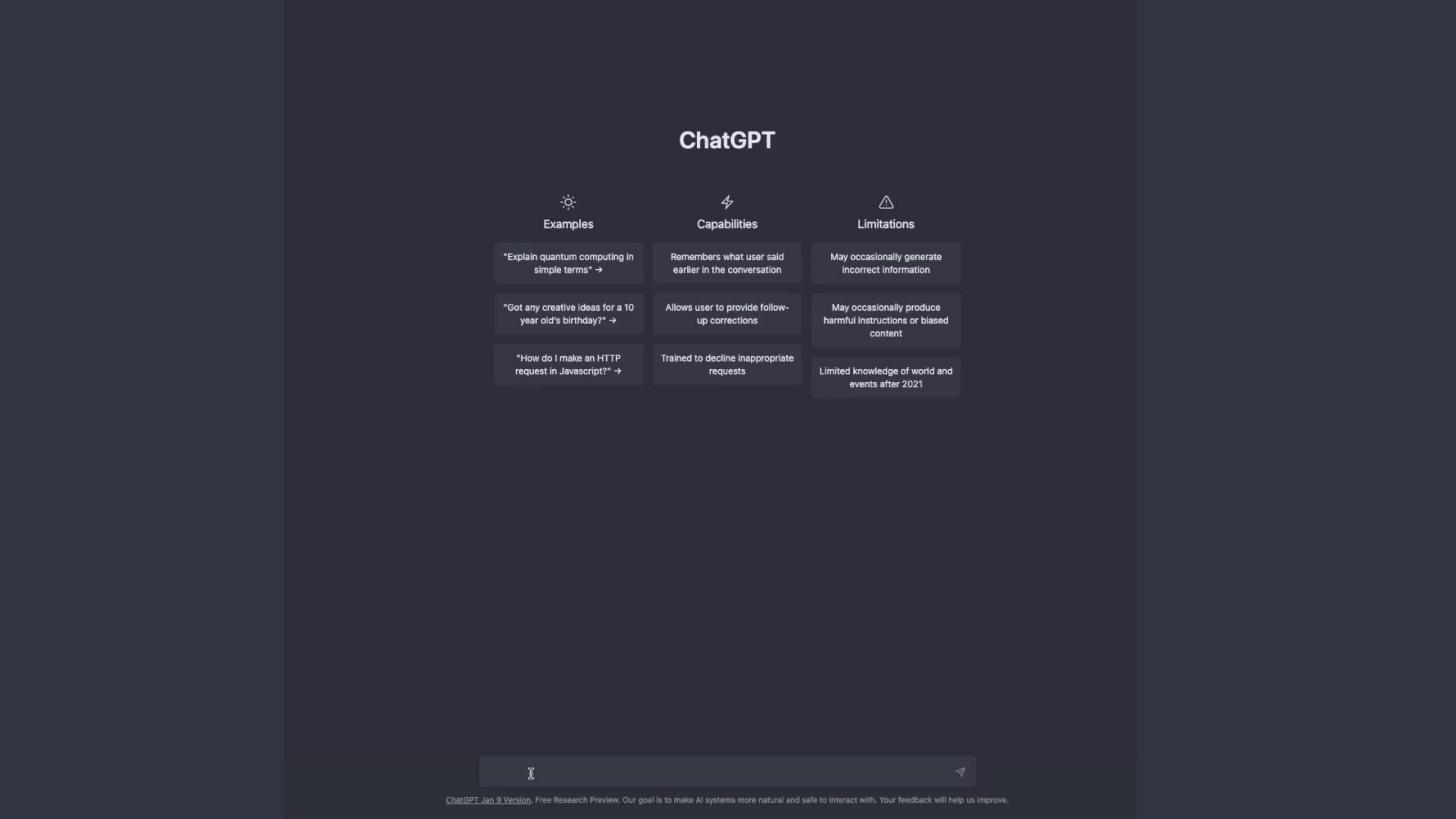
The bridge to possible

ACI Troubleshooting: A deep dive into PBR

Kristof De Brouwer
Technical Leader ACI TAC EMEAR
CCIE #23336



ChatGPT

Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →

Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests

Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

ChatGPT Jan 9 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

Cisco Webex App

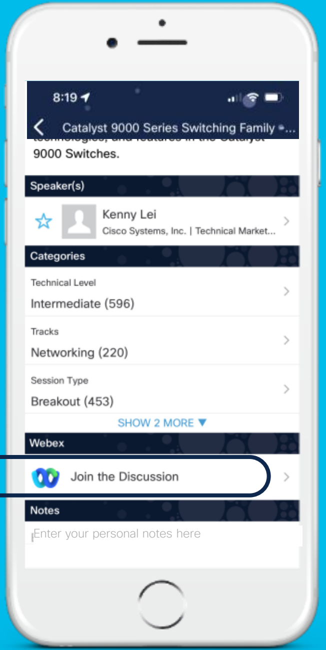
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated
until February 24, 2023.



8:19

Catalyst 9000 Series Switching Family ...

9000 Switches.

Speaker(s)

Kenny Lei
Cisco Systems, Inc. | Technical Market...

Categories

Technical Level
Intermediate (596)

Tracks
Networking (220)

Session Type
Breakout (453)

SHOW 2 MORE ▼

Webex

Join the Discussion

Notes

Enter your personal notes here

Agenda

- Introduction
- PBR Firewall insertion in ACI Multipod
  - East-West
  - North-South
- PBR Firewall insertion in ACI Multisite
  - East-West
  - North-South

Acronyms/Definitions

Reference Slide Icon ➔

Acronyms	Definitions	Acronyms	Definitions
EPG and EP	Endpoint Group and Endpoint	BD	Bridge Domain
FW	Firewall	Zoning-rule	Refer to a permit/deny/redirect rule between two pcTag on a leaf
LB	Load Balancer	Redir-info	Redirect info – refers to relevant info to apply redirect including VMAC to redirect, VIP and Service BD
PBR	Policy Based Redirect	SNAT	Source NAT
L3out	Layer 3 out		
North-South	Refer to traffic between EPG and L3out		
East-West	Refer to traffic between EPG or within EPG		
Ext EPG	External EPG aka EPG part of a L3out		
pcTag	Policy Tag		
sclass	Source class or pcTag of source		
dclass	Destination class of pcTag of destination		
VNID	VXLAN network identifier – refer to either a BD or a VRF in ACI		

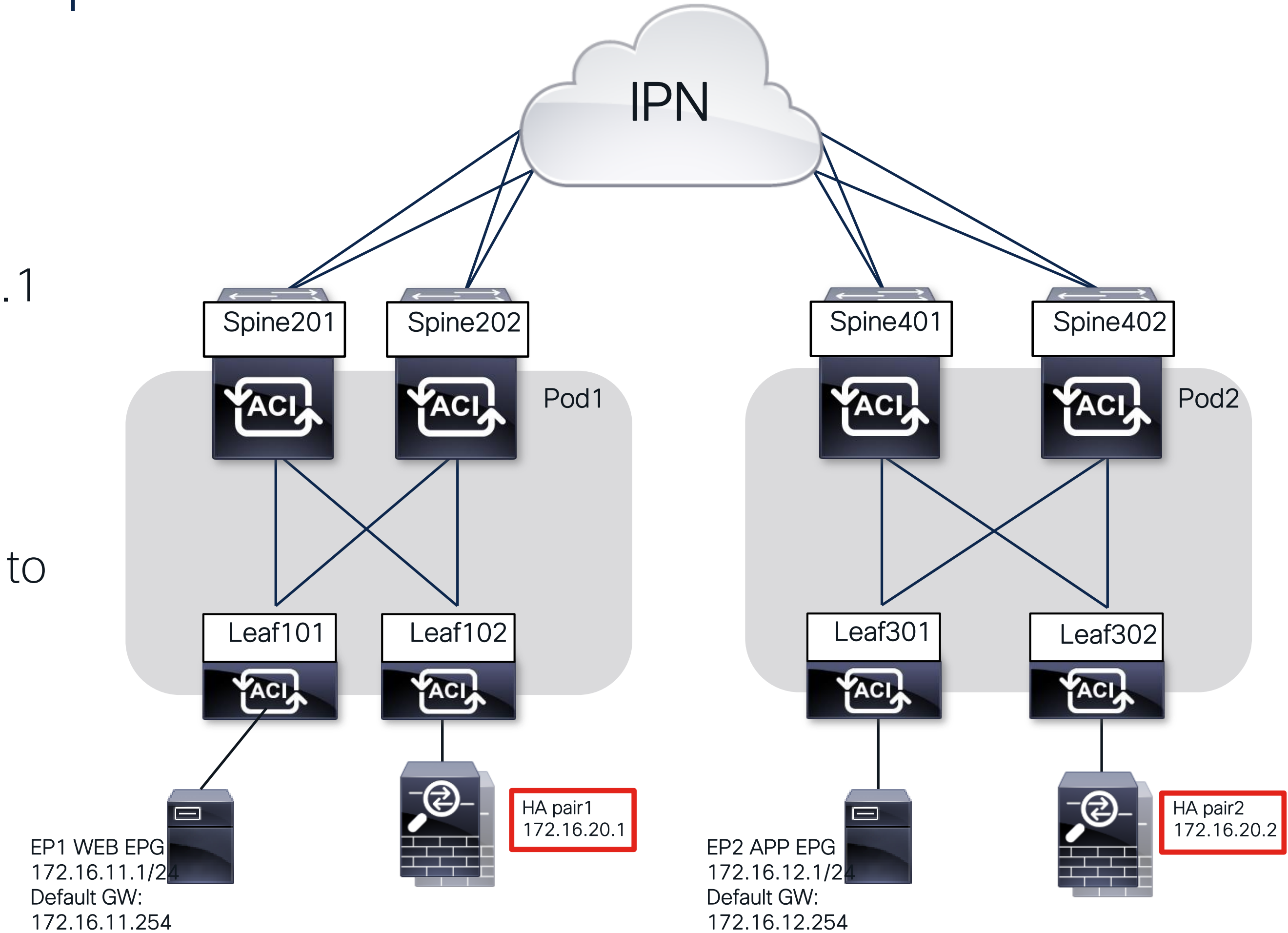
Multipod East-West Symmetric PBR

Topology – Multipod East-West
Symmetric PBR

Routed flow between 172.16.11.1 to 172.16.12.1

Redirected to one the Firewall HA pair

FW are one-arm attached to ACI



Packet – symmetric PBR

Packet walk Consumer WEB to Provider APP Epg

5

Return from firewall on service leaf hits permit rule to egress leaf

4

On service leaf, it is a pure Layer 2 packet to the firewall

3

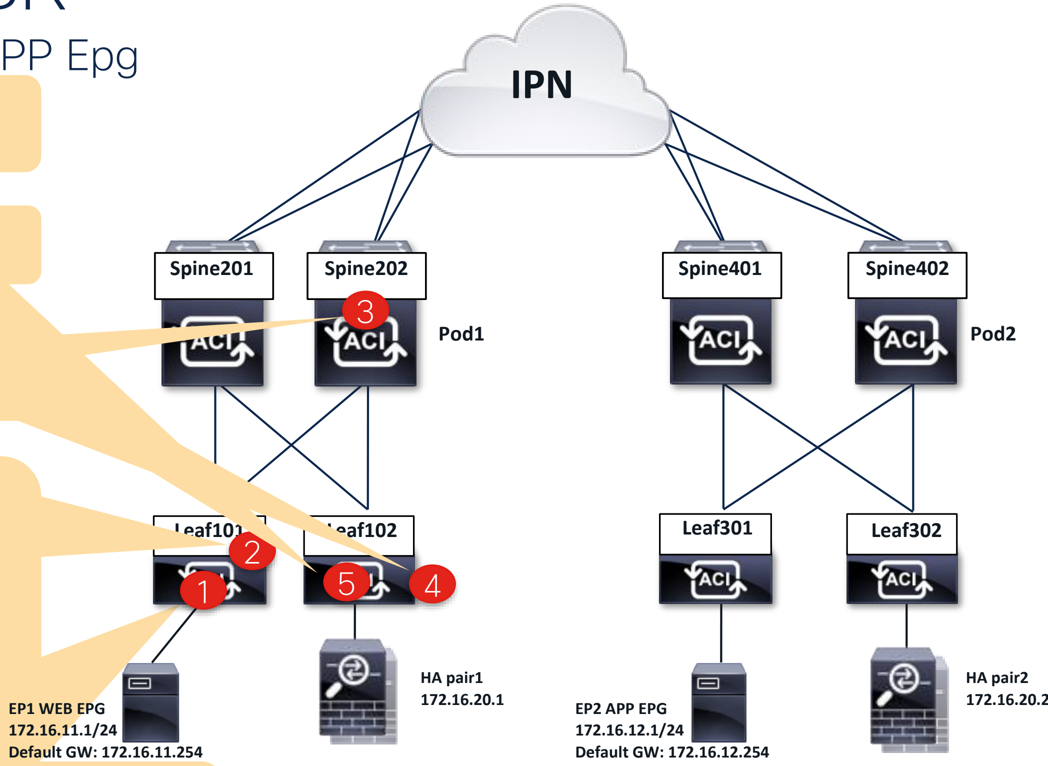
COOP lookup in BD VNID for Redirect mac and will send it toward service leaf

2

Leaf doing the redirect (101 or 301)
DMAC is rewritten to Firewall Pair1 or 2 .
No Mac lookup happening on leaf.
Packet is encapsulated to Service BD VNID and send to vxlan tunnel to anycast-mac on spine.

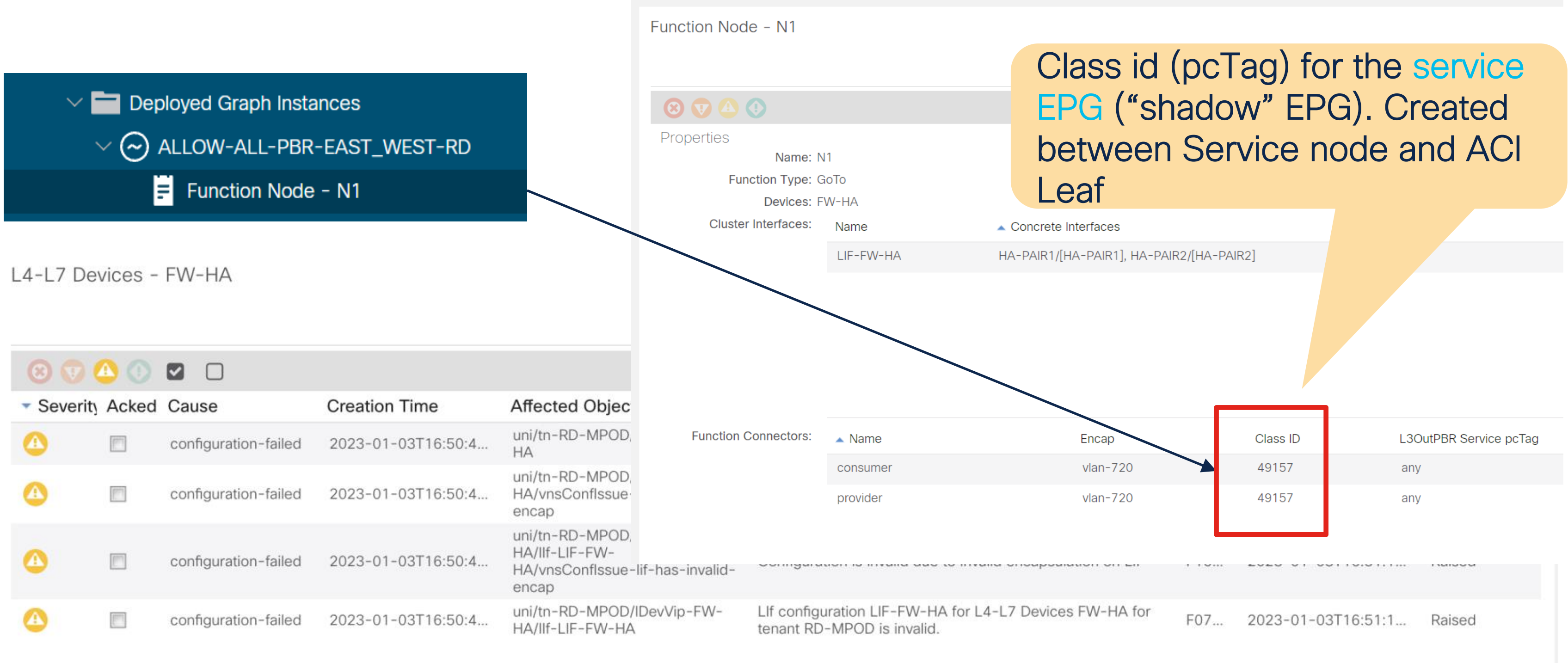
1

Ingress leaf
if EP is known → redirect
if EP is unknown redirect will happen on egress leaf (301)



Check 1 – Is the Graph deployed

Once Config is completed (Contract, Serv Graph Template, device selection policies.,)



- Deployed Graph Instances
  - ALLOW-ALL-PBR-EAST_WEST-RD
    - Function Node - N1

L4-L7 Devices - FW-HA

Severity	Acked	Cause	Creation Time	Affected Object				
⚠	□	configuration-failed	2023-01-03T16:50:4...	uni/tn-RD-MPOD/ HA				
⚠	□	configuration-failed	2023-01-03T16:50:4...	uni/tn-RD-MPOD/ HA/vnsConfIssue- encap				
⚠	□	configuration-failed	2023-01-03T16:50:4...	uni/tn-RD-MPOD/ HA/lIf-LIF-FW- HA/vnsConfIssue-lif-has-invalid- encap				
⚠	□	configuration-failed	2023-01-03T16:50:4...	uni/tn-RD-MPOD/lDevVip-FW- HA/lIf-LIF-FW-HA	LIf configuration LIF-FW-HA for L4-L7 Devices FW-HA for tenant RD-MPOD is invalid.	F07...	2023-01-03T16:51:1...	Raised

Function Node - N1

Properties

Name: N1

Function Type: GoTo

Devices: FW-HA

Cluster Interfaces:

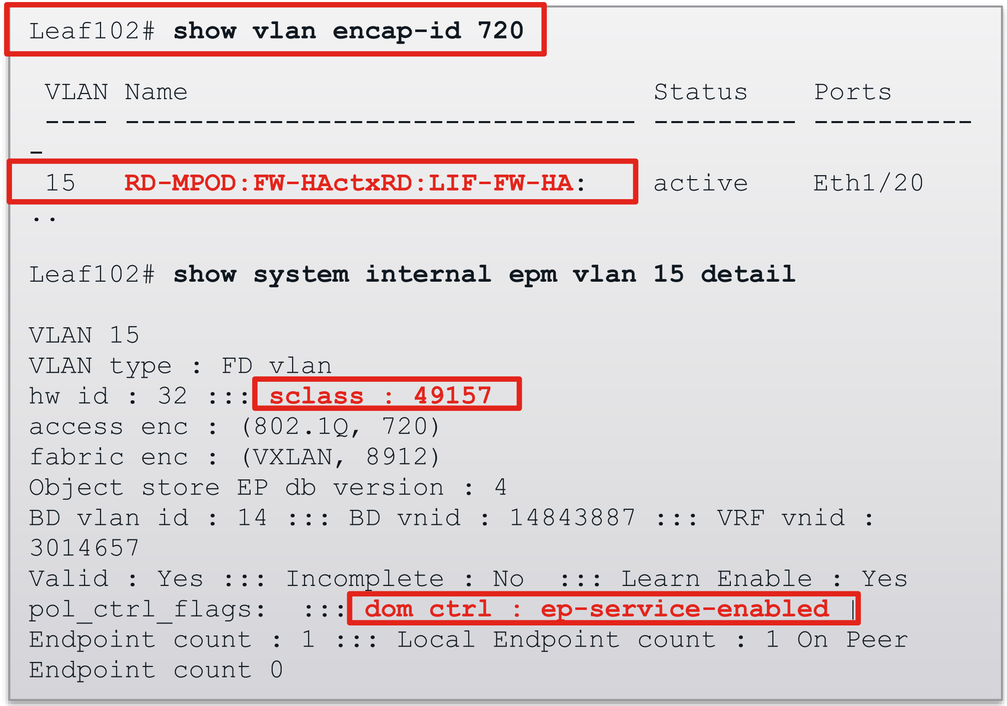
Name	Concrete Interfaces
LIF-FW-HA	HA-PAIR1/[HA-PAIR1], HA-PAIR2/[HA-PAIR2]

Function Connectors:

Name	Encap	Class ID	L3OutPBR Service pcTag
consumer	vlan-720	49157	any
provider	vlan-720	49157	any

Class id (pcTag) for the service EPG ("shadow" EPG). Created between Service node and ACI Leaf

Check 2 – Is the Service EPG deployed

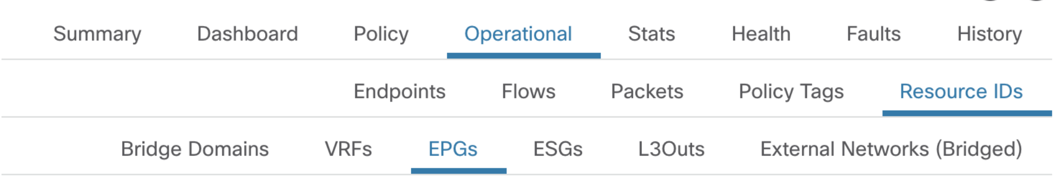


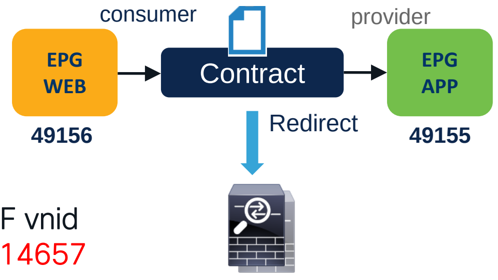
```
Leaf102# show vlan encap-id 720

 VLAN Name                             Status    Ports
 ---- -------------------------------- --------- ----------
-
 15   RD-MPOD:FW-HActxRD:LIF-FW-HA:    active    Eth1/20
..

Leaf102# show system internal epm vlan 15 detail

VLAN 15
VLAN type : FD vlan
hw id : 32 ::: sclass  :  49157
access enc : (802.1Q, 720)
fabric enc : (VXLAN, 8912)
Object store EP db version : 4
BD vlan id : 14 ::: BD vnid : 14843887 ::: VRF vnid :
3014657
Valid : Yes ::: Incomplete : No  ::: Learn Enable : Yes
pol_ctrl_flags:  ::: dom ctrl  :  ep-service-enabled  |
Endpoint count : 1 ::: Local Endpoint count : 1 On Peer
Endpoint count 0
```

- FW cluster interface is using the defined encap vlan-720.
- Service VLAN is deployed on the service leafs and is using the correct service EPG pcTag (sclass 49157).
- The VLAN is marked as a service EPG.

Check 3 - Zoning-rules

Take note of all vnid and sclass involved





VRF vnid 3014657

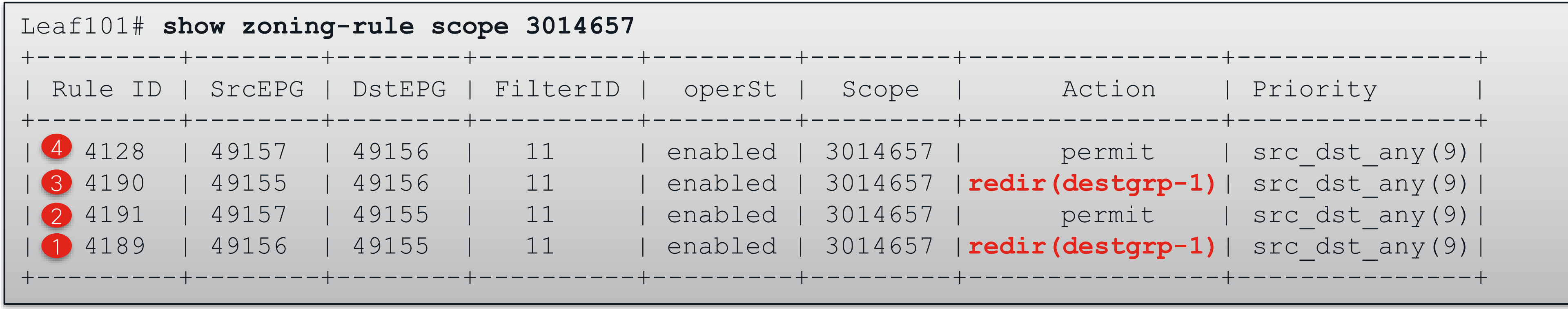
pcTag 49157 (service epg)

One-Arm FW

Expected zoning-rules:

1. Cons to Prov : 49156 to 49155 : REDIRECT
2. Shadow to Prov : 49157 to 49155 : PERMIT
3. Prov to Cons : 49155 to 49156 : REDIRECT
4. Shadow to Cons : 49157 to 49156 : PERMIT

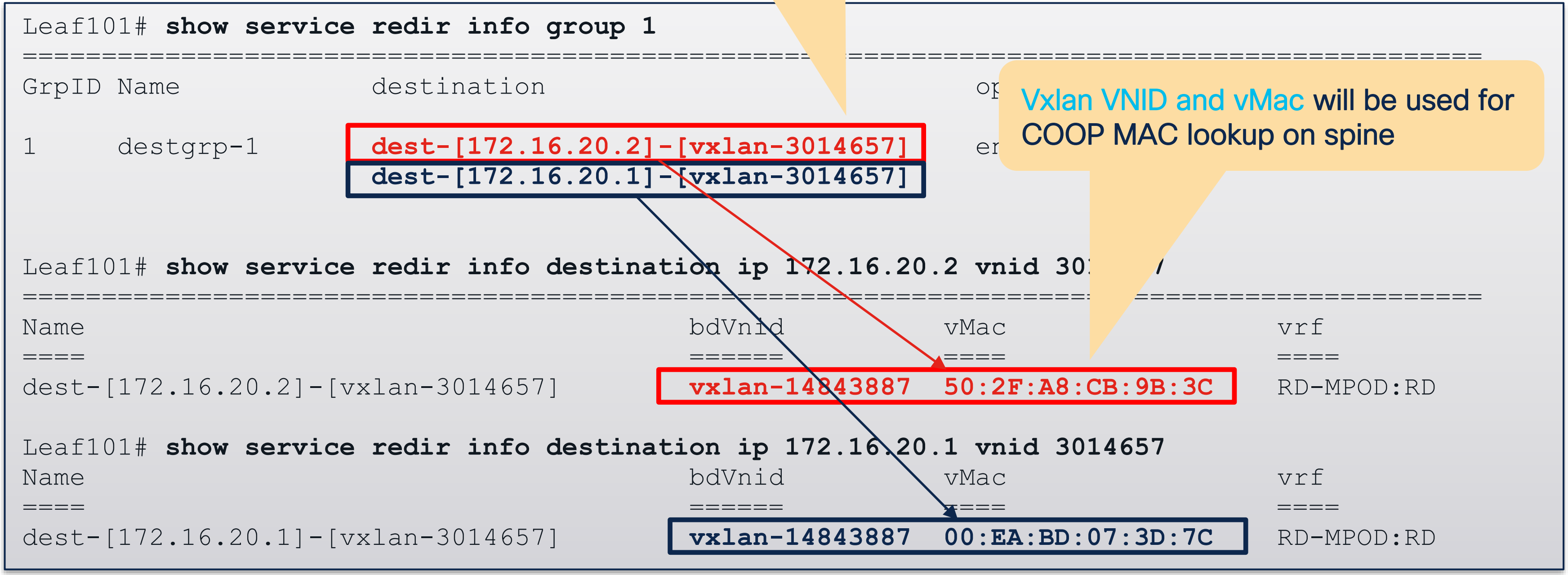
Note it may be all rules are not on the same leaf



```
Leaf101# show zoning-rule scope 3014657
+---------+--------+--------+----------+---------+---------+-----------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID | operSt  |  Scope  |      Action     | Priority       |
+---------+--------+--------+----------+---------+---------+-----------------+----------------+
| 4 4128  | 49157  | 49156  |    11    | enabled | 3014657 |      permit     | src_dst_any(9) |
| 3 4190  | 49155  | 49156  |    11    | enabled | 3014657 |redir(destgrp-1) | src_dst_any(9) |
| 2 4191  | 49157  | 49155  |    11    | enabled | 3014657 |      permit     | src_dst_any(9) |
| 1 4189  | 49156  | 49155  |    11    | enabled | 3014657 |redir(destgrp-1) | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+-----------------+----------------+
```

Check 4 - Redirect info

Redir group should have the VIP of each HA pair



```
Leaf101# show service redir info group 1
==============================================================================
GrpID Name          destination                                  op

1     destgrp-1     dest-[172.16.20.2]-[vxlan-3014657]           er
                    dest-[172.16.20.1]-[vxlan-3014657]


Leaf101# show service redir info destination ip 172.16.20.2 vnid 30     7
==============================================================================
Name                                bdVnid         vMac             vrf
====                                ======         ====             ====
dest-[172.16.20.2]-[vxlan-3014657]  vxlan-14843887  50:2F:A8:CB:9B:3C  RD-MPOD:RD

Leaf101# show service redir info destination ip 172.16.20.1 vnid 3014657
Name                                bdVnid         vMac             vrf
====                                ======         ====             ====
dest-[172.16.20.1]-[vxlan-3014657]  vxlan-14843887  00:EA:BD:07:3D:7C  RD-MPOD:RD
```

Vxlan VNID and vMac will be used for COOP MAC lookup on spine

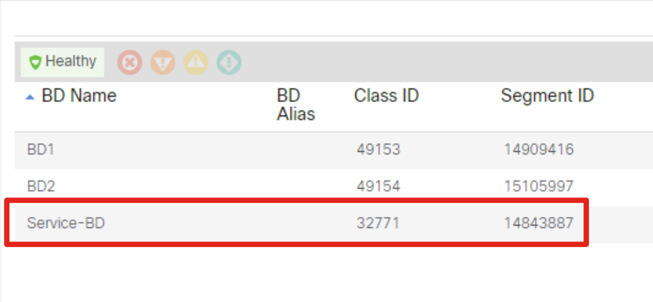
Check 5 – Coop DB on Spine

Verify COOP DB if hashing gives you FW MAC

```
Spine201# show coop internal info  repo ep key 14843887 00:EA:BD:07:3D:7C
Repo Hdr Checksum : 46240
Repo Hdr record timestamp : 10 12 2022 14:37:13 505028097
Repo Hdr last pub timestamp : 10 12 2022 14:37:13 507060173
Repo Hdr last dampen timestamp : 01 01 1970 00:00:00 0
Repo Hdr dampen penalty : 0
Repo Hdr flags : IN_OBJ EXPORT ACTIVE
EP bd vnid : 14843887
EP mac :  00:EA:BD:07:3D:7C
flags : 0x80
repo flags : 0x122
Vrf vnid : 3014657
PcTag : 0x1008004
EVPN Seq no : 0
Remote publish timestamp: 01 01 1970 00:00:00 0
Snapshot timestamp: 10 12 2022 14:37:13 505028097
Tunnel nh : 10.0.0.67
MAC Tunnel  : 10.0.0.67
TX Status: COOP_TX_DONE
Damp penalty: 30
Damp status: NORMAL
Leaf 0 Info
IPv4 Repo H
Real IPv4 EP : 172.16.20.1
```

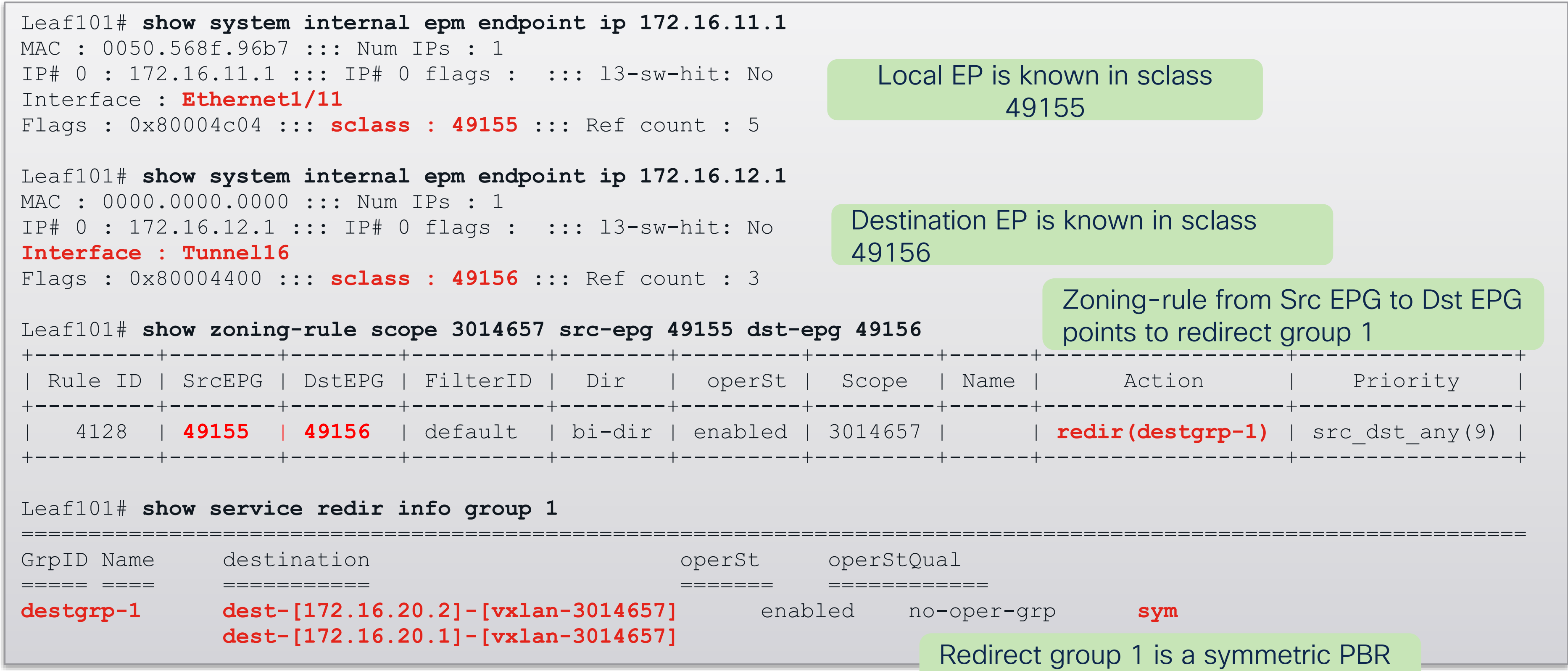
```
S1P1-Spine201# acidiag fnvread | egrep 10.0.0.67
       102           1          S1P1-Leaf102       FDO223007G7        10.0.0.67/32     leaf             active    0
```

FW VMAC in service BD VNID



BD Name	BD Alias	Class ID	Segment ID
BD1		49153	14909416
BD2		49154	15105997
Service-BD		32771	14843887

Example Check ingress leaf



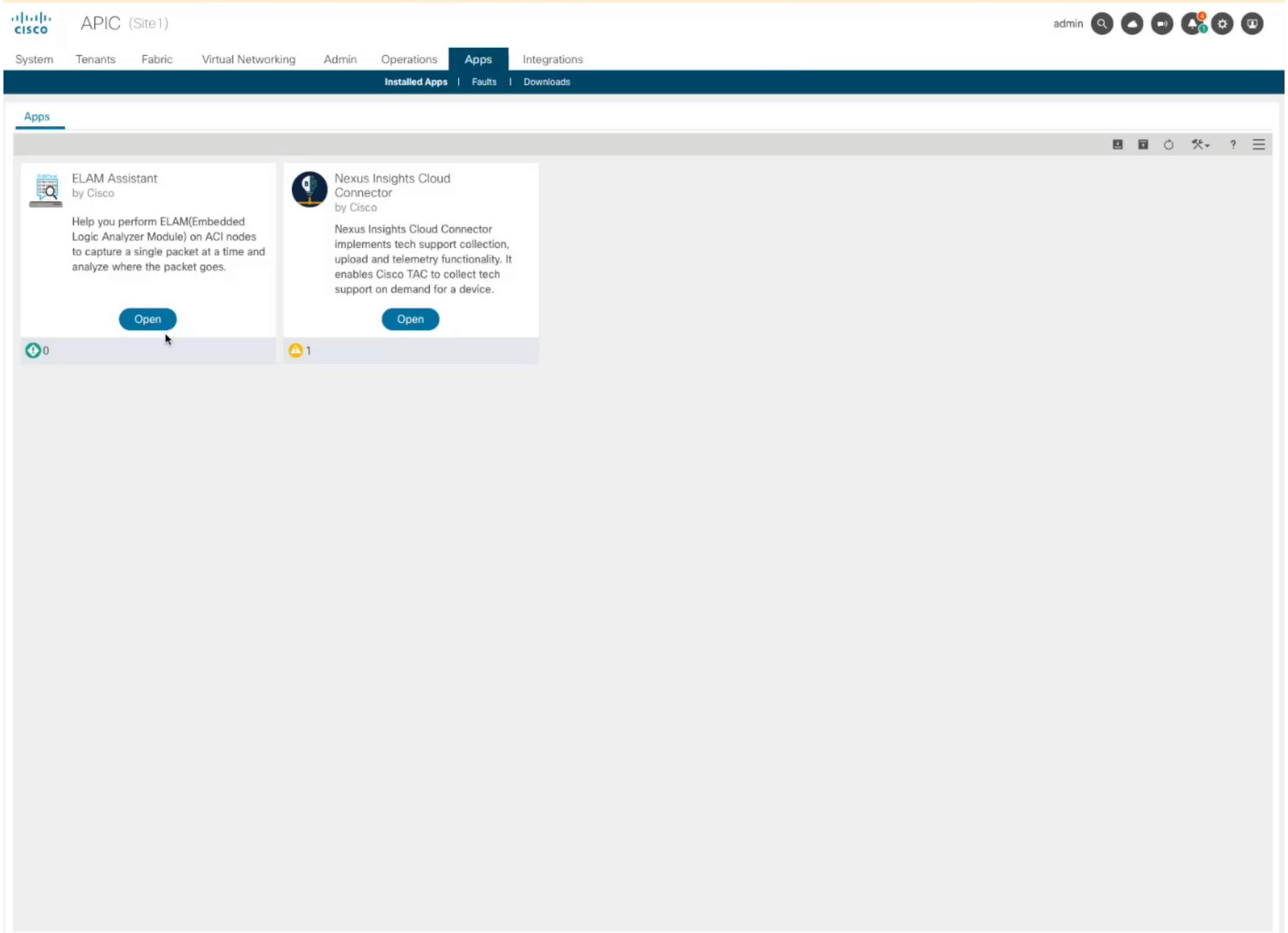
```
Leaf101# show system internal epm endpoint ip 172.16.11.1
MAC : 0050.568f.96b7 ::: Num IPs : 1
IP# 0 : 172.16.11.1 ::: IP# 0 flags :  ::: l3-sw-hit: No
Interface : Ethernet1/11
Flags : 0x80004c04 ::: sclass : 49155 ::: Ref count : 5

Leaf101# show system internal epm endpoint ip 172.16.12.1
MAC : 0000.0000.0000 ::: Num IPs : 1
IP# 0 : 172.16.12.1 ::: IP# 0 flags :  ::: l3-sw-hit: No
Interface : Tunnel16
Flags : 0x80004400 ::: sclass : 49156 ::: Ref count : 3

Leaf101# show zoning-rule scope 3014657 src-epg 49155 dst-epg 49156
+---------+--------+--------+----------+--------+---------+---------+------+-------------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  Dir   |  operSt |  Scope  | Name |       Action      |    Priority    |
+---------+--------+--------+----------+--------+---------+---------+------+-------------------+----------------+
|   4128  | 49155  | 49156  | default  | bi-dir | enabled | 3014657 |      | redir(destgrp-1)  | src_dst_any(9) |
+---------+--------+--------+----------+--------+---------+---------+------+-------------------+----------------+

Leaf101# show service redir info group 1
==========================================================================================================
GrpID Name      destination                          operSt      operStQual
===== ====      ===========                          ========    ============
destgrp-1       dest-[172.16.20.2]-[vxlan-3014657]      enabled     no-oper-grp       sym
                dest-[172.16.20.1]-[vxlan-3014657]
```

Local EP is known in sclass 49155

Destination EP is known in sclass 49156

Zoning-rule from Src EPG to Dst EPG points to redirect group 1

Redirect group 1 is a symmetric PBR with two destination IPs assigned

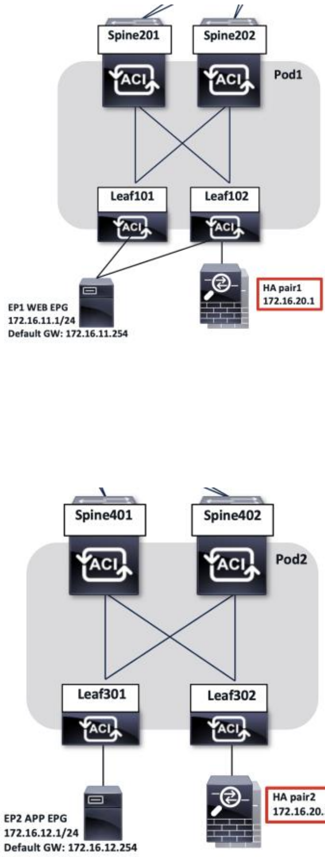


ELAM Assistant
by Cisco

Help you perform ELAM(Embedded Logic Analyzer Module) on ACI nodes to capture a single packet at a time and analyze where the packet goes.

Nexus Insights Cloud Connector
by Cisco

Nexus Insights Cloud Connector implements tech support collection, upload and telemetry functionality. It enables Cisco TAC to collect tech support on demand for a device.



Datapath Troubleshooting Tool:

ftriage from APIC CLI (Example SW Release 5.2(3))

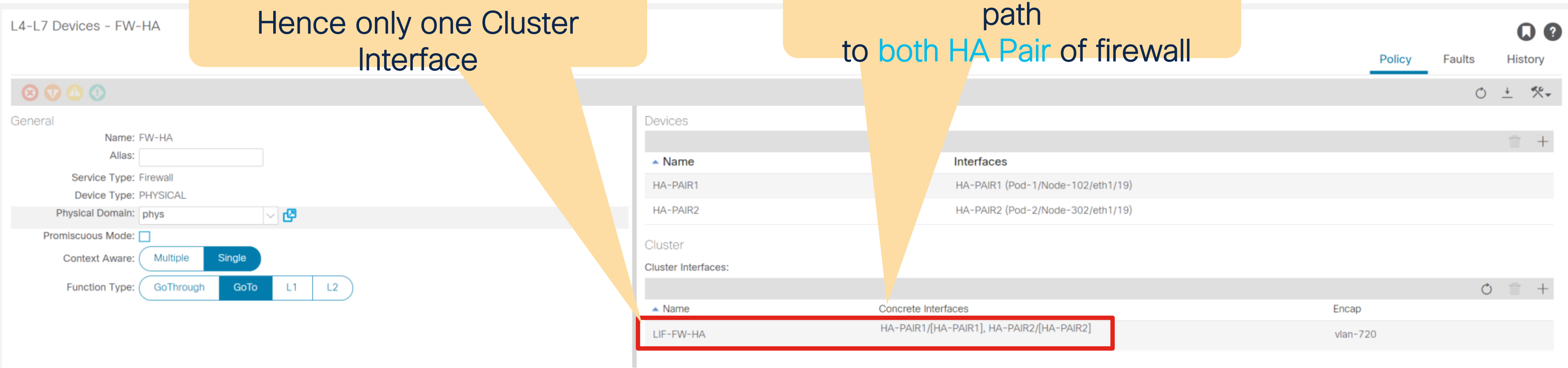
Before service device

```
Apic1# ftriage route -ii LEAF:101 -sip 172.16.11.2 -dip 172.16.12.2
2023-01-27 08:28:41,179 INFO     ftriage:     main:1295 L3 packet Seen on  S1P1-Leaf101 Ingress: Eth1/11 Egress: Eth1/49  Vnid: 14909416
2023-10-27 08:29:27,042 INFO     ftriage:  unicast:1543 S1P1-Leaf101: traffic is redirected to vnid:14843887 mac:00:EA:BD:07:3D:7C via tenant:RD-
     MPOD graph:EAST_WEST contract: ALLOW-ALL-PBR
2023-01-27 08:30:18,974 INFO     ftriage:     main:1333 S1P1-Spine201: Incoming Packet captured with Outer [SIP:10.0.0.67, DIP:10.0.72.65] ....
     Inner [SIP:172.16.11.2, DIP:172.16.12.2]
2023-01-27 08:31:28,056 INFO     ftriage:  unicast:2196 S1P1-Spine201: EP is known in COOP (DIPo = 10.0.0.67)
2023-01-27 08:31:41,494 INFO     ftriage:     main:958  Found peer-node S1P1-Leaf102 and IF: Eth1/49 in candidate list
2023-01-27 08:31:51,918 INFO     ftriage:       ep:128  S1P1-Leaf102: pbr traffic with dmac: 00:EA:BD:07:3D:7C
2023-01-27 08:32:06,748 INFO     ftriage:     main:1796 Packet is Exiting fabric with peer-device: POD1-router1 and peer-port: Ethernet1/19
2023-01-27 08:32:06,753 INFO     ftriage: acigraph:646   found matching devicenode:N1 ldev:FW-HA dev:HA-PAIR1HA-PAIR1uni/tn-RD-MPOD/lDevVip-FW-
     HA/cDev-HA-PAIR1/cIf-[HA-PAIR1]
2023-01-27 08:32:06,754 INFO     ftriage:  unicast:2739 S1P1-Leaf102:  PBR first pass is done and trafic is sent to service device: node:N1
     ldev:FW-HA dev:HA-PAIR1
2023-01-27 08:32:06,754 INFO     ftriage:  unicast:2741 S1P1-Leaf102:  expected traffic to return from: topology/pod-1/paths-102/pathep-[eth1/19]
     encap:720
```

After service device

```
2023-01-27 08:32:21,224 INFO     ftriage:     main:1821 pbr return path, nxt_nifs {S1P1-Leaf102: ['Eth1/19']}, nxt_dbg_f_n ig, nxt_inst ig, eg_ifs
     Eth1/19, Vnid: 720
2023-01-27 08:32:33,581 INFO     ftriage:     main:1295 L3 packet Seen on  S1P1-Leaf102 Ingress: Eth1/19 Egress: Eth1/49  Vnid: 3014657
2023-01-27 08:33:14,060 INFO     ftriage:     main:958  Found peer-node S1P1-Spine201 and IF: Eth1/2 in candidate list
```

Config Gotcha - L4/L7 devices for Symmetric PBR



Here we use one-arm
Hence only one Cluster Interface

Cluster interface contains path
to both HA Pair of firewall

L4-L7 Devices - FW-HA

General

Name: FW-HA

Alias:

Service Type: Firewall

Device Type: PHYSICAL

Physical Domain: phys

Promiscuous Mode: □

Context Aware: Multiple [✓ Single]

Function Type: GoThrough [✓ GoTo] L1 L2

Devices

Name	Interfaces
HA-PAIR1	HA-PAIR1 (Pod-1/Node-102/eth1/19)
HA-PAIR2	HA-PAIR2 (Pod-2/Node-302/eth1/19)

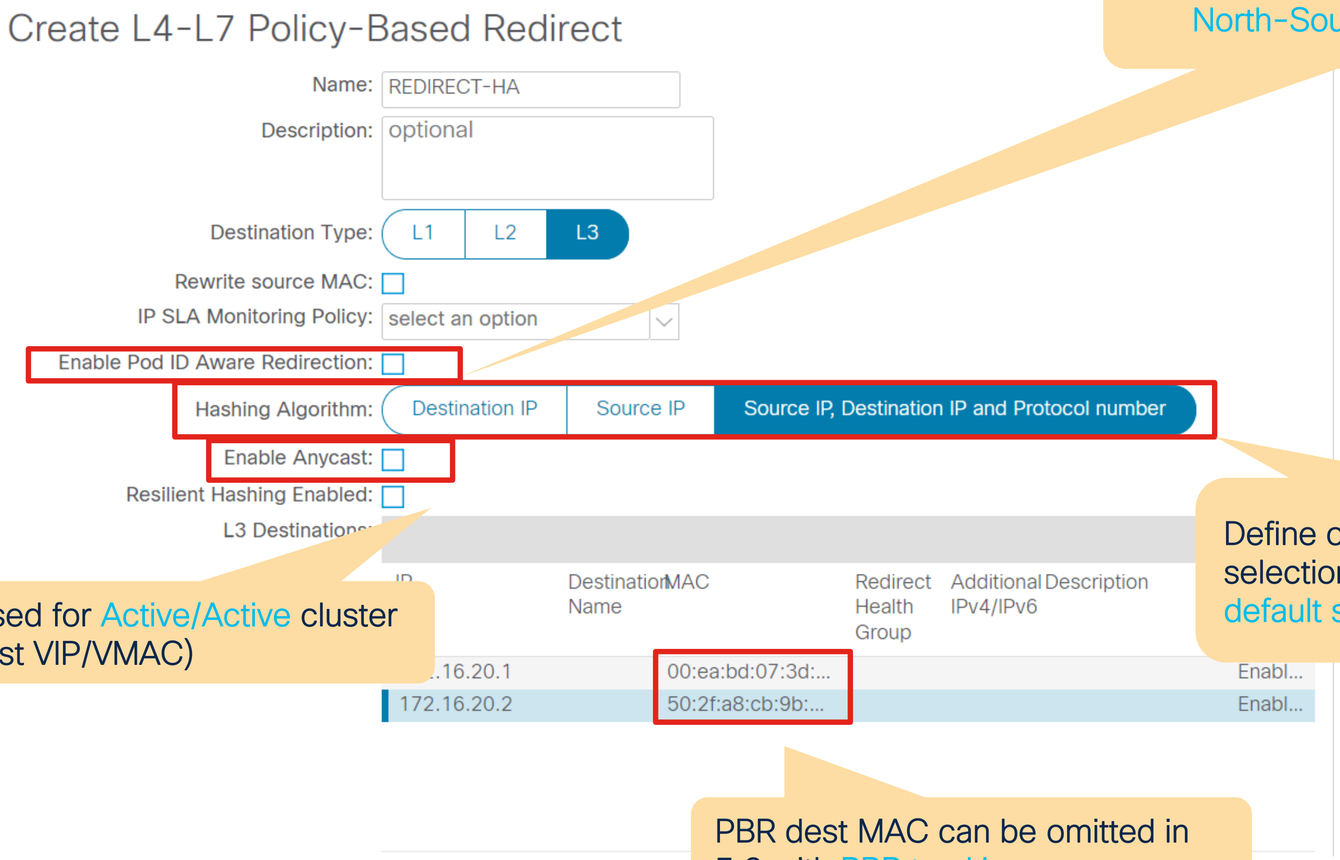
Cluster

Cluster Interfaces:

Name	Concrete Interfaces	Encap
LIF-FW-HA	HA-PAIR1/[HA-PAIR1], HA-PAIR2/[HA-PAIR2]	vlan-720

Note a single L4/L7 cluster interface is representing both HA pair
Symmetric PBR will select HA Pair1 or HA Pair2 based on hashing

Config Gotcha - Redirect policy



Create L4-L7 Policy-Based Redirect

Name: REDIRECT-HA

Description: optional

Destination Type: L1 | L2 | [✓ L3]

Rewrite source MAC: □

IP SLA Monitoring Policy: select an option

Enable Pod ID Aware Redirection: □

Hashing Algorithm: Destination IP | Source IP | [✓ Source IP, Destination IP and Protocol number]

Enable Anycast: □

Resilient Hashing Enabled: □

L3 Destinations:

IP	Destination Name	MAC	Redirect Health Group	Additional IPv4/IPv6	Description	
.16.20.1		00:ea:bd:07:3d:...				Enabl...
172.16.20.2		50:2f:a8:cb:9b:...				Enabl...

Should only be considered in North-South PBR scenario

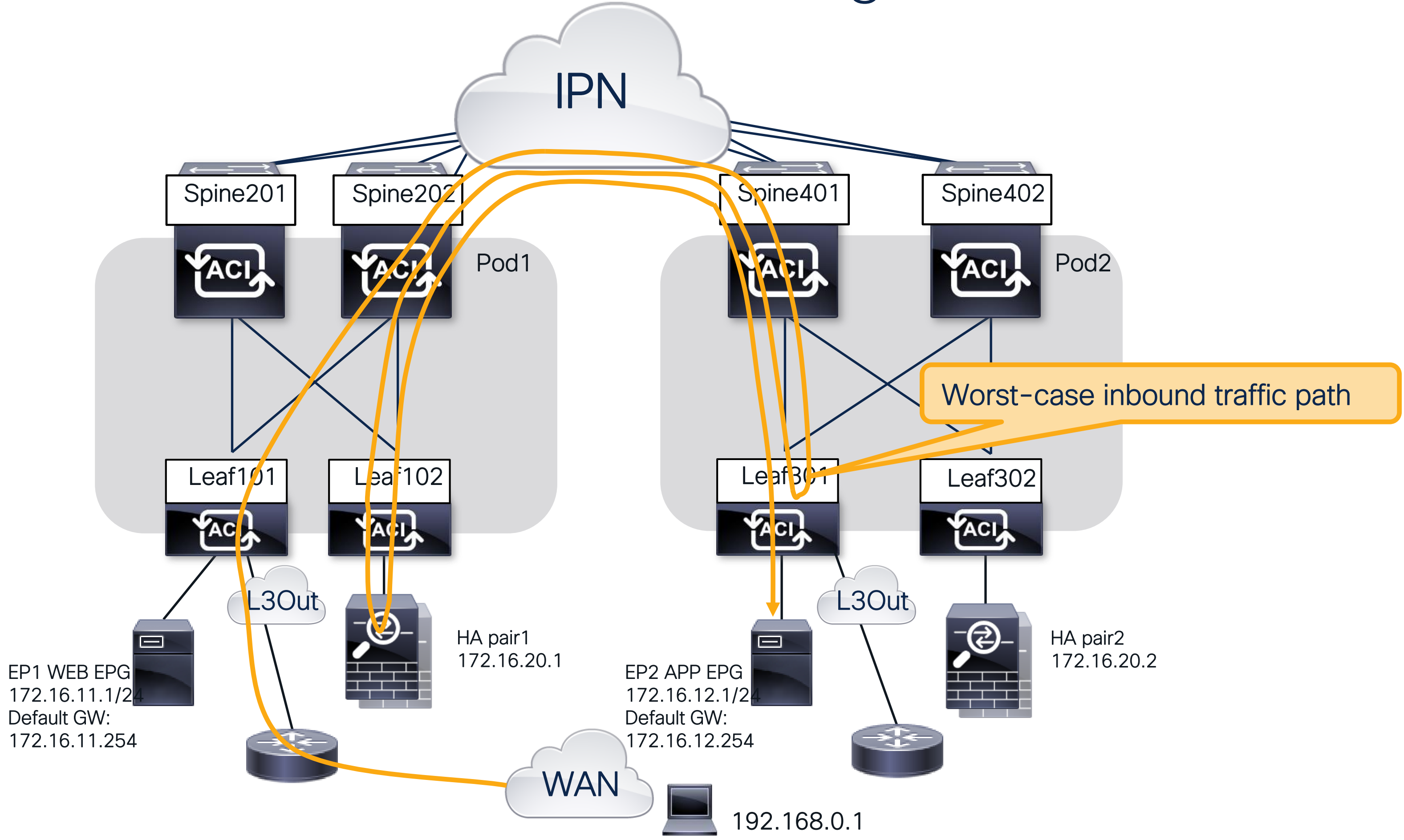
Define our hash for PBR next-hop selection (recommended to keep default src/dst/proto)

Only used for Active/Active cluster (Anycast VIP/VMAC)

PBR dest MAC can be omitted in 5.2 with PBR tracking

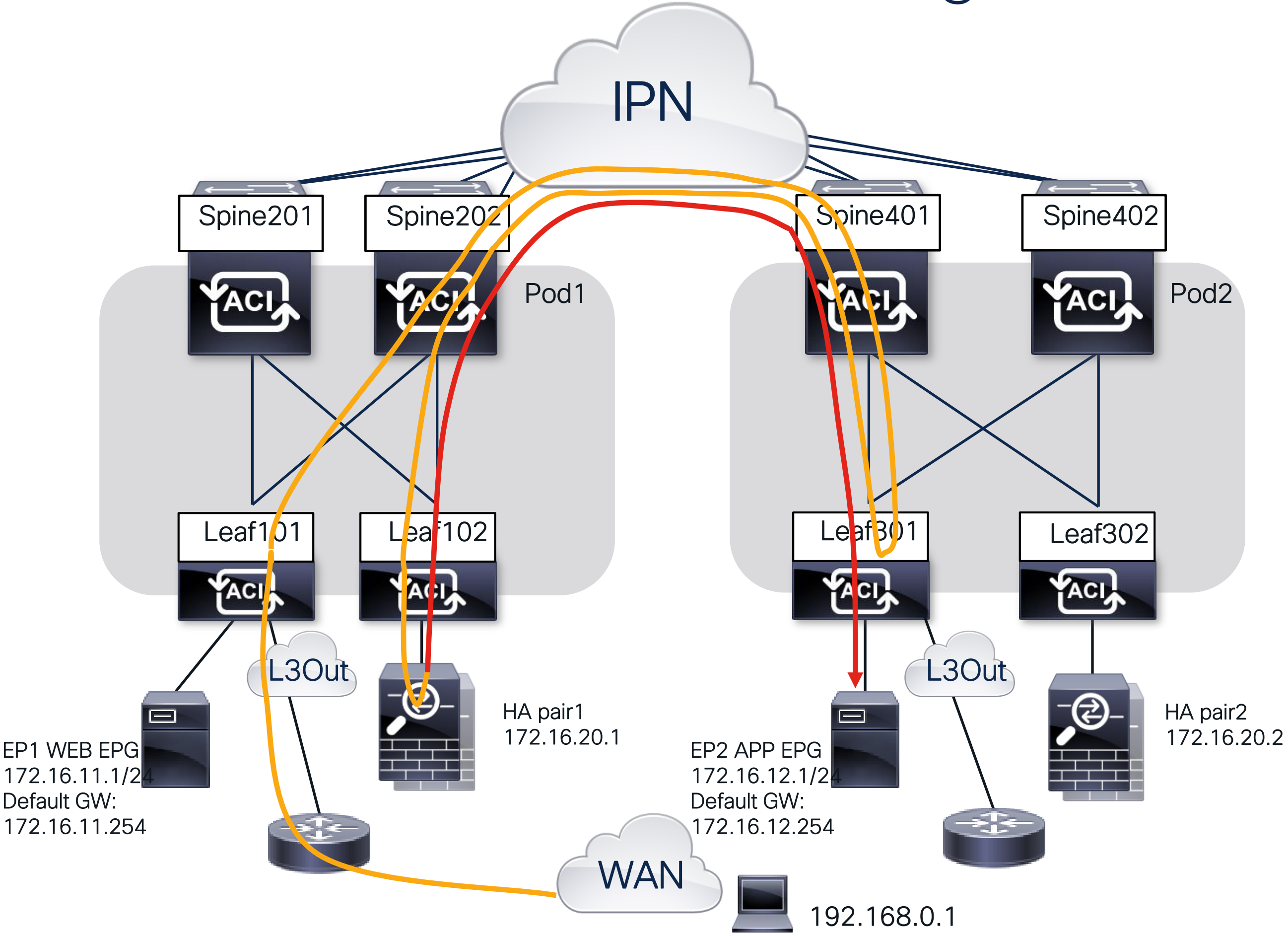
Multipod North-South Location-based PBR

Multipod North-South PBR – Challenge

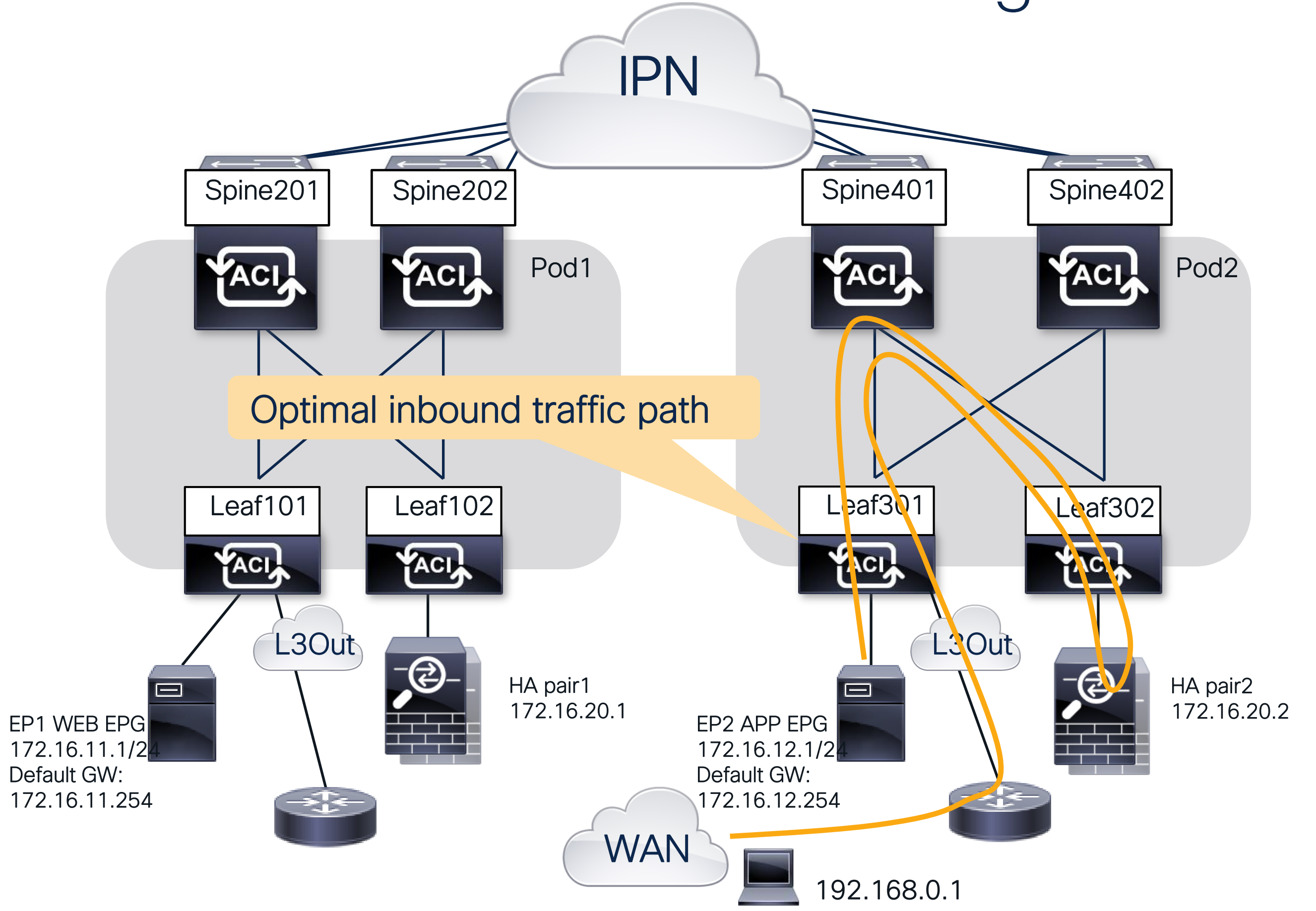


Worst-case inbound traffic path

Multipod North-South PBR – Challenge



Multipod North-South PBR – Challenge

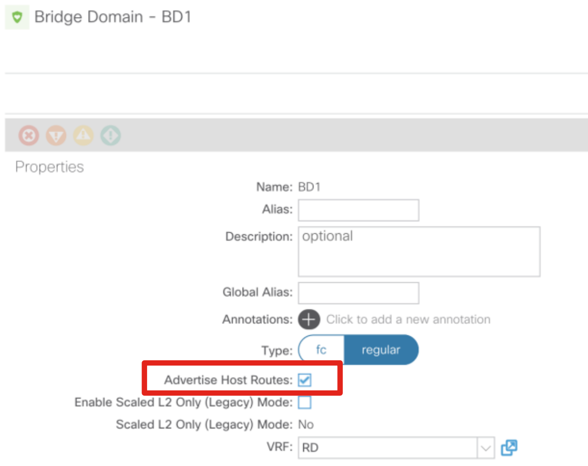


Optimal inbound traffic path

Multipod North-South PBR

Enable Host Route Advertisement

Starting 4.x we can configure an BD to advertise /32 host routes for Pod local Endpoints on its L3Out.



Bridge Domain - BD1

Properties

Name: BD1

Alias:

Description: optional

Global Alias:

Annotations: Click to add a new annotation

Type: fc [✓ regular]

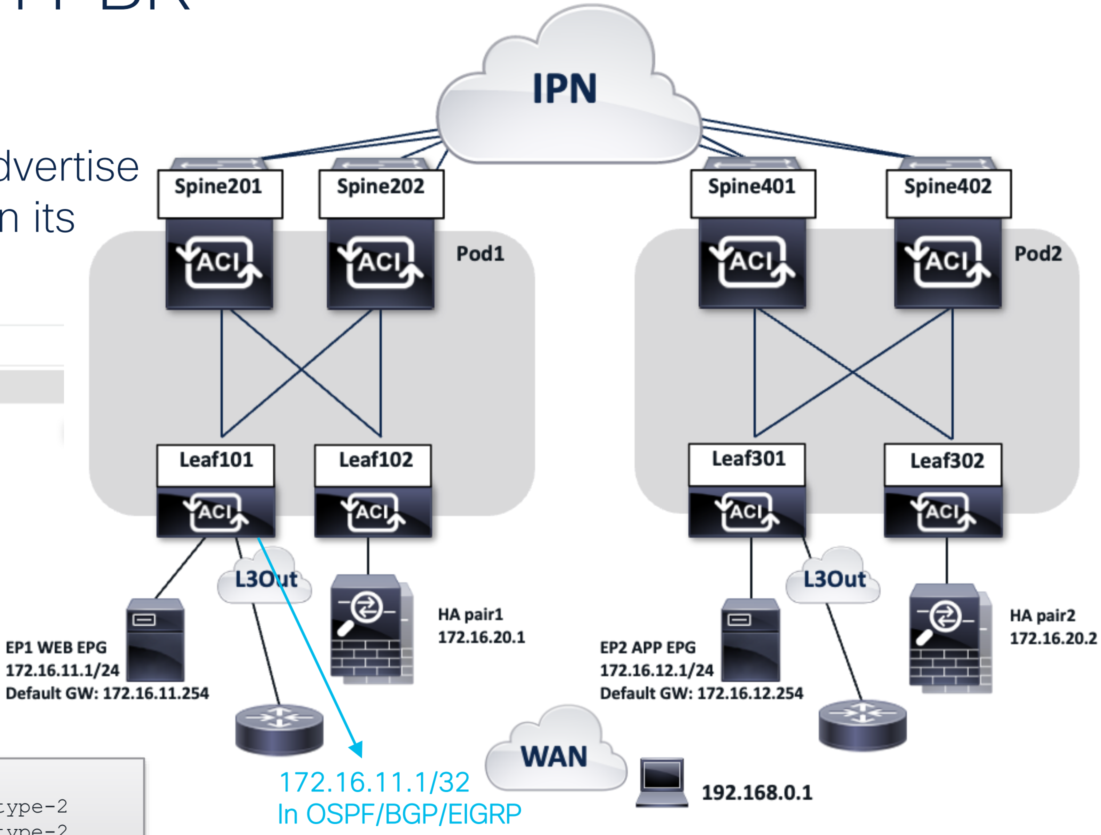
Advertise Host Routes: ☑

Enable Scaled L2 Only (Legacy) Mode: □

Scaled L2 Only (Legacy) Mode: No

VRF: RD

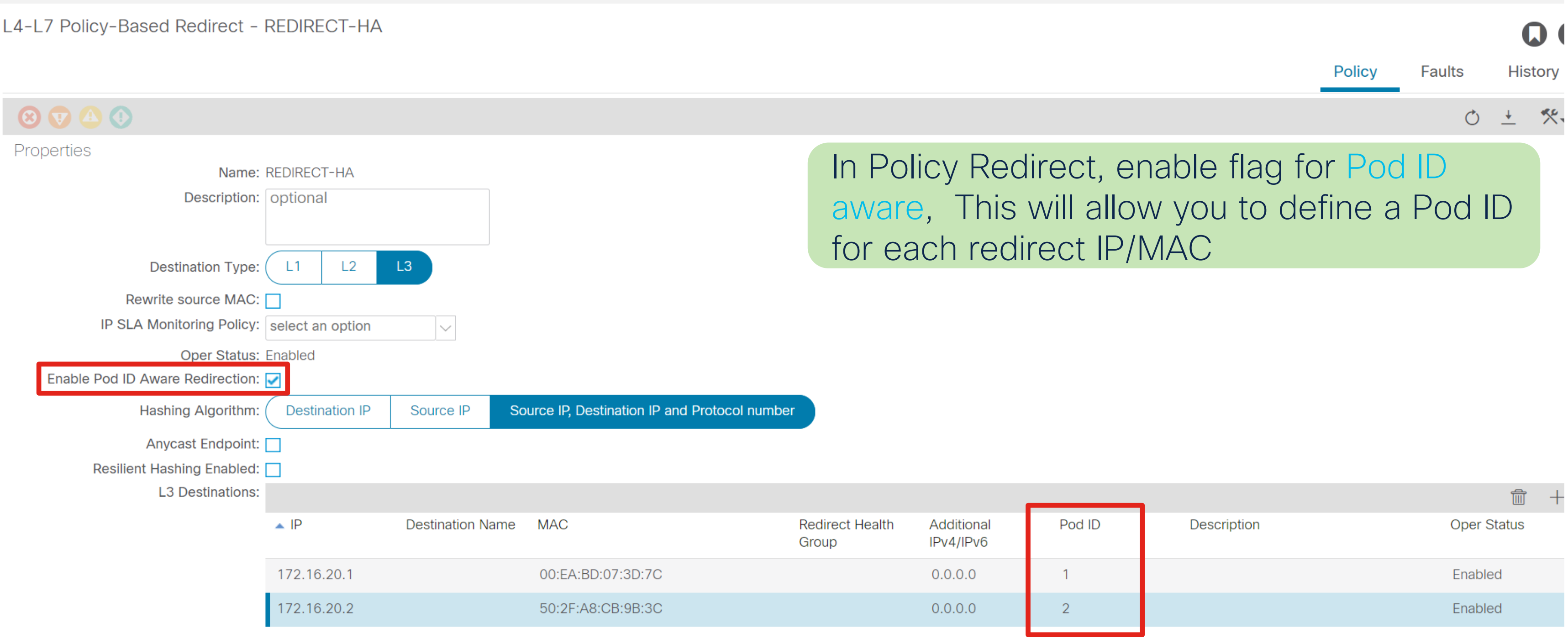
```
172.16.11.0/24, ubest/mbest: 2/0
    *via 192.168.1.1, Vlan920, [110/20], 00:01:44, ospf-1, type-2
    *via 192.168.1.3, Vlan920, [110/20], 00:01:44, ospf-1, type-2
172.16.11.1/32, ubest/mbest: 1/0
    *via 192.168.1.1, Vlan920, [110/1], 01:43:18, ospf-1, type-2
```



172.16.11.1/32
In OSPF/BGP/EIGRP

Multipod North-South PBR

Enable Pod ID Aware Redirection



L4-L7 Policy-Based Redirect - REDIRECT-HA

Properties

Name: REDIRECT-HA

Description: optional

Destination Type: L1 L2 [✓ L3]

Rewrite source MAC: □

IP SLA Monitoring Policy: select an option

Oper Status: Enabled

Enable Pod ID Aware Redirection: ☑

Hashing Algorithm: Destination IP Source IP [✓ Source IP, Destination IP and Protocol number]

Anycast Endpoint: □

Resilient Hashing Enabled: □

L3 Destinations:

IP	Destination Name	MAC	Redirect Health Group	Additional IPv4/IPv6	Pod ID	Description	Oper Status
172.16.20.1		00:EA:BD:07:3D:7C		0.0.0.0	1		Enabled
172.16.20.2		50:2F:A8:CB:9B:3C		0.0.0.0	2		Enabled

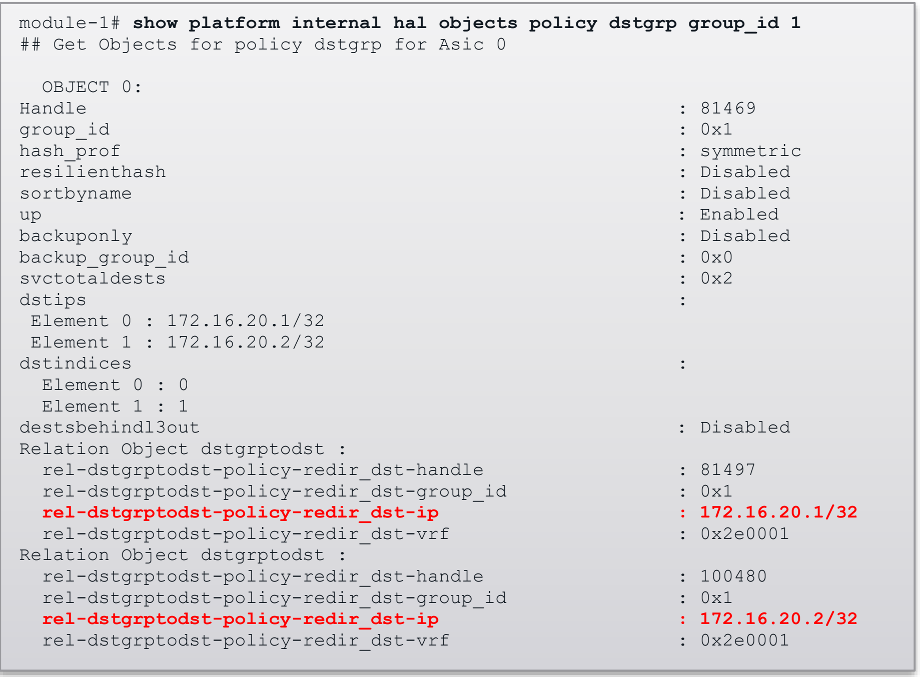
In Policy Redirect, enable flag for Pod ID aware, This will allow you to define a Pod ID for each redirect IP/MAC

Note that there are no visible changes in service redir info . However HAL will show the change

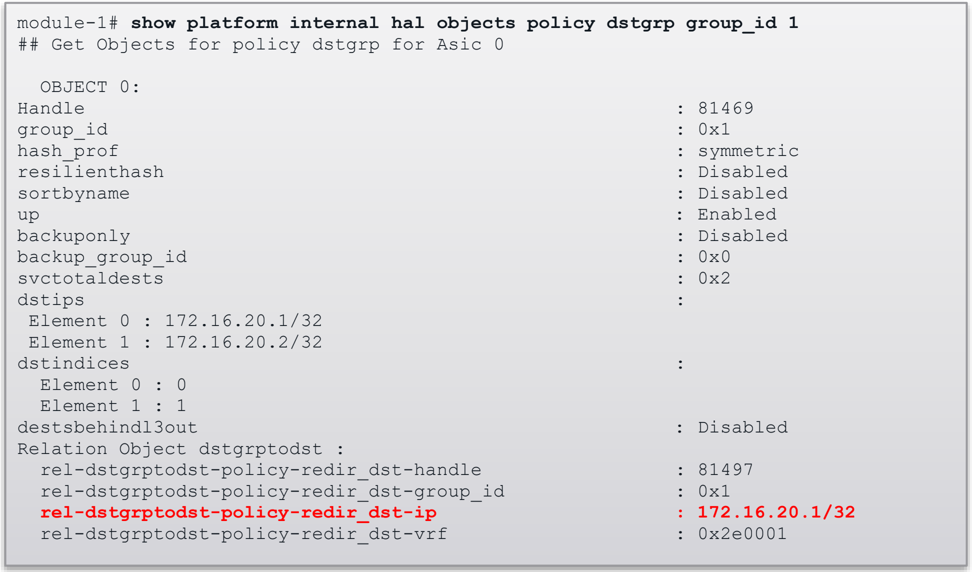
Changes in hardware (Leaf Pod1 shown)

Before enabling Pod aware redirection
On leaf 101 we see both redirect destinations
(group id comes from zoning-rule)

After enabling Pod aware
On leaf 101 we only see local 172.16.20.1
In the hash list



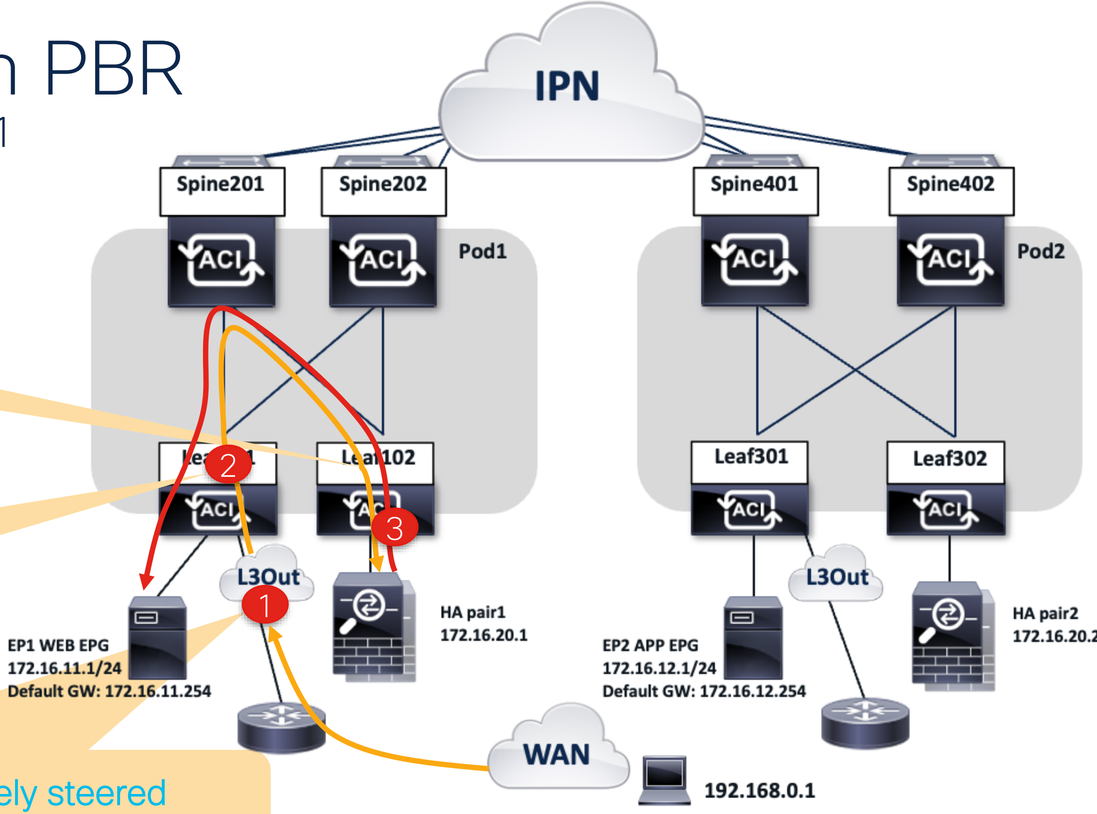
```
module-1# show platform internal hal objects policy dstgrp group_id 1
## Get Objects for policy dstgrp for Asic 0

  OBJECT 0:
Handle                                                      : 81469
group_id                                                    : 0x1
hash_prof                                                   : symmetric
resilienthash                                               : Disabled
sortbyname                                                  : Disabled
up                                                          : Enabled
backuponly                                                  : Disabled
backup_group_id                                             : 0x0
svctotaldests                                               : 0x2
dstips                                                      :
 Element 0 : 172.16.20.1/32
 Element 1 : 172.16.20.2/32
dstindices                                                  :
  Element 0 : 0
  Element 1 : 1
destsbehindl3out                                            : Disabled
Relation Object dstgrptodst :
  rel-dstgrptodst-policy-redir_dst-handle                   : 81497
  rel-dstgrptodst-policy-redir_dst-group_id                 : 0x1
  rel-dstgrptodst-policy-redir_dst-ip                       : 172.16.20.1/32
  rel-dstgrptodst-policy-redir_dst-vrf                      : 0x2e0001
Relation Object dstgrptodst :
  rel-dstgrptodst-policy-redir_dst-handle                   : 100480
  rel-dstgrptodst-policy-redir_dst-group_id                 : 0x1
  rel-dstgrptodst-policy-redir_dst-ip                       : 172.16.20.2/32
  rel-dstgrptodst-policy-redir_dst-vrf                      : 0x2e0001
```



```
module-1# show platform internal hal objects policy dstgrp group_id 1
## Get Objects for policy dstgrp for Asic 0

  OBJECT 0:
Handle                                                      : 81469
group_id                                                    : 0x1
hash_prof                                                   : symmetric
resilienthash                                               : Disabled
sortbyname                                                  : Disabled
up                                                          : Enabled
backuponly                                                  : Disabled
backup_group_id                                             : 0x0
svctotaldests                                               : 0x2
dstips                                                      :
 Element 0 : 172.16.20.1/32
 Element 1 : 172.16.20.2/32
dstindices                                                  :
  Element 0 : 0
  Element 1 : 1
destsbehindl3out                                            : Disabled
Relation Object dstgrptodst :
  rel-dstgrptodst-policy-redir_dst-handle                   : 81497
  rel-dstgrptodst-policy-redir_dst-group_id                 : 0x1
  rel-dstgrptodst-policy-redir_dst-ip                       : 172.16.20.1/32
  rel-dstgrptodst-policy-redir_dst-vrf                      : 0x2e0001
```

Multipod North-South PBR

Packet flow from External to 172.16.11.1



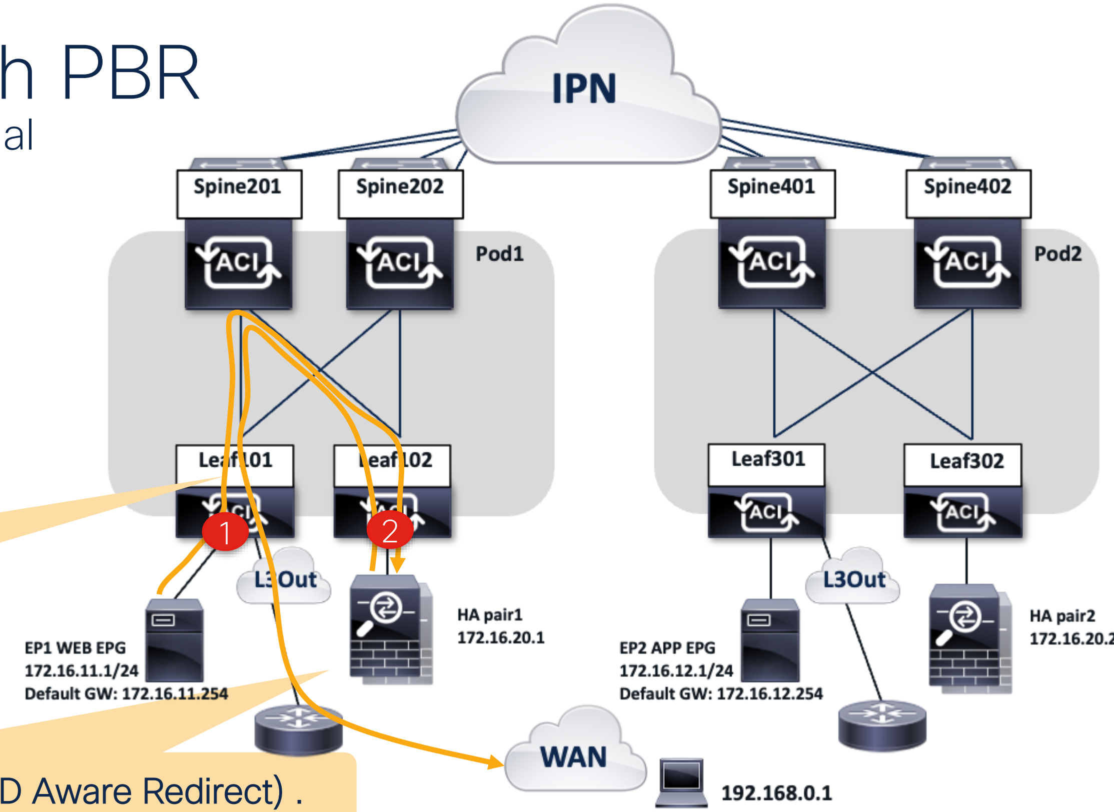
3 Traffic from HA pair1 goes to EP1 on Leaf 101 (permit rule)

2 Ingress packet on an L3Out is redirected on the egress Leaf. Leaf 101 will redirect the traffic ALWAYS to HA pair1 (Pod ID Aware redirect)

1 Traffic originating from an external client is selectively steered towards the Pod on which the destination EP resides.

Multipod North-South PBR

Return Path from 172.16.11.1 to External



2 Routing lookup occurs on service leaf 102 for external client IP.

1 Ingress leaf 101 redirects to Local HA pair1 (Pod ID Aware Redirect) .

Multipod North-South PBR – Optimization

How to avoid this hair pinning across the IPN ?

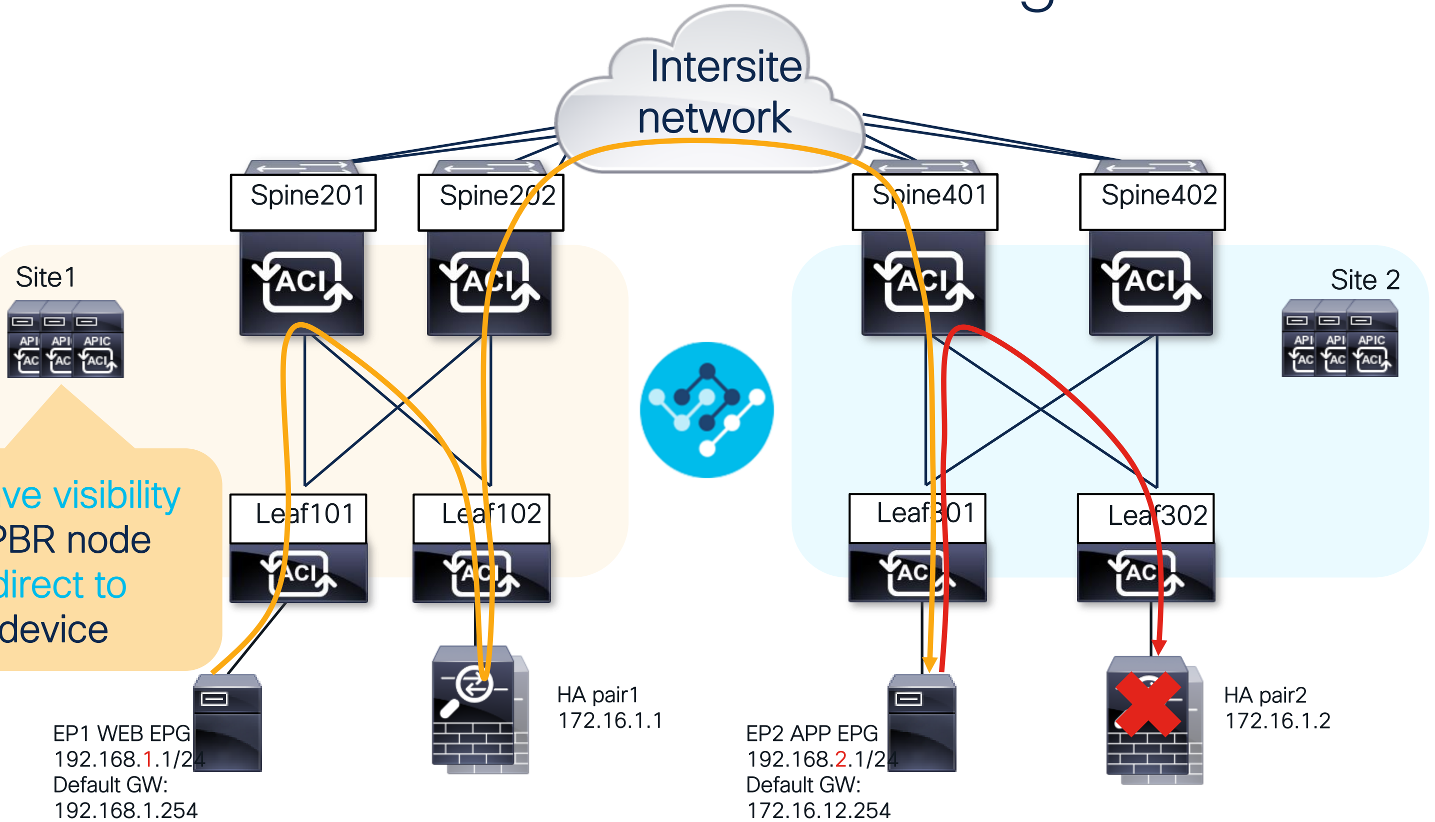
Host based routing (HBR) (4.0 and plus)
Location Aware PBR (3.1 and plus)

- If we have multiple PBR service nodes, it's load-balanced based on Source IP, Destination IP and Protocol Type by default. Hash tuple is configurable, but we don't have capability to select local PBR service node. In 3.1, we have option to prefer local pod PBR node (multipod fabric only)
- It is recommended (not mandatory) that Location aware PBR be used for North-South firewall integration with host route advertisement.
- **Location aware PBR CANNOT be used for EAST-WEST** traffic, this will lead to asymmetric forwarding (each flow direction using a different FW pair)
- It can't be used for Transit Routing (L3out to L3out).

Multi-Site PBR

Multi-Site East-West PBR

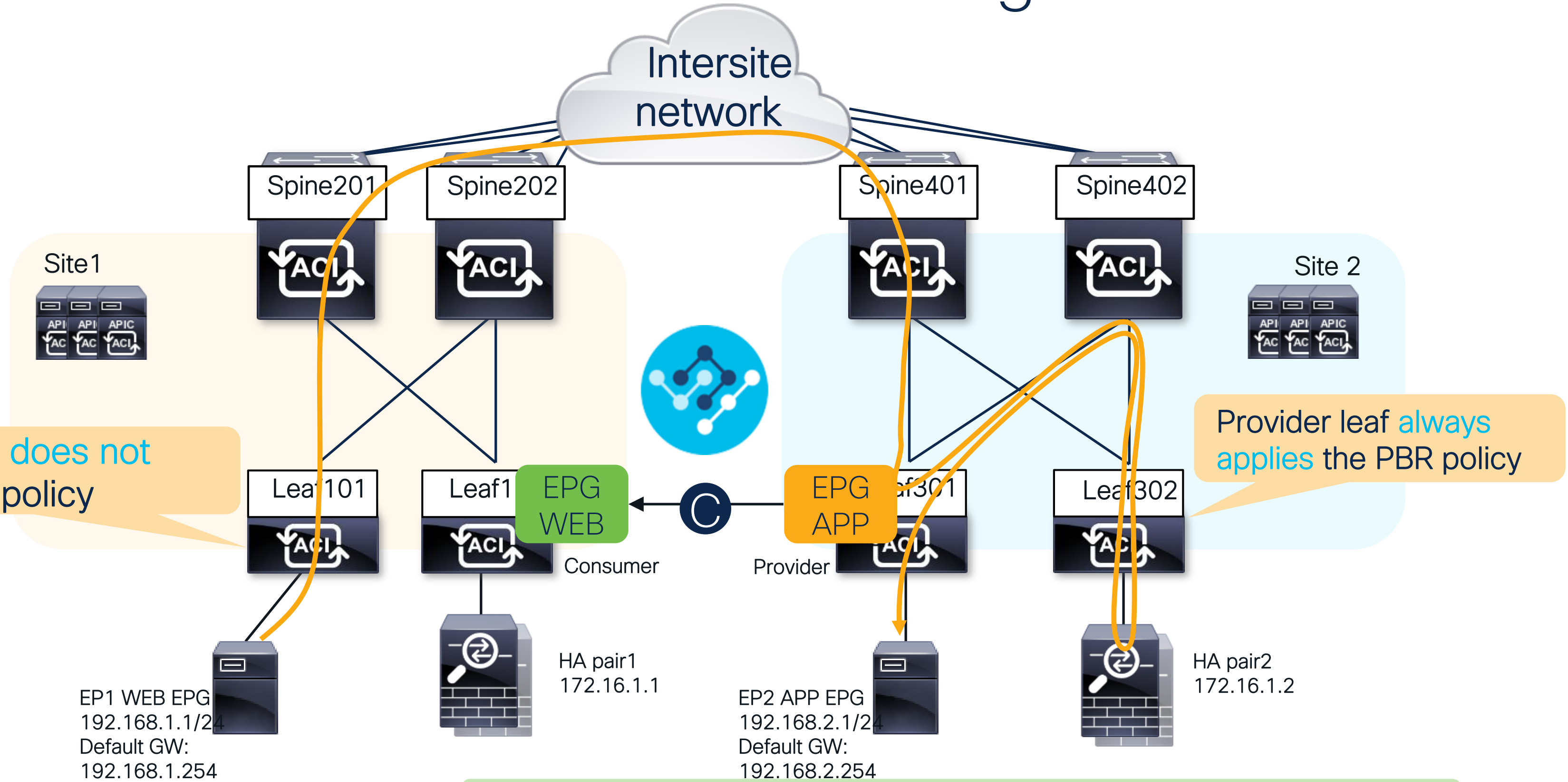
Multi-Site PBR – East-West Challenge



APICs do not have visibility on remote site PBR node and can only redirect to local site L4/L7 device

How can we ensure redirect is symmetric (same site in both directions)?

Multi-Site PBR – East-West Challenge

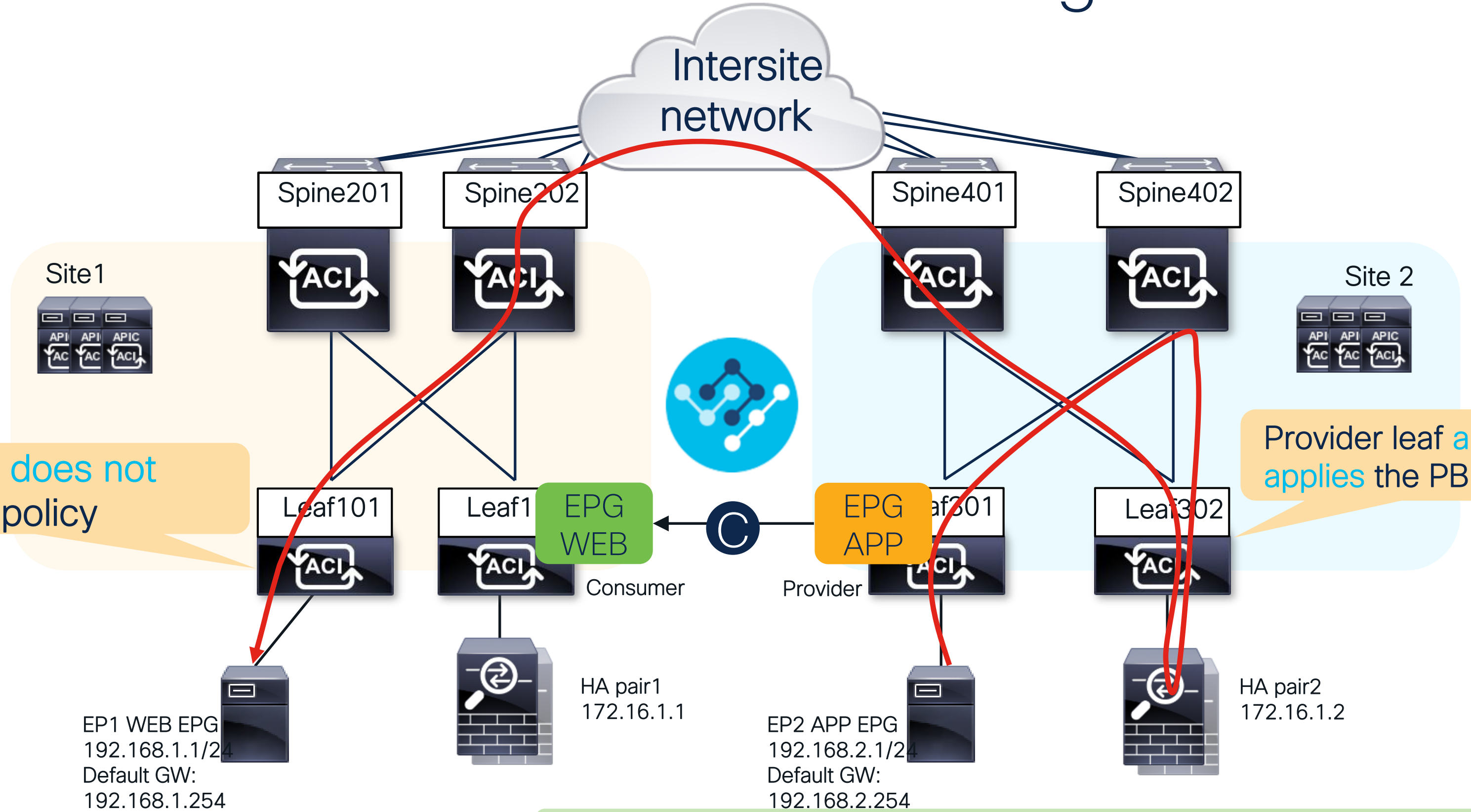


Consumer leaf does not apply the PBR policy

Provider leaf always applies the PBR policy

IP subnet must be configured under the consumer EPG.

Multi-Site PBR – East -West Challenge

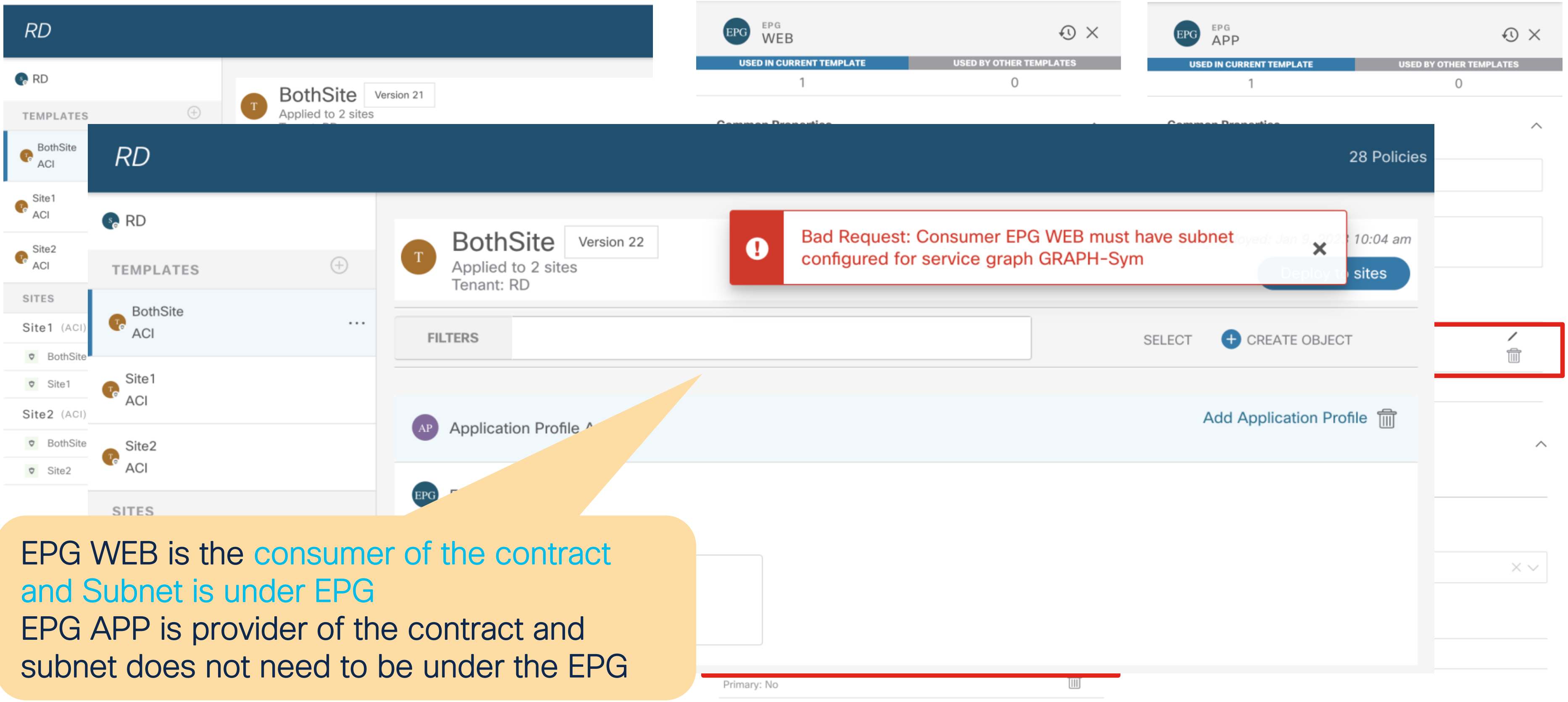


Consumer leaf does not apply the PBR policy

Provider leaf always applies the PBR policy

IP subnet must be configured under the consumer EPG.

Config Gotcha Multi-Site PBR – East-West

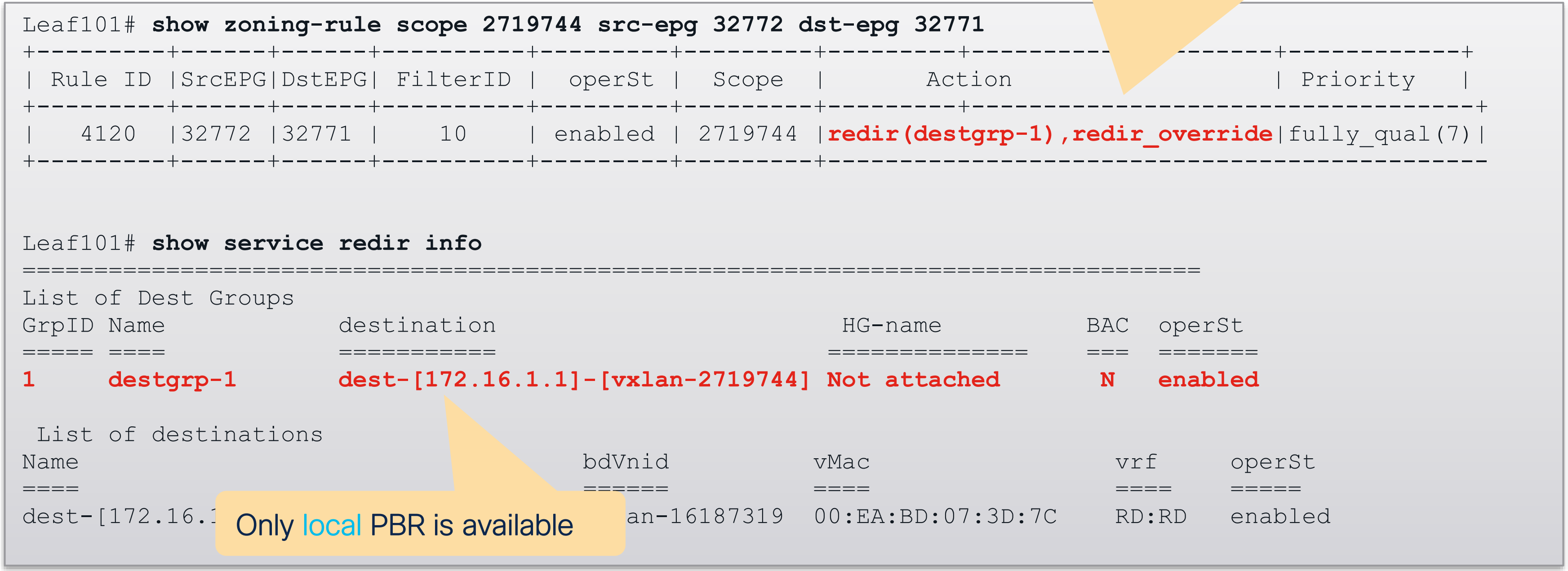


Bad Request: Consumer EPG WEB must have subnet configured for service graph GRAPH-Sym

EPG WEB is the consumer of the contract and Subnet is under EPG
EPG APP is provider of the contract and subnet does not need to be under the EPG

Consumer to Provider
Ingress Consumer leaf zoning-rule – site1

Unless the destination EP is local redir_override rule will be used(bypass PBR and do not mark policy)

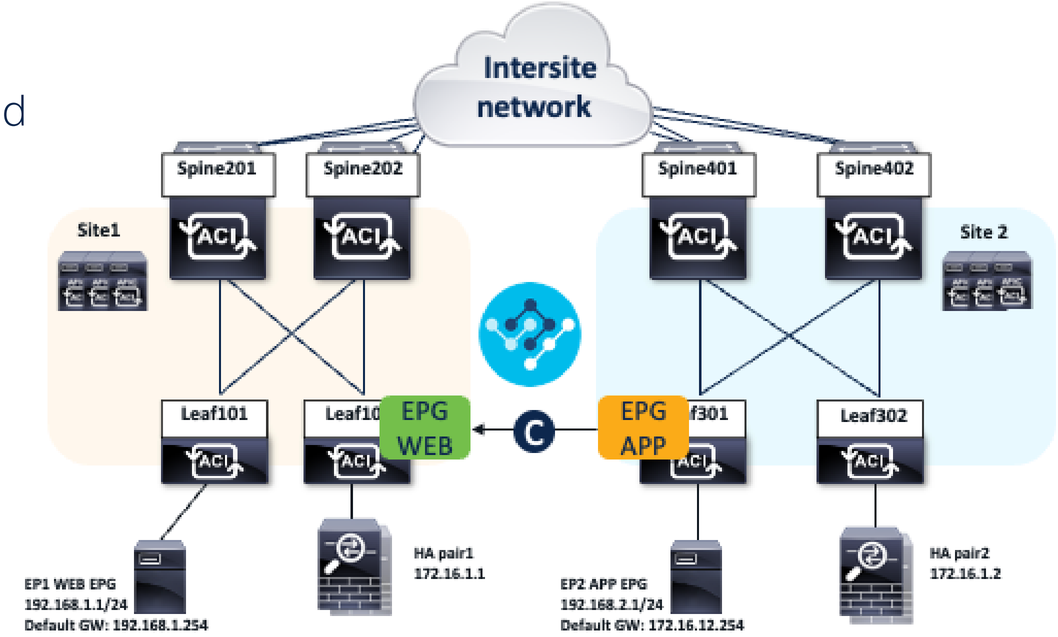


```
Leaf101# show zoning-rule scope 2719744 src-epg 32772 dst-epg 32771
+---------+------+------+----------+---------+---------+---------+-----------           -------+------------+
| Rule ID |SrcEPG|DstEPG| FilterID |  operSt |  Scope  |      Action                          | Priority   |
+---------+------+------+----------+---------+---------+---------+-----------------------------------------+
|   4120  |32772 |32771 |    10    | enabled | 2719744 |redir(destgrp-1),redir_override|fully_qual(7)|
+---------+------+------+----------+---------+---------+------------------------------------------------


Leaf101# show service redir info
==================================================================================
List of Dest Groups
GrpID Name          destination                       HG-name          BAC  operSt
===== ====          ===========                       ==============   ===  =======
1     destgrp-1     dest-[172.16.1.1]-[vxlan-2719744] Not attached      N   enabled

 List of destinations
Name                                       bdVnid         vMac                  vrf      operSt
====                                       ======         ====                  ====     =====
dest-[172.16.1                             an-16187319    00:EA:BD:07:3D:7C     RD:RD    enabled
```

Only local PBR is available

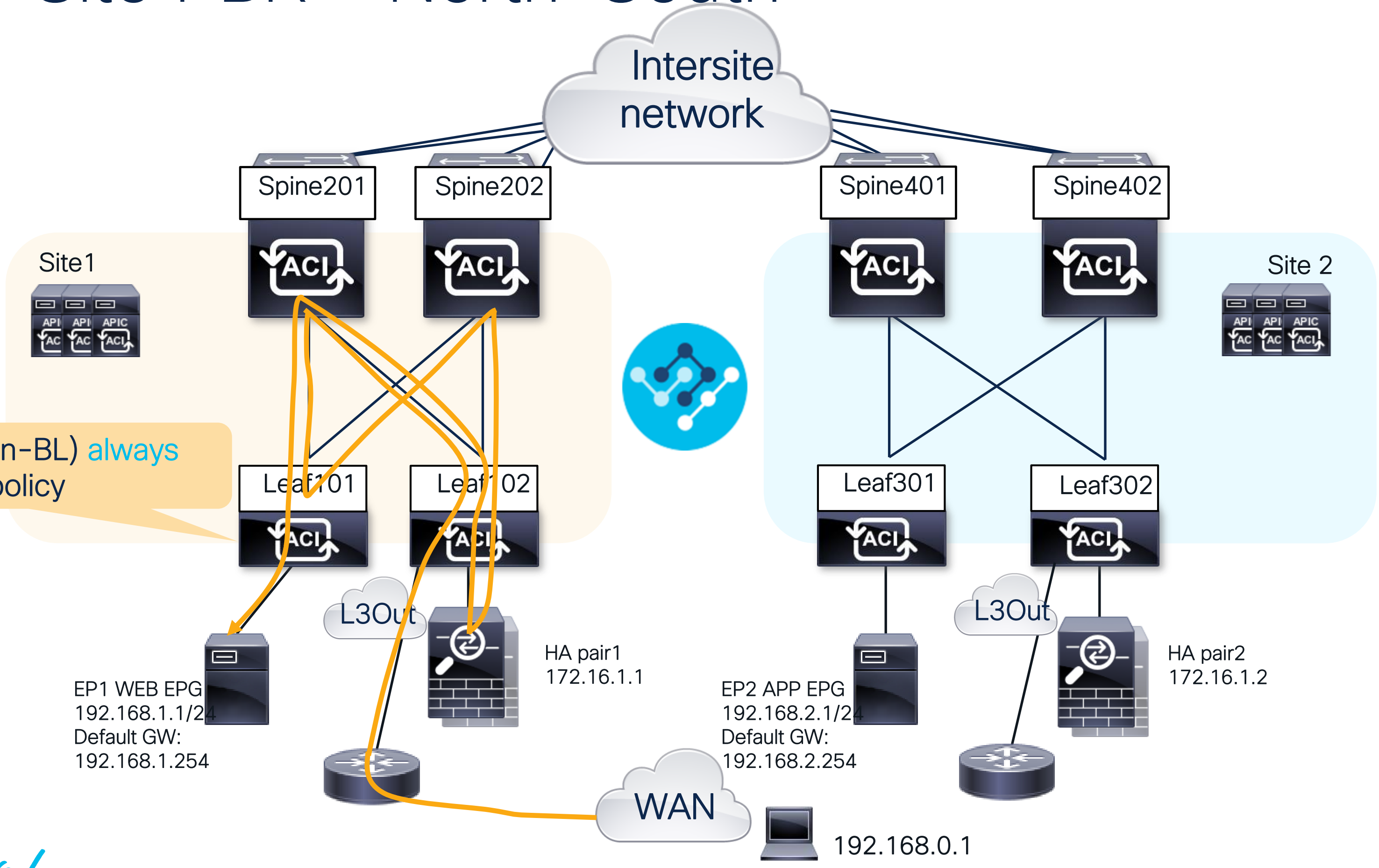
Multi-Site PBR – East-West

- Multisite PBR requirement
  - ✓ Consumer subnet must be configured under the consumer EPG
  - A site can only redirect to site local PBR Devices
- Rule: we need to go through the same Firewall pair in both directions
- Solution:
  - Redirect happens on the site where
    the provider endpoint is.



Multi-Site North-South PBR

Multi-Site PBR – North-South



Compute leaf (non-BL) always applies the PBR policy

Multi-Site PBR – North-South

Endpoint to L3Out

3

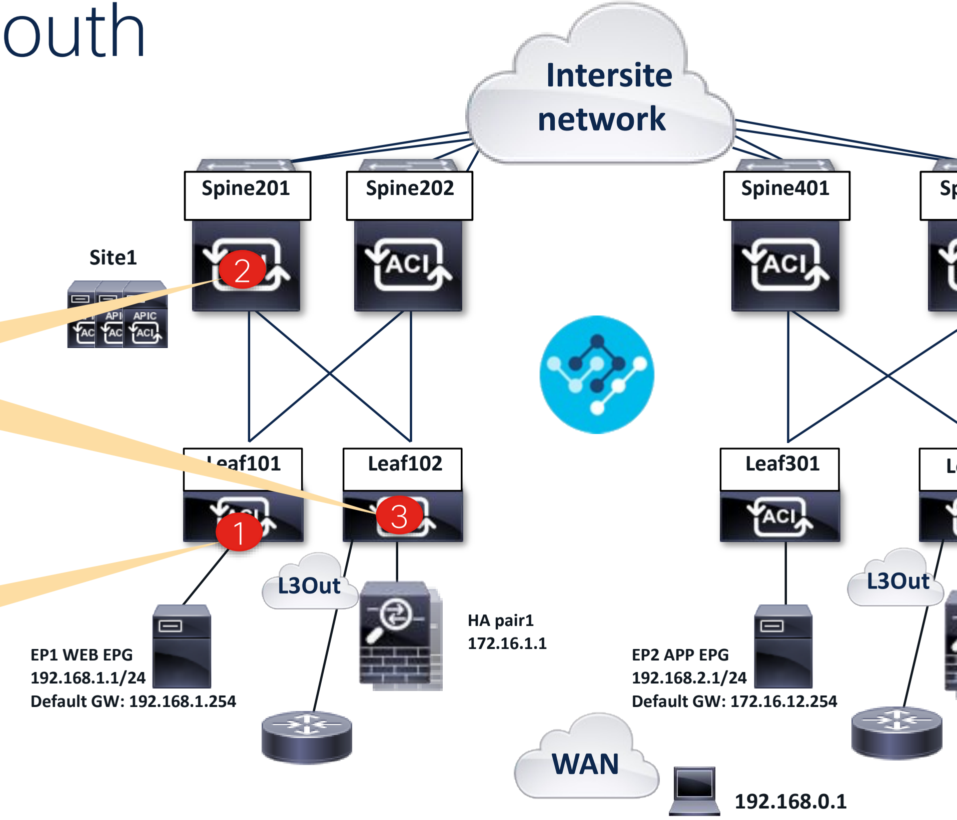
Back from FW a permit rule allows to reach L3Out which may be local (likely) or remote site L3Out depending on routing table.

2

Spine performs COOP lookup for VMAC in service BD and sends it to service leaf, who forwards it to FW

1

Ingress leaf will always apply redirect (dclass or dest pcTag from zoning prefix).
No override rule from EP to L3 out.
Redirect will be to local site HA pair (HA pair1)

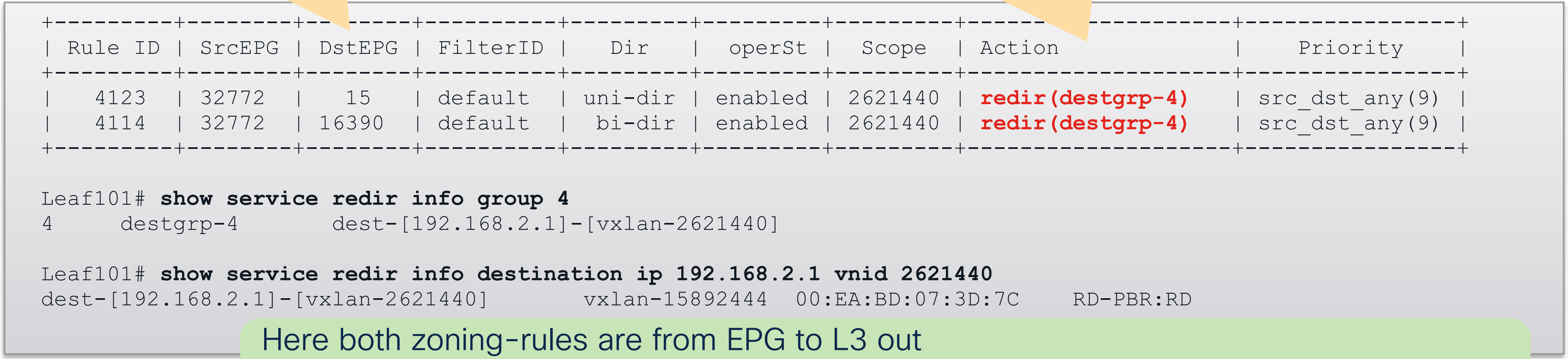


Multi-Site PBR – North-South

Zoning Rule EPG to L3Out on compute leaf

Compute leaf dclass to reach L3Out will either by 15 (0.0.0.0/0 prefix) or external EPG pcTag (specific prefix here 16390).
As VRF enforcement is ingress, dclass is always known

In all case the rule is always redirect with no option override → redirect always apply on this leaf



```
+---------+--------+--------+----------+---------+---------+---------+--------------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |   Dir   |  operSt |  Scope  | Action             |    Priority    |
+---------+--------+--------+----------+---------+---------+---------+--------------------+----------------+
|   4123  | 32772  |   15   | default  | uni-dir | enabled | 2621440 | redir(destgrp-4)   | src_dst_any(9) |
|   4114  | 32772  | 16390  | default  |  bi-dir | enabled | 2621440 | redir(destgrp-4)   | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+---------+--------------------+----------------+

Leaf101# show service redir info group 4
4     destgrp-4       dest-[192.168.2.1]-[vxlan-2621440]

Leaf101# show service redir info destination ip 192.168.2.1 vnid 2621440
dest-[192.168.2.1]-[vxlan-2621440]       vxlan-15892444  00:EA:BD:07:3D:7C    RD-PBR:RD
```

Here both zoning-rules are from EPG to L3 out
We will use one or the other depending on the zoning-rule subnet in the external EPG (0.0.0.0/0 or specific subnet)

Multi-Site PBR – North-South

L3out to endpoint

3

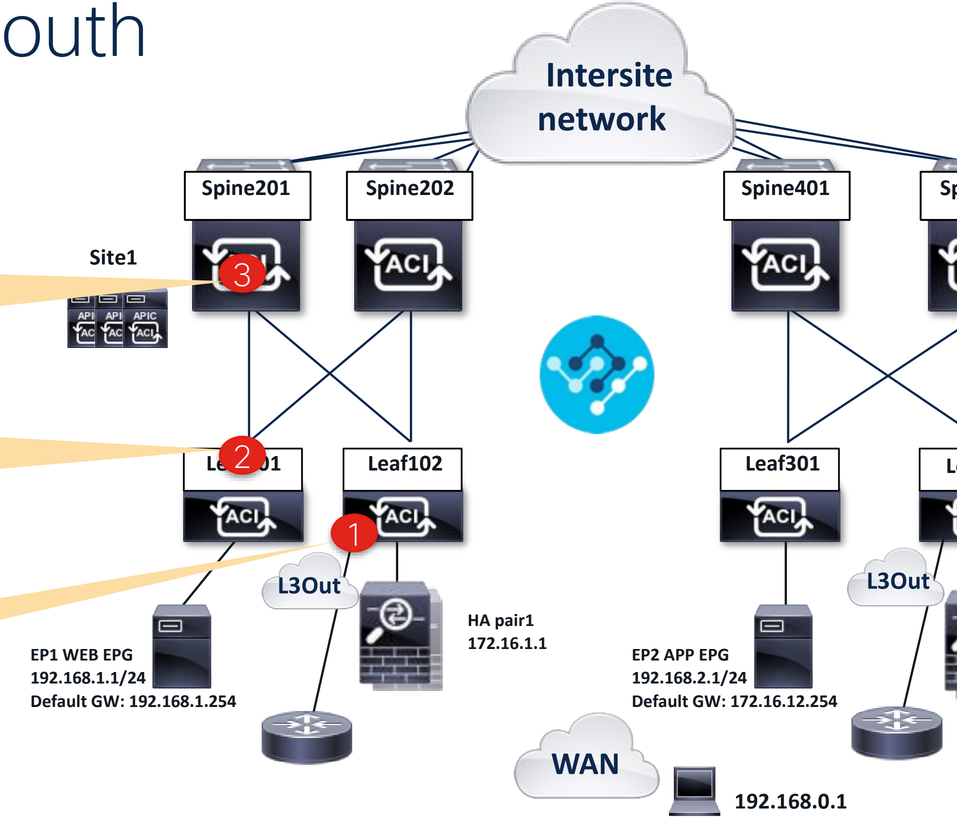
Spine performs COOP lookup for VMAC in service BD and sends it to service leaf, who forwards it to FW.

2

Compute leaf applies redirect to HA pair of server site (here HA pair1 in site1). Compute leaf will send it to spiny-mac proxy

1

On Border Leaf traffic hits REDIR+OVERRIDE rule as destination EP is not local. BL will not redirect, traffic follows regular forwarding to reach compute leaf

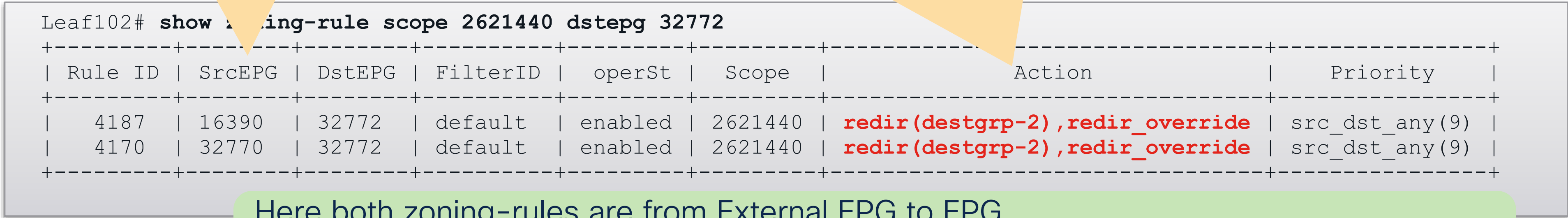


Multi-Site PBR – North-South

Zoning Rule L3Out EPG on ingress Border Leaf

On border leaf sclass from L3out will either be VRF pcTag 32770 (0.0.0.0/0 prefix) or External EPG pcTag 16390 (specific prefix).

Zoning-rules are always redir +override, so BL will apply permit override unless the destination EP is also local



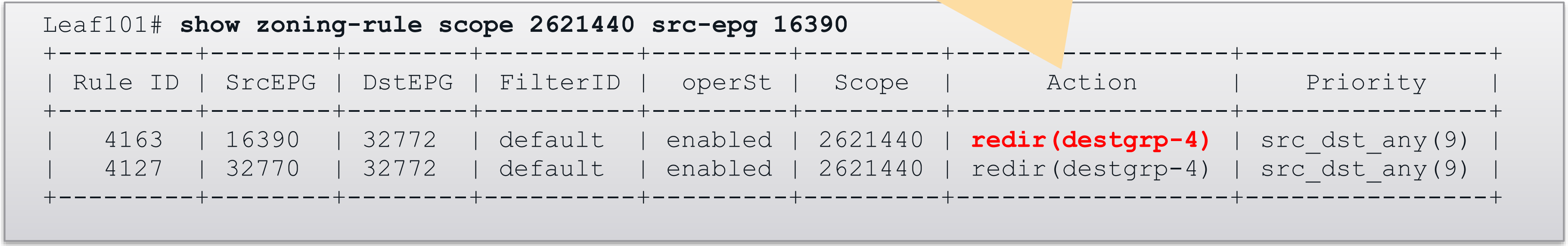
```
Leaf102# show zoning-rule scope 2621440 dstepg 32772
+---------+--------+--------+----------+---------+---------+-------------------------------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |                Action               |    Priority    |
+---------+--------+--------+----------+---------+---------+-------------------------------------+----------------+
|   4187  | 16390  | 32772  | default  | enabled | 2621440 | redir(destgrp-2),redir_override | src_dst_any(9) |
|   4170  | 32770  | 32772  | default  | enabled | 2621440 | redir(destgrp-2),redir_override | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+-------------------------------------+----------------+
```

Here both zoning-rules are from External EPG to EPG
We will use one or the other depending on the zoning-rule subnet in the external EPG (0.0.0.0/0 or specific subnet)

Multi-Site PBR – North-South

Zoning Rule L3Out EPG on compute leaf

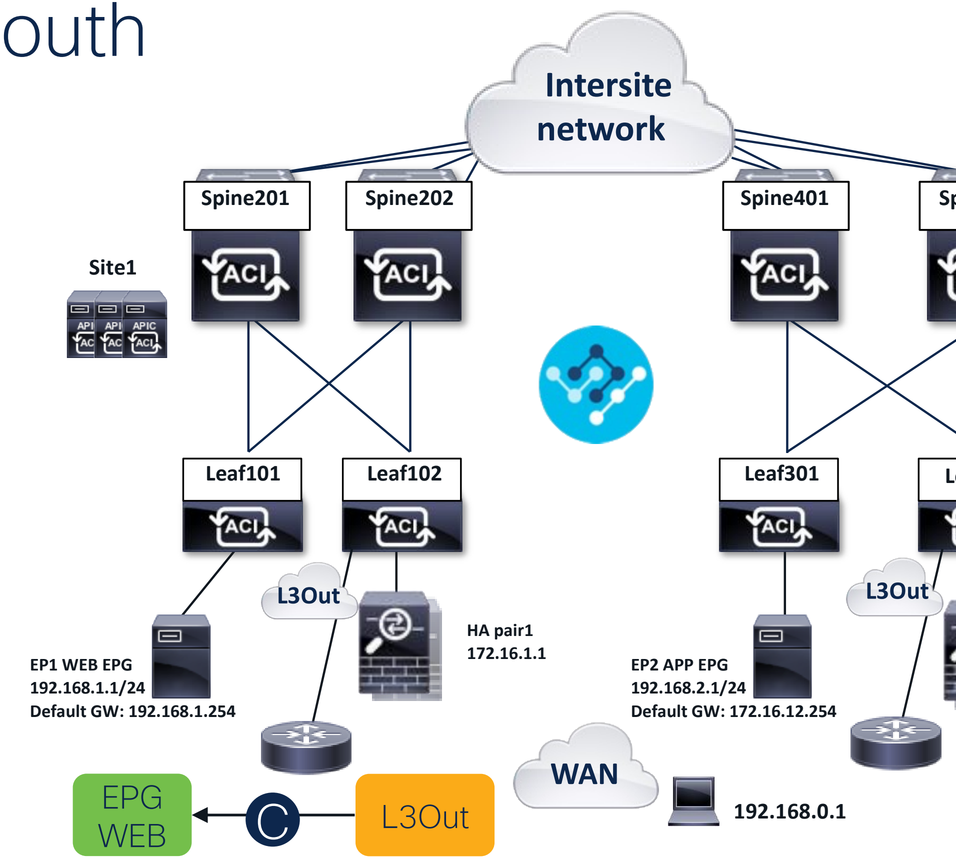
On compute leaf only redirect Action is present in rule, so we will always redirect here



```
Leaf101# show zoning-rule scope 2621440 src-epg 16390
+---------+--------+--------+----------+---------+---------+-------------------+-----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |       Action      |     Priority    |
+---------+--------+--------+----------+---------+---------+-------------------+-----------------+
|   4163  | 16390  | 32772  | default  | enabled | 2621440 | redir(destgrp-4)  | src_dst_any(9)  |
|   4127  | 32770  | 32772  | default  | enabled | 2621440 | redir(destgrp-4)  | src_dst_any(9)  |
+---------+--------+--------+----------+---------+---------+-------------------+-----------------+
```

Multi-Site PBR – North-South

- Multisite PBR requirement
  - A site can only redirect to site local PBR Devices
- Rule: Redirect happens on compute leaf, not Border Leaf
- Solution for North-South:
  - Provider or consumer location does not matter
  - What matters is Compute and Border leaf
  - Only ingress vrf enforcement is supported (default). Need to ensure all compute leaf have a zoning-rule to apply the contract for an external prefix



Multi-Site PBR – One Side Summary

Rule East-West

EPG – pcTag (sclass)	EPG pcTag (dclass)	Action	Remark
Consumer	Provider	REDIRECT + OVERRIDE	To ensure redirect is one on site where provider EP sits
Service EPG	Provider	Permit	
Provider	Consumer	REDIRECT	Redirect always done on provider ingress leaf
Service EPG	Consumer	Permit	

Rule North-South

EPG – pcTag (sclass)	EPG pcTag (dclass)	Action	Remark
Server EPG	External EPG	REDIRECT	Coming from EP we redirect directly on ingress server leaf
Service EPG	External EPG	Permit	
External EPG	Server EPG	REDIRECT + OVERRIDE	Coming from L3 out we do NOT redirect but we override to be apply redirect on site of incoming server EP
Service EPG	Server EPG	Permit	

Multisite PBR – Summary

- We need to ensure traffic symmetry across site
- APIC cluster do not have visibility on remote site PBR node and can only redirect to local site L4/L7 device
    - How can we ensure redirect is symmetric (same site in both direction)

- Implementation is the following (post 4.x)
    - East-West – Redirect ALWAYS applied in the site where Provider EP sits.
        - Extra requirement – Consumer EPG should have subnet under them
    - North-South – Redirect is always apply on Server leaf site (non BL)

Unidirection PBR Load Balancer with no SNAT

Load Balancer with no SNAT

Traffic from Client to Server through Load Balancer

3

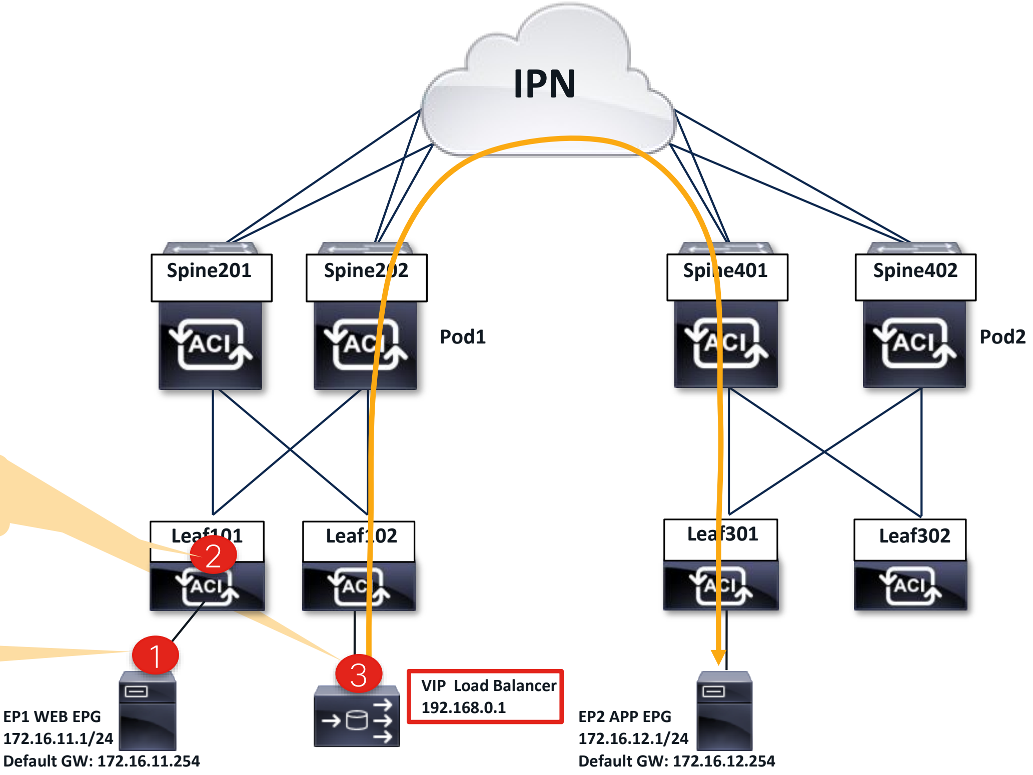
LB does rewrite DIP but not source IP (NO SNAT)
Src IP: 172.16.11.1
Src MAC: VMAC LB
Dest IP: 172.16.12.1 - Real Server
Dest MAC: Anycast MAC

2

No PBR routing needed to VIP

1

Traffic from Client
Src IP: 172.16.11.1
Src MAC: MAC EP1
Dest IP: 192.168.0.1 - VIP
Dest MAC: Anycast MAC



Load Balancer with no SNAT

Return traffic from Server to Client through Load Balancer

3

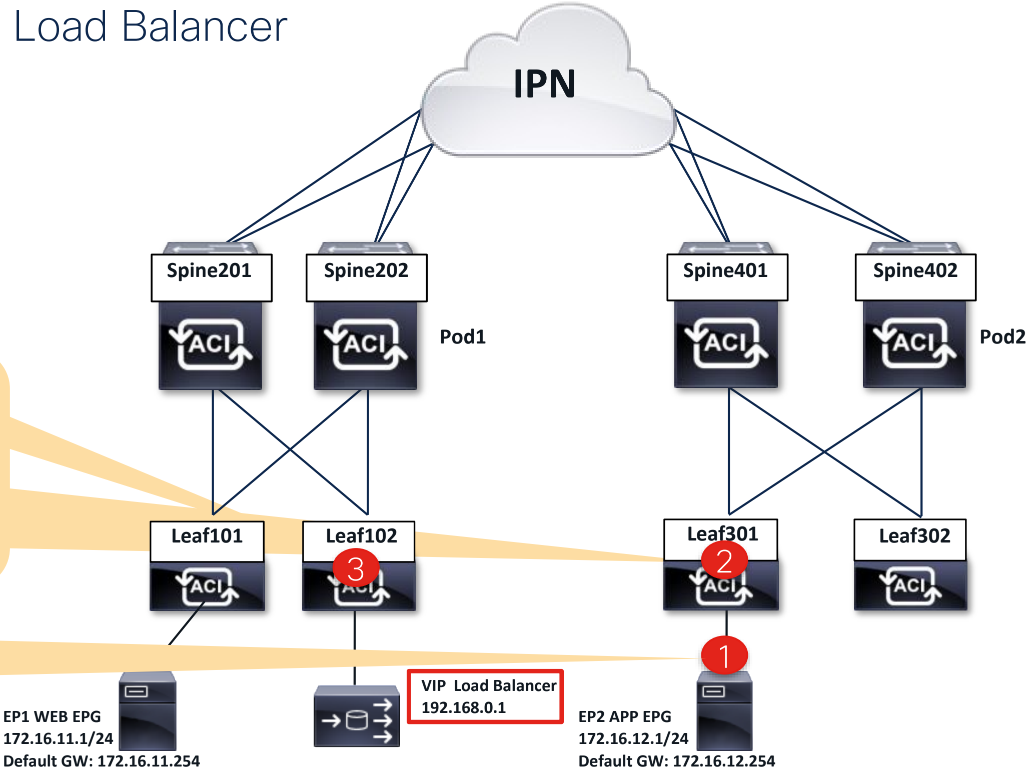
Traffic from Service node (After PBR)
Src IP: 172.16.21.1 (VIP)
Src MAC: Leaf MAC
Dest IP: 172.16.11.1
Dest MAC: EP A

2

Return traffic will hit zoning-rule for redirect
DMAC is rewritten to LB VMAC.
No Mac lookup happening on leaf.
Packet is encapsulated to Service BD VNID and send to vxlan tunnel to anycast-mac on spine.

1

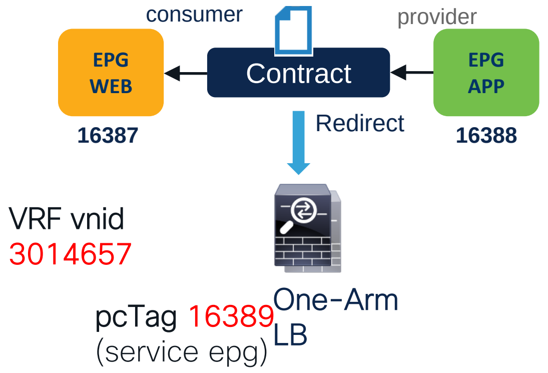
Real server replies directly to Client IP, so not to the VIP (NO SNAT)
Traffic will bypass LB in return direction, unless PBR is used



Zoning-rule

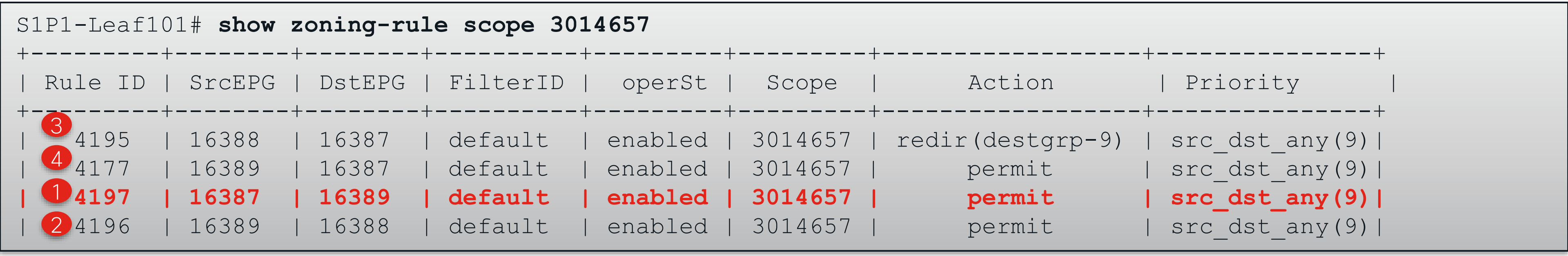
Note : There is no rule from Consumer to Provider.
As a result we have an SG with NO PBR leg, so we apply the rule directly to service EPG
However there is no direct contract between Web and service EPG

Make note all all vnid and sclass involved



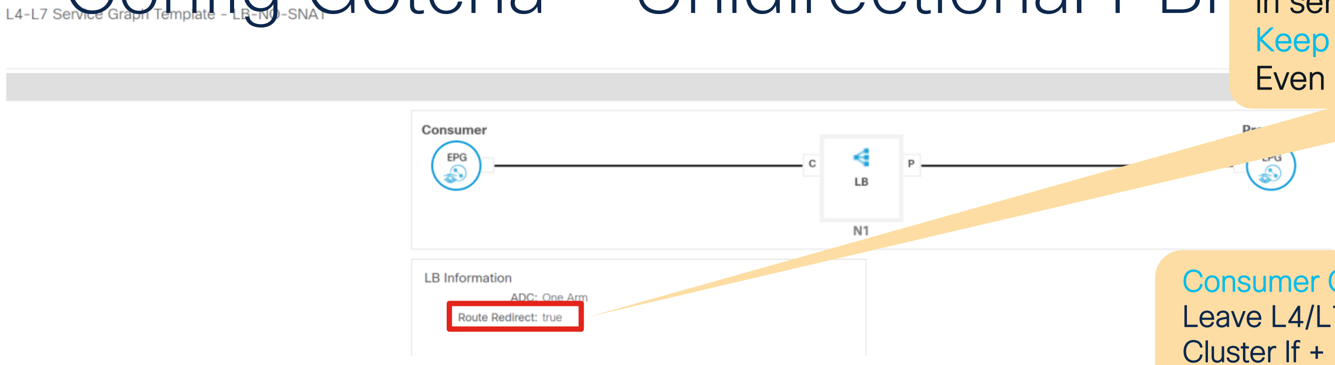
Expected zoning-rules:

1. Cons to Prov (replaced by Consumer to Service EPG because DIP is VIP in service EPG) : 16387 to 16389 : PERMIT
2. Shadow to Prov : 16389 to 16388 : PERMIT
3. Prov to Cons : 16388 to 16387 : REDIRECT
4. Shadow to Cons : 16389 to 16387 : PERMIT



```
S1P1-Leaf101# show zoning-rule scope 3014657
+---------+--------+--------+----------+---------+---------+------------------+---------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |      Action      | Priority      |
+---------+--------+--------+----------+---------+---------+------------------+---------------+
|  4195   | 16388  | 16387  | default  | enabled | 3014657 | redir(destgrp-9) | src_dst_any(9)|
|  4177   | 16389  | 16387  | default  | enabled | 3014657 |      permit      | src_dst_any(9)|
|  4197   | 16387  | 16389  | default  | enabled | 3014657 |      permit      | src_dst_any(9)|
|  4196   | 16389  | 16388  | default  | enabled | 3014657 |      permit      | src_dst_any(9)|
```

Config Gotcha – Unidirectional PBR



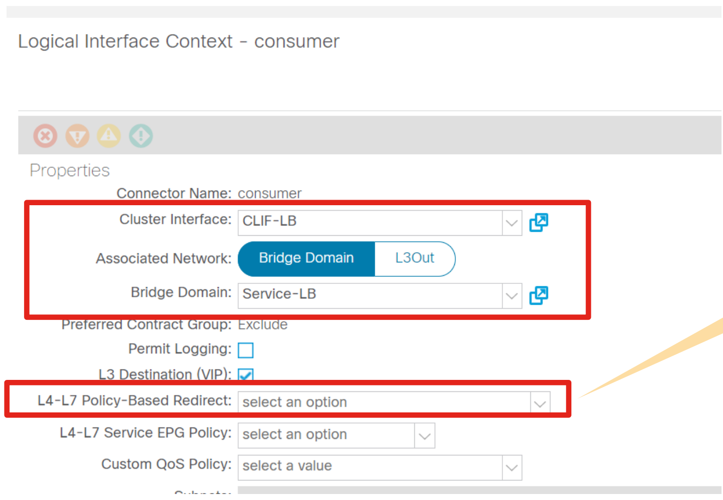
L4-L7 Service Graph Template - LB-NO-SNAT

LB Information

ADC: One Arm

Route Redirect: true

In service graph template
Keep route redirect : True
Even if only one leg needs redirect



Logical Interface Context - consumer

Properties

Connector Name: consumer

Cluster Interface: CLIF-LB

Associated Network: Bridge Domain | L3Out

Bridge Domain: Service-LB

Preferred Contract Group: Exclude

Permit Logging: □

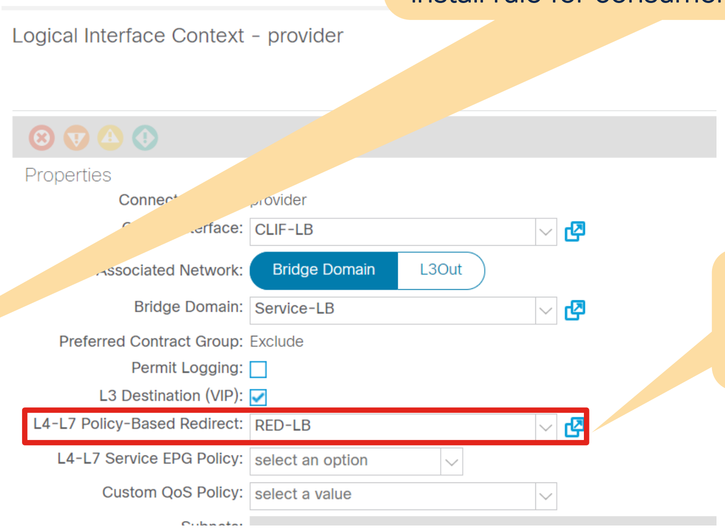
L3 Destination (VIP): ☑

L4-L7 Policy-Based Redirect: select an option

L4-L7 Service EPG Policy: select an option

Custom QoS Policy: select a value

Consumer Connector (no PBR):
Leave L4/L7 redirect empty
Cluster If + Service BD will instruct ACI to install rule for consumer to reach service EGP



Logical Interface Context - provider

Properties

Connect provider

terface: CLIF-LB

ssociated Network: Bridge Domain | L3Out

Bridge Domain: Service-LB

Preferred Contract Group: Exclude

Permit Logging: □

L3 Destination (VIP): ☑

L4-L7 Policy-Based Redirect: RED-LB

L4-L7 Service EPG Policy: select an option

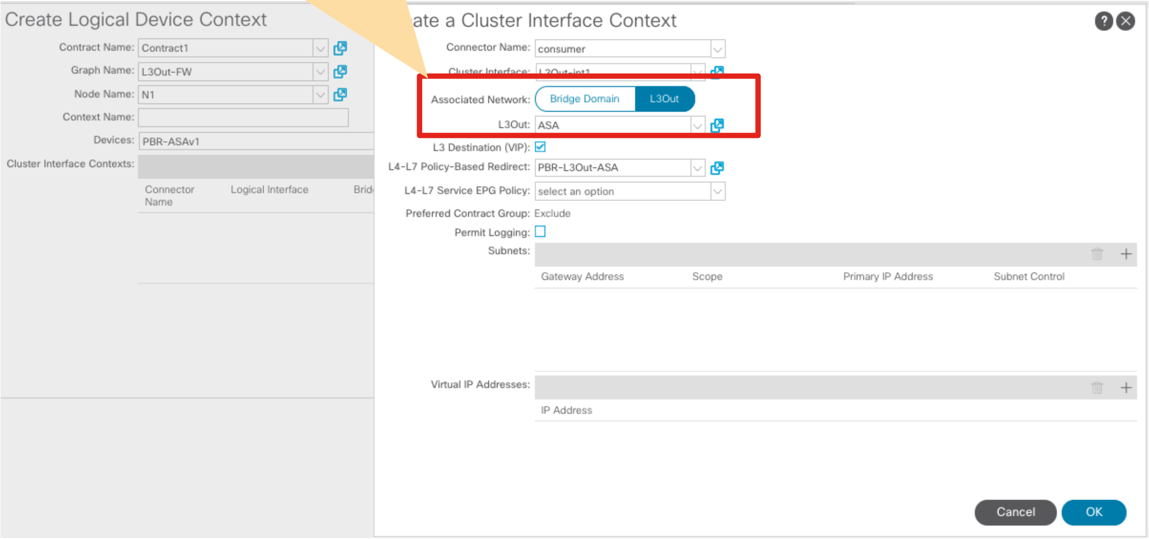
Custom QoS Policy: select a value

Provider Connector
PBR as usual

PBR on L3Out

Config Gotcha – PBR on L3Out

Associated Network: L3Out
L3Out: Select your External EPG



Create Logical Device Context

Contract Name: Contract1
Graph Name: L3Out-FW
Node Name: N1
Context Name:
Devices: PBR-ASAv1
Cluster Interface Contexts:

Connector Name	Logical Interface	Brid

ate a Cluster Interface Context

Connector Name: consumer
Cluster Interface: L3Out-int1
Associated Network: Bridge Domain | L3Out
L3Out: ASA
L3 Destination (VIP): ☑
L4-L7 Policy-Based Redirect: PBR-L3Out-ASA
L4-L7 Service EPG Policy: select an option
Preferred Contract Group: Exclude
Permit Logging: □
Subnets:

Gateway Address	Scope	Primary IP Address	Subnet Control

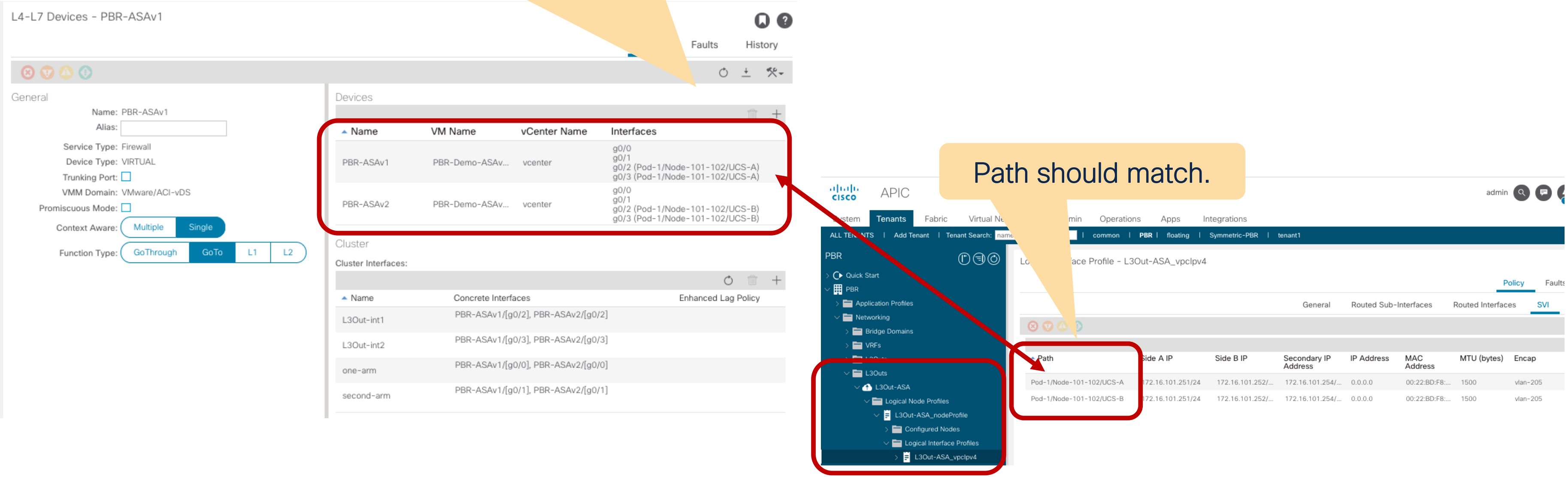
Virtual IP Addresses:

IP Address

Other configurations are the same with PBR Destination in a BD

Config Gotcha – Path in L4-L7 device

In this example, g0/2 and g0/3 are used for PBR destinations in an L3Out where Path configuration is required.



General

Name: PBR-ASAv1
Alias:
Service Type: Firewall
Device Type: VIRTUAL
Trunking Port: □
VMM Domain: VMware/ACI-vDS
Promiscuous Mode: □
Context Aware: Multiple [✓ Single]
Function Type: GoThrough [✓ GoTo] L1 L2

Devices

Name	VM Name	vCenter Name	Interfaces
PBR-ASAv1	PBR-Demo-ASAv...	vcenter	g0/0 g0/1 g0/2 (Pod-1/Node-101-102/UCS-A) g0/3 (Pod-1/Node-101-102/UCS-A)
PBR-ASAv2	PBR-Demo-ASAv...	vcenter	g0/0 g0/1 g0/2 (Pod-1/Node-101-102/UCS-B) g0/3 (Pod-1/Node-101-102/UCS-B)

Cluster

Cluster Interfaces:

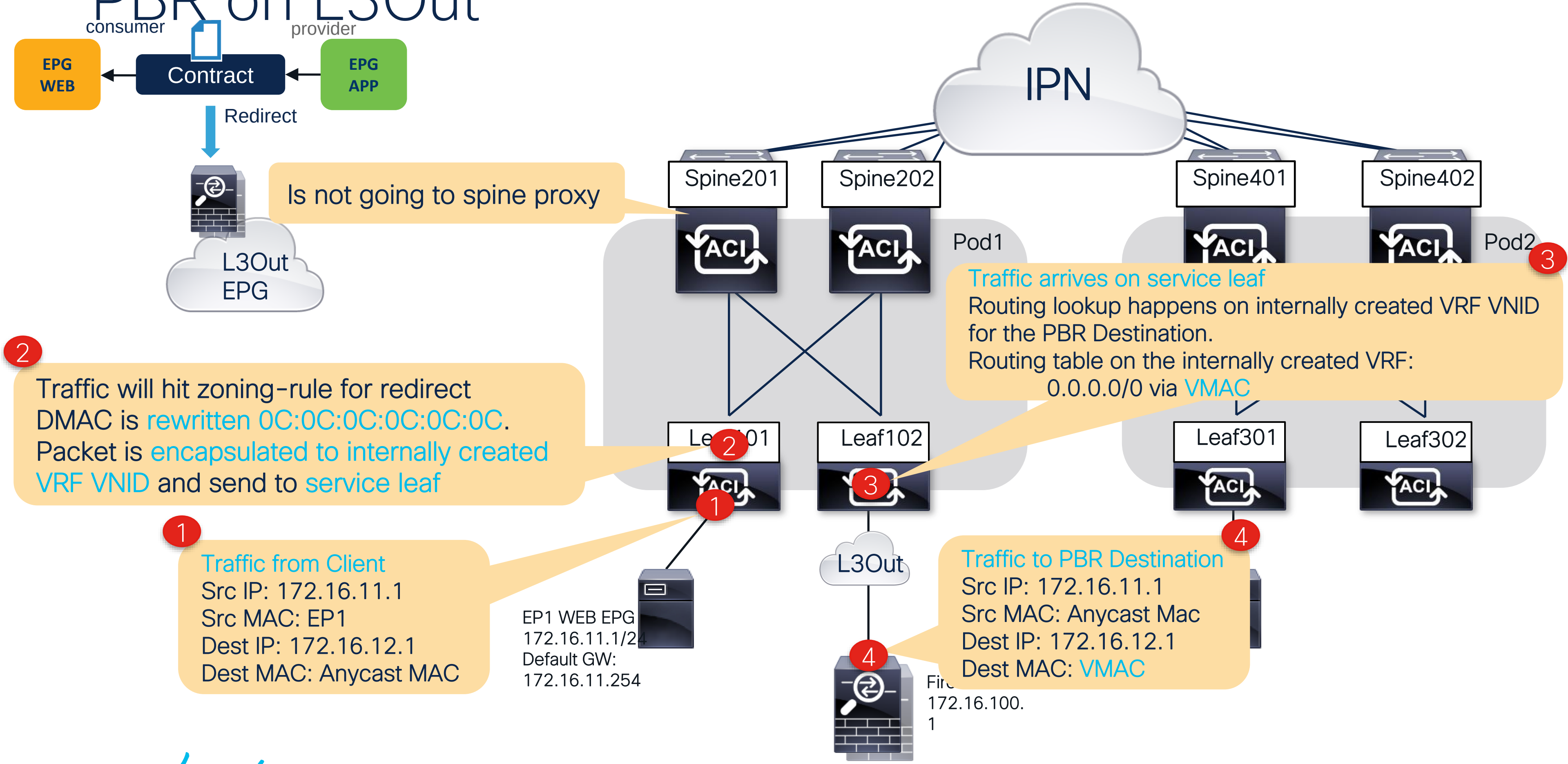
Name	Concrete Interfaces	Enhanced Lag Policy
L3Out-int1	PBR-ASAv1/[g0/2], PBR-ASAv2/[g0/2]	
L3Out-int2	PBR-ASAv1/[g0/3], PBR-ASAv2/[g0/3]	
one-arm	PBR-ASAv1/[g0/0], PBR-ASAv2/[g0/0]	
second-arm	PBR-ASAv1/[g0/1], PBR-ASAv2/[g0/1]	

Path should match.

Path	Side A IP	Side B IP	Secondary IP Address	IP Address	MAC Address	MTU (bytes)	Encap
Pod-1/Node-101-102/UCS-A	172.16.101.251/24	172.16.101.252/...	172.16.101.254/...	0.0.0.0	00:22:BD:F8:...	1500	vlan-205
Pod-1/Node-101-102/UCS-B	172.16.101.251/24	172.16.101.252/...	172.16.101.254/...	0.0.0.0	00:22:BD:F8:...	1500	vlan-205

Even with virtual service nodes, paths must be configured in L4-L7 device and it must be matched with the paths of logical interfaces of the L3Out. Otherwise, APIC raises fault.

PBR on L3Out



Is not going to spine proxy

2 Traffic will hit zoning-rule for redirect
DMAC is rewritten 0C:0C:0C:0C:0C:0C.
Packet is encapsulated to internally created VRF VNID and send to service leaf

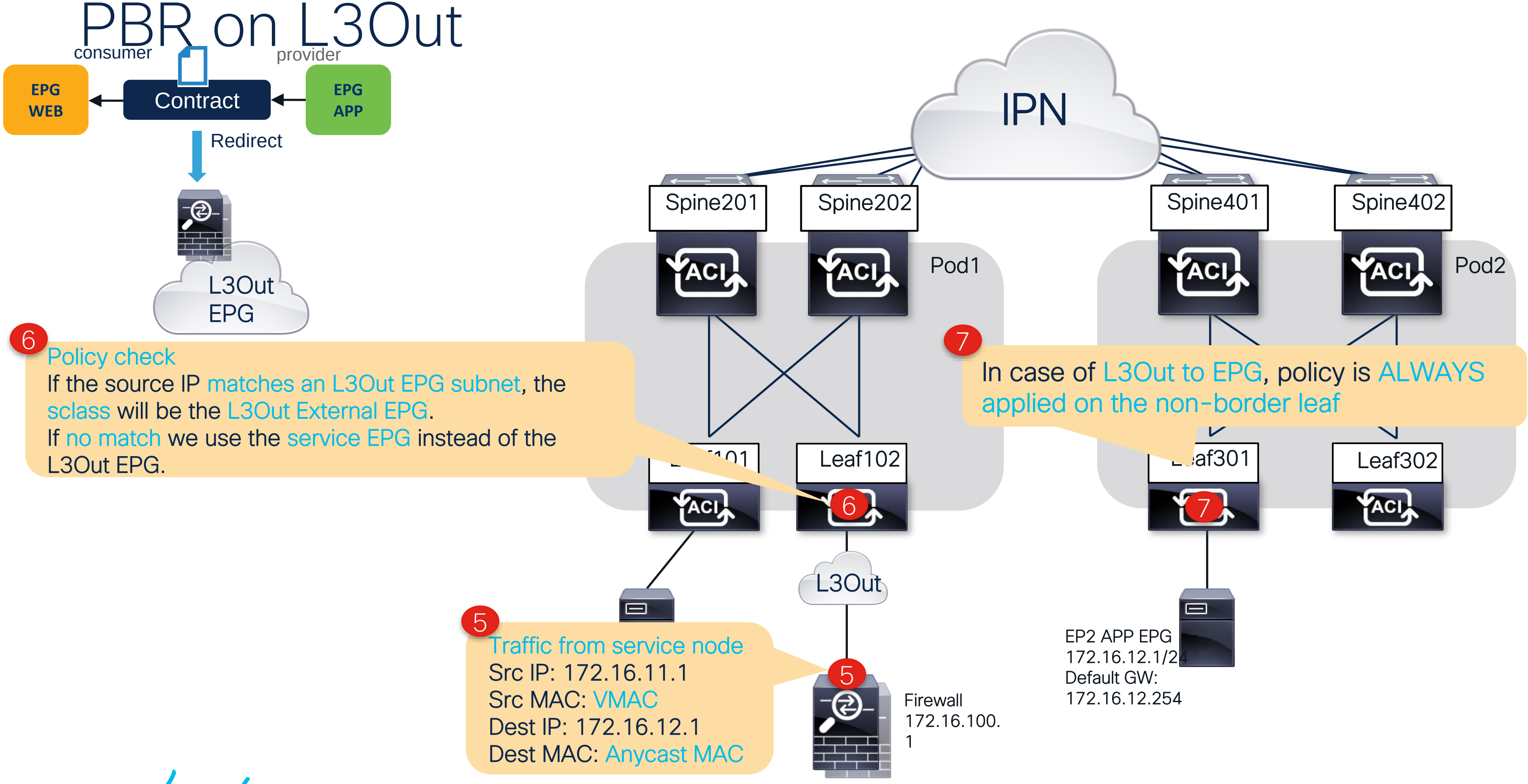
1 Traffic from Client
Src IP: 172.16.11.1
Src MAC: EP1
Dest IP: 172.16.12.1
Dest MAC: Anycast MAC

3 Traffic arrives on service leaf
Routing lookup happens on internally created VRF VNID for the PBR Destination.
Routing table on the internally created VRF:
0.0.0.0/0 via VMAC

4 Traffic to PBR Destination
Src IP: 172.16.11.1
Src MAC: Anycast Mac
Dest IP: 172.16.12.1
Dest MAC: VMAC

cisco Live!

PBR on L3Out

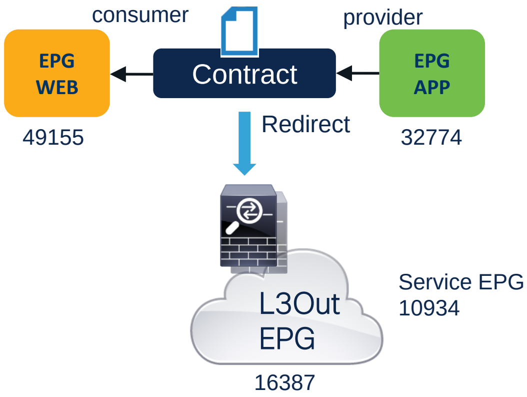


6 Policy check
If the source IP matches an L3Out EPG subnet, the sclass will be the L3Out External EPG.
If no match we use the service EPG instead of the L3Out EPG.

7 In case of L3Out to EPG, policy is ALWAYS applied on the non-border leaf

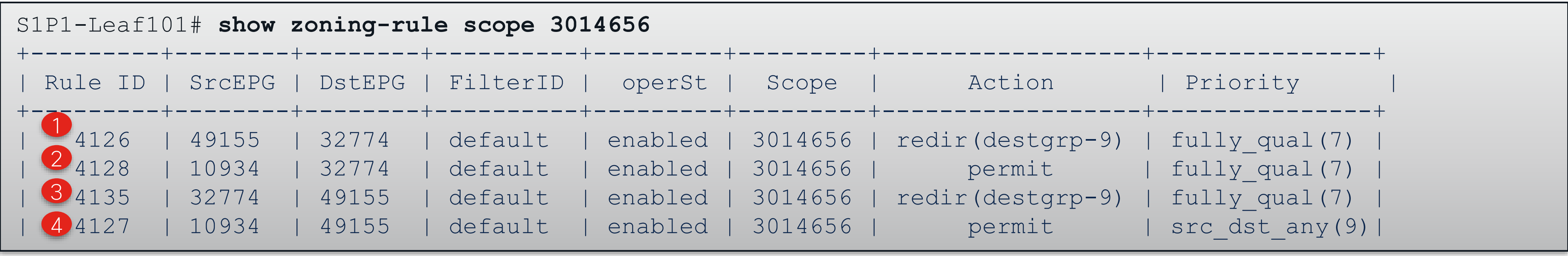
5 Traffic from service node
Src IP: 172.16.11.1
Src MAC: VMAC
Dest IP: 172.16.12.1
Dest MAC: Anycast MAC

PBR on L3Out



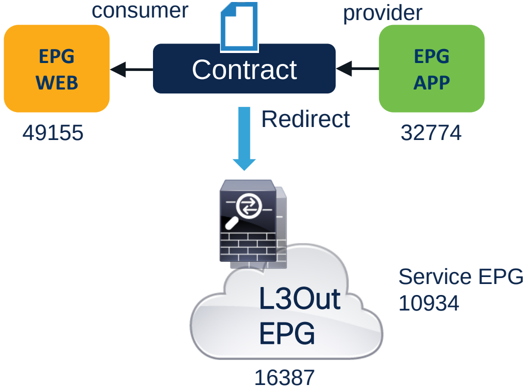
Expected zoning-rules:

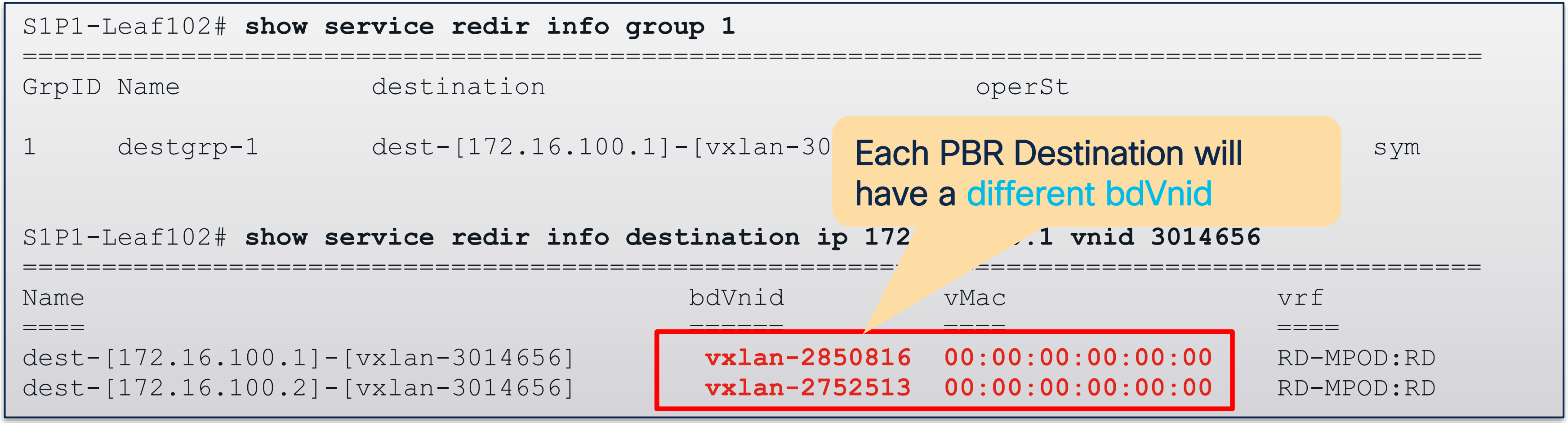
1. Cons to Prov : 49155 to 32774 : Redirect
2. Shadow to Prov : 10934 to 32774 : PERMIT
3. Prov to Cons : 32774 to 49155 : REDIRECT
4. Shadow to Cons : 10934 to 49155 : PERMIT



```
S1P1-Leaf101# show zoning-rule scope 3014656
+---------+--------+--------+----------+---------+---------+-------------------+---------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |       Action      | Priority      |
+---------+--------+--------+----------+---------+---------+-------------------+---------------+
|  4126   | 49155  | 32774  | default  | enabled | 3014656 | redir(destgrp-9)  | fully_qual(7) |
|  4128   | 10934  | 32774  | default  | enabled | 3014656 |       permit      | fully_qual(7) |
|  4135   | 32774  | 49155  | default  | enabled | 3014656 | redir(destgrp-9)  | fully_qual(7) |
|  4127   | 10934  | 49155  | default  | enabled | 3014656 |       permit      | src_dst_any(9)|
```

PBR on L3Out



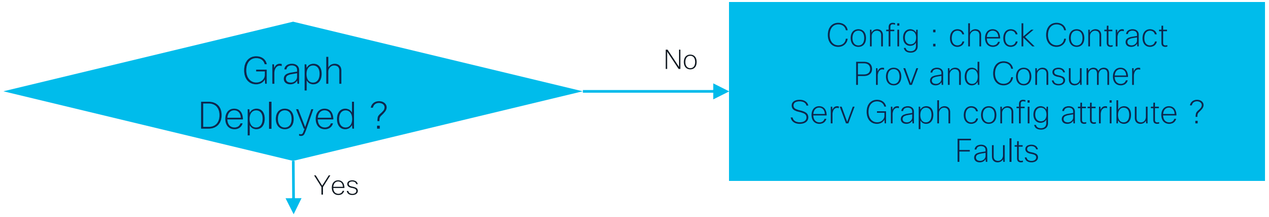


```
S1P1-Leaf102# show service redir info group 1
==========================================================================================
GrpID Name            destination                                                   operSt

1     destgrp-1       dest-[172.16.100.1]-[vxlan-30                                 sym


S1P1-Leaf102# show service redir info destination ip 172     .1 vnid 3014656
==========================================================================================
Name                                  bdVnid          vMac                vrf
====                                  ======          ====                ====
dest-[172.16.100.1]-[vxlan-3014656]   vxlan-2850816   00:00:00:00:00:00   RD-MPOD:RD
dest-[172.16.100.2]-[vxlan-3014656]   vxlan-2752513   00:00:00:00:00:00   RD-MPOD:RD
```

Each PBR Destination will have a different bdVnid

Summary

Troubleshooting PBR checklist


Graph Deployed ?
No
Config : check Contract Prov and Consumer Serv Graph config attribute ? Faults
Yes

Take note of pcTag for Cons/Provider and service(s) EPG

Run the Basic Checks

Follow the packet path with ftriage or elam assistant

Remember the expected packet path based on your fabric (single pod/multipod/multisite) East-West vs North-South

Summary – PBR and firewall deployment options

Firewall integration model	Multipod East-West	Multipod North-South	Multisite East-West	Multisite North-South
Active and standby across Pod/Site	OK - Simple PBR	OK – simple PBR	NOK	NOK
Active/Active FW across POD	OK with anycast PBR	OK with anycast PBR	NOK	NOK
Active/Standby per pod site	OK symmetric PBR	OK either symmetric PBR or pod aware (+option Host based routing)	OK with PBR – Redirect on provider site	OK with PBR – Redirect on Server leaf site

Packet Format

1 From EP1 to ACI Leaf 101

Inner			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC EP1	DMAC ACI

2 From Leaf 101 to Fabric in case EP 2 is unknown in leaf 101

Inner				Outer			
DIP 172.16.12.1	SIP 172.16.11.1	xx	xx	VNID VRF	Sclass EPG1	DIP Anycast-v4	SIP Leaf101 PTEP

3 After Redirect (either by Leaf 101 or by leaf 301

Inner				Outer			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC EP1	DMAC FW	VNID Service BD	Sclass EPG1	DIP Anycast-Mac	SIP Leaf 101

4 After FW MAC coop lookup in service BD

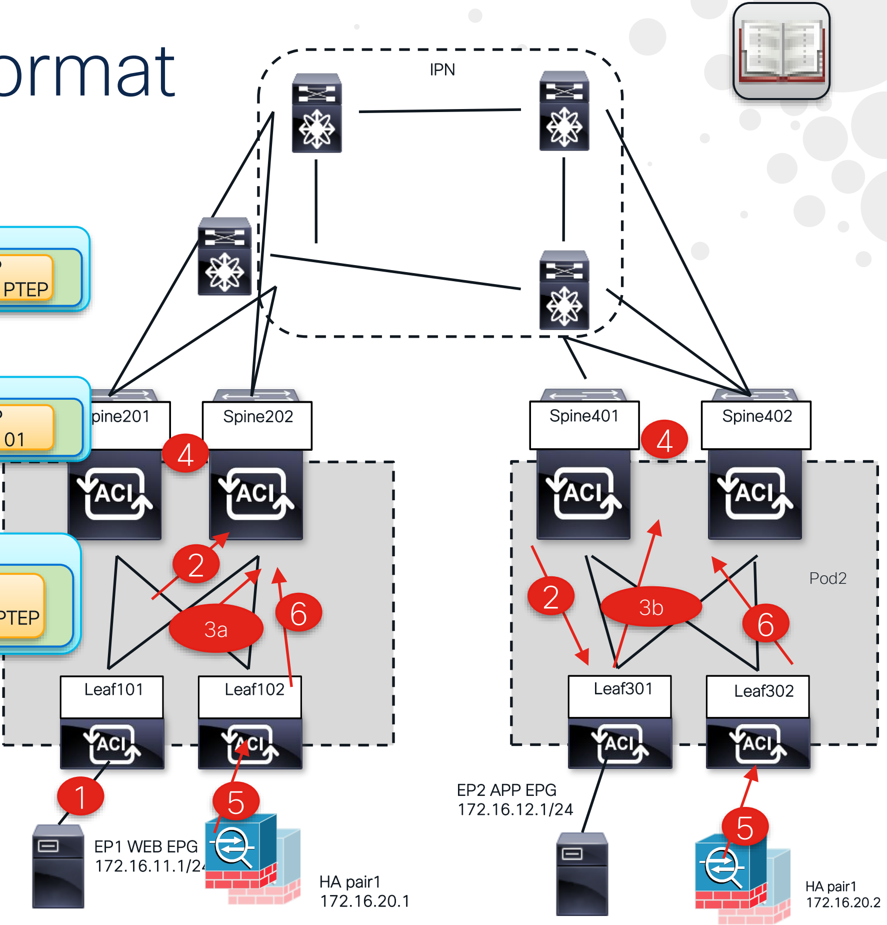
Inner				Outer			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC EP1	DMAC FW	VNID Service BD	Sclass EPG1	DIP - PTEP FW Leaf or Remote Pod	SIP Leaf101 PTEP

5 Coming back from FW on leaf 102 or 302

Inner			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC FW	DMAC ACI

6 After FW, between Fw leaf and destination leaf 301

Inner							
DIP 172.16.12.1	SIP 172.16.11.1	xxx	xxx	VNID VRF	Sclass Shadow +No Learn	DIP - PTEP Leaf 301 or acast-v4	SIP - PTEP FW Leaf



Complete your Session Survey

- Please complete your session survey after each session. Your feedback is important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (open from Thursday) to receive your Cisco Live t-shirt.
- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Session Catalog and clicking the "Attendee Dashboard" at https://www.ciscolive.com/emea/learn/sessions/session-catalog.html

Continue Your Education

- Visit the Cisco Showcase for related demos.
- Book your one-on-one Meet the Engineer meeting.
- Attend any of the related sessions at the DevNet, Capture the Flag, and Walk-in Labs zones.
- Visit the On-Demand Library for more sessions at ciscolive.com/on-demand.

cisco The bridge to possible

Thank you

cisco Live!

cisco Live!
ALL IN