whether a packet is in-profile or out-of-profile. For color-aware mode, the CIR bucket does not change the packet profile state.

In order to ensure that the overall amount of in-profile traffic takes into account both the explicit and dynamic in-profile packets, tokens from the explicit in-profile packets are allowed to fill the bucket above the CBS threshold. By doing this, dynamically profiled packets are only marked as in-profile after the token level representing dynamically in-profile and explicit in-profile packets have fallen below the CBS threshold (as the bucket drains). Note that explicitly marked out-of-profile packets remain out-of-profile, so the bottom of the bucket can be considered to be an implicit burst threshold for these packets. This is shown in Figure 373.

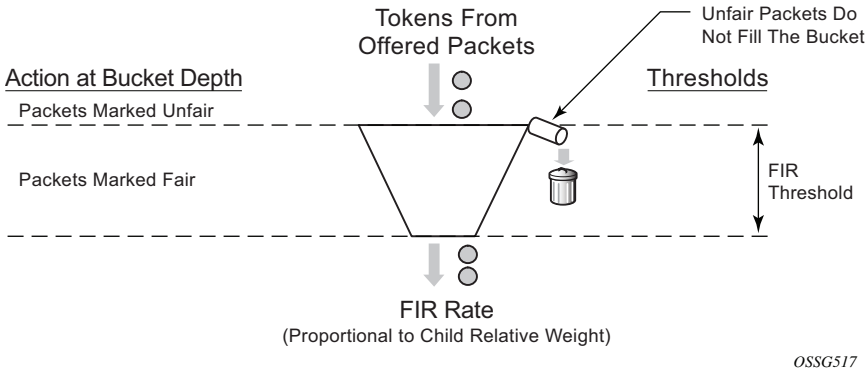
Figure 373 **Committed Information Rate (CIR) Bucket**



As the depths of the PIR and CIR buckets (MBS and CBS, respectively) are configured independently it is possible to have, for example, the CBS to be larger than the MBS (which is not possible for a queue). This could result in traffic being discarded because it is non-conforming at the PIR bucket but would have been conforming at the CIR bucket. Conversely, if the CBS is smaller than the MBS and the PIR=CIR traffic can be forwarded as out-of-profile, which would not be the case with a queue.

The FIR bucket is controlled by the system and is only used in hierarchical policing scenarios to determine a child's fair access to the available capacity at a parent priority level relative to other children at the same level. This bucket is only used when there is more than one child policer assigned to a given parent policer priority level. The drain rate of the FIR bucket is dynamically set proportionally to the weight configured for the child. This is shown in Figure 374.

Figure 374 **Fair Information Rate (FIR) Bucket**



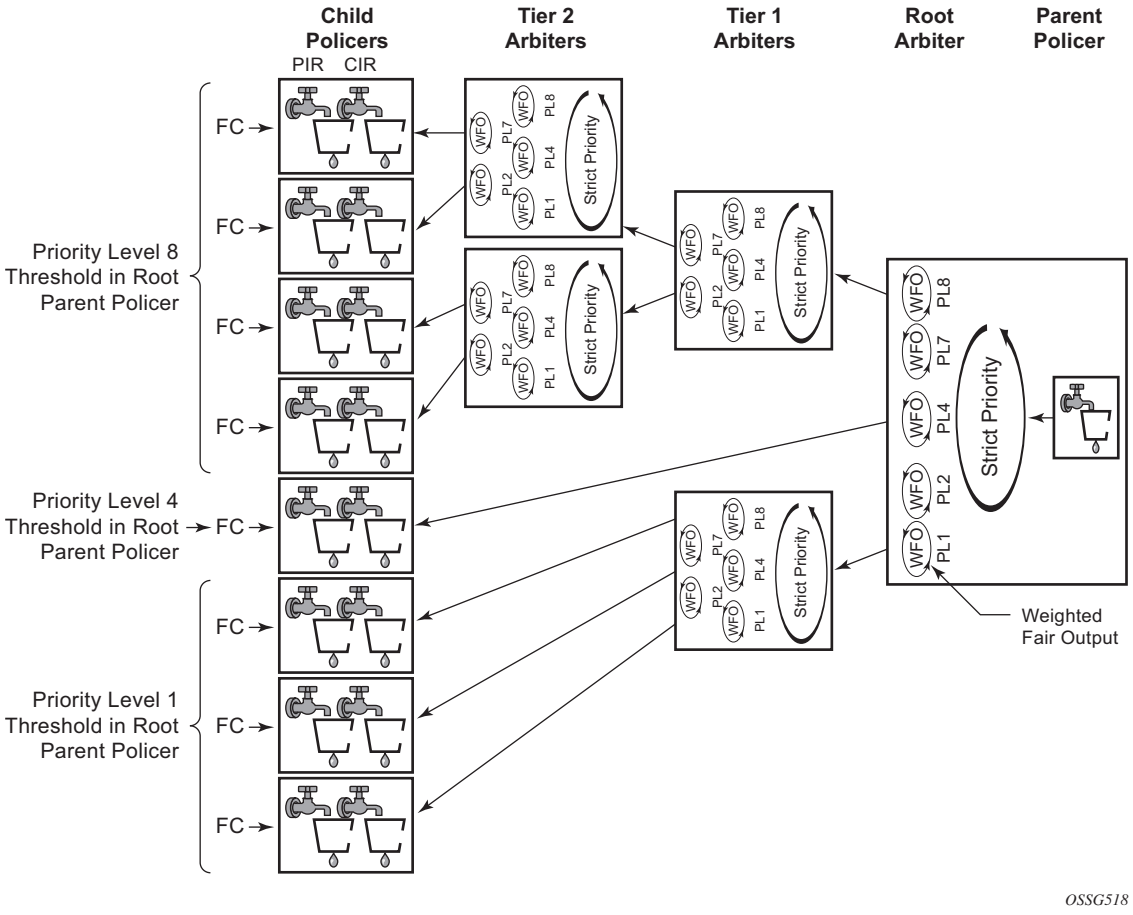
Hierarchical Policing

Policers can be used standalone or with a parent policer to provide hierarchical policing. Up to four stages can be configured in the hierarchy: the child policer, tier 1 and 2 intermediate arbiters, and a root arbiter (which is associated with the parent policer). The arbiters are logical entities that distribute bandwidth at a particular tier to their children in a priority level order, see Figure 375.

This may result in the drain rates for the child policer buckets being modified, so each child policer PIR and CIR bucket has an administrative rate value (what it is configured to) and an operational rate value (the current operating rate) based on the bandwidth distribution by the parent arbiters.

Each stage in the hierarchy connects to its parent at a priority level and a weight. There are eight available priorities which are serviced in a strict order (8 to 1, highest to lowest, respectively). The weight is used to define relative fairness when multiple children are configured in the same priority level. Note that the child access to parent policer burst capacity is governed by the level at which the child ultimately connects into the root arbiter, not by its connection level at any intermediate arbiters.

Figure 375 **Policer and Arbiter Hierarchy**



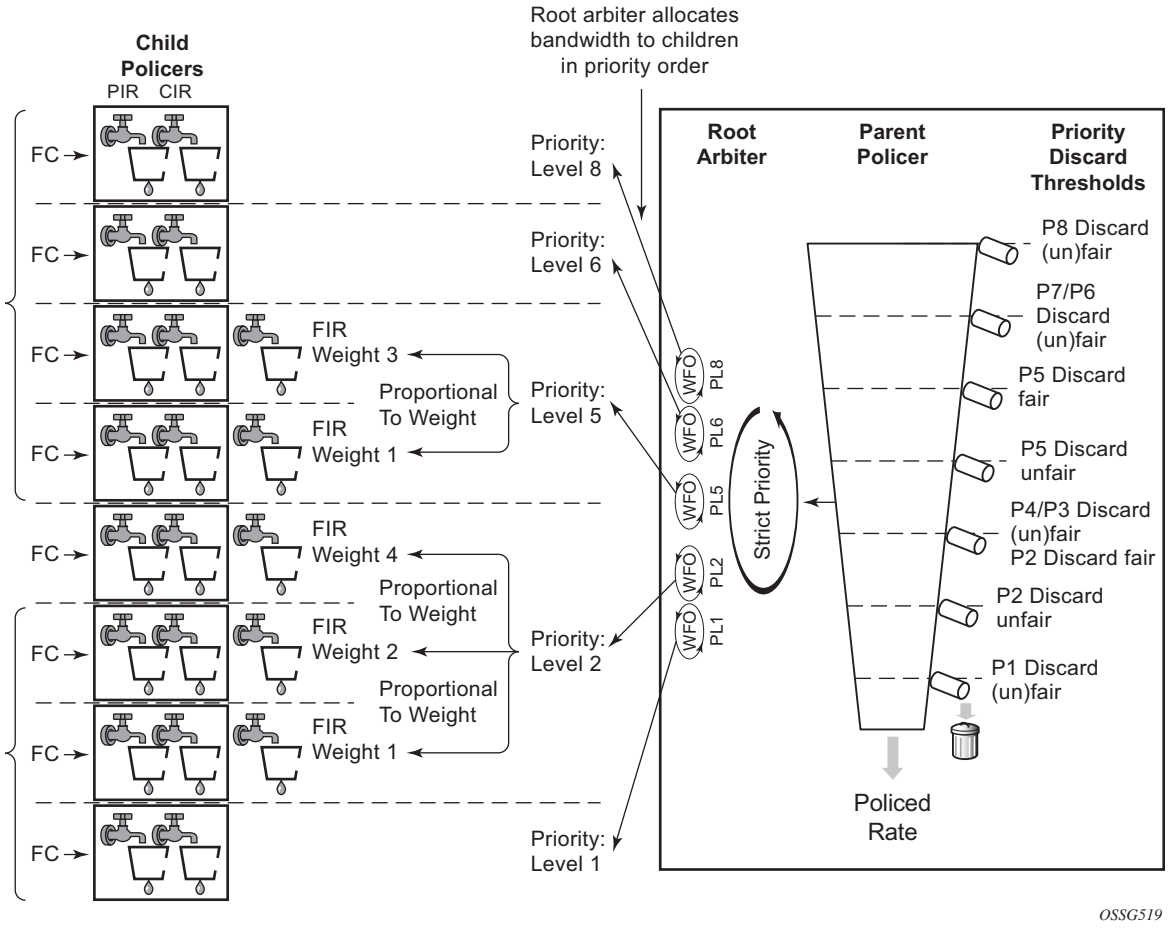
The final configuration aspect to consider is the parent policer, specifically its multiple thresholds and how they relate to the child policers. See Figure 376.

There are 8 priority levels at the parent policer, each having an associated discard-fair and discard-unfair threshold.

The discard-fair threshold is the upper burst limit for all tokens (consequently, all packets) at the given priority, all traffic at a given priority level is discarded when its tokens arrive with this threshold being exceeded. The discard-fair thresholds enable prioritization at the parent policer by having the burst capacity for each priority threshold be larger (or equal) to those of lower priorities. For example, referring to Figure 376, the priority 6 (P6) discard-fair threshold is larger than the priority 5 (P5) discard-fair threshold with the result that even if the priority 5 and below traffic is overloading the parent policer, the priority 6 traffic has burst capacity available in order to allow some of its packets to conform and get forwarded through the parent policer.

Note that if a packet is discarded at the parent policer, the discard needs to be reflected in the associated child policer, this is achieved by also logically removing the related tokens from the child policer buckets.

Figure 376 **Parent Policer and Root Arbiter**



Each priority also has a discard-unfair threshold which discards only unfair traffic of that priority, remembering that fair and unfair are determined by the FIR bucket based on the relative weights of the children.

By default, if there are no children configured at a given priority level then both its discard-fair and discard-unfair thresholds are set to zero bytes above the previous priority's discard-fair threshold.

If there is only a single child at a priority level, the discard-fair will be greater than the previous priority's discard-fair value (by an amount equal to the maximum of the min-thresh-separation and the mbs-contribution, see below) but the discard-unfair will be the same as the previous priority's discard-fair threshold (there is no need for a fairness function when there is only a single child at that priority).

If there is more than one child at a priority level, the discard-unfair threshold will be greater than the previous priority's discard-fair threshold by min-thresh-separation (see below) and the discard-fair threshold will be adjusted upwards by an amount equal to mbs-contribution minus min-thresh-separation.

The result can be summarized as follows:

- With no children at a priority level, the discard fair and unfair thresholds match the values of the previous priority.
- If there are at least two children at a priority level, the discard-unfair burst capacity equals min-thresh-separation.
- The burst capacity for a given priority level with at least one child equals the mbs-contribution, unless this is less than min-thresh-separation in which case the min-thresh-separation is used.

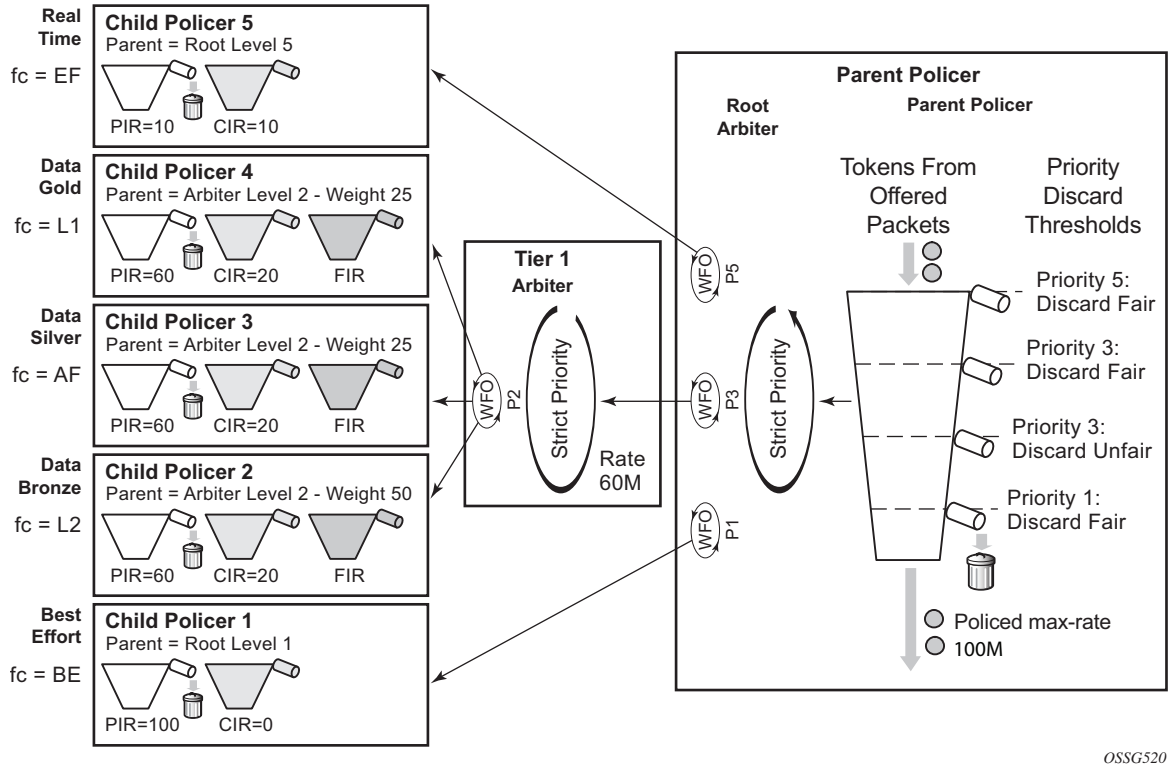
The burst tolerance for each threshold is its own burst capacity plus the sum of the burst capacities of all lower thresholds. Referring to Figure 376, the total burst capacity for priority 6 is the sum of the burst capacities for priorities 1 to 6. Note that the burst for a given FC is normally controlled by the burst allowed at the child PIR threshold, not by the parent policer.

As the burst capacity at the parent policer for a given priority level can change when adding or removing children at lower priority levels, a parameter (fixed) is available per priority threshold which causes the discard-fair and discard-unfair thresholds to be non-zero and so greater than the previous priority's thresholds, calculated as above, even when there are no children at that priority level. An exception to this is when the mbs-contribution is set to zero with the fixed parameter configured, in which case both the discard unfair and fair for that priority level are set to zero bytes above the previous level's thresholds (which results in the corresponding traffic being dropped).

A specific configuration and associated show output is included below to highlight the different threshold options described above.

The QoS example shown in Figure 377 is used to describe the configuration of CFHP.

Figure 377 **Configuration Example**



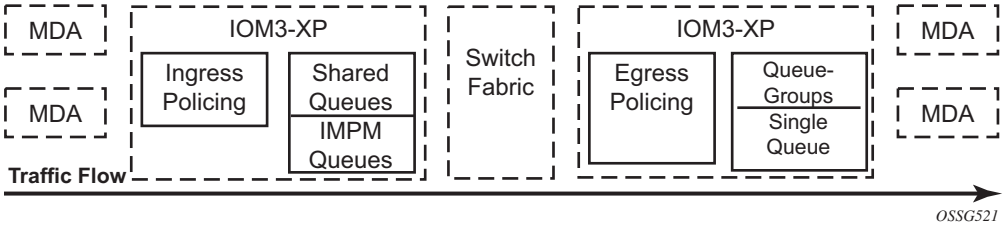
Five classes of services are accepted, each with a specific CIR and PIR. The data classes, bronze, silver and gold (L2/AF/L1), have a relative weighting of 50/25/25 at priority Level 2 of an intermediate arbiter which is constrained to 60Mbps. At the parent policer, the real time traffic (EF) is defined at level 5, with the data classes at Level 3 and a best effort class (BE) at Level 1. The overall traffic is constrained to 100Mbps at the parent policer. Only unicast traffic is policed in this example.

This example focuses on ingress policing, however, the configuration of policers, arbiters and the parent policer at the egress is almost identical to that at the ingress, the only difference being the particular statistics that can be collected.

There is a difference between ingress and egress policing in terms of how the ingress traffic accesses the switch fabric and the egress traffic access the port after it has been policed. In both cases, unicast access is enabled through a set of policer-output-queues, which are shared-queues at the ingress and queue-groups at the egress (at the egress, user defined queue-groups can be used). It is also possible to use a single service queue to access the egress port. Ingress multipoint traffic accesses the switch fabric using the Ingress Multicast Path Management (IMPM) queues.

This is shown in Figure 378 on an IOM3-XP (other line cards have the same logic).

Figure 378 **Post Policing Queues**



The differences between the ingress and egress policing configuration will be highlighted in the associated sections.

Configuration

To achieve the QoS shown in Figure 377, configure a SAP-ingress QoS policy to define the child policers and a policer-control-policy to define the intermediate arbiter and the root arbiter/parent policer. As this example is for ingress, the unicast traffic will pass through a set of shared queues called policer-output-queues, which could be modified if required.

Policers

Policers control the CIR and PIR rates for each of the traffic classes and are defined in a SAP-ingress QoS policy. The focus here are parameters related to policing.

The configuration of a child (or standalone) policer is similar to that of a queue.

```
config>qos>sap-ingress# policer policer-id [create]
    description "description-string"
    adaptation-rule [pir {max | min | closest}] [cir {max | min | closest}]
    stat-mode {no-stats|minimal|offered-profile-no-cir|
           offered-priority-no-cir|offered-profile-cir|offered-priority-cir|
           offered-total-cir|offered-limited-profile-cir}
    rate {max | kilobits-per-second} [cir {max | kilobits-per-second}]
    percent-rate pir-percent [cir cir-percent]
    mbs size [bytes | kilobytes]
    cbs size [bytes | kilobytes]
    high-prio-only [default | percent-of-mbs]
    parent {root | arbiter-name} [level level] [weight weight-within-level]
    packet-byte-offset {add bytes | subtract bytes}
```

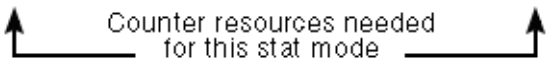
Parameters:

- description — This configures a text string, up to 80 characters, which can be used to describe the use of the policy.
- adaptation-rule — The hardware supports distinct values for the rates. This parameter tells the system how the rate configured should be mapped onto the possible hardware values. min results in the next higher hardware value being used, max results in the next lower hardware value being used and closest results in the closest available hardware value being used. As can be seen, it is possible to set the adaptation-rule independently for the CIR and PIR.
  Default: closest
- stat-mode — This defines the traffic statistics collected by the policer, summarized in Table 27.

Table 27 **Policer stat-mode**

stat-mode	Ingress		Egress	
no-stats	0	Neither policer nor parent arbiter account are required.	0	Neither policer nor parent arbiter accounting are required.
minimal (default)	1	Basic policer accounting (default).	1	Basic policer accounting (default).
offered-profile-no-cir	2	All ingress packets are either in-profile or out-of-profile.	2	Accounting for egress offered profile is required. No visibility for CIR profile state output.
offered-priority-no-cir	2	Ingress packet burst priority accounting is the primary requirement.		N/A
offered-limited-profile-cir	3	Ingress color-aware profiling is in use but packets are not being classified as in-profile.		N/A
offered-profile-cir	4	Ingress color-aware profiling is in use and packets are undefined or classified as out-of-profile or in-profile.	4	Egress profile reclassification is performed.
offered-priority-cir	4	Ingress policer is used in color-blind mode and ingress packet priority and CIR state output accounting is needed.		N/A
offered-total-cir	2	Ingress priority and ingress profile accounting is not needed. CIR profiling is in use.	2	Offered profile visibility is not required (such as, all offered packets have the same profile) and CIR profiling is in use.

Table 27 **Policer stat-mode**

stat-mode	Ingress	Egress
		

- rate and cir — The rate defines the PIR and the cir defines the CIR, both are in Kbps. The parameters rate and percent-rate are mutually exclusive and will overwrite each other when configured in the same policy.
  Range: PIR=1 to 20,000,000 Kbps or max ; CIR=0 to 20,000,000 Kbps or max
  Default: rate(PIR)=max ; cir=0
- percent-rate and cir — The percent-rate defines the PIR and the cir defines the CIR with their values being a percentage of the maximum policer rate of 20Gbps. The parameters rate and percent-rate are mutually exclusive and will overwrite each other when configured in the same policy.
  Range: pir-percent = [0.01..100.00]; cir-percent = [0.00..100.00]
  Default: pir-percent = 100; cir-percent = 0.00
- mbs and cbs — The mbs defines the MBS for the PIR bucket and the cbs defines the CBS for the CIR bucket, both can be configured in bytes or kilobytes.
  Note that the PIR MBS applies to high burst priority packets (these are packets whose classification match criteria is configured with priority high at the ingress and are in-profile packets at the egress).
  Range: mbs=0 to 4194304 bytes; cbs=0 to 4194304 bytes
  Note: mbs=0 prevents any traffic from being forwarded.
  Default: mbs=10ms of traffic or 64KB if PIR=max; cbs=10ms of traffic or 64KB if CIR=max
- high-prio-only — This defines a second burst threshold within the PIR bucket to give a maximum burst size for low burst priority packets (these are packets whose classification match criteria is configured with priority low at the ingress and are out-of-profile packets at the egress). It is configured as a percentage of the MBS.
  Default: 10%
- parent — This parameter is used when hierarchical policing is being performed and points to the parent arbiter (which could be the root arbiter or an intermediate arbiter), giving the level to which this policer connects to its parent arbiter and its relative weight compared to other children at the same level. Note that for a child policer to be associated with a parent, its stat-mode cannot be no-stats.
  Range: level=1 to 8; weight=1 to 100
  Default: level=1; weight=1

- packet-byte-offset — This changes the packet size used for accounting purposes, both in terms of the CIR and PIR rates and what is reported in the statistics. The change can either add or subtract a number of bytes. For example:
  - To have the policer work on Layer 2 frame size including inter-frame gap and preamble, add 20 bytes.
  - To have the policer work on IP packet size instead of the default layer 2 frame size, subtract the encapsulation overhead:

    14 bytes L2 + 4bytes VLAN ID + 4 bytes FCS = 22 bytes

    Range: add-bytes=0 to 31; sub-bytes=1 to 32
    Default: add-bytes=0; sub-bytes=0

A FC must be assigned to the policer in order for the policer to be instantiated (allocating a hardware policer).

By default, any unicast traffic assigned to the FC at the ingress will be processed by the policer, non-unicast traffic would continue to use the multipoint queue. At the egress all traffic assigned to the FC is processed by the policer (as there is no distinction between unicast and non-unicast traffic at the egress).

If required, non-unicast traffic can be policed in IES/VPRN and VPLS services at the ingress (note: all Epipe traffic is treated as unicast). Within an IES/VPRN service, multicast traffic can be assigned to a specific ingress policer on a PIM enabled IP interface. When the service is VPLS, broadcast, unknown unicast and multicast traffic can be individually assigned to ingress policers. In each of these cases, the policers used could be separate from the unicast policer, resulting in the instantiation of additional hardware policers, or a single policer could be used for multiple traffic types (this differs from the queuing implementation where separate queue types are used for unicast and non-unicast traffic).

```
config>qos>sap-ingress>fc#
    broadcast-policer <policer-id>
    unknown-policer <policer-id>
    multicast-policer <policer-id>
```

As mentioned above, the ingress policed unicast traffic passes through a set of shared-queues (policer-output-queues) to access the switch fabric with the multipoint traffic using the IMPM queues.

When policers are required at the egress, a SAP-egress policy is used. The configuration of the policers is almost identical to that used in the SAP-ingress policy, the only difference being the available stat-modes (as shown above).

At the egress, the policed traffic can also be directed to a specific queue-group (instead of the default policer-output-queues) and to a specific queue within that queue-group, as follows:

```
config>qos>sap-egress>fc# policer <policer-id> [group <queue-group-name> [queue
<queueid>]]
```

It is also possible to direct the egress policed traffic to a single service queue if specific egress queuing is required, as follows:

```
config>qos>sap-egress>fc# policer <policer-id> queue <queue-id>
```

Multiple egress policers in a SAP-egress policy can use the same local queue and other forwarding classes can directly use the same local queue that is being used by policers.

Parent Policer and Arbiters

The parent policer and its associated root arbiter, together with the tier 1 and 2 arbiters, are configured within a policer-control-policy.

```
config>qos# policer-control-policy policy-name [create]
    description description-string
    root
        max-rate {kilobits-per-second | max}
        priority-mbs-thresholds
            min-thresh-separation size [bytes|kilobytes]
            priority level
                mbs-contribution size [bytes|kilobytes] [fixed]
    tier 1
        arbiter arbiter-name [create]
            description escription-string
            rate {kilobits-per-second|max}
            parent root [level priority-level] [weight weight-within-level]
    tier 2
        arbiter arbiter-name [create]
            description description-string
            rate {kilobits-per-second | max}
            parent {root|arbiter-name} [level priority-level] [weight weight-within-
level]
```

Parameters:

- description — This configures a text string, up to 80 characters, which can be used to describe the use of the policy.

- root — This section defines the configuration of the parent policer and the root arbiter.
  - max-rate — This defines the policed rate of the parent policer, the rate at which the bucket is drained. It is defined in Kbps with an option to use max, in which case the maximum possible rate is used.
    Range: 1 to 20,000,000Kbps or max
    Default: max
  - priority-mbs-thresholds
    This section defines the thresholds used for the 8 priorities available in the parent policer.
    - min-thresh-separation — This defines the minimum separation between any two active thresholds in the parent policer in units of bytes or kilobytes.

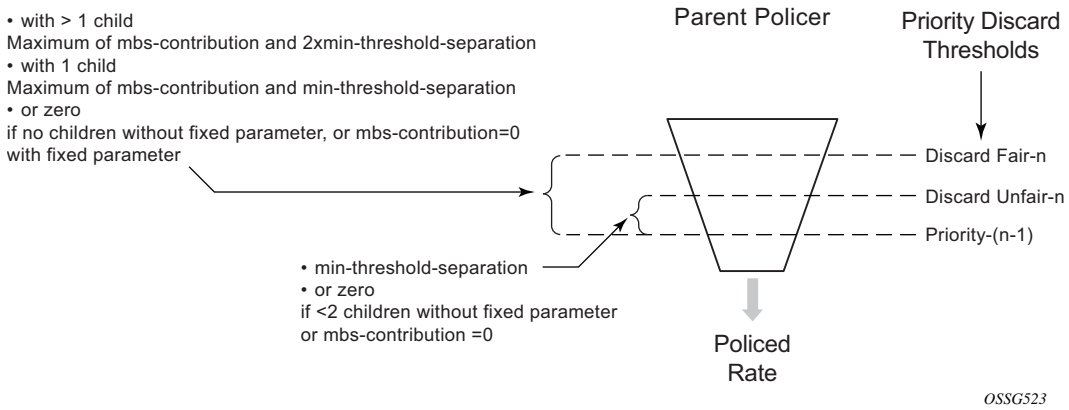
      It should be set to a value greater than the maximum packet size used for traffic passing through the policer. This ensures that a single packet arriving in the parent policer will not cause the depth of tokens to cross two burst thresholds, if this did happen it would result in the prioritization failing as a given priority level could be starved of burst capacity by a lower priority traffic.

      This parameter is also used as the burst capacity for each priority level's unfair packets.
      Range: 0 to 4194304 bytes
      Default: 1536 bytes
    - mbs-contribution — This is normally used to define the amount of packet burst capacity required at the parent policer for a particular priority level with at least one child, keeping in mind that the total capacity is the sum of this plus that of all lower thresholds. The actual burst capacity used depends also on the setting of min-thresh-separation, as described earlier. This permits the tuning of the burst capacity at the parent for any children at a given priority level. A conservative setting would ensure that the burst at the parent policer for a given priority is the sum of the bursts of all children at that priority. Less conservative settings could use a lower value and assume some level of oversubscription.

      The use of the fixed parameter causes both the fair and unfair discard thresholds to be non-zero even when there are no children assigned to this priority level (unless the mbs-contribution is set to zero).
      Range: 0 to 4194304
      Default: 8192 bytes

      The relationship between these two parameters is shown in Figure 379.

Figure 379 **Parent Policer Thresholds**



- tier

  This section defines the configuration of any intermediate tier 1 or 2 arbiters.

  – arbiter

    This specifies the name of the arbiter.

    - description — This configures a text string, up to 80 characters, which can be used to describe the use of the policy.
    - rate — This defines the rate of the arbiter, it is the maximum rate at which the arbiter will distribute burst capacity to its children. It is defined in Kbps with an option to use max, in which case the maximum possible rate is used.
      Range: 1 to 20,000,000Kbps or max
      Default: max
    - parent — This parameter is used when hierarchical policing is being performed and points to the parent arbiter (which could be the root arbiter or a tier 1 arbiter), giving the level to which this arbiter connects to its parent arbiter and its relative weight compared to other children at the same level.
      Range: level=1 to 8; weight=1 to 100
      Default: level=1; weight=1

Access to Switch Fabric and Egress Port

After the traffic has been processed by the policers it must pass through a set of queues in order to access the switch fabric at the ingress or the port at the egress.

For the ingress unicast traffic, there is a set of shared-queues (one queue per FC for each possible switch fabric destination) called policer output queues. Note that only their queue characteristics can be configured, the FC to queue mapping is fixed. Also, the PIR/CIR rates only affect the packet scheduling, they do not alter the packet profile state. The details of shared-queues are beyond the scope of this note.

```
config>qos# shared-queue "policer-output-queues"
    description description-string
    fc <fc-name> [create]
        broadcast-queue <queue-id>
        multicast-queue <queue-id>
        queue <queue-id>
        unknown-queue <queue-id>
    queue queue-id [queue-type] [multipoint] [create]
        cbs percent
        mbs percent
        high-prio-only percent
        pool pool-name
        rate percent [cir percent]
```

Multipoint traffic uses the IMPM queues to access the switch fabric. For the egress to the port, either a queue-group or a single service queue is used. There is a default queue-group called policer-output-queues or a user configured queue-group can also be used.

As mentioned above, when a policer is assigned to a specific queue-group (default or user defined) it is optionally possible to configure explicitly the queue to be used. Within the queue-group it is also possible to redirect a FC for policed traffic to a specific queue, using the FC parameter. The preference of the FC to queue mapping is (in order, highest to lowest):

1. Explicitly configured in SAP-egress FC definition
2. Mapped using FC parameter within queue-group definition
3. Default is to use queue 1

```
config>qos>qgrps>egr# queue-group queue-group-name [create]
    description description-string
    queue queue-id [queue-type] [create]
        adaptation-rule [pir adaptation-rule] [cir adaptation-rule]
        burst-limit size [bytes|kilobytes]
        cbs size-in-kbytes
        high-prio-only percent
        mbs size [bytes|kilobytes]
        parent scheduler-name [weight weight] [level level] [cir-weight cir-weight]
                              [cir-level cir-level]
        percent-rate pir-percent [cir cir-percent]
        pool pool-name
        port-parent [weight weight] [level level] [cir-weight cir-weight]
                    [cir-level cir-level]
        rate pir-rate [cir cir-rate]
        xp-specific
```

```
            wred-queue [policy slope-policy-name]
    fc fc-name [create]
        queue queue-id
```

The default policer-output-queues queue-group consists of two queues; queue 1 being best-effort and queue 2 being expedite. The lowest four FCs (BE, L2, AF, L1) are assigned to queue 1 and the highest four queues (H2, EF, H1, NC) are assigned to queue 2. It may be important to change the queue 2 definition in the queue-group to have CIR=PIR when there are other best-effort queues using a non-zero CIR on the same egress port. This ensures that the policed traffic using queue 2 will be scheduled before any other best-effort within CIR traffic. It also results in the queue CBS being non-zero, allowing the queue 2 traffic access to reserved buffer space.

```
A:PE-1>config>qos# queue-group-templates egress queue-group "policer-output-queues"
A:PE-1>cfg>qos>qgrps>egr>qgrp# info detail
----------------------------------------------
                    description "Default egress policer output queues."
                    queue 1 best-effort create
                        no parent
                        no port-parent
                        adaptation-rule pir closest cir closest
                        rate max cir 0
                        cbs default
                        mbs default
                        high-prio-only default
                        no pool
                        xp-specific
                            no wred-queue
                        exit
                        no burst-limit
                    exit
                    queue 2 expedite create
                        no parent
                        no port-parent
                        adaptation-rule pir closest cir closest
                        rate max cir 0
                        cbs default
                        mbs default
                        high-prio-only default
                        no pool
                        xp-specific
                            no wred-queue
                        exit
                        no burst-limit
                    exit
                    fc af create
                        queue 1
                    exit
                    fc be create
                        queue 1
                    exit
                    fc ef create
                        queue 2
                    exit
```

```
fc h1 create
    queue 2
exit
fc h2 create
    queue 2
exit
fc l1 create
    queue 1
exit
fc l2 create
    queue 1
exit
fc nc create
    queue 2
exit
```

The remaining details of queue-groups are beyond the scope of this section.

Applying the SAP Ingress and Policer Control Policy

The SAP ingress policy and policer control policy are both applied under the associated SAP. After applying these, it is possible to override the configuration of specific policers and/or the policer control policy. This is shown below. The parameter values are the same as detailed for the policies, as above.

```
config>service><service>#
    sap sap-id [create]
        [ingress|egress]
            qos policy-id
            policer-control-policy policy-name
            policer-override
                policer policer-id [create]
cbs size [bytes|kilobytes]
     mbs size [bytes|kilobytes]
packet-byte-offset {add add-bytes | subtract sub-bytes}
rate {rate | max} [cir {max | rate}
percent-rate <pir-percent> [cir <cir-percent>]
stat-mode stat-mode
            policer-control-override [create]
                max-rate {rate | max}
                priority-mbs-thresholds
min-threshold-separation size [bytes | kilobytes]
priority level
                        mbs-contribution size [bytes | kilobytes]
```

The SAP ingress policy and policer control policy required for the configuration example in Figure 377 is shown below.

```
#------------------------------------------------
echo "QoS Policy Configuration"
```

```
#--------------------------------------------------
    qos
        policer-control-policy "cfhp-1" create
            root
                max-rate 100000
            exit
            tier 1
                arbiter "a3" create
                    parent "root" level 3
                    rate 60000
                exit
            exit
        exit
        sap-ingress 10 create
            queue 1 create
            exit
            queue 11 multipoint create
            exit
            policer 1 create
                stat-mode offered-total-cir
                parent "root"
                rate 100000
                high-prio-only 0
            exit
            policer 2 create
                stat-mode offered-total-cir
                parent "a3" level 2 weight 50
                rate 60000 cir 20000
                high-prio-only 0
            exit
            policer 3 create
                stat-mode offered-total-cir
                parent "a3" level 2 weight 25
                rate 60000 cir 20000
                high-prio-only 0
            exit
            policer 4 create
                stat-mode offered-total-cir
                parent "a3" level 2 weight 25
                rate 60000 cir 20000
                high-prio-only 0
            exit
            policer 5 create
                stat-mode offered-total-cir
                parent "root" level 5
                rate 10000 cir 10000
                high-prio-only 0
            exit
            fc "af" create
                policer 3
            exit
            fc "be" create
                policer 1
            exit
            fc "ef" create
                policer 5
            exit
            fc "l1" create
                policer 4
```

```
            exit
            fc "l2" create
                policer 2
            exit
            dot1p 1 fc "be"
            dot1p 2 fc "l2"
            dot1p 3 fc "af"
            dot1p 4 fc "l1"
            dot1p 5 fc "ef"
        exit
```

Traffic is classified based on dot1p values, each of which is assigned to an individual FC which in turn is assigned to a policer. The policer rates are configured as required for the example with an appropriate stat-mode. Default values are used for the policer burst thresholds. As all FCs are low burst priority by default, the high-prio-only has been set to zero in order to allow the traffic to use all of the MBS available at the PIR bucket.

Policers 2, 3 and 4 are parented to the arbiter “a3” with the required weights and at a single level (Level 2). In this example it does not matter which level of “a3” is used to parent these policers, the important aspect is the level at which “a3” is parented to the root. Consequently, these policers use the Level 3 parent policer thresholds (not the level they are parented on a“a3” not Level 2). Arbiter “a3” has a rate of 60Mbps so that its children cannot exceed this rate (except up to the burst tolerances).

Policers 1 and 5 are directly parented to the root arbiter, together with tier 1 arbiter “a3”.

The total capacity for the 5 traffic streams is constrained to 100Mbps by the parent policer, again with the default burst tolerances at the root arbiter.

The SAP-ingress and policer-control-policies are applied to a SAP within an Epipe.

```
#--------------------------------------------------
echo "Service Configuration"
#--------------------------------------------------
    service
        epipe 1 customer 1 create
            sap 1/1/3:1 create
                ingress
                    policer-control-policy "cfhp-1"
                    qos 10
                exit
            exit
            sap 1/1/4:1 create
            exit
            no shutdown
        exit
    exit
```

The following configuration is used to highlight the relative thresholds in the parent policer when a priority level has 0, 1 or 2 associated children, both with and without using the fixed parameter.

```
------------------------------------------------
echo "QoS Policy Configuration"
#------------------------------------------------
    qos
       policer-control-policy "cfhp-2" create
            root
                max-rate 100000
                priority-mbs-thresholds
                    min-thresh-separation 256 bytes
                    priority 1
                        mbs-contribution 1 kilobytes
                    exit
                    priority 2
                        mbs-contribution 1 kilobytes
                    exit
                    priority 3
                        mbs-contribution 1 kilobytes
                    exit
                    priority 4
                        mbs-contribution 1 kilobytes fixed
                    exit
                    priority 5
                        mbs-contribution 1 kilobytes fixed
                    exit
                    priority 6
                        mbs-contribution 1 kilobytes fixed
                    exit
                exit
            exit
        exit
        sap-ingress 20 create
            queue 1 create
            exit
            queue 11 multipoint create
            exit
            policer 1 create
                parent "root" level 2
            exit
            policer 2 create
                parent "root" level 3
            exit
            policer 3 create
                parent "root" level 3
            exit
            policer 4 create
                parent "root" level 5
            exit
            policer 5 create
                parent "root" level 6
            exit
            policer 6 create
                parent "root" level 6
            exit
            fc "af" create
```

```
                policer 3
            exit
            fc "be" create
                policer 1
            exit
            fc "ef" create
                policer 6
            exit
            fc "h2" create
                policer 5
            exit
            fc "l1" create
                policer 4
            exit
            fc "l2" create
                policer 2
            exit
        exit
#--------------------------------------------------
echo "Service Configuration"
#--------------------------------------------------
    service
        epipe 2 customer 1 create
            sap 1/1/3:2 create
                ingress
                    policer-control-policy "cfhp-2"
                    qos 20
                exit
            exit
            sap 1/1/4:2 create
            exit
            no shutdown
        exit
```

A policer-control-policy can also be applied under a multi-service site (MSS) so that the hierarchical policing applies to traffic on multiple SAPs, potentially from different services. The MSS can only be assigned to a port, which could be a LAG, but it is not possible to assign an MSS to a card. When MSS are used, policer overrides are not supported.

```
config>service><service>#
    service
        customer customer-id [create]
            multi-service-site customer-site-name [create]
                assignment port port-id
                egress
                    policer-control-policy name
                ingress
                    policer-control-policy name
        service-type
            sap sap-id
                multi-service-site customer-site-name
                ingress
                    qos policy-id
                egress
                    qos policy-id
```

Show Output

After configuring the example as described in the previous section, steady state traffic was sent through the Epipe to overload each of the policers and the show output below was collected. This output focuses on the policer and arbiter details.

The following shows the policers on the SAP and their current state.

```
A:PE-1# show qos policer sap 1/1/3:1
===============================================================================
Policer Information (Summary), Slot 1
===============================================================================
-------------------------------------------------------------------------------
Name                FC-Maps       MBS       HP-Only A.PIR    A.CIR
Direction                         CBS       Depth   O.PIR    O.CIR    O.FIR
-------------------------------------------------------------------------------
1->1/1/3:1->1
Ingress             be            124 KB    0 KB    100000   0
                                  0 KB      82      30000    0        30000
1->1/1/3:1->2
Ingress             l2            76 KB     0 KB    60000    20000
                                  25 KB     77846   30000    20000    30000
1->1/1/3:1->3
Ingress             af            76 KB     0 KB    60000    20000
                                  25 KB     77824   15000    15000    15000
1->1/1/3:1->4
Ingress             l1            76 KB     0 KB    60000    20000
                                  25 KB     77868   15000    15000    15000
1->1/1/3:1->5
Ingress             ef            12800 B   0 KB    10000    10000
                                  12800 B   12834   10000    10000    10000
===============================================================================
A:PE-1#
```

The output above shows the configured values for the policers, e.g. PIR and CIR, together with their operational (current) state, such as PIR, CIR and FIR. The depth of each of the PIR buckets is also shown.

The detailed state of each policer can be seen by adding the parameter detail. The following is the output for policer 3.

```
A:PE-1# show qos policer sap 1/1/3:1 ingress detail
...
===============================================================================
Policer Info (1->1/1/3:1->3), Slot 1
===============================================================================
Policer Name       : 1->1/1/3:1->3
Direction          : Ingress            Fwding Plane       : 1
FC-Map             : af
Depth PIR          : 77842 Bytes        Depth CIR          : 25618 Bytes
Depth FIR          : 77842 Bytes
MBS                : 76 KB              CBS                : 25 KB
Hi Prio Only       : 0 KB               Pkt Byte Offset    : 0
```

```
Admin PIR          : 60000 Kbps         Admin CIR          : 20000 Kbps
Oper PIR           : 15000 Kbps         Oper CIR           : 15000 Kbps
Oper FIR           : 15000 Kbps
Stat Mode          : offered-total-cir
PIR Adaption       : closest            CIR Adaption       : closest
Parent Arbiter Name: a3
-------------------------------------------------------------------------------
Arbiter Member Information
-------------------------------------------------------------------------------
Offered Rate       : 45800 Kbps
Level              : 2                  Weight             : 25
Parent PIR         : 15000 Kbps         Parent FIR         : 15000 Kbps
Consumed           : 15000 Kbps
-------------------------------------------------------------------------------
===============================================================================...
A:PE-1#
```

Notice that the above output shows the depth of the PIR, CIR and FIR buckets together with their operational rates. This can be used to explain the operation of the policers in this example and is discussed later in this section.

The stat-mode of offered-total-cir configured on policer 3 results in these statistics being collected.

```
A:PE-1# show service id 1 sap 1/1/3:1 stats
===============================================================================
...
-------------------------------------------------------------------------------
Sap per Policer stats
-------------------------------------------------------------------------------
                        Packets                 Octets

Ingress Policer 1 (Stats mode: offered-total-cir)
Off. All              : 2690893                 172217152
Dro. InProf           : 0                       0
Dro. OutProf          : 967465                  61917760
For. InProf           : 0                       0
For. OutProf          : 1723428                 110299392

Ingress Policer 2 (Stats mode: offered-total-cir)
Off. All              : 2690988                 172223232
Dro. InProf           : 0                       0
Dro. OutProf          : 909492                  58207488
For. InProf           : 1178507                 75424448
For. OutProf          : 602989                  38591296
...
```

The following output is included for reference and shows the statistics which are collected for each of the ingress and egress stat-modes.

```
PE-1# show service id 2 sap 1/1/1:2 stats
...
-------------------------------------------------------------------------------
```

```
Sap per Policer stats
-------------------------------------------------------------------------------
                        Packets                 Octets

Ingress Policer 1 (Stats mode: no-stats)

Ingress Policer 2 (Stats mode: minimal)
Off. All              : 0                       0
For. All              : 0                       0
Dro. All              : 0                       0

Ingress Policer 3 (Stats mode: offered-profile-no-cir)
Off. InProf           : 0                       0
Off. OutProf          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0

Ingress Policer 4 (Stats mode: offered-priority-no-cir)
Off. HiPrio           : 0                       0
Off. LowPrio          : 0                       0
For. HiPrio           : 0                       0
For. LoPrio           : 0                       0
Dro. HiPrio           : 0                       0
Dro. LowPrio          : 0                       0

Ingress Policer 5 (Stats mode: offered-profile-cir)
Off. InProf           : 0                       0
Off. OutProf          : 0                       0
Off. Uncolor          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0

Ingress Policer 6 (Stats mode: offered-priority-cir)
Off. HiPrio           : 0                       0
Off. LowPrio          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0

Ingress Policer 7 (Stats mode: offered-total-cir)
Off. All              : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0

Ingress Policer 8 (Stats mode: offered-limited-profile-cir)
Off. OutProf          : 0                       0
Off. Uncolor          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
```

```
Egress Policer 1 (Stats mode: no-stats)

Egress Policer 2 (Stats mode: minimal)
Off. All              : 0                       0
For. All              : 0                       0
Dro. All              : 0                       0

Egress Policer 3 (Stats mode: offered-profile-no-cir)
Off. InProf           : 0                       0
Off. OutProf          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0

Egress Policer 4 (Stats mode: offered-profile-cir)
Off. InProf           : 0                       0
Off. OutProf          : 0                       0
Off. Uncolor          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0

Egress Policer 5 (Stats mode: offered-total-cir)
Off. All              : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
===============================================================================
```

It is possible to show the policer-control-policy details and the SAPs with which it is associated, as shown here.

```
A:PE-1# show qos policer-control-policy cfhp-1
===============================================================================
QoS Policer Control Policy
===============================================================================
Policy-Name        : cfhp-1
Description        : (Not Specified)
Min Threshold Sep  : Def

-------------------------------------------------------------------------------
Priority MBS Thresholds
-------------------------------------------------------------------------------
Priority            MBS Contribution
-------------------------------------------------------------------------------
1                   none
2                   none
3                   none
4                   none
5                   none
6                   none
7                   none
8                   none
```

```
-------------------------------------------------------------------------------
Tier/Arbiter                        Lvl/Wt      Rate      Parent
-------------------------------------------------------------------------------
  root                              N/A         100000    None
1 a3                                3/1         60000     root

===============================================================================
A:PE-1# show qos policer-control-policy "cfhp-1" association

===============================================================================
QoS Policer Control Policy
===============================================================================
Policy-Name        : cfhp-1
Description        : (Not Specified)

-------------------------------------------------------------------------------
Associations
-------------------------------------------------------------------------------
Service-Id         : 1 (Epipe)          Customer-Id        : 1
 - SAP : 1/1/3:1 (Ing)

===============================================================================
A:PE-1
```

The following command shows the policer hierarchy, including the child policers and their relationship to the intermediate arbiter (a3) and the root arbiter. It can be used to monitor the status of the child policers in the hierarchy. The output shows the assigned, offered and consumed capacity for each policer.

```
A:PE-1# show qos policer-hierarchy sap 1/1/3:1
===============================================================================
Policer Hierarchy - Sap 1/1/3:1
===============================================================================
Ingress Policer Control Policy : cfhp-1
Egress Policer Control Policy  :
-------------------------------------------------------------------------------
root (Ing)
|
| slot(1)
|
|--(A) : a3 (Sap 1/1/3:1)
|   |
|   |--(P) : Policer 1->1/1/3:1->4
|   |   |
|   |   |    [Level 2 Weight 25]
|   |   |    Assigned PIR:15000      Offered:45800
|   |   |    Consumed:15000
|   |   |
|   |   |    Assigned FIR:15000
|   |
|   |--(P) : Policer 1->1/1/3:1->3
|   |   |
|   |   |    [Level 2 Weight 25]
|   |   |    Assigned PIR:15000      Offered:45800
|   |   |    Consumed:15000
```

```
|   |   |
|   |   |    Assigned FIR:15000
|   |
|   |--(P) : Policer 1->1/1/3:1->2
|   |   |
|   |   |    [Level 2 Weight 50]
|   |   |    Assigned PIR:30000      Offered:45800
|   |   |    Consumed:30000
|   |   |
|   |   |    Assigned FIR:30000
|
|--(P) : Policer 1->1/1/3:1->5
|   |
|   |    [Level 5 Weight 1]
|   |    Assigned PIR:10000      Offered:10000
|   |    Consumed:10000
|   |
|   |    Assigned FIR:10000
|
|--(P) : Policer 1->1/1/3:1->1
|   |
|   |    [Level 1 Weight 1]
|   |    Assigned PIR:30000      Offered:45800
|   |    Consumed:30000
|   |
|   |    Assigned FIR:30000
root (Egr)
|
No Active Members Found on slot 1
===============================================================================
A:PE-1#
```

The complete information about the policer hierarchy can be seen by adding the detail parameter, as shown below, with alternative parameters to select more specific information.

- root-detail — Rates, depth and thresholds for the root arbiter.
- thresholds — CBS, MBS and high-prio-only thresholds with associated rates of child policers.
- priority-info — Discard-fair and discard-unfair thresholds, with number of associated children, for each of the root priority levels.
- depth — Parent policer and child PIR buckets depth, with PIR and FIR rate information.
- arbiter — Specific information of a given arbiter.
- port — For use with LAGs in different line cards or using adapt-qos link.

The output adds a good representation of the root arbiter thresholds, indicating the priority levels, discard-unfair and discard-fair thresholds, and how many child policers are associated with each level. It also includes the current depth of the child policer PIR buckets and the parent policer bucket.

```
A:PE-1# show qos policer-hierarchy sap 1/1/3:1 detail
===============================================================================
Policer Hierarchy - Sap 1/1/3:1
===============================================================================
Ingress Policer Control Policy : cfhp-1
Egress Policer Control Policy  :
-------------------------------------------------------------------------------
Legend :
(*) real-time dynamic value
(w) Wire rates
-------------------------------------------------------------------------------
root (Ing)
|
| slot(1)
|    MaxPIR:100000
|    ConsumedByChildren:100000
|    OperPIR:100000      OperFIR:100000
|
|    DepthPIR:8111 bytes
|  Priority 8
|    Oper Thresh Unfair:17408       Oper Thresh Fair:25600
|    Association count:0
|  Priority 7
|    Oper Thresh Unfair:17408       Oper Thresh Fair:25600
|    Association count:0
|  Priority 6
|    Oper Thresh Unfair:17408       Oper Thresh Fair:25600
|    Association count:0
|  Priority 5
|    Oper Thresh Unfair:17408       Oper Thresh Fair:25600
|    Association count:1
|  Priority 4
|    Oper Thresh Unfair:9728        Oper Thresh Fair:17408
|    Association count:0
|  Priority 3
|    Oper Thresh Unfair:9728        Oper Thresh Fair:17408
|    Association count:3
|  Priority 2
|    Oper Thresh Unfair:0           Oper Thresh Fair:8192
|    Association count:0
|  Priority 1
|    Oper Thresh Unfair:0           Oper Thresh Fair:8192
|    Association count:1
|
|--(A) : a3 (Sap 1/1/3:1)
|   |    MaxPIR:60000
|   |    ConsumedByChildren:60000
|   |    OperPIR:60000       OperFIR:60000
|   |
|   |    [Level 3 Weight 1]
|   |    Assigned PIR:60000      Offered:60000
|   |    Consumed:60000
|   |
|   |    Assigned FIR:60000
|   |
|   |--(P) : Policer 1->1/1/3:1->4
|   |   |    MaxPIR:60000        MaxCIR:20000
|   |   |    CBS:25600           MBS:77824
|   |   |    HiPrio:0
```

```
|   |   |   Depth:77876
|   |   |
|   |   |   OperPIR:15000        OperCIR:15000
|   |   |   OperFIR:15000
|   |   |   PacketByteOffset:0
|   |   |   StatMode: offered-total-cir
|   |   |
|   |   |   [Level 2 Weight 25]
|   |   |   Assigned PIR:15000      Offered:45800
|   |   |   Consumed:15000
|   |   |
|   |   |   Assigned FIR:15000
|   |
|   |--(P) : Policer 1->1/1/3:1->3
|   |   |   MaxPIR:60000        MaxCIR:20000
|   |   |   CBS:25600           MBS:77824
|   |   |   HiPrio:0
|   |   |   Depth:77834
|   |   |
|   |   |   OperPIR:15000        OperCIR:15000
|   |   |   OperFIR:15000
|   |   |   PacketByteOffset:0
|   |   |   StatMode: offered-total-cir
|   |   |
|   |   |   [Level 2 Weight 25]
|   |   |   Assigned PIR:15000      Offered:45800
|   |   |   Consumed:15000
|   |   |
|   |   |   Assigned FIR:15000
|   |
|   |--(P) : Policer 1->1/1/3:1->2
|   |   |   MaxPIR:60000        MaxCIR:20000
|   |   |   CBS:25600           MBS:77824
|   |   |   HiPrio:0
|   |   |   Depth:77848
|   |   |
|   |   |   OperPIR:30000        OperCIR:20000
|   |   |   OperFIR:30000
|   |   |   PacketByteOffset:0
|   |   |   StatMode: offered-total-cir
|   |   |
|   |   |   [Level 2 Weight 50]
|   |   |   Assigned PIR:30000      Offered:45800
|   |   |   Consumed:30000
|   |   |
|   |   |   Assigned FIR:30000
|
|--(P) : Policer 1->1/1/3:1->5
|   |   MaxPIR:10000        MaxCIR:10000
|   |   CBS:12800           MBS:12800
|   |   HiPrio:0
|   |   Depth:12854
|   |
|   |   OperPIR:10000        OperCIR:10000
|   |   OperFIR:10000
|   |   PacketByteOffset:0
|   |   StatMode: offered-total-cir
|   |
|   |   [Level 5 Weight 1]
```

```
|   |    Assigned PIR:10000      Offered:10000
|   |    Consumed:10000
|   |
|   |    Assigned FIR:10000
|
|--(P) : Policer 1->1/1/3:1->1
|   |    MaxPIR:100000       MaxCIR:0
|   |    CBS:0               MBS:126976
|   |    HiPrio:0
|   |    Depth:135
|   |
|   |    OperPIR:30000        OperCIR:0
|   |    OperFIR:30000
|   |    PacketByteOffset:0
|   |    StatMode: offered-total-cir
|   |
|   |    [Level 1 Weight 1]
|   |    Assigned PIR:30000      Offered:45800
|   |    Consumed:30000
|   |
|   |    Assigned FIR:30000


root (Egr)
|
No Active Members Found on slot 1

===============================================================================
A:PE-1#
```

The output above gives the depth of the parent policer, which can be used with the output below to explain the operation of the policing in this example.

```
A:PE-1# show qos policer sap 1/1/3:1 detail | match expression "Slot | Bytes | Kbps"
Policer Info (1->1/1/3:1->1), Slot 1
Depth PIR          : 153 Bytes          Depth CIR          : 0 Bytes
Depth FIR          : 153 Bytes
Admin PIR          : 100000 Kbps        Admin CIR          : 0 Kbps
Oper PIR           : 30000 Kbps         Oper CIR           : 0 Kbps
Oper FIR           : 30000 Kbps
Offered Rate       : 45800 Kbps
Parent PIR         : 30000 Kbps         Parent FIR         : 30000 Kbps
Consumed           : 30000 Kbps
Policer Info (1->1/1/3:1->2), Slot 1
Depth PIR          : 77828 Bytes        Depth CIR          : 25624 Bytes
Depth FIR          : 77828 Bytes
Admin PIR          : 60000 Kbps         Admin CIR          : 20000 Kbps
Oper PIR           : 30000 Kbps         Oper CIR           : 20000 Kbps
Oper FIR           : 30000 Kbps
Offered Rate       : 45800 Kbps
Parent PIR         : 30000 Kbps         Parent FIR         : 30000 Kbps
Consumed           : 30000 Kbps
Policer Info (1->1/1/3:1->3), Slot 1
Depth PIR          : 77858 Bytes        Depth CIR          : 25634 Bytes
Depth FIR          : 77858 Bytes
Admin PIR          : 60000 Kbps         Admin CIR          : 20000 Kbps
Oper PIR           : 15000 Kbps         Oper CIR           : 15000 Kbps
```

```
Oper FIR           : 15000 Kbps
Offered Rate       : 45800 Kbps
Parent PIR         : 15000 Kbps         Parent FIR         : 15000 Kbps
Consumed           : 15000 Kbps
Policer Info (1->1/1/3:1->4), Slot 1
Depth PIR          : 77838 Bytes        Depth CIR          : 25614 Bytes
Depth FIR          : 77838 Bytes
Admin PIR          : 60000 Kbps         Admin CIR          : 20000 Kbps
Oper PIR           : 15000 Kbps         Oper CIR           : 15000 Kbps
Oper FIR           : 15000 Kbps
Offered Rate       : 45800 Kbps
Parent PIR         : 15000 Kbps         Parent FIR         : 15000 Kbps
Consumed           : 15000 Kbps
Policer Info (1->1/1/3:1->5), Slot 1
Depth PIR          : 12814 Bytes        Depth CIR          : 12814 Bytes
Depth FIR          : 12814 Bytes
Admin PIR          : 10000 Kbps         Admin CIR          : 10000 Kbps
Oper PIR           : 10000 Kbps         Oper CIR           : 10000 Kbps
Oper FIR           : 10000 Kbps
Offered Rate       : 10000 Kbps
Parent PIR         : 10000 Kbps         Parent FIR         : 10000 Kbps
Consumed           : 10000 Kbps
A:PE-1#
```

From the output above, it can be seen that the offered rate for policers 1-4 is 45800Kbps, in fact it is the same for policer 5 but this is capped at the admin PIR rate, 10000Kbps.

The depth of the parent policer is only 8111 bytes, so this is not causing any discarding of priority 2-5 traffic at the parent policer as their discard thresholds are all above this value. Therefore the drops in policers 2-5 are all occurring in the child policers.

Policer 5 is consuming all of its operational capacity (PIR, CIR and FIR), and it can be seen that the level of the PIR bucket is 12814 bytes, which is slightly above its MBS of 12800 bytes. The level of the PIR bucket will oscillate around the MBS value as tokens are added to exceed the threshold (causing discards) then the draining reduces the level to just below the threshold (allowing forwarding).

Policers 2-4 are functioning in the same way as policer 5, as can be seen from their PIR bucket levels (levels are 77828 bytes with MBS of 77824), resulting in the PIR buckets constraining the rates of the traffic through these policers. This is happening because the arbiter “a3” is distributing its 60000Kbps in the configured ratio to these policers, which changes the operational PIR to 30000Kbps for policer 2 and 15000Kbps for policers 3 and 4, all being below the offered traffic rate. A similar effect can be seen with the CIR rates and bucket depths, as the operational CIR rate

of policer 2 has reached its administrative value with those of policer 3 and 4 being constrained by the operational PIR. The CIR bucket depths are just above the CBS, again this will oscillate causing traffic to both in-profile and out-of-profile. As this is steady state traffic, the operational FIR rates for these policers have settled to match their operational PIR rates.

Policer 1 is also discarding traffic at the PIR bucket but it is also discarding traffic at the parent policer. This can be seen by the fact that policer 1 PIR depth is nowhere near its MBS whereas the parent policer level is just below the priority 1 discard-fair threshold. The level of the parent policer bucket will oscillate around this threshold causing policer 1 traffic to be discarded, which in turn is reflected back into the level of tokens in the policer 1 PIR bucket.

As this example is based on ingress unicast policing, the traffic exits the policers and then accesses the switch fabric using a set of shared-queue (policer-output-queues). The parameters for these queues can be seen using the following **show** command.

```
A:PE-1# show qos shared-queue "policer-output-queues" detail
===============================================================================
QoS Shared Queue Policy
===============================================================================
-------------------------------------------------------------------------------
Shared Queue Policy (policer-output-queues)
-------------------------------------------------------------------------------
Policy         : policer-output-queues
Description    : Default Policer Output Shared Queue Policy

-------------------------------------------------------------------------------
Queue CIR       PIR       CBS     MBS     HiPrio  Multipoint Pool-Name
-------------------------------------------------------------------------------
1     0         100       1       50      10      FALSE
2     25        100       3       50      10      FALSE
3     25        100       10      50      10      FALSE
4     25        100       3       25      10      FALSE
5     100       100       10      50      10      FALSE
6     100       100       10      50      10      FALSE
7     10        100       3       25      10      FALSE
8     10        100       3       25      10      FALSE
9     0         100       1       50      10      TRUE
10    25        100       3       50      10      TRUE
11    25        100       10      50      10      TRUE
12    25        100       3       25      10      TRUE
13    100       100       10      50      10      TRUE
14    100       100       10      50      10      TRUE
15    10        100       3       25      10      TRUE
16    10        100       3       25      10      TRUE

-------------------------------------------------------------------------------
FC    UCastQ    MCastQ    BCastQ  UnknownQ
-------------------------------------------------------------------------------
be    1         9         9       9
l2    2         10        10      10
af    3         11        11      11
l1    4         12        12      12
h2    5         13        13      13
```

```
ef    6         14        14      14
h1    7         15        15      15
nc    8         16        16      16

-------------------------------------------------------------------------------
Associations
-------------------------------------------------------------------------------
Service : 1             SAP : 1/1/3:1
===============================================================================
A:PE-1#
```

For egress policing, policed traffic can access the exit port by a queue-group, the default being called policer-output-queues. The following shows the parameters for these queues.

```
A:PE-1# show qos queue-group "policer-output-queues" detail
===============================================================================
QoS Queue-Group Ingress
===============================================================================
===============================================================================
QoS Queue-Group Egress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : policer-output-queues
Description    : Default egress policer output queues.

-------------------------------------------------------------------------------
Q  CIR Admin PIR Admin CBS         HiPrio PIR Lvl/Wt    Parent    BurstLimit(B)
   CIR Rule  PIR Rule  MBS                CIR Lvl/Wt    Wred-Queue     Slope
   Named-Buffer Pool
-------------------------------------------------------------------------------
1  0         max       def         def    1/1           None          default
   closest   closest   def                0/1           disabled      default
   (not-assigned)
2  0         max       def         def    1/1           None          default
   closest   closest   def                0/1           disabled      default
   (not-assigned)

===============================================================================
Queue Group Ports (access)
===============================================================================
Port                Sched Pol           Acctg Pol Stats     Description
-------------------------------------------------------------------------------
1/1/3                                   0         No
1/1/4                                   0         No
-------------------------------------------------------------------------------

===============================================================================
Queue Group Ports (network)
===============================================================================
Port                Sched Pol           Acctg Pol Stats     Description
-------------------------------------------------------------------------------
No Matching Entries

===============================================================================
```

```
Queue Group Sap FC Maps
===============================================================================
Sap Policy     FC Name             Queue Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
A:PE-1#
```

The following output shows the relative thresholds in the parent policer when a priority level has 0, 1 or 2 associated children, both with and without using the fixed parameter.

```
A:PE-1# show qos policer-hierarchy sap 1/1/3:2 ingress priority-info
===============================================================================
Policer Hierarchy - Sap 1/1/3:2
===============================================================================
Ingress Policer Control Policy : cfhp-2
-------------------------------------------------------------------------------
root (Ing)
|
| slot(1)
|  Priority 8
|    Oper Thresh Unfair:4352        Oper Thresh Fair:5120
|    Association count:0
|  Priority 7
|    Oper Thresh Unfair:4352        Oper Thresh Fair:5120
|    Association count:0
|  Priority 6
|    Oper Thresh Unfair:4352        Oper Thresh Fair:5120
|    Association count:2 fixed
|  Priority 5
|    Oper Thresh Unfair:3328        Oper Thresh Fair:4096
|    Association count:1 fixed
|  Priority 4
|    Oper Thresh Unfair:2304        Oper Thresh Fair:3072
|    Association count:0 fixed
|  Priority 3
|    Oper Thresh Unfair:1280        Oper Thresh Fair:2048
|    Association count:2
|  Priority 2
|    Oper Thresh Unfair:0           Oper Thresh Fair:1024
|    Association count:1
|  Priority 1
|    Oper Thresh Unfair:0           Oper Thresh Fair:0
|    Association count:0


===============================================================================
A:PE-1#
```

Where

- Priority Level 1 has no children so both its fair and unfair thresholds are 0.

- Priority Level 2 has one child so its unfair threshold is 0 and its fair threshold is at the configured mbs-contribution [1024 bytes] (given that this is larger than the min-thresh-separation).
- Priority Level 3 has two children so its unfair threshold is equal to the min-thresh-separation plus the fair threshold of priority 2 [256+1024=1280 bytes]. Its fair threshold is effectively the mbs-contribution plus the fair threshold of priority 2 [1024+1024=2048 bytes] (given that the mbs-contribution is larger than 2x min-thresh-separation).
- Priorities 4, 5 and 6 have the fixed parameter configured. Even though priority 4 has no children, priority 5 has only one child and priority 6 has two children, all three priorities have the same incremental values for their unfair and fair discard threshold. This result in
  - Priority 4's unfair threshold being equal to the min-thresh-separation plus the fair threshold of priority 3 [256+2048=2304 bytes]. Its fair threshold is effectively the mbs-contribution plus the fair threshold of priority 3 [1024+2048=3072 bytes] (given that the mbs-contribution is larger than 2x min-thresh-separation).
  - Priority 5's unfair threshold being equal to the min-thresh-separation plus the fair threshold of priority 4 [256+3072=3328 bytes]. Its fair threshold is effectively the mbs-contribution plus the fair threshold of priority 4 [1024+3072=4096 bytes] (given that the mbs-contribution is larger than 2x min-thresh-separation).
  - Priority 6's unfair threshold being equal to the min-thresh-separation plus the fair threshold of priority 5 [256+4096=4352 bytes]. Its fair threshold is effectively the mbs-contribution plus the fair threshold of priority 5 [1024+4096=5120 bytes] (given that the mbs-contribution is larger than 2x min-thresh-separation).

Note that the above parameter values were chosen to exactly match available hardware values to simplify the output.

Conclusion

This note has described the configuration of Class Fair Hierarchical Policing for SAPs. This hardware policing provides low latency ingress and egress prioritized traffic control with the ability to provide fairness between child policers at the same parent policer priority level.

FP and Port Queue Groups

This chapter provides information about FP and port queue groups.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to the 7950 XRS-16c/20, 7750 SR-7/12, 7750 SR-a4/8, 7750 SR-c4/12, and 7450 ESS-6/6v/7/12 platforms and assumes only FP2- and higher-based line cards are used.

Port queue groups can be configured on FP1 line cards but not all functions described in this chapter are supported on FP1 line cards. FP queue groups are not supported on FP1 line cards because FP queue groups rely on hardware policing.

The configuration was tested on release 13.0.R7. There are no other specific prerequisites for this configuration.

Overview

Queue groups provide flexible QoS control beyond that available by default for SAPs and network interfaces.

Many applications require detailed QoS control for SAPs, with aggregated QoS control across the core on network interfaces. Queue groups allow the reverse, specifically aggregated QoS control for multiple SAPs and per-network interface QoS control. This is summarized in Table 28.

Table 28 **Default QoS and Queue Group Comparison**

	Default	Queue group
SAPs	• Per-SAP ingress queues and policers • Per-SAP egress queues and policers	• Policers for a set of SAPs at ingress • Queues for a set of SAPs at egress
Network interfaces	• Per-MDA ingress network queues • Per-port egress network queues	• Policers per network interface at ingress • Policers and queues per network interface at egress

Queue groups were introduced in SR OS release 7 as a mechanism of grouping a set of queues and enhanced in subsequent releases. Release 10 added the ability to configure policers within queue groups and introduced a more flexible configuration where the queue group template could be applied multiple times with different instances. The term queue groups was retained even though they can now contain queues, policers, or both. This chapter describes this new configuration.

Queue groups can be used for other applications than those listed in Table 28; for example:

- Pseudowire (PW) QoS

  Providing QoS control for spoke SDPs in the various pipe services, different types of VPLS services, and associated IES/VPRN interfaces (see chapter Pseudowire QoS).
- Carrier Supporting Carrier (CSC) services

  Providing QoS control for CSC network interfaces in VPRN services.
- Ingress QoS control on VPRN network interfaces

  Providing for ingress QoS control of unicast traffic into a VPRN over automatically created or manually created bindings in a VPRN service.
- Network egress QoS

  Providing queue rates in kbps instead of the percentage of the port rate, and queue CBS/MBS in kbytes instead of a fractional percentage of the pool size.

When egress queue groups are used for SAPs, the groups provide a similar functionality to multi-service sites in that both can provide an overall rate for a set of SAPs. However, multi-service sites also provide per-SAP QoS control, whereas the queue groups do not.

Queue groups are not applicable to subscriber management except when egress policing is used.



Note: The queue groups described in this chapter are different from those that are configured when using a high-scale Ethernet MDA (HS-MDAv2).

Configuration

Following the description of the overall configuration of queue groups, their configuration is illustrated using examples on a SAP, a network interface, and with egress subscriber policing.

The steps required to configure a queue group are summarized as follows:

- Create a queue group template:
  - Ingress or egress
- Apply an instance of the queue group template to:
  - FP ingress
    - Access or network
  - Ethernet egress port
    - Access or network
- Redirect traffic to a policer or queue in the FP or port queue group instance per forwarding class (FC) within a:
  - SAP ingress or egress QoS policy
  - Network QoS policy for both ingress and egress
- Consider post-egress policer queuing

Creating Queue Group Templates

Queue group templates are configured separately for ingress and egress; an ingress queue group template with the same name as an egress queue group template is a different object.

Ingress queue group templates can contain policers or queues, but not both.

Egress queue group templates can contain policers or queues, or both; queue 1 is created by default and cannot be deleted. A forwarding class (FC) can also be configured to determine the mapping of the FC of the forwarded traffic to a policer or queue.

To summarize the use of policers with respect to queue groups:

- Ingress FP queue groups can only use policers.
- Port egress access queue groups cannot use policers.
- Network ingress and egress applications can only use policers in a queue group.

The configuration of an ingress and egress queue group template is as follows:

```
configure
    qos
        queue-group-templates
            ingress
                queue-group <queue-group-name> [create]
                    description <description-string>
                    policer <policer-id> [create]
                    queue <queue-id> [multipoint] [<queue-type>]
                                     [<queue-mode>] [create]
            egress
                queue-group <queue-group-name> [create]
                    description <description-string>
                    fc <fc-name> [create]
                        queue <queue-id>
                    policer <policer-id> [create]
                    queue <queue-id> [<queue-type>] [create]
```

The configuration of the policer and queue is similar to the configuration in a SAP ingress and egress QoS policy. However, the SAP ingress QoS policy allows the configuration of a percent-rate and the SAP egress QoS policy allows the configuration of an avg-frame-overhead, neither of which is available in a queue group template. An ingress queue group template supports the creation of up to 32 queues and 32 policers, while an egress queue group template supports the creation of up to 8 queues and 8 policers.

Because an egress queue group template uses kbps for the CIR/PIR rates and bytes/kbytes for the CBS/MBS, unlike a network-queue QoS policy, a network egress interface can use them when an egress queue group template is applied.

The system instantiates the following queue group templates by default. Table 30 shows the ingress queue group templates:

Table 29 **Queue Group Templates - Ingress**

Group name	Description
_tmnx_nat_ing_q_grp	NAT/LNS Ingress Queue Group Template
_tmnx_nat_ing_q_grp_v2	NAT/LNS Ingress Queue Group Template for ISAv2
_tmnx_lns_esm_ing_q_grp	LNS ESM Ingress Queue Group Template

Table 30 shows the egress queue group templates:

Table 30 **Queue Group Templates - Egress**

Group name	Description
_tmnx_nat_egr_q_grp	NAT/LNS Egress Queue Group Template
policer-output-queues	Default egress policer output queues
_tmnx_nat_egr_q_grp_v2	NAT/LNS Egress Queue Group Template for ISAv2
_tmnx_lns_esm_egr_q_grp	LNS ESM Egress Queue Group Template

This chapter will only discuss the policer-output-queues queue group from the preceding table, an instance of which is created on all egress access and hybrid ports, to be used for egress policed traffic.

Applying Queue Group Templates

Ingress

An ingress queue group template containing only policers can be applied to an FP ingress. When applied, an instance identifier must be specified, which represents the instantiated instance of the related queue group template.

If an attempt is made to configure a queue group template containing queues for an FP ingress, the following error message appears:

```
*A:PE-1>config#  card 5 fp 1 ingress access queue-group "qg2" instance 1 create
MINOR: CHMGR #1164 Cannot attach a Queue Group containing queues
```

Because 7750 SR-a4/8 platforms do not support SAP or network hardware policers, FP ingress queue groups are not supported on those platforms.

Queue group templates containing only queues can be applied to port ingress. However, only one such queue group can be applied per ingress port.

An ingress template can be applied to an FP ingress as either an ingress access queue group or an ingress network queue group, or both. In each case:

- An accounting policy can be applied.
- Statistics collection can be enabled.
- A description can be configured.
- A policer control policy can be applied, containing parent arbiters for the queue group policers, and its parameters can be overridden.
- Policer parameters configured within the queue group template can also be overridden.



Note: When the queue group applies to a tunnel object that can move between different network interfaces, and consequently different network ingress FPs (for example, PW QoS), a network ingress queue group instance must be applied to each FP ingress that could be used.

The configuration of ingress queue group templates for ingress FPs is as follows:

```
configure
    card <slot-number>
        fp [<fp-number>]
            ingress
                access
                    queue-group <queue-group-name>
                                instance <[1..65535]> [create]
                        accounting-policy <acct-policy-id>
                        collect-stats
                        description <description-string>
                        policer-control-override [create]
                            max-rate {<rate> | max}
                            priority-mbs-thresholds
                                min-thresh-separation <size> [bytes | kilobytes]
                                priority <level>
                                   mbs-contribution <size> [bytes | kilobytes]
                        policer-control-policy
                                <policer-control-policy-name>
                        policer-override
                            policer <policer-id> [create]
                network
                    queue-group <queue-group-name>
                                instance <[1..65535]> [create]
                        accounting-policy <acct-policy-id>
                        collect-stats
                        description <description-string>
                        policer-control-override [create]
                            max-rate {<rate> | max}
                            priority-mbs-thresholds
```

```
            min-thresh-separation <size> [bytes | kilobytes]
            priority <level>
               mbs-contribution <size> [bytes | kilobytes]
    policer-control-policy
            <policer-control-policy-name>
    policer-override
        policer <policer-id> [create]
```

Note: Configure FP ingress access and network queue group instances consistently across FPs relating to a LAG.

A single ingress template containing only queues can also be applied to an Ethernet port ingress, which can be used only for access ingress:

- An accounting policy can be applied.
- Statistics collection can be enabled.
- A description can be configured.
- Queue parameters configured within the queue group template can be overridden.
- A scheduler policy can be applied containing parent schedulers for the queue group queues.

The configuration of an ingress queue group template for an ingress port is as follows:

```
configure
    port <port-id>
        ethernet
            access
                ingress
                    queue-group <queue-group-name> [create]
                        accounting-policy <acct-policy-id>
                        collect-stats
                        description <description-string>
                        queue-overrides
                            queue <queue-id> [create]
                        scheduler-policy <scheduler-policy-name>
```

Egress

An egress queue group template containing policers or queues, or both, can be applied to an Ethernet port network egress. Only queue group templates containing queues (not policers) can be applied to an Ethernet port access egress. When applied, an instance identifier can be specified, which represents the instantiated instance of the related queue group template; the default being 1.

If an attempt is made to configure a queue group template containing policers for an Ethernet port access egress, the following error message appears:

```
*A:PE-1# configure port 5/1/5 ethernet access egress queue-
group "qg1" instance 1 create
MINOR: CLI Could not create/change "qg1".
MINOR: PMGR #1324 Cannot attach a Queue Group containing policers
```

An egress template can be applied as either an egress access queue group or an egress network queue group, or both. In each case:

- An accounting policy can be applied.
- Statistics collection can be enabled.
- A description can be configured.
- Queue parameters configured within the queue group template can be overridden.
- A scheduler policy can be applied containing parent schedulers for the queue group queues.

When applied to the Ethernet access egress, a host-match can be configured, which is used to select a queue group for subscriber egress policed traffic; see later in this chapter for details.

When applied to the Ethernet network egress, a policer control policy can be applied, which contains parent arbiters for the queue group policers.



Note: When the queue group applies to a tunnel object that can move between different network interfaces, and consequently different network egress ports (for example, PW QoS), a network egress queue group instance must be applied to each network egress port that could be used.

The configuration of egress queue group templates for egress ports is as follows:

```
configure
    port <port-id>
        ethernet
            access
                egress
                    queue-group <queue-group-name> [create]
                                [instance <instance-id>]
                        accounting-policy <acct-policy-id>
                        agg-rate
                            limit-unused-bandwidth
                            queue-frame-based-accounting
                            rate <kilobits-per-second>
                        collect-stats
                        description <description-string>
                        host-match dest <destination-string> [create]
```

```
            queue-overrides
                queue <queue-id> [create]
            scheduler-policy <scheduler-policy-name>
network
     queue-group <queue-group-name> [create]
                 [instance <instance-id>]
            accounting-policy <acct-policy-id>
            agg-rate
                limit-unused-bandwidth
                queue-frame-based-accounting
                rate <kilobits-per-second>
            collect-stats
            description <description-string>
            policer-control-policy
                    <policer-control-policy-name>
            queue-overrides
                queue <queue-id> [create]
            scheduler-policy <scheduler-policy-name>
```

Note: When port egress queue groups are used with a LAG, the system enforces a consistent configuration between the ports (based on only the configuration of the primary LAG port).

Redirecting Traffic to a Queue Group Queue or Policer

There are multiple ways to redirect traffic to a queue group queue or policer. The redirection is always on a per-FC basis so that different FCs within the same policy can use a mix of a local SAP policer, a local SAP or network queue, or a queue group policer or queue (where applicable).

Redirection to a queue group queue or policer is not supported for subscribers; attempting to do so will display the following errors:

```
*A:PE-1>config>qos>sap-ingress# fc "af" policer 1 fp-redirect-group
MINOR: QOS #1489 Cannot assign a queue-group because SLA profile references to this
policy exist

*A:PE-1>config>qos>sap-egress# fc af policer 1 port-redirect-group-queue
MINOR: QOS #1628 Cannot assign a queue-group because SLA profile references to this
policy exist

*A:PE-1>config>qos>sap-egress# fc af queue 1 port-redirect-group-queue
MINOR: QOS #1628 Cannot assign a queue-group because SLA profile references to this
policy exist

*A:PE-1>config>subscr-mgmt>sla-prof>ingress# qos 40
MINOR: SUBMGR #1110 Cannot assign Qos Policy, QOS Plcy contains queue-
group references

*A:PE-1>config>subscr-mgmt>sla-prof>egress# qos 40
MINOR: SUBMGR #1110 Cannot assign Qos Policy, QOS Plcy contains queue-
```

```
group references
```

SAP Ingress Redirection

Traffic can be redirected per FC to a policer in an FP ingress access queue group, using the fp-redirect-group parameter in the SAP ingress QoS policy, as follows:

```
configure
    qos
        sap-ingress <policy-id> [create]
            fc <fc-name> [create]
                policer <policer-id> fp-redirect-group
```

The policer-id is referencing a policer in the FP ingress queue group.

The queue group name and instance to be used is not yet configured, allowing this QoS policy to be applied to objects using different FP ingress queue group instances, which results in greater flexibility.

Compare this preferred configuration to the original configuration used for ingress port queue groups, as follows, where the queue group name is specified within the SAP ingress QoS policy and the instance is not available:

```
configure
    qos
        sap-ingress <policy-id> [create]
            fc <fc-name> [create]
                queue <queue-id> group <queue-group-name>
```

In the preferred configuration, redirection is completed by applying the SAP ingress QoS policy to the SAP with the queue group name and instance to be used, as follows:

```
configure
    service
        {apipe|cpipe|epipe|fpipe|ipipe} <service-id>
            sap <sap-id>
                ingress
                    qos <policy-id> fp-redirect-group <queue-group-name>
                                                instance <instance-id>
        {ies|vprn} <service-id>
            interface <ip-int-name>
                sap <sap-id>
                ingress
                    qos <policy-id>
                        fp-redirect-group <queue-group-name>
                                  instance <instance-id>
```

SAP Egress Redirection

Traffic can be redirected per FC to a queue in a port egress access queue group, using the port-redirect-group parameter, as follows:

```
configure
    qos
        sap-egress <policy-id> [create]
            fc <fc-name> [create]
                queue <queue-id> port-redirect-group-queue
```

The queue-id is referencing a queue in the port egress queue group.

The queue group name and instance to be used is not yet configured, allowing this QoS policy to be applied to objects using different port egress queue groups.

Compare this preferred configuration to the original configuration, which allows the queue group name with an instance to be specified:

```
queue <id> {group <grp-name> [instance instance-id]}
```

In the preferred configuration, redirection is completed by applying the SAP egress QoS policy to the SAP with the queue group name and instance to be used, as follows:

```
configure
    service
        {apipe|cpipe|epipe|fpipe|ipipe} <service-id>
            sap <sap-id>
                egress
                    qos <policy-id>
                        port-redirect-group <queue-group-name>
                                        instance <instance-id>
        {ies|vprn} <service-id>
            interface <ip-int-name>
                sap <sap-id>
                    egress
                        qos <policy-id>
                            port-redirect-group <queue-group-name>
                                            instance <instance-id>
```

Network Ingress Redirection

Traffic can be redirected per FC to a policer in an FP ingress network queue group, using the fp-redirect-group parameter, as follows:

```
configure
    qos
        network <network-policy-id> [create]
```

```
            ingress
                fc <fc-name>
                    fp-redirect-group <policer-type> <policer-id>
```

The policer-id is referencing a policer in the FP ingress queue group.

The traffic usage for each policer type for the supported services is shown in Table 4.

Table 31 **Network Ingress FP Queue Group Policer Usage**

Policer Type	Usage
broadcast-policer	Broadcast traffic for PW QoS in a VPLS service
mcast-policer	Multipoint traffic (except for ingress QoS control on VPRN network interfaces and CSC services where the IP multicast traffic uses the ingress network queues or queue group related to the network interface) Multicast traffic for PW QoS in a VPLS service
unknown-policer	Unknown traffic for PW QoS in a VPLS service
policer	Unicast traffic

The queue group name and instance to be used is not yet configured, allowing this QoS policy to be applied to objects using different FP ingress queue group instances.

The redirection is completed by applying the network QoS policy to the required object with the queue group name and instance to be used, as follows (only the network interface redirection is shown):

```
configure
    router
        interface <interface-name>
            qos <network-policy-id>
                        ingress-fp-redirect-group <queue-group-name>
                        ingress-instance <instance-id>
```

Network Egress Redirection

Traffic can be redirected per FC to a policer or a queue in a port egress network queue group, using the port-redirect-group parameter, as follows:

```
configure
    qos
        network <network-policy-id> [create]
            egress
                fc <fc-name>
```

```
                    port-redirect-group {queue <queue-id> |
                                 policer <plcr-id> [queue <queue-id>]}
```

The queue-id and policer-id are referencing a queue and a policer in the port egress queue group.

The queue group name and instance to be used are not yet configured, allowing this QoS policy to be applied to objects using different port egress queue groups.

The redirection is completed by applying the network QoS policy to the required object with the queue group name and instance to be used, as follows (only the network interface redirection is shown):

```
configure
    router
        interface <interface-name>
            qos <network-policy-id>
                        egress-fp-redirect-group <queue-group-name>
                        egress-instance <instance-id>
```

Note: Non-IPv4/non-IPv6/non-MPLS packets are not subject to the redirection to an egress queue group instance and will remain on the regular port network egress queues. When using an egress port scheduler, parent the related regular network port queues to appropriate port scheduler priority levels, to ensure the required operation under port congestion. This is important for protocol traffic, such as LACP, EFMOAM, ETH-CFM, ARP, and IS-IS, which by default use the FC nc regular network port queue.

Post-egress Policer Queuing

The queuing of traffic exiting an egress policer is described as follows for SAP egress and network egress.

SAP egress policed traffic exits a port, using either a local queue or an egress queue group instance queue. If an egress queue group instance queue is to be used, the policed traffic can be mapped to the queue in the following ways, listed in reverse preference order (first is lowest preference):

1. By default, SAP egress policed traffic exits using a queue in the policer-output-queues queue group instance when only a policer is configured:

```
configure
    qos
        sap-egress <policy-id> [create]
            fc <fc-name> [create]
                policer <policer-id>
```

The system always creates this egress queue group template and applies it to all access and hybrid ports (for use by the access part of the hybrid port) as instance 1. The queue group template contains the mapping for all FCs to one of the two created queues to determine which queue is to be used.

The default configuration of the policer-output-queues group template is as follows (without the queue details):

```
configure
    qos
        queue-group-templates
            egress
                queue-group "policer-output-queues" create
                    description "Default egress policer
                                 output queues."
                    queue 1 best-effort create
                    exit
                    queue 2 expedite create
                    exit
                    fc af create
                        queue 1
                    exit
                    fc be create
                        queue 1
                        exit
                    fc ef create
                        queue 2
                    exit
                    fc h1 create
                        queue 2
                    exit
                    fc h2 create
                        queue 2
                    exit
                    fc l1 create
                        queue 1
                    exit
                    fc l2 create
                        queue 1
                    exit
                    fc nc create
                        queue 2
                    exit
                exit
```

This queue group template cannot be deleted, but its configuration can be modified. The configuration commands under a port access egress queue group instance are also available under the policer-output-queues queue group instance.

2. The FC in the SAP egress QoS policy can be mapped to a policer with its traffic redirected to a port access egress queue group, as follows:

```
configure
    qos
        sap-egress <policy-id> [create]
            fc <fc-name> [create]
```

```
            policer <policer-id> port-redirect-group-queue
```

The QoS policy must be applied to the SAP with a redirection to a queue group instance. By default, queue 1 in the queue group will be used, but other queues can be created and used by creating an FC-to-queue mapping to them.

3. Similar to step 2, the FC in the SAP egress QoS policy can be mapped to a policer with its traffic redirected to a port access egress queue group instance and the queue within that queue group instance to be used, as follows:

```
configure
    qos
        sap-egress <policy-id> [create]
            fc <fc-name> [create]
                policer <policer-id> port-redirect-group-queue queue <id>
```

The QoS policy must again be applied to the SAP with a redirection to a queue group instance. In this case, the queue specified with the port-redirect-group-queue will be used.

Network egress can only use policers that are created within an egress queue group instance, and this requires the egress FC to have the port-redirect-group parameter configured.

With the following configuration, traffic exits the port on the network egress queue to which its FC is mapped (not on a queue group queue):

```
configure
    qos
        network <network-policy-id> [create]
            egress
                fc <fc-name>
                    port-redirect-group policer <plcr-id>
```

If a queue is specified on the port-redirect-group statement, as follows, the traffic exits on the referenced queue within the port network egress queue group instance:

```
configure
    qos
        network <network-policy-id> [create]
            egress
                fc <fc-name>
                    port-redirect-group policer <plcr-id> queue <queue-id>
```

Configuration Examples

The following configuration examples are for the use of queue groups with a SAP, network interface, and egress policed traffic subscriber traffic. Different queue group templates and instances have been used to highlight the flexibility of the configuration.

SAP Configuration Example

In this example, a SAP is created in an IES service using ingress queue group qg1 instance 1 and egress queue group qg1 instance 1.

First, the queue group templates are configured, as follows:

```
configure
    qos
        queue-group-templates
            ingress
                queue-group "qg1" create
                    policer 1 create
                    exit
                    policer 2 create
                    exit
                exit
            exit
            egress
                queue-group "qg1" create
                    queue 1 best-effort create
                    exit
                    queue 2 expedite create
                    exit
                exit
                queue-group "policer-output-queues" create
                    queue 3 best-effort create
                    exit
                    fc l2 create
                        queue 3
                    exit
                exit
```

The policer-output-queues queue group template has been modified to add an extra queue and map FC l2 to it. This is used by the traffic mapped to policer 1 in the SAP egress QoS policy.

The ingress template is applied to card 5 fp 1 to create an FP access queue group instance, as follows:

```
configure
    card 5
        fp 1
```

```
            ingress
                access
                    queue-group "qg1" instance 1 create
                    exit
                exit
```

The egress template is applied to port 5/1/5 to create a port access queue group instance, as follows:

```
configure
    port 5/1/5
        ethernet
            mode hybrid
            encap-type dot1q
            access
                egress
                    queue-group "qg1" instance 1 create
                    exit
                exit
```

The SAP ingress QoS policy is created to redirect FC af to policer 1 and FC ef to policer 2 in the FP ingress queue group. To emphasize that the redirection is per FC and can coexist with locally mapped queues or policers, FC be is mapped to the local queue 1 and FC l2 to local queue 2, as follows:

```
configure
    qos
        sap-ingress 10 create
            queue 1 create
            exit
            queue 2 create
            exit
            queue 11 multipoint create
            exit
            fc "af" create
                policer 1 fp-redirect-group
            exit
            fc "be" create
                queue 1
            exit
            fc "ef" create
                policer 2 fp-redirect-group
            exit
            fc "l2" create
                queue 2
            exit
            dot1p 0 fc "be"
            dot1p 1 fc "l2"
            dot1p 2 fc "af"
            dot1p 3 fc "ef"
        exit
```

The SAP egress QoS policy is created to redirect FC af to queue 1 in the port egress queue group and FC ef to local policer 2, with its traffic exiting through queue 2 in the port egress queue group. FC be is mapped to the local queue 1 and FC l2 to local policer 1. The policer 1 traffic will exit using queue 3 of the policer-output-queues queue group, because the mapping for FC l2 has been modified in its template, as follows:

```
configure
    qos
        sap-egress 10 create
            queue 1 create
            exit
            policer 1 create
            exit
            policer 2 create
            exit
            fc af create
                queue 1 port-redirect-group-queue
            exit
            fc be create
                queue 1
            exit
            fc ef create
                policer 2 port-redirect-group-queue queue 2
            exit
            fc l2 create
                policer 1
            exit
        exit
```

The IES service is created with an interface on SAP 5/1/5:1, using the preceding ingress and egress QoS policies and queue group instances. A second interface (not shown) is used to forward the traffic to and from this interface, as follows:

```
configure
    service
        ies 1 customer 1 create
           interface "PE-1-int1-2" create
                address 10.1.2.1/24
                sap 5/1/5:1 create
                    ingress
                        qos 10 fp-redirect-group "qg1" instance 1
                    exit
                    egress
                        qos 10 port-redirect-group "qg1" instance 1
                    exit
                exit
            exit
            no shutdown
        exit
```

The following output shows the configuration and the QoS state after sending traffic through the service.

The traffic statistics are either counted in the SAP queues and policers, or in the queue group instance queues and policers, but not in both. However, summary statistics per SAP are available when using FP ingress queue groups, as shown.

The queue group templates can be shown. The output shows the details related to the ingress queue group template:

```
*A:PE-1# show qos queue-group "qg1" ingress detail
===============================================================================
QoS Queue-Group Ingress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : qg1
Description    : (Not Specified)
-------------------------------------------------------------------------------
Q  Mode    CIR Admin  PIR Admin  CBS          HiPrio  PIR Lvl/Wt Parent
           CIR Rule   PIR Rule   MBS                  CIR Lvl/Wt BurstLimit(B)
           Named-Buffer Pool     Pkt Bt Ofst          Adv Config Policy Name
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group Ports
===============================================================================
Port                Sched Pol           Acctg Pol Stats     Description
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group Sap FC Maps
===============================================================================
Sap Policy     FC Name             Queue (id type)
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group FP Maps
===============================================================================
Card Num       Fp Num              Instance            Type
-------------------------------------------------------------------------------
5              1                   1                   Access
-------------------------------------------------------------------------------
Entries found: 1
-------------------------------------------------------------------------------
===============================================================================
Queue Group Policer
===============================================================================
Policer Id     : 1
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : max                           Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : 0
Profile Capped : Disabled
StatMode       : minimal
===============================================================================
```

```
Policer Id     : 2
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : max                           Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : 0
Profile Capped : Disabled
StatMode       : minimal
===============================================================================
*A:PE-1#
```

The following output shows the details related to the egress queue group template:

```
*A:PE-1# show qos queue-group "qg1" egress detail
===============================================================================
QoS Queue-Group Egress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : qg1
Description    : (Not Specified)
-------------------------------------------------------------------------------
Q CIR Admin  PIR Admin  CBS          HiPrio PIR Lvl/Wt Parent     BurstLimit(B)
  CIR Rule   PIR Rule   MBS                 CIR Lvl/Wt Wred-Queue Slope
  Named-Buffer Pool     Pkt Bt Ofst         Adv Config Policy Name
-------------------------------------------------------------------------------
1 0          max        def          def    1/1        None       default
  closest    closest    def                 0/1        disabled   default
  (not-assigned)        add 0               (not-assigned)
2 0          max        def          def    1/1        None       default
  closest    closest    def                 0/1        disabled   default
  (not-assigned)        add 0               (not-assigned)
===============================================================================
Queue Group FC Mapping
===============================================================================
FC Name                                Queue-Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
===============================================================================
Queue Group Ports (access)
===============================================================================
Port      Sched Pol          Acctg Pol Stats Description          QGrp-Instance
-------------------------------------------------------------------------------
5/1/5                        0         No                         1
-------------------------------------------------------------------------------
===============================================================================
Queue Group Ports (network)
===============================================================================
Port   Sched Pol    Policer-Ctrl-Pol  Acctg Pol Stats Description QGrp-Instance
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Qos Sap-Egress FC Group-Queue References
===============================================================================
```

```
Sap Policy     FC Name             Queue Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Qos Sap-Egress FC Port-Redirect-Group-Queue References
===============================================================================
Sap Policy     FC Name             Queue Id
-------------------------------------------------------------------------------
10             af                  1
10             ef                  2
-------------------------------------------------------------------------------
Entries found: 2
-------------------------------------------------------------------------------
===============================================================================
Queue Group Policer
===============================================================================
No Matching Entries
-------------------------------------------------------------------------------
HSMDA PIR Admin  Packet  WRR     MBS       Slope Plcy        WRR Plcy
Queue PIR Rule   Offset  Weight            Max Class         Burst Lmt
-------------------------------------------------------------------------------
1     max        add 0   1       default   default           n/a
      closest                              8                 default
2     max        add 0   1       default   default           n/a
      closest                              8                 default
3     max        add 0   1       default   default           n/a
      closest                              8                 default
4     max        add 0   1       default   default           n/a
      closest                              8                 default
5     max        add 0   1       default   default           n/a
      closest                              8                 default
6     max        add 0   1       default   default           n/a
      closest                              8                 default
7     max        add 0   1       default   default           n/a
      closest                              8                 default
8     max        add 0   1       default   default           n/a
      closest                              8                 default
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The preceding output shows that the ingress template has an instance 1, which is applied to card 5 fp 1, and the egress template has an instance 1, which is applied to port 5/1/5.

The remapping of FC l2 to queue 3 in the policer-output-queues queue group template is as follows:

```
*A:PE-1# show qos queue-group "policer-output-queues" egress detail | match post-
lines 12 "Queue Group FC Mapping"
Queue Group FC Mapping
===============================================================================
FC Name                                 Queue-Id
-------------------------------------------------------------------------------
af                                      1
be                                      1
ef                                      2
```

```
h1                                  2
h2                                  2
l1                                  1
l2                                  3
nc                                  2
===============================================================================
*A:PE-1#
```

For SAP ingress QoS policy 10, the redirection is true for FC af using policer 1 and FC ef using policer 2 to an FP ingress queue group, as follows:

```
*A:PE-1# show qos sap-ingress 10 detail | match post-
lines 5 expression "af Unicast|ef Unicast"
FC                : af Unicast
-------------------------------------------------------------------------------
Policer           : 1                 Queue             : def
Queue-Group       : None
FP Queue-Group    : True
FC                : ef Unicast
-------------------------------------------------------------------------------
Policer           : 2                 Queue             : def
Queue-Group       : None
FP Queue-Group    : True
```

In the following output for SAP egress QoS policy 10:

- The redirection is true for FC af using queue 1 in the queue group instance, to be specified when the policy is applied to a SAP.
- The redirection is true for FC ef using local policer 2 with its traffic exiting queue 2 in the queue group instance, to be specified when the policy is applied to a SAP.
- FC l2 is using local policer 1 with its traffic exiting using queue 3 in the policer-output-queues port access egress queue group.

```
*A:PE-1# show qos sap-egress 10 detail | match post-lines 6 expression "Queue-Group"
FC  Queue  Queue-Group                      InstanceId   SapBReDir  Plcr
-------------------------------------------------------------------------------
be  1      n/a                              n/a          False      None
l2  3      policer-output-queues            1            False      1
af  1      n/a                              n/a          True       None
ef  2      n/a                              n/a          True       2
```

The QoS information, including the queue group instances being used for the IES service SAP, is shown as follows:

```
*A:PE-1# show service id 1 sap 5/1/5:1 detail | match post-lines 4 QOS
QOS
-------------------------------------------------------------------------------
Ingress qos-policy : 10                     Egress qos-policy : 10
Ingress FP QGrp    : qg1                    Egress Port QGrp  : qg1
Ing FP QGrp Inst   : 1                      Egr Port QGrp Inst: 1
```

In the following output for the egress SAP, FC ef is using policer 2 and its traffic exits using queue 2 in queue group qg1 instance 1, while FC l2 is using policer 1 and its traffic exits using queue 3 of the policer-output-queues queue group instance 1:

```
*A:PE-1# show qos policer sap 5/1/5:1 egress detail | match post-
lines 4 "Policer Info"
Policer Info (1->5/1/5:1->1), Slot 5
===============================================================================
Policer Name       : 1->5/1/5:1->1
Direction          : Egress             Fwding Plane       : 1
FC->[QGrp:Inst->]Q : l2->policer-output-queues:1->3
Policer Info (1->5/1/5:1->2), Slot 5
===============================================================================
Policer Name       : 1->5/1/5:1->2
Direction          : Egress             Fwding Plane       : 1
FC->[QGrp:Inst->]Q : ef->qg1:1->2
```

The associations of the port access egress queue group are shown as follows:

```
*A:PE-1# show port 5/1/5 queue-group "qg1" instance 1 egress access associations
===============================================================================
Ethernet port 5/1/5 Access Egress queue-group
===============================================================================
Queue-Group Name   : qg1
Instance-Id        : 1
-------------------------------------------------------------------------------
Subscriber-Host Queue-Group Associations
-------------------------------------------------------------------------------
No associations
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
SAP Based Queue-Group Association
-------------------------------------------------------------------------------
Service-Id               SAP                      Qos Policy-Id
-------------------------------------------------------------------------------
1 (IES)                  5/1/5:1                  10
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Qos Policy Based Queue-Group Association
-------------------------------------------------------------------------------
FC Name   Qos Policy-Id       Local Policer-Id    Local Queue-Id      QGrp-Q
-------------------------------------------------------------------------------
No associations
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
*A:PE-1#
```

After traffic is sent through the service, the FP ingress access queue group policer statistics are as follows:

```
*A:PE-1# show card 5 fp 1 ingress queue-
group "qg1" instance 1 mode access statistics
===============================================================================
Card:5  Acc.QGrp: qg1  Instance: 1
===============================================================================
```

```
Group Name    : qg1
Description   : (Not Specified)
Pol Ctl Pol   : None                    Acct Pol      : None
Collect Stats : disabled
-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
                        Packets                 Octets
Ing. Policer:  1  Grp: qg1 (Stats mode: minimal)
Off. All              : 1000                    128000
Dro. All              : 0                       0
For. All              : 1000                    128000
Ing. Policer:  2  Grp: qg1 (Stats mode: minimal)
Off. All              : 1000                    128000
Dro. All              : 0                       0
For. All              : 1000                    128000
===============================================================================
*A:PE-1#
```

The traffic sent through the port egress access queue group queues is as follows:

```
*A:PE-1# show port 5/1/5 queue-group "qg1" instance 1 egress access statistics
-------------------------------------------------------------------------------
Ethernet port 5/1/5 Access Egress queue-group
-------------------------------------------------------------------------------
                        Packets                 Octets
Egress Queue: 1 Group: qg1 Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 1000                    128000
Out Profile dropped   : 0                       0
Egress Queue: 2 Group: qg1 Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 1000                    128000
Out Profile dropped   : 0                       0
-------------------------------------------------------------------------------
*A:PE-1#
```

Finally, the FC l2 traffic using the egress policer 1 is in queue 3, in the policer-output-queues queue group instance statistics, as follows:

```
*A:PE-1# show port 5/1/5 queue-group "policer-output-
queues" instance 1 egress access statistics
-------------------------------------------------------------------------------
Ethernet port 5/1/5 Access Egress queue-group
-------------------------------------------------------------------------------
                        Packets                 Octets
Egress Queue: 1 Group: policer-output-queues Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 0                       0
Out Profile dropped   : 0                       0
Egress Queue: 2 Group: policer-output-queues Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 0                       0
```

```
Out Profile dropped   : 0                       0
Egress Queue: 3 Group: policer-output-queues Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 1000                    128000
Out Profile dropped   : 0                       0
-------------------------------------------------------------------------------
*A:PE-1#
```

The number of valid ingress packets received on a SAP, or subscribers on that SAP, in the sap-stats output, are shown as follows. The received valid counter includes both the local SAP counters and the counters from the related FP ingress queue group instance. This is useful to display SAP-level traffic statistics when forwarding classes in a SAP ingress policy have been redirected to an ingress queue group.

```
*A:PE-1# show service id 1 sap 5/1/5:1 sap-stats | match post-
lines 6 "Forwarding Engine Stats"
Forwarding Engine Stats
Dropped               : 0                       0
Received Valid        : 4000                    512000
Off. HiPrio           : 0                       0
Off. LowPrio          : 2000                    256000
Off. Uncolor          : 0                       0
Off. Managed          : 0                       0
```

Traffic forwarded through FP ingress access, port ingress access, and port egress access queue groups can be monitored, as follows:

```
monitor card <slot-number> fp <fp-number>  ingress {access|network} queue-group
<queue-group-name> instance <instance-id> [interval <seconds>][repeat<repeat>]
policer <policer-id> [absolute | percent-rate [<reference-rate>]]

monitor port queue-group <queue-group-name> ingress <access> ingress-queue <ingress-
queue-id> [interval <seconds>] [repeat <repeat>] [absolute|rate]

monitor port queue-group <queue-group-name> egress <access> [instance <instance-
id>] [egress-queue <egress-queue-id>] [interval <seconds>] [repeat <repeat>]
[absolute|rate]
```

The summary of the queue groups applied to a port is shown as follows:

```
*A:PE-1# show port 5/1/5 queue-group summary
===============================================================================
Port queue-group summary
===============================================================================
Access-egress queue groups:
----------------------------
qg1
policer-output-queues
Total number of access-egress queue groups  : 2
Network-egress queue groups:
----------------------------
Total number of network-egress queue groups : 0
Access-ingress queue groups:
----------------------------
```

```
Total number of access-ingress queue groups : 0
===============================================================================
*A:PE-1#
```

The total usage of queue groups is shown as follows:

```
*A:PE-1# show qos queue-group summary
===============================================================================
Queue-group instances per card
===============================================================================
card      port-acc-ing  port-acc-egr  port-nw-egr   fp-acc-ing    fp-nw-ing
-------------------------------------------------------------------------------
1         0             0             0             0             0
2         0             0             0             0             0
3         0             0             0             0             0
4         0             0             0             0             0
5         0             3             0             1             0
-------------------------------------------------------------------------------
Total ingress QG templates per system :  4
Total egress  QG templates per system :  5
===============================================================================
*A:PE-1#
```

The preceding output includes the created ingress template plus the three system-created ingress templates (making four in total), and the created egress template plus the four system-created egress templates (making five in total). There is one applied FP access ingress queue group instance on card 5. There are three port access egress queue group instances (the applied queue group instance and two instances of the policer-output-queues queue group), one on each access port used for IES service interfaces.

Network Interface Configuration Example

A network interface is created using ingress queue group qg2 instance 2 and egress queue group qg2 instance 2. If the goal is to provide per-network interface QoS on a single port, each network interface would be configured on a separate VLAN.

First, the queue group templates are configured, as follows:

```
configure
    qos
        queue-group-templates
            ingress
                queue-group "qg2" create
                    policer 1 create
                    exit
                    policer 2 create
                    exit
                exit
            exit
            egress
```

```
            queue-group "qg2" create
                queue 1 best-effort create
                exit
                queue 2 expedite create
                exit
                policer 1 create
                exit
                policer 2 create
                exit
            exit
        exit
```

The ingress template is applied to card 5 fp 1 to create an FP network queue group instance, as follows:

```
configure
    card 5
        fp 1
            ingress
                network
                    queue-group "qg2" instance 2 create
                    exit
                exit
```

The egress template is applied to port 5/1/5 to create a port network queue group instance, as follows:

```
configure
    port 5/1/5
        ethernet
            mode hybrid
            encap-type dot1q
            network
                egress
                    queue-group "qg2" instance 2 create
                    exit
                exit
```

The network QoS policy is created to:

- Ingress

  Redirect FC af to policer 1 and FC ef to policer 2 in the FP ingress queue group. FC be and FC l2 continue to use their default mapping to the local network ingress queues 1 and 2, respectively.

- Egress

  Redirect FC af to queue 1 in the port egress queue group and FC ef to policer 2 in the port egress queue group, with its traffic exiting through queue 2 of the port egress queue group. FC l2 is redirected to policer 1 in the port egress queue group, with its traffic exiting through the default network egress queue mapped by FC l2; that is, queue 2. FC be continues to use the default network egress queue 1.

```
configure
    qos
        network 10 create
            ingress
                dot1p 0 fc be profile out
                dot1p 1 fc l2 profile out
                dot1p 2 fc af profile out
                dot1p 3 fc ef profile in
                fc af
                    fp-redirect-group policer 1
                exit
                fc ef
                    fp-redirect-group policer 2
                exit
            exit
            egress
                fc af
                    port-redirect-group queue 1
                exit
                fc ef
                    port-redirect-group policer 2 queue 2
                exit
                fc l2
                    port-redirect-group policer 1
                exit
            exit
        exit
```

The network interface is created on port 5/1/5:2 using the preceding network QoS policy with the ingress and egress being redirected to created queue group instances. A second interface (not shown) is used to forward the traffic to and from this network interface.

```
configure
    router Base
        interface "PE-1-int2-2"
            address 10.2.2.1/24
            port 5/1/5:2
            qos 10 egress-port-redirect-group "qg2" egress-instance 2
                   ingress-fp-redirect-group "qg2" ingress-instance 2
            no shutdown
        exit
```

The following output shows the details of the configuration and the QoS state after sending traffic through the network interface.

The traffic statistics are either counted in the network interface queues or in the queue group instance queues and policers, but not in both.

The queue group templates can be shown. The output shows the details related to the ingress queue group template:

```
*A:PE-1# show qos queue-group "qg2" ingress detail
===============================================================================
QoS Queue-Group Ingress
```

```
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : qg2
Description    : (Not Specified)
-------------------------------------------------------------------------------
Q  Mode    CIR Admin  PIR Admin  CBS          HiPrio  PIR Lvl/Wt Parent
           CIR Rule   PIR Rule   MBS                  CIR Lvl/Wt BurstLimit(B)
           Named-Buffer Pool     Pkt Bt Ofst          Adv Config Policy Name
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group Ports
===============================================================================
Port                Sched Pol           Acctg Pol Stats     Description
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group Sap FC Maps
===============================================================================
Sap Policy     FC Name             Queue (id type)
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group FP Maps
===============================================================================
Card Num       Fp Num              Instance            Type
-------------------------------------------------------------------------------
5              1                   2                   Network
-------------------------------------------------------------------------------
Entries found: 1
-------------------------------------------------------------------------------
===============================================================================
Queue Group Policer
===============================================================================
Policer Id     : 1
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : max                           Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : 0
Profile Capped : Disabled
StatMode       : minimal
===============================================================================
Policer Id     : 2
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : max                           Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : 0
Profile Capped : Disabled
StatMode       : minimal
===============================================================================
*A:PE-1#
```

The following output shows the details related to the egress queue group template:

```
*A:PE-1# show qos queue-group "qg2" egress detail
===============================================================================
QoS Queue-Group Egress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : qg2
Description    : (Not Specified)
-------------------------------------------------------------------------------
Q CIR Admin  PIR Admin  CBS          HiPrio PIR Lvl/Wt Parent     BurstLimit(B)
  CIR Rule   PIR Rule   MBS                 CIR Lvl/Wt Wred-Queue Slope
  Named-Buffer Pool     Pkt Bt Ofst         Adv Config Policy Name
-------------------------------------------------------------------------------
1 0          max        def          def    1/1        None       default
  closest    closest    def                 0/1        disabled   default
  (not-assigned)        add 0               (not-assigned)
2 0          max        def          def    1/1        None       default
  closest    closest    def                 0/1        disabled   default
  (not-assigned)        add 0               (not-assigned)
===============================================================================
Queue Group FC Mapping
===============================================================================
FC Name                                Queue-Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
===============================================================================
Queue Group Ports (access)
===============================================================================
Port      Sched Pol          Acctg Pol Stats Description          QGrp-Instance
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group Ports (network)
===============================================================================
Port   Sched Pol    Policer-Ctrl-Pol  Acctg Pol Stats Description QGrp-Instance
-------------------------------------------------------------------------------
5/1/5                                      No                 2
-------------------------------------------------------------------------------
===============================================================================
Qos Sap-Egress FC Group-Queue References
===============================================================================
Sap Policy     FC Name             Queue Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Qos Sap-Egress FC Port-Redirect-Group-Queue References
===============================================================================
Sap Policy     FC Name             Queue Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
Queue Group Policer
===============================================================================
Policer Id     : 1
Description    : (Not Specified)
```

```
PIR Adptn      : closest                      CIR Adptn    : closest
Parent         : none                         Level        : 1
Weight         : 1                            Adv. Cfg Plcy: none
Admin PIR      : max                          Admin CIR    : 0
CBS            : def                          MBS          : def
Hi Prio Only   : def                          Pkt Offset   : 0
Profile Capped : Disabled
StatMode       : minimal
===============================================================================
Policer Id     : 2
Description    : (Not Specified)
PIR Adptn      : closest                      CIR Adptn    : closest
Parent         : none                         Level        : 1
Weight         : 1                            Adv. Cfg Plcy: none
Admin PIR      : max                          Admin CIR    : 0
CBS            : def                          MBS          : def
Hi Prio Only   : def                          Pkt Offset   : 0
Profile Capped : Disabled
StatMode       : minimal
-------------------------------------------------------------------------------
HSMDA PIR Admin  Packet  WRR     MBS       Slope Plcy        WRR Plcy
Queue PIR Rule   Offset  Weight            Max Class         Burst Lmt
-------------------------------------------------------------------------------
1     max        add 0   1       default   default           n/a
      closest                              8                 default
2     max        add 0   1       default   default           n/a
      closest                              8                 default
3     max        add 0   1       default   default           n/a
      closest                              8                 default
4     max        add 0   1       default   default           n/a
      closest                              8                 default
5     max        add 0   1       default   default           n/a
      closest                              8                 default
6     max        add 0   1       default   default           n/a
      closest                              8                 default
7     max        add 0   1       default   default           n/a
      closest                              8                 default
8     max        add 0   1       default   default           n/a
      closest                              8                 default
-------------------------------------------------------------------------------
===============================================================================
*A:PE-1#
```

The preceding output shows that the ingress template has an instance 2, which is applied to card 5 fp 1, and the egress template has an instance 2, which is applied to port 5/1/5.

The details of network QoS policy 10 shows the redirection to the ingress and egress queue groups, as follows:

```
*A:PE-1# show qos network 10 detail | match expression "Egress Forwarding Class
Mapping|Ingress Forwarding Class Mapping|FC Value|Redirect"
Egress Forwarding Class Mapping
FC Value         : 0                    FC Name          : be
Redirect Grp Q   :  None                Redirect Grp Plcr:  None
FC Value         : 1                    FC Name          : l2
Redirect Grp Q   :  None                Redirect Grp Plcr:  1
```

```
FC Value         : 2                    FC Name          : af
Redirect Grp Q   :  1                   Redirect Grp Plcr:  None
FC Value         : 3                    FC Name          : l1
Redirect Grp Q   :  None                Redirect Grp Plcr:  None
FC Value         : 4                    FC Name          : h2
Redirect Grp Q   :  None                Redirect Grp Plcr:  None
FC Value         : 5                    FC Name          : ef
Redirect Grp Q   :  2                   Redirect Grp Plcr:  2
FC Value         : 6                    FC Name          : h1
Redirect Grp Q   :  None                Redirect Grp Plcr:  None
FC Value         : 7                    FC Name          : nc
Redirect Grp Q   :  None                Redirect Grp Plcr:  None
Ingress Forwarding Class Mapping
FC Value                : 0             FC Name                 : be
Redirect UniCast Plcr   : None          Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 1             FC Name                 : l2
Redirect UniCast Plcr   : None          Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 2             FC Name                 : af
Redirect UniCast Plcr   : 1             Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 3             FC Name                 : l1
Redirect UniCast Plcr   : None          Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 4             FC Name                 : h2
Redirect UniCast Plcr   : None          Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 5             FC Name                 : ef
Redirect UniCast Plcr   : 2             Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 6             FC Name                 : h1
Redirect UniCast Plcr   : None          Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
FC Value                : 7             FC Name                 : nc
Redirect UniCast Plcr   : None          Redirect MultiCast Plcr : None
Redirect BroadCast Plcr : None          Redirect Unknown Plcr   : None
*A:PE-1#
```

The preceding output shows that:

- Ingress
  - FC af is redirected to unicast policer 1 in the queue group instance, to be specified when the policy is applied to a network interface.
  - FC ef is redirected to unicast policer 2 in the queue group instance, to be specified when the policy is applied to a network interface.
- Egress
  - FC af is redirected to queue 1 in the queue group instance, to be specified when the policy is applied to a network interface.
  - FC ef is redirected to policer 2 and queue 2 in the queue group instance, to be specified when the policy is applied to a network interface, so that the policer 2 traffic exits using queue 2 in the port network egress queue group instance.

The queue group instances used by the network interface are shown as follows:

```
*A:PE-1# show router interface "PE-1-int2-2" detail | match post-lines 3 "QoS Queue-
Group Redirection Details"
QoS Queue-Group Redirection Details
-------------------------------------------------------------------------------
Ingress FP QGrp  : qg2                  Egress Port QGrp  : qg2
Ing FP QGrp Inst : 2                    Egr Port QGrp Inst: 2
```

After traffic is sent through the network, it can be shown in the FP ingress network queue group policers, as follows:

```
*A:PE-1# show card 5 fp 1 ingress queue-group "qg2" instance 2 mode network
statistics
===============================================================================
Card:5  Net.QGrp: qg2  Instance: 2
===============================================================================
Group Name    : qg2
Description   : (Not Specified)
Pol Ctl Pol   : None                    Acct Pol      : None
Collect Stats : disabled
-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
                        Packets                 Octets
Ing. Policer:  1  Grp: qg2 (Stats mode: minimal)
Off. All              : 1000                    128000
Dro. All              : 0                       0
For. All              : 1000                    128000
Ing. Policer:  2  Grp: qg2 (Stats mode: minimal)
Off. All              : 1000                    128000
Dro. All              : 0                       0
For. All              : 1000                    128000
===============================================================================
*A:PE-1#
```

The traffic sent through the port egress network queue group queues can also be shown, as follows:

```
*A:PE-1# show port 5/1/5 queue-group "qg2" instance 2 egress network statistics
-------------------------------------------------------------------------------
Ethernet port 5/1/5 Network Egress queue-group
-------------------------------------------------------------------------------
                        Packets                 Octets
Egress Queue:  1    Group: qg2    Instance-Id:  2
In Profile forwarded  : 0                     0
In Profile dropped    : 0                     0
Out Profile forwarded : 1000                  128000
Out Profile dropped   : 0                     0
Egress Queue:  2    Group: qg2    Instance-Id:  2
In Profile forwarded  : 0                     0
In Profile dropped    : 0                     0
Out Profile forwarded : 1000                  128000
Out Profile dropped   : 0                     0
Egress Policer:  1  Group: qg2  Instance-Id: 2
Stats mode: minimal
```

```
Off. All              : 1000                    128000
Dro. All              : 0                       0
For. All              : 1000                    128000
Egress Policer:  2  Group: qg2  Instance-Id: 2
Stats mode: minimal
Off. All              : 1000                    128000
Dro. All              : 0                       0
For. All              : 1000                    128000
-------------------------------------------------------------------------------
*A:PE-1#
```

In the preceding output, the traffic on queue 1 is FC af, on policer 1 is FC l2, and on policer 2 is FC ef, with the post-policed traffic on queue 2.

Finally, the FC l2 traffic using the network egress queue group policer 1 can be shown in queue 2 of the default network egress queues, as follows:

```
*A:PE-1# show port 5/1/5 detail | match post-lines 4 "Egress Queue  2"
Egress Queue  2              Packets                 Octets
     In Profile forwarded  :    0                       0
     In Profile dropped    :    0                       0
     Out Profile forwarded :    1000                    128000
     Out Profile dropped   :    0                       0
```

Traffic forwarded through both FP ingress network and port egress network queue groups can be monitored, as follows:

```
monitor card <slot-number> fp <fp-number>  ingress {access|network} queue-group
<queue-group-name> instance <instance-id> [interval <seconds>][repeat<repeat>]
policer <policer-id> [absolute | percent-rate [<reference-rate>]]

monitor port queue-group <queue-group-name> egress <access> [instance <instance-id>]
[egress-queue <egress-queue-id>] [interval <seconds>] [repeat <repeat>]
[absolute|rate]
```

The summary of the queue groups applied to a port is shown as follows:

```
*A:PE-1# show port 5/1/5 queue-group summary

===============================================================================
Port queue-group summary
===============================================================================
Access-egress queue groups:
----------------------------
policer-output-queues
Total number of access-egress queue groups  : 1

Network-egress queue groups:
----------------------------
qg2
Total number of network-egress queue groups : 1

Access-ingress queue groups:
----------------------------
Total number of access-ingress queue groups : 0
```

```
===============================================================================
*A:PE-1#
```

The total usage of queue groups is shown as follows:

```
*A:PE-1# show qos queue-group summary

===============================================================================
Queue-group instances per card
===============================================================================
card      port-acc-ing  port-acc-egr  port-nw-egr   fp-acc-ing    fp-nw-ing
-------------------------------------------------------------------------------
1         0             0             0             0             0
2         0             0             0             0             0
3         0             0             0             0             0
4         0             0             0             0             0
5         0             2             1             0             1
-------------------------------------------------------------------------------
Total ingress QG templates per system :  4
Total egress  QG templates per system :  5
===============================================================================
*A:PE-1#
```

The preceding output includes the created ingress template plus the three system-created ingress templates (making four in total), and the created egress template plus the four system-created egress templates (making five in total). There is the one applied FP network ingress queue group instance on card 5. There is the one created port network egress queue group instance. There are also two port access egress queue group instances, which are the policer-output-queues queue group instances associated with the access port used for IES service interface (not discussed) and the access side of the hybrid port 5/1/5.

Egress Policed Subscriber Configuration Example

Queue groups are only applicable to subscribers for egress policed traffic. By default, subscriber egress policed traffic exits the port using a queue in the egress access policer-output-queues queue group instance. The queue used is determined by the FC mapping in the policer-output-queues queue group template. This is the same default operation as in the SAP example.

The subscriber policed traffic can be sent to a different queue group instance using the inter-dest-id and a host-match (described when applying an egress access queue group template to a port), which represent an intermediate destination, such as a downstream DSLAM or GPON OLT. The inter-dest-id can be associated with a subscriber host when it is created; this would usually be received from the DHCP, or RADIUS server, or Diameter Gx (Policy and Rule Charging Function), or the local user database, or configured under a static host.

An alternative to host matching on an inter-dest-id is to match on the top VLAN tag when a QinQ SAP is configured using a default inter-dest-id.

A default inter-dest-id can be configured in IES, VPRN, and VPLS services, and under an MSAP policy, as follows:

```
configure
    service
        {ies|vprn} <service-id>
            subscriber-interface <ip-int-name>
                group-interface <ip-int-name> [create]
                    sap <sap-id>
                        sub-sla-mgmt
                              def-inter-dest-id {<inter-dest-string>|use-top-q}


configure
    service
        vpls <service-id>
            sap <sap-id>
                sub-sla-mgmt
                    def-inter-dest-id {<inter-dest-string>|use-top-q}


configure
    subscriber-mgmt
        msap-policy <msap-policy-name> [create]
                sub-sla-mgmt
                    def-inter-dest-id {<inter-dest-string>|use-top-q}
```

The egress queue group template is configured as follows:

```
configure
    qos
        queue-group-templates
            egress
                queue-group "qg3" create
                    queue 1 best-effort create
                    exit
                    queue 2 expedite create
                    exit
                    fc af create
                        queue 1
                    exit
                    fc ef create
                        queue 2
                    exit
                exit
            exit
```

The egress template is applied to port 5/1/5 to create a port access queue group instance. A host match is configured under the created queue group instance on an egress access port. In the following, the host match dslam-1 is used:

```
configure
    port 5/1/5
```

```
        ethernet
            mode hybrid
            encap-type dot1q
            access
                egress
                    queue-group "qg3" instance 1 create
                        host-match dest "dslam-1" create
                    exit
                exit
```

A host-match can only be configured under instance 1 of a port access egress queue group; if its configuration is attempted on a different instance, the following error is displayed:

```
*A:PE-1# configure port 5/1/5 ethernet access egress queue-group "qg3" instance 3
create host-match dest another-dslam create
MINOR: PMGR #1337 Host match entries only supported on port access egress queue
groups with system default instance 1
*A:PE-1#
```

The subscriber host uses a SAP egress QoS policy in an egress SLA profile to map FCs to egress queue and policers. The SAP egress QoS policy is created with FC af using policer 1 and FC ef using policer 2, as follows:

```
configure
    qos
        sap-egress 30 create
            queue 1 create
            exit
            queue 2 create
            exit
            policer 1 create
            exit
            policer 2 create
            exit
            fc af create
                policer 1
            exit
            fc be create
                queue 1
            exit
            fc ef create
                policer 2
            exit
            fc l2 create
                queue 2
            exit
        exit
```

To redirect the subscriber egress policed traffic to the access egress queue group qg3 instance 1 on port 5/1/5, which is configured with the host-match, an inter-dest-id is configured for the created subscriber static host, as follows:

```
configure
    service
```

```
vprn 3 customer 1 create
    route-distinguisher 65536:200
    subscriber-interface "sub-int-1" create
        address 10.3.2.1/24
        group-interface "group-int-1" create
            arp-populate
            sap 5/1/5:3 create
                sub-sla-mgmt
                    def-sub-profile "basic-sub"
                    def-sla-profile "basic-sla"
                    multi-sub-sap 200
                    single-sub-parameters
                        profiled-traffic-only
                    exit
                    no shutdown
                exit
                static-host ip 10.3.2.2 mac 00:00:10:03:02:02 create
                    inter-dest-id "dslam-1"
                    sla-profile "basic-sla"
                    sub-profile "basic-sub"
                    subscriber "sub1"
                    no shutdown
                exit
            exit
```

The host match configured under the egress queue group instance is shown as follows:

```
A:PE-1# show port 5/1/5 queue-group "qg3" instance 1 access | match post-
lines 3 Host-Matches
Host-Matches
-------------------------------------------------------------------------------
Dest: dslam-1
-------------------------------------------------------------------------------
```

When the subscriber host is created, the inter-dest-id for the subscriber host is shown as follows:

```
A:PE-1# show service active-subscribers subscriber "sub1" detail | match expression
"Subscriber sub1 |Sub. Int Dest Id"
Subscriber sub1 (basic-sub)
Sub. Int Dest Id : "dslam-1"
```

The inter-dest-id dslam-1 is matched against the host-match destination configured on the access egress group instances on the port on which the host is being created. If a match is found, the subscriber egress policed traffic will use that egress queue group instance, with the actual queue used being selected by the FC-to-queue mapping in the related queue group template. Otherwise, the default policer-output-queues queue group instance will be used.

The egress queue group instance subscriber host associations are shown as follows:

```
A:PE-1# show port 5/1/5 queue-group egress "qg3" associations | match post-
lines 6 Subscriber-Host
```

```
Subscriber-Host Queue-Group Associations
------------------------------------------------------------------------------
svc-id : 3 (VPRN)
 sap   : 5/1/5:3
 subscr: sub1
 ip    : 10.3.2.2
 mac   : 00:00:10:03:02:02  pppoe-sid: N/A
```

The following output shows that FC ef is using policer 2 and its traffic exits using queue 2 in queue group qg3, while FC af is using policer 1 and its traffic exits using queue 1 in queue group qg3.

```
A:PE-1# show qos policer subscriber "sub1" egress detail | match post-
lines 4 "Policer Info"
Policer Info (Sub=1:1 3->5/1/5:3->2), Slot 5
==============================================================================
Policer Name       : Sub=1:1 3->5/1/5:3->2
Direction          : Egress             Fwding Plane       : 1
FC->[QGrp:Inst->]Q : ef->qg3->2
Policer Info (Sub=1:1 3->5/1/5:3->1), Slot 5
==============================================================================
Policer Name       : Sub=1:1 3->5/1/5:3->1
Direction          : Egress             Fwding Plane       : 1
FC->[QGrp:Inst->]Q : af->qg3->1
```

After traffic is sent through the subscriber, the egress policed traffic can be shown in the port egress access queue group instance, as follows:

```
A:PE-1# show port 5/1/5 queue-group "qg3" instance 1 access egress statistics

------------------------------------------------------------------------------
Ethernet port 5/1/5 Access Egress queue-group
------------------------------------------------------------------------------
                        Packets                 Octets

Egress Queue: 1 Group: qg3 Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 1000                    128000
Out Profile dropped   : 0                       0
Egress Queue: 2 Group: qg3 Instance: 1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 1000                    128000
Out Profile dropped   : 0                       0
------------------------------------------------------------------------------
A:PE-1#
```

The traffic statistics are either counted in the subscriber queues or policers, or in the queue group instance queues, but not in both. However, summary statistics per SAP are available when using FP ingress queue groups.

The number of valid ingress packets received on a SAP, or subscribers on that SAP, can be shown in the sap-stats output, as follows. The received valid counter includes both the local SAP counters and the counters from the related FP ingress queue group instance. This is useful to display SAP-level traffic statistics when forwarding classes in a SAP ingress policy have been redirected to an ingress queue group.

```
*A:PE-1# show service id 3 sap 5/1/5:3 sap-stats | match post-
lines 6 "Forwarding Engine Stats"
Forwarding Engine Stats
Dropped               : 0                      0
Received Valid        : 4000                   512000
Off. HiPrio           : 0                      0
Off. LowPrio          : 0                      0
Off. Uncolor          : 0                      0
Off. Managed          : 0                      0
```

Traffic forwarded through port egress access queue groups can be monitored, as follows:

```
monitor port queue-group <queue-group-name> egress <access> [instance <instance-id>]
[egress-queue <egress-queue-id>] [interval <seconds>] [repeat <repeat>]
[absolute|rate]
```

The summary of the queue groups applied to a port is shown as follows:

```
A:PE-1# show port 5/1/5 queue-group summary

===============================================================================
Port queue-group summary
===============================================================================
Access-egress queue groups:
----------------------------
qg3
policer-output-queues
Total number of access-egress queue groups  : 2

Network-egress queue groups:
----------------------------
Total number of network-egress queue groups : 0

Access-ingress queue groups:
----------------------------
Total number of access-ingress queue groups : 0
===============================================================================
A:PE-1#
```

The total usage of queue groups is shown as follows:

```
A:PE-1# show qos queue-group summary

===============================================================================
Queue-group instances per card
===============================================================================
card      port-acc-ing  port-acc-egr  port-nw-egr   fp-acc-ing    fp-nw-ing
-------------------------------------------------------------------------------
```

```
1          0              0              0              0              0
2          0              0              0              0              0
3          0              0              0              0              0
4          0              0              0              0              0
5          0              3              0              0              0
-------------------------------------------------------------------------------
Total ingress QG templates per system :  3
Total egress  QG templates per system :  5
===============================================================================
A:PE-1#
```

The preceding output includes the three system-created ingress templates, and the created egress template plus the four system-created egress templates (making five in total). There are three port access egress queue group instances (the applied queue group instance and two instances of the policer-output-queues queue group), one on each access port used for VPRN service interfaces.

Conclusion

This chapter described the use of queue groups as a mechanism to provide an aggregate QoS control for multiple SAPs and per-network interface QoS control. The configuration steps and commands are described, followed by example configurations on a SAP, network interface, and for egress policed traffic subscriber traffic.

Pseudowire QoS

This chapter describes pseudowire QoS configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This example is applicable to the 7950 XRS-16c/20, 7750 SR-7/12, 7750 SR-c4/12 and 7450 ESS-6/6v/7/12 platforms when all network IP interfaces are on IOM3-XP/IMM (FP2 and above) hardware.

The information and configuration is based on release 11.0.R4. There are no specific prerequisites for this configuration.

Overview

A pseudowire (PW) provides a virtual connection across an IP or MPLS network between services configured on provider edge (PE) devices. In SR OS release R10.0.R1, and later, it is possible to provide specific QoS to either a single pseudowire or a multiple pseudowires. This is supported for the following:

- SDP
  - MPLS
  - GRE
- Epipe
  - Including vc-switching and dynamic MS-PW
  - PBB-epipe
  - BGP-VPWS (release 11.0.R1 and later)
- VPLS
  - Mesh and spoke SDP

- LDP signaled pseudowires
- BGP-AD signaled pseudowires
- I-VPLS, B-VPLS
- R-VPLS
- BGP-VPLS

• Spoke termination on IES/VPRN (both Epipe and Ipipe)
• Apipe (from R10.0.R4)
• Cpipe (from R10.0.R4)
• Fpipe (from R10.0.R4)
• Ipipe (from R10.0.R4)

It is supported at ingress on both Ethernet and POS/TDM ports on an IOM3-XP/IMM and only on Ethernet ports at egress.

Bandwidth control is achieved using queue-groups which are implemented per flexpath (FP) at the ingress and per port at the egress (these being relative to the data path through the system), as shown in Figure 380 and Figure 381, respectively.

Figure 380 **Ingress PW QoS**

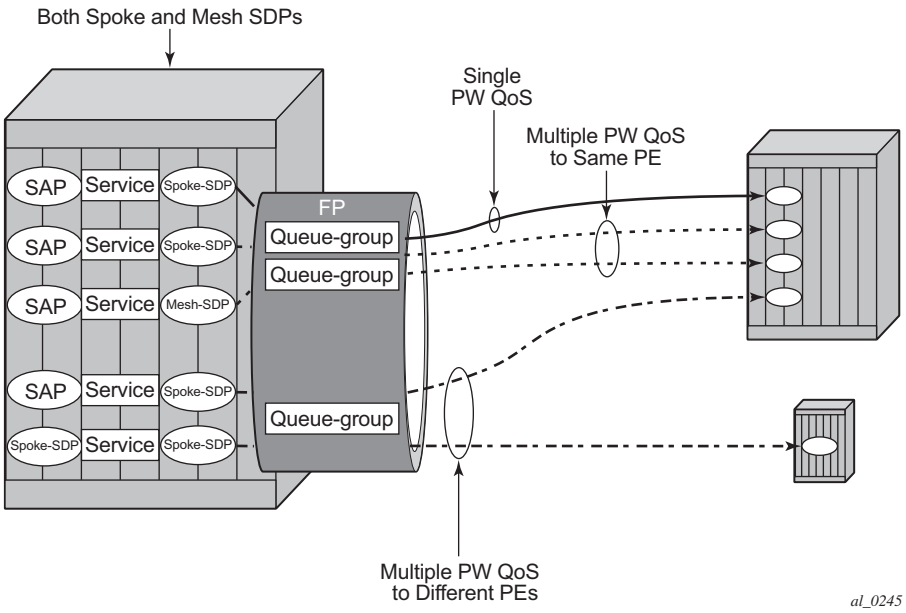
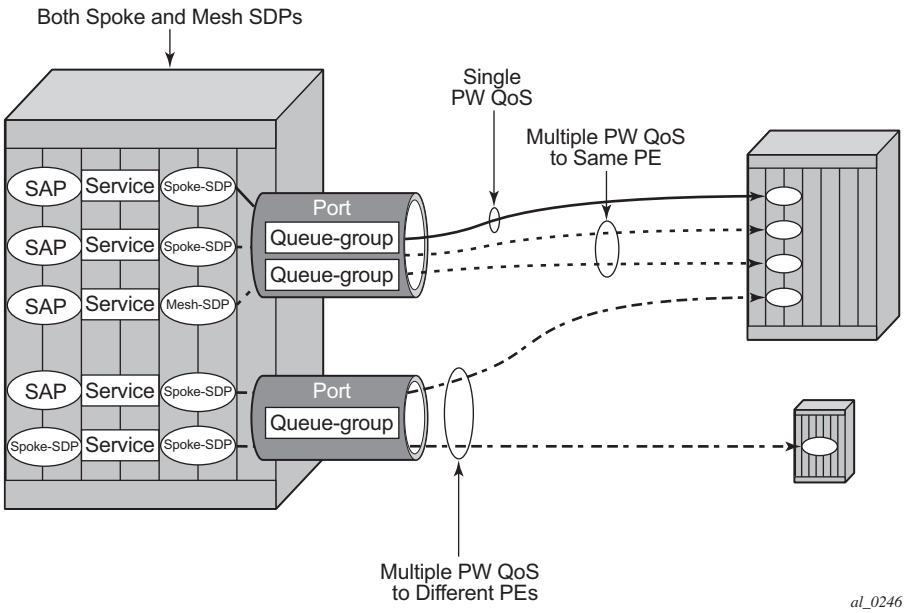


Figure 381 **Egress PW QoS**



Bandwidth control is applied independently for ingress and egress, and can be set up for a single pseudowire or for multiple pseudowires where the remote services are located on a single PE or on multiple PEs.

It is possible to benefit from Hierarchical QoS which can be configured under the queue-groups, but this is beyond the scope of this chapter.

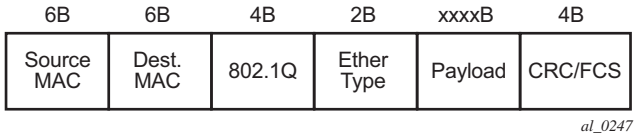
The ingress and egress classification and egress marking is configured by applying a network QoS policy to each pseudowire.

Ingress QoS

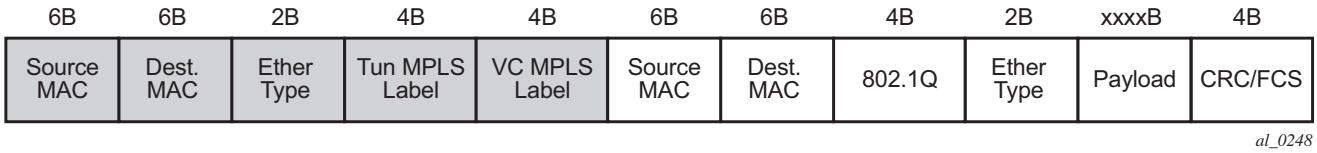
Ingress QoS is achieved using a queue group which is applied to an ingress FP on a card. Queue groups applied to an FP can only contain policers, not queues. The network QoS policy applied to the pseudowire redirects forwarding classes (FCs) to the individual queue group (unicast or multipoint) policers. The actual queue group to be used is defined separately to the network QoS policy, thereby allowing the network QoS policies to be independent from the queue groups used and therefore both are reusable.

Ingress bandwidth control does not take into account the outer Ethernet header, the MPLS labels/control word or GRE headers, or the FCS of the incoming frame. The configuration allows an offset to be added or subtracted from the received frame size in order to change the actual length used for the bandwidth control.
For example: if the same ingress rate is configured on a pseudowire (without a control word) and a dot1q SAP, what packet-byte-offset needs to be used on the pseudowire in order to achieve the same throughput as on the SAP?

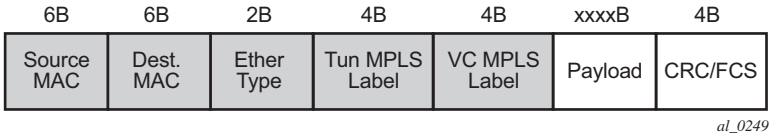
- SAP — The following shows the bytes in the frame that are used by default on a policer for the rate at a SAP ingress.



- VPLS Pseudowire — For a tagged (**vc-type vlan**) pseudowire, it would be necessary to add 4 bytes using the packet-byte-offset applied to the ingress policer in order to achieve the same throughput as on the SAP. This compensates for the omission of the FCS that is included on the SAP and so needs to be added.



- VPRN Pseudowire — For an Ipipe (**vc-type** ipipe) pseudowire, it would be necessary to add 22 bytes using the packet-byte-offset to the ingress policer to achieve the same throughput as on the SAP. This compensates for the omission of the source and destination MAC addresses (12 bytes), Ether type (2 bytes), VLAN tag (4 bytes) and the FCS (4 bytes) that are included on the SAP and so needs to be added.



The ingress classification is configured in the ingress section of the network QoS policy and is based on the outer encapsulation header only, the outer Ethernet header (dot1p/DE), MPLS labels (EXP), or GRE headers (DSCP). At an egress LER, the ler-use-dscp is applicable only to IES and VPRN pseudowires.

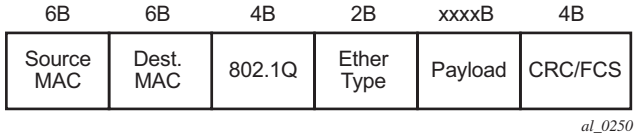
Egress QoS

Egress QoS is achieved using a queue group which is applied to an egress port. Queue groups applied to a port can contain both policers and queues. The network QoS policy applied to the pseudowire redirects forwarding classes (FCs) to the individual queue group policers/queues. The actual queue group to be used is defined separately to the network QoS policy, thereby allowing the network QoS policies to be independent from the queue groups used and therefore both are reusable.

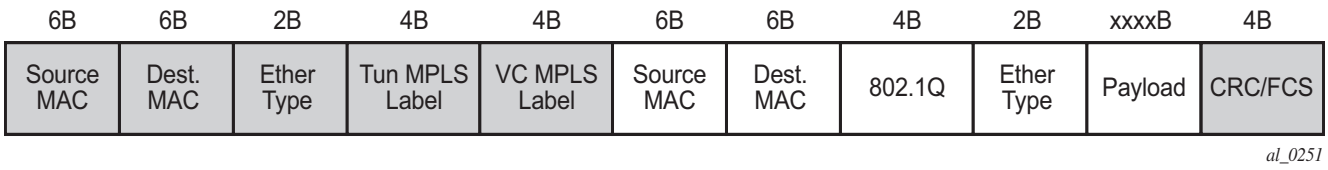
Egress bandwidth control takes into account the outer Ethernet header, MPLS labels/control word, or GRE headers, and the FCS of the outgoing frame. The configuration allows an offset to be added or subtracted from the sent frame size in order to affect the actual length used for the bandwidth control.

For example, if the same egress rate is configured on a pseudowire (without a control word) and a dot1q SAP, what packet-byte-offset needs to be used on the pseudowire in order to achieve the same throughput as on the SAP?

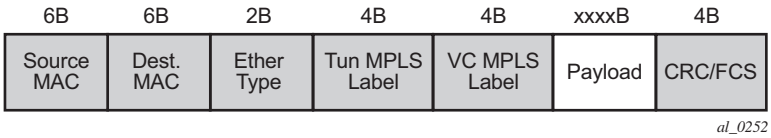
- SAP — The following shows the bytes in the frame that are used by default on a policer/queue at a SAP egress.



- VPLS Pseudowire — For a tagged (**vc-type vlan**) pseudowire, it would be necessary to subtract 22 bytes using the packet-byte-offset applied to the egress policer/queue applied to achieve the same throughput as on the SAP. This compensates for the MPLS header (source and destination MAC addresses (12 bytes), Ether type (2 bytes), two labels (8 bytes)) that is not included on the SAP and needs to be subtracted.



- VPRN Pseudowire — For an Ipipe (**vc-type ipipe**) pseudowire, it would be necessary to subtract 4 bytes using the packet-byte-offset applied to the egress policer/queue applied to achieve the same throughput as on the SAP. This compensates for the MPLS header (source and destination MAC addresses (12 bytes), Ether type (2 bytes), two labels (8 bytes)) that is not included on the SAP so is subtracted, and the source and destination MAC addresses (12 bytes), dot1q header (4 bytes) and Ether type (2 bytes) of the SAP frame which needs to be added. This results in subtracting 4 bytes.

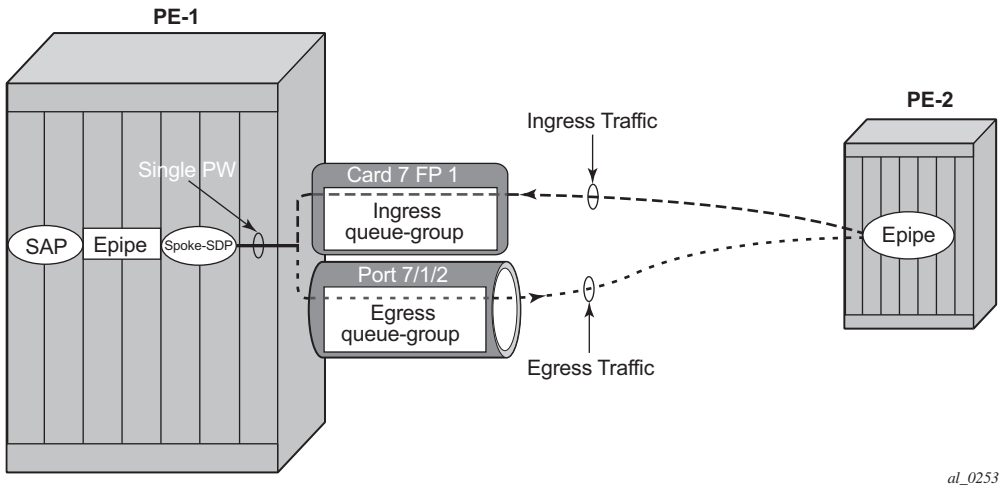


The egress classification and marking is configured in the egress section of the network QoS policy. DSCP/prec egress reclassification is supported in SR OS release R10.0.R4, and later, for IES and VPRN spoke SDPs. The egress marking affects the outer encapsulation header, the outer Ethernet header (dot1p/DE), MPLS labels (EXP) or GRE headers (DSCP).

Configuration

The configuration of pseudowire QoS is described using an Epipe pseudowire. The topology is shown in Figure 382.

Figure 382 **Example Epipe Pseudowire Topology**



The following prerequisite configuration is assumed to be in place:

- Hardware provisioning
- IP address and routing
- MPLS protocols
- SDP
- Epipe service, including the SAP
- SAP QoS policies

Traffic is sent across a virtual leased line between PE-1 and PE-2 using Epipes with a pseudowire configured as a spoke SDP on each PE. The QoS is applied to the pseudowire at the ingress and egress of PE-1.

The following configuration is required for applying pseudowire QoS:

- Create the ingress and egress queue groups.
  These contain the ingress policer and egress policer/queue definitions.
- Create an instance of the ingress queue group on the ingress FP and instance of the egress queue group on the port that will be used for the pseudowire traffic.
- Create a network QoS policy to redirect the traffic to the ingress and egress queue groups, and to perform the ingress classification and egress marking.
- Apply the network QoS policy, together with the reference to the ingress and egress queue group instances, to the spoke SDP representing the pseudowire.

The traffic consists of two bidirectional flows, one in FC BE and one in FC EF. At the ingress of the pseudowire, each FC is assigned to its own policer, whereas at the egress of the pseudowire, FC BE is assigned to a queue and FC EF is assigned to a policer.

Although this example makes use of both ingress and egress queue groups, the focus is pseudowire QoS, so the full details of queue group configuration are not covered.

Create the Ingress and Egress Queue Groups

Queue groups are created using templates, which are separate for ingress and egress. The following shows the queue group templates configured.

```
configure qos
        queue-group-templates
            ingress
                queue-group "ingress-queue-group" create
                    policer 1 create
                        rate 6000
                        packet-byte-offset add 4
```

```
                exit
                policer 2 create
                    rate 4000
                    packet-byte-offset add 4
                exit
            exit
        exit
        egress
            queue-group "egress-queue-group" create
                queue 1 best-effort create
                    rate 6000
                    xp-specific
                        packet-byte-offset subtract 22
                    exit
                exit
                policer 1 create
                    rate 4000
                    packet-byte-offset subtract 22
                exit
            exit
        exit
    exit
```

The ingress queue group has two policers associated with it; policer 1 will be used for the FC BE traffic and policer 2 will be used for the FC EF traffic. The configuration of policers in an ingress queue group is the same as that in a sap-ingress QoS policy, with the exception that the percent-rate is not supported within the queue group.

In order to achieve the same ingress throughput as that when applying the same rates to policers on a dot1q tagged SAP, the packet-byte-offset adds 4 bytes to the packet length for both policers.

The egress queue group has one queue (queue 1) that will be used for the FC BE traffic and one policer (policer 1) that will be used for the FC EF traffic. The configuration of policers in an egress queue group is the same as that in a sap-egress QoS policy, with the exception that the percent-rate is not supported within the queue group. The configuration of queues in an egress queue group is the same as in a sap-egress QoS policy, with the exception that the avg-frame-overhead is not supported within the queue group.

In order to achieve the same egress throughput as that when applying the same rates to policers/queues on a dot1q tagged SAP, the packet-byte-offset subtracts 22 bytes from the packet length for both the policer and queue.

Rates have been configured such that the ingress and egress capacity of the BE traffic is 6Mb/s and 4Mb/s for the EF traffic.

Create the Ingress FP and Egress Port Queue Group Instances

The queue group templates are then applied as individual instances to the ingress FP and egress port; using instances allows the reuse of the same template.

Following is the ingress FP configuration. From a QoS perspective, it is also possible to configure a policer-control-policy under the ingress queue group in order to perform hierarchical policing. In release R11.0R4, and later, the configuration supports overrides for both the policer-control-policy parameters and some of the queue group policer parameters.

```
configure
    card 7
        card-type imm5-10gb-xfp
        mda 1
            no shutdown
        exit
        fp 1
            ingress
                network
                    queue-group "ingress-queue-group" instance 1 create
                    exit
                exit
            exit
        exit
        no shutdown
    exit
```

Following is the egress port configuration. From a QoS perspective, it is also possible to configure under the egress queue group a policer-control-policy in order to perform hierarchical policing, a scheduler-policy in order to perform hierarchical shaping and overrides for some of the queue group queue parameters.

```
configure
    port 7/1/2
        ethernet
            network
                egress
                    queue-group "egress-queue-group" instance 1 create
                    exit
                exit
            exit
        exit
        no shutdown
    exit
```

If there are redundant network interfaces over which the pseudowire traffic can enter or exit the system, it is necessary to configure any ingress FP and egress port queue groups consistently across all possible interfaces to be used by the pseudowire to ensure the QoS is always applied. If a queue group configuration was omitted, the pseudowire would not be subject to the QoS defined in that queue group.

If a LAG is used, the system only allows the egress port queue group to be added or removed from the LAG primary port, thereby keeping the LAG configuration consistent. However, this is not possible at the ingress as the queue-group is applied at the FP, so it is necessary to ensure that the ingress queue group is applied consistently on all FPs corresponding to the configured LAG.

Create the Network QoS Policy

A network QoS policy is created to redirect ingress and egress traffic to the respective queue groups, and perform ingress classification (in this example).

The redirection to the queue group policer/queue is performed per FC.

At ingress, traffic can be redirected to policers (being the same or different policers) based on the traffic type. Unicast traffic is redirected to a policer specified by the policer command and will use the ingress shared policer-output-queues to access the switch fabric. All multipoint traffic is redirected to the policer specified by the multicast-policer command (for example, with a pseudowire configured in a VPLS service, all broadcast, unknown, and multicast traffic will use this policer). The multipoint traffic accesses the switch fabric using the Ingress Multicast Path Management queues. It is possible to individually redirect one traffic type (unicast or multipoint) within an FC to a queue group policer while allowing the other traffic type to use default network queues.

At egress, traffic can be redirected to a queue or to a policer. The policed traffic will exit the egress port using one of the default network queues (with the queue chosen by FC assignment) or optionally can use a queue in the egress queue group if configured in the port-redirect-group command following the policer parameter.

Any FC not redirected to a queue-group, will continue to use the regular default network ingress and egress queues.

The syntax for the FC redirection is as follows.

```
config# qos
  network <network-policy-id> [create]
      ingress
          fc <fc-name>
              fp-redirect-group multicast-policer <policer-id>
              fp-redirect-group policer <policer-id>
```

```
      egress
          fc <fc-name>
              port-redirect-group {queue <queue-id>|
policer <policer-id> [queue <queue-id>]}
```

The required commands are shown below.

```
configure qos
        network 10 create
            ingress
                lsp-exp 5 fc ef profile in
                fc be
                    fp-redirect-group policer 1
                exit
                fc ef
                    fp-redirect-group policer 2
                exit
            exit
            egress
                fc be
                    port-redirect-group queue 1
                exit
                fc ef
                    port-redirect-group policer 1
                exit
            exit
        exit
```

At ingress, the FC BE and FC EF traffic are redirected to the two policers in the queue-group applied to the FP. At egress, the two FCs are redirected to the queue and policer in the queue group applied to the egress port.

The ingress classification required here is for the traffic which is received with exp=5 to be in FC EF.

Apply Network QoS Policy with Queue Group Instances to the Spoke SDP

To apply the QoS to the pseudowire, the following commands can be used, dependent on the service type.

```
config# service {apipe|cpipe|epipe|fpipe|ipipe} <service-id>
  spoke-sdp <sdp-id:vc-id>
     ingress
       qos <network-policy-id> fp-redirect-group <queue-group-name>
                               instance <instance-id>
     egress
       qos <network-policy-id> port-redirect-group <queue-group-name>
                               instance <instance-id>
```

```
config# service {ies|vprn} <service-id>
  interface <ip-int-name>
    spoke-sdp <sdp-id:vc-id>
      ingress
        qos <network-policy-id> fp-redirect-group <queue-group-name>
                                instance <instance-id>
      egress
        qos <network-policy-id> port-redirect-group <queue-group-name>
                                instance <instance-id>


config# service vpls <service-id>
  {spoke-sdp|mesh-sdp} <sdp-id:vc-id>
     ingress
       qos <network-policy-id> fp-redirect-group <queue-group-name>
                               instance <instance-id>
     egress
       qos <network-policy-id> port-redirect-group <queue-group-name>
                               instance <instance-id>
```

For services using BGP auto-discovery to signal the pseudowire, the QoS configuration is included in the pseudowire template.

```
config# service pw-template <policy-id>
     ingress
       qos <network-policy-id> fp-redirect-group <queue-group-name>
                               instance <instance-id>
     egress
       qos <network-policy-id> port-redirect-group <queue-group-name>
                               instance <instance-id>
```

To propagate changes in a pw-template to existing BGP-AD pseudowires, it is necessary to use the following command:

```
tools perform service eval-pw-template policy-id
```

The allow-service-impact parameter is not required for changing the ingress or egress QoS definition as these do not affect the operational state of the pseudowire.

QoS applied directly to a pseudowire, using the preceding commands, takes precedence over any QoS applied to the network interface (using a network QoS policy with or without queue group redirection).

Each time a pseudowire uses a network egress port queue group, an FP resource is allocated. This only requires that the pseudowire egress QoS is configured with a port-redirect-group, and will occur even if there are no FCs redirected using a port-redirect-group within the configured network QoS policy. The resources used can be seen using the **tools dump system-resources** command and is listed under Egr Network Queue Group Mappings which is part of the total for the “Dynamic Service Entries “.

As an Epipe is used in this example, QoS is configured directly under a spoke SDP.

```
configure service
        epipe 1 customer 1 create
            spoke-sdp 1:1 vc-type vlan create
                ingress
                    qos 10 fp-redirect-group "ingress-queue-group" instance 1
                exit
                egress
                    qos 10 port-redirect-group "egress-queue-group" instance 1
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The created network QoS policy is applied at both ingress and egress, with the ingress referencing the ingress queue group instance applied to the FP and the egress referencing the egress queue group instance applied to the port.

Show Output

The configured ingress queue group can be shown, including the details of the configured policers and where it is applied, as follows.

```
*A:PE-1# show qos queue-group "ingress-queue-group" ingress detail
===============================================================================
QoS Queue-Group Ingress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : ingress-queue-group
Description    : (Not Specified)
-------------------------------------------------------------------------------
---snip---
===============================================================================
Queue Group FP Maps
===============================================================================
Card Num       Fp Num              Instance            Type
-------------------------------------------------------------------------------
7              1                   1                   Network
-------------------------------------------------------------------------------
Entries found: 1
-------------------------------------------------------------------------------
===============================================================================
Queue Group Policer
===============================================================================
Policer Id     : 1
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : 6000                          Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : 4
```

```
Profile Capped : Disabled
StatMode       : minimal
===============================================================================
Policer Id     : 2
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : 4000                          Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : 4
Profile Capped : Disabled
StatMode       : minimal
```

Similar information can be shown for the egress queue group, including the details of the configured queue and policer and again where it is applied.

```
*A:PE-1# show qos queue-group "egress-queue-group" egress detail
===============================================================================
QoS Queue-Group Egress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : egress-queue-group
Description    : (Not Specified)

-------------------------------------------------------------------------------
Q  CIR Admin PIR Admin CBS         HiPrio PIR Lvl/Wt    Parent    BurstLimit(B)
   CIR Rule  PIR Rule  MBS                CIR Lvl/Wt    Wred-Queue     Slope
   Named-Buffer Pool                      Adv Config Policy Name
-------------------------------------------------------------------------------
1  0         6000      def         def    1/1           None          default
   closest   closest   def                0/1           disabled      default
   (not-assigned)                         (not-assigned)
---snip---
===============================================================================
Queue Group Ports (network)
===============================================================================
Port   Sched Pol    Policer-Ctrl-Pol  Acctg Pol Stats Description QGrp-Instance
-------------------------------------------------------------------------------
7/1/2                                      No                1
-------------------------------------------------------------------------------
---snip---
===============================================================================
Queue Group Policer
===============================================================================
Policer Id     : 1
Description    : (Not Specified)
PIR Adptn      : closest                       CIR Adptn    : closest
Parent         : none                          Level        : 1
Weight         : 1                             Adv. Cfg Plcy: none
Admin PIR      : 4000                          Admin CIR    : 0
CBS            : def                           MBS          : def
Hi Prio Only   : def                           Pkt Offset   : -22
Profile Capped : Disabled
StatMode       : minimal
```

```
---snip---
```

The following command shows where the ingress queue group has been applied.

```
*A:PE-1# show qos queue-group ingress association
===============================================================================
QoS Queue-Group Ingress
===============================================================================
---snip---
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : ingress-queue-group
Description    : (Not Specified)
---snip---
===============================================================================
Queue Group FP Maps
===============================================================================
Card Num       Fp Num              Instance            Type
-------------------------------------------------------------------------------
7              1                   1                   Network
-------------------------------------------------------------------------------
Entries found: 1
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

The following command shows where the egress queue group has been applied.

```
*A:PE-1# show qos queue-group egress association
===============================================================================
QoS Queue-Group Egress
===============================================================================
-------------------------------------------------------------------------------
QoS Queue Group
-------------------------------------------------------------------------------
Group-Name     : egress-queue-group
Description    : (Not Specified)
---snip---
===============================================================================
Queue Group Ports (network)
===============================================================================
Port   Sched Pol    Policer-Ctrl-Pol  Acctg Pol Stats Description QGrp-Instance
-------------------------------------------------------------------------------
7/1/2                                      No                1
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

The following command shows the ingress queue group applied to the FP on card 7.

```
*A:PE-1# show card 7 fp 1 ingress queue-group "ingress-queue-group" instance 1
mode network
===============================================================================
Card:7  Net.QGrp: ingress-queue-group  Instance: 1
===============================================================================
```

```
Group Name    : ingress-queue-group
Description   : (Not Specified)
Pol Ctl Pol   : None                    Acct Pol      : None
Collect Stats : disabled
```

The following command show the details of the policers in the ingress FP queue group.

```
*A:PE-1# show qos policer card 7 fp 1 queue-group "ingress-queue-group" instance 1
network detail
===============================================================================
Policer Info (Net-FPQG-1-ingress-queue-group:1->1), Slot 7
===============================================================================
Policer Name       : Net-FPQG-1-ingress-queue-group:1->1
Direction          : Ingress            Fwding Plane       : 1
Depth PIR          : 0 Bytes            Depth CIR          : 0 Bytes
Depth FIR          : 0 Bytes
MBS                : 7680 B             CBS                : 0 KB
Hi Prio Only       : 768 B              Pkt Byte Offset    : 4
Admin PIR          : 6000 Kbps          Admin CIR          : 0 Kbps
Oper PIR           : 6000 Kbps          Oper CIR           : 0 Kbps
Oper FIR           : 6000 Kbps
Stat Mode          : minimal
PIR Adaption       : closest            CIR Adaption       : closest
Adv.Cfg Plcy       : None               Profile Capped     : disabled
Parent Arbiter Name: (Not Specified)
-------------------------------------------------------------------------------
Arbiter Member Information
-------------------------------------------------------------------------------
Offered Rate       : 0 Kbps
Level              : 0                  Weight             : 0
Parent PIR         : 0 Kbps             Parent FIR         : 0 Kbps
Consumed           : 0 Kbps
-------------------------------------------------------------------------------
===============================================================================
Policer Info (Net-FPQG-1-ingress-queue-group:1->2), Slot 7
===============================================================================
Policer Name       : Net-FPQG-1-ingress-queue-group:1->2
Direction          : Ingress            Fwding Plane       : 1
Depth PIR          : 0 Bytes            Depth CIR          : 0 Bytes
Depth FIR          : 0 Bytes
MBS                : 5 KB               CBS                : 0 KB
Hi Prio Only       : 512 B              Pkt Byte Offset    : 4
Admin PIR          : 4000 Kbps          Admin CIR          : 0 Kbps
Oper PIR           : 4000 Kbps          Oper CIR           : 0 Kbps
Oper FIR           : 4000 Kbps
Stat Mode          : minimal
PIR Adaption       : closest            CIR Adaption       : closest
Adv.Cfg Plcy       : None               Profile Capped     : disabled
Parent Arbiter Name: (Not Specified)
-------------------------------------------------------------------------------
Arbiter Member Information
-------------------------------------------------------------------------------
Offered Rate       : 0 Kbps
Level              : 0                  Weight             : 0
Parent PIR         : 0 Kbps             Parent FIR         : 0 Kbps
Consumed           : 0 Kbps
```

```
-------------------------------------------------------------------------------
===============================================================================
Network Interface Association
-------------------------------------------------------------------------------
No Association Found.
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
SDP Association
-------------------------------------------------------------------------------
Policer Info (1->1:1->10), Slot 7
-------------------------------------------------------------------------------
SDP Association Count : 1
-------------------------------------------------------------------------------
```

The details of the queue and policer in the egress queue group applied to port 7/1/2 can also be shown as follows.

```
*A:PE-1# show port 7/1/2 queue-group egress "egress-queue-group" network instance 1
===============================================================================
Ethernet port 7/1/2 Network Egress queue-group
===============================================================================
Group Name    : egress-queue-group Instance-Id   : 1
Description   : (Not Specified)
Sched Policy  : None               Acct Pol      : None
Collect Stats : disabled           Agg. Limit    : -1

Queues
-------------------------------------------------------------------------------
Queue-Group   : egress-queue-group Instance-Id   : 1      Queue-Id   : 1
Description   : (Not Specified)
Admin PIR     : 6000*              Admin CIR     : 0*
PIR Rule      : closest*           CIR Rule      : closest*
CBS           : def*               MBS           : def*
Hi Prio       : def*

Policers
-------------------------------------------------------------------------------
Queue-Group   : egress-queue-group Instance-Id   : 1      Policer-Id   : 1
Description   : (Not Specified)
Admin PIR     : 4000*              Admin CIR     : 0*
PIR Rule      : closest*           CIR Rule      : closest*
CBS           : def*               MBS           : def*
Hi Prio       : def*

* means the value is inherited
```

The network QoS policy can be shown with the details of the configured FC redirection and ingress classification used on the pseudowire, as follows.

```
*A:PE-1# show qos network 10 detail
===============================================================================
QoS Network Policy
===============================================================================
-------------------------------------------------------------------------------
Network Policy (10)
-------------------------------------------------------------------------------
Policy-id        : 10                    Remark           : False
```

```
Forward Class    : be                      Profile          : Out
LER Use DSCP     : False
Description    : (Not Specified)
---snip---
-------------------------------------------------------------------------------
LSP EXP Bit Map                          Forwarding Class               Profile
-------------------------------------------------------------------------------
5                                        ef                             In
---snip---
-------------------------------------------------------------------------------
Egress Forwarding Class Mapping
-------------------------------------------------------------------------------
FC Value         : 0                       FC Name          : be
- DSCP Mapping
Out-of-Profile   : be                      In-Profile       : be
---snip---
DE Mark          :  None
Redirect Grp Q   :  1                      Redirect Grp Plcr:  None

---snip---
FC Value         : 5                       FC Name          : ef
---snip---
DE Mark          :  None
Redirect Grp Q   :  None                   Redirect Grp Plcr:  1

-------------------------------------------------------------------------------
Ingress Forwarding Class Mapping
-------------------------------------------------------------------------------
FC Value                : 0             FC Name                 : be
Redirect UniCast Plcr   : 1             Redirect MultiCast Plcr : None

---snip---
FC Value                : 5             FC Name                 : ef
Redirect UniCast Plcr   : 2             Redirect MultiCast Plcr : None
---snip---
```

The details of the configuration of the pseudowire QoS can be seen when showing the details of the SDP within the Epipe service, as follows.

```
*A:PE-1# show service id 1 sdp 1:1 detail
===============================================================================
Service Destination Point (Sdp Id : 1:1) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 1:1  -(192.0.2.2)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id            : 1:1                    Type             : Spoke
Spoke Descr     : (Not Specified)
VC Type           : VLAN                   VC Tag           : 0
Admin Path MTU    : 0                      Oper Path MTU    : 9190
Delivery          : MPLS
Far End           : 192.0.2.2
Tunnel Far End    : 192.0.2.2              LSP Types        : LDP
Hash Label        : Disabled               Hash Lbl Sig Cap : Disabled
Oper Hash Label   : Disabled

Admin State       : Up                     Oper State       : Up
```

```
---snip---
Ingress Qos Policy : 10                       Egress Qos Policy : 10
Ingress FP QGrp    : ingress-queue-group      Egress Port QGrp  : egress-queue*
Ing FP QGrp Inst   : 1                        Egr Port QGrp Inst: 1
```

The usage of the “Egr Network Queue Group Mappings” out of the total number of “Dynamic Service Entries” can be seen with the following command. Only one egress QoS pseudowire redirection has been configured.

```
*A:PE-1# tools dump system-resources
Resource Manager info at 005 m 07/31/13 13:11:03.355:

Hardware Resource Usage for Slot #7, CardType imm5-10gb-xfp, Cmplx #0:
                                |   Total   | Allocated |    Free
  ------------------------------+-----------+-----------+------------
---snip---
         Dynamic Service Entries |      65535|          1|      65534
                Subscriber Hosts |           |          0|
            Encap Group Members  |           |          0|
Egr Network Queue Group Mappings |           |          1|
```

It is possible to show the statistics on the ingress FP queue group used by the pseudowire.

```
*A:PE-1# show card 7 fp 1 ingress queue-group "ingress-queue-group" instance 1 mode
network statistics

===============================================================================
Card:7  Net.QGrp: ingress-queue-group  Instance: 1
===============================================================================
Group Name    : ingress-queue-group
Description   : (Not Specified)
Pol Ctl Pol   : None                    Acct Pol      : None
Collect Stats : disabled
-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
                        Packets                 Octets

Ing. Policer:  1  Grp: ingress-queue-group (Stats mode: minimal)
Off. All              :         184275                  23587200
Dro. All              :         36801                   4710528
For. All              :         147474                  18876672

Ing. Policer:  2  Grp: ingress-queue-group (Stats mode: minimal)
Off. All              :         184274                  23587072
Dro. All              :         85955                   11002240
For. All              :         98319                   12584832
```

Similar statistics can be shown for the egress port queue group used by the pseudowire.

```
*A:PE-1# show port 7/1/2 queue-group egress "egress-queue-group" network statistics
instance 1
-------------------------------------------------------------------------------
Ethernet port 7/1/2 Network Egress queue-group
-------------------------------------------------------------------------------
                        Packets                 Octets
Egress Queue:  1    Group: egress-queue-group    Instance-Id:  1
In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 150989                  19326592
Out Profile dropped   : 37123                   4751744

Egress Policer:  1  Group: egress-queue-group  Instance-Id: 1
Stats mode: minimal
Off. All              : 188421                  24117888
Dro. All              : 87894                   11250432
For. All              : 100527                  12867456
```

Monitor commands are available to see the statistics (and rates on egress port queue group). As an example, the following shows the utilization on the queue and policer in the egress queue-group.

```
*A:PE-1# monitor port 7/1/2 queue-group "egress-queue-group" instance 1 egress network
egress-queue 1 repeat 1 rate
===============================================================================
Monitor Port Queue-Group Egress Network Queue Statistics
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
                        Packets                 Octets

In Profile forwarded  : 0                       0
In Profile dropped    : 0                       0
Out Profile forwarded : 299113                  38286464
Out Profile dropped   : 74155                   9491840
-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
                        Packets/sec             Octets/sec              % Port
                                                                        Util.

In Profile forwarded  : 0                       0                       0.00
In Profile dropped    : 0                       0                       0.00
Out Profile forwarded : 5863                    750436                  0.06
Out Profile dropped   : 1466                    187609                  0.01
===============================================================================

*A:PE-1# monitor port 7/1/2 queue-group "egress-queue-group" instance 1 egress network
policer 1 repeat 1 rate
===============================================================================
Monitor Port Queue-Group Egress Network Policer Statistics
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
```

```
                        Packets                 Octets

Off. All              : 454750                  58208000
Dro. All              : 212181                  27159168
For. All              : 242569                  31048832

-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
                        Packets/sec             Octets/sec              % Port
                                                                        Util.

Off. All              : 7326                    937716                  0.07
Dro. All              : 3419                    437609                  0.03
For. All              : 3907                    500108                  0.04
===============================================================================
*A:PE-1#
```

As mentioned, the egress policer (FC EF) traffic exits the egress port by default using the related network queue on the port, as follows.

```
*A:PE-1# show port 7/1/2 detail
===============================================================================
Ethernet Interface
===============================================================================
Description        : 10-Gig Ethernet
Interface          : 7/1/2                   Oper Speed       : 10 Gbps
Link-level         : Ethernet                Config Speed     : N/A
Admin State        : up                      Oper Duplex      : full
Oper State         : up                      Config Duplex    : N/A
---snip---
===============================================================================
Queue Statistics
===============================================================================
-------------------------------------------------------------------------------
---snip---
Egress Queue  6              Packets                 Octets
     In Profile forwarded  :    0                      0
     In Profile dropped    :    0                      0
     Out Profile forwarded :    102381                 15357150
     Out Profile dropped   :    0                      0
```

The throughput achieved using the preceding configuration can be verified in the traffic generator output. Port 202/1 is connected to PE-1 and port 204/1 is connected to PE-2.

Port	Tx Test Packets	Rx Test Packets	Tx Test Octets	Rx Test Octets	Tx Test Throughput (Mb/s)	Rx Test Throughput (Mb/s)	Rx Packet Loss	Average Latency (us)
All Ports	29296	19531	3749888	2499968	29.999	20.000	n/a	15512.18
202/1	14648	9765	1874944	1249920	15.000	9.999	n/a	39.28
202/1->204/1, BE traffic	7324	5860	937472	750080	7.500	6.001	1464	51609.56
202/1->204/1, EF traffic	7324	3906	937472	499968	7.500	4.000	3418	39.13
204/1	14648	9766	1874944	1250048	15.000	10.000	n/a	30983.50
204/1->202/1, BE traffic	7324	5859	937472	749952	7.500	6.000	1465	39.28
204/1->202/1, EF traffic'	7324	3906	937472	499968	7.500	4.000	3418	39.27

Conclusion

This example has shown the configuration and monitoring of pseudowire QoS, providing a powerful QoS solution for pseudowire applications. QoS can be applied independently to the ingress and/or egress of a single pseudowire or multiple pseudowires.

QoS Architecture and Basic Operation

This chapter provides information about QoS architecture and basic operation.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information in this chapter is applicable to all of the Nokia 7x50 platforms and assumes only FP2-and higher-based line cards are used. The chapter was initially written for SR OS release 9.0.R3. The CLI in the current edition corresponds to SROS release 14.0.R2.

Overview

The 7x50 platforms provide an extensive Quality of Service (QoS) capability for service provider solutions. QoS is a system behavior to provide different traffic with different amounts of resources, including buffer memory and queue serving time.

By allocating system resources with certain degrees of guarantee, the bandwidth can be used more efficiently and more controllably. Lack of buffer memory leads to packet drop, while a smaller amount of queue serving time normally means longer delay for the packet and may cause buffer memory to be completely consumed and eventually also lead to packet drop.

In a system, such as the 7x50 platform, different types of traffic contend for the same resources at several major points, such as the ingress to the switch fabric and the egress out of a physical port. In a multi-node network, QoS is achieved on hop by hop basis. Thus, QoS needs to be configured individually but with the consistency across the whole network.

This chapter is focused on the configuration of the basic QoS, namely the use of queues to shape traffic at the ingress and egress of the system. More sophisticated aspects will be referenced where appropriate but their details are beyond the scope of this chapter. Other topics not included are Hierarchical QoS scheduling, egress port-scheduler, queue-groups, named buffer pools, WRED-per-queue, LAGs, high scale MDA, QoS for ATM/FR and Enhanced Subscriber Management.

QoS Components

QoS consists of four main components:

- Classification
- Buffering (enqueuing)
- Scheduling (dequeuing)
- Remarking

These are also the fundamental building blocks of the QoS implementation in the 7x50. Ingress packets, classified by various rules, belong to one of eight Forwarding Classes (FC). An FC can be viewed a set of packets which are treated in a similar manner within the system (have the same priority level and scheduling behavior). Each FC has a queue associated with it and each queue has a set of parameters controlling how buffer memory is allocated for packets; if a packet is enqueued (placed on the queue) a scheduler will control the way the packet gets dequeued (removed from the queue) relative to other queues. When a packet exits an egress port, a remarking action can be taken to guarantee the next downstream device will properly handle the different types of traffic.

Configuration

Policies

QoS policies are used to control how traffic is handled at distinct points in the service delivery model within the device. There are different types of QoS policies catering to the different QoS needs at each point. QoS policies only take effect when applied to a relevant entity (Service Access Point (SAP) or network port/interface) so by default can be seen as templates with each application instantiating a new set of related resources.

The following QoS policies are discussed:

- Ingress/egress QoS Policies — For classification, queue attributes and remarking.
- Slope policies — Define the RED slope definitions.
- Scheduler policies — Determine how queues are scheduled (only the default scheduling is included here).

Access, Network, and Hybrid Ports

The system has two different types of interfaces: access and network.

- A network interface will classify packets received from the network core at ingress and remark packets sent to the core at egress. Aggregated differentiated service QoS is performed on network ports, aggregating traffic from multiple services into a set of queues per FC.
- An access interface connects to one or more customer devices; it receives customer packets, classifies them into different FCs at ingress and remarks packets according to FCs at egress. Since an access interface needs application awareness, it has many more rules to classify the ingress packets. Access and network also differ in how buffer memory is handled, as will be made clear when discussing the buffer management. Here the QoS is performed per SAP.

Access interfaces (SAPs) are configured on access ports and network interfaces are configured on network ports. A third type of port is available, the hybrid port, which supports both access and network interfaces on the same port.

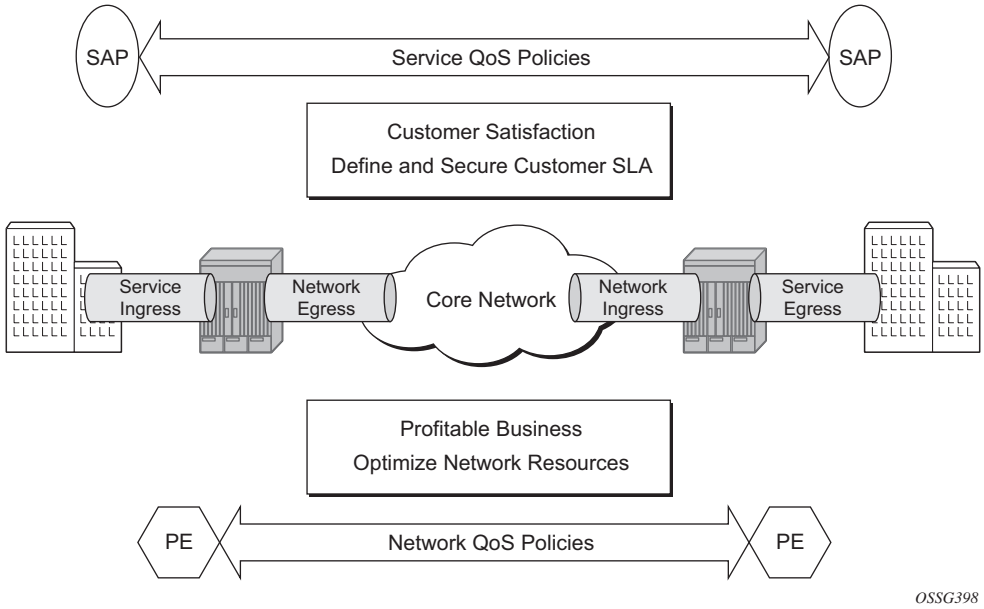
Hybrid ports are only supported on Ethernet ports and optionally with a single-chassis LAG. They must be configured to use VLANs (either single (dot1q encapsulation) or double (QinQ encapsulation) tagging) with each VLAN mapping to either the access or network part of the port. This allows the classification to associate incoming traffic with the correct port type and service. Port based traffic, such as LACP, CCM and EFM, uses a system queue on an access port, but the default network queues on a network or hybrid port.

Customer traffic follows the path shown below:

[service ingress ➔ network egress]	➔	[network ingress ➔ network egress]	➔	[network ingress ➔ service egress]
ingress PE		transit P		egress PE

The network administrator needs to make sure that QoS is configured correctly at each point using the appropriate QoS policies (Figure 383).

Figure 383 **Service and Network QoS Policies**



Service Ingress QoS Policy

The SAP ingress policies are created in the *qos* context of the CLI and require a unique identifier (from 1 to 65535). The default *sap-ingress* policy has identifier 1.

Classification

Services can be delineated at the SAP ingress by

- A physical port (null encapsulated) or
- An encapsulation on the physical port, for example a VLAN ID on an Ethernet port or a DLCI on a Frame Relay port.

The following configuration is an example of an IES service created with an IP interface on VLAN 2 of port 3/2/10 (IOM 3, MDA 2, port 10) and has SAP ingress QoS policy 10 applied.

```
configure
```

```
service
    ies 1 customer 1 create
        interface "int-access" create
            address 192.168.1.1/30
            sap 1/1/1:1 create
                ingress
                    qos 10
                exit
            exit
        exit
        no shutdown
    exit
```

As traffic enters the port, the service can be identified by the VLAN tag (and unwanted packets dropped). The ingress service QoS policy applied to the SAP maps traffic to FCs, and thus to queues, and sets the enqueuing priority. Mapping flows to FCs is controlled by comparing each packet to the match criteria in the QoS policy. The match criteria that can be used in ingress QoS policies can be combinations of those listed in Table 32. When a packet matches two criteria (802.1p priority and DSCP) it is the lowest precedence value that is used to map the packet to the FC.

Table 32 **SAP Ingress Classification Match Criteria**

Match Precedence	Match Criteria		
1	IPv4 fields match criteria: • Destination IP address/prefix including prefix list • Destination port/range • DSCP value • IP fragment • Protocol type (TCP, UDP, etc.) • Source port/range • Source IP address/prefix including prefix list	IPv6 fields match criteria: • Destination IP address/prefix • Destination port/range • DSCP value • Next header • Source port/range • Source IP address/prefix	MAC fields match criteria: • Frame type [802dot3|802dot2-llc|802dot2-snap|ethernetII|atm] • ATM VCI value • IEEE 802.1p value/mask • Source MAC address/mask • Destination MAC address/mask • EtherType value • IEEE 802.2 LLC SSAP value/mask • IEEE 802.2 LLC DSAP value/mask • IEEE 802.3 LLC SNAP OUI zero or non-zero value • IEEE 802.3 LLC SNAP PID value
	Note: For an ingress QoS policy, either IP match criteria or MAC match criteria can be defined, not both.		
2	DSCP		

Table 32 **SAP Ingress Classification Match Criteria**

Match Precedence	Match Criteria
3	IP precedence
4	LSP EXP
5	IEEE 802.1p priority and/or Drop Eligibility Indicator (DEI)
6	Default forwarding class for non-matching traffic

It is possible to match MAC criteria on VPLS/Epipe SAPs and IP criteria on IP interface SAPs. However, it is also possible to classify on MAC criteria on an IP interface SAP and conversely to classify on IP criteria on VPLS/Epipe SAPs. When MPLS labeled traffic is received on a VPLS/Epipe SAP, it is possible to match on either of the LSP EXP bits (outer label) or the MAC criteria.

A SAP can be configured to have no VLAN tag (null encapsulated), one VLAN tag (dot1q encapsulated) or two VLAN tags (QinQ encapsulated). The configuration allows the selection of which VLAN tag to match against for the 802.1p bits, using the keyword **match-qinq-dot1p** with the keyword **top** or **bottom**.

The following example configuration shows match QinQ traffic with dot1p value 1 in the top VLAN tag entering the QinQ SAP in Epipe service 1 and assign it to FC **af** using queue 2.

```
configure
    qos
        sap-ingress 10 create
            queue 2 create
            exit
            fc "af" create
                queue 2
            exit
            dot1p 1 fc "af"
        exit


configure
    service
        epipe 2 customer 1 create
            sap 1/1/2:1.2 create
                ingress
                    qos 10
                exit
                ingress
                    match-qinq-dot1p top
                exit
            exit
            no shutdown
        exit
```

The classification of traffic using the default, **top** and **bottom** keyword parameters is summarized in Table 33. A TopQ SAP is a QinQ SAP where only the outer (top) VLAN tag is explicitly specified (sap 1/1/1:10.* or sap 1/1/1:10.0).

Table 33 **QinQ Dot1p Bit Classification**

Port/SAP Type	Existing Packet Tags	Pbits Used for Match		
		Default	Match Top	Match Bottom
Null	None	None	None	None
Null	Dot1P (VLAN-ID 0)	Dot1P PBits	Dot1P PBits	Dot1P PBits
Null	Dot1Q	Dot1Q PBits	Dot1Q PBits	Dot1Q PBits
Null	TopQ BottomQ	TopQ PBits	TopQ PBits	BottomQ PBits
Null	TopQ (No BottomQ)	TopQ PBits	TopQ PBits	TopQ PBits
Dot1Q	None (Default SAP)	None	None	None
Dot1Q	Dot1P (Default SAP VLAN-ID 0)	Dot1P PBits	Dot1P PBits	Dot1P PBits
Dot1Q	Dot1Q	Dot1Q PBits	Dot1Q PBits	Dot1Q PBits
QinQ/TopQ	TopQ	TopQ PBits	TopQ PBits	TopQ PBits
QinQ/TopQ	TopQ BottomQ	TopQ PBits	TopQ PBits	BottomQ PBits
QinQ/QinQ	TopQ BottomQ	BottomQ PBits	TopQ PBits	BottomQ Pbits

The Drop Eligibility Indicator (DEI) bit (IEEE 802.1ad-2005 and IEEE 802.1ah (PBB)) can be used to indicate the in/out profile state of the packet, this will be covered later in the discussion on profile mode.

Ingress traffic with a local destination (for example, OSPF hellos) is classified by the system automatically and uses a set of dedicated system queues.

After the traffic has been classified, the next step is to assign it to a given FC. There are 8 pre-defined FCs within the system which are shown in Table 34 (the FC identifiers are keywords and do not have a fixed relationfship with the associated Differentiated Services Code Points (DSCP)).

Table 34 **Forwarding Classes**

FC Identifier	FC Name	Default Scheduling Priority
NC	Network Control	Expedited

Table 34 **Forwarding Classes**

FC Identifier	FC Name	Default Scheduling Priority
H1	High-1	Expedited
EF	Expedited	Expedited
H2	High-2	Expedited
L1	Low-1	Best Effort
AF	Assured	Best Effort
L2	Low-2	Best Effort
BE	Best Effort	Best Effort

When an FC is configured for a classification, it must first be created in the configuration. One of the FCs can be also configured to be the default in case there is no explicit classification match and by default this FC is **be**.

Normally, once traffic is assigned to an FC at the ingress it remains in that FC throughout its time within the system. Re-classification of IP traffic at a SAP egress is possible, but is beyond the scope of this chapter. The FC used at egress can also be specified to be different than that used at ingress by configuring **egress-fc** under the FC configuration in the SAP ingress policy.

Packets also have a state of being in-profile or out-of-profile which represents their drop precedence within the system, therefore there can be up to 8 distinct per hop behavior (PHB) classes with two drop precedences.

Buffering (Enqueuing)

Once a packet is assigned to a certain forwarding class, it will try to get a buffer in order to be enqueued. Whether the packet can get a buffer is determined by the instantaneous buffer utilization and several attributes of the queue (such as Maximum Burst Size (MBS), Committed Burst Size (CBS) and high-prio-only) that will be discussed in more detail later in this chapter. If a packet cannot get a buffer for whatever reason, the packet will get dropped immediately.

As traffic is classified at the SAP ingress it is also assigned an enqueuing priority, which can be high or low. This governs the likelihood of a packet being accepted into a buffer and so onto a queue, and is managed using the queue's high-prio-only parameter and the buffer pools weighted random early detection (WRED) slope policies. Traffic having a high enqueuing priority has more chance of getting a buffer than traffic with low enqueuing priority. The enqueuing priority is specified with the classification using the parameter *priority*, and a default enqueuing priority can be configured, its default being low.

Enqueuing priority is a property of a packet and should not to be confused with scheduling priority, expedited or best-effort, which is a property of a queue.

The following configuration shows an example where all packets with dot1p value 3 are classified as **ef** and have their enqueuing priority set to high, while all other packets are classified as **af** with a low enqueuing priority.

```
configure
    qos
        sap-ingress 10 create
            fc "af" create
            exit
            fc "ef" create
            exit
            dot1p 3 fc "ef" priority high
            default-fc "af"
            default-priority low # this is the default
        exit
```

Each forwarding class is associated with at most one unicast queue. In the case of a VPLS service, each FC can also be assigned a single multipoint queue at ingress, or for more granular control, separate queues for broadcast, multicast and unknown traffic. Since each queue maintains forward/drop statistics, it allows the network operator to easily track unicast, broadcast, multicast and unknown traffic load per forwarding class. Separate multicast queues can also be assigned for IES/VPRN services which have IP multicast enabled.

This results in an Epipe SAP having up to 8 ingress queues, an IES/VPRN SAP having up to 16 ingress queues and a VPLS SAP having up to 32 ingress queues. Each queue has a locally significant (to the policy) identifier, which can be from 1 to 32.

The default SAP ingress QoS policy (id=1) has two queues; queue 1 for unicast traffic and queue 11 for multipoint traffic, and is assigned to every ingress SAP at service creation time. Equally, when a new (non-default) SAP ingress policy is created, queue 1 and queue 11 are automatically created with all FCs assigned to both by default. Additional queues must be created before being assigned to a FC, with

multipoint queues requiring the **multipoint** keyword. When a SAP ingress policy is applied to a SAP, physical hardware queues on the FP are allocated for each queue with a FC assigned (if no QoS policy is explicitly configured, the default policy is applied). Multipoint queues within the SAP ingress policy are ignored when applied to an Epipe SAP or an IES/VPRN SAP which is not configured for IP multicast.

The mechanism described here uses a separate set of queues per SAP. For cases where per-SAP queuing is not required it is possible to use port based queues, known as *queue-groups*, which reduces the number of queues required, as described in chapter *FP and Port Queue Groups*.

Scheduling (Dequeuing)

A queue has a priority which affects the relative scheduling of packets from it compared to other queues. There are two queue priorities: expedited and best-effort, with expedited being the higher. When creating a queue, one of these priorities can be configured, thereby explicitly setting the queue's priority. Alternatively, the default is auto-expedite in which case the queue's priority is governed by the FCs assigned to it, as shown in Table 34. If there is a mix of expedited and best-effort FCs assigned, the queue is deemed to be best-effort.

The following configuration displays an example that ensures that EF traffic is treated as expedited by assigning it to new unicast and multicast queues.

```
configure
    qos
        sap-ingress 10 create
            queue 3 expedite create
            exit
            queue 13 multipoint expedite create
            exit
            fc "ef" create
                queue 3
                multicast-queue 13
            exit
            default-fc "ef"
        exit
```

Once a packet gets a buffer and is queued, it will wait to be served and sent through the switch fabric to its destination port by the hardware scheduler. There are two scheduler priorities: expedited or best-effort, corresponding to the queue's priority. The expedited hardware schedulers are used to enforce priority access to internal switch fabric destinations with expedited queues having a higher preference than best-effort queues. Queues of the same priority get equally serviced in round robin fashion by the associated scheduler.

When a queue gets its turn to be serviced, the scheduler will use the operational Peak Information Rate (PIR) and Committed Information Rate (CIR) attributes of the queue to determine what to do with the packet.

- The scheduler does not allow queues to exceed their configured PIR. If the packet arrival rate for a given queue is higher than the rate at which it is drained, the queue will fill. If the queue size (in bytes or Kbytes) reaches its defined MBS all subsequent packets will be discarded, this is known as tail drop.
- If the dequeue rate is below or equal to the operational CIR, the packet will be forwarded and marked as **in-profile**.
- If the dequeue rate is below or equal to the operational PIR but higher than the CIR, the packet will be forwarded but marked as **out-of-profile**.

Out-of-profile packets have a higher probability of being dropped when there is congestion somewhere in the downstream path. Packets that are marked as out-of-profile will also be treated differently at the network egress and service egress.

These marking actions are known as color marking (green for in-profile and yellow for out-of-profile). Using the default queue setting of **priority-mode**, as described above, the in/out-of-profile state of a packet is determined from the queue scheduling state (within CIR or above CIR, as described later) at the time that the packet is dequeued. An alternative queue mode is **profile-mode**.

Profile Mode

A queue is created with profile mode when the aim is that the in/out-of-profile state of packets is determined by the QoS bits of the incoming packets, this is known as color-aware (as opposed to color-unaware for priority mode).

As part of the classification, the profile state of the packets is explicitly configured. To provide granular control, it is possible to configure FC sub-classes with each having a different profile state, while inheriting the other parameters from their parent FC (for example the queue, in order to avoid out of order packets). The FC subclasses are named *fc.sub-class*, where *sub-class* is a text string up to 29 characters (though normally the words **in** and **out** are used for clarity). Any traffic classified without an explicit profile state is treated as if the queue were in priority mode.

When using the profile mode, the DEI in the Ethernet header can be used to classify a packet as in-profile (DEI=0) or out-of-profile (DEI=1).

The following configuration shows traffic with dot1p 3 is set to in-profile, dot1p 2 to out-of-profile and the profile state of dot1p 0 depends on the scheduling state of the queue.

```
configure
    qos
        sap-ingress 20 create
            queue 2 profile-mode create
            exit
            fc "af" create
                queue 2
            exit
            fc "af.in" create
                profile in
            exit
            fc "af.out" create
                profile out
            exit
            dot1p 0 fc "af"
            dot1p 2 fc "af.out"
            dot1p 3 fc "af.in"
        exit
```

The difference between a queue configured in priority (default) and profile mode is summarized in Table 35 (within/above CIR is described later).

Table 35 **Queue Priority vs. Profile Mode**

	Priority Mode	Profile Mode
Packet In-Profile/ Out-of-Profile state	Determined by state of the queue at scheduling time. Within CIR – In Profile Above CIR – Out Profile	Explicitly stated in FC or subclass classification. If not, then defaults to state of the queue at scheduling time
Packet High/Low Enqueuing Priority	Explicitly stated in FC classification. If not, then defaults to Low priority	Always follows state of in-profile/out-of-profile determined above In-profile = High Priority Out-Profile = Low Priority If not set = High Priority

Remarking

Remarking at the service ingress is possible when using an IES or VPRN service. The DSCP/precedence field can be remarked for in-profile (**in-remark**) and out-of-profile (**out-remark**) traffic as defined above for queues in either priority mode or profile mode. If configured for other services, the remarking is ignored. If remarking is performed at the service ingress, then the traffic is not subject to any egress remarking on the same system.

The following configuration displays an example classifying traffic to 10.0.0.0/8 as FC **ef** in-profile and remark its DSCP to **ef**.

```
configure
    qos
        sap-ingress 30 create
            queue 2 profile-mode create
            exit
            fc "ef" create
                queue 2
                in-remark dscp ef
                profile in
            exit
            ip-criteria
                entry 10 create
                    match
                        dst-ip 10.0.0.0/8
                    exit
                    action fc "ef"
                exit
            exit
        exit
```

Service Egress QoS Policy

The service egress uses a SAP egress QoS policy to define how FCs map to queues and how a packet of an FC is remarked. SAP egress policies are created in the CLI qos context and require a unique identifier (from 1 to 65535). The default SAP egress policy has identifier 1.

Once a service packet is delivered to the egress SAP, it has following attributes:

- Forwarding class, determined from classification at the ingress of the node.
- In/out-of-profile state from the service ingress or network ingress.

Similar to the service ingress enqueuing process, it is possible that a packet cannot get a buffer and thus gets dropped. Once on an egress queue, a packet is scheduled from the queue based on priority of the queue (expedited or best-effort) and the scheduling state with respect to the CIR/PIR rates (the profile state of the packet [in/out] is not modified here). Egress queues do not have a priority/profile mode and have no concept of multipoint.

Only one queue exists in the default SAP egress QoS policy (id=1) and also when a new *sap-egress* policy is created, this being queue 1 which is used for both unicast traffic and multipoint traffic. All FCs are assigned to this queue unless otherwise explicitly configured to a different configured queue. When a SAP egress policy is applied to a SAP, physical hardware queues on the FP are allocated for each queue with FC assigned (if no QoS policy is explicitly configured, the default policy is applied).

As mentioned earlier, re-classification at a SAP egress is possible based on the packet's dot1p, DSCP or precedence values or using IP or IPv6 criteria matching, similar to the functionality at SAP ingress.

SAP egress also supports two additional profiles, inplus-profile and exceed-profile. Both can be assigned to a packet using egress reclassification and the exceed-profile can be assigned to a packet in an egress policer configured with the **enable-exceed-pir** command.

Traffic originated by the system (known as self generated traffic) has its FC and marking configured under router/sgt-qos (for the base routing) or under service/vprn/sgt-qos (for a VPRN service). This is beyond the scope of this chapter.

Remarking

At the service egress, the dot1p/DEI can be remarked for any service per FC with separate marking for in/out/exceed profile if required (inplus-profile packets are marked with the same value as in-profile packets and exceed-profile packets are marked with the same value as out-of-profile packets if not explicitly configured). The DEI bit can also be forced to a specific value (using the **de-mark force** command). When no **dot1p/de-mark** is configured, the ingress dot1p/DEI is preserved; if the ingress was untagged, the dot1p/DEI bit is set to 0.

The following configuration shows a remark example with different FCs with different dot1p values. FC **af** also differentiates between in/out-of-profile and then remarks the DEI bit accordingly based on the packet's profile.

```
configure
    qos
        sap-egress 10 create
            queue 1 create
                rate 20000
            exit
            queue 2 create
                rate 10000 cir 5000
            exit
            queue 3 create
                rate 2000 cir 2000
            exit
            fc af create
                queue 2
                dot1p in-profile 3 out-profile 2
                de-mark
            exit
            fc be create
                queue 1
                dot1p 0
            exit
            fc ef create
```

```
                queue 3
                dot1p 5
            exit
        exit
```

If QinQ encapsulation is used, the default is to remark both tags in the same way. However, it is also possible to remark only the top tag using the **qinq-mark-top-only** parameter configured under the SAP egress.

The following configuration shows a remark example with only the dot1p/DEI bits in top tag of a QinQ SAP.

```
configure
    service
        vpls 3 customer 1 create
            sap 1/1/2:2.2 create
                egress
                    qos 20
                    qinq-mark-top-only
                exit
            exit
            no shutdown
        exit
```

For IES and VPRN services, the DSCP/precedence field can also be remarked based on the in/out/exceed-profile state (with inplus-profile packets marked with the same values as in-profile packets, and exceed-profile packets marked with the same value as out-of-profile packets if not explicitly configured) of the packets (and only if no ingress remarking was performed).

The following configuration shows DSCP values for FC **af** based on in/out-of-profile traffic.

```
configure
    qos
        sap-egress 20 create
            queue 2 create
            exit
            fc af create
                queue 2
                dscp in-profile af41 out-profile af43
            exit
        exit
```

Network Ports

The QoS policies relating to the network ports are divided into a network and a network-queue policy. The network policy covers the ingress and egress classification into FCs and the egress remarking based on FCs, while the network-queue policy covers the queues parameters and the FC to queue mapping. The logic behind this is that there is only one set of queues provisioned on a network port, whereas the use of these queues is configured per network IP interface. This in turn determines where the two policies can be applied. Network ports are used for IP routing and switching, and for GRE/MPLS tunneling.

Network QoS Policy

The network QoS policy has an ingress section and an egress section. It is created in the *qos* context of the CLI and requires a unique identifier (from 1 to 65535). The default network policy has identifier 1. Network QoS policies are applied to IP interfaces configured on a network port.

The following configuration shows an example to apply different network QoS policies to two network interfaces.

```
configure
    router
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/3:1
            qos 28
        exit
        interface "int-PE-1-PE-3"
            address 192.168.13.1/30
            port 1/1/4
            qos 18
        exit
```

Classification

Classification is available at both ingress and egress.

The ingress section defines the classification rules for IP/MPLS packets received on a network IP interface. The rules for classifying traffic are based on the incoming QoS bits (Dot1p, DSCP, EXP [MPLS experimental bits]). The order in which classification occurs relative to these fields is:

1. IPv4 and IPv6 match criteria for IP packets

2. EXP (for MPLS packets) or DSCP (for IP packets)
   Dot1p/DEI bit (network ports do not support QinQ encapsulation)
3. default action (default = fc be profile out)

The configuration specifies the QoS bits to match against the incoming traffic together with the FC and profile (in/out) to be used (it is analogous to the SAP profile-mode in that the profile of the traffic is determined from the incoming traffic, rather than the CIR configured on the queue). A **default-action** keyword configures a default FC and profile state.

The IPv4 and IPv6 criteria matching only applies to the outer IP header of *non-tunneled* traffic, except for traffic received on an RFC 6037 MVPN tunnel for which classification on the outer IP header only is supported, and is only supported on network interfaces.

For tunneled traffic (GRE or MPLS), the match is based on the outer encapsulation header unless the keyword **ler-use-dscp** is configured. In this case, traffic received on the router terminating the tunnel that is to be routed to the base router or a VPRN destination is classified based on the encapsulated packet DSCP value (assuming it is an IP packet) rather than its EXP bits.

The ability exists for an egress LER to signal an implicit-null label (numeric value 3). This informs the previous hop to send MPLS packets without an outer label and so is known as penultimate hop popping (PHP). This can result in MPLS traffic being received at the termination of an LSP without any MPLS labels. In general, this would only be the case for IP encapsulated traffic, in which case the egress LER would need to classify the incoming traffic using IP criteria.

The egress section also defines the classification rules based on both the DSCP and precedence values in a packet to re-assign the packet's FC and profile (inplus/in/out/exceed).

Remarking

The egress section of the network policy defines the remarking of the egress packets, there is no remarking possible at the network ingress. The egress remarking is configured per FC and can set the related dot1p/DEI (explicitly or dependent on in/out-of-profile), DSCP (dependent on in/out-of-profile) and EXP (dependent on in/out-of-profile; inplus-profile packets are marked with the same values as in-profile packets and exceed-profile packets are marked with the same value as out-of-profile packets).

The traffic exiting a network port is either tunneled (in GRE or MPLS) or IP routed.

For tunneled traffic exiting a network port, the remarking applies to the DSCP/EXP bits in any tunnel encapsulation headers (GRE/MPLS) pushed onto the packet by this system, together with the associated dot1p/DEI bits if the traffic has an outer VLAN tag. For MPLS tunnels, the EXP bits in the entire label stack are remarked.

Note: Strictly speaking this is marking (as opposed to remarking) as the action is adding QoS information rather than changing it.

A new outer encapsulation header is pushed onto traffic at each MPLS transit label switched router as part of the label swap operation.

For VPLS/Epipe services no additional remarking is possible. However, for IES/VPRN/base-routing traffic, the remarking capabilities at the network egress are different at the first network egress (egress on the system on which the traffic entered by a SAP ingress) and subsequent network egress in the network (egress on the systems on which the traffic entered through another network interface).

At the first network egress, the DSCP of the routed/tunneled IP packet can be remarked, but this is dependent on two configuration settings:

- The trusted state of the ingress (service/network) interface and
- The **remarking** keyword in the network QoS policy at the network egress. The configuration combinations are summarized in Table 36.

This is in addition to the remarking of any encapsulation headers and, as stated earlier, is not performed if the traffic was remarked at the service ingress.

For traffic exiting a subsequent network egress in the network, only the IP routed traffic can be remarked, again this is dependent on the ingress trusted state and egress remarking parameter.

There is one addition to the above to handle the marking for IP-VPN Option-B in order to remark the EXP, DSCP and dot1p/DEI bits at a network egress, this being **remarking force**. Without this, only the EXP and dot1p/DEI bits are remarked. This does not apply to label switched path traffic switched at a label switched router.

Table 36 **Network QoS Policy DSCP Remarking**

Ingress	Trusted State	Remarking Configuration	Marking Performed
IES	Untrusted (default)	remarking	Yes
		no remarking (default)	Yes
	Trusted	remarking	Yes
		no remarking (default)	No

Table 36 **Network QoS Policy DSCP Remarking (Continued)**

Ingress	Trusted State	Remarking Configuration	Marking Performed
Network	Untrusted	remarking	Yes
		no remarking (default)	Yes
	Trusted (default)	remarking	Yes
		no remarking (default)	No
VPRN	Untrusted	remarking	Yes
		no remarking (default)	Yes
	Trusted (default)	remarking	No
		no remarking (default)	No

The following configuration shows a ingress network classification for DSCP EF explicitly, with a default action for the remainder of the traffic and use the DSCP from the encapsulated IP packet if terminating a tunnel. Remark the DSCP values for FC **af** and **ef** and remark all traffic (except incoming VPRN traffic) at the egress. Apply this policy to a network interface.

```
configure
    qos
        network 20 create
            ingress
                default-action fc af profile out
                ler-use-dscp
                dscp ef fc ef profile in
            exit
            egress
                remarking
                fc af
                    no dscp-in-profile
                    dscp-out-profile af13
                    lsp-exp-in-profile 6
                    lsp-exp-out-profile 5
                exit
                fc ef
                    dscp-in-profile af41
                exit
            exit
        exit

configure
    router
        interface "int-PE-1-PE-4"
            address 192.168.14.1/30
            port 1/2/1
            qos 20
        exit
```

The following configuration shows the trusted IES interface.

```
configure
    service
        ies 1 customer 1 create
            interface "int-access" create
                address 192.168.1.1/30
                tos-marking-state trusted
                sap 1/1/1:1 create
                exit
            exit
            no shutdown
        exit
```

The network QoS ingress and egress sections also contain the configuration for the use of FP-based policers and port-based queues by queue-groups which are out of scope of this chapter.

Network Queue Policy

The network queue QoS policy defines the queues and their parameters together with the FC to queue mapping. The policies are named, with the default policy having the name **default**. Network queues policies are applied under **config > card > mda > network > ingress** for the network ingress queues though only one policy is supported per MDA, so when a new policy is applied under one MDA, it is automatically applied under the other MDA on the same FP. At egress, network queue policies are applied under Ethernet: **config > port > ethernet > network**, POS: **config > port > sonet-sdh > path > network**, TDM: **config > port > tdm > e3 | ds3 > network** for the egress.

The following configuration shows an ingress and egress network-queue policy.

```
configure
    card 1
        card-type iom3-xp
        mda 1
            mda-type m4-10gb-xp-xfp
            network
                ingress
                    queue-policy "network-queue-1"
                exit
            exit
            no shutdown
        exit

configure
    port 1/1/3
        ethernet
            encap-type dot1q
            network
```

```
                queue-policy "network-queue-1"
            exit
        exit
        no shutdown
    exit
```

Up to 16 queues can be configured in a network-queue policy, each with a queue-type of best-effort, expedite, or auto-expedite. A new network-queue policy contains two queues, queue 1 for unicast traffic and queue 9 for multipoint traffic and by default all FCs are mapped to these queues. There is no differentiation for broadcast, multicast and unknown traffic. If the policy is applied to the egress, then any multipoint queues are ignored. As there are 8 FCs, there would be up to 8 unicast queues and 8 multipoint queues, resulting in 16 ingress queues and 8 egress queues. Normally, the network queue configuration is symmetric (the same queues/FC-mapping at the ingress and egress).

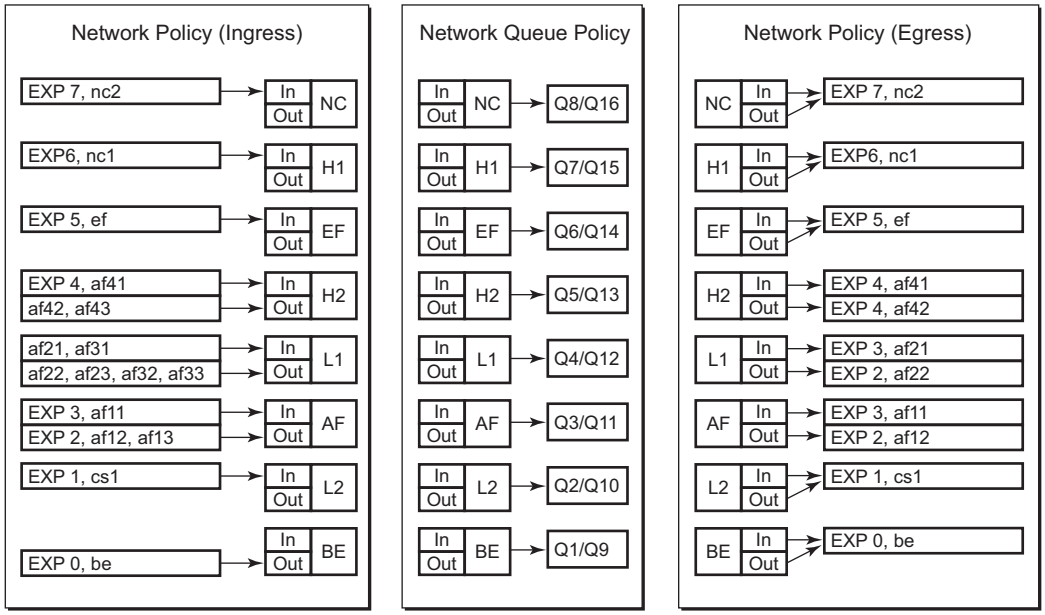
The following configuration defines a network-queue policy with FC **af** and **ef** assigned to queues 2 and 3 for unicast traffic, and queue 9 for multipoint traffic.

```
configure qos
        network-queue "network-queue-1" create
            queue 1 create
                mbs 50
                high-prio-only 10
            exit
            queue 2 create
            exit
            queue 3 create
            exit
            queue 9 multipoint create
                mbs 50
                high-prio-only 10
            exit
            fc af create
                multicast-queue 9
                queue 2
            exit
            fc ef create
                multicast-queue 9
                queue 3
            exit
        exit
```

Summary of Network Policies

Figure 384 displays the default network policies with respect to classification, FC to queue mapping and remarking.

Figure 384 **Visualization of Default Network Policies**



Queue Management

The policies described so far define queues but not the characteristics of those queues which determine how they behave. This section describes the detailed configuration associated with these queues. There are two aspects:

- Enqueuing packets onto a queue
  - buffer pools
  - queue sizing
  - Weighted Random Early Detection (WRED)
- Dequeuing packets from a queue
  - queue rates
  - scheduling

Enqueuing Packets: Buffer Pools

The packet buffer space is divided equally between ingress and egress. For line cards using a 50G FP2 for both ingress and egress traffic, the proportion of ingress versus egress buffer space can be modified using the following command:

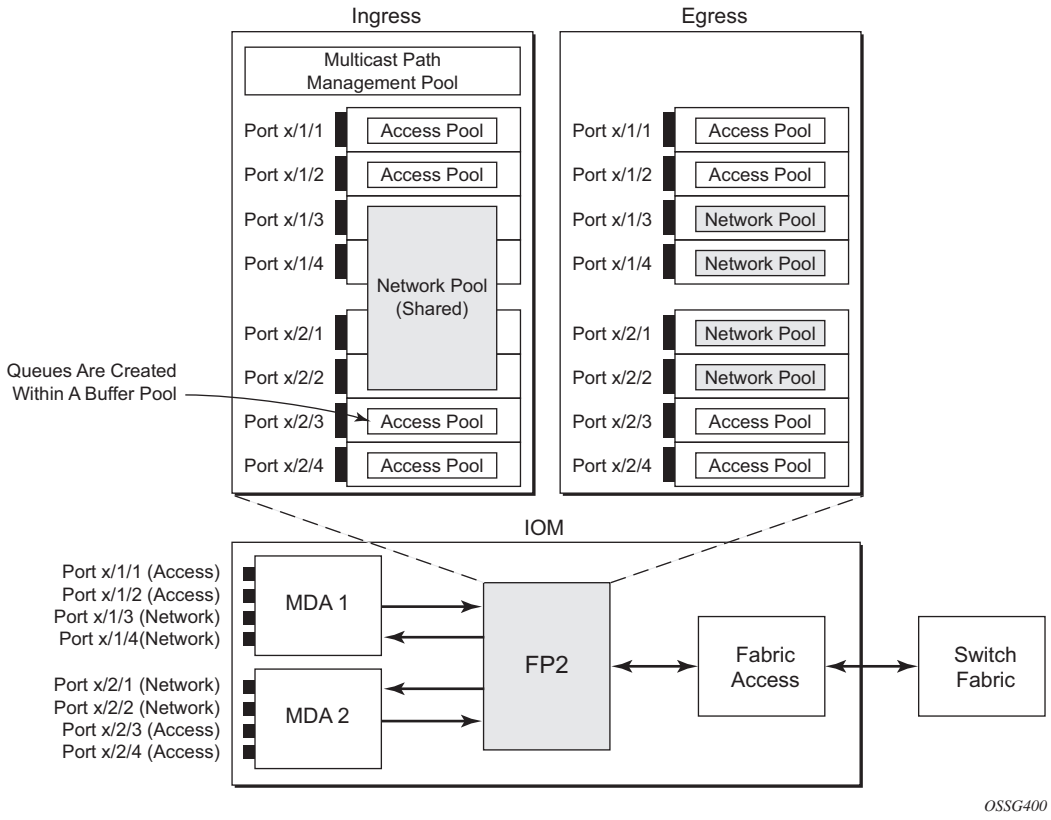
```
configure
    card <slot-number>
        ingress-buffer-allocation <percentage>
```

The ingress buffer allocation percentage can be configured from 20% up to 80%.

Beyond that, by default there is one pool for network ingress per FP2/IOM, with one pool per access ingress port and one pool per access/network egress port. (The FP2 chipset is used in the IOM3-XP/IMM and in the 7750 SR-c 12/4.) This is shown in Figure 385. This segregation provides isolation against buffer starvation between the separate pools. An additional ingress pool exists for managed multicast traffic (the multicast path management pool) but this is beyond the scope of this chapter.

The buffer management can be modified using named buffer pools and/or WRED-per-queue pools which are out of scope of this chapter.

Figure 385 **Default Buffer Pools**



The size of the pools is based on the MDA type and the speed/type (access or network) of each port. Buffer space is allocated in proportion to the active bandwidth of each port, which is dependent on:

- The actual speed of the port
- Bandwidth for configured channels only (on channelized cards)
- Zero for ports without queues configured

This calculation can be tuned separately for ingress and egress, without modifying the actual port speed, using the port/modify-buffer-allocation-rate. Changing the port's ingress-rate or egress-rate will also modify its buffer sizes.

The following configuration changes the relative size for the ingress/egress buffer space on port 1/1/10 to 50% of the default.

```
configure
    port 1/1/3
        modify-buffer-allocation-rate
            ing-percentage-of-rate 50
            egr-percentage-of-rate 50
        exit
```

Each of the buffer pools created is further divided into a section of reserved buffers and another of shared buffers, see Figure 387. The amount of reserved buffers is calculated differently for network and access pools. For network pools, the default is approximately the sum of the CBS (committed burst size) values defined for all of the queues within the pool. The reserved buffer size can also be statically configured to a percentage of the full pool size (ingress: **config > card > mda > network > ingress > pool**; egress: **config > port > network > egress > pool**). For access pools, the default reserved buffer size is 30% of the full pool size and can be set statically to an explicit value (ingress: **config > port > access > ingress > pool**; egress: **config > port > access > egress > pool**).

The following configuration sets the reserved buffer size to 50% of the egress pool space on a network port.

```
configure
    port 1/1/3
        network
            egress
                pool
                    resv-cbs 50
                exit
            exit
        exit
```

On an access port, the reserved buffer size is set to 50% of the egress pool space, as follows:

```
configure
    port 1/1/1
        access
            egress
                pool
                    resv-cbs 50
                exit
            exit
        exit
```

Both the total buffer and the reserved buffer sizes are allocated in blocks (discrete values of Kbytes). The pool sizes can be seen using the **show pools** command.

It is possible to configure alarms to be triggered when the usage of the reserved buffers in the buffer pools reaches a certain percentage. Two alarm percentages are configurable, amber and red, **amber-alarm-threshold** <*percentage*> and **red-alarm-threshold** <*percentage*>. The percentage range is 1 — 1000.

- The percentage for the red must be at least as large as that for the amber.
- The alarms are cleared when the reserved CBS drops below the related threshold.
- When the amber alarm is enabled, dynamic reserved buffer sizing can be used; after the amber alarm is triggered the reserved buffer size is increased or decreased depending on the CBS usage. This requires a non-default resv-cbs to be configured together with a step and max value for the amber-alarm-action parameters. As the reserved CBS usage increases above the amber alarm percentage, the reserved buffer size is increased in increments defined by the step, up to a maximum of the max. If the CBS usage decreases, the reserved buffer size is reduced in steps down to its configured size.
- As the reserved buffer size changes, alarms will continue to be triggered at the same color (amber or red) indicating the new reserved buffer size. The pool sizing is checked at intervals, so it can take up to one minute for the alarms and pool re-sizing to occur.

The following displays a configuration for access ingress and egress pools.

```
configure
    port 1/1/1
        access
            ingress
                pool
                    amber-alarm-threshold 25
                    red-alarm-threshold 50
                    resv-cbs 20 amber-alarm-action step 5 max 50
                exit
            exit
            egress
                pool
                    amber-alarm-threshold 25
                    red-alarm-threshold 25
                    resv-cbs 20 amber-alarm-action step 5 max 50
                exit
            exit
        exit
```

The following is an example alarm that is triggered when the amber percentage has been exceeded and the reserved buffer size has increased from 20% to 25%:

```
82 2016/04/25 14:21:52.42 UTC MINOR: PORT #2050 Base Resv CBS Alarm
"Amber Alarm: CBS over Amber threshold: ObjType=port Owner=1/1/1 Type=accessIngress
 Pool=default NamedPoolPolicy= ResvSize=672 SumOfQ ResvSize=138 Old ResvCBS=20
New ResvCBS=25 Old ResvSize=528"
```

When a port is configured to be a hybrid port, its buffer space is divided into an access portion and a network portion. The split by default is 50:50 but it can be configured on a per port basis.

```
configure port 1/1/1
        ethernet
            mode hybrid
            encap-type dot1q
        exit
        hybrid-buffer-allocation
            ing-weight access 70 network 30
            egr-weight access 70 network 30
        exit
```

Enqueuing Packets: Queue Sizing

Queue sizes change dynamically when packets are added to a queue faster than they are removed, without any traffic the queue depth is zero. When packets arrive for a queue there will be request for buffer memory which will result in buffers being allocated dynamically from the buffer pool that the queue belongs to.

A queue has four buffer size related attributes: MBS, CBS, high-prio-only, and hi-low-prio-only, which affect packets only during the enqueuing process.

- Maximum Burst Size (MBS) defines the maximum buffer size that a queue can use. If the actual queue depth is larger than the MBS, any incoming packet will not be able to get a buffer and the packet will be dropped. This is defined in bytes or Kbytes for access queues with a configurable non-zero minimum of 1byte or a default (without configuring the MBS) of the maximum between 10ms of the PIR or 64Kbytes. A value of zero will cause all packets to be dropped. MBS is a fractional percentage (xx.xx%) of pool size for network queues with defaults varying dependent on the queue (see default network-queue policy for default values). The MBS setting is the main factor determining the packet latency through a system when packets experience congestion. Queues within an egress queue group can have their MBS configured with as target packet queue delay in milliseconds.
- Committed Burst Size (CBS) defines the maximum guaranteed buffer size for an incoming packet. This buffer space is effectively reserved for this queue as long as the CBS is not oversubscribed (such the sum of the CBS for all queues using this pool does not exceed its reserved buffer pool size). For access queues, the CBS is defined in Kbytes with a configurable non-zero minimum of 6Kbytes or a default (without configuring the CBS) of the maximum between 10ms of the CIR or 6Kbytes. It is a fractional percentage (xx.xx%) of pool size for network queues with defaults varying dependent on the queue (see default network-queue policy for default values). Regardless of what is configured, the CBS attained normally

will not be larger than the MBS. One case where CBS could be configured larger than MBS is for queues on LAGs, because in some cases the CBS is shared among the LAG ports (LAG QoS is not covered in this chapter). If the MBS and CBS values are configured to be equal (or nearly equal) this will result in the CBS being slightly higher than the value configured.

- High-prio-only. As a queue can accept both high and low enqueuing priority packets, a high enqueuing priority packet should have a higher probability to get a buffer. High-prio-only is a way to achieve this. Within the MBS, high-prio-only defines that a certain amount of buffer space will be exclusively available for high enqueuing priority packets. At network ingress and all egress buffering, highpriority corresponds to in-profile and low priority to out-of-profile. At service ingress, enqueuing priority is part of the classification. The high-prio-only is defined as a percentage of the MBS, with the default being 10%. A queue being used only for low priority/out-of-profile packets would normally have this set to zero. The high-prio-only could be considered to be an MBS for low enqueuing/ out-of-profile packets.
- Hi-low-prio-only. There is an additional threshold, hi-low-prio-only, at egress which is equivalent to an MBS for exceed-profile packet. When the queue depth is beyond the hi-low-prio-only depth, the exceed-profile packets are dropped. The hi-low-prio-only is defined as a percentage of the MBS, with the default being 20%.

As with the buffer pools, the MBS, CBS, high-prio-only, and hi-low-prio-only values attained are based on a number of discrete values (not always an increment of 3Kbytes). The values for these parameters can be seen using the **show pools** command.

The MBS changes dynamically for queues in an egress queue group when H-QoS is used and the queue command **dynamic-mbs** is configured. This results in the MBS being modified in the ratio of the operational PIR to the admin PIR which derives an operational MBS. The ratio also affects the high-prio-only and hi-low-prio-only drop tails, and the WRED slopes if a slope policy is applied ot the queue. The configured CBS is used as the minimum operational MBS and the maximum MBS is capped by the maximum admin MBS (1GB).

As packets are added to a queue they will use the available CBS space, in which case they are placed in the reserved portion in the buffer pool. Once the CBS is exhausted, packets use the shared buffer pool space up to the hi-low-prio-only threshold (for exceed-profile packets), the high-prio-only threshold (for out-of-profile packets), or the maximum MBS size (for inplus-profile and in-profile packets).

The following configuration shows a queue with a specific MBS, CBS and disable high-prio-only.

```
configure
    qos
```

```
        sap-ingress 10 create
            queue 1 create
                mbs 10000
                cbs 100
                high-prio-only 0
            exit
        exit
```

Queue depth monitoring aims to give more visibility to the operator of the queue depths when traffic is bursty. It is a polling mechanism that is by default disabled. Queue depth monitoring can be enabled as a queue override on a service SAP or on a queue group.

The following command enables queue depth monitoring on SAP 1/1/1:11 in VPLS 10:

```
configure service vpls 10 sap 1/1/1:11 ingress queue-override queue 1 create monitor-
depth
```

The result of the queue depth monitoring is presented in the form of occupancy ranges of 10% on the queue depth for each configured queue with the percentage of polls seen in each occupancy range, as follows:

```
*A:PE-1# show service id 10 sap 1/1/1:11 queue-depth
===============================================================================
Queue Depth Information (Ingress SAP)
===============================================================================
No Matching Entries
===============================================================================

===============================================================================
Queue Depth Information (Egress SAP)
===============================================================================
-------------------------------------------------------------------------------
Name                       : 10->1/1/1:11->1
MBS                        : Def

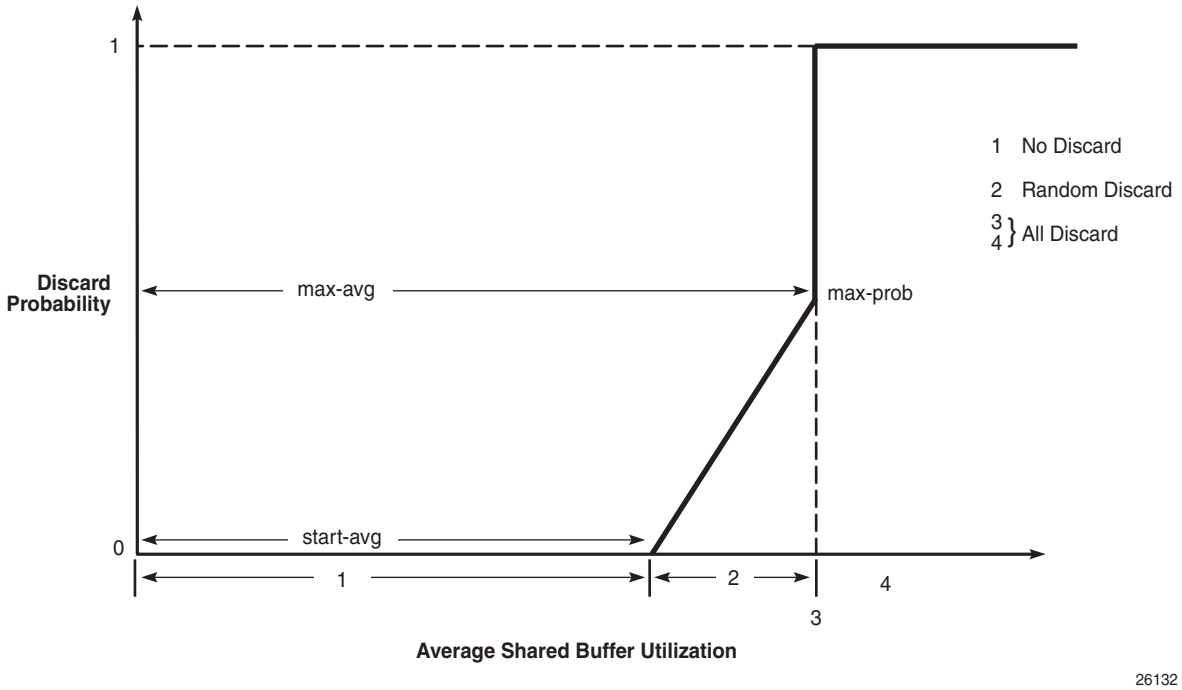
-------------------------------------------------------------------------------
Queue Depths (percentage)
-------------------------------------------------------------------------------
0%-10% 11%-20% 21%-30% 31%-40% 41%-50% 51%-60% 61%-70% 71%-80% 81%-90% 91%-100%
-------------------------------------------------------------------------------
68.21  3.64    3.43    3.47    3.86    3.22    3.86    2.78    3.78    3.66
-------------------------------------------------------------------------------
Average Elapsed Time      : 0d 00:39:11
Wghtd Avg PollEgr Interval: 100 ms
-------------------------------------------------------------------------------
*A:PE-1#
```

Enqueuing Packets: Weighted Random Early Detection (WRED)

In order to gracefully manage the use of the shared portion of the buffer pool, WRED can be configured on that part of the pool, and therefore applies to all queues in the shared pool as it fills. WRED is a congestion avoidance mechanism designed for TCP traffic. This chapter will only focus on the configuration of WRED. WRED-per-queue is an option to have WRED apply on a per egress queue basis, but is not covered here.

WRED is configured by a slope-policy which contains two WRED slope definitions, a high-slope which applies WRED to high enqueuing priority/in-profile packets and a low-slope which applies WRED to low enqueuing priority/out-of-profile packets. Both have the standard WRED parameters: start average (start-avg), maximum average (max-avg) and maximum probability (max-prob), and can be enabled or disabled individually. WRED slope policies also contain definitions for two slopes which are only applicable to access and network egress; a highplus-slope which applies WRED to inplus-profile packets and an exceed-slope which applies WRED to exceed-profile packets. The WRED slope characteristics are shown in Figure 386.

Figure 386 **WRED Slope Characteristics**



A time-average-factor parameter can be configured per slope-policy which determines the sensitivity of the WRED algorithm to shared buffer utilization fluctuations (the smaller the value makes the average buffer utilization more reactive to changes in the instantaneous buffer utilization). The slope-policy is applied on a network port under **config > card > mda > network > ingress > pool** and **config > port > network > egress > pool** and on an access port under **config > port > access > ingress > pool** and **config > port > access > egress > pool**.

WRED is usually configured for assured and best-effort service traffic with premium traffic not typically being subject to WRED as it is always given preferential treatment and should never be dropped.

The following configuration defines a WRED slope policy and applies it to an ingress access port (the highplus and exceed slopes are ignored at ingress).

```
configure
    qos
        slope-policy "slope1" create
            exceed-slope
                shutdown
                start-avg 30
                max-avg 55
                max-prob 80
            exit
            high-slope
                start-avg 80
                max-avg 100
                max-prob 100
                no shutdown
            exit
            highplus-slope
                shutdown
                start-avg 85
                max-avg 100
                max-prob 80
            exit
            low-slope
                max-avg 100
                start-avg 80
                max-prob 100
                no shutdown
            exit
            time-average-factor 12
        exit

configure
    port 1/1/1
        access
            ingress
                pool
                    slope-policy "slope1"
                exit
            exit
        exit
```

The queue sizing parameters and buffer pools layout for ingress is shown in Figure 387.

Figure 387 **Ingress Buffer Pools and Queue Sizing**

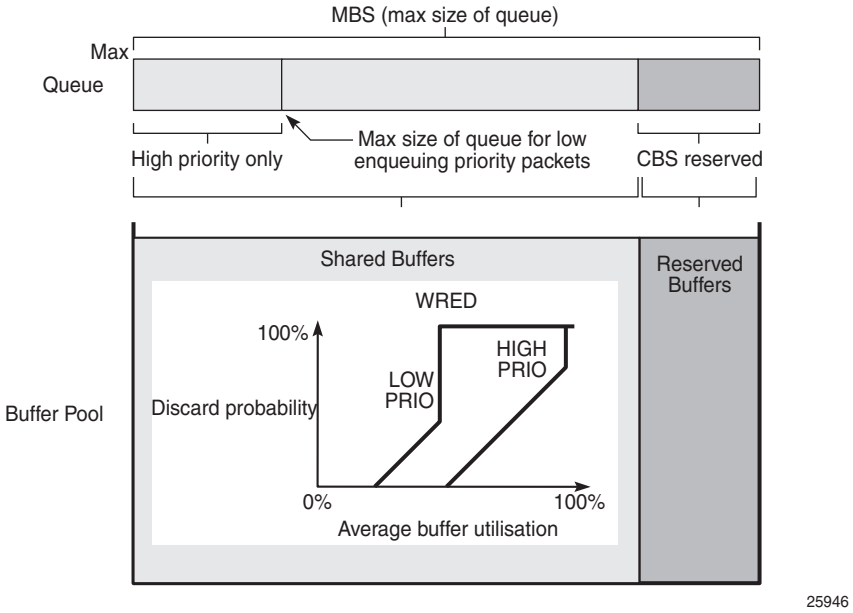
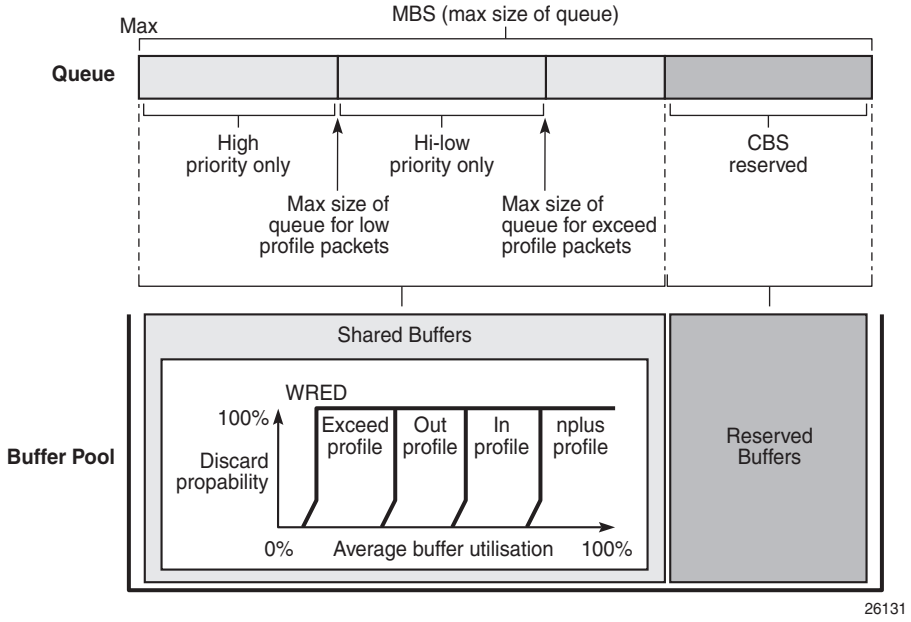


Figure 388 shows the queue sizing parameters and buffer pools layout for egress.

Figure 388 **Egress Buffer Pools and Queue Sizing**



Dequeuing Packets: Queue Rates

A queue has two rate attributes: PIR and CIR. These affect packets only during the dequeue process.

- PIR — If the instantaneous dequeue rate of a queue reaches this rate, the queue is no longer served. Excess packets will be discarded eventually when the queue reaches its MBS/high-prio-only/hi-low-prio-only sizes. The PIR for access ports can be set in Kb/s with a default of **max** or as a percentage (see below). For network ports, the PIR is set as a percentage of 390000Kb/s for ingress queues and of the port speed for egress queues, both with a default of 100%.
- CIR — The CIR is used to determine whether an ingress packet is in-profile or out-of-profile at the SAP ingress. It is also used by the scheduler in that queues operating within their CIRs will be served ahead of queues operating above their CIRs. The CIR for access ports can be set in Kb/s with a default of zero or as a percentage (see below). For network ports, it is set as a percentage of 390000Kb/s for ingress queues and of the port speed for egress queues, with defaults varying dependent on the queue.

A percentage rate can be used in the sap-ingress and sap-egress policies, and can be defined relative to the local-limit (the parent scheduler rate) or the port-limit (the rate of the port on which the SAP is configured, including any egress-rate configured). The parameters rate and percent-rate are mutually exclusive and will overwrite each other when configured in the same policy. The following example shows a percent-rate configured as a port-limit.

```
configure
    qos
        sap-egress 10 create
            queue 1 create
                percent-rate 50.00 cir 10.00 port-limit
            exit
```

The PIR and CIR rates are shown in Figure 389.

The queues operate at discrete rates supported by the hardware. If a configured rate does not match exactly one of the hardware rates an adaptation rule can be configured to control whether the rate is rounded up or down or set to the closest attainable value. The actual rate used can be seen under the operational PIR/CIR (O.PIR/O.CIR) in the **show pools** command output.

The following configuration shows a queue with a PIR, CIR and adaptation rule.

```
configure
    qos
        sap-ingress 20 create
            queue 2 profile-mode create
                adaptation-rule pir max cir min
                rate 10000 cir 5000
            exit
```

By default, the rates apply to packet bytes based on packet accounting, which for Ethernet includes the Layer 2 frame plus the FCS. An alternative is frame accounting which adds the Ethernet inter-frame gap, preamble and start frame delimiter.

Dequeuing Packets: Scheduling

Once a packet is placed on a queue, it is always dequeued from the queue by a scheduler. The scheduling order of the queues dynamically changes depending on whether a queue is currently operating below or above its CIR, with expedited queues being serviced before best-effort queues. This results in a default scheduling order of (in strict priority).

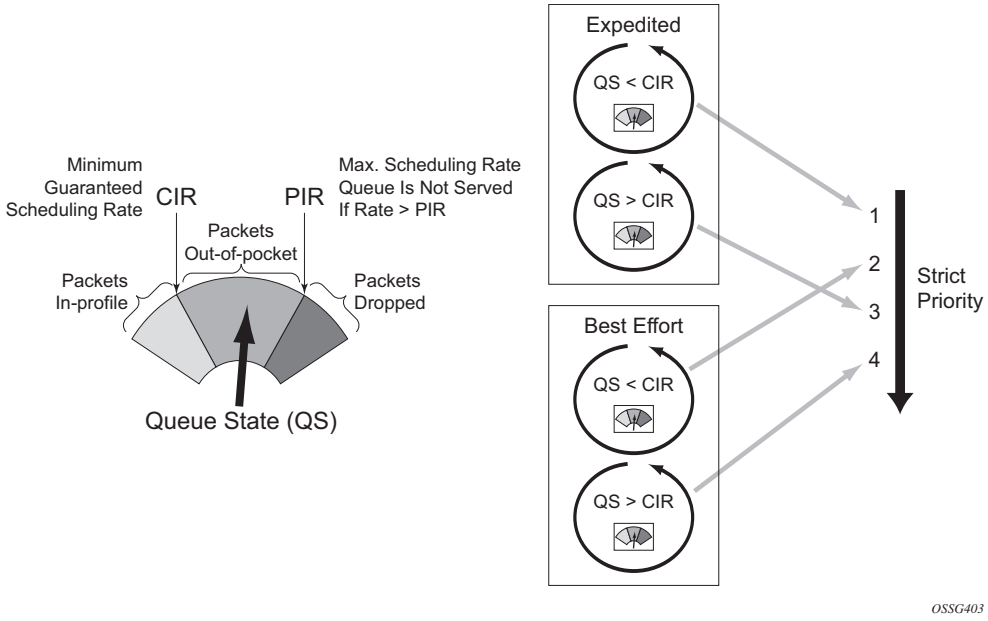
1. Expedited queues operating below CIR
2. Best-effort queues operating below CIR

3. Expedited queues operating above CIR
4. Best-effort queues operating above CIR

This is displayed in Figure 389.

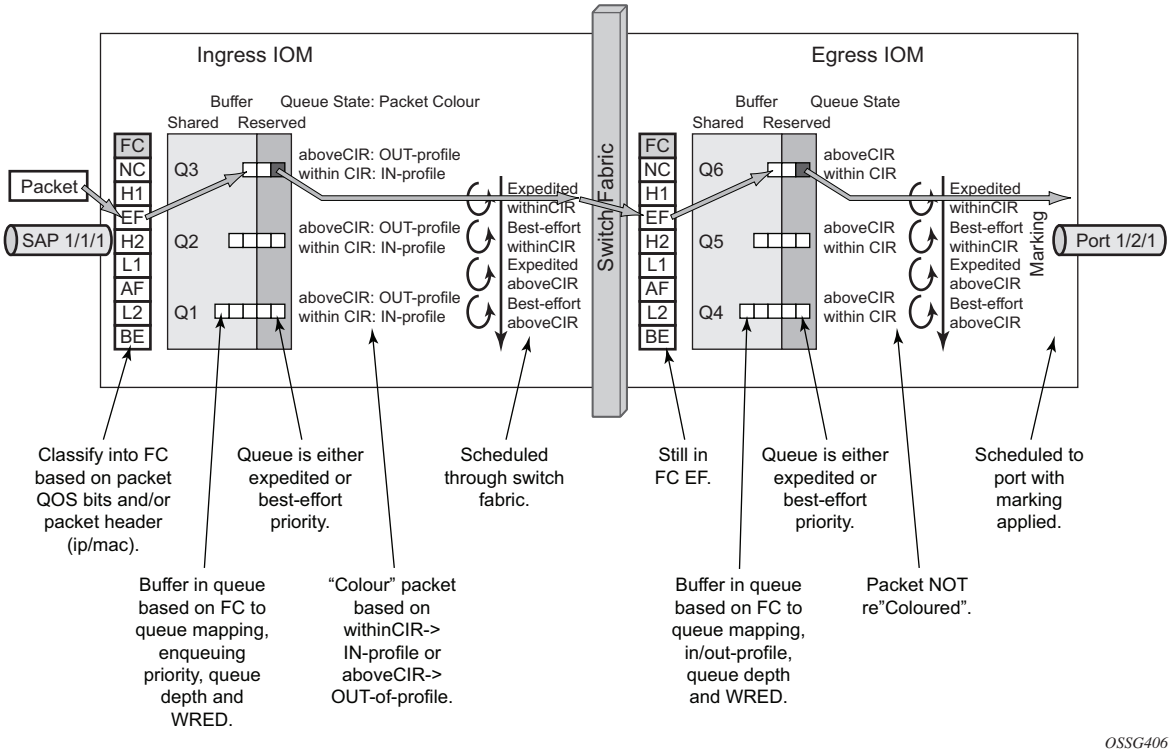
The scheduling orperation can be modified using hierarchical QoS (with a scheduler-policy or port-scheduler-policy) which is out of scope of this chapter.

Figure 389 **Scheduling (Dequeuing Packets from the Queue)**



The overall QoS actions at both the ingress and egress IOMs are shown in Figure 390.

Figure 390 **IOM QoS Overview**



Show Output

The following displays **show** command output for:

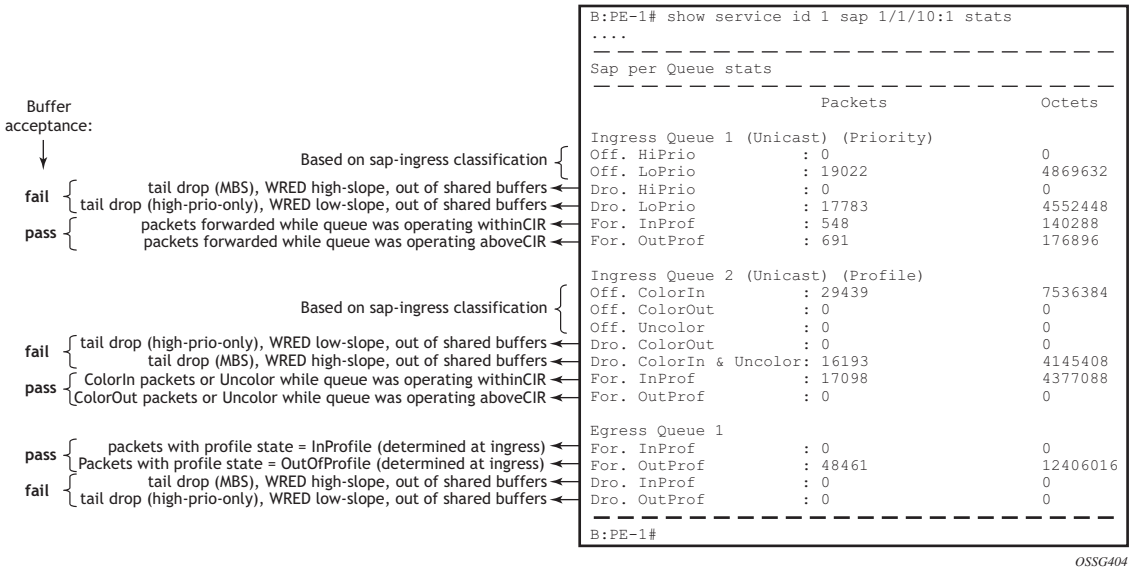
- SAP queue statistics
- port queue statistics
- per-port aggregate egress-queue statistics monitoring
- access-ingress pools

The **show pools** command output for network-ingress and network/access-egress is similar to that of access-ingress and is not included here.

SAP Queue Statistics

The following output shows an example of the ingress and egress statistics on a SAP for an IES service (without multicast enabled, therefore no ingress multicast queue). There are two ingress queues, one being in priority mode and the other in profile mode. An explanation of the statistics is given for each entry.

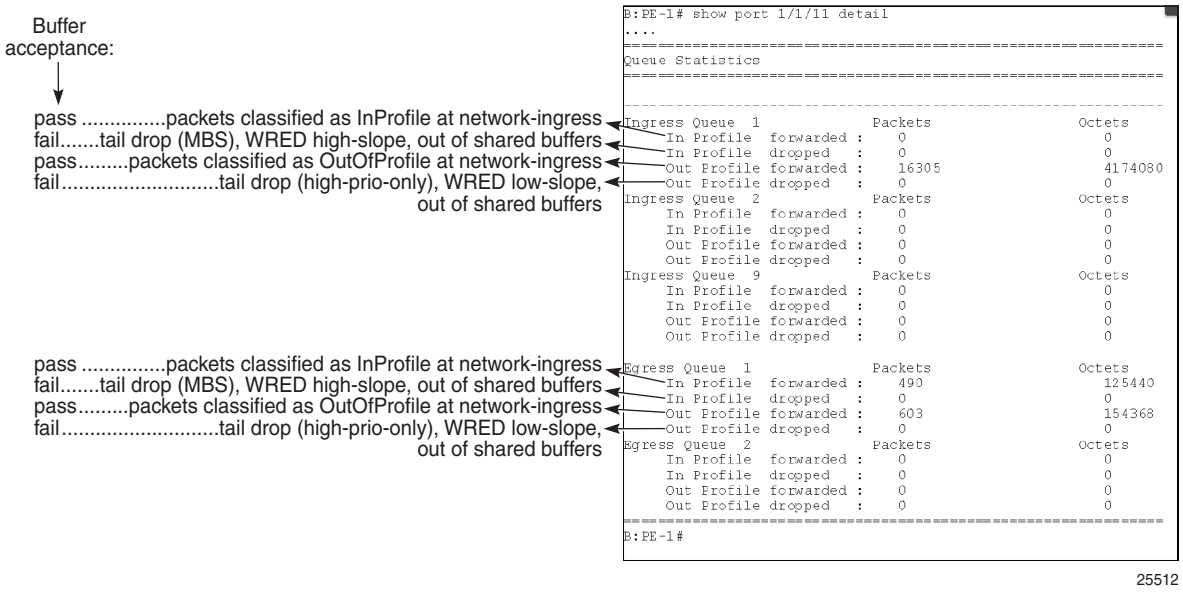
Figure 391 **Ingress and Egress SAP Queue Statistics for an IES Service**



Port Queue Statistics

This output shows an example of the ingress and egress network port statistics. There are two unicast ingress queues (1 and 2) and one multicast ingress queue (9) with two egress queues. An explanation of the statistics is given for each entry.

Figure 392 **Ingress and Egress Network Port Queue Statistics**



Per-Port Aggregate Egress-Queue Statistics Monitoring

Per-port aggregate egress-queue statistics show the number of forwarded and the number of dropped packets for in-profile and out-of-profile packets. All egress queues on the port are monitored: SAP egress, network egress, subscriber egress, egress queue group queues, system queues.

Per-port aggregate egress-queue statistics monitoring is enabled with the following command:

```
configure port 1/1/1 monitor-agg-egress-queue-stats
```

The collected statistics can be displayed as follows:

```
*A:PE-1# show port 1/1/1 statistics egress-aggregate

===============================================================================
Port 1/1/1 Egress Aggregate Statistics on Slot 1
===============================================================================
                        Forwarded                Dropped                  Total
-------------------------------------------------------------------------------
PacketsIn                       0                      0                      0
PacketsOut                5251690                      0                5251690
OctetsIn                        0                      0                      0
OctetsOut               357114942                      0              357114942
===============================================================================
```

```
*A:PE-1#
```

Access-Ingress Pools

This output shows an example of the default pools output for access-ingress. It includes the pools sizes, WRED information and queue parameters for each queue in the pool.

For this particular output, queue 2 on SAP 5/1/1:1 is being over-loaded which is causing its queue depth to be 67087296 bytes, made up of 64509 Kbytes from the shared pool (in use) and 1008 Kbytes from the reserved pool (in use). The output shows the pool total in usage as 65517 Kbytes, which is the sum of the shared and reserved pool in use. Sometimes the sum and total could be different by the size of one buffer, however, this is due to the dynamics of the buffer allocation which uses a ‘sliding-window’ mechanism and may therefore not always be perfectly aligned.

It can be seen that the high, low, and exceed WRED slopes are enabled and their instantaneous drop probability is shown 100% and their max averages are 64512 Kbytes, 46080 Kbytes, and 27648 Kbytes, respectively – this shows that the reserved portion of the buffer pool on this port is exhausted causing WRED to drop the packets for this queue.

The admin and operational PIR on the overloaded queues is 10Mb/s with CIR values of zero.

```
*A:PE-1# show pools 5/1/1 access-ingress

===============================================================================
Pool Information
===============================================================================
Port                 : 5/1/1
Application          : Acc-Ing          Pool Name             : default
CLI Config. Resv CBS : 30%(default)
Resv CBS Step        : 0%               Resv CBS Max          : 0%
Amber Alarm Threshold: 0%               Red Alarm Threshold   : 0%
-------------------------------------------------------------------------------
Utilization                   State       Start-Avg    Max-Avg    Max-Prob
-------------------------------------------------------------------------------
High-Slope                    Up                70%        70%        100%
Low-Slope                     Up                50%        50%         80%
Exceed-Slope                  Up                30%        30%         80%
Time Avg Factor      : 12
Pool Total           : 132096 KB
Pool Shared          : 92160 KB         Pool Resv             : 39936 KB

High Slope Start Avg : 64500 KB         High Slope Max Avg    : 64512 KB
Low Slope Start Avg  : 46068 KB         Low Slope Max Avg     : 46080 KB
Excd Slope Start Avg : 27636 KB         Excd Slope Max Avg    : 27648 KB
```

```
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
Current Resv CBS    Provisioned    Rising         Falling        Alarm
%age                all Queues     Alarm Thd      Alarm Thd      Color
-------------------------------------------------------------------------------
30%                 1020 KB        NA             NA             Green
Pool Total In Use    : 65517 KB
Pool Shared In Use   : 64509 KB         Pool Resv In Use      : 1008 KB
WA Shared In Use     : 64509 KB

Hi-Slope Drop Prob   : 100              Lo-Slope Drop Prob    : 100
Excd-Slope Drop Prob : 100

===============================================================================
Queue : 1->5/1/1:1->1
===============================================================================
FC Map           : be l2 af l1 h2 h1 nc
Tap              : 5/1
Admin PIR        : 10000000             Oper PIR         : Max
Admin CIR        : 0                    Oper CIR         : 0
Admin MBS        : 12288 KB             Oper MBS         : 12288 KB
Hi Prio Only     : 1344 KB              Hi Low Prio Only : 2496 KB
CBS              : 12 KB                Depth            : 0
Slope            : not-applicable
===============================================================================

===============================================================================
Queue : 1->5/1/1:1->2
===============================================================================
FC Map           : ef
Tap              : 5/1
Admin PIR        : 10000                Oper PIR         : 10000
Admin CIR        : 0                    Oper CIR         : 0
Admin MBS        : 132096 KB            Oper MBS         : 132096 KB
Hi Prio Only     : 0 KB                 Hi Low Prio Only : 27648 KB
CBS              : 1008 KB              Depth            : 67087296 B
Slope            : not-applicable
===============================================================================

===============================================================================
Queue : 28->5/1/1:28->1
===============================================================================
FC Map           : be l2 af l1 h2 ef h1 nc
Tap              : 5/1
Admin PIR        : 10000000             Oper PIR         : Max
Admin CIR        : 0                    Oper CIR         : 0
Admin MBS        : 12288 KB             Oper MBS         : 12288 KB
Hi Prio Only     : 1344 KB              Hi Low Prio Only : 2496 KB
CBS              : 0 KB                 Depth            : 0
Slope            : not-applicable
===============================================================================

===============================================================================
Queue : 28->5/1/1:28->11
===============================================================================
FC Map           : be l2 af l1 h2 ef h1 nc
Tap              : MCast
Admin PIR        : 10000000             Oper PIR         : Max
```

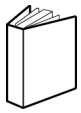
```
Admin CIR        : 0                    Oper CIR         : 0
Admin MBS        : 12288 KB             Oper MBS         : 12288 KB
Hi Prio Only     : 1344 KB              Hi Low Prio Only : 2496 KB
CBS              : 0 KB                 Depth            : 0
Slope            : not-applicable
===============================================================================
===============================================================================

*A:PE-1#
```

Conclusion

This chapter has described the basic QoS functionality available on the Nokia 7x50 platforms, specifically focused on the FP2 chipset. This comprises of the use of queues to shape traffic at the ingress and egress of the system and the classification, buffering, scheduling and remarking of traffic on access, network, and hybrid ports.

Customer Document and Product Support



Customer documentation

Customer Documentation Welcome Page



Technical Support

Product Support Portal



Documentation feedback

Customer Documentation Feedback

© 2016 Nokia.
3HE 11598 AAAA TQZZA 01