NORTEL NETWORKS™

Nortel Networks Multiservice Switch 7400/15000/20000

IP Technology Fundamentals

NN10600-800

Document status: Standard
Document issue: 6.1S2
Document date: November 2004
Product release: Release 6.1
Job function: Operations
Type: NTP
Language type: U.S. English

Contents

About this document

This user guide describes virtual routers, the Internet protocol (IP), and other protocols and services related to IP in Nortel Networks Multiservice Switch systems.

This guide is for anyone who performs the following tasks: planning, installing and provisioning, operating and maintaining.

What's new in this document

The following features were added to this document:

- 4-port 10/100BaseT Ethernet FP (page 10)
- 8-port 10/100 BaseT Ethernet FP (page 10)
- Hitless IP CP switchover (page 11)
- Multi-hop EBGP (page 11)
- Protected default route (page 11)
- VR Control plane protection (CPP) (page 11)
- IP multicast (page 11)
- MD5 Authentication (page 11)
- VIPR on 4-port Gigabit Ethernet, 4-port 10/100 BaseT Ethernet, and 8-port 10/100 BaseT Ethernet FPs (page 11)
- Virtual media inter-VR hardware connectivity (page 12)
- Virtual router redundancy protocol (VRRP) 4-port Gigabit Ethernet, 4-port 10/100 BaseT Ethernet, and 8-port 10/100 BaseT Ethernet FPs (page 12)

Other changes made to this document include the following:

- Updated table Minimum supported MTU sizes (page 37) to include Ethernet VLAN and port-mode details.
- Updated sections in Multiservice Switch IP fundamentals (page 16) and IP over ATM (page 46) to include more detailed explanations regarding IP VRs and IP over ATM.

- Updated section CP switchover (page 73) with the removal of the term Flat VR when referencing VIPR.
- Moved the chapters "IP class of service (CoS)", "Multiservice Switch IP differentiated services", "IP flow filters", and "IP CoS to IP DiffServ Migration" to NN10600-590 *Nortel Networks Multiservice Switch 7400/15000/20000 Layer 3 Traffic Management Fundamentals*.
- The terms Passport and PVG have been rebranded in conjunction with the new Nortel Networks' brand simplified naming format. Passport is now referred to as the Nortel Networks Multiservice Switch, and PVG is now Media Gateway 7480/15000. For more information on the product rebranding, refer to NN10600-000 *Nortel Networks Multiservice Switch 7400/15000/20000 What's New in PCR6.1*.
- Updated section ATM MPE over soft PVCs (page 47) to explain two other ways of connecting ATM MPE soft PVCs.
- For CR Q00866544, updated section Forwarding classes and route preferences (page 88).
- For CR Q00877198, updated sections Overview of Multiservice Switch RIP (page 93), RIP policies (page 93), Migrating from RIPv1 to RIPv2 (page 94), Migrating from RIP to OSPF (page 98), and OSPF areas (page 100) to ensure that customers are aware of certain details regarding RIP and OSPF functionality.
- Moved the chapter ISIS to NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals*.

4-port 10/100BaseT Ethernet FP

The following sections were updated:

- Virtual LAN (page 32)
- Overview of Multiservice Switch IP over Ethernet (page 71)
- Overview of VRRP (page 132)
- VRRP virtual routers (page 133)
- Router redundancy (page 134)

8-port 10/100 BaseT Ethernet FP

The following sections were added:

- Virtual LAN (page 32)
- Router redundancy with VIPR (page 135)

The following sections were updated:

- IP over Ethernet (page 71)
- Overview of Multiservice Switch IP over Ethernet (page 71)

- Overview of VRRP (page 132)
- VRRP virtual routers (page 133)
- Router redundancy (page 134)
- Router availability (page 137)
- Application and feature names for Multiservice Switch IP (page 17)

Hitless IP CP switchover

The following section was updated:

- CP switchover (page 73)

Multi-hop EBGP

The following section was added:

- Single-hop and multi-hop BGP (page 109)

Protected default route

The following section was added:

- Protected default route (page 124)

VR Control plane protection (CPP)

The following sections were added:

- IP security mechanisms (page 152)
- Control plane protection (CPP) (page 153)

IP multicast

The following section was added:

- IP multicast (page 126)

MD5 Authentication

The following section was added:

- MD5 authentication (page 157)

VIPR on 4-port Gigabit Ethernet, 4-port 10/100 BaseT Ethernet, and 8-port 10/100 BaseT Ethernet FPs

The following sections were updated:

- Application and feature names for Multiservice Switch IP (page 17)
- IP processor cards (page 18)
- Virtual LAN (page 32)

Virtual media inter-VR hardware connectivity

The following section was updated:

- Multiservice Switch virtual media (page 28)

Virtual router redundancy protocol (VRRP) 4-port Gigabit Ethernet, 4-port 10/100 BaseT Ethernet, and 8-port 10/100 BaseT Ethernet FPs

The following section was update:

- Virtual router redundancy protocol (page 132)

Procedure conventions

This document uses the following procedure conventions:

- You can enter commands using full component and attribute names, or you can abbreviate them. The commands used in the procedures contain the full component and attribute names in the first instance. In the second instance, the component and attribute names are abbreviated. For more information on abbreviating component and attribute names, see NN10600-060 *Nortel Networks Multiservice Switch 7400/15000/20000 Component Reference*. All component and attribute names are formatted in italics.
- The introduction of every procedure states whether you must perform the procedure in operational mode or provisioning mode. For more information on these modes, see Operational mode (page 12) or Provisioning mode (page 13).
- When you complete a procedure, you can verify your changes and then activate them as the new node configuration. For more information on completing configuration changes and exiting provisioning mode, see Completing configuration changes (page 13).

Operational mode

Procedures contained within this document can be performed when the Nortel Networks Multiservice Switch node is in operational mode or provisioning mode. When you initially log into the node, you are in operational mode. Multiservice Switch nodes use the following command prompt when you are in operational mode:

```
#>
```

where:
is the current command number

In operational mode, you work with operational components and attributes. In operational mode, you can

- list operational components and display operational attributes to determine the current operating parameters for the node
- control the state of parts of the node by locking and unlocking components
- set certain operational attributes and enter commands to perform diagnostic tests

Provisioning mode

To change from operational mode to provisioning mode, use the start Prov command. Only one user can be in provisioning mode at a time. Nortel Networks Multiservice Switch nodes use the following command prompt whenever you are in provisioning mode:

```
PROV #>
```

where:
is the current command number

In provisioning mode, you work with the provisionable components and attributes that contain the current and future configurations of the node. You can add and delete components, and display and set provisionable attributes. For information on completing the configuration changes, exiting provisioning mode, and returning to operational mode see Completing configuration changes (page 13).

For information on operational and provisionable attributes, see NN10600-060 *Nortel Networks Multiservice Switch 7400/15000/20000 Component Reference*.

Completing configuration changes

Several procedures in this document ask that you complete the configuration changes. When you complete the configuration changes, you are activating the configuration changes, confirming that you want to activate them, and saving the changes. Follow this procedure in provisioning mode when asked to complete the configuration changes. See the section Provisioning mode (page 13) for more information.

1. Verify that the provisioning changes you have made are acceptable:

   `check Prov`

   Correct any errors and then verify the provisioning changes again.

2. If you want to store the provisioning changes in a file, save the provisioning view with portable formats:

   `save -f(<filename>) -portable Prov`

3 If you want these changes as well as other changes made in the edit view to take effect immediately, activate and commit the provisioning changes:

```
activate Prov
confirm Prov
commit Prov
```

4 End the provisioning session:

```
end Prov
```

Request for comments (RFCs)

The following Requests for Comments (RFCs) containing information related to IP are available from numerous sources including Internet Network Information Center (NIC) servers:

- RFC761, *DoD standard Transmission Control Protocol*
- RFC768, *User Datagram Protocol*
- RFC791, *Internet Protocol*
- RFC792, *Internet Control Message Protocol*
- RFC793, *Transmission Control Protocol*
- RFC815, *IP Datagram Reassembly Algorithms*
- RFC821, *Simple Mail Transfer Protocol*
- RFC826, *An Ethernet Address Resolution Protocol*
- RFC854, *Telnet Protocol Specifications*
- RFC904, *Exterior Gateway Protocol Formal Specification*
- RFC950, *Internet Standard Subnetting Procedure*
- RFC951, *Bootstrap Protocol (BootP)*
- RFC959, *File Transfer Protocol*
- RFC1009, *Requirements for Internet Gateways*
- RFC1038, *Draft Revised IP Security Option*
- RFC1042, *Standard for Transmission of IP Datagrams over IEEE 802 Networks*
- RFC1122, *Requirements for Internet Hosts - Communication Layers*
- RFC1157, *Management Information Base for Network Management of TCP/IP-based Internets*
- RFC1213, *Management Information Base for Network Management of TCP/IP-based internets: MIB-II*
- RFC1253, *OSPF Version 2 Management Information Base*
- RFC1354, *IP Forwarding Table MIB*

- RFC1517, *Applicability Statement For the Implementation of Classless Inter-Domain Routing (CIDR)*
- RFC1518, *An Architecture for IP Address Allocation with CIDR*
- RFC1519, *Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*
- RFC1541, *Dynamic Host Configuration Protocol*
- RFC1577, *Classical IP and ARP over ATM*
- RFC1583, *OSPF Version 2*
- RFC1657, *Border Gateway Protocol version 4 (BGP-4) MIB*
- RFC1701, *Generic Routing Encapsulation*
- RFC1702, *Generic Routing Encapsulation over IPv4 networks*
- RFC1723, *RIP Version 2 Carrying Additional Information*
- RFC1724, *RIP Version 2 MIB Extension*
- RFC1745, *BGP4/IDRP for IP-OSFP Interaction*
- RFC1771, *Border Gateway Protocol 4 (BGP-4*)
- RFC1772, *Application of the Border Gateway Protocol in the Internet*
- RFC2003, *IP Encapsulation within IP*
- *RFC2236, Internet Group Management Protocol (IGMP), version 2* router functionality
- RFC2334, *Server Cache Synchronization Protocol*
- RFC2338, *Virtual Router Redundancy Protocol*
- *RFC2362, Protocol Independent Multicast - Sparse Mode (PIM-SM)*
- RFC2474, *DiffServ Field Definition*
- RFC2597, *Assured Forwarding PHB Group*
- RFC3246, *An Expedited Forwarding PHB*

Multiservice Switch IP fundamentals

Internet Protocol (IP) enables Nortel Networks Multiservice Switch nodes to provide IP virtual private network (VPN) capabilities across Multiservice Switch networks. Multiservice Switch networks use virtual routers (VRs) to provide IP connectivity between Multiservice Switch nodes. Every node can support numerous VRs.

Multiservice Switch supports the VIPR solution and two VPN solutions, RFC2764 and RFC2547. See the NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals* for conceptual information about RFC2764 and RFC2547. For information about VIPR, see Multiservice Switch virtual routers (page 21).

VIPR (Virtual IP Router) solution: A VIPR solution is a managed IP service solution used in a single customer environment; either carrier or enterprise. It may involve 1 or more virtual routers, depending on addressing requirements, or routing requirements.

IP VPN solution: An IP VPN is a managed IP service offered by a carrier to an enterprise customer. The IP VPN service provides secure and reliable connectivity, management, and addressing (equivalent to that available on a private network) over a shared public network infrastructure.

Multiservice Switch IP can simultaneously manage different software applications and types of traffic. When services are running on a common network facility, Multiservice Switch nodes allow you to consolidate bandwidth usage. Multiservice Switch offers a feature-rich IP interconnect service that provides high reliability and advanced packet security.

Navigation

Application and feature names for Multiservice Switch IP

The table lists the functionality provided by Nortel Networks Multiservice Switch IP nodes, the associated software application name, and the associated feature name. Use this information when you need to know the software application to download and feature name to link to a logical processor type (LPT). For information about downloading application software to a Nortel Networks Multiservice Switch node, see NN10600-270 *Nortel Networks Multiservice Switch 7400/15000/20000 Software Installation*.

Application and feature names for Multiservice Switch IP

Functionality	Software application name	Provisionable feature name
IP	ip	ip
ATM MPE	wanDte	atmMpe
ATM MPE soft PVCs	wanDte	atmMpe, atmMpeSpvc
Frame relay DTE	wanDte	FrameRelayDte
IP-optimized DLCI	frameRelay	frUniIpOptimized
Ethernet	ip, ethernet	ip, ethernetMedia
Ethernet	ip, ethernet	ip
Point-to-point protocol (PPP)	wanDte	PPP
IP class of service (CoS)	ip	ipCos
IP differentiated services (DiffServ) 1	ip	ipDiffServ
IP flow filters	ip	ipFilter
Border gateway protocol	ip	BGP
1 Feature ipDiffServ is required for the DiffServ profile for the interface (*Vr Ip DiffServ*) but not for the DiffServ domain for the router (*Vr Dsd*).		

IP processor cards

Function processors (FPs) provide interface ports that connect network communications facilities to Nortel Networks Multiservice Switch nodes. FPs support and execute real-time processes that are essential to service delivery.

IP services running on Multiservice Switch 15000 and Multiservice Switch 20000 nodes require CP3 CP and an FQM-based, PQC2.0-based, PQC12-based, or GQM-based FP. The CP2 CP, CQC-based, PQC12-based Ethernet, and SBIC-based FPs for IP are supported on Multiservice Switch 7400 nodes only. For more information on the FPs over which the IP service operates, see NN10600-551 *Nortel Networks Multiservice Switch 7400/15000/20000 FP Configuration Reference*.

The VPN extender card (VpnXc) is a special server card that you can use to increase the scalability of IP VPN services. This card has its own dedicated processor and memory and acts as the IP VPN control plane, hosting all IP VPN virtual routers (VCGs and customer VRs). Note that the Management VR is supported only on the CP. For more information on the VPN extender card, see NN10600-170 *Nortel Networks Multiservice Switch 7400 Hardware Description* and NN10600-120 *Nortel Networks Multiservice Switch 15000/20000 Hardware Description*.

IP protocol suite

Transmission control protocol/Internet protocol (TCP/IP) is a group of protocols that defines a common set of rules and standards that enable networks and hosts to communicate. IP is the routed, or network layer, protocol of TCP/IP and is one of the most popular internetworking protocols. Most internetworks support TCP/IP, whether or not TCP/IP end systems are present.

When you add an *Ip* component the system automatically adds supporting transmission control protocol/internet protocol (TCP/IP) processes such as address resolution protocol (ARP), internet control message protocol (ICMP), relay broadcast (RelayBC), user datagram protocol (UDP), and transmission control protocol (TCP).

Nortel Networks Multiservice Switch nodes support the following router management software applications, all of which are part of the TCP/IP architecture:

- Internet control message protocol (ICMP) (page 19)
- Transmission control protocol (TCP) (page 19)
- User datagram protocol (UDP) (page 19)
- File transfer protocol (FTP) (page 19)
- Telnet (page 19)

Internet control message protocol (ICMP)

Internet control message protocol (ICMP) provides feedback from an IP router or gateway to a source host. ICMP messages are sent in several situations: for example, to report resource or routing problems or to report a shorter available route to a destination. Nortel Networks Multiservice Switch systems use ICMP echoes and echo replies to verify the reachability of routers or end systems. See RFC792 for more information.

Transmission control protocol (TCP)

Transmission control protocol (TCP) is a connection-oriented transport-layer protocol. TCP provides reliable, robust, and adaptable data transfer between end-system upper layer protocols. TCP assumes that simple, potentially unreliable, data transmission services are available from lower-level protocols. See RFC793 for more information.

User datagram protocol (UDP)

User datagram protocol (UDP) defines the use of unacknowledged datagrams. UDP packets are useful for very low-priority data or for very high-reliability networks. UDP is also useful when an application already provides an integrity function and does not need to duplicate that function by using TCP. See RFC768 for more information.

File transfer protocol (FTP)

The file transfer protocol (FTP) provides a robust file transfer mechanism for data transfer between IP hosts. FTP transfers files between the file system on the node and a UNIX server. Once a connection is established, the node requests the appropriate account information (including security information) before establishing a session. See RFC959 for more information.

Telnet

Telnet allows a valid user access to a terminal or command process on a remote system such as the operator process in a Nortel Networks Multiservice Switch system. The Multiservice Switch system supports both Telnet client and server connections. See RFC854 for more information.

IP addressing protocols

A virtual router uses IP addressing protocols to map an IP address to the correct physical address when it needs to send data across a physical network. Nortel Networks Multiservice Switch nodes support the following IP addressing protocols:

- Address resolution protocol (ARP) (page 20)
- Reverse ARP (RARP) (page 20)
- Proxy ARP (page 20)
- Inverse ARP (InARP) (page 21)

- Bootstrap protocol (BOOTP) (page 21)

Address resolution protocol (ARP)

The address resolution protocol (ARP) is a mechanism for mapping 32-bit IP addresses to 48-bit Ethernet hardware addresses. The hardware address is a concatenation, or joining, of two numbers: a vendor ID number, centrally assigned by the IANA, and a unique serial number assigned by the vendor for each hardware unit. This hardware address, termed the media access control (MAC) address usually has significance only on the Ethernet wire.

Nortel Networks Multiservice Switch system implementation of ARP supports the following capabilities:

- removal of out-of-date ARP cache data
- configurable cache data timeout
- translation of encapsulation information between Ethernet and IEEE 802.3 networks

Ethernet and frame relay media support ARP. For more information about ARP, see RFC826.

Reverse ARP (RARP)

Reverse address resolution protocol (RARP) determines or assigns a particular station's IP address when only the station's MAC address is known. There are many reasons why an end system does not already have an IP address. The end system can be a diskless workstation homed off a server. Or, the end system can be a portable computer belonging to an itinerant employee sharing a pool of IP addresses with other itinerant employees. Nortel Networks Multiservice Switch systems cannot currently act as a RARP server. RFC903 defines RARP.

Proxy ARP

The proxy ARP is used to help an IP device locate a destination device, when the destination device is on a remote IP network or wire. When a source station broadcasts an ARP request on the local wire, and there is no station matching the destination IP address on the wire, the source does not receive an ARP response from the actual destination. Instead, the router derives the destination's IP wire address and searches for a match in its IP routing table. If the destination IP wire address is present in the routing table, the router responds with its own MAC address, in effect telling the source that the router's MAC address is the destination station's MAC address. The source IP station has no idea that the destination is on another wire. Nortel Networks Multiservice Switch systems fully support proxy ARP. RFC1027 defines proxy ARP.

Inverse ARP (InARP)

The inverse address resolution protocol (InARP) is used to determine a remote router's IP address on a particular ATM or frame relay connection. This is the local ATM or frame relay address of a permanent virtual circuit (PVC) to a remote router. Nortel Networks Multiservice Switch systems fully supports InARP. RFC1293 defines InARP.

For more information, see Inverse ARP scalability (page 30).

Bootstrap protocol (BOOTP)

The bootstrap protocol (BOOTP) is a UDP/IP-based protocol which allows a booting host to configure itself dynamically and without user supervision. BOOTP provides a means to notify a host of:

- its assigned IP address
- the IP address of a boot server host
- the name of a file to be loaded into memory and executed
- the local subnet mask
- the local time offset
- the addresses of default routers
- the addresses of various Internet servers

Nortel Networks Multiservice Switch systems support the BOOTP relay agent functionality described in RFC951 and RFC1542.

Multiservice Switch virtual routers

Nortel Networks Multiservice Switch virtual routers (VRs) provide IP connectivity between Multiservice Switch nodes and CE devices. Virtual routers can be used to:

- provide IP connectivity for Network Management, in an existing Multiservice Switch Layer 2 network (for example ATM or Frame Relay).
- provide IP connectivity between CE devices over a WAN network; in a carrier or enterprise scenario. It takes advantage of the strong Multiservice Switch Layer 2 WAN capabilities.
- allow a carrier to offer secure and reliable IP VPN services to customers.

Multiservice Switch virtual routers also provide a software emulation of physical routers. A VR has two main functions:

- constructing routing tables describing the paths to networks or subnetworks
- forwarding or switching packets to the final destination network or subnetwork

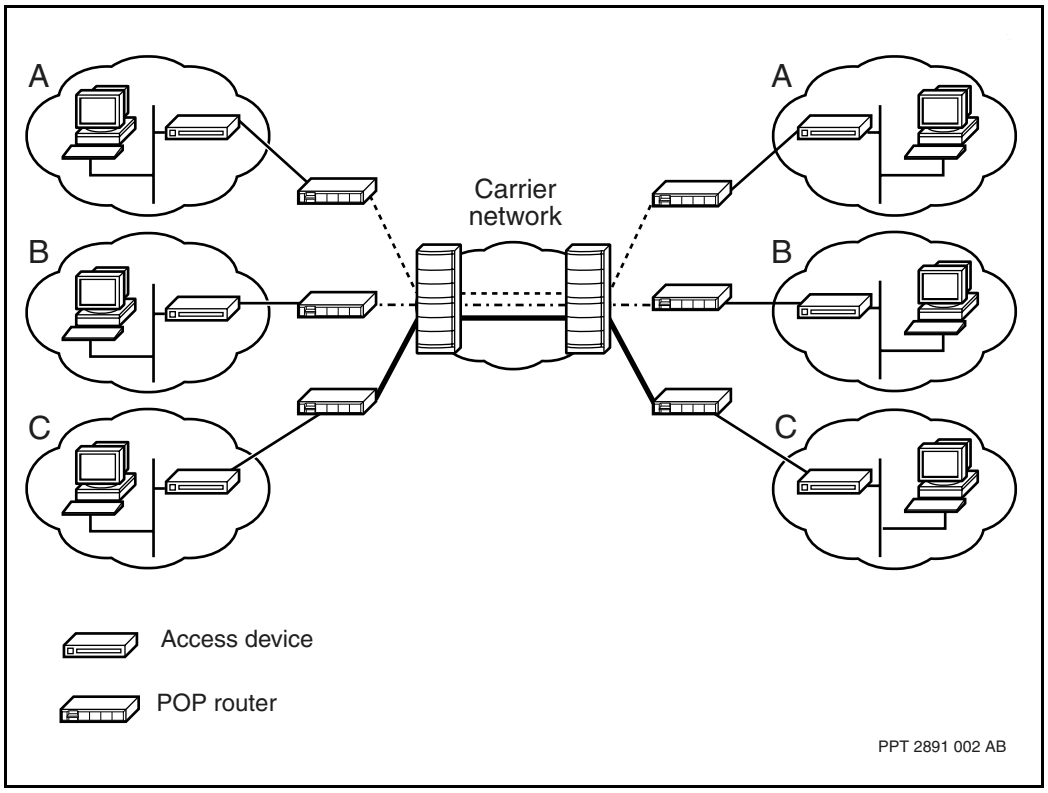
Each VR is an instance of a routing protocol used over a unique set of IP ports, point-to-point protocol (PPP) sessions, frame relay data link connection identifiers (DLCIs), and ATM virtual circuits (VCs). A VR coexists with other Multiservice Switch facilities on the same node.

Virtual routers on Multiservice Switch nodes can perform the functions of independent physical routers, forwarding packets to the correct destination while isolating each customer's traffic. Virtual routers provide a cost-effective alternative to using many separate hardware routers to provide multiple customer routing over a shared network. Carriers can therefore share backbone networks more effectively. See the figures Traditional router configuration (page 23) and Multiservice Switch virtual router configuration (page 23) to see how VRs eliminate the need to use separate physical routers.

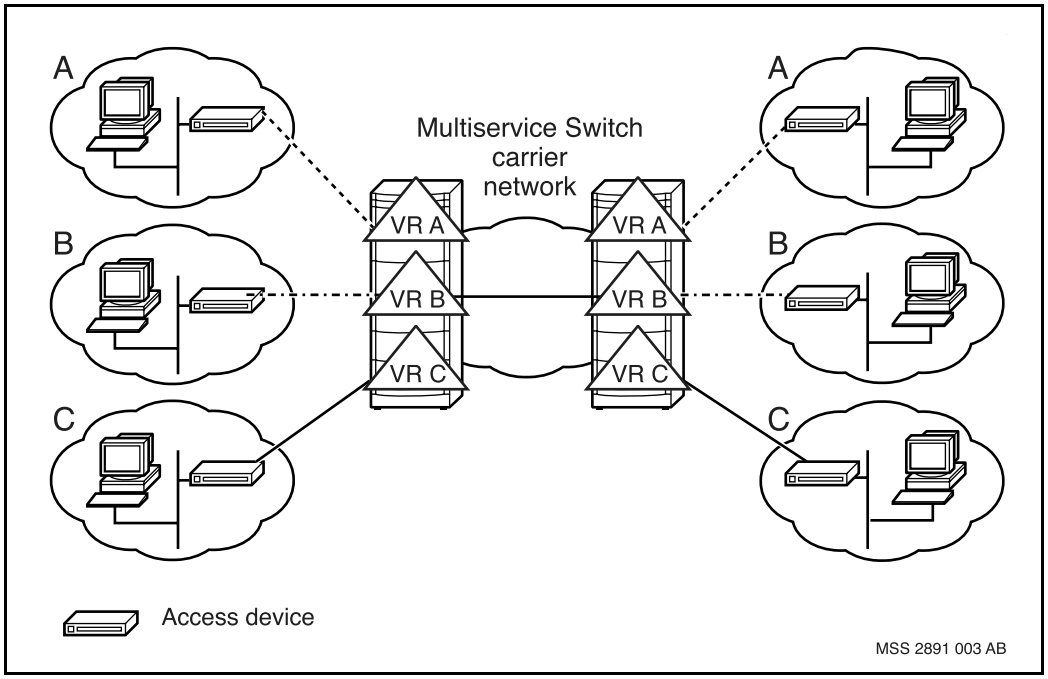
VRs have independent IP routing tables and are isolated from each other. These separate routing capabilities provide each enterprise customer with the appearance of a dedicated router that guarantees isolation from other customers while running on shared switching and transmission resources. This means that a customer's IP addressing space can overlap with another customer's address space. The IP addresses need only be unique within a customer's domain.

Multiservice Switch nodes can support multiple virtual routers. Using multiple VRs on a node enables carriers to support multiple isolated networks on the same platform by assigning each network to its own virtual router. See the figure Multiservice Switch virtual router configuration (page 23) for more information.

Traditional router configuration



Multiservice Switch virtual router configuration



For more information on virtual routers and their functionality, see the following sections:

- Management virtual router (page 24)
- Customer virtual router (page 24)
- Customer Edge (CE) IP device (page 25)
- Virtual connection gateway (page 25)
- Virtual router memory management (page 26)
- Source routing option (page 27)
- Cache table size (page 27)

Management virtual router

The management VR is a Nortel Networks Multiservice Switch virtual router that provides a single point of external entry into the node. You can also use the management VR to manage all customer VRs that reside on the node. The figure Management access for customer VRs (page 25) illustrates the use of a management VR.

The first VR you create on a Multiservice Switch node becomes, by default, the management VR. This means that even on a node running a single VR, that VR has all the features associated with the management VR. Once you activate your provisioning view, you cannot designate any other VR as the management VR.

A single TCP agent running under the management VR allows external access to the node from a workstation running network management system software through telnet, using TCP or FTP. You can also manage all VRs on the node through a single SNMP agent running under the management VR.

For information about configuring a management virtual router, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Customer virtual router

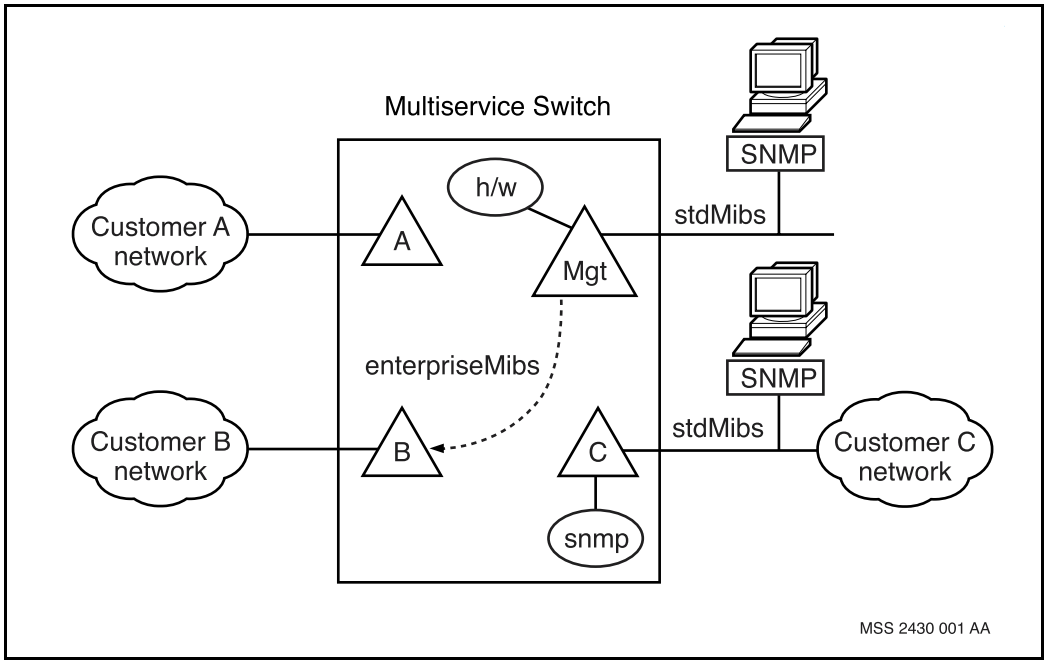
Configuring customer VRs is the same as the management VR with the exception that the customer VRs are restricted as to the protocols and interfaces they support. Management access for customer VRs is disabled meaning that users cannot set up Telnet sessions to any of the interfaces on the customer VR.

You can manage these VRs through SNMP or Nortel Networks Multiservice Data Manager on each VR (stdMibs), or enterpriseMibs on the management VR. The figure Management access for customer VRs (page 25) illustrates the management of customer VRs.

Attention: For security purposes, you should restrict SNMP access to customer VRs to designated personnel only.

For information about configuring a customer virtual router, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Management access for customer VRs



Customer Edge (CE) IP device

IP enables Nortel Networks Multiservice Switch nodes to connect with Customer Edge (CE) IP devices and provide IP virtual private network (VPN) capabilities across Multiservice Switch networks. Multiservice Switch networks use VRs to provide IP connectivity between Multiservice Switch nodes and with CE devices. Every node can support numerous VRs.

Virtual connection gateway

In a typical Nortel Networks Multiservice Switch IP VPN implementation, CE routers connect to a customer VR assigned to that enterprise. Each customer VR on the node connects to a common VR for the node, called the Virtual Connection Gateway (VCG).

The VCG aggregates traffic from the customer VRs and provides a single outbound connection into the wide area network (WAN) for all individual customer traffic on the node. The VCGs link all Multiservice Switch nodes that provide IP VPN functionality, and provide connectivity between customer VRs in the same IP VPN through point-to-multipoint (PTMP) IP tunnels.

The independence of customer VRs' traffic and control plane data is maintained within the VCG: tunnels provide logical separation of customer VR data traffic, and all customer VR routing data managed by the VCG is done on a per-customerVR basis, so as to preserve the requirement that customerVR address spaces need only be unique within a customer's domain.

For information about configuring a virtual connection gateway, see NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management*.

Virtual router memory management

The number of routes in a VR's routing database affects its memory consumption. A memory limit is assigned to a VR by setting the value of attribute *Vr Mm vrMaxHeapSpace* as a percentage of total available CP heap memory. As memory usage for the VR increases and the pre-defined thresholds are crossed, alarms are generated. See the table VR memory thresholds (page 26).

If memory usage reaches 101% of *Vr Mm vrMaxHeapSpace*, the VR is automatically locked. When this occurs, either reduce the number of routes propagated to the VR (for example, through route summarization) or reconfigure *Vr Mm vrMaxHeapSpace* to a larger value. Then, manually unlock the VR so that it can provide service again. You need to take these steps before unlocking the VR or else it will continue to exhaust its allocated memory and repeat the locking behavior.

It is strongly recommended that *vrMaxHeapSpace* be left to its default value (100%) for VRs configured as virtual connection gateways (VCGs).

VR memory thresholds

Percentage used of *Vr Mm vrMaxHeapSpace*	**Network action**
101	Set critical alarm and lock VR
99	Clear critical alarm and replace with major alarm Note: VR must be unlocked manually.

(1 of 2)

VR memory thresholds (continued)

Percentage used of *Vr Mm vrMaxHeapSpace*	Network action
90	Set major alarm
85	Clear major alarm and replace with minor alarm
80	Set minor alarm
75	Clear minor alarm

(2 of 2)

Source routing option

Source routing is an option specified in the IP header that allows the originator of a packet to specify a particular route to its destination.

You can enable or disable the processing of input datagrams that have a source IP option on a VR basis using the *sourceRoute* provisionable attribute. See NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management* for information about how to configure this option.

Cache table size

The cache management system (CMS) allows you to configure the IP local cache table size. This enables you to fine tune network performance by provisioning cache table size based on resource demand.

When planning your CMS

- determine the optimum memory requirements for all LPs, and adjust your cache table sizes accordingly. Carefully consider the type and amount of traffic being run on LPs.
- ensure an increase in cache table size does not adversely impact IP traffic (as long as the cache table sizes are optimized as discussed above).
- be aware that a decrease in cache table size can impact IP traffic in the case where the number of cache entries is larger than the newly provisioned cache table size permits. In this situation, the related protocol traffic is blocked during the adjustment period. However, if cache table sizes are optimized, there is no adverse impact on IP traffic.
- consult the *Passport IP VPN Engineering Guidelines* for information on cache table entry allocation behavior and recommendations.

Nortel Networks Multiservice Switch systems create the local IP cache table on a logical processor (LP) as soon as the first inbound protocol port is enabled on that LP. The system creates cache tables using default values. For

provisioning information, see the section on configuring IP on a virtual router in NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

The CMS also offers local cache monitoring and control capabilities through the component administration system (CAS) standard interface. The *Cache* component is a dynamic subcomponent of the *Ip* component. It represents the IP cache table on an LP and contains the operational attributes that allow for cache table monitoring. For information on monitoring IP cache tables, see the section on monitoring and testing in NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Multiservice Switch virtual media

You can configure connectivity between some of your VRs on a Nortel Networks Multiservice Switch node. By default, VRs on a node are completely isolated from one another for security purposes. VRs can be connected using hardware or software connections.

Hardware connections between VRs

You can physically link different VR ports using a hairpin connector. This type of connection provides maximum performance, but uses two physical ports per connection.

Software connections between VRs using PVG software

If the card type at each end of the connection supports Multiservice Switch Voice Gateway (PVG) and the PVG software is loaded, you can set the Custspec component on both VRs to either PVG or CDMA to enable a hardware datapath between the VRs. Hardware datapath connections are used only for tandem Multiservice Switch queue controller (PQC) based FP-to-FP traffic. Hardware datapath connections operate in one of three modes:

- interVrConnection, which is used for interconnectivity between VRs on the same Multiservice Switch module
- alwaysUpInterface, which is used to provide connectivity for VSP cards
- alwaysUpSummary, which is used to provide connectivity for local public tunnel endpoints

Virtual media on Multiservice Switch nodes using hardware datapaths support the following routing and forwarding functions:

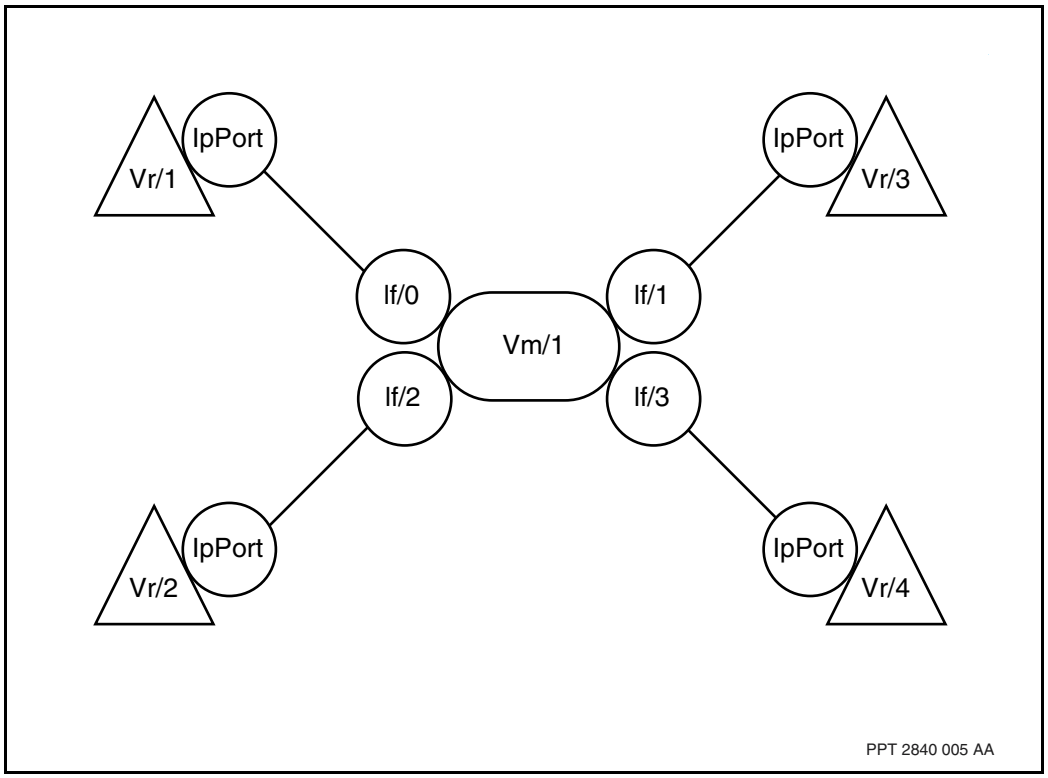
- IP ARP and IP ARP Reply
- IP datagram forwarding through ARP and static route definitions
- RIP

Software connections between VRs without PVG software

If either card at either end of the connection does not support Multiservice Switch Voice Gateway (PVG) or you do not have PVG software loaded, you can emulate a physical connection between VRs by configuring IP-only connectivity in the software using the *Interface* (*If*) subcomponent of the *VirtualMedia* (*Vm*) component. This type of connection supports Call Processor traffic, and is suitable only for low levels of traffic due to the processing load placed on the cards.

The *Vm If* component provides virtual (as opposed to physical) next-hop functionality between VRs. You can enable connectivity between different VRs by linking them through an IP port to different instances of the *If* subcomponent under the same *Vm* component. VR connectivity through a software link (page 29) illustrates the relationship between the *Vm* component and the VRs.

VR connectivity through a software link



You can add a *Vm* component if you want to provision an always-up IP interface. A virtual media application is not associated with a physical port. Since logical IP interfaces under the virtual media application are defined independently of any physical media, they remain up even though individual

links to the node might lose connectivity. An IP address associated with the virtual media protocol is always reachable as long as the node itself remains connected to the network.

Virtual media on Multiservice Switch nodes using software datapaths support the following routing and forwarding functions:

- IP ARP and IP ARP Reply
- IP datagram forwarding through ARP and static route definitions
- OSPF
- RIP
- BGP-4

For information about configuring virtual media, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Inverse ARP scalability

You can reduce the number of inverse ARP requests generated by a Nortel Networks Multiservice Switch node by linking an individual IP logical interface to a particular frame relay or ATM connection. When used on a customer VR in an IP VPN, this configuration improves IP VPN scalability.

Background

When a frame relay or ATM connection comes up, the virtual router linked to that connection is notified. The virtual router then sends an inverse ARP request to the other end.

When there is no explicit association between an IP logical connection and the connection, Nortel Networks Multiservice Switch systems do not know if an individual connection is associated with only one IP subnet. Therefore, when there is more than one IP logical interface configured on a protocol port, an inverse ARP request is sent across the connection for each IP logical interface on the protocol port.

Inverse ARP scalability description

You can reduce the number of inverse ARP requests by giving each CE its own subnet where each connection is known as a subconnection. When a frame relay or ATM subconnection comes up, the virtual router linked to that subconnection is notified and it sends only one inverse ARP request to the other end.

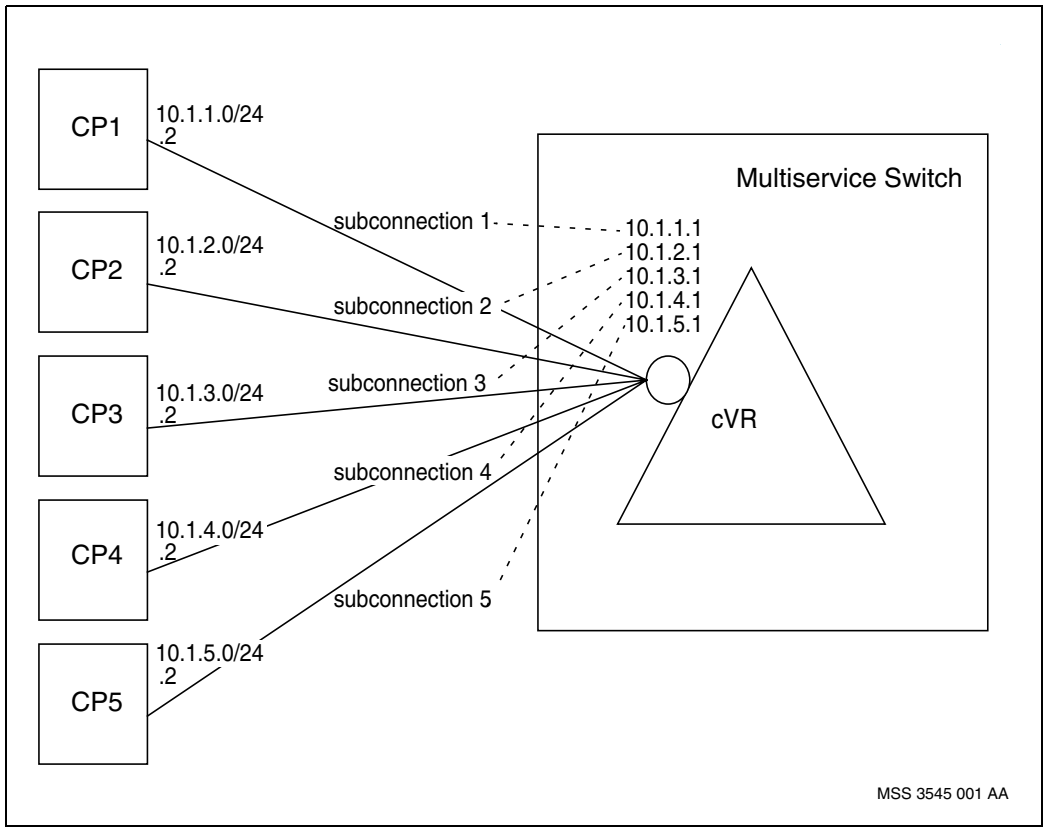
This configuration is shown in the figure Example of inverse ARP scalability (page 31) and includes the following steps:

1 Configure multiple frame relay or ATM connections on the same protocol port where each connection leads to a different CE device.

2 Configure an individual IP logical interface against the port for each CE device (that is, for each subconnection).

3 Using attribute *Vr ProtocolPort IpPort LogicalIf linkToMediaConnection*, link each IP logical interface to the corresponding subconnection to the CE device.

Example of inverse ARP scalability



When you configure OSPF on IP logical interfaces with links to individual subconnections, be aware of the following:

- It is recommended that OSPF be configured in non-broadcast mode. Broadcast/multicast mode causes hello packets to each IP logical interface to be sent over all connections. This causes unnecessary network traffic and may raise unnecessary traps or alarms.
- If an individual connection goes down, it is not detected immediately by OSPF unless all connections attached to the protocol port are down. This occurs because OSPF detects interface state changes at the protocol port

level, not at the individual connection level. Instead, when an individual connection goes down, the failure is detected after the router dead interval (attribute *Vr Pp IpPort LogicalIf OspfIf rtrDeadInt*) corresponding to the neighbor node on the far end of the connection has expired.

Virtual LAN

Virtual LAN (VLAN) provides virtual connections on an Ethernet interface to a VR. A service provider or enterprise customer can map individual VLANs within the port to different VRs instead of mapping a whole Ethernet port to a VR. The use of VLANs optimizes Ethernet port usage and port bandwidth.

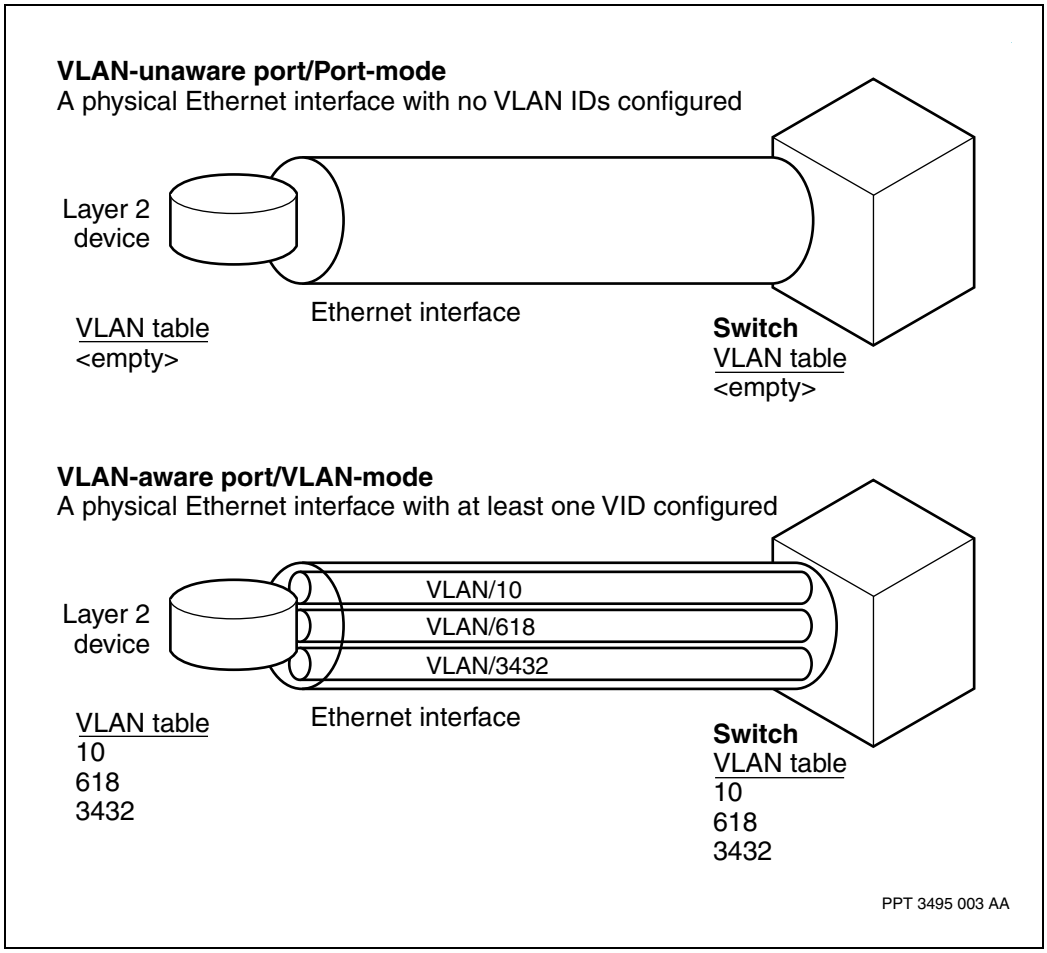
Virtual LAN (VLAN) provides virtual connections on an Ethernet interface to a VR. A service provider or enterprise customer can map individual VLANs within the port to one or more VRs instead of mapping a whole Ethernet port to a VR. The use of VLANs optimizes Ethernet port usage and port bandwidth.

An Ethernet interface is logically channelized into separate VLANs by VLAN identifiers (VIDs). Each VID identifies a different service on the Ethernet interface. An Ethernet interface operates in VLAN-mode, when at least one VID is configured on that interface. The interface is then considered VLAN-aware.

When no VIDs are configured on the Ethernet interface, the interface is VLAN-unaware and operating in port-mode. For more information about operating modes, see Ethernet services operating modes (page 33).

For information about managing traffic with VLANs, see Ethernet traffic treatment (page 33).

Ethernet services operating modes



Ethernet traffic treatment

Ingress and egress network access traffic is managed according to the operating mode of the Ethernet interface. Network access traffic operates in port-mode or VLAN-mode, these modes dictate the behavior of the Ethernet interface. For additional information about the treatment of traffic types, see Treatment of IP media traffic types (page 34).

Both Ethernet v2.0 and IEEE802.3 layer-2 header encapsulation are accepted by the interface. However, only Ethernet v2.0 is transmitted from the interface. Traffic can be tagged in the following ways:

- Untagged: Ethernet traffic without an IEEE 802.1Q tag header.
- Priority-tagged: Ethernet traffic that has an IEEE 802.1Q tag header with the VID set to 0. User-priority bits (p-bits) in the header specify the priority of the packet.
- VLAN-tagged: Ethernet traffic that has an IEEE 802.1Q tag header with the VID set between 2 and 4094.

- VLAN-tagged: Ethernet traffic that has an IEEE 802.1Q tag header with the VID set between 1 and 4094.

An Ethernet interface operating in port-mode accepts only untagged and priority-tagged traffic. VLAN-tagged traffic is discarded in port-mode. Priority-tagged traffic received from the Ethernet interface must conform to IEEE 802.1Q layer-2 header encapsulation. On transmission, all traffic sent from the Ethernet interface is untagged.

All multicast traffic, both group multicast address and broadcast address frames, received from the Ethernet interface is accepted without filtering. NG: Paragraph modified and moved to page 62.

Attention: For the 8-port 10/100BaseT Ethernet FP, ensure that all devices on a LAN segment connected to this FP are controlled to avoid flooding the LAN segment with multicast packets that may result in denial of service.

Attention: For the 4-port 10/100BaseT Ethernet and 8-port 10/100BaseT Ethernet FPs, it is important to ensure that all devices on a LAN segment connected to this FP are configured to avoid flooding the LAN segment with multicast packets that may result in denial of service.

For more information regarding denial of service and control plane protection, see Control plane protection (CPP) (page 153).

Treatment of IP media traffic types

Traffic type	Ingress		Egress	
	Port-mode	VLAN-mode	Port-mode	VLAN-mode
untagged	accepted	not accepted (see note)	always transmitted	never transmitted
priority-tagged	accepted	not accepted (see note)	never transmitted	never transmitted
VLAN-tagged	discarded	• accepted for configured VIDs • discarded if VIDs are not configured	never transmitted	always transmitted based on configured VID

(1 of 2)

Treatment of IP media traffic types (continued)

Traffic type	Ingress		Egress	
	Port-mode	VLAN-mode	Port-mode	VLAN-mode
Attention: For the VIPR solution, untagged, priority-tagged, and port-VLAN ID tagged traffic are not accepted.				
Attention: For the VIPR solution, untagged, priority-tagged, and port-VLAN ID tagged traffic are not accepted by an Ethernet interface operating in VLAN mode.				
(2 of 2)				

An Ethernet interface operating in VLAN-mode manages untagged, priority-tagged, and VLAN-tagged traffic based on the configuration of the interface. The Ethernet interface can be configured to accept or discard any of these types of traffic. Both untagged and priority-tagged types of traffic are accepted if the port-VLAN is configured on the interface, otherwise this traffic is discarded.

The port-VLAN is a reserved VLAN that defines the treatment for untagged, priority-tagged, and port-VLAN ID tagged traffic on a VLAN-aware Ethernet interface. The port-VLAN ID tagged traffic has a port-VLAN identifier (PVID). The PVID represents the default VLAN identifier on the Multiservice Switch and is always set to 1.

VLAN-tagged traffic is treated in three ways:

1 Traffic tagged with the reserved VID (4095) is discarded.

2 Untagged, priority-tagged, and port-VLAN tagged traffic are all treated in the same way. Traffic is accepted if the port-VLAN is configured on the Ethernet interface.

Attention: The port-VLAN is not supported on the 8-port 10/100BaseT Ethernet FP.

Attention: The port-VLAN is not supported for IP.

3 Traffic tagged with a VID that has a corresponding VLAN configured on the Ethernet interface is accepted, otherwise that traffic is discarded.

All broadcast traffic is accepted by the Ethernet FPs. Only group multicast traffic supported by the Multiservice Switch is accepted by the 4-port 10/100BaseT Ethernet, 8-port 10/100BaseT Ethernet, and 4-port gigabit Ethernet FPs. All other group multicast traffic is ignored by the 4-port 10/100BaseT Ethernet, 8-port 10/100BaseT Ethernet, and 4-port gigabit Ethernet FPs.

IP virtual private networks (VPNs)

An IP VPN is a managed IP service offered by a carrier to an enterprise customer. The IP VPN service provides secure and reliable connectivity, management, and addressing (equivalent to that available on a private network) over a shared public network infrastructure.

See NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals* and NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management* for more information on the Nortel Networks Multiservice Switch IP VPN service.

Provisioning MTU size

The maximum transmission size (MTU) is the largest unit of data that a network's physical medium can transmit. It can be set at the media link or the protocol port. The default MTU size is set as follows:

- 9188 for AtmMpe
- 9180 for IP tunnels
- 1604 for FrDte
- 1500 for Ethernet
- MRU (maximum receive unit) of 18000 for PPP
- must be manually provisioned on protocol port

You can provision lower MTU values for certain media. Smaller MTU sizes can help control jitter in a real-time voice stream because small voice packets are no longer delayed by large data packets.

The minimum MTU value depends on the type of processor card type. See the table Minimum supported MTU sizes (page 37).

Set MTU size at

- *AtmMpe mtu*
- *Vr Ip Tunnel Msep mtu*
- *FrDte Rg mtuSize*
- *Ppp Link configInitialMru*
- *Vr ProtocolPort IpPort mtu*
- *Lp Eth maxFrameSize*

Attention: Setting the maximum frame size for the Ethernet interface, also sets the MTU for the upper layer protocols.

If set at the protocol port, the MTU must be within the valid range of the *IpPort* media type. If both the media and the protocol port MTU are set, the lowest of the two values becomes the MTU.

Attention: PQC-based FPs used to have the same minimum supported MTU sizes as CQC and SBIC-based FPs. Setting the MTU size on PQC-based FPs to less than its previous minimum (576 for IP tunnels and the protocol port and 262 for FrDte) can have a real time impact on FPs and throughput, so when using these small sizes contact your Nortel Networks representative for detailed engineering assistance.

Minimum supported MTU sizes

Media	CQC and SBIC	PQC	FQM
AtmMpe	256	256	n/a
IP tunnels	576	100	n/a
FrDte	262	80	n/a
PPP	68	68	n/a
Protocol port	576	80	80
Ethernet/port-mode	1518	1518	1518
Ethernet/VLAN-mode	n/a	1518	1518

Related information for Multiservice Switch IP

This section describes where to find information related to the following topics:

- IP media (page 37)
- IP routing protocols (page 39)
- IP features (page 40)

IP media

Nortel Networks Multiservice Switch nodes can provide customer access to the carrier network using the media listed in the table Multiservice Switch supported access media (page 38).

Multiservice Switch supported access media

Node type	Supported access media
Multiservice Switch 7400	ATM, frame relay using FrDte, frame relay using IP-optimized DLCIs, PPP, 10BaseT Ethernet, 100BaseT Ethernet
Multiservice Switch 15000 and Multiservice Switch 20000	ATM, frame relay using FrDte, frame relay using IP-optimized DLCIs, PPP, gigabit Ethernet

Multiservice Switch-supported core media are ATM, frame relay using FrDte, and MPLS.

Protocol ports represent physical instances of data link or media protocols. When you configure protocol ports, you must link them to the corresponding media. You can configure protocol port designations that follow a descriptive numbering convention to allow easy recognition of protocol port-attached media.

The table Where to find IP media information (page 38) tells you where to find more information about specific IP media.

Where to find IP media information

Media	Fundamentals	Configuration	RFC
IP over ATM	IP over ATM (page 46)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC1483
IP over frame relay	IP over frame relay using frame relay DTE (page 56) IP over frame relay using IP-optimized DLCIs (page 67)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC2427 RFC1490
IP over Ethernet	IP over Ethernet (page 71)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC894
	(1 of 2)		

Where to find IP media information (continued)

Media	Fundamentals	Configuration	RFC
IP over point-to-point protocol (PPP)	IP over point-to-point protocol (PPP) (page 74)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC1661
IP over multiprotocol label switching (MPLS)	NN10600-445 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Operations: Multiprotocol Label Switching*	NN10600-445 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Operations: Multiprotocol Label Switching*	RFC2702

(2 of 2)

IP routing protocols

Nortel Networks Multiservice Switch nodes support static routes as well as interior and exterior dynamic routing protocols. Interior routing protocols determine best paths within an autonomous system or enterprise. Exterior routing protocols determine best paths between autonomous systems. You can configure multiple dynamic routing protocols on one virtual router.

The table Where to find IP routing information (page 39) lists the IP traffic routing methods that Multiservice Switch systems support. It also tells you where to find fundamental and configuration information about specific IP traffic routing methods.

Where to find IP routing information

Routing method	Fundamentals	Configuration	RFC
Routing information protocol (RIP)	Routing information protocol (RIP) (page 93)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC1723 RFC1724
Open shortest path first protocol (OSPF)	Open shortest path first (OSPF) protocol (page 100)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC2178

(1 of 2)

Where to find IP routing information (continued)

Routing method	Fundamentals	Configuration	RFC
Border gateway protocol 4 (BGP-4)	Border gateway protocol 4 (BGP-4) (page 107)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC1771 RFC1772 RFC1745
Static routes	Static routes (page 123)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	n/a
Bootstrap protocol (BOOTP)	Bootstrap protocol (BOOTP) (page 21)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC951 RFC1542
Virtual router redundancy protocol (VRRP)	Virtual router redundancy protocol (page 132)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC2338

(2 of 2)

IP features

The table Where to find IP feature information (page 40) tells you where to find more information about specific Nortel Networks Multiservice Switch IP features.

Where to find IP feature information

Service	Fundamentals	Configuration	RFC
IP multicast	IP multicast (page 126)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC2236 RFC2262
IP class of service (CoS), including IP CoS to QoS mapping	NN10600-590 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Layer 3 Traffic Management Fundamentals*	NN10600-591 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Layer 3 Traffic Management Configuration*	RFC2474

(1 of 2)

Where to find IP feature information (continued)

Service	Fundamentals	Configuration	RFC
IP flow control	NN10600-590 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Layer 3 Traffic Management Fundamentals*	NN10600-591 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Layer 3 Traffic Management Configuration*	RFC2267
IP tunnels between IP networks	IP tunnels (page 140)	NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/ 20000 IP Configuration Management*	RFC1701 RFC1702 RFC2003
IP differentiated services	NN10600-590 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Layer 3 Traffic Management Fundamentals*	NN10600-591 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Layer 3 Traffic Management Configuration*	RFC2474 RFC3246 RFC2597
IP accounting	IP accounting (page 148)	NN10600-560 *Nortel Networks Multiservice Switch 7400/15000/ 20000 Accounting*	n/a

(2 of 2)

Planning Multiservice Switch IP configuration

This section describes the things you need to consider when planning your IP configuration in Nortel Networks Multiservice Switch networks.

Navigation

- Network considerations (page 42)
- Mapping the IP network (page 42)
- Multiservice Switch IP configuration sequence (page 44)

Network considerations

If your network plan includes other routers (from any manufacturer), you need to complete the following steps:

- Select a routing protocol. In cases where you are integrating the Multiservice Switch system into an existing network, choose the routing protocol to conform or interoperate with the existing routers.
- Gather relevant information about the networks on the other side of the remote routers including server addresses and special needs.
- If your network connects to other networks that are not under the control of your organization, you must plan security firewalls to prevent unauthorized access to the network.

Mapping the IP network

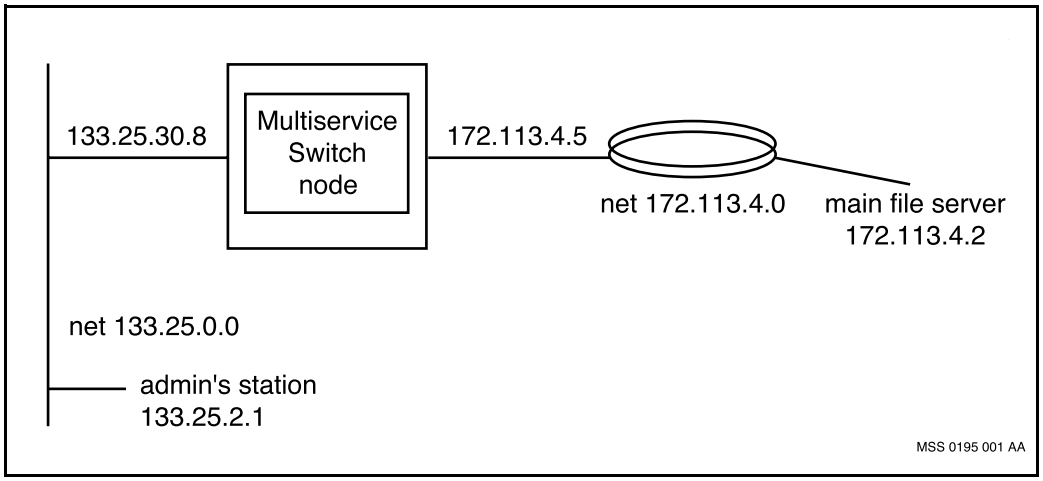
It is very important to have a usable representation of your network before configuring IP. If IP is already in use in your network, you probably only need a rough diagram showing the network numbers you need and the IP addresses assigned to the ports. Simple network diagram (page 43) illustrates a simple network map.

The networks shown in Simple network diagram (page 43) are established and only need to be joined by the Nortel Networks Multiservice Switch network. The only information that the installers and administrators need to

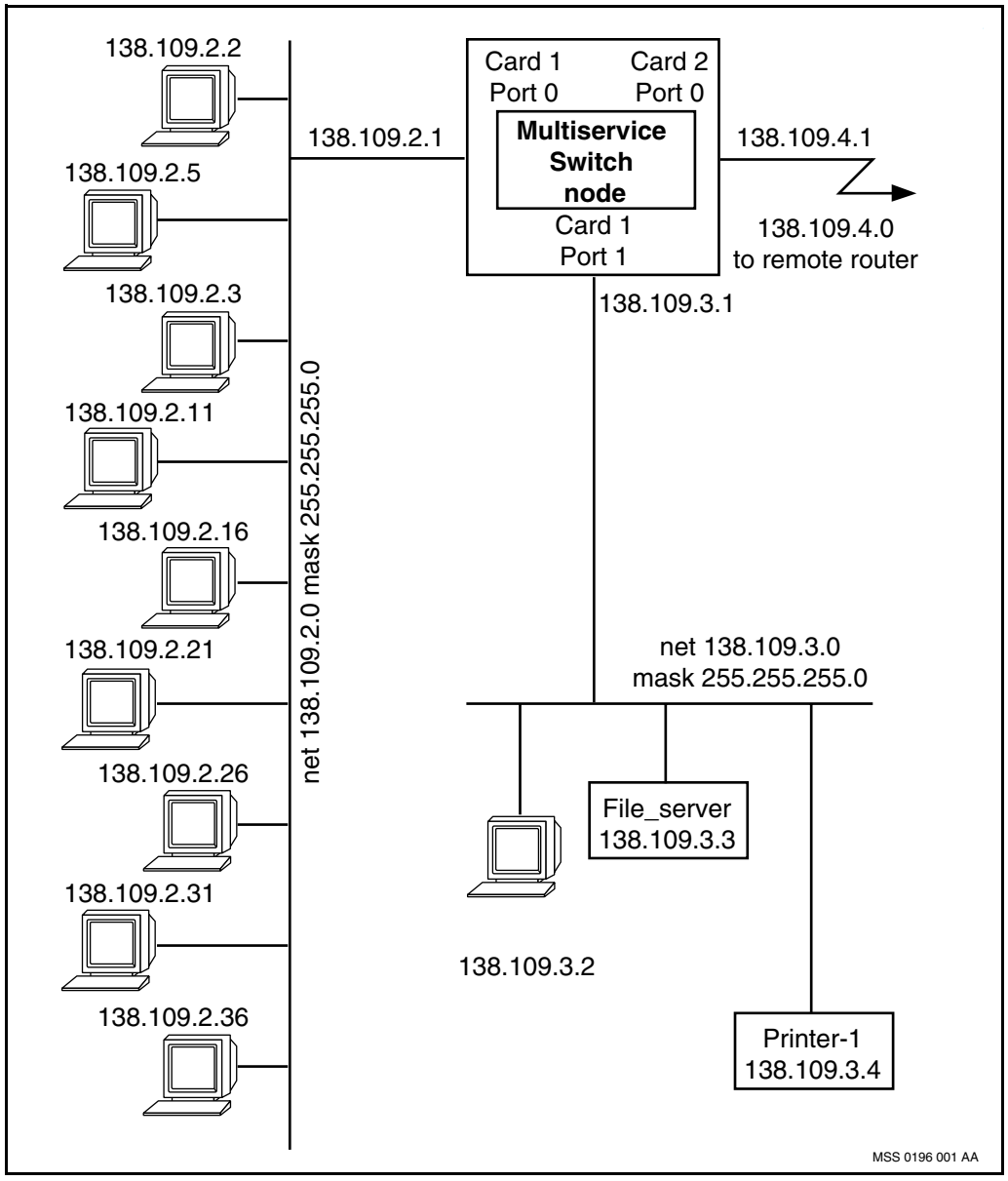
understand the network are the network addresses for the ports and the routing protocol currently in use. Each connected segment must have a unique network or subnetwork number.

However, if you are introducing IP at the same time you are installing the Multiservice Switch system, you can benefit from a map showing each node and its IP address. Detailed network diagram (page 44) is an example of one page of such a network map.

Simple network diagram



Detailed network diagram



Multiservice Switch IP configuration sequence

The table Multiservice Switch IP configuration sequence (page 45) provides a high-level view of the IP configuration process. You can use it to plan the end-to-end configuration of IP on Multiservice Switch nodes.

You might find it more efficient to download all the software for IP, virtual routers, and access media together, then proceed to configure all the LPs, LPTs, and FPs you need to operate IP, virtual routers, IP routing protocols, and IP services.

You must also configure all required IP media, such as ATM MPE, frame relay DTE, and point-to-point protocol (PPP), before you can configure IP. When you configure IP you will link the protocol ports of the virtual router to the IP media.

Multiservice Switch IP configuration sequence

Order	Task
1	Download all required software. See NN10600-270 *Nortel Networks Multiservice Switch 7400/15000/20000 Software Installation*.
2	Configure all required LPs and LPTs. See NN10600-270 *Nortel Networks Multiservice Switch 7400/15000/20000 Software Installation*.
3	Configure all required FPs. See NN10600-551 *Nortel Networks Multiservice Switch 7400/15000/20000 FP Configuration Reference*, NN10600-175 *Nortel Networks Multiservice Switch 7400 Hardware Installation, Maintenance, and Upgrade*.
4	Configure all required IP media. See NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.
5	Configure virtual routers and IP. See NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.
6	Configure all required routing protocols. See NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.
7	Configure IP features. See NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

IP over ATM

This section describes the implementation of IP over ATM in Nortel Networks Multiservice Switch networks.

Navigation

Overview of ATM MPE

The ATM multiprotocol encapsulation (MPE) interface is an access service that allows IP encapsulation over ATM in accordance with RFC1483. You can use the ATM MPE service to transmit IP traffic to interconnected external routers and other Nortel Networks Multiservice Switch virtual routers over an ATM network.

For information about FPs that support the ATM MPE service, see NN10600-551 *Nortel Networks Multiservice Switch 7400/15000/20000 FP Configuration Reference*.

ATM MPE media

The ATM MPE service allows IP traffic to be transmitted across the ATM network using the following two types of ATM MPE media:

- permanent virtual circuits (PVCs)
- soft PVCs

ATM MPE over PVCs

In a PVC, all the connection points through the network are defined, or nailed up. In this ATM MPE medium, standard encapsulation (using RFC1483) allows the system to interoperate with any other RFC1483 implementation (Nortel Networks or other vendor's equipment).

Nortel Networks Multiservice Switch ATM MPE service follows the specifications detailed in RFC1483, which describes two methods for carrying connectionless network traffic over ATM Adaptation Layer 5 (AAL5). The first method is Logical Link Control (LLC) encapsulation, where multiple upper layer protocols (ULPs) are carried over a single ATM Virtual Channel Connection (VCC). The second method is Virtual Circuit (VC) encapsulation, where only the IP protocol is permitted for each ATM VCC.

When you run the ATM MPE service on a CQC-based ATM FP, you must run the ATM MPE service in conjunction with an ILS Forwarder FP. IP forwarding over ATM on a CQC-based ATM FP alone is restricted to network management connectivity only.

ATM MPE over soft PVCs

The ATM MPE medium allows you to transmit IP traffic over soft PVCs in an ATM private network-to-network interface (PNNI) network. In a soft PVC, only the endpoints of the PVC are defined. PNNI routing provides route selection through the network between the endpoints. ATM MPE medium interoperates only within Nortel Networks Multiservice Switch networks. Soft PVCs provides simpler overall provisioning: only endpoints; not every hop through the ATM cloud. And, if there are multiple paths through the ATM PNNI cloud between the endpoints, PNNI provides a level of protection against link or node failures within the ATM PNNI cloud.

After the soft PVC is established, the dynamic component *AtmConnection* (*AtmCon*) is created by the system under the *Ac* component at both ends of the connection. The *AtmCon* component links to the ATM VCC through the *AtmIf Vcc Ep* component.

Carriers typically use soft PVCs to connect virtual connection gateways (VCG) across the backbone, for increased reliability through the ATM PNNI cloud. However, any customer VR in an ATM PNNI network can be connected over soft PVCs. ATM soft PVCs support only the LLC encapsulation method. For more information on IP over soft PVCs, see NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals* and NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management*.

There may be configurations where it is desired to connect an ATM MPE soft PVC in the following ways:

- The ATM MPE source PVC establishes an ATM connection through a PNNI network and terminates the connection on an ATM interface instead of an ATM MPE destination endpoint. In this case, the connection established terminates on an ATM UNI interface as a signalled or provisioned destination PVC. The destination VCC on the ATM UNI interface can be Vcc/0.32 to Vcc/0.255.

In this configuration, the ATM traffic management parameters for the ATM connection are configured using the ATM MPE source endpoint, and includes the peak cell rate and service category. If the destination end of the connection is configured using a provisioned destination PVC, the traffic management parameters for the destination VCC can be overridden. The peak cell rate and service category of both the ATM MPE source and the ATM UNI interface destination VCC should match.

- An ATM UNI interface source PVC establishes an ATM connection through a PNNI network and terminates the connection on an ATM MPE destination PVC.

In this configuration, there is more flexibility in what can be provisioned for the ATM traffic management parameters since more options are available for an ATM interface source PVC. Also, the VCC instances on the ATM UNI interface are not restricted to the range of Vcc/0.32 to Vcc/0.255 as in the ATM MPE source PVC to ATM UNI interface configuration.

For more information about ATM soft PVCs, see NN10600-702 *Nortel Networks Multiservice Switch 7400/15000/20000 ATM Routing and Signalling Fundamentals*.

For the steps used to provision these configurations, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Encapsulation methods

There are two methods for carrying connectionless network interconnect traffic over ATM Adaptation Layer 5 (AAL5):

- LLC encapsulation (page 48)
- VC encapsulation (page 50)

LLC encapsulation is supported on ATM MPE over PVCs and soft PVCs. VC encapsulation is supported on ATM MPE over PVCs only.

LLC encapsulation

Logical link control (LLC) encapsulation allows the ATM virtual circuits (VCs) associated with the ATM MPE interface to carry multiple protocols. For more information on LLC encapsulation, see RFC2684.

For more information on LLC encapsulation in Nortel Networks Multiservice Switch nodes, see

- LLC encapsulation for routed protocols (page 49)
- LLC encapsulation for bridged protocols (page 49)

LLC encapsulation for routed protocols

The protocol of the protocol data unit (PDU) is identified by prefixing the PDU with an IEEE 802.2 LLC header. The first octet is the Destination Service Access Point (DSAP), the second octet is the Source Service Access Point (SSAP), and the third octet is the Control (Ctrl) field. These fields indicate the type of PDU that follows.

An LLC header is followed by a SNAP header. A SNAP header exists when the LLC header has a value of 0xAA-AA-03. The first three octets of the SNAP header represent the Organizationally Unique Identifier (OUI) field; the next two octets represent the Protocol Identifier (PID) field.

The meaning of the PID field value depends on the OUI field value. For example, if the OUI field has the value 0x000000, the PID specifies an EtherType. The EtherType for IP, for example, is 0x0800.

LLC encapsulation for bridged protocols

LLC encapsulation for bridged protocols allows Nortel Networks Multiservice Switch nodes to internetwork with a bridge over an ATM link. Currently the only type of supported bridged media over ATM MPE is Ethernet. Although Multiservice Switch systems do not provide bridging functionality, it can perform the following:

- receive bridged Ethernet LLC encapsulation packets and provide IP forwarding on them
- transmit bridged Ethernet LLC encapsulation packets into the bridge network

As in LLC encapsulation for routed protocols, the LLC header must be equal to 0xAA-AA-03. The LLC header is followed by a SNAP header. The Organizationally Unique Identifier (OUI) field in the SNAP header must be the 802.1 organization code 0x00-80-C2. The type of the bridged media must be specified by the Protocol Identifier (PID) field. The PID must also identify whether the original Frame Check Sequence (FCS) is preserved within the bridged protocol data unit (PDU). See the table LLC encapsulation for bridged Ethernet PDUs (page 50).

Do the following to enable LLC encapsulation for bridged protocols:

- Set the *AtmMpe encaptype* attribute to llcBridgeEncap. Also, it is recommended that you set the *AtmMpe maxTransmissionUnit* attribute to 1524.
- Ensure that the other end of the connection is a bridged port running over an ATM interface.

When *AtmMpe encaptype* is set to llcBridgeEncap, a MAC address is automatically assigned to the protocol port of the VR to which the *AtmMpe* component is linked. Address Resolution Protocol (ARP) is then used on demand to discover the layer 2 addresses of the Ethernet hosts that internetwork with the Multiservice Switch node through the remote bridge.

Bridged termination is supported only on PQC2-based OC-3 and OC-12 FPs.

LLC encapsulation for bridged Ethernet PDUs

LLC header	OUI	PID	Padding	Supported media
0xAA-AA-03	0x00-80-C2	0x00-01 with preserved FCS	0x00-00	Ethernet
0xAA-AA-03	0x00-80-C2	0x00-07 without preserved FCS	0x00-00	Ethernet

VC encapsulation

Virtual circuit (VC) encapsulation allows the ATM virtual circuits (VCs) associated with the ATM MPE interface to carry one (and only one) protocol. Therefore, no protocol identifier is required since the Virtual Channel Connection (VCC) distinguishes between different protocols.

You configure the protocol type that is carried over a VCC, and must ensure that the protocol type is configured to the same value at both ends of the connection.

If the encapsulation type is *IpVcEncap*, Address Resolution Protocol (ARP) is not supported on that ATM MPE service. Since ARP is a protocol distinct from IP, no ARP packets can be transported on the ATM MPE service. If you use VC encapsulation, you must configure static ARP entries to ensure IP connectivity across the ATM network.

Inverse ARP on ATM

Inverse ARP provides a method for dynamically discovering the IP address at the remote end of a VCC. When inverse ARP is absent (for example, when the remote end does not support inverse ARP or VC encapsulation is used), the IP address of the remote end must be provisioned.

Attention: A full implementation of RFC1577 is not used, just the use of inverse ARP.

For more information related to inverse ARP on ATM, see Inverse ARP scalability (page 30).

Frame forwarding for IP traffic

Nortel Networks Multiservice Switch 7400 ATM MPE nodes support VCCs that terminate on CQC-based ATM FPs and on ATM IP FPs. If you are using CQC-based ATM FPs, you must configure the ATM MPE service in conjunction with an ILS Forwarder FP. Multiservice Switch 15000 and Multiservice Switch 20000 nodes support IP forwarding on ATM IP FPs only.

For more information, see the following sections:

- Frame forwarding on CQC-based ATM FPs for Multiservice Data Manager connectivity (page 51)
- Frame forwarding using the ILS Forwarder FP (page 52)
- Frame forwarding on ATM IP FPs (page 53)

Frame forwarding on CQC-based ATM FPs for Multiservice Data Manager connectivity

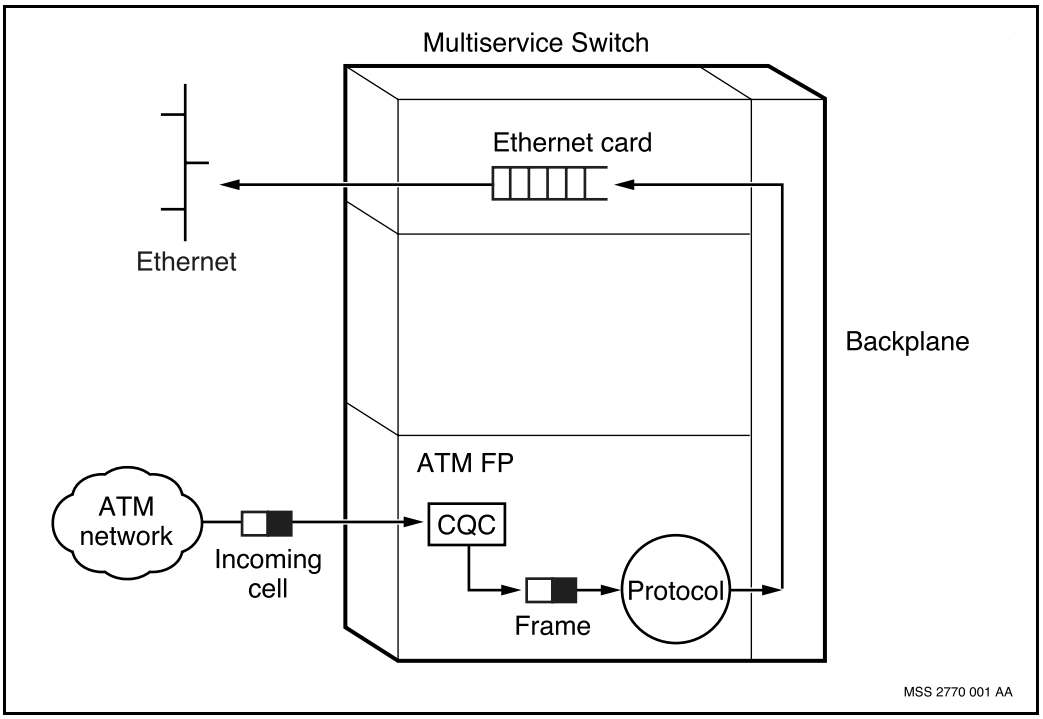
To support Multiservice Data Manager connectivity only, you can configure the ATM MPE service on a CQC-based ATM FP alone, for IP forwarding over ATM.

By default, Nortel Networks Multiservice Switch 7400 node frame forwarding decisions for IP traffic over ATM are made on the ATM FP's cell queue controller (CQC). The applicable protocol stack resides on the ATM FP. See the figure Frame forwarding on an ATM FP (page 52) for an illustration of the frame forwarding process when you use Multiservice Switch 7400 CQC-based ATM FP.

There are no special configuration procedures to enable this capability. If you do not configure the *ilsForwarder* attribute under the *AtmMpe* component and link it to an *Lp IlsForwarder* component, the ATM FP performs all the tasks necessary to handle the encapsulated IP frames. If you delete an existing *IlsForwarder* component from the current software view, all references to the *AtmMpe* component are also removed; this enables the ATM FP's frame forwarding functions.

For information about using Multiservice Switch 7400 CQC-based ATM FP in conjunction with an ILS Forwarder FP, see Frame forwarding using the ILS Forwarder FP (page 52).

Frame forwarding on an ATM FP



Frame forwarding using the ILS Forwarder FP

The ILS Forwarder FP is designed specifically for handling frames, and enhances the frame handling capability of frames coming in to a Multiservice Switch 7400 CQC-based ATM FP. You can use an ILS Forwarder FP in conjunction with a CQC-based ATM FP to provide higher frame forwarding performance. In addition, other services running on the CQC-based ATM FP do not have to share resources (such as CPU and memory) with the ATM FP's frame forwarding service.

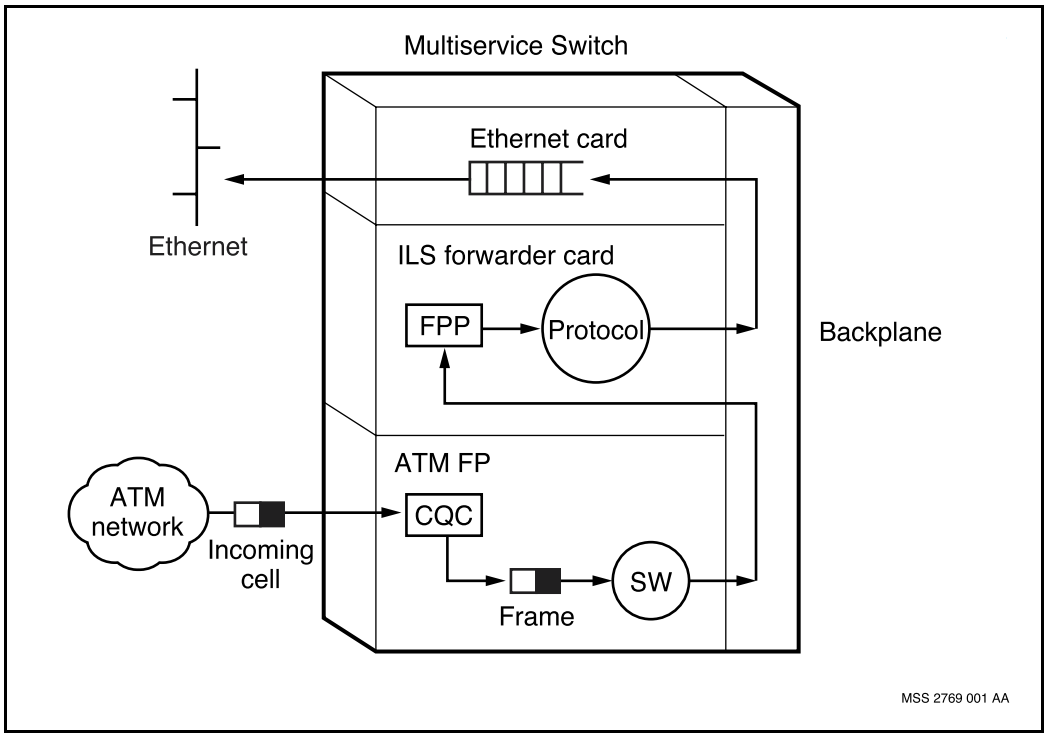
An ILS Forwarder FP makes the forwarding decisions, where

- one ATM FP links with multiple ILS Forwarder FPs
- multiple ATM FPs link to one ILS Forwarder FP

When you configure the *ilsForwarder* attribute under the *AtmMpe* component and link it to an *Lp IlsForwarder* component, all ATM MPE traffic that arrives on the CQC-based ATM FP is forwarded directly to the ILS Forwarder FP. The applicable IP protocol stack resides on the ILS Forwarder FP, and forwarding decisions are made with the assistance of the ILS Forwarder FP's fast packet processor (FPP) hardware.

The figure Frame forwarding on an ILS Forwarder card (page 53) illustrates the frame forwarding process when you use an ILS Forwarder FP.

Frame forwarding on an ILS Forwarder card



ILS Forwarder FP restrictions for ATM MPE

The ILS Forwarder FP has the following restrictions on the *AtmMpe* component:

- The total throughput of all the ATM connections forwarding to an ILS Forwarder FP should not exceed the maximum throughput of the ILS Forwarder FP.
- The maximum size (4475 bytes) of the frame that an *AtmMpe* component can accept is limited to the maximum size of the frame that passes through the FPP on the ILS Forwarder FP.
- The ILS Forwarder FP is supported on Nortel Networks Multiservice Switch 7400 nodes only. It is not supported on Multiservice Switch 15000 and Multiservice Switch 20000 nodes.

Frame forwarding on ATM IP FPs

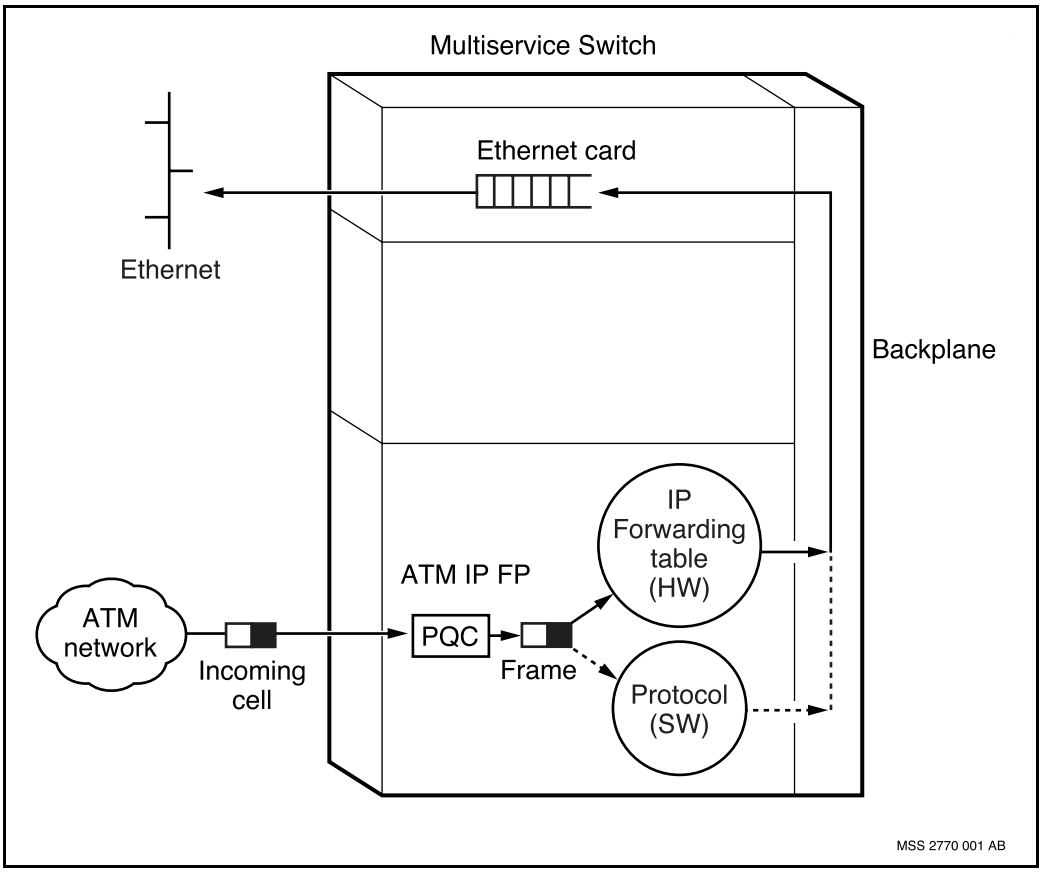
The ATM IP FP performs both hardware and software forwarding functions. Hardware forwarding is much faster than software forwarding, and operates independently of the software forwarding function. The ATM IP FP has specialized hardware to forward IP packets autonomously: it supports hardware lookups into IP forwarding tables, so it can forward almost all IP packets without the aid of the CP or the ATM IP FP's processor. See the figure Frame forwarding on an ATM IP FP (page 55) for an illustration of the frame forwarding process when you use an ATM IP FP.

To route IP packets properly, the hardware forwarding function must have a full IP forwarding table for each configured instance of a virtual router (VR). The IP software distributes a copy of these tables to each ATM IP FP. You can limit the number of routes stored in the ATM IP FP's hardware IP forwarding tables by configuring the *ipRoutesPoolCapacity* attribute under the *Lp Eng Fcrc Pqc Ov* component. This value applies to the entire LP, so all virtual routers on the LP should be taken into account and the capacity for IP routes set accordingly.

The ATM IP FP's software forwarding function processes traffic for routes that are not found in the hardware. Packets remain in the software datapath as long as the condition that caused them to take the software datapath persists. This can occur as a result of IP packet fragmentation, or when the maximum size of the hardware forwarding table is exceeded. Once the condition clears, traffic flow reverts back to the hardware path.

IP software running on an ATM IP FP does not require the use of an ILS Forwarder card. The *ilsForwarder* attribute under the *AtmMpe* component has no bearing on traffic received on the ATM IP FP. You can, however, use an ILS Forwarder card with Nortel Networks Multiservice Switch 7400 CQC-based ATM FPs in the same node as ATM IP FPs. If you set the *ilsForwarder* attribute under the *AtmMpe* component, IP packets that arrive on CQC-based ATM FPs are sent to the ILS Forwarder for processing, and packets that arrive on ATM IP FPs are processed locally.

Frame forwarding on an ATM IP FP



IP over frame relay using frame relay DTE

This section describes the Nortel Networks Multiservice Switch implementation of IP over frame relay using frame relay DTE. This is an alternative to IP over frame relay using IP-optimized DLCIs (page 67) as an access media. This section includes the following topics:

Navigation

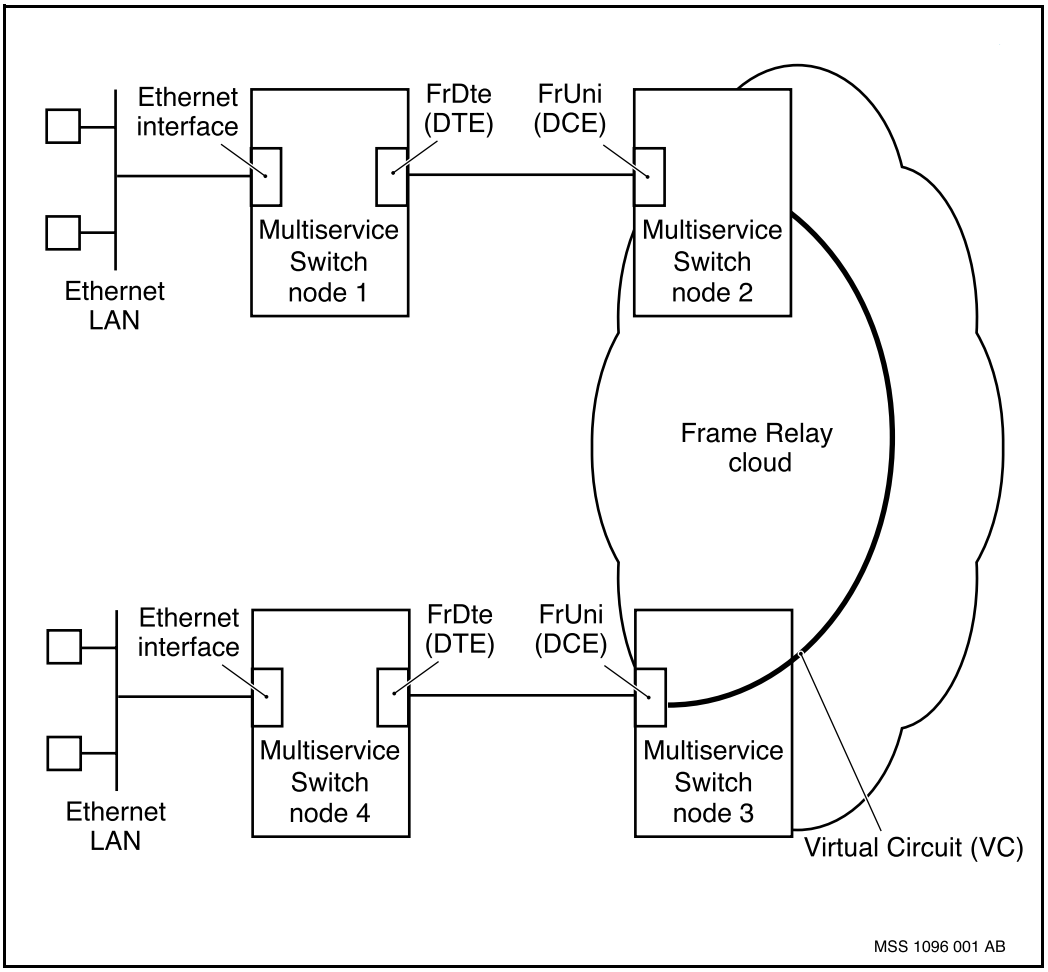
Overview of Multiservice Switch frame relay DTE (FrDte)

Nortel Networks Multiservice Switch frame relay connection, called a virtual circuit (VC), is provided through a standard interface between the user device and the network. The interface is called the user-to-network interface (UNI). The connection that attaches UNI to the VC is provided by data circuit terminating equipment (DCE). The connection that attaches UNI to a device is provided by data terminating equipment (DTE). The figure Frame relay overview (page 57) illustrates how frame relay works. This section describes the access software associated with the DTE endpoint of frame relay UNI. This access software is referred to as frame relay DTE and the provisionable component of the software is *FrameRelayDte* (*FrDte*).

Encapsulation as defined in RFC2427 and RFC1490 is supported. This specifies recognition of the control field and a one byte padding field. It specifies use of the NLPID for IP protocol and SNAP encapsulation for other protocols. Multiservice Switch systems support only IP protocol.

Inverse ARP for IP protocol is supported as specified in RFC1293.

Frame relay overview



Data link connection identifiers (DLCIs)

The FrDte software is capable of automatically setting up Data Link Connection Identifier (DLCI) dynamic subcomponents for dynamically learned circuits. This feature is enabled or disabled through the *acceptUndefinedDlci* attribute located under the *FrDte* component. When enabled, a *DynamicDlci* subcomponent will be created when the frame relay network notifies the FrDte of a new permanent virtual circuit (PVC) through the LMI protocol and a corresponding *StaticDlci* subcomponent does not exist. A *DynamicDlci* subcomponent will also be created when a frame is received on the FrDte interface over a PVC which does not yet exist.

DynamicDlci subcomponents are always linked to the *RemoteGroup/1* instance (a mandatory component) and inherit the attributes provisioned under the *DynamicDlciDefaults* (*DynDlciDefs*) subcomponent. A *DynamicDlci* subcomponent can be removed using the Clear verb or by replacing it with a *StaticDlci* subcomponent. Although a *DynamicDlci* can have a Committed

Information Rate (CIR) enforced on egress frames sent out on it, it cannot have a *HibernationQueue* (Hq) subcomponent to buffer frames in violation of the rate enforcement policy. Only *StaticDlci* components, on a Multiservice Switch 7400 with SBIC-based FPs, can have *Hq* subcomponents.

You can provision dynamic DLCIs using commands if you are in operational mode. Static DLCIs are provisioned using commands in provisioning mode. See Operational mode (page 12) and Provisioning mode (page 13). For more information, see the following:

- Local management interface (LMI) (page 58)
- Remote groups (page 59)

Local management interface (LMI)

LMI is used between a frame relay end station and the local node that is directly attached. (A Nortel Networks Multiservice Switch node acts as a frame relay DTE end station or as the local node.) It allows each end of the frame relay UNI service to verify that the other end is operational, and also allows the end station to learn from the local node which PVCs are active. LMI is provisioned through the *LinkManagementInterface* (*Lmi*) component. Several standards apply:

- Vendor Forum LMI support:
  - Frame Relay Specification with Extensions, Doc. No. 001-208966.
  - “Section 4: Physical Interfaces” is dependent on the attached link.
  - “Section 5: Data Link Interface” in entirety. The default value of dN1=1604.
  - “Section 6: LMI” in entirety with the frame relay interface defined as “DTE”.
  - “Section 7: Optional Extensions” supports the PVC status of the update status message. The D, R, and PVC bandwidth fields are ignored.
- ITU-T, Annex A is supported, with the following exceptions:
  - “Section A.3.3: PVC Status” is supported except that only two-byte DLCIs are recognized.
  - “Section A.6: Optional Network Procedures” is not supported.
  - “Section A.7: System Parameters” is supported as follows: full compliance with default parameters (timer T392 pertains to the network and is not applicable to frame relay DTE); full Status Polling Counter N391 is set to 6 polling cycles; Error Threshold N392 is set to 3 errors; Monitored Events Count N392 is set to 4 events; Link Integrity Verification Polling Timer T391 is set to 10 seconds.
- ANSI T1.617, Annex D is supported, with the following exceptions:

— "Section D.3.3: PVC Status" is supported except that only two-byte DLCIs are recognized.

— "Section D.6: Optional Network Procedures" is not supported.

— "Section D.7: System Parameters" is supported as follows: full compliance with default parameters (timer T392 pertains to the network and is not applicable to frame relay DTE); full Status Polling Counter N391 is set to 6 polling cycles: Error Threshold N392 is set to 3 errors; Monitored Events Count N392 is set to 4 events; link Integrity Verification Polling Timer T391 is set to 10 seconds.

Remote groups

Nortel Networks Multiservice Switch IP over frame relay supports remote groups as follows:

- Multicast frames are transmitted across each PVC in the associated remote group.
- Each frame relay DTE remote group is modeled as a fully connected mesh network by IP. If a network is not fully connected, it can be divided into smaller subnetworks until it is fully connected under each remote group.

FrDte to FrUni connectivity

The frame relay user-to-network interface (FrUni) is the standard interface between the user device and the network. The FrDte is the frame relay interface into an IP network. There are three methods of implementing FrDte to FrUni connectivity: physical (hairpin), logical, and direct. The table FrDte to FrUni connectivity on Multiservice Switch FPs (page 59) summarizes which FPs support each method.

FrDte to FrUni connectivity on Multiservice Switch FPs

	Multiservice Switch 7400 SBIC	Multiservice Switch 7400 MSA32	Multiservice Switch 15000 and Multiservice Switch 20000
Physical (hairpin)	supported	not supported	not supported
Logical	supported	supported	supported
Direct	not supported	supported	supported
Attention: Where supported, a direct connection is the recommended method.			

See the following sections for more information:

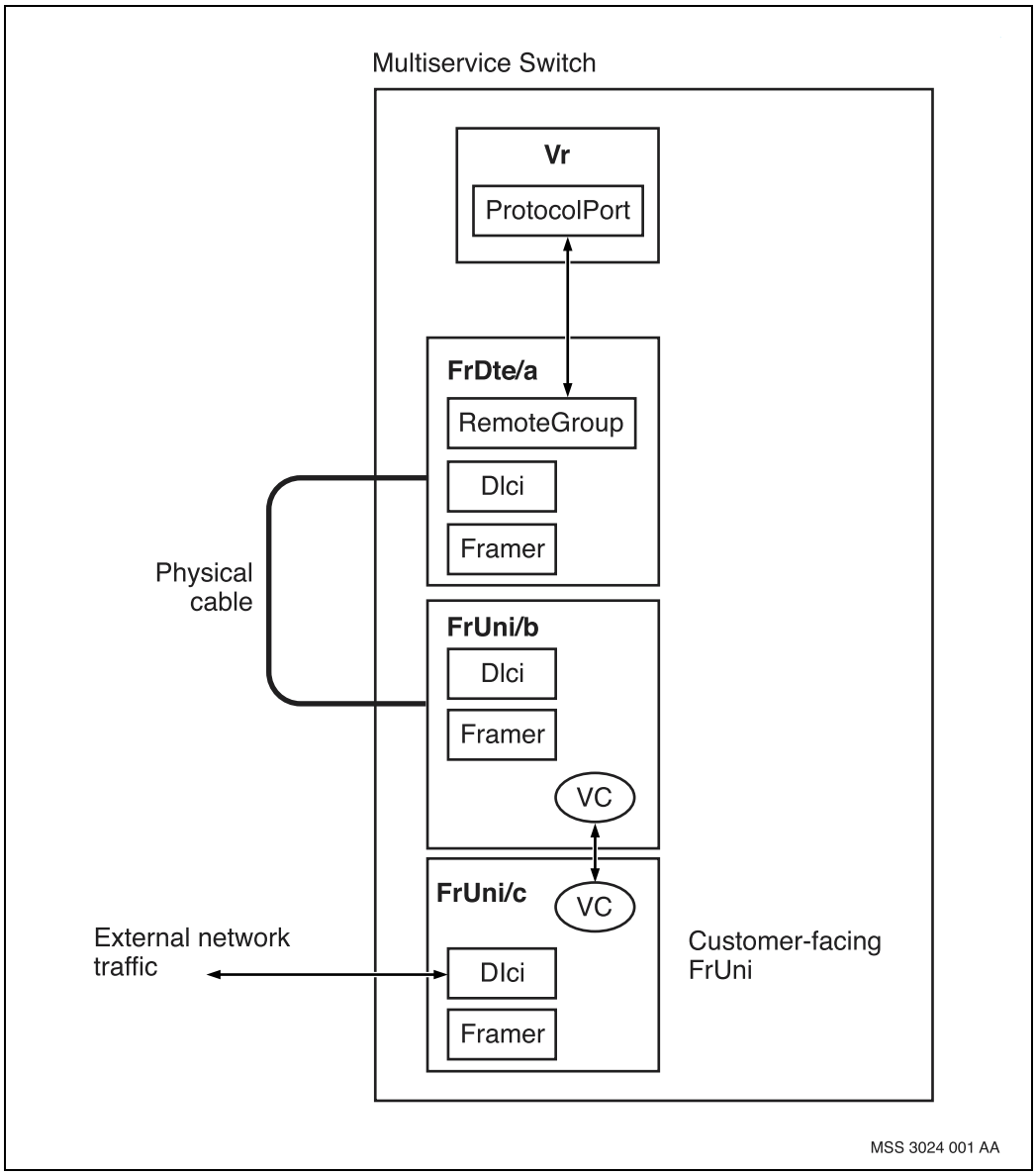
- Physical (hairpin) connection (page 60)
- Logical connection (page 61)

- Direct connection (page 62)

Physical (hairpin) connection

Use this configuration if you are using a hardware connection to link the FrDte and FrUni. In this configuration, each interface must be linked to a physical port through their respective *Framer* components. In addition, the two ports must be physically linked through a cable. Configure a PVC to the customer-facing FrUni on the Nortel Networks Multiservice Switch node.

FrDte to FrUni connection with a physical link



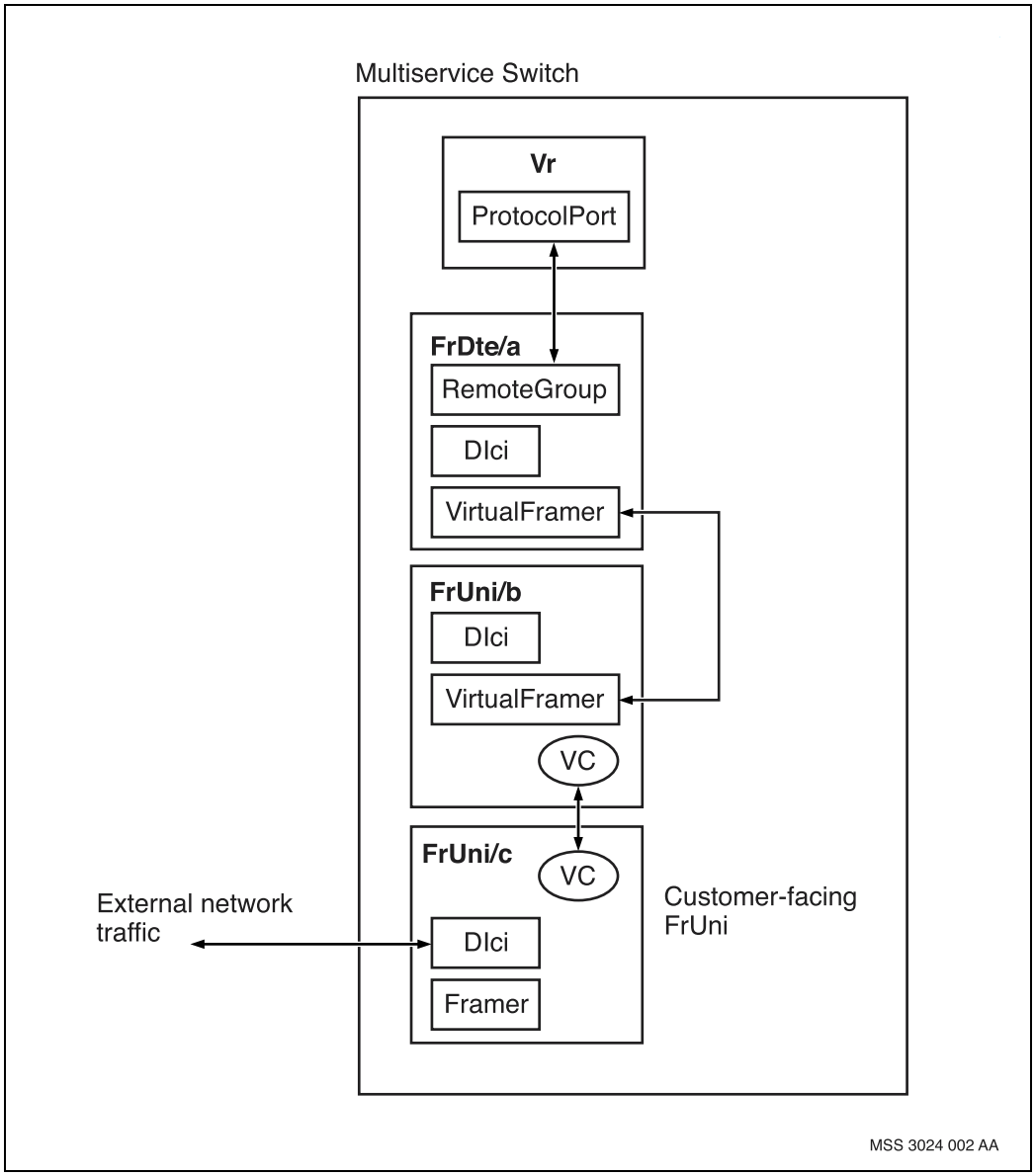
Logical connection

You can conserve physical ports by using an internal software connection to link the FrDte and FrUni interfaces through their respective *VirtualFramer* (*VFramer*) components. Configure a PVC to the customer-facing FrUni on the Nortel Networks Multiservice Switch node.

If you use a logical connection, you need to delete the Framer components, which are automatically created on installation, and add VFramer components in their place. The components linked by the VFramer components must reside on the same card.

CoS to QoS mappings over a single DLCI are not supported on the logically connected FrDte and FrUni ports.

FrDte to FrUni connection with a logical link



Direct connection

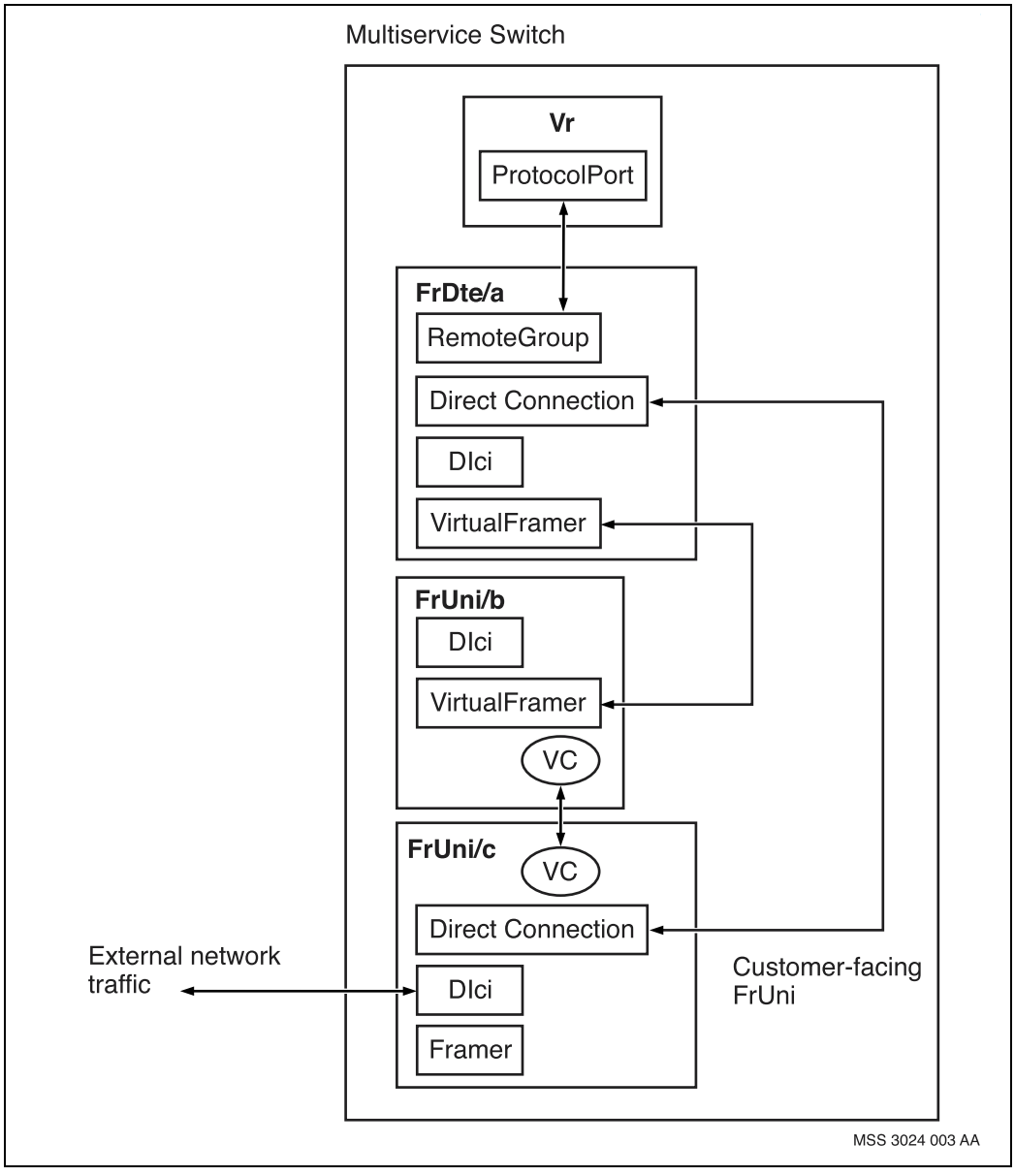
This configuration enables you to create an alternative data path (direct connection) between the FrDte and the customer-facing FrUni using a single DLCI. Using a direct connection causes a significant improvement in performance.

The direct connection uses the same frame relay and IP configuration as a logical connection, except that you need to add DirectConnection (Dconn) components to the FrDte and customer-facing FrUni and link them.

Attention: When you use a direct connection, the FrDte and FrUni must reside on the same card.

Attention: When you use a direct connection, you must use a single DLCI for CoS to QoS mapping. See NN10600-591 *Nortel Networks Multiservice Switch 7400/15000/20000 Layer 3 Traffic Management Configuration* for more information.

FrDte to FrUni connection with a direct connection link



Congestion control

The primary objective of congestion control is to maintain the quality of service by minimizing frame discards, warning end users of congestion, and minimizing the possibility of one end user consuming network resources at the expense of other users. It is recommended that all traffic management, including congestion control, be performed at the entry point of the network, on the port of the customer-facing FrUni. See the figure FrDte to FrUni connection with a physical link (page 60) for an example of a customer-facing FrUni.

The following are the congestion control mechanisms found in this software:

- The Committed Information Rate (CIR) is the rate that the network agrees to transfer information over a a virtual circuit under typical conditions. CIR enforcement prevents one user from utilizing an inequitable share of the network resource. In FrDte, CIR enforcement is a provisionable feature on a per DLCI basis. *committedburst* (*bc*), *committedInformationRate* (*cir*), Excess Information Rate (EIR), and Time Measurement Interval (Tc) are used to define this feature.

  In this feature, CIR enforcement is applied on the egress data path (that is, frames sent out the interface to the WAN cloud) on a per DLCI basis as provisioned by the user. The ingress data path (that is, frames received on the interface from the WAN cloud) is not monitored by the FrDte for rate enforcement.

- The Explicit Congestion Notification (ECN) bits are two bits in the Q.922 address part in the frame header that may be set by the frame relay network to notify the user that frames are encountering congestion. The objective of ECN is to reduce demands on resources so the network can return to normal operation mode. There are two types of congestion notification: Forward Explicit Congestion Notification (FECN) and Backward Explicit Congestion Notification (BECN). The user may also set the FECN or BECN bit.

  Unlike DE bit, FECN and BECN bits are not set by FrDte during processing of the frame. All arriving frames at FrDte, the FECN and BECN bits are noticed and counted in appropriate statistics. However, no action is taken to throttle traffic on the basis of these indications.

- The Discard Eligibility (DE) bit in Q.922 address part of the frame header is also used for congestion management. The DE bit determines whether a frame should be discarded in preference to other frames. It may be set by either the user or the service. The discard of discard eligible frames during congestion ensures that priority is given to frames within the CIR limit.

- *LinkEmissionQueue* (*Leq*), which is an optional component under *FrDte*, provides services that help to control congestion by putting limits on the

size of the queue, the maximum number of multicast packets allowed to transmit, and a time to live limit for each packet in the queue. One use of this feature is rate shaping. The *Leq* is supported only on Nortel Networks Multiservice Switch 7400 nodes with SBIC-based FPs.

Committed information rate (CIR)

The CIR is the rate (in bit/s) that the network agrees to transfer information over a PVC under typical conditions. On the FrDte interface, CIR enforcement is applied on the egress data path (that is, frames sent out the interface to the frame relay cloud) on a per DLCI basis as provisioned by the user. The ingress data path (that is, frames received on the interface from the frame relay cloud) is not monitored by the FrDte for rate enforcement.

There are five attributes (provisionable on each *Dlci* subcomponent) that control the rate shaping performed on each PVC:

- *rateEnforcement* (*re*) - this attribute enables or disables rate enforcement on the PVC - if it is disabled, the attributes *committedInformationRate* (*cir*), *committedBurst* (*Bc*), *excessBurst* (*be*), and *excessBurstAction* (*beAction*) cannot be used to define rate enforcement.
- *committedInformationRate* (*cir*) - this attribute is the CIR (in bit/s) to be enforced on egress frames sent over the PVC.
- *committedBurst* (*bc*) - this attribute is the committed burst size (in bits) for the PVC. This value represents the maximum amount of data that can be sent as a burst of traffic over the PVC at a rate above the CIR.
- *excessBurst* (*be*) - this attribute is the excess burst size (in bits) for the PVC. This value represents an additional amount of data which may be sent in excess of the BC amount at a rate above the CIR. Traffic operating in this range may have the Discard Eligible flag set in the frame relay header if provisioned to do so by the excessBurstAction attribute.
- *excessBurstAction* (*beAction*) - this attribute controls whether the frame relay header Discard Eligible flag is set in frames which are exceeding the committed burst level (that is, traffic operating in the excess burst range). Frames marked as Discard Eligible are discarded at the discretion of the network.

The Leaky Bucket algorithm implements the rate enforcement mechanism. This algorithm enforces conformance to a single rate.

Traffic that violates the rate shaping policy is typically discarded. However on a Nortel Networks Multiservice Switch 7400 node with SBIC-based FPs, a *HibernationQueue* (*Hq*) subcomponent can be provisioned under a *StaticDlci* component as a temporary holding area for violating frames. While at least one frame remains on the *Hq*, all frames being forwarded out that PVC are

automatically placed on the *Hq*. This is done to preserve ordering. Frames are then forwarded from the *Hq* in the order they were placed on the queue (higher priority frames first) as the rate shaping policy allows.

The *Hq* provides the same services as the *LinkEmissionQueue* (*Leq*) with the following noted exceptions:

- an *Hq* holds frames destined to be sent out a particular PVC. Therefore, the Indexed and Balanced service classes are not applicable
- the Hardware Forced flag is ignored

All data frames which make it through the CIR enforcement are placed on the high priority link queue for transmission. Frames that do not undergo CIR enforcement (that is, *rateEnforcement* is disabled for the PVC) are placed on the normal priority link queue for transmission and are also subject to be placed on the Link Emission Queue if one is provisioned.

IP over frame relay using IP-optimized DLCIs

This section describes Nortel Networks Multiservice Switch implementation of IP over frame relay using IP-optimized DLCIs. THis is an alternative to IP over frame relay using frame relay DTE (page 56) as an access media. This section includes the following topics:

Navigation

Overview of IP-optimized DLCIs

An IP-optimized data link connection identifier (DLCI) can directly bind to a virtual router protocol port. This type of DLCI is linked to the Nortel Networks Multiservice Switch frame relay user-to-network interface (FRUNI), which eliminates the need for a frame relay DTE as described in IP over frame relay using frame relay DTE (page 56) and simplifies provisioning.

IP-optimized DLCIs can be combined with existing DLCI types (PVC, SVC, or SPVC across DPRS, or BnxIwf DLCI) so that a single FRUNI can have several DLCIs that are PVCs across DPRS and several IP-optimized DLCIs that are linked to a protocol port. The ability to have several types of DLCIs on a single interface facilitates the migration to IP-optimized DLCIs.

Due to its flexibility and performance, using IP-optimized DLCIs is the preferred method for frame relay access to an IP VPN. For IP VPN information, see NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals* and NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management*.

Frame relay congestion notification

Existing frame relay congestion notification is used to signal local congestion to the CE device. In the egress direction, the CE is notified with BECN when the receive queue is congested, and with FECN when the transmit queue is congested.

In the ingress direction, FECN and BECN are counted by the DLCI but do not trigger any special behavior. When local congestion occurs, frames tagged with discard eligibility (DE) bits are discarded before frames whose DE bits are not set.

For more information on FECN and BECN, see NN10600-900 *Nortel Networks Multiservice Switch 7400/15000/20000 Frame Relay Technology Fundamentals*.

LMI and A-bit status

Nortel Networks Multiservice Switch frame relay local management interface (LMI) can run network side procedure (NSP), user side procedure (USP), or both procedures at the same time. All existing protocols (Vendor Forum, ITU, and ANSI) are supported and can be used with IP-optimized DLCIs with no restrictions. For more information, see:

- Network side procedure (page 68)
- User side procedure (page 69)

Network side procedure

When the LMI is running the network side procedure, it must periodically report the state of its DLCI. Attribute *FrUni Dlci aBitReasonTolf* is set as shown in the table A-bit status and reason signaled to the CE device (page 68).

An IP-optimized DLCI is reported active when it is unlocked and the protocol port to which it is attached is active.

A-bit status and reason signaled to the CE device

aBitReasonTolf	Cause
notApplicable	Used when aBitStatusTolf is active.
localLmiError	Used when the local FRUNI is down.
localLinkDown	Used when the local link is down.
pvcSpvcDown	Used when the DLCI is locked or its protocol port is disabled.

(1 of 2)

A-bit status and reason signaled to the CE device (continued)

aBitReasonToIf	Cause
remoteUserSignaled remoteLinkDown remoteLmiError userNotAuthorized resourceNotAvailable dlciCollisionAtNni	The remaining values of *aBitReasonToIf* are not used by IP-optimized DLCI.

(2 of 2)

User side procedure

When IP-optimized DLCI is deployed at the edge of the network, the LMI should be set to perform the network side procedure. However you can set the LMI to run the user side procedure by setting attribute *FrUni Lmi side* to user or both.

With the user side procedure, the LMI periodically polls the local CE for the status of each DLCI. The status reported by the LMI (network side procedure) is combined with the local status (user side procedure) to create the PVC state reported to the virtual router.

Attribute *FrUni Dlci aBitReasonFromIf* is set as shown in the table A-bit status and reason signaled from the CE device (page 69).

A-bit status and reason signaled from the CE device

aBitReasonFromIf	Cause
notApplicable	Used when the status of the DLCI is active. An active status is reported to the virtual router.
remoteUserSignaled	Used when the remote user is down. An inactive status is reported to the virtual router.
localLmiError	Used when the LMI is down. An inactive status is reported to the virtual router.

(1 of 2)

A-bit status and reason signaled from the CE device (continued)

aBitReasonFromIf	Cause
localLinkDown	Used when the link is down. An inactive status is reported to the virtual router.
missingFromLmiReport	Used when the LMI is up, but the adjacent LMI was not reported by the adjacent LMI. An inactive status is reported to the virtual router.

(2 of 2)

IP over Ethernet

This section describes the implementation of IP over Ethernet on Nortel Networks Multiservice Switch nodes.

Navigation

Overview of Multiservice Switch IP over Ethernet

Ethernet is an access media to an IP forwarding and routing service that enables the use of Ethernet media on a VR protocol port, in combination with legacy media.

IP over Ethernet supports IP encapsulation over Ethernet. There are three types of Ethernet available:

- 10BaseT is available on Multiservice Switch 7400 nodes at a rate of up to 10 megabits per second per Ethernet port for IP routing and forwarding.
- 100BaseT is available on Multiservice Switch 7400 nodes at a rate of up to 100 megabits per second per Ethernet port for IP routing and forwarding.
- Gigabit Ethernet is available on Multiservice Switch 15000 and Multiservice Switch 20000 nodes at a rate of up to 1 gigabit per second per Ethernet port for IP routing and forwarding.

The following are supported for all types of IP over Ethernet on Multiservice Switch nodes:

- full compliance with RFC894
- TCP and UDP transport layer protocols
- ICMP and ARP control protocols
- static routes

- RIPv2, OSPF, and BGP-4 routing protocols
- IP class of service (CoS)
- IP differentiated services

IP datapath interworking

IP datapath interworking between Ethernet ports and the following ports is supported over the backplane:

- ATM MPE media ports on CQC-based, PQC-based, and GQM-based ATM FPs depending on Ethernet type. See the table Supported ATM FP types for IP datapath interworking (page 72) for details.
- Voice Services Processors 2 and 3 (VSP2 and VSP3) ports for Media Gateway VoIP applications. See NN10600-780 *Nortel Networks Media Gateway 7480/15000 Technology Fundamentals* for more information.
- CP3 OAM Ethernet 100BaseT ports for network management traffic on Nortel Networks Multiservice Switch 15000 and Multiservice Switch 20000 nodes, and CP2 on Multiservice Switch 7400 nodes.

For more information on FPs, see NN10600-551 *Nortel Networks Multiservice Switch 7400/15000/20000 FP Configuration Reference*.

For more information on CPs, see NN10600-120 *Nortel Networks Multiservice Switch 15000/20000 Hardware Description*.

Supported ATM FP types for IP datapath interworking

Ethernet type	ATM FP type	PEC code
10/100BaseT	8-port DS1 (see note)	NTNQ49
	8-port E1 (see note)	NTNQ50
	2-port OC-3/STM-1 ATM IP	NTNQ65, NTNQ66
	32-port DS1 MSA	NTNQ71
	32-port E1 MSA	NTNQ73

(1 of 2)

Supported ATM FP types for IP datapath interworking (continued)

Ethernet type	ATM FP type	PEC code
Gigabit Ethernet	12-port E3 ATM	NTHR25DA
	16-port OC3/STM-1 ATM	NTHW21, NTHW31
	1-port OC-12/STM-4 ATM	NTHR29DA
	4-port OC-12/STM-4 ATM (PQC12)	NTHW86xx
	4-port OC-3 ATM (multimode) (PQC2)	NTHR17DA
	4-port OC-3 ATM (multimode) (PQC1)	NTHR17CA
	4-port OC-3 ATM (multimode) (PQC12)	NTHW05AA
Attention: The 4-port 10/100BaseT Ethernet FP and 8-port 10/100BaseT Ethernet FP are not compatible with the 8-port DS1 and 8-port E1 FPs.		
(2 of 2)		

IP packet sizes

The table IP packet sizes (page 73) shows the minimum and maximum supported sizes for gigabit Ethernet packets. Maximum transmission unit (MTU) refers to the maximum IP packet size carried by the Ethernet frames.

The minimum packet size that is transported by the Ethernet link is 64 bytes. If a packet that is less than 64 bytes is received over the plane from another FP, the FP pads the packet to 64 bytes with little or no impact on performance.

IP packet sizes

Packet type	Minimum (octets)	Maximum (octets)	MTU (octets)
Ethernet V2.0	64	1600	1500
Ethernet 802.3 LLC-SNAP	64	1600	1492

CP switchover

For the VIPR solution, CP switch-over (CPSO) for both static and OSPF learned routes causes minimal impact for Gigabit Ethernet media-attached service. Routes remain active and datapath forwarding experiences a minimal outage. ARP entries are refreshed but not cleared.

Over a cold CP switchover, administrative states of all relevant components are reset to their default upon CPSO. Over a hot CP switchover, administrative states of all relevant components are maintained at their pre-switchover settings. In both cases, however, operational states will reflect the current operational status of the component.

IP over point-to-point protocol (PPP)

This section describes the implementation of Nortel Networks Multiservice Switch IP over point-to-point protocol (PPP).

Navigation

Overview of IP over PPP

Point-to-point protocol (PPP) is a link layer protocol (LLP). It provides a simple, reliable method of transporting IP datagrams by encapsulating them into a high-level data link (HDLC) frame.

PPP is designed for simple links that transport packets between two peers. These links provide full-duplex (simultaneous and bi-directional) operation and are assumed to deliver packets in order. PPP can provide a common solution for easy connection of a wide variety of hosts, bridges, and routers.

Although PPP operates at the link layer, which is layer 2 of the OSI model, it performs several functions for the network layer, which is layer 3. Its unit of information consists of 8-bit bytes with no parity. The WAN link required for PPP does not require modem status signalling. Any type of leased line including copper, fiber optic, microwave or satellite can be used to transmit PPP messages.

PPP consists of three main functional components:

- a method for encapsulating multi-protocol datagrams
- a link control protocol (LCP) for establishing, configuring, and testing the data-link connection

- a family of network control protocols (NCPs) for establishing and configuring different network protocols

IP over PPP Multiservice Switch implementation

IP over PPP is supported on Nortel Networks Multiservice Switch 7400 (both SBIC-based and MSA32 FPs), Multiservice Switch 15000, and Multiservice Switch 20000 nodes. For more information, see NN10600-551 *Nortel Networks Multiservice Switch 7400/15000/20000 FP Configuration Reference*.

Attention: On MSA32 FP, Multiservice Switch 15000 node and Multiservice Switch 20000 FPs, IP over PPP is supported on PQC2.0 or later FPs.

The following features are supported in the Multiservice Switch system implementation of IP over PPP:

- On all supported FPs, PPP's maximum receive unit (MRU) is 18,000 octets.
- Cold standby equipment protection is provided on the 4-port DS3CH FR FP.
- A maximum of 200 PPP services are supported per Multiservice Switch 7400 node MSA32 FP, and per Multiservice Switch 15000 and Multiservice Switch 20000 FPs.
- Weighted fair queuing is supported on Multiservice Switch 7400 MSA32FPs, and on Multiservice Switch 15000 and Multiservice Switch 20000 FPs.

Link transmission and monitoring features

On SBIC-based FPs only, the *Ppp Leq* component provides a link emission queuing (LEQ) feature, which allows the service to assign transmit priorities to ensure that selected high priority packets are transmitted before lower priority packets. In addition, LEQ controls the maximum number of packets and bytes that can be present in the queue at any time, and the amount of time that each packet can remain in the queue. These parameters can be tailored to a particular PPP implementation through provisioning. The LEQ is typically used for low data rate connections with delay sensitive traffic.

Using the *Ppp Link continuityMonitor* attribute, the link continuity monitoring (LCM) feature can continually confirm the link connection with the peer PPP application. LCM uses LCP echo packets to continually communicate with the peer. If more than five echo packets in succession are not received from the peer, the link is considered unusable, marked disabled, and attempts renegotiation of LCP.

The *Ppp Link configMagicNumber* attribute provides a method for detecting looped back links.

For Nortel Networks Multiservice Switch 15000 and Multiservice Switch 20000, and Multiservice Switch 7400 MSA32 FPs, direct hardware forwarding allows, for example, an IP packet to be routed directly to the link queue of the card where PPP resides without software intervention on the FP. This dramatically improves full duplex packet switching performance over PPP. In order to use direct hardware forwarding, the link quality monitor (LQM) must be disabled. Do this with the *Ppp Lqm configStatus* attribute. On Multiservice Switch 7400 MSA32 FPs, and on Multiservice Switch 15000 and Multiservice Switch 20000 FPs, these statistics are not enabled and should not be used for monitoring the link.

Weighted fair queuing (WFQ) provides a mechanism to queue traffic into separate traffic flows according to traffic class definitions, and ensures that low priority traffic has some opportunity to transmit. For Multiservice Switch 15000 and Multiservice Switch 20000 (4pDS3Ch and 1pSTM1Ch), and Multiservice Switch 7400 MSA32 FPs, the Ppp Ewfq component provides configurable weighted fair queuing using a priority guaranteed WFQ algorithm based on a frame rate for IP traffic over PPP. This component provides the ability to reconfigure the WFQ table to redistribute the transmission opportunities among the four egress traffic queues. For each of the four egress queues the transmission opportunity values range from 0% to 100%. The default values for the four queues are 91%, 3%, 3%. and 3% for queues 0 to 3 respectively.

For more information on provisioning the features in this section, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

PPP Framer statistics

There are twelve statistics available under the *PPP Framer* component and on SBIC-based FPs, these statistics are all valid. But on Nortel Networks Multiservice Switch 7400 MSA32, and Multiservice Switch 15000 and Multiservice Switch 20000 FPs, these statistics should never be used as they are not valid for FPs that are not SBIC-based.

Two of these statistics, *frmToIf* and *frmFromIf*, have equivalent counts available under the *Vr IfTableEntry* component. For the remaining ten statistics, there are no equivalent statistics on Multiservice Switch 7400 MSA32, and Multiservice Switch 15000 and Multiservice Switch 20000 FPs. See the table PPP Framer statistics (page 76) for details.

PPP Framer statistics

PPP-related statistics available on SBIC-based FPs under the *PPP Framer* component	Equivalent statistics on Multiservice Switch 7400 MSA32, and Multiservice Switch 15000 and Multiservice Switch 20000 FPs
frmToIf	Vr IfTableEntry ifInUcastPkts
frmFromIf	Vr IfTableEntry ifOutUcastPkts
aborts	None
crcErrors	None
lrcErrors	None
nonOctetErrors	None
overruns	None
underruns	None
largeFrmErrors	None
frmModeErrors	None
outOfSequenceFrm	None
repeatedFrm	None

PPP outages

On all supported FPs, PPP experiences a brief outage during some provisioning changes. Specifically, for any protocol port that is linked to a *Ppp* component, a PPP outage occurs if any component or attribute under the *Vr Pp IpPort* component is changed. The outage is a result of PPP renegotiating the Network Control Protocol (NCP) layer. Effectively, any provisioning change to a PPP-linked *IpPort* component is treated as a critical change.

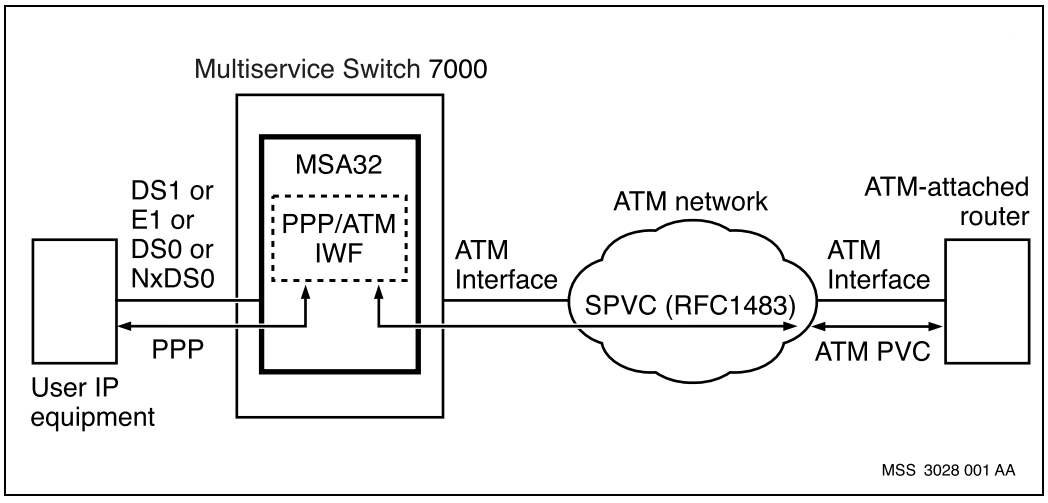
The NCP layer goes down for a short period, typically less than one second under non-stress conditions, and then automatically re-establishes itself. A potential effect is that the PPP link's routing protocol, for example, OSPF, may experience a brief outage.

Point-to-point protocol (PPP)/ATM interworking for Multiservice Switch 7400 nodes

The PPP/ATM interworking feature on MSA32 function processors, enables IP transport between PPP-attached user devices and ATM attached routers as shown in figure PPP/ATM interworking on MSA32 (page 78).

In the ingress direction, the IP packets received on DS1 or E1 ports (or DS0 or NxDS0 channels) are extracted from the PPP protocol, re-encapsulated into ATM (RFC1483), and forwarded onto ATM SVPCs toward ATM-attached routers. In the egress direction, ATM-encapsulated IP packets are extracted, encapsulated into PPP, and forwarded on DS1 or E1 ports (or DS0 or NxDS0 channels). Routing or switching does not take place, and there is a 1-to-1 relationship between the port or channel carrying the PPP protocol and a corresponding ATM SPVC.

PPP/ATM interworking on MSA32



Navigation

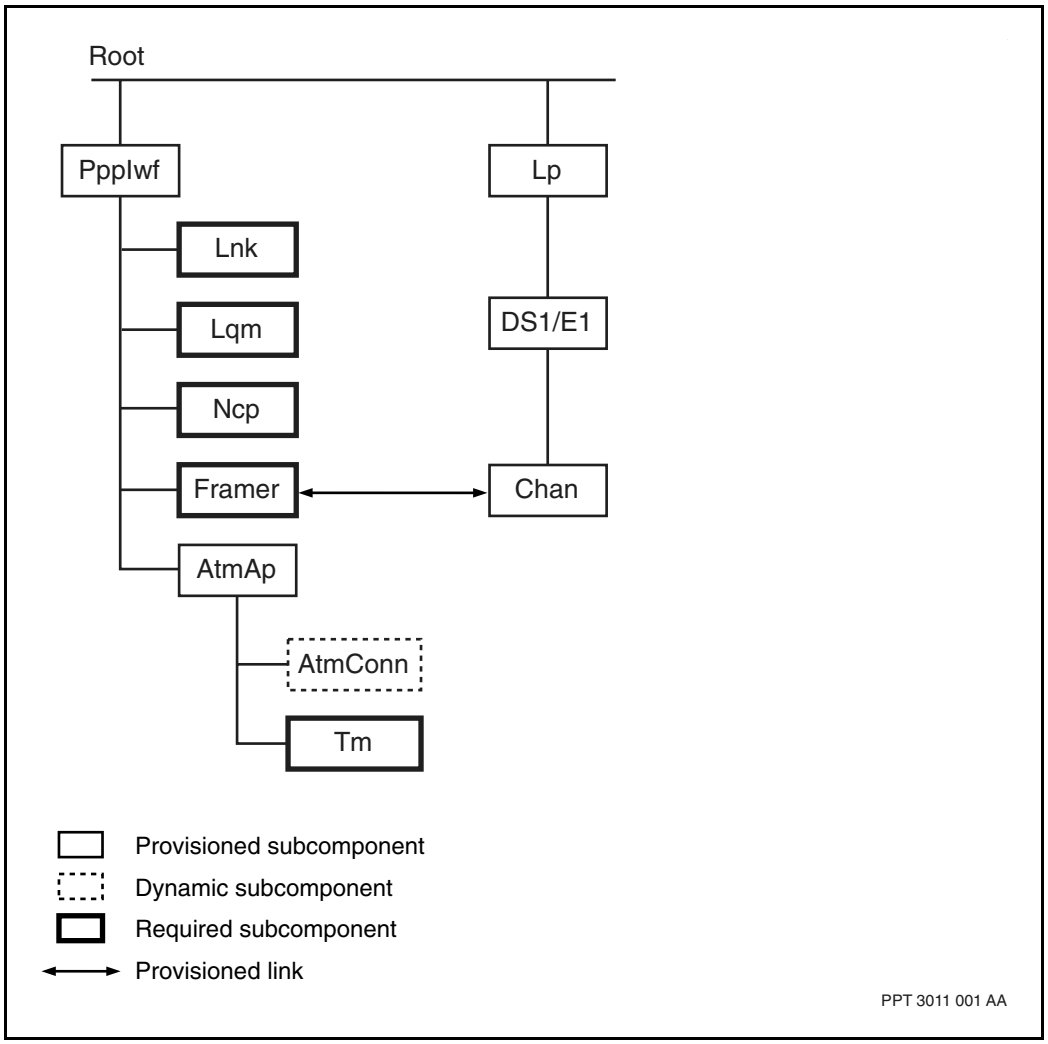
- Software architecture of PPP/ATM interworking (page 79)

- Components and attributes of PPP/ATM interworking (page 80)

Software architecture of PPP/ATM interworking

The figure Component hierarchy for PPP/ATM interworking (page 79)shows the component hierarchy for the PPP/ATM interworking function.

Component hierarchy for PPP/ATM interworking



The components supporting the PPP/ATM interworking function are as follows:

- *PppIwf* defines an instance of the PPP interworking function which provides the means for PPP to interwork with different forms of layer two protocols. For example, AtmMpe. The component instance number is a unique number.
- *Lnk* contains all the attributes related to the Ppp link.

- *Lqm* contains all the operational attributes for the Ppp link quality monitor.
- *Ncp* contains all the operational attributes of the Ppp network control protocols (NCP).
- *Fr*amer controls link layer framing for application components that send and receive data on a link interface. It is also through this component that an application component is associated with a specific hardware link interface.
- *AtmAp* represents the Soft PVC between *PppIwf* and *AtmIf*.
- *AtmConn* displays where the data traffic for this connection is directed.
- *Tm* contains ATM QoS information used by the SPVC connection to request a particular QoS from ATM.

Components and attributes of PPP/ATM interworking

This section describes the components and attributes of the PPP/ATM interworking function:

- PppIwf/n component (page 80)
- PppIwf/n AtmAdaptationPoint component (page 80)
- PppIwf/n AtmAp TrafficManagement component (page 81)

PppIwf/n component

This component defines an instance of the PPP interworking function.

Configurable attributes for PppIwf/n

Attributes	Description
customerIdentifier (cid)	Identification of the customer who owns the CES.
ifAdminStatus	The desired state of the interface.
ifIndex	This is the index for the IfEntry.

PppIwf/n AtmAdaptationPoint component

This component is used to set up Soft PVCs to transmit and receive encapsulated IP data from the ATM interface. It also performs encapsulation of the IP data on the PPP port.

Configurable attributes for AtmAdaptationPoint

Attributes	Description
maxTransmissionUnit	The size of the largest datagram that can be sent on the Soft PVC.
encapType	The RFC1483 encapsulation type to be used.
localAddress	The local NSAP address.
calledVpiVci	The identity of the PVC at the remote ATM node on which the soft PVC connection terminates.
addressToCall	The remote NSAP address called a PNNI address.
firstRetryInterval	The time to wait in seconds, before attempting to establish the connection after the first failed attempt.
retryLimit	The maximum number of consecutive unsuccessful connection setup attempts that may be made before further attempts are abandoned.

Ppplwf/n AtmAp TrafficManagement component

This component is used to request a particular QoS from ATM.

Configurable attributes for Ppplwf/n AtmAp TrafficManagement

Attributes	Description
atmServiceCategory	The desired ATM service category for an SPVC connection.
peakCellRate	The desired peak cell rate for an SPVC connection.

IP routing management

To determine the optimum route to a destination, routers in a network must exchange route information. This function is accomplished by dynamic routing protocols and static routes. After you have configured IP and virtual routers in Nortel Networks Multiservice Switch system, you must configure static routes or one or more dynamic routing protocols to enable the exchange of route information. The exchange of route information between two different routing protocols is called route redistribution.

This section is intended to help you plan the redistribution of route information among dynamic routing protocols operating on Multiservice Switch VRs.

Navigation

- Overview of IP routing management (page 82)
- Route preferences (page 86)
- Example routing topologies (page 89)
- IP differentiated services for routing packets (page 91)

For background information about IP routing, see the following sections:

- Static routes (page 123)
- Routing information protocol (RIP) (page 93)
- Open shortest path first (OSPF) protocol (page 100)
- Border gateway protocol 4 (BGP-4) (page 107)

Overview of IP routing management

You can configure multiple dynamic routing protocols on a single VR to connect networks that use different dynamic routing protocols. For example, you can run RIP on one subnetted network, and OSPF on another subnetted network, and exchange routing information between them in a controlled fashion. Import and export policies control the exchange of information between protocols and between routers.

This section covers the following topics:

Nortel Networks Multiservice Switch VRs support simultaneous operation of one instance of

- routing information protocol (RIP). See Routing information protocol (RIP) (page 93).
- open shortest path first (OSPF) protocol. See Open shortest path first (OSPF) protocol (page 100).
- border gateway protocol 4 (BGP-4). See Border gateway protocol 4 (BGP-4) (page 107).

Each of these dynamic routing protocols collects different types of information and reacts to topology changes in its own way. For example, RIP uses a hop-count metric to compute the shortest paths, while OSPF uses an interface-cost metric. RIP and OSPF are interior routing protocols used by routers to exchange information within a single autonomous system. BGP-4 is an exterior routing protocol (EGP) used by routers to exchange information among different autonomous systems, although it can also function as an interior routing protocol. BGP-4 uses an AS_PATH length metric to compute the shortest paths. Multiservice Switch nodes also use static routes and local routes from a directly-connected interface to exchange information between virtual routers.

In the case where routing information is being exchanged between different networks that use different dynamic routing protocols, there are many configuration options that enable the exchange of routing information. For more information on some of these configuration options, see Example routing topologies (page 89).

Routing policies

Routing policy is a set of rules used to control the exchange of information between protocols and between routers. A protocol uses routing policy to determine which routes it learns from a peer to install in the routing database (RDB) and which information from the RDB to announce to other routers in the network.

All of the routing protocols have a default policy. In addition, each of the routing protocols can be configured with many different import and export policies on top of its default policy. If no other policies are configured, the routing protocol uses the actions contained in the default policy to filter route information.

A routing protocol has two sets of routing policies:

- import policy determines if a route learned from a peer can be installed in the RDB (used) or cannot be installed in the RDB (ignored)
- export policy determines if a route in the RDB can be sent to a peer (send) or cannot be sent to a peer (blocked)

Routing protocols on Nortel Networks Multiservice Switch nodes use import and export polices in the following ways:

- RIP (versions 1 and 2) uses both import and export policies
- BGP-4 uses both import and export policies. By default, BGP-4 on Multiservice Switch nodes accept all routes from any internal peer and rejects all routes from any external peer. BGP-4 also blocks any routes to its exterior BGP (EBGP) peers and advertises all routes learned from EBGP to its interior BGP (IBGP) peers.
- OSPF uses only export policy. It accepts all advertisements and therefore does not need to filter routes with import policy. However, it uses export policy to determine which non-OSPF routes can be sent to its neighbors.

Import and export statements use attributes (also called selection keys) to define information found in the protocol update packets or RDB. This information helps to discriminate between packets or table entries. Two examples of selection keys are the IP address and the protocol. See Route preferences (page 86) for more information on selection keys.

You can use multiple policy statements when configuring a router. The IP forwarding table uses a best match policy, where policy decisions are made from the longest (that is, the most specific) match between the policies and the routes stored in the RDB. To determine which match is most specific between policies, the selection keys available to each protocol's import or export statements get a rating integer to order their importance. If two policies with different selection keys both match in selection criteria, then the rating integers of the selection keys that are set on each policy are added up. The one with the highest sum of rating integers is the best match policy. The selection keys available and their respective rating integer are listed with the import and export statements of each protocol.

Flow of routing information

Each dynamic IP routing protocol installs its routes in the routing database (RDB). The RDB stores all sources of routing information. The routing table manager (RTM) manages the RDB.

As part of the filtering process, the routing protocol applies local import policy to the information learned from a peer and places the best route to each destination in the RDB. Thus, the RDB contains the best routes computed by each protocol.

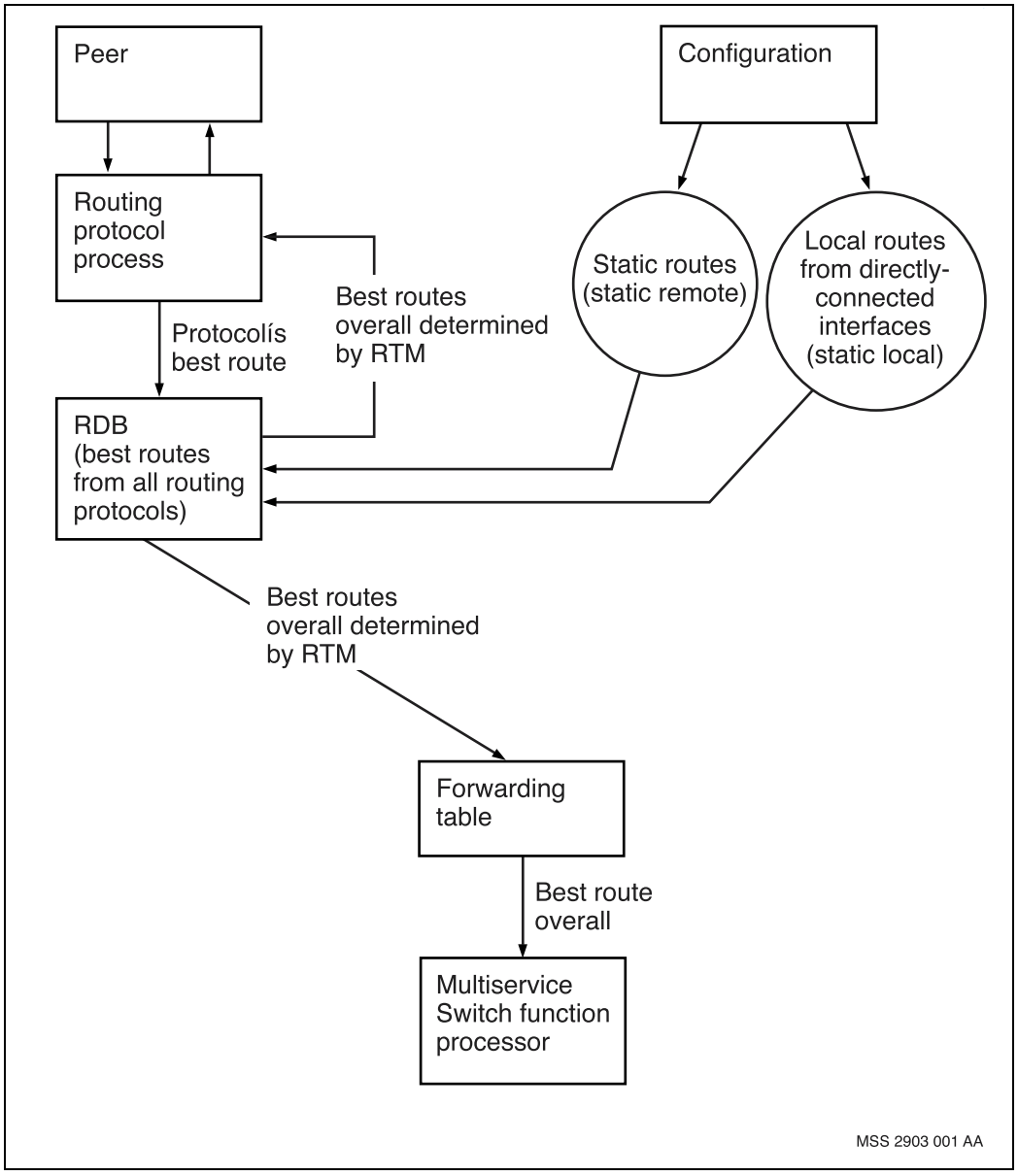
The RTM examines the RDB and selects the best overall route to each destination. If there is only one route to a given destination in the RDB, then it is the best overall route. If there are multiple routes to a given destination in the RDB, for example a route installed by OSPF and a route installed by BGP, then the RTM selects the best overall route according to a specific order. For more information on preferences in route selection, see Route preferences (page 86).

The RTM informs each routing protocol of the best overall routes. Each routing protocol applies its export policy to determine if it will send these routes to peers. See Routing policies (page 83) for more information.

The RTM also places the best overall routes in the IP forwarding table of the Nortel Networks Multiservice Switch node. The IP forwarding table stores the routes used by the forwarding process. The routes in the forwarding table are also sent to the node's function processors.

The Flow of routing information through the Multiservice Switch system (page 86) illustrates how information flows from peers through the routing protocol, to the RDB, then to each routing protocol as well as to the IP forwarding table and the node's FPs.

Flow of routing information through the Multiservice Switch system



Route preferences

See the following sections for information on route preferences and their purpose:

- Overview of route preferences (page 87)
- Forwarding classes and route preferences (page 88)
- Override recommendations (page 89)

For information on the provisioning procedures used to modify the route preference for each routing protocol, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management* and NN10600-445 *Nortel Networks Multiservice Switch 7400/15000/20000 Operations: Multiprotocol Label Switching*.

Overview of route preferences

You can have more than one valid route to the same destination. When this happens, the routing manager uses the route preference to distinguish the different routes. The route preference is a number that is assigned to the route entry and it represents the dependability of the routing information. The lower the number, the more dependable the routing information. See the table Default route preference values (page 87).

You can assign different values to a route preference, allowing you to engineer how IP traffic is routed through your network. It is highly recommended that you do not change the default route preference unless you have done thorough engineering and modelling of the network and you fully understand the potential impacts.

When you change a route preference using a provisioning procedure, it is recommended that the change is applied to all virtual routers in the network/ autonomous system during low traffic periods in order to avoid potential routing loops. You do not need to stop and restart the protocol for the change to take effect.

If you use the same value for more than one route preference, the routing protocols' metrics are used to determine which routing protocol is used. The protocol that has the best metric is used.

Default route preference values

Protocol	Default route preference
Local, static discard	0 *
MPLS	10
Reserved for RIP migration	20 *
Reserved for RIP migration	21 *
OSPF internal	30
BGP external	70
Static remote	72
OSPF external type 1	80

(1 of 2)

Default route preference values (continued)

Protocol	Default route preference
RIP	82
OSPF external type 2	120
BGP internal	122
BGP aggregate	124 *
Reserved for internal use	254 *
UN_PREF	255 **
* reserved default values for internal protocols that are non-configurable	
** used by the software and represents routes that are never put in the forwarding table	
(2 of 2)	

Forwarding classes and route preferences

Each route preference belongs to one of three forwarding classes. The lower the value of the forwarding class, the more important the routing information. The table Forwarding classes (page 89) shows how a route preference is assigned a forwarding class. It is recommended that whenever you change a route preference, to keep the new value within its existing forwarding class.

The forwarding class has an effect on routing. Specifically, a new more specific route can only be inserted into the forwarding table as long as less specific routes do not exist in a more important forwarding class.

Example

Assume an OSPF internal route (forwarding class 0) exists for address 10.1.0.0/16 and BGP internal (forwarding class 1) finds a route for address 10.1.1.0/24.

The BGP route, which is more specific, is not allowed in the forwarding table because its forwarding class is less important than the forwarding class of the OSPF route.

Forwarding classes

Route preference range	Forwarding class
0-63	0
64-127	1
128-255	2

There are 2 exceptions to that rule:

- Locally provisioned static routes (more specific route) are inserted into the forwarding table even if a less specific route does exist in a more important forwarding class
- Default route (learned via IP routing protocol) is inserted into the forwarding table even if a less specific route does exist in a more important forwarding class

Override recommendations

In addition to changing the route preference, some protocols have an override facility that allows you to prefer one protocol over another. When you use these overrides, it is recommended that you use the following values. It is assumed that all other protocols are using their original route preference values.

- To prefer BGP internal routes over OSPF internal routes, the recommended setting for attribute *ibgpRtePref* is 6.
- To prefer BGP external routes over OSPF internal routes, the recommended setting for attribute *ebgpRtePref* is 6.
- To prefer static remote routes over OSPF internal routes, the recommended setting for attribute *staticRemoteRtePreference* is 5.

For information on the provisioning commands to use these overrides, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Example routing topologies

There are many configuration options that enable the redistribution of route information between different dynamic routing protocols. Some common redistribution scenarios include:

- Route redistribution between two interior routing protocols within a single autonomous system (AS) (page 90)
- Route redistribution from an interior routing protocol to EBGP (page 91)

Route redistribution between two interior routing protocols within a single autonomous system (AS)

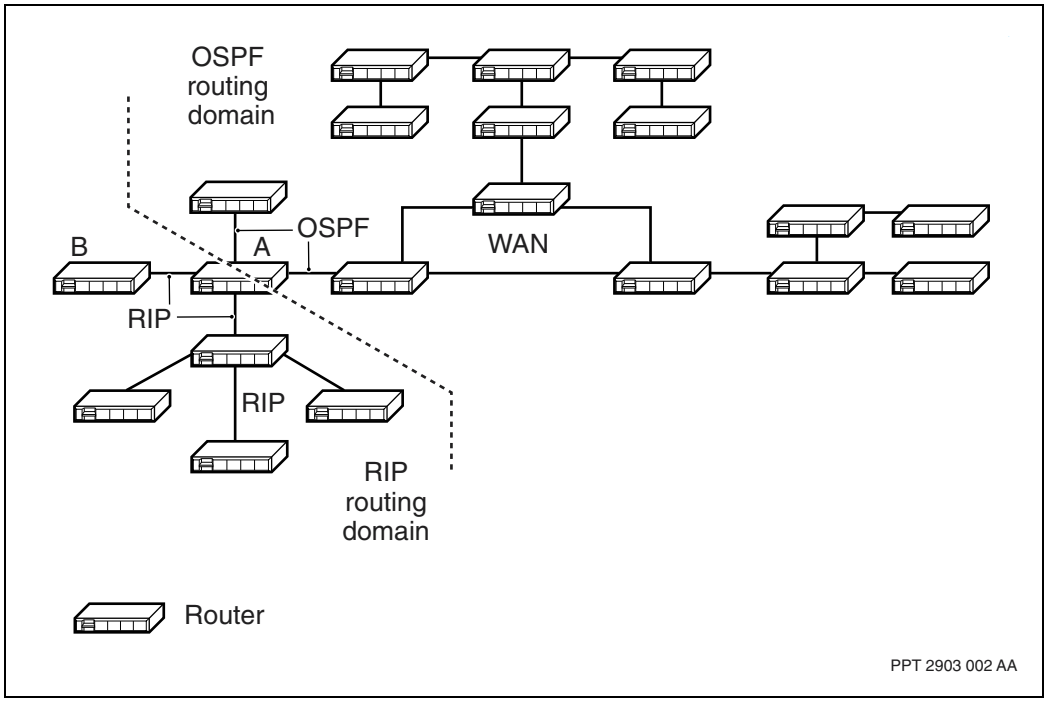
An autonomous system (AS) is a series of gateways or routers that fall under a single administrative entity and cooperate using the same Interior Gateway Protocol (IGP). The figure Route redistribution between two interior routing protocols within a single AS (page 90) illustrates the redistribution of route information from a RIP routing domain to an OSPF routing domain.

In this example, router B sends RIP routes to router A. Router A's RIP process applies import policy to the routes it receives from router B. It might reject some routes. The RIP process selects the best routes from among the routes received from all RIP peers and installs these routes in the RDB.

The RTM selects the best routes overall from the RDB and installs them in the forwarding table. Some of the best overall routes might be RIP routes.The routes installed in the forwarding table are sent to the Nortel Networks Multiservice Switch node FP and to the other routing protocols operating on router A.

The OSPF process on router A receives the routes sent to it by the RTM. The OSPF process applies export policy to these routes to determine if it will send them to OSPF neighbors. For more information on how OSPF manages the routes it receives, see Route redistribution from an interior routing protocol to EBGP (page 91).

Route redistribution between two interior routing protocols within a single AS



Route redistribution from an interior routing protocol to EBGP

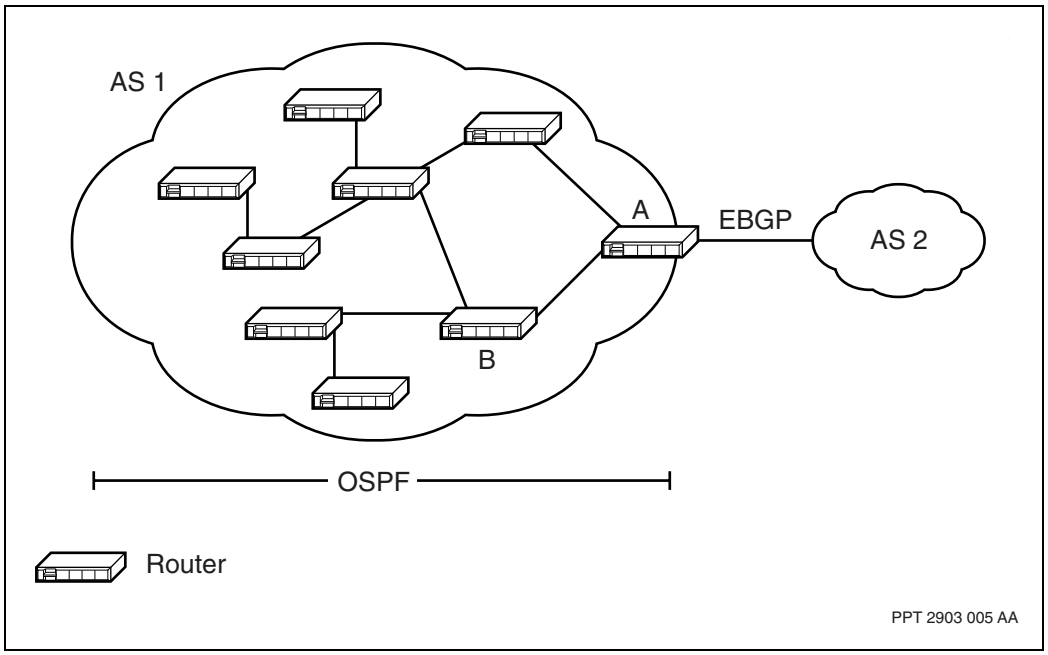
The figure Route redistribution from an interior routing protocol to EBGP (page 91) illustrates the redistribution of route information from an OSPF routing domain to a BGP-4 routing domain.

In this example, the OSPF process running on router A is exchanging link state advertisements (LSAs) with its neighbors in AS 1. Since OSPF has no import policies, it cannot filter these LSAs. OSPF computes the shortest paths to all destinations described in the LSAs and installs these routes in the RDB.

The RTM selects the best routes overall from the RDB and installs them in the forwarding table. Some of these overall best routes may be OSPF routes. The routes installed in the forwarding table are sent to the Nortel Networks Multiservice Switch node FP and to the other routing protocols operating on router A.

The BGP process operating on router A receives the routes sent to it by the RTM. The BGP process applies export policy to these routes to determine if it will send them to the BGP peers in AS 2.

Route redistribution from an interior routing protocol to EBGP



IP differentiated services for routing packets

The Differentiated services code point field of routing packets is controlled by the *VirtualRouter Ip dscpRoutingSource* attribute. See the *Vr Ip* section of the NTP NN10600-060 *Nortel Networks Multiservice Switch 7400/15000/20000 Component Reference* for more information on the *dscpRoutingSource*

attribute. See NN10600-590 *Nortel Networks Multiservice Switch 7400/15000/20000 Layer 3 Traffic Management Fundamentals* for more information on IP Differentiated Services.

Routing information protocol (RIP)

This section describes Nortel Networks Multiservice Switch implementation of the routing information protocol (RIP).

Navigation

Overview of Multiservice Switch RIP

RIP is a routing protocol based on the Bellman-Ford (or distance vector) algorithm. Neighboring routers send routing updates containing a list of destination addresses and their costs to reach them. From this information, the RIP router calculates the shortest (least costly) path to each destination. RIP is a simple protocol suitable for small networks with a maximum diameter of 15 hops. (Destinations with metrics greater than 15 are considered unreachable.)

Nortel Networks Multiservice Switch nodes support RIP version 1 (RIPv1) and some of the extensions included in RIP version 2 (RIPv2). RIPv1 uses a fixed subnet mask rule for different classes of IP addresses. Multiservice Switch RIPv2 adds support for variable length subnet masks, multicasting, and next hop specification, as described by RFC2453.

RIP policies

RIP import policies define which learned set of routing information is given to the RIP routing process, as well as which metrics to use. You can assign non-default metrics so that traffic must use a particular route unless that route becomes unusable. You can also use import policies to define which routing processes (for example, OSPF) provide routing information to the RIP routing process.

RIP export policies define how to advertise routing information on specific interfaces. You can assign non-default metrics so that traffic must use a particular route unless that route becomes unusable. RIP export policies also define which routing processes, learned from a specific interface, can be exported. RIP export policies are optional.

Attention: If you provision a RIP export policy make sure you are aware of the export policies of the “send all staticLocal routes” and “send all RIP learned routes” as well as their possible impact on other configured policies.

For more information on import and export policy, see Routing policies (page 83). To configure import and export policies for RIP, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

RIP can have both import and export policies.

RIP import policies control usage of routes learned from RIP neighbors. Routes may be ignored, used, or used with an override metric. Setting override metrics for imported routes allows you to force traffic to use a particular route unless that route becomes unavailable. If no import policy is configured, all routes learned from RIP neighbors are used.

RIP export policies control advertisement of routing information to the router's RIP neighbors. This includes routing information learned from RIP as well as information learned from other routing processes (for example, OSPF). Routes may be blocked, sent, or sent with a modified metric value. Modifying the metric value may be required to prevent large metric values calculated by other routing protocols from being considered unreachable by RIP, which treats all destinations with metrics larger than 15 as unreachable. If no export policy is configured, no routes from other routing processes are advertised to RIP neighbors, but all RIP routes and local routes are advertised.

For more information on import and export policy, see “Routing policies” (page 83). To configure import and export policies for RIP, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Migrating from RIPv1 to RIPv2

RIPv2 adds several additional capabilities which RIPv1 does not support. The primary advantages of RIPv2 are more efficient use of IP addresses through variable length subnet masks, more efficient routing through specification of the next hop, and reduced processing load on hosts not listening to RIPv2 packets through multicasting routing updates.

To migrate Nortel Networks Multiservice Switch network nodes from RIPv1 to RIPv2, all the nodes must be running a software release that supports RIPv2 (R5.1 and later). Migrate all Multiservice Switch nodes on a link-by-link basis, until all the nodes are set to RIPv2.

The following example describes the sequence of steps involved in migrating from RIPv1 to RIPv2 using two Multiservice Switch nodes. The steps correspond to the figure Example RIPv1 to RIPv2migration from one node to another (page 96).

1 Add RIPv2 to Multiservice Switch 2, but provision the RIP interface on Multiservice Switch 2 to be backwards compatible with RIPv1 on Multiservice Switch 1.

   For example, set the ifConfSend attribute on the RIP interface of Multiservice Switch 2 to v2b.

   See NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management* for details on configuring the *ifConfSend* and *ifConfReceive* attributes under the *RipIf* component. See the table Example migration: RIP behavior on two nodes with different RIP configurations (page 97) for the meaning of the values of the *ifConfSend* and *ifConfReceive* attributes.

   The table Example migration: RIP behavior on two nodes with different RIP configurations (page 97) also illustrates the behavior of the RIP interface on these two nodes for different combinations of attribute values provisioned for the *ifConfSend* and *ifConfReceive* attributes. This table can be useful in helping you planning your migration. The attribute values appear in the table in italics.

2 Configure the RIP interface on Multiservice Switch 1 to support RIPv2 only.

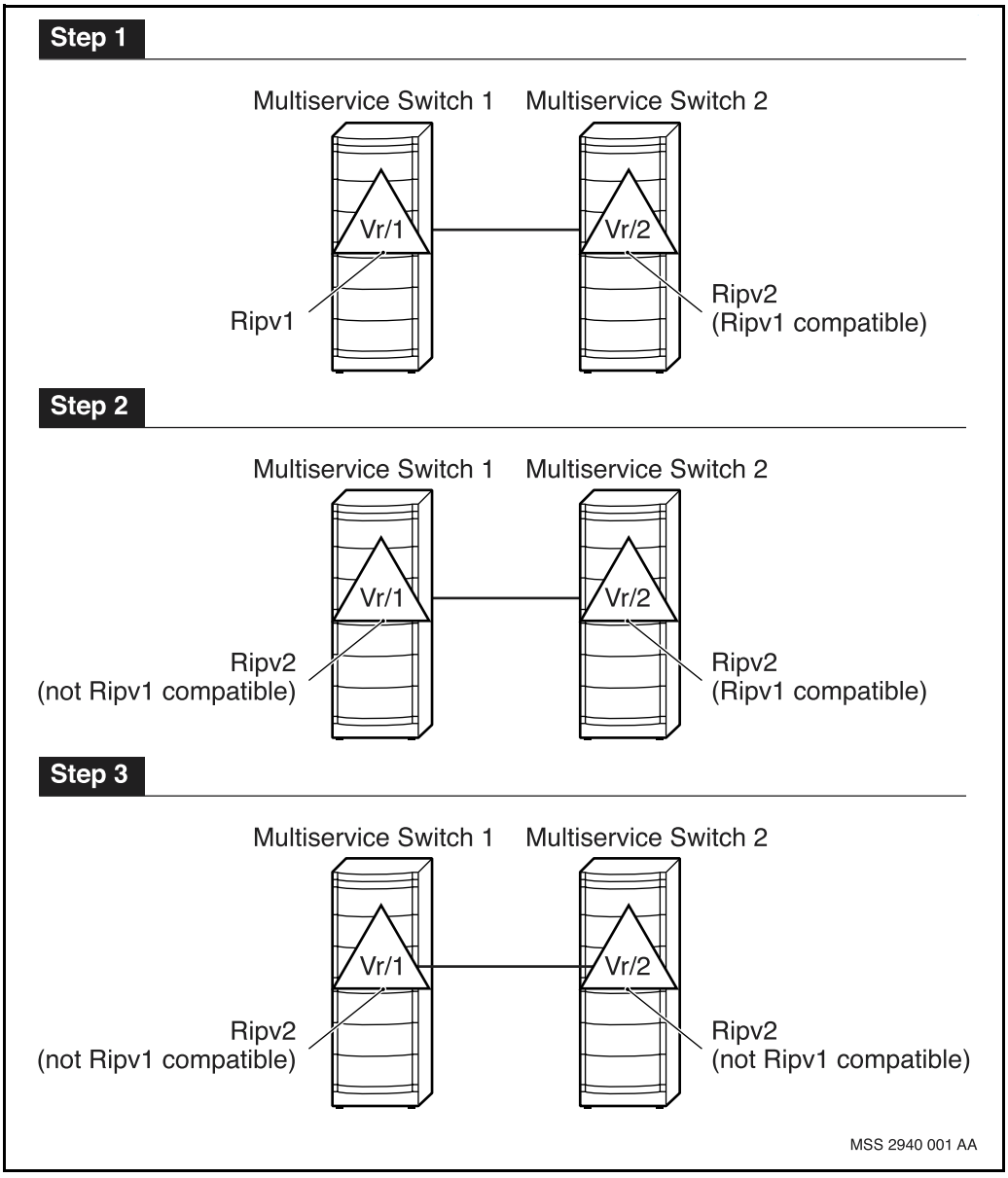
   For example, set the ifConfSend attribute on the RIP interface of Multiservice Switch 1 to v2, and the ifConfReceive attribute to v2.

3 Change the RIP interface on Multiservice Switch 2 to support RIPv2 only.

   For example, set the ifConfSend attribute on the RIP interface of Multiservice Switch 2 to v2, and the ifConfReceive attribute to v2.

4 Remove all RIPv1 components from Multiservice Switch 2.

Example RIPv1 to RIPv2migration from one node to another



Example migration: RIP behavior on two nodes with different RIP configurations

ifConfSend attribute value on Multiservice Switch 1 (Vr/1) (transmitting)	**ifConfReceive attribute value on Multiservice Switch 2 (Vr/2) (receiving)**			
	***v1* (RIP 1)**	***v2* (RIP 2)**	***both* (RIP 1 or 2)**	***reject* (do not accept)**
silent (do not send)	No transmission	No transmission	No transmission	No transmission/ updates are rejected.
v1 (RIP 1)	RIP 1 updates broadcast by Vr/ 1. RIP 1 updates accepted by Vr/ 2.	RIP 1 updates broadcast by Vr/ 1. RIP 1 updates rejected by Vr/2.	RIP 1 updates broadcast by Vr/ 1. RIP 1 updates accepted by Vr/ 2. The Vr/2 RIP interface processes the updates as RIP 1 updates.	RIP 1 updates broadcast by Vr/ 1. RIP 1 updates are rejected by Vr/2.

(1 of 2)

Example migration: RIP behavior on two nodes with different RIP configurations (continued)

ifConfSend attribute value on Multiservice Switch 1 (Vr/1) (transmitting)	ifConfReceive attribute value on Multiservice Switch 2 (Vr/2) (receiving)			
	***v1* (RIP 1)**	***v2* (RIP 2)**	***both* (RIP 1 or 2)**	***reject* (do not accept)**
v2b (RIP 1 compatible)	RIP 2 updates broadcast by Vr/1. RIP 2 updates accepted by Vr/2. The Vr/2 RIP interface processes the RIP 2 updates as RIP 1 updates. (The Vr/2 RIP interface ignores the subnet mask and next hop fields in the RIP 2 update.)	RIP 2 updates broadcast by Vr/1. RIP 2 updates accepted by Vr/2. Vr/2 RIP interface will process the subnet mask and next hop fields in the RIP 2 updates.	RIP 2 updates broadcast by Vr/1. RIP 2 updates accepted by Vr/2. Vr/2 RIP interface will process the subnet mask and next hop fields in the RIP 2 updates.	RIP 2 updates broadcast by Vr/1. RIP 2 updates are rejected by Vr/2.
v2 (RIP 2)	RIP 2 updates are multicast by Vr/1. Because the Vr/2 RIP interface is set for RIP 1 only, Vr/1 will not send RIP 2 updates to Vr/2.	RIP 2 updates are multicast by Vr/1. RIP 2 updates are accepted by Vr/2. Vr/2 RIP interface will process the subnet mask and next hop fields in the RIP 2 updates.	RIP 2 updates are multicast by Vr/1. RIP 2 updates are accepted by Vr/2. Vr/2 RIP interface will process the subnet mask and next hop fields in the RIP 2 updates.	Updates are rejected

(2 of 2)

Migrating from RIP to OSPF

RIP is a simple protocol designed for small networks with a maximum diameter of 15 hops. For larger networks and where more sophisticated routing, greater administrative control, and quicker convergence after topological changes are required, a protocol such as OSPF is preferred.

Use one of the following methods to migrate your network from RIP to OSPF. Using a route preference is preferable to the *migrateRip* attribute because enabling and disabling *migrateRip* restarts the protocol.

Attention: After enabling *migrateRip*, any subsequent route preference changes are not enabled until *migrateRip* is disabled.

Using a route preference

Change the route preference of either RIP or OSPF internal so that RIP routes are preferred over OSPF internal routes. The route preference change must be done on all RIP and OSPF virtual routers in the network/autonomous system. See Route preferences (page 86) for more information.

Using migrateRip

When making a transition from RIP to OSPF, the *migrateRip* attribute of the *Ospf* and *Rip* components can be enabled during the transition phase. This process allows both RIP and OSPF routes to be learned, but gives preference to RIP-learned routes. When OSPF has proven its stability, disable the *migrateRip* attribute to allow OSPF to take over route selection duties.

All RIP and OSPF virtual routers in the network/autonomous system must have the *migrateRip* attribute in the same state and the changeover must be made quickly to prevent routing loops.

For more information about migrating from RIP to OSPF using *migrateRip*, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Open shortest path first (OSPF) protocol

This section describes Nortel Networks Multiservice Switch implementation of the open shortest path first (OSPF) protocol.

Navigation

Overview of OSPF

OSPF is a link-state route management protocol. Neighboring routers send routing updates, called link state advertisements (LSAs), that include the cost of internal and external routes. From information provided by neighboring routers, the OSPF process creates a tree-like map of the network with itself at the root. From this tree, it creates a routing table using the LSAs to calculate the shortest paths.

OSPF is defined in RFC1583.

This section covers the following topics:

OSPF areas

An autonomous system (AS) is a series of gateways or routers that fall under a single administrative entity and cooperate using the same Interior Gateway Protocol (IGP). An OSPF autonomous system (AS) can be subdivided into areas. An area is a contiguous collection of networks and hosts. The topology

of an area is invisible to routers outside the area, and the topologies of other areas are unknown from within an area. As a result, the topological database of each router in the AS can be different from other routers within the AS.

There are four OSPF area types: transit, stub, summary stub, and not-so-stubby. The backbone area is a special case of a transit area.

Transit areas can carry data traffic that neither originates nor terminates in the area. Routers in transit areas learn about all ASBRs and external routes in the OSPF domain, in addition to the topology of their own area and summary information about other areas. If an active virtual link transits an area, then the area must be a transit area.

Stub areas are “leaf” areas which cannot carry transit traffic. Internal routers in stub areas have no knowledge of the network outside of the area, which allows them to have smaller link-state databases. An area can be configured as a stub area when there is a single exit point or the choice of exit point does not have to be made based on the external destination. Stub area border routers (ABRs) which have an active connection to the backbone area can advertise a default route into the stub area which can be used for routing to destinations external to the area. Stub areas cannot include an ASBR connecting the AS to an external AS.

Summary stub areas are stub areas which allow the importing of summary information about other areas. This allows more informed inter-area routing while still limiting the amount of information required to be stored by internal routers in the area.

Not-so-stubby areas (NSSA) are stub areas which allow the importing of summary information about other areas and may include an ASBR connecting the AS to an external AS. This allows routers in the area to have routing information for external destinations reached through their own area, while still blocking information about external destinations reachable through other areas. External information learned from an ASBR in an NSSA area is propagated throughout the OSPF AS.

The backbone area for an AS is a specialized transit area with area ID 0.0.0.0. It is responsible for distributing routing information between non-backbone areas. The backbone always contains all ABRs, and may also contain internal routers. The backbone must be contiguous or have virtual links provisioned to connect the disjoint segments.

OSPF supports two main configurations of areas:

- Hierarchical OSPF divides the internetwork into contiguous logical areas, with each area usually corresponding to a community of interest. Each area, including the area 0 backbone, runs a separate copy of the OSPF

algorithm. Each router in a particular area knows how to reach every subnet in that area as well as how to reach the backbone. Once the datagram is given to the backbone, it is the backbone's job to route the datagram to the destination area. The destination area's interior routers know how to complete the routing to the destination end system. Hierarchical OSPF is characterized by smaller routing tables because of OSPF partitioning, but usually requires more router hardware.

- Collapsed backbone refers to the practice of having only one routing area defined, with all routers and networks in the AS belonging to the same area. Collapsed backbone configurations are ideal for small (under 20 routers) IP networks.

OSPF routing types

The routes that are learned from the OSPF protocol can be divided into two types, internal and external. Internal routes are internal to the OSPF routing domain and external routes are learned from other routing protocols and are for destinations outside the OSPF routing domain. By default, internal routes are preferred over external routes.

External routes are imported into the OSPF routing domain by autonomous system boundary routers (ASBRs). There are two types of external routes:

- OSPF external type 1, where the metric assigned to the external part of the route has the same order as the metric assigned to the internal part (source to ASBR) of the route. Add the internal and external costs to calculate the total cost of the route. For example, using hop-count as the metric in the OSPF routing domain, you can import RIP routes as external type 1 routes.
- OSPF external type 2, where the metric assigned to the external part of the route is more significant than the metric assigned to the internal part of the route. The total cost of the route is equal to the external cost.

By default, type 1 routes are preferred over type 2 routes. For more information on route preferences, see Route preferences (page 86).

OSPF router types

Nortel Networks Multiservice Switch nodes can function as any one of four types of OSPF routers. To configure any of these router types, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

- Internal routers are routers whose directly-connected networks belong to the same area. There can be internal routers in the backbone area if all of their interfaces are in the backbone. It is sufficient to configure a single instance of the *AreaTable* subcomponent for internal or routers, since they are part of only one area.

- Area border routers (ABR) connect to one or more areas and the backbone. Area border routers can condense or summarize the topological data of their attached areas for distribution on the backbone. The backbone in turn distributes the information to other areas. A single instance of the AreaTable subcomponent is sufficient for internal or backbone routers, since they are part of only one area. You must configure instance for each area of the *AreaTable* subcomponent for area border routers.
- Backbone routers are routers that have an interface to the backbone (including ABRs). A backbone router that has connections only to other backbone routers is also considered an internal router. A single instance of the *AreaTable* subcomponent is sufficient for backbone routers, since they are part of only one area.
- Autonomous system boundary routers (ASBR) exchange routing information with routers belonging to other autonomous systems. The path to each ASBR is known by every router in the autonomous system (AS), except in the case of stub areas. See OSPF areas (page 100) for more information. This classification is completely independent of the previous classifications. ASBRs can be internal routers or area border routers, and may or may not participate in the backbone.

OSPF virtual links

OSPF does not allow a non-contiguous backbone (area 0.0.0.0), nor does it allow a router to connect two or more areas, becoming an area border router, unless the router is part of the backbone. One solution is to provision a virtual link reaching from the isolated area border router, through one of the attached areas, to the backbone. For information on configuring virtual links, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

OSPF export policy

Export policies define how routing information is shared between routing processes (for example, between OSPF and RIP). Export policies are optional and are not present in cases where you don't want to share routing information. For information on configuring OSPF export policy, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Since OSPF normally uses IP routing information from all sources, OSPF import policies are not used. OSPF allows only export policies.

For more information on import and export policies, see Routing policies (page 83).

OSPF equal-cost multipath routing

OSPF supports an equal-cost multipath routing function where equal-cost paths are used in load-sharing mode. Equal-cost multipath routing supports up to three next hop addresses. For more information on equal-cost multipath routing, see Static routes (page 123).

OSPF optimization

See the following sections for more information on various ways to optimize OSPF:

- Optimizing OSPF memory allocation (page 104)
- Hitless OSPF for CP/VpnXc switchover (page 104)

There is additional information about OSPF optimization in Inverse ARP scalability (page 30).

Optimizing OSPF memory allocation

You can optimize OSPF memory allocation by configuring estimates for the number of routes in the network. Nortel Networks Multiservice Switch nodes use these numbers to calculate queue sizes for OSPF. In this way, you can make more effective use of memory by allocating only what is necessary.

You can set estimated values for the number of

- internal OSPF routes (routes learned from OSPF neighbors)
- external OSPF routes (routes learned from BGP-4 and RIP)
- OSPF areas in the AS
- OSPF interfaces in each area
- OSPF neighbors for each OSPF interface

For information on configuring route estimates, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*. If you do not set values for the parameters described, Multiservice Switch nodes use default values to calculate OSPF queue sizes. These default values result in queue sizes similar to those allocated to the IP service prior to the introduction of the user-configurable parameters.

Hitless OSPF for CP/VpnXc switchover

Hitless OSPF refers to the ability to maintain a synchronized OSPF instance on the standby card. Its purpose is to further secure the core of an IP VPN. On a VCG-based IP VPN network configuration, hitless OSPF can be enabled on both VCGs and/or customer VRs. On a direct VR-to-VR configuration, it can be enabled on any virtual router.

Hitless OSPF can protect OSPF against failure in the following spaces: from the CE to the customer VR, and from VCG to VCG.

See NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals* and NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management* for more information on IP VPNs.

Hitless OSPF is available on the following:

- On a Nortel Networks Multiservice Switch node that has either a control processor (CP) or both a CP and a VPN extender card (VpnXc) along with their respective standby modules
- On IP traffic over ATM MPE protocol ports on PQC2 and PQC12-based FPs
- On IP traffic over IP-optimized DLCI protocol ports on PQC2 and PQC12-based FPs
- On access connections with supported media (ATM MPE and IP-optimized DLCI) that use OSPF

Background

All routers running OSPF in the network maintain identical databases of link state advertisements (LSAs). OSPF accomplishes this by having routers synchronize their link state databases with neighboring routers by exchanging LSAs. Whenever the LSAs change in the database, the IP routes are recalculated. These new routes and other IP routing protocol routes make up the routing table.

Since an OSPF instance establishes relationships with its neighbors, an unplanned restart of OSPF affects not only the restarting router but also the general behavior of IP traffic within an OSPF network. IP traffic is interrupted by an unplanned OSPF restart due to two main reasons:

- In the restarting router's routing table, OSPF-discovered IP routes are identified and removed, which limits its forwarding ability.
- The OSPF neighbors detect the restart and inform the network that the OSPF instance has restarted. This causes traffic to be redirected around the restarting router.

Hitless OSPF description

Enabling the hitless OSPF feature allows information such as the neighbor relationships and the link state database to be preserved during a planned or unplanned switchover. In particular, remote OSPF routers within the network, especially neighbors, do not detect any change in the topology. Generated IP routes prior to the switchover are not affected, allowing the standby card to seamlessly assume normal neighboring interactions.

The benefits of hitless OSPF are apparent when the switchover is executed on the card (CP or VpnXc) where OSPF resides. If OSPF resides on the VpnXc and a CP switchover is executed, OSPF is unaffected whether or not hitless OSPF is enabled. However under the following conditions, OSPF remains unaffected only when hitless OSPF is implemented:

- OSPF resides on the VpnXc and a VpnXc switchover is executed
- OSPF resides on the CP and a CP switchover is executed

Enable hitless OSPF by setting attribute *Vr Ip Ospf spareInstance* to enable. In order for hitless OSPF to function, you must set attribute *Shelf cpEquipmentProtection* to hot.

Migrating from RIP to OSPF

If your routing requirements change, you can migrate from RIP to OSPF. Use one of the following methods to migrate your network from RIP to OSPF. Using a route preference is preferable to the *migrateRip* attribute because enabling and disabling *migrateRip* restarts the protocol.

Attention: After enabling *migrateRip*, any subsequent route preference changes are not enabled until *migrateRip* is disabled.

Using a route preference

Change the route preference of either RIP or OSPF internal so that RIP routes are preferred over OSPF internal routes. The route preference change must be done on all RIP and OSPF virtual routers in the network/autonomous system. See Route preferences (page 86) for more information.

Using migrateRip

When making a transition from RIP to OSPF, the *migrateRip* attribute of the *Ospf* and *Rip* components can be enabled during the transition phase. This process allows both RIP and OSPF routes to be learned, but gives preference to RIP-learned routes. When OSPF has proven its stability, disable the *migrateRip* attribute to allow OSPF to take over route selection duties.

All RIP and OSPF virtual routers in the network/autonomous system must have the *migrateRip* attribute in the same state and the changeover must be made quickly to prevent routing loops.

For more information about migrating from RIP to OSPF using *migrateRip*, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Border gateway protocol 4 (BGP-4)

This section describes Nortel Networks Multiservice Switch implementation of the border gateway protocol 4 (BGP-4).

Navigation

Overview of BGP-4

The border gateway protocol 4 (BGP-4) is a routing protocol for exchanging network reachability information between autonomous systems (ASs). BGP-4 is replacing the external gateway protocol (EGP) as the routing protocol of the Internet backbone.

BGP-speaking routers establish peer relationships with other BGP-speaking routers over a TCP connection and exchange routing information. Using this reachability information, the BGP-speakers construct a map of AS connectivity that allows them to eliminate routing loops and enforce policy decisions at the AS level. BGP-4 also supports classless inter-domain routing as described in RFC1519, and route aggregation.

BGP-4 is defined in RFC1771 and RFC1772.

This section covers the following topics:

BGP-4 peers

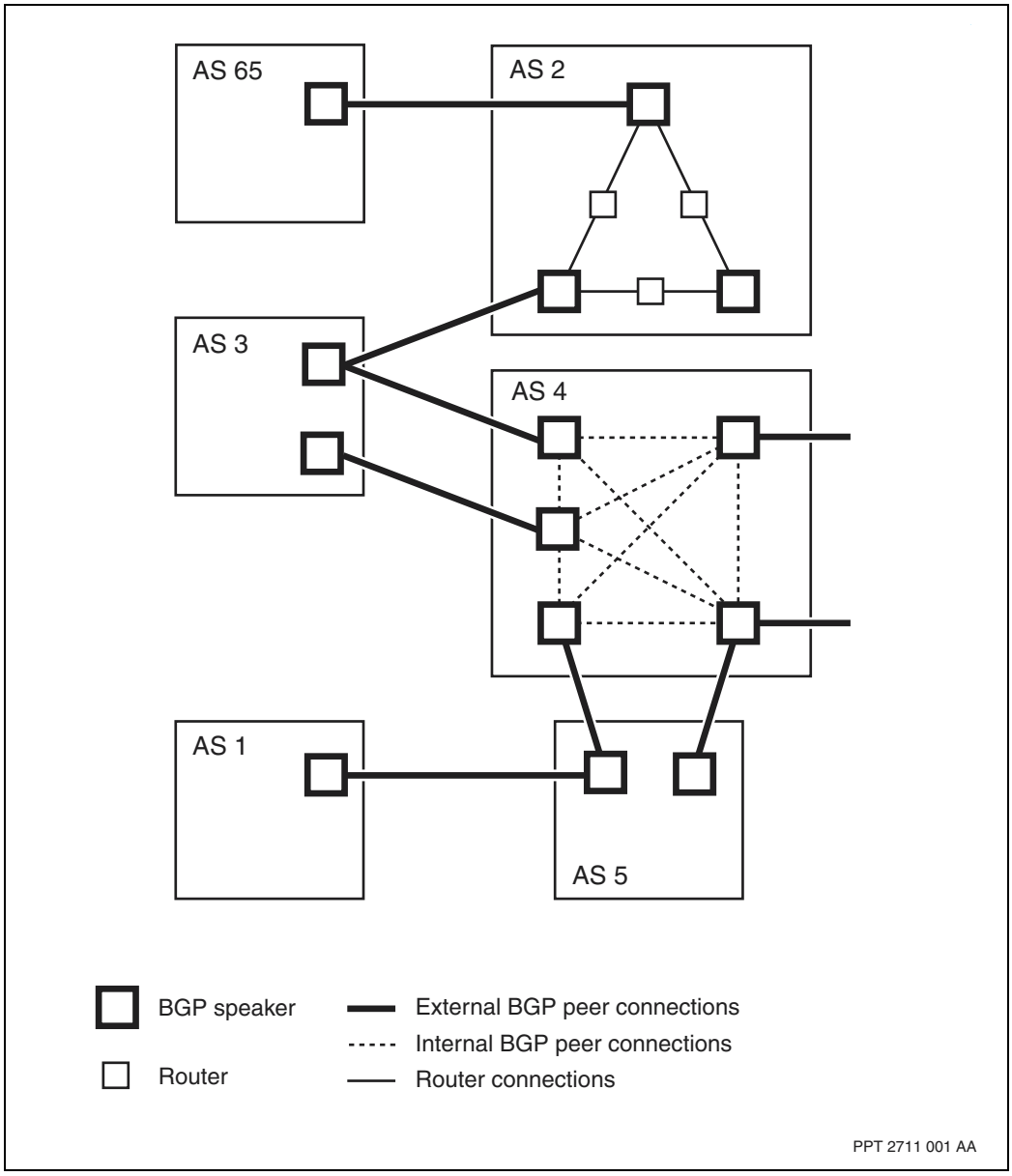
BGP-4 peers are a pair of BGP-4 speakers that exchange routing information about reachable destinations in different ASs. If the two BGP-4 speakers are in the same AS, they are internal BGP (IBGP) peers. If they are in different ASs, they are external BGP (EBGP) peers. The figure BGP-4 peering (page 109) shows an example of BGP peer relationships.

BGP-4 sets up peer relationships between BGP-speaking routers over a TCP connection. TCP exchanges the peer connection requests, responses, and route updates. When the peer relationship is first established, the BGP-4 speakers exchange the entire BGP routing table. After the initial exchange of the complete routing table, peers exchange only route changes whenever network topology changes. In the meantime, peers exchange periodic keep alive messages to confirm connectivity.

Any given AS can have several different types of peerings with another AS:

- stub AS singly-homed, for example, AS 65 and AS 1 in the figure BGP-4 peering (page 109).
- multi-homed non-transit A, for example, AS 5 in the figure BGP-4 peering (page 109).
- multi-homed transit AS, for example, AS 2, AS3, and AS 4 in the figure BGP-4 peering (page 109).

BGP-4 peering



Single-hop and multi-hop BGP

Nortel Networks Multiservice Switch nodes support single-hop and multi-hop BGP configurations.

In a single-hop BGP configuration, routes are distributed to remote BGP peers located on a directly attached network. The figure Single-hop BGP configuration (page 110) shows router A and router B in a single-hop BGP configuration.

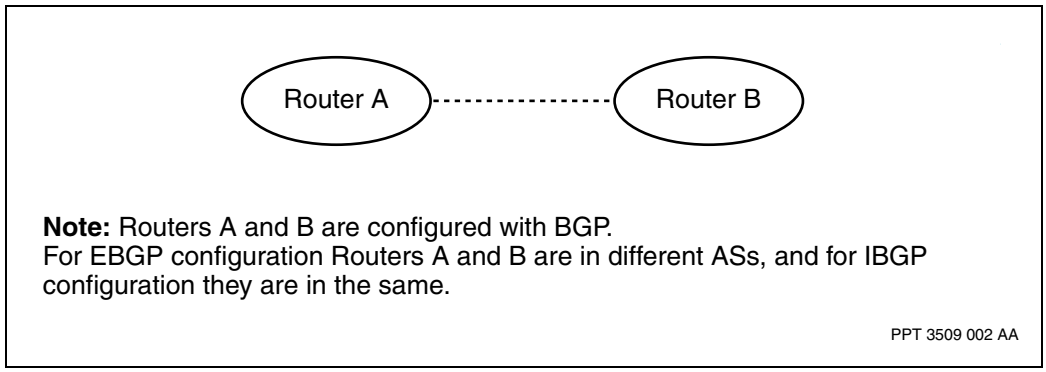
In a multi-hop BGP configuration, routes are distributed with BGP connections established between peers that are not on a directly attached network and can span across multiple hops. The figure Single-path multi-hop BGP configuration (page 111) shows router A and router B in a single path multi-hop BGP configuration.

Multi-hop BGP peering can be direct or virtual. Direct peering refers to using real interface addresses between routers, which are BGP peers. Virtual peering refers to using virtual interface addresses on the routers, which are BGP peers. A virtual interface address can be for example, an always-up IP interface address or a loopback IP address. Because virtual peering is not tied to a specific physical interface and always stays up, it is ideal to use when redundancy is required. Virtual peering configuration is recommended when multiple paths are available and can provide redundancy for BGP route distribution. When one path fails, another one can be selected without losing BGP peer connection. For multi-path multi-hop BGP, see Multi-path multi-hop BGP configuration (page 111).

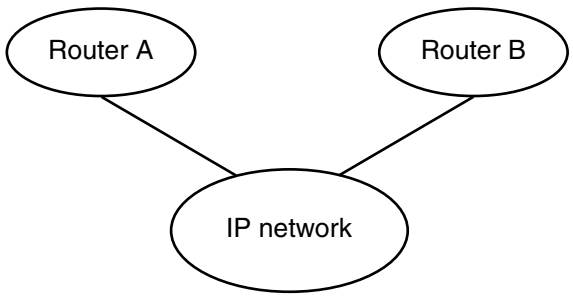
Configure routers in a multi-path multi-hop BGP configuration to provide:

- redundancy between peers so that peers are reachable by multiple paths. Only one path is used at a time.
- peering over multiple hops

Single-hop BGP configuration



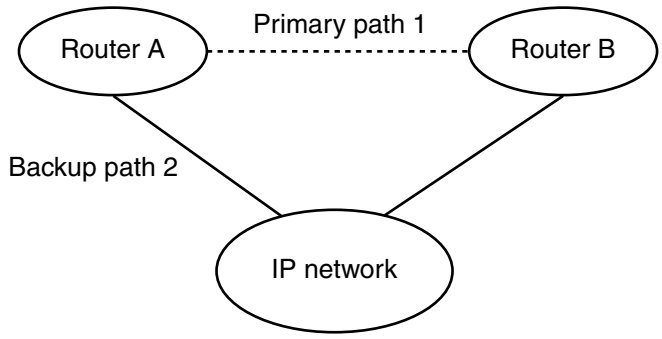
Single-path multi-hop BGP configuration



Note: Routers A and B are configured with BGP.
For EBGP configuration Routers A and B are in different ASs, and for IBGP configuration they are in the same.

PPT 3509 001 AA

Multi-path multi-hop BGP configuration



Note: Routers A and B are configured with BGP.
For EBGP configuration Routers A and B are in different ASs, and for IBGP configuration they are in the same.

PPT 3509 003 AA

BGP-4 updates

BGP-4 peers exchange routing information about reachable destinations through network layer reachability information (NLRI) update messages. If route information changes, BGP-4 informs its peers by withdrawing the invalid routes and advertising new route information.

To conserve bandwidth and processing resources, BGP-4 uses incremental updates. BGP-4 also uses route aggregation to represent information about a group of networks as a single entity. In addition to learning routes from peers, a BGP-4 router can originate routing updates to advertise networks that belong to its own AS.

BGP-4 path attributes

Path attributes associated with the NLRI updates provide detailed information about an advertised route. Routers can use path attribute information when making policy decisions.

To avoid routing loops, BGP-4 prepends the AS number of the BGP speaker to the AS path attribute of a route when that route is advertised to an EBGP peer.

Under the AS path attribute, you can configure

- path attributes
- AS weights. For more information seeAS weights (page 118).
- private AS number removal. For more information, see Private AS number removal (page 120).

The table BGP-4 path attributes (page 112) lists the path attributes supported on Nortel Networks Multiservice Switch nodes.

BGP-4 path attributes

Path attribute	Description
Origin	Specifies whether an NLRI originated from an interior gateway protocol (IGP), an exterior gateway protocol (EGP), or an unknown protocol, usually a static route (incomplete).
AS path	Contains a list of all the ASs traversed by the NLRI.
Next hop	Indicates the IP address of the AS border router to use to reach the route advertised in the NLRI.
Multi exit discriminator (MED)	Specifies a preferred entry point for traffic coming back to the advertising AS. A lower MED value indicates a higher preference.
Local preference	Specifies a preferred route in an NLRI update between internal BGP peers. A higher value indicates a higher preference.
Atomic aggregate	Indicates that BGP-4 has performed some aggregation on the NLRI and removed previous path information.

(1 of 2)

BGP-4 path attributes (continued)

Path attribute	Description
Aggregator	Indicates the AS and BGP-4 peer that performed the last aggregation on a route. A BGP-4 peer that performs route aggregation adds the aggregator path attribute to the NLRI.
Communities	Associates one or more properties with the route.
Extended communities	Used in IP VPN auto discovery to carry the VPN ID and VPN peering topology value.
MP reach NRLI	Used in IP VPN auto discovery to carry public/private address mapping information.
Originator ID	Indicates the router ID of the BGP speaker that originated the route in the AS. This attribute is inserted by the route reflector.
Cluster list	Contains a list of the clusters traversed by the route. This attribute is inserted by the route reflector.

(2 of 2)

BGP-4 routing policies

A BGP-4 policy is a set of rules that determine which routes BGP-4 uses or advertises to other peers. BGP-4 uses import policies to filter NLRI updates received from other BGP-4 peers. BGP-4 export policies determine which NLRI updates to advertise to other BGP-4 peers.

Policies consist of keys and actions. A routing policy specifies an action for routing information received from BGP peers defined by a key.

For more information, see the following sections:

- BGP-4 import policy (page 113)
- BGP-4 export policy (page 115)

BGP-4 import policy

By applying import policies, BGP-4 allows certain routing information into the IP routing database or blocks it. Through import policies, you can configure BGP-4 to

- use or ignore an update by setting the *usageFlag* attribute
- set a route preference for IBGP and EBGP routes using the *ibgpRtePreference and ebgpRtePreference* attributes
- add a community number to the routing information using the *appendCommunity* attribute

- set a preference for updates to be used by the local BGP instance using the *localPreference* attribute

As routing updates arrive, BGP-4 applies the policy with the most specific key matches. When using path matching expressions, the *expressPreference* attribute acts as the tie-breaker between equally valid expressions.

By default, BGP-4 on Nortel Networks Multiservice Switch nodes reject all routes from external BGP (EBGP) peers and accepts all routes from internal BGP (IBGP) peers. See the table BGP-4 import keys (page 114) for a summary of BGP-4 import policy criteria.

BGP-4 import keys

Import key	Associated attribute	Description
Network prefix and length	*network*	Applies the policy action to NLRI updates that match, or are more specific than, the specified network.
Peer AS number	*peerAS*	Applies the policy action to updates from peers in a specified AS.
Peer IP address	*peerIpAddress*	Applies the policy action to updates from a peer specified by an IP address.
Originating AS number	*originAs*	Applies the policy action to NLRI originating from the specified AS.
Originating protocol	*originProtocol*	Applies the policy action to routes with the specified origin protocol.
AS path	*asPathExpression*	Applies the policy action to routes with an AS path that matches the specified expression.
Community path	*communityExpression*	Applies the policy action to routes with a communities attribute that matches the specified expression.

For example, a Multiservice Switch node configured with two BGP-4 import policies, x and y, receives a routing update originating from an EBGP peer along the AS path 1 2 3 4. The two policy keys are defined as follows in the table Example BGP-4 import policy key definition (page 115).

Example BGP-4 import policy key definition

Keys for import policy x	Keys for import policy y
Protocol = all	Protocol = all
AS path = 1 2 3	AS path = 2
expressPreference = 90	expressPreference = 100

The path matches both AS path expression keys. However, the expression for policy y has a greater preference value. BGP-4 applies the actions of policy y.

BGP-4 export policy

By applying export policies, BGP-4 distributes certain IP routing information to other BGP-4 peers. Through export policies, you can configure BGP-4 to

- send or block routing updates by setting the *advertiseStatus* attribute
- include the preferred entry point to the AS in updates to external peers by setting the *sendMultiExitDiscToEbgp* attribute. The MED value sent is the value specified in the *muliExitDisc* attribute.
- send your community number to BGP peers through the *sendCommunity* attribute
- alter the outgoing route's AS path by setting the *insertDummyAs* attribute
- set a preference for routes sent out to internal peers through the localPreference attribute

As BGP-4 sends out routing updates, it applies the policy with the most specific key matches. When using path-matching expressions, the *expressPreference* attribute acts as the tie breaker between equally valid expressions.

By using the export policies of other protocols, BGP-4 can distribute BGP-4-learned routes into other routing protocols such as OSPF, RIP, and EGP.

BGP-4 export keys

Export key	Associated attribute	Description
Peer AS number (see Note)	*peerAS*	Applies the policy action to routing updates of a specific protocol.
Peer IP address	*peerIpAddress*	Applies the policy action to updates destined for a peer specified by its IP address.
		(1 of 2)

BGP-4 export keys (continued)

Export key	Associated attribute	Description
Protocol	*protocol*	Applies the policy action to routes learned from peers defined by specific protocols.
Local RIP interface	*ripInterface*	Applies the policy action to routes learned from a local RIP interface (defined by an IP address).
Protocol-specific peers	*egpAs, bgpAs, ripNeighbor, ospfTag*	Applies the policy action to routes learned from peers defined by specific protocols.
Network prefix and length	*Network*	Applies the policy action to NLRI updates that match, or are more specific than, the specified network.
AS path	*matchAsPath*	Applies the policy action to routes with an AS path that matches the specified expression.
Community path	*matchCommunity*	Applies the policy action to routes with a communities attribute that matches the specified expression.
Attention: To configure an export policy for an internal BGP peer, specify its local AS number in the *peerAs* attribute; otherwise, the internal default export policy is preferred.		
(2 of 2)		

For example, a Nortel Networks Multiservice Switch node configured with two export policies, a and b, has a route with an AS path of 123, learned from an external peer. The two policy keys are defined as follows in the table Example BGP-4 export policy key definition (page 116):

Example BGP-4 export policy key definition

Keys for export policy a	Keys for export policy b
Protocol = all AS path = 1 2 3	Protocol = EBGP

BGP-4 would apply the actions of policy b. In this example, the protocol key matched in policy b is more specific than the AS path key of policy a, and therefore has greater weight.

BGP-4 route selection

This section describes the mechanisms that BGP-4 uses to handle routing information. It covers the following topics:

BGP-4 routing information bases (RIBs)

BGP-4 keeps track of all BGP-4 updates in a BGP-4 route database. When there are multiple routes to the same destination, BGP-4 selects the best route and places it in the IP forwarding table of the Nortel Networks Multiservice Switch node.

Each BGP-4 speaker maintains three sets of routing information bases (RIBs):

- The Indb RIB contains all NLRI updates received from IBGP and EBGP peers.
- The Localdb RIB contains routes that have been selected by the BGP-4 decision process and propagated from the Indb.
- The Outdb RIB contains all NLRI updates advertised by the BGP-4 instance to its IBGP and EBGP peers.

BGP-4 stores all incoming routes learned from BGP-4 peers in the Indb. The BGP-4 instance applies local import policies to all routes stored in the Indb, and filters out some of the routes. Only the routes that remain are potential candidates for the next stage of the BGP-4 process. For more information on import policy, see BGP-4 import policy (page 113).

BGP-4 then selects the most preferred route in the Indb for propagation to the Localdb. To select the best BGP-4 route, BGP-4 uses the tie breaking rules For more information, see Tie-breaking rules (page 117).

BGP-4 then applies its local export policies to the routes in the Localdb, and filters out some of the routes. BGP-4 stores the remaining routes in the Outdb, and advertises them to internal and external BGP-4 peers. For more information on export policy, see BGP-4 export policy (page 115).

Tie-breaking rules

If a route specifies an unreachable next hop, BGP-4 does not use it. However, if more than one route to a destination is available, BGP-4 uses the criteria in the table Tie-breaking rules (page 118) in the order shown to choose the best BGP-4 route.

Tie-breaking rules

Order	Rule
1	BGP-4 uses the route with the highest calculated local preference.
2	If the local preferences are the same for each route, BGP-4 uses the route with the lowest weight in the AS path or the shortest AS path.For more information, see AS weights (page 118).
3	If the weights and the AS paths are the same for each route, BGP-4 uses the route with the lowest multiExitDiscriminator attribute.
4	If the multiExitDiscriminator attributes are the same for each route, BGP-4 uses the lowest cost route to the next hop.
5	If the interior cost is the same for each route, BGP-4 uses the route learned from an external peer over the route from an internal peer.
6	If the criteria in rules 1 to 5 are true, BGP-4 uses the route with the lowest BGP-4 router ID. You set the BGP-4 router ID in the *bgpIdentifier* attribute of the *Bgp* component. The BGP-4 router ID must be unique in the BGP-4 network.

AS weights

You can set a preference for an AS and discriminate against other ASs by using AS weights. The AS weight attribute is local to the BGP-4 speaker, and is not propagated in route advertisements.

You can assign a weight to each AS in an AS path attribute. The weight is an integer value between 0 and 255 inclusive. BGP-4 prefers the path with the lowest weight. BGP-4 considers the value 255 as infinity and does not use that path. If you do not assign a weight to an AS, BGP-4 uses 128 as the default weight.

You can configure a BGP-4 instance to add a dummy AS to a path to decrease its preference by setting the *insertDummyAs* attribute under the *Bgp Export* component.

BGP-4 optimization

You can use optional subcomponents to optimize the BGP-4 configuration. For more information on these subcomponents and their functions, see the following sections:

- Route aggregation (page 119)
- Route reflection (page 119)
- BGP-4 communities (page 120)
- Private AS number removal (page 120)

- Dynamic default aggregation (DDA) mode (page 121)

Route aggregation

Through the aggregate policy, BGP-4 combines the characteristics of different routes and advertises this combination as a single route. Aggregation reduces the data a BGP-4 speaker stores and exchanges with another BGP-4 speaker. See RFC1771 for a complete list of rules for route aggregation.

Route reflection

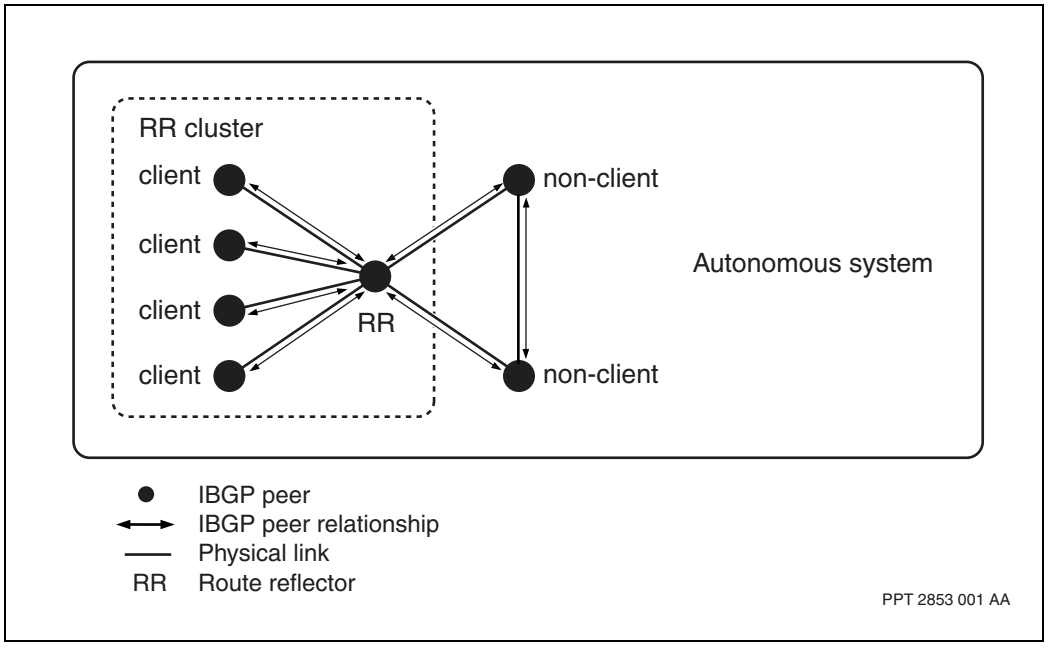
BGP requires that all the IBGP speakers be fully meshed. Route reflection is an IBGP peering mechanism that reduces the IBGP mesh. This implementation of route reflection is compliant with RFC1966.

When you configure an IBGP speaker as a route reflector, you make it the focal point for internal BGP sessions. Multiple BGP routers in an AS have a peer relationship with a route reflector as either clients or non-clients. See the figure BGP-4 route reflection (page 119). The route reflector and client peers form a cluster. Unlike client peers, non-client peers must be fully meshed with each other.

A route reflector advertises the best IBGP route based on the following rules:

- If the route is received from a non-client peer, advertise to client peers only
- If the route is received from a client peer, advertise to all non-client and client peers, except for the originator of the route

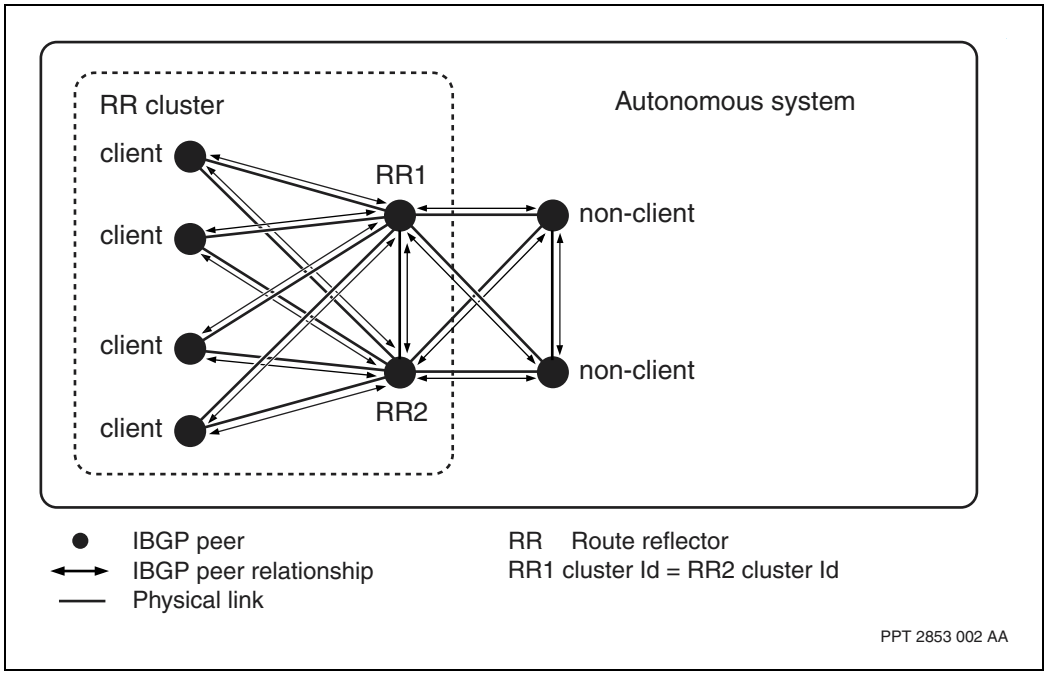
BGP-4 route reflection



BGP-4 allows redundant route reflectors in each cluster, and multiple clusters in an AS. See the figure BGP-4 route reflection with redundant reflector (page 120). The cluster identifier of both route reflectors is set to the same IP address using the *routeReflectorCluster* attribute.

It is recommended that you configure at most two route reflectors per cluster.

BGP-4 route reflection with redundant reflector



BGP-4 communities

A BGP-4 community is a group of destinations sharing some common property. RFC1997 defines the BGP communities attribute (note that in this instance, attribute refers to a configured decimal identifier of a community). Each system administrator may define which community a destination belongs to. The communities attribute adds another route filtering mechanism, giving users greater flexibility in defining import and export policies.

Private AS number removal

Every BGP-4 speaker belongs to an AS. Under normal circumstances, the AS number must be unique if routes from that AS are advertised to the Internet. However, as described in RFC1930, an enterprise customer that is homed to a single carrier need not be allocated a globally unique AS number, but may use a private AS number. When the carrier advertises its customer's routes to the Internet, the private AS number must be removed from the AS path attribute in the routing updates.

You can configure BGP-4 on Nortel Networks Multiservice Switch nodes to remove any AS numbers from the AS path attribute that are within the private AS range of 64 512 to 65 535. This applies to routes advertised through EBGP peering only. It does not apply to internal BGP peers.

Dynamic default aggregation (DDA) mode

The dynamic default aggregation (DDA) feature allows a Nortel Networks Multiservice Switch node to aggregate routes learned from an external BGP peer to a single default route. The node then propagates the DDA route to other nodes through IBGP peering.

For more information, see the following sections:

- DDA routes (page 121)
- DDA mode for Internet access (page 121)

DDA routes

You can activate an aggregate policy for incoming routes learned from a specific EBGP peer.

When the first valid route is learned from the EBGP peer, the policy generates a single default route. This DDA route is the only route considered valid among all those learned from the EBGP peer. If the import policy permits, BGP-4 propagates the DDA route to the routing database and then to the IP forwarding table. The route is tagged as a BGP external route whose next hop is the remote IP interface of the EGBP peer.

When BGP-4 generates the single DDA route, it sets the AS path attribute to the AS number of the EBGP peer and the next hop attribute to the IP address of the EBGP peer. In addition, it sets the multi-exit discriminator (MED) attribute to the value specified in the *defaultInAggMed* attribute (under the *Vr Ip Bgp Peer* component).

If you configure an OSPF export policy that specifies BGP external routes (by setting the *protocol* attribute under the *Vr Ip Ospf Export* component), BGP-4 advertises the DDA route into the OSPF domain. Therefore, there is a DDA route exported into OSPF from each node that runs DDA mode for a configured EBGP peer.

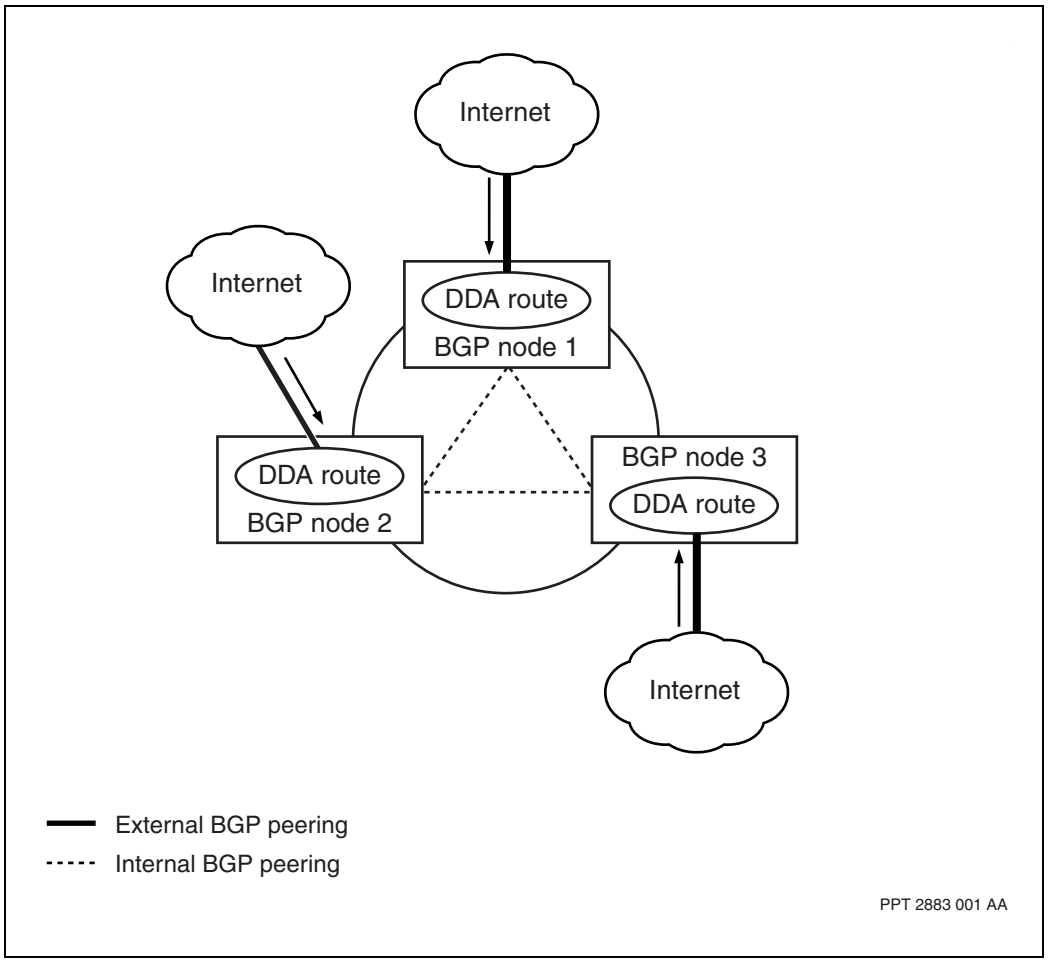
DDA mode for Internet access

The DDA feature allows for full connectivity to the Internet without the necessity of maintaining large forwarding tables. A Nortel Networks Multiservice Switch node that connects to the Internet can use DDA mode to aggregate all routes learned from its Internet EBGP peer. The node can then propagate the DDA route to other BGP nodes through IBGP peering.

For example, in the figure DDA mode on EBGP peers (page 122), BGP nodes 1, 2 and 3 connect to the Internet through EBGP peering, and to each other through IBGP peering. Through DDA, each node aggregates all incoming routes from the Internet into a single default route.

By default, this DDA route is advertised over each IBGP peer to other BGP nodes. Thus, each node learns one DDA route from its EBGP peering with the Internet and two DDAs from its IBGP peering with the other BGP nodes. The DDA route that is learned through EBGP peering is preferred, but the propagation of DDA routes from other IBGP peers ensures redundant access to the Internet without the risks associated with routing loops. The DDA route is withdrawn if the node loses its connectivity to its external peer.

DDA mode on EBGP peers



Static routes

This section describes Nortel Networks Multiservice Switch implementation of static routes.

Navigation

Overview of static routes

Static route definition allows Nortel Networks Multiservice Switch systems to specifically identify routes to remote IP networks or hosts. The definition includes a destination address, address mask, and one or more next hop addresses (gateways).

Equal-cost multipath routing

Static routes (and OSPF) support an equal-cost multipath routing function where equal-cost paths are used in a load-sharing mode. Equal-cost multipath routing has the following limitations:

- It allows a maximum of three next hop addresses.
- The entire route is discarded if, in a multiple-hop address, any next-hop IP address cannot be resolved through the address resolution protocol (ARP) process.

Static route definition

The *Static* subcomponent of the *Ip* component allows you to add, delete, and modify static route information. You can configure one static route instance on each VR.

The *RouteEntry* subcomponent of the *Static* subcomponent is used to define static routes by specifying one host, a subnetwork, or a network. There is one *RouteEntry* subcomponent for each static route. The *RouteEntry* subcomponent has one subcomponent: called *NextHop*. You must provision at least one *NextHop* component for each *RouteEntry* component.

Discard route entry

The *DiscardRouteEntry* subcomponent of the *Static* component is an optional subcomponent used to identify destination networks and nodes that do not receive packets through IP. The system discards packets addressed to these destinations immediately. No notification is sent to the sending host that Nortel Networks Multiservice Switch systems have discarded the packets. There is one *DiscardRouteEntry* subcomponent for each route that you wish to restrict. For provisioning information about discard route entry, see the procedure Configuring discard route entry for specific destinations in NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

Protected default route

Protected default route (PDR) is a mechanism for sparing at the IP / L3 layer in order to provide hitless (less than one second) IP forwarding support on the IP default route, for 4pGe line failures, card failures and software migrations. A PDR is defined as the IP default route provisioned with a set of 2 or more nextHops (each using a unique local IP interface) which can be used to forward the IP packet. The operational state of these nextHops / interfaces of the PDR is monitored, and the forwarding information is optimally managed to enable route reprogramming within one second in the case of an active nextHop / interface failure.

The PDR will be implemented only on:

- non-ECMP-load-balanced static default IP route (IP default route with multiple nextHops each with different metrics; such that only one nextHop is used for forwarding, at any one point in time)
- routes with nextHops over 4pGe ports.

Hitless (less than one second) support for PDR over the 4pGe FP is supported for the following scenarios:

- Ge port / link failure,
- Ge LAG group failure,
- 4pGe card failure,
- CP failure / switchover, and
- Hitless Software Migration (HSM).

Provisioning protected default route

Configure a PDR by provisioning a *Vr Ip Static RouteEntry*/0.0.0.0,0.0.0.0,0 component and setting a new provisionable attribute, *protected*, to yes. The default for the protected attribute is no, to preserve legacy behavior.

A protected *Vr Ip Static RouteEntry/0.0.0.0,0.0.0.0,0* must have at least 2 *NextHop* subcomponents using unique protocolPorts defined; and may have up to a maximum of 8 *NextHop* subcomponents. Note that provisioning additional *nextHops* to an existing PDR does not result in outage.

There can be up to 8 VR instances with protected *Vr Ip Static RouteEntry*/0.0.0.0,0.0.0.0,0 instances defined.

In order to support the outage requirement of less than one second during a HSM event, the Multiservice Switch 15000 node that has a VR with the PDR present must also have static address resolution protocol (ARP) entries provisioned against the PDR nextHops that are reachable over the migration active (MA) Ge FP links. Dynamic ARP entries must be used against the PDR nextHops that are reachable over the service active (SA) Ge FP links for this PDR VR.

Static routes provisioned on an adjacent router with nextHops for the Multiservice Switch 15000 VR hosting the PDR must have their nextHops set in such a way that the nextHops against the SA card are preferred (assigned a lower metric) over the nextHops against the MA card on the Multiservice Switch 15000 VR hosting the PDR.

For information on configuring PDR, see the VR static route configuration' chapter in NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*.

IP multicast

This section describes Nortel Networks Multiservice Switch implementation of IP multicast.

Navigation

Multiservice Switch nodes support IP multicast as defined in the following RFCs:

- *RFC2362, Protocol Independent Multicast - Sparse Mode (PIM-SM)*

Attention: PIM Border Router is currently not supported.

- *RFC2236, Internet Group Management Protocol (IGMP), version 2* router functionality

Multiservice Switch multicast

- supports IGMP version 2 router functionality.
- interworks with IGMP version 1 or version 2 hosts.

Attention: Multiservice Switch multicast does not support the IGMP version 1 router functionality, nor can it interwork with an IGMP version 1 router (that is, it cannot share a LAN segment).

Overview of IP multicast

IP multicast is receiver initiated, in that receivers may join a group at any time. Receivers use the Internet group management protocol (IGMP) to inform local routers about groups they are interested in receiving. Multicast routers use a multicast routing protocol to build trees to group members. Multicast packets are forwarded along these trees. Trees may be either rooted at a source or at a core node in the network. The IP multicast model is receiver oriented in that receivers control the 'building of tree' by joining groups. In order to send to a group, a host need not perform any routing operation. A host simply sources multicast packets on its interfaces, which are then forwarded by multicast routers to all group members.

Supported media

Nortel Networks Multiservice Switch nodes support multicast forwarding over the following media:

- Ethernet (100BaseT)
- ATM MPE
- FrDte

Dense and sparse mode protocols

IP multicast routing protocols generally fall into two categories, depending on the assumption of the distribution of the receiving hosts. A group is considered dense if there are many receivers within a region. A group is considered sparse if receivers are sparsely distributed across a larger area, or where a small group of receivers is concentrated in one portion of the network. DVMRP, PIM-DM, MOSPF are all dense mode protocols. BGMP, CBT and PIM-SM are sparse mode protocols.

Dense mode protocols assume dense membership within a region, and either flood group membership information (MOSPF), or multicast packets (DVMRP, PIM-DM) using the flood and prune model. Sparse mode protocols assume sparse membership over a larger region, and do not flood group information or data packets. Sparse mode protocols use an explicit joining mechanism.

Attention: Multiservice Switch nodes only support PIM-SM.

Source specific and shared trees

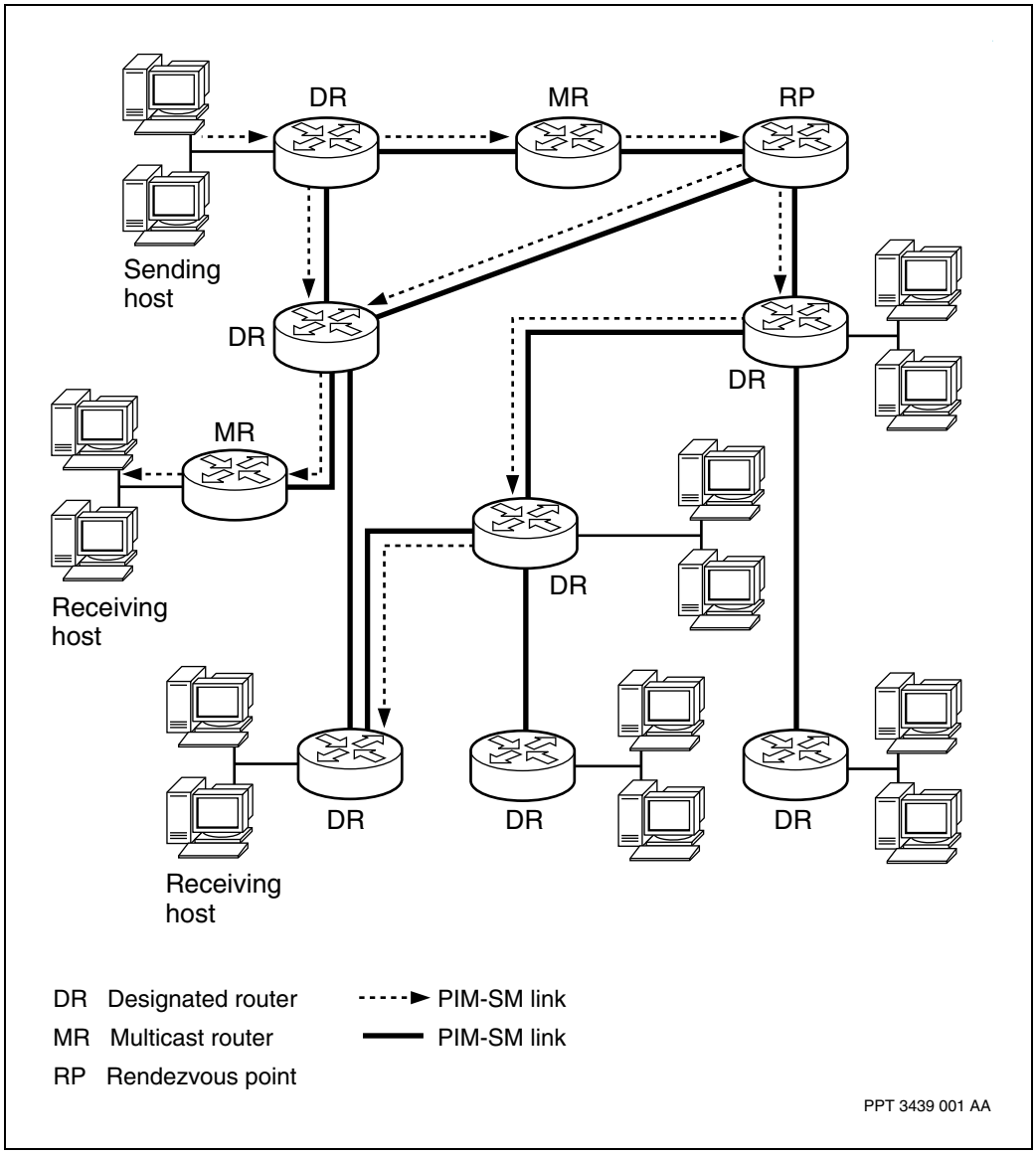
Multicast routing protocols build spanning trees to all group members for forwarding of multicast data packets. These spanning trees may be either source-based trees (DVMRP, PIM-DM, MOSPF) or shared trees (CBT, BGMP, PIM-SM).

PIM-SM can build both source-specific and shared distribution trees. Shared trees are used by default, but when a particular data threshold is reached, PIM-SM may change to source-specific trees.

Source-based trees are rooted at a source and span to all group members. If N-N communication is desired (for example: conferencing) when a source-specific tree routing protocol is in use, then multiple source-specific trees comprise a given group. These trees have the ability to offer lowest delay, but with the cost of extra state. Source-specific trees require keeping (S,G) state, or state that is per source per group and thus scales as the number of sources per group. Source-specific trees may also be based on source prefixes, (Sp,G), and represent all the sources summarized by the prefix. Source-specific trees are almost always used in dense mode routing protocols and are inherently uni-directional.

Shared trees are used in sparse mode routing protocols. They require the definition of a root node, which acts as a sink for all source data and group joins. The root node, also known as the core node or rendezvous point (RP), is used as a meeting point between sources and receivers. When a source transmits, its data is sent towards the root node; when a host joins, its join is propagated towards the root node. Shared trees do not guarantee the lowest delay, but can offer state savings within routers.

IP multicast - RP and shared tree for a multicast group



IGMP

The Internet group management protocol (IGMP) [DEERING] is the protocol used by hosts to communicate their desired group memberships to their local multicast router. IGMP exists as part of a host's IP implementation, as well as part of a router's multicast implementation. IGMP uses a query / response mechanism to solicit membership status from hosts. Periodically, routers send a query message, to which hosts respond with a report per group desired. When a router receives a report from a host, the action it takes depends upon the multicast routing protocol in use.

IGMPv1 [DEERING] was the first implemented version of IGMP, and exists in many host (and router) implementations. IGMPv2 [FENNER1] added support for a fast leave message so that hosts can inform the network immediately when they wish to leave a group. This message improves the multicast “leave latency” (with IGMPv1, routers simply time out the state for a group). IGMPv3 [CAIN1] enhancements support source-specific join and leave messages. This allows hosts to individually join or leave sources or sets of sources.

PIM-SM

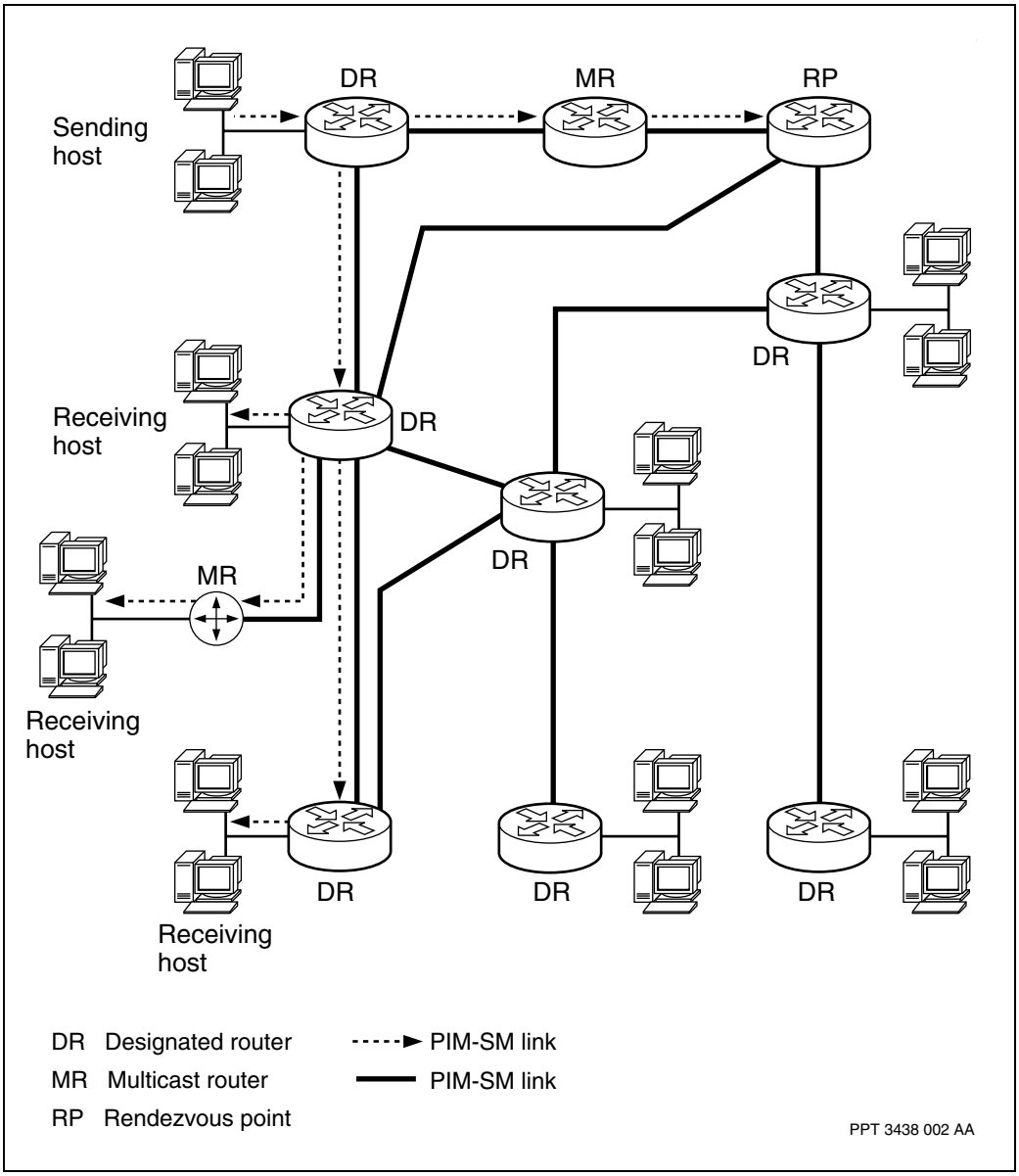
Protocol independent multicast sparse mode (PIM-SM) is a multicast routing protocol designed for use in networks where receivers make up a small portion of the overall network topology. Under PIM-SM, multicast traffic is distributed to only those routers with group members downstream. PIM-SM builds soft-state, multicast-group based, shared trees for forwarding multicast traffic. PIM-SM employs the use of a rendezvous point (RP), where receivers and sources meet. The RP is the node toward which the join/prune and register messages are sent for the shared tree. Each group has one specified RP. Routers with downstream receivers wishing to join a group must explicitly inform the RP of the join. This in turn requires RPs to maintain state information pertaining to group membership.

Routers running PIM-SM respond to changing group membership by issuing join or prune messages towards an RP. Spanning trees created under the PIM-SM model are defined by the actions of joining and pruning.

PIM-SM has the flexibility to operate using a shared tree or a shortest-path tree (SPT). Use of the SPT can only be initiated by an RP, or by a PIM router with locally attached hosts. A Nortel Networks Multiservice Switch node acting as an RP will initiate the SPT by default, on receipt of the first packet from a source.

A Multiservice Switch node acting as a last-hop DR can be configured to join the SPT on receipt of a first packet from a source. Multiservice Switch nodes do not support the use of traffic thresholds to trigger SPT joins.

IP multicast - RP and shortest path tree for a multicast group



Multicast domains

Nortel Networks Multiservice Switch nodes support the use of up to four independent multicast domains per virtual router. Domains are used to partition a network into sections where multicast operates independently under the control of the same or different multicast routing protocols. Forwarding of multicast traffic between domains requires the use of a multicast border router which supports the multicast routing protocols used in the neighboring domains. Multiservice Switch nodes do not currently support the multicast border router functionality.

Virtual router redundancy protocol

This section describes Nortel Networks Multiservice Switch 7400/15000/20000 node implementation of the virtual router redundancy protocol (VRRP).

Navigation

Overview of VRRP

Nortel Networks Multiservice Switch 7400/15000/20000 nodes use VRRP version 2, to provide router redundancy and availability to IP routing. Router redundancy is achieved with VRRP virtual routers (VRs). Router availability by monitoring critical IP interfaces is available on the 2-port 100BaseT Ethernet FP, 4-port 10/100BaseT Ethernet FP, 4-port gigabit Ethernet FP, and 8-port 10/100BaseT Ethernet FP only. RFC2338 describes VRRP in detail. Nortel Networks MSS 7400/15000/20000 implementation of VRRP supports:

- IP over Ethernet with the 2-port 100BaseT Ethernet FP
- IP over Ethernet with the 4-port 10/100BaseT Ethernet FP
- IP over Ethernet with the 4-port gigabit Ethernet FP
- IP over Ethernet with the 8-port 10/100BaseT Ethernet FP
- multiple instances of VRRP virtual routers on each node's VR

Attention: The term 'virtual router' and the VRRP protocol describe different entities. To minimize the confusion, the terms VR and VRRP VR are used to differentiate between the two.

VRRP virtual routers

Implementing VRRP involves creating a VRRP virtual router (VR) made up of two or more routers in the same subnet sharing IP addresses and a virtual MAC address. Within the VRRP virtual router, one router (for example, a Nortel Networks Multiservice Switch node's VR) will act as the master and the others as backups. To an end-host, this VRRP virtual router appears as a single router. VRRP virtual router (page 134) depicts this arrangement. The VRRP routers communicate with each other using IP multicasts through the local Ethernet interfaces (Multiservice Switch node VR LAN protocol port).

VRRP VRs can communicate using the local LAN media. The protected VRs configuration does not have an impact on the VRRP functionality. A VRRP VR can consist of Multiservice Switch VRs configured as either virtual IP routers (VIPR with individual WAN connectivity per VR) or as RFC2764-based IP-VPN routers (customer VRs aggregated through a virtual connection gateway). Alternatively, a VRRP VR can consist of Multiservice Switch VRs and non-Multiservice Switch routers compliant to RFC2338.

Typically, VRs are on different Multiservice Switch nodes and are connected over an Ethernet LAN segment. Alternatively, any of the following configurations can be used:

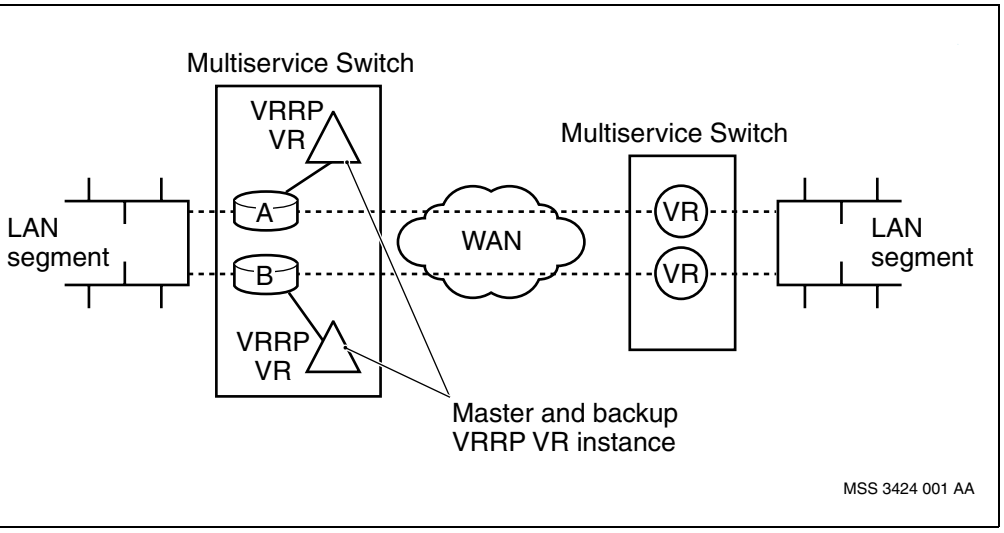
- A single Multiservice Switch that has two VRs with the same IP address reachability. Each VR has a VRRP VR instance. One VRRP VR instance is the master, and the other is the backup. Each VRRP VR instance is on a different FP.
- A single Multiservice Switch that has two VRs with the same IP address reachability. Each VR has a VRRP VR instance. One VRRP VR instance is the master, and the other is the backup. Each VRRP VR instance is on a different port on the same FP.

Also, VRRP on one Multiservice Switch VR can interwork with an external router implementing RFC 2338-compliant VRRP on the same Ethernet LAN/VLAN segment.

Attention: The IP VPN RFC2764 solution is only available on the 2-port 100BaseT Ethernet FP.

Each VRRP VR has a priority value that determines if it will act as a master or backup. The VRRP master router typically owns the IP addresses of the VRRP VR and has a priority of 255. If none of the VRRP VRs own an IP address, the VRRP VR with the higher priority is the master. In the case of equal priority, the higher interface IP address is the master.

VRRP virtual router



Router redundancy

You can configure a single Nortel Networks Multiservice Switch VR with multiple VRRP VRs. This allows one Multiservice Switch VR to participate in more than one VRRP VR. How you configure router redundancy depends on the unique characteristics of your network. The figure, Example router redundancy topologies (page 135), depicts three possible scenarios where a LAN segment uses multiple routers for load balancing and static routes to end hosts.

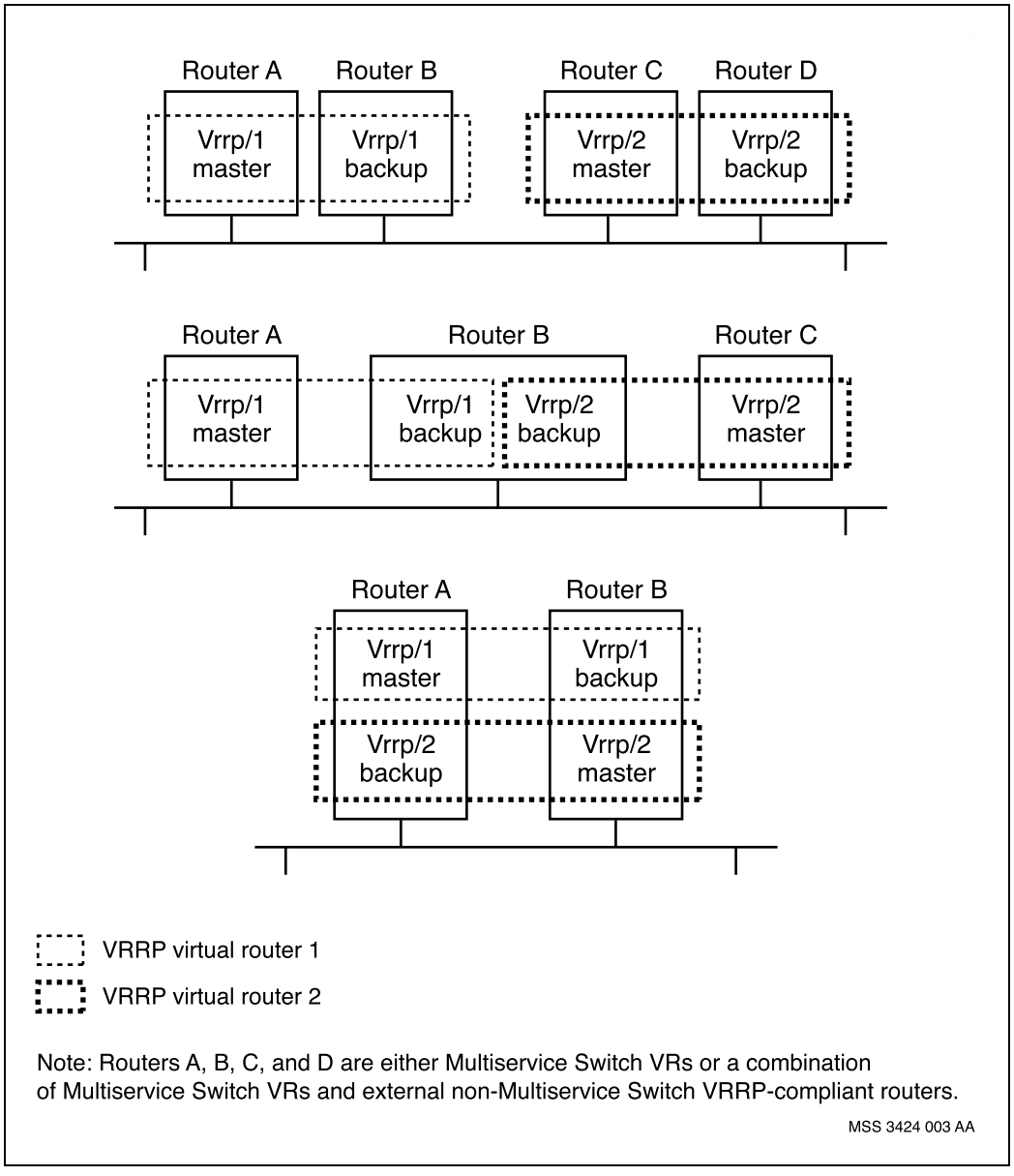
Attention: All three examples are available on the 2-port 100BaseT Ethernet FP. Only the first example is available on the 4-port 10/100BaseT Ethernet FP, 4-port gigabit Ethernet FP, and 8-port 10/100BaseT Ethernet FP. See Example router redundancy topologies (page 135).

VRRP provides redundancy on the Ethernet interface in both port-mode and VLAN-mode. When a protocol port with VRRP is associated with an Ethernet interface that is operating in port-mode, VRRP redundancy functionality behavior is unchanged. When a protocol port with VRRP is associated with an Ethernet interface that is operating in VLAN-mode, VRRP functionality behaves the same as port-mode, but only for that specific VLAN. A VLAN that is linked to a protocol port without VRRP configured is not redundant.

Attention: On the 4-port 10/100BaseT Ethernet, 4-port gigabit Ethernet, and 8-port 10/100BaseT Ethernet FPs, only one instance of VRRP per protocol port is supported. Interior gateway protocols, other than static protocols, on the same protocol port as VRRP must be in passive mode. Also, load balancing is not supported on these FPs.

For information about VRRP router redundancy with VIPR, see Router redundancy with VIPR (page 135).

Example router redundancy topologies



Note: Routers A, B, C, and D are either Multiservice Switch VRs or a combination of Multiservice Switch VRs and external non-Multiservice Switch VRRP-compliant routers.

MSS 3424 003 AA

Router redundancy with VIPR

Nortel Networks Multiservice Switch 7400/15000/20000 nodes providing VIPR solution with VRRP on the same LAN/VLAN segment are designated as master VRRP VR and backup VR. The master VRRP VR enables its virtual MAC (VMAC) on the Ethernet interface and the backup VRRP VR disables the same VMAC on its Ethernet interface. The master VRRP VR will then route traffic from the LAN/VLAN that is destined to the VMAC. After receiving an

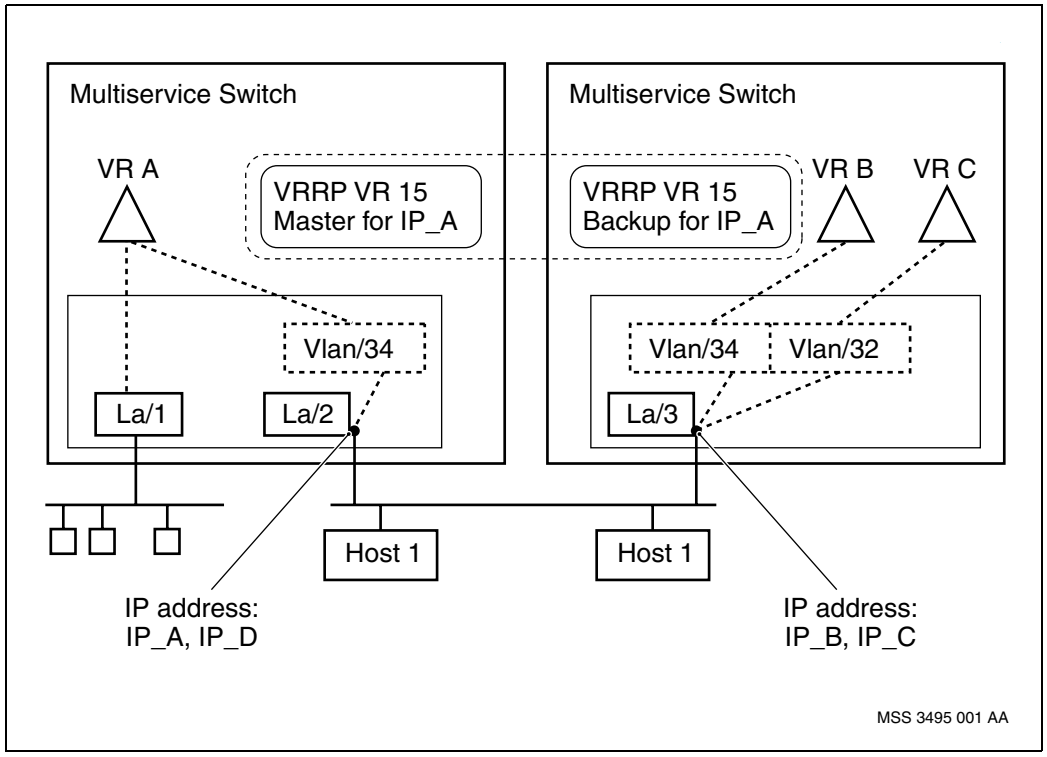
ARP message from the master VRRP VR, the hosts start sending traffic destined to the configured default route with an Ethernet header containing a MAC DA set to the VMAC of VRRP. At intermittent intervals, the master transmits a heartbeat control packet that is multicast onto the LAN/VLAN segment. The backup VRRP VR expects to receive that heartbeat message within a configured time interval, after which it assumes the master VRRP VR is no longer providing service on the LAN/VLAN segment.

If the master VRRP VR fails, the backup VRRP VR assumes the master role and enables the VMAC on its Ethernet interface. The new master (originally the backup) VRRP VR would then start routing traffic from the hosts that are sending traffic to MAC DA equal to the VRRP VR VMAC.

If the original master VRRP VR recovers, it re-establishes its master role by sending out heartbeat messages at configured advertisement intervals. When the active master VRRP VR receives the heartbeat, it reverts to the backup role again. The active master VRRP VR relinquishes its master role because the heartbeat message from the original master has a higher priority.

The figure, VRRP configuration to provide Multiservice Switch VIPR VR redundancy with VLANs (page 137), depicts an example of two Multiservice Switch nodes providing the VIPR solution with VRRP redundancy. In this instance, VR A and VR B are on the same VLAN segment, identified by VID=34. Both VRs are backed up by VRRP VR 15. The VRRP VR instance providing redundancy to VR A is elected master, while the VRRP VR instance providing redundancy to VR B assumes the backup role. The master enables its VMAC on the Ethernet interface and the backup disables the same VMAC on its Ethernet interface. VIPR VR A routes traffic from La/2 Vlan/34. Hosts 1 and 2 receive an ARP from the master VRRP VR, associating IP address, IP_A, with its VMAC as the MAC DA. The hosts send traffic to the statically configured default route with an Ethernet header containing a MAC DA set to the VMAC of VRRP VR 15.

VRRP configuration to provide Multiservice Switch VIPR VR redundancy with VLANs



Router availability

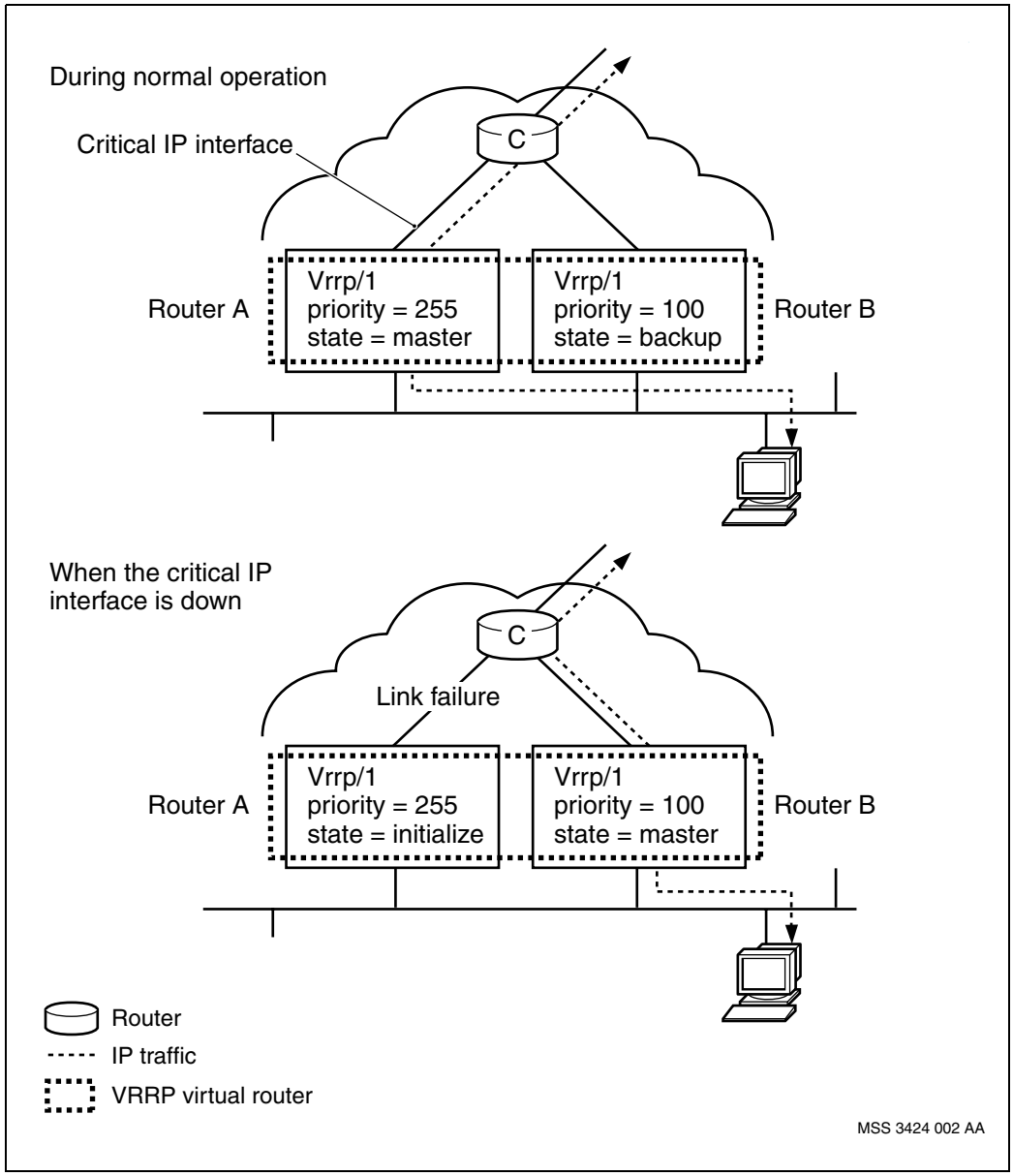
Specific IP interfaces can be monitored by a VRRP router. These interfaces are called critical IP interfaces. A VRRP router can be operational but have a key interface that is down, resulting in the loss of an IP traffic flow. By configuring VRRP with critical IP interfaces, you can better guarantee router availability for the IP traffic.

When a critical interface goes down or becomes locked, the current master goes into an initialize state. One of the VRRP backup routers, routing via a different interface, becomes the new master until the interface comes back up or is unlocked.

Critical IP interfaces can be passing IP traffic over media other than Ethernet. “Critical IP interfaces” (page 138) depicts a scenario where VRRP and a critical IP interface is used to provide router availability.

Attention: Critical IP interfaces are not supported when VRRP is configured for the RFC2547 IP VPN solution.

Critical IP interfaces



The VRRP process

When operational, VRRP VRs are in one of three states: master, backup, or initialize. Virtual routers (VRs) in the master state perform the routing duties for addresses associated with the VRRP virtual router. Virtual routers (VRs) in the backup state monitor the availability of the master router. The priority parameter of a VRRP VR determines if it acts as a master or backup. A VR is in the initialize state when its *Vrrp* component is locked, or when its IP

interface or linked critical IP interface is down. Table Summary of the VRRP virtual router states in relation to network conditions (page 139) summarizes the states of the VRRP VRs under different conditions.

Summary of the VRRP virtual router states in relation to network conditions

network condition	VRRP virtual router state and activities	
	VRRP virtual router A priority = 255	**VRRP virtual router B priority = 100**
start up	master	backup
normal	master As master, VRRP VR A multicasts messages at a configured advertisement interval, advertising to the backup VRRP VR or VRRP VRs that it is operational.	backup As backup, VRRP VR B listens for the multicast advertisement. When it receives the advertisement message with a higher priority, an advertisement wait timer resets.
VRRP VR A failure	na	master VRRP VR B transitions to master when the advertisement wait timer expires.
critical IP interface goes down	initialize Advertises to backup VRRP VR to take over as master VRRP VR.	master
critical IP interface comes back up	master Advertises to backup VRRP VR that it is resuming the role as master VRRP VR.	backup Transitions to backup state upon receiving startup multicast from VRRP VR A.
operator locks VRRP VR or IP interface is down	initialize Advertises to backup to take over as master VRRP VR.	master Transitions to master state upon receiving advertisement from VRRP VR A that it relinquished its role as master VRRP VR, or after the advertisement wait timer expires.

IP tunnels

This section describes Nortel Networks Multiservice Switch implementation of IP tunnels.

Navigation

Overview of IP tunnels

The IP tunnels feature enables you to connect two physically separate networks that share the same address space through an IP network with a different address space by encapsulating the original packet in an IP header. This outer IP header contains the routing information to traverse, or tunnel through, a network with a different address space. Nortel Networks Multiservice Switch nodes implement IP tunnel functionality through the *Tunnel* component and *StaticEndPoint* subcomponent.

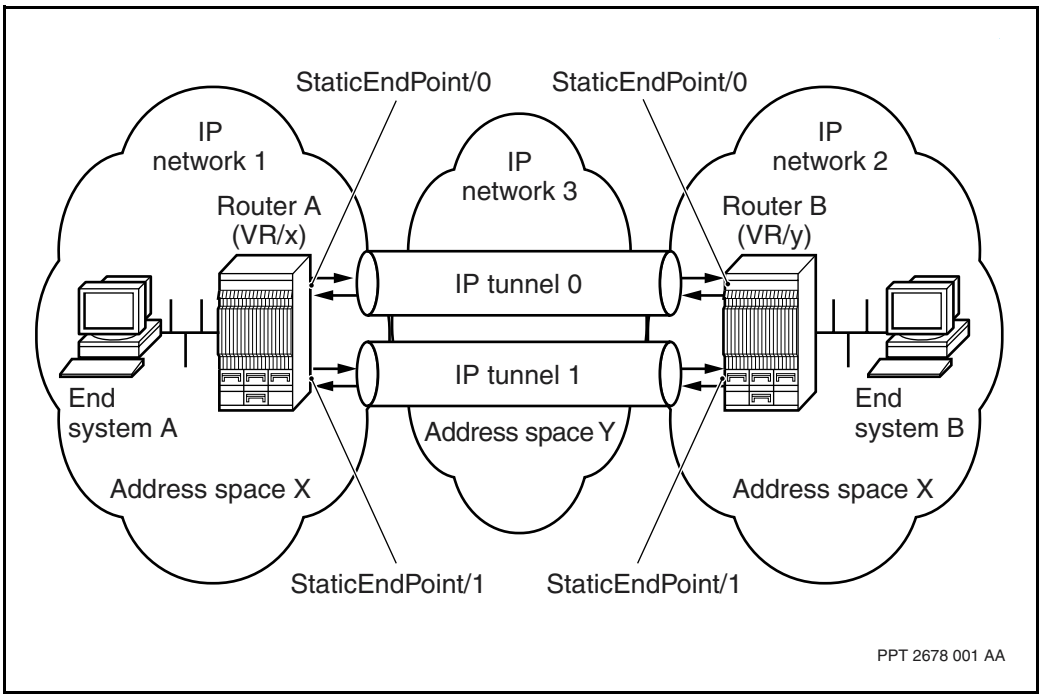
Multiservice Switch nodes support point-to-point (PTP) tunnels and point-to-multipoint (PTMP) tunnels.

The IP tunneling feature supports static point-to-point tunnels. IP tunnel example (page 141) illustrates the following characteristics of static point-to point tunnels:

- You identify tunnel end points using a *StaticEndPoint* component.
- You identify the source address of a tunnel using the *sourceAddress* attribute under the *StaticEndPoint* component. You identify the destination address of a tunnel using the *destinationAddress* attribute under the *StaticEndPoint* component. The source and destination addresses are what define a tunnel instance.You must manually provision source and destination addresses on each node in the shared address space.

- The tunnels themselves are subnets of the address space of the network being tunneled through which, in the figure IP tunnel example (page 141), is IP Network 3.

IP tunnel example



Encapsulation techniques

IP tunneling, regardless of the protocol of the originating network, is made possible by adding an outer IP header to the original packet. This outer IP header contains the routing information to traverse, or tunnel through, a network with a different address space.

You can use two encapsulation methods for point-to-point or point-to-multipoint tunnel traffic originating in IP networks, but the method must be the same at both ends of the tunnel:

- IP in IP encapsulation as defined in RFC 2003. For more information see IP in IP encapsulation (page 141).
- Generic Routing Encapsulation (GRE) over IP as defined in RFC1702. For more information see Generic routing encapsulation (GRE) (page 142).

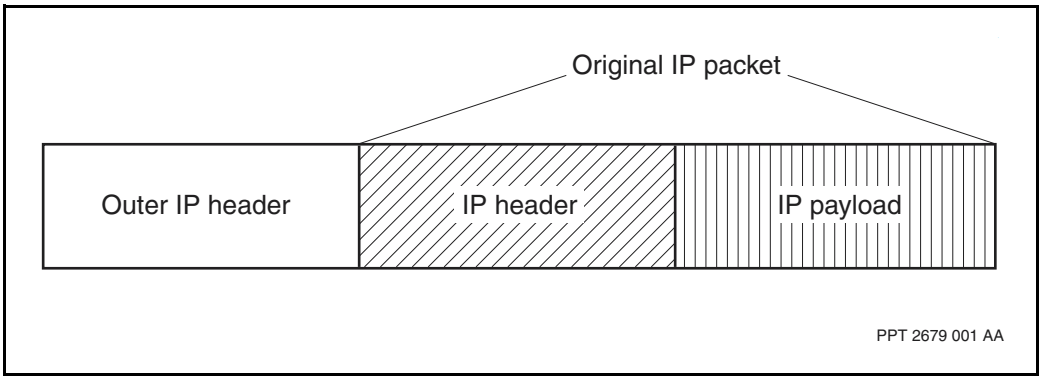
IP in IP encapsulation

IP in IP is the encapsulation method for tunnel traffic originating in IP networks. In this method, the Nortel Networks Multiservice Switch node in the originating IP network adds a new (outer) IP header, which contains the routing information to tunnel through the network with a different address space, to the original IP packet. The Multiservice Switch *Vr* component adds

this header before entering the tunnel. The *Vr* component in the IP network at the far end of the tunnel strips off the outer IP header as the packet exits the tunnel. The original packet is then forwarded to its destination as usual. IP in IP encapsulation format (page 142) shows the format of an IP packet encapsulated in IP.

Attention: IP in IP tunnel encapsulation is not compatible with RFC2003 when using an ATM IP FP as a backbone FP.

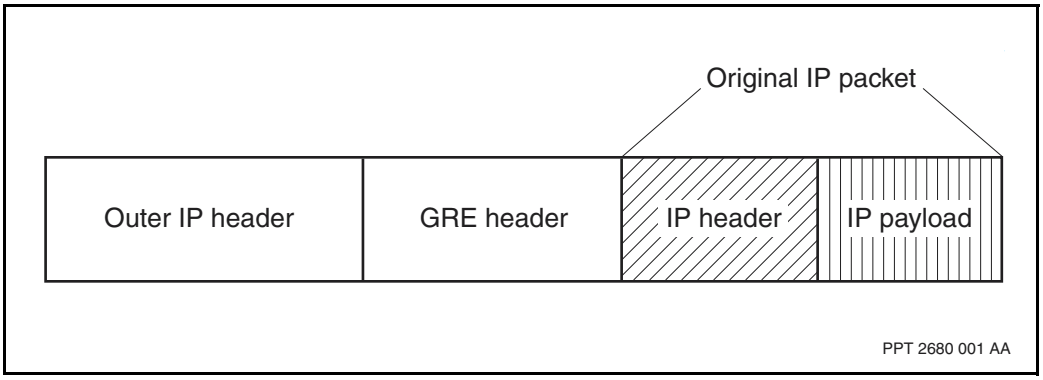
IP in IP encapsulation format



Generic routing encapsulation (GRE)

GRE is the other encapsulation method for tunnel traffic originating in IP networks. The Nortel Networks Multiservice Switch *Vr* component in the originating IP network adds a GRE header and then an outer IP header to the original IP packet before it enters the tunnel. The network being tunneled through treats the encapsulated packet as a regular IP packet while it is in transmission through the tunnel. The *Vr* component in the IP network at the far end of the tunnel strips off the outer IP header and the GRE header as the packet exits the tunnel. The original packet is then forwarded to its destination as usual.

GRE over IP encapsulation format for IP

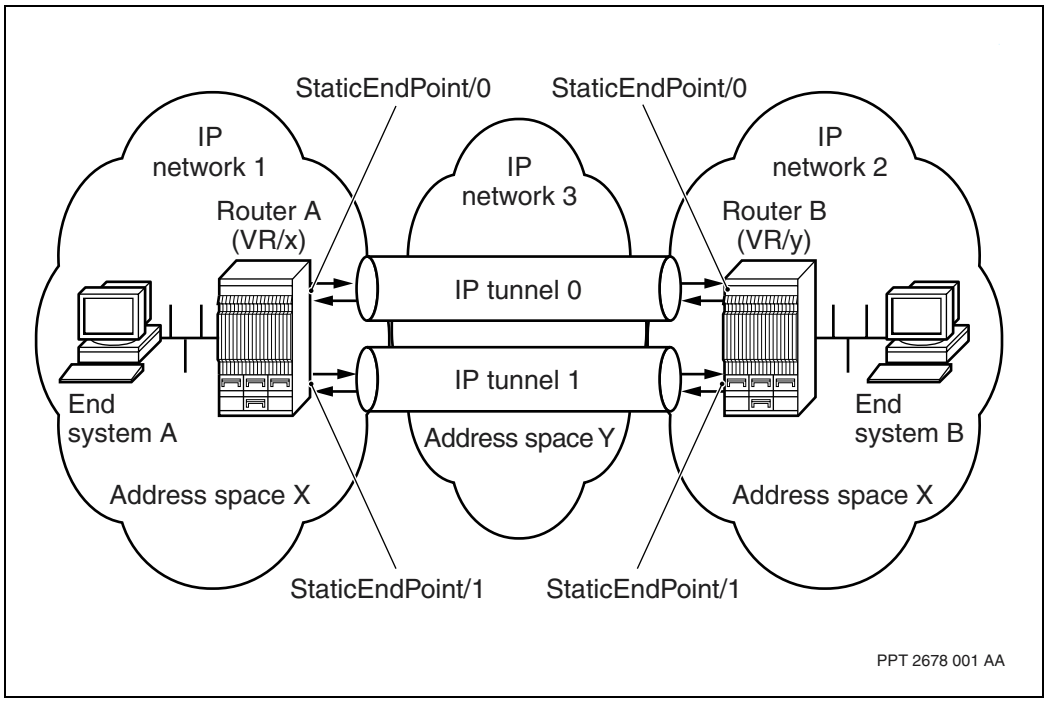


Point-to-point tunnels

Nortel Networks Multiservice Switch IP service uses point-to-point (PTP) IP tunnels to provide connectivity between customer VRs that reside on different nodes.

The figure Point-to-point IP tunnels (page 143) provides an example of two IP networks sharing address space X connected through two IP tunnels through a third IP network on address space Y.

Point-to-point IP tunnels



You define a tunnel instance by its source and destination addresses. The source address at one end of the tunnel must be the same value as the destination address at the other end of the tunnel. In addition, the source address at one end and the destination address at the other end must belong to the same address space. The tunnels themselves are subnets of the address space of the network being tunneled through which, in the figure Point-to-point IP tunnels (page 143), is IP Network 3.

You must manually configure source and destination addresses on each Multiservice Switch node in the shared address space. For example, in the figure Point-to-point IP tunnels (page 143) an operator must configure the source and destination addresses on the node in IP network 1, and another operator must configure the same in IP network 2.

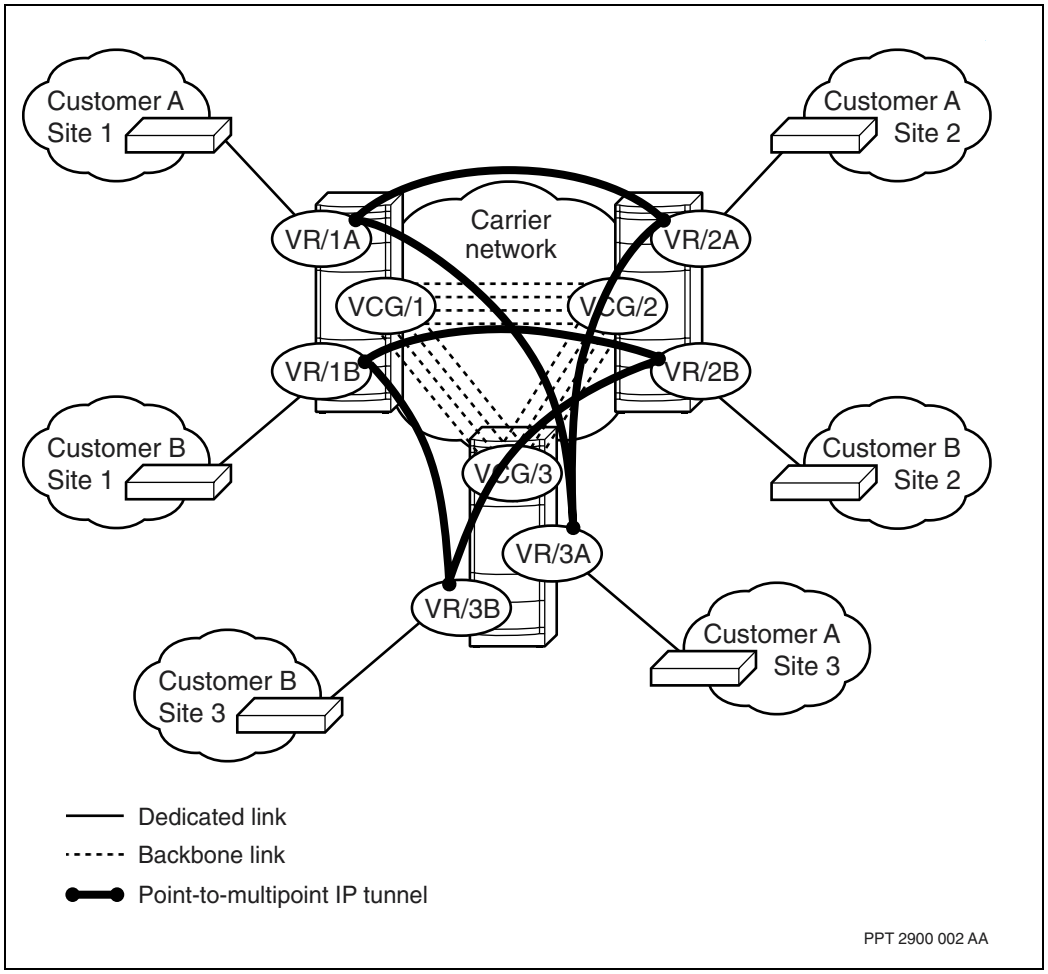
For IP VPN accounting, IP layer 3 usage measurements are collected for VRs connected by PTP tunnels. With IP accounting enabled, tunnel encapsulation and decapsulation counts are collected for PTP tunnel configurations on ATM functional processors. These statistics are categorized by CoS.

For information on configuring PTP IP tunnels, see NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*. For information on IP accounting, see IP accounting (page 148) and NN10600-560 *Nortel Networks Multiservice Switch 7400/15000/20000 Accounting*.

Point-to-multipoint tunnels

Nortel Networks Multiservice Switch IP virtual private network (VPN) service uses point-to-multipoint (PTMP) IP tunnels through a virtual connection gateway (VCG) to provide connectivity between customer VRs that reside on different nodes. An IP VPN consists of multiple customer VRs, each representing a private customer VPN site. VCGs on different Multiservice Switch nodes connect to each other through backbone logical connections. See the figure VCG-based IP VPN with point-to-multipoint IP tunnels (page 145).

VCG-based IP VPN with point-to-multipoint IP tunnels



For full site-to-site connectivity, the carrier must configure the source and multiple destination addresses of the PMTP tunnel on every customer VR in the IP VPN. The customer VR performs IP in IP encapsulation (as defined in RFC2003) at the ingress. The VCG performs decapsulation at the egress. With IP VPN accounting enabled, the IP tunnel encapsulation and decapsulation counts are collected for PTMP tunnel configurations on ATM IP functional processors. These accounting statistics are further broken by CoS.

Attention: IP in IP tunnel encapsulation is not compatible with RFC2003 when using an ATM IP FP as a backbone FP.

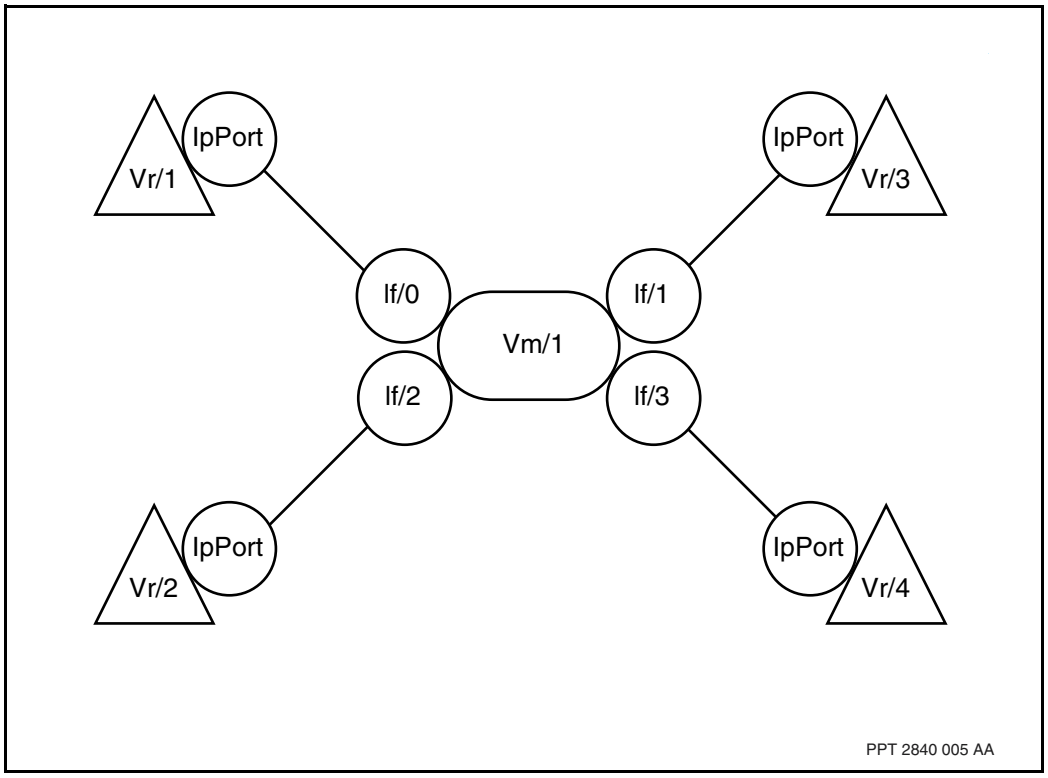
For more information about PTMP IP tunnels, IP VPN, and VCGs, see NN10600-581 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Technology Fundamentals* and NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management*.

Nortel Networks Multiservice Switch virtual media

You can configure connectivity between some of your VRs on a Nortel Networks Multiservice Switch node. By default, VRs on a node are completely isolated from one another, for security purposes. You can create connectivity through the use of a hairpin connector to physically link different VRs ports. This method, however, is wasteful of physical resources.

Instead, you can emulate a physical connection between VRs by configuring IP-only connectivity in the software using the *Interface* (*If*) subcomponent of the *VirtualMedia* (*Vm*) component. The *Vm If* component provides virtual (as opposed to physical) next-hop functionality between VRs. You can enable connectivity between different VRs by linking them through an IP port to different instances of the *If* subcomponent under the same *Vm* component. VR connectivity through a software link (page 147) illustrates the relationship between the *Vm* component and the VRs.

VR connectivity through a software link



You can add a *Vm* component if you want to provision an always-up IP interface for applications such as open shortest path first (OSPF) protocol, routing information protocol (RIP), and border gateway protocol (BGP). A virtual media application is not associated with a physical port. Since logical IP interfaces under the virtual media application are defined independently of any physical media, they remain up even though individual links to the node might lose connectivity. An IP address associated with the virtual media protocol is always reachable as long as the node itself remains connected to the network.

Virtual media on Multiservice Switch nodes support the following routing and forwarding functions:

- IP ARP and IP ARP Reply
- IP datagram forwarding through ARP and static route definitions
- OSPF
- RIP
- BGP-4

IP accounting

IP accounting allows you to collect, record and report usage measurements for each customer virtual router (VR) that is part of an IP virtual private network (VPN). The IP accounting statistics provide a breakdown of the volume of IP packets sent and received by VPN customers. IP accounting statistics are collected for three VPN configurations:

- point-to-point (PTP) tunnels
- point-to-multipoint (PTMP) tunnels
- layer 2 data connections (virtual circuits)

Accounting records are generated for each VR within these VPN configurations. VPN site-to-site information is generated for VRs within a VPN connected through IP tunnels. Local site information is available for VRs connected to the network through direct data link connections. The breakdown of the IP traffic based on tunnel usage and aggregate packet counts gives the carrier the ability to analyze the IP packets sent and received by VPN customers.

Layer 3 usage statistics are generated for VRs connected by IP tunnels. IP tunnel encapsulation and decapsulation counts are collected for PTP and PMPT tunnel configurations. If IP class of service (CoS) is applied to the customer traffic, the statistics are categorized into four possible IP traffic classifications.The CoS breakdown is generated at tunnel entry. Source and destination counts are provided at tunnel exit.

Attention: Tunnel ingress statistics are recorded as outbound statistics for the protocol port at which the tunnel originates. Similarly, tunnel egress statistics are recorded as inbound statistics for the protocol port at which the tunnel terminates.

Network statistics are provided for layer 2 connections. These statistics are collected for AtmMpe, IP-optimized DLCI, FrDte, and PPP media. Since the VPN site address is unknown, the accounting records gathered at the

outgoing traffic ports are aggregate statistics of the number of packets received and sent by the VR to the network. These statistics are broken down by CoS.

For more information, see NN10600-560 *Nortel Networks Multiservice Switch 7400/15000/20000 Accounting*.

For more information on IP accounting, see the following sections:

The CoS breakdown is generated at tunnel ingress and tunnel egress. There is information on tunnel source address, tunnel destination address, and packet counts per CoS.

Navigation

- IP accounting fundamentals (page 149)
- Collecting records (page 150)
- Troubleshooting IP accounting (page 151)

IP accounting fundamentals

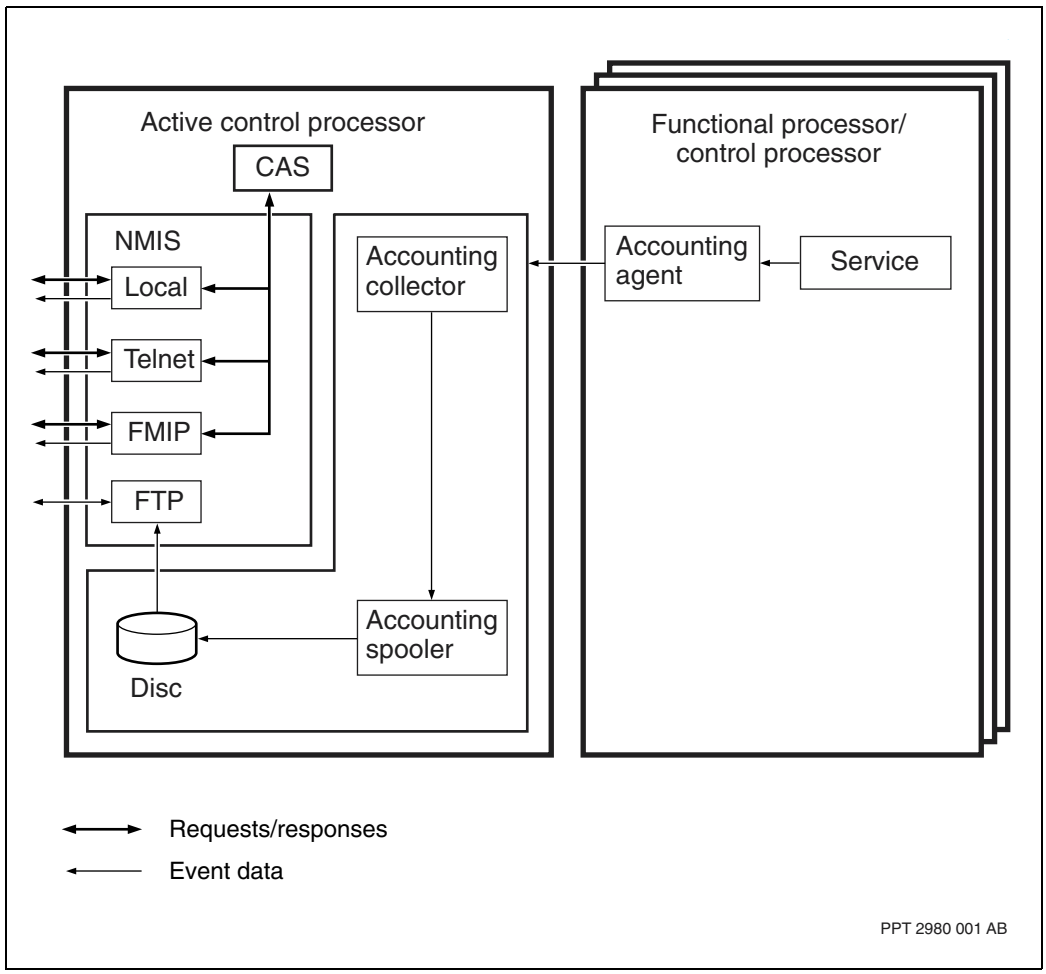
Nortel Networks Multiservice Switch systems manages the IP accounting information by controlling, recording, and reporting the usage statistics. The IP accounting system provides the statistical data to the nodal and network management system through the interactions between Multiservice Data Manager workstations and the disk on the control processor (CP).

The IP accounting information is managed by an accounting controller, located on the CP card in the Multiservice Switch node. The accounting controller works with Multiservice Data Manager to:

- start and stop the collection of usage statistics
- identify the usage statistics collected
- create and format the accounting records

The figure Accounting structure (page 150) illustrates the interactions of the subsystems on the node and the path followed to collect the accounting data at the protocol port to the output of the accounting record at the management data provider (MDP).

Accounting structure



Collecting records

The accounting record presents the usage statistics collected over a specified time. A record is generated for each protocol port, for each source address, and for each destination address for PTP and PTMP tunnels.

An IP accounting record contains the following information:

- accounting collection reasons
- static data such as VR name, source and destination address, virtual private network identifier (VPN ID), and protocol port identity
- usage data such as collection time and duration, and packet and byte counts

For further information on accounting collection reasons see attribute *Vr Ip accountCollection* in NN10600-060 *Nortel Networks Multiservice Switch 7400/15000/20000 Component Reference*.

An accounting record is generated for each protocol port with accounting enabled. The carrier can determine the record collection time by configuring the time of day accounting (TODA) or by allowing the system to collect statistics at the default 12-hour period. For more information on TODA and accounting collection times, see NN10600-560 *Nortel Networks Multiservice Switch 7400/15000/20000 Accounting*.

Troubleshooting IP accounting

Specific circumstances can affect the collection of the accounting records. During a CP switchover, for example, all IP VPN accounting statistics are cleared and the records for that accounting interval are lost.

The configuration of the protocol ports within the VPN determines if single-ended or double-ended accounting is possible. For layer 2 data connections, only single-ended accounting is possible. For IP tunnel configurations, if accounting is enabled at both ends of the tunnel, then two accounting records are generated for each tunnel (double-ended accounting).

IP security mechanisms

The Nortel Networks Multiservice Switch is based on the VxWorks real-time operating system, so a virus or worm will typically not affect a Multiservice Switch node itself. Most Multiservice Switch nodes operate normally even when a connected network is infected with a virus or worm; however, problems can occur due to the possible flooding of the nodes with large amounts of IP traffic across a large range of unique destination addresses (DAs). These destination IP addresses may or may not exist on the network, which in turn can cause excessive CPU utilization as the Multiservice Switch attempts to forward or discard these packets in addition to wasting bandwidth within the links themselves.

The exact response of the Multiservice Switch varies depending on a number of factors including the capacity of the network, the network configuration and the type of the node's functional processors (FPs) involved. The only permanent solution to these problems is to remove the virus or worm from the infected network; however, there are some measures that can be taken to alleviate the effects on the Multiservice Switch itself. The goal is to block or drop the virus traffic as soon as possible, within the network, and not allow it to be forwarded, thus cutting down on the amount of illegitimate traffic within the network.

There are three protection mechanisms to help prevent denial of service (DoS) attacks at the IP layer. These protection mechanisms can be used together to form a more complete solution. The control plane protection (CPP) feature prevents one VR that is experiencing a DoS attack from exhausting CPU resources, which will affect the other VRs or VRFs. In protect mode, once CPP detects an excessive traffic flow, the packets are discarded in hardware. The discard route entry prevents packets that are not routable and are not covered by CPP to be discarded in hardware, which reduces the load on the CPU. Finally, MD5 cryptographic authentication on the routing protocols (OSPF, BGP, LDP) prevents unauthorized parties from forming peer relationships with a virtual router. This can prevent attackers from introducing faulty routing information into the Multiservice Switch.

All three solutions help reduce the CPU load.

Navigation

Control plane protection (CPP)

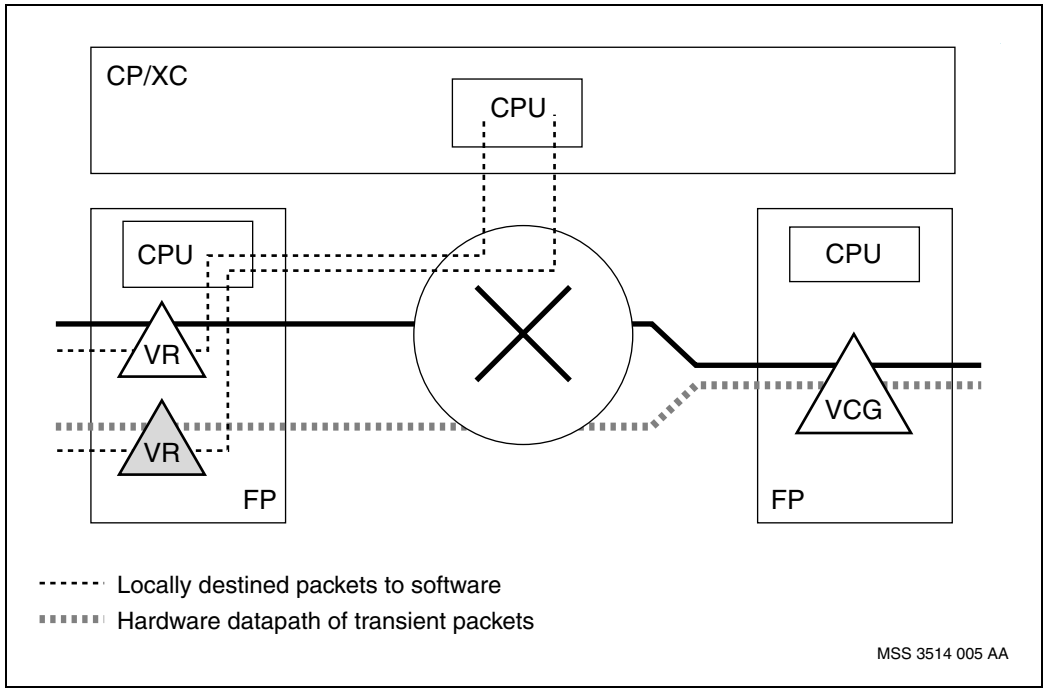
The Nortel Networks Multiservice Switch IP Control Plane Protection (CPP) helps ensure that VRs have fair access to the processor on the FP, which in turn prevents congestion on the CP/VpnXc by monitoring traffic terminating on the node. If an excessively large volume of traffic is sent to a destination address (DA) on the node, the traffic can (optionally) be discarded in hardware at the FP-level before it gets the opportunity to consume any shared system resources. A VPN is an example of a shared resource that should not be impacted by another VPN with an excessive volume of locally destined traffic.

This is achieved by automatically monitoring local DAs on VRs with the CPP component provisioned. If the rate of packets received by the DA is greater than the provisioned value, CPP is notified of the VR and DA that exceeded the threshold, so that it can monitor the DA.

For provisioning information about CPP, see the section Configuring VR CPP in NN10600-801 *Nortel Networks Multiservice Switch 7400/15000/20000 IP Configuration Management*, and sections Configuring RTR CPP and Configuring VRF CPP in NN10600-582 *Nortel Networks Multiservice Switch 7400/15000/20000 VPN Configuration Management*.

The hardware datapath of tandem packets is not impacted by this feature, as seen in High level view of CPP in an RFC2764 example (page 154).

High level view of CPP in an RFC2764 example



Benefits

Control plane protection provides a number of benefits:

- locally destined traffic with abnormally high traffic volumes can be automatically alarmed, shut off, and restarted without operator intervention
- potential denial of service (DoS) attacks are identified and rate-limited without requiring operator intervention or investigation
- the feature is optionally enabled per VR
- operators can manually re-enable traffic to an affected DA of a VR
- tandem traffic or other DAs on an FP with an isolated DA will not be affected by the VRs experiencing DoS attacks
- ability to notify the user of excessive volumes on a per-DA/VR/FP basis
- the feature can be enabled without discard to allow service providers to set a value and study whether the assigned flow rate is accurate, or whether there are excessive flow rates during normal operations

Monitoring process

There are three possible operational modes of CPP: disabled, study, and protect. The disabled mode offers the flexibility of pre-configuring the feature without enabling the monitoring process.

In order to monitor frames, the operational mode must be either study or protect. Other parameters that need to be configured are as follows:

- the flow rate, in packets per second, that triggers monitoring of a DA
- a grace period to calculate the average flow rate to avoid isolating DAs during bursty traffic; a value of zero, when mode is protect, indicates that traffic will be discarded immediately
- an isolation timer which indicates the amount of time that traffic to the DA should be discarded; a value of zero indicates to permanently discard traffic until the condition is cleared by an operator

When the provisioning is activated, all packets destined for local DAs on the VRs with the CPP component will have flow rate information kept on them after the first packet is received by the DA.

Protect mode

The hardware monitoring notifies the CPP agent on the FP once the hardware detects that a DA has exceeded the configured threshold.

Upon notification, the following steps are performed:

1 Sample the DA traffic for the duration of the grace period and calculate the average flow rate.

2 If the average flow rate exceeds the threshold, start discarding the DA traffic in hardware for the duration of the isolation timer and generate a minor SET alarm.

3 Prior to the end of the isolation time, resample the DA traffic to calculate the average flow rate. If it still exceeds the threshold, double the isolation time and continue discarding traffic. This step is repeated at most one more time at which point the isolation time would be four times the configured value.

4 If the DA traffic exceeds the threshold after discarding and resampling the DA traffic three times, then the DA is permanently disabled. A major SET alarm is generated and the operator needs to clear the problem before the DA traffic can resume.

5 If the DA traffic does not exceed the threshold when the average flow rate is calculated in any of the above steps, then generate a CLEAR alarm. The traffic rate is deemed to be acceptable, and hardware monitoring resumes.

Study mode

During the study mode, traffic is never discarded. The isolation time signifies how often to sample/monitor the flow.

The hardware monitoring notifies the CPP agent on the FP once the hardware detects that a DA has exceeded the configured threshold. Upon notification, the following steps are performed:

1 Sample the DA traffic for the duration of the grace period and calculate the average flow rate.

2 If the average flow rate exceeds the threshold, log the event and start the isolation timer which is used as an inter-sampling period only.

3 Prior to the end of the isolation time, resample the DA traffic to calculate the average flow rate and move to the previous step.

4 A spooled statistic record per DA per FP is generated every 15 minutes, summarizing the traffic flow rate of exceeding DAs during that 15 minute period.

5 The process continues until the DA traffic goes below the threshold. In that case, the traffic rate is deemed to be acceptable, and hardware monitoring resumes

VR support

Flow rate monitoring can be performed on DAs associated with any of the VRs supported by Multiservice Switch IP:

- Customer VRs (CVRs)
- VPN Route Forwarders (VRFs)
- Virtual Carrier Gateways (VCGs)
- Routers (RTRs)
- Management VRs (mVRs)
- VIPR VRs

It is recommended that IP CPP on the VR only be enabled on VRs connected to customer equipment, namely CVRs, VRFs, and VIPR. This will avoid poorly engineered or configured networks from losing their core routing links.

The OAM Ethernet port on a PQC-based CP of an mVR cannot use the CPP feature as the OAM Ethernet port uses a software datapath only. However, other OAM ports on the mVR such as a PQC-based port or ethernet FPs can use the CPP feature.

Misbehaviors

When a VR exceeds its programmed flow rate threshold in protect mode, the alarm that is generated will indicate which VR, DA, the average flow rate, and LP the excessive flow originated on.

This data can be used to trace the origins of the excessive traffic, to determine if the flow rate should be increased for the CPP component of the VR, and/or to implement an IP flow filtering. For more information about IP flow filters, see NN10600-590 *Nortel Networks Multiservice Switch 7400/15000/20000 Layer 3 Traffic Management Fundamentals*.

Discard route

When a packet is received on a PQC-based card with a destination IP address that is not known to the forwarding table, a software exception occurs and the packet is processed through the software data path.

The software data path is engineered and designed for exceptional packet processing, such as the first time a packet for a particular subnet is forwarded, ARP processing, OSPF hello packets.

Some viruses are designed to send a high volume of packets to random or sequential addresses. If this happens, and these addresses are not known to the forwarding table, all these packets will be excepted to software.

If the volume is high enough, software processing congestion can occur, and valid packets that maintain the state dynamic protocols such as OSPF can be lost.

A default discard route will avoid this potential problem. A default discard route will discard, in hardware, all packets for which there is no known route. There will be no software exception. Statistics will be updated to track the discarded packets.

A consequence of this is that the Nortel Networks Multiservice Switch will not generate any ICMP Destination Unreachable Error messages. These are normally sent to the originator (source IP address) of a packet when a destination address is not known to the router. The value of sending the ICMP Destination Unreachable messages has to be balanced against the risk of a virus attack that affects the services on an FP. In many applications, the value of sending ICMP Destination Unreachable messages is limited.

For more information about discard route entry, see “Discard route entry” (page 124).

MD5 authentication

This feature introduces protection of OSPF, BGP and LDP peering relationships via MD5 authentication.

The MD5 authentication functionality introduced in this feature differs significantly depending on which protocol is being authenticated. For example, authentication for the BGP and LDP protocols is controlled on a

peering relationship basis and involves the provisioning of a single key per peer/neighbor, whereas OSPF is controlled on a per interface/subnet basis and involves the provisioning of multiple keys per interface/subnet.

MD5 is a one-way hash algorithm that takes a message of arbitrary length and produces a 128-bit message digest or checksum. When used in a protocol authentication scenario, the MD5 algorithm is combined with the use of a secret key, to provide security against message spoofing. The routers at either end of a protocol peering session are each provisioned with a shared, secret key. The message digest is calculated as a function of the protocol packet and the secret key. Subsequently, only the protocol packet and the message digest are transmitted. The other end must combine the received protocol packet with its own copy of the key to verify the received message digest. The actual key is never sent over the network, and as MD5 is a one-way algorithm it is impractical to attempt to discover the key from the transmitted data.

The goal of the MD5 authentication mechanism is solely to ensure that protocol peering relationships are only formed with legitimate peers. It is important to understand what kind of security the MD5 mechanism provides and what attacks it will not protect against. The MD5 authentication mechanism does not protect against bogus routing information originating from a legitimate authenticated peer. If the security of a legitimate peer router has been compromised by some other means and its routing database has been corrupted, then that compromised router can now advertise this faulty information to us, even though our session with it is authenticated. The MD5 authentication mechanism does not provide specific protection against replay attacks. The only protection against a replay attack is provided by the normal TCP sequence number processing. A replay attack could be mounted against a protocol session running over TCP, but only in very carefully timed circumstances.

The security offered by this feature relies heavily on the quality of the key information that is used to compute the message digest value. Key information is extremely sensitive data and must be transmitted and stored with caution. The fewer people who are aware of the key data the better. Customers should define keys that are not easily guessed, and keys should not be shared among multiple peering relationships in order to avoid a “break one, break all” scenario. If all these other methods fail and the key is broken, then reducing key lifetimes can partially mitigate the effects by limiting the window of opportunity for a rogue keyholder. In order to encourage the frequent changing of key values, a key change should not be a session critical event. That is, a method of transitioning keys is required to enable customers to move to a new key value without terminating an existing protocol peering relationship. The MD5 authentication feature supports smooth transitioning of key values for OSPF, BGP and LDP sessions between two Multiservice Switches.

Nortel Networks Multiservice Switch 7400/15000/20000

IP Technology Fundamentals

Publication: NN10600-800
Document status: Standard
Document issue: 6.1S2
Document date: November 2004
Product release: Release 6.1
Job function: Operations
Type: NTP
Language type: U.S. English