Planning and Engineering — Network Design
Avaya Virtual Services Platform 9000

3.2
NN46250-200, 03.04
April 2012

Notice

While reasonable efforts have been made to ensure that the information in this document is complete and accurate at the time of printing, Avaya assumes no liability for any errors. Avaya reserves the right to make changes and corrections to the information in this document without the obligation to notify any person or organization of such changes.

Documentation disclaimer

"Documentation" means information published by Avaya in varying mediums which may include product information, operating instructions and performance specifications that Avaya generally makes available to users of its products. Documentation does not include marketing materials. Avaya shall not be responsible for any modifications, additions, or deletions to the original published version of documentation unless such modifications, additions, or deletions were performed by Avaya. End User agrees to indemnify and hold harmless Avaya, Avaya's agents, servants and employees against all claims, lawsuits, demands and judgments arising out of, or in connection with, subsequent modifications, additions or deletions to this documentation, to the extent made by End User.

Link disclaimer

Avaya is not responsible for the contents or reliability of any linked Web sites referenced within this site or documentation provided by Avaya. Avaya is not responsible for the accuracy of any information, statement or content provided on these sites and does not necessarily endorse the products, services, or information described or offered within them. Avaya does not guarantee that these links will work all the time and has no control over the availability of the linked pages.

Warranty

Avaya provides a limited warranty on its Hardware and Software ("Product(s)"). Refer to your sales agreement to establish the terms of the limited warranty. In addition, Avaya's standard warranty language, as well as information regarding support for this Product while under warranty is available to Avaya customers and other parties through the Avaya Support Web site: http://support.avaya.com. Please note that if you acquired the Product(s) from an authorized Avaya reseller outside of the United States and Canada, the warranty is provided to you by said Avaya reseller and not by Avaya.

Licenses

Copyright

Third-party components

Certain software programs or portions thereof included in the Product may contain software distributed under third party agreements ("Third Party Components"), which may contain terms that expand or limit rights to use certain portions of the Product ("Third Party Terms"). Information regarding distributed Linux OS source code (for those Products that have distributed the Linux OS source code), and identifying the copyright holders of the Third Party Components and the Third Party Terms that apply to them is available on the Avaya Support Web site: http://support.avaya.com/Copyright.

Trademarks

The trademarks, logos and service marks ("Marks") displayed in this site, the Documentation and Product(s) provided by Avaya are the registered or unregistered Marks of Avaya, its affiliates, or other third parties. Users are not permitted to use such Marks without prior written consent from Avaya or such third party which may own the Mark. Nothing contained in this site, the Documentation and Product(s) should be construed as granting, by implication, estoppel, or otherwise, any license or right in and to the Marks without the express written permission of Avaya or the applicable third party.

Avaya is a registered trademark of Avaya Inc.

All non-Avaya trademarks are the property of their respective owners, and "Linux" is a registered trademark of Linus Torvalds.

Downloading Documentation

For the most current versions of Documentation, see the Avaya Support Web site: http://support.avaya.com.

Contact Avaya Support

Avaya provides a telephone number for you to use to report problems or to ask questions about your Product. The support telephone number is 1-800-242-2121 in the United States. For additional support telephone numbers, see the Avaya Web site: http://support.avaya.com.

Contents

Chapter 1: Purpose of this document.... 7

Chapter 2: New in this release.... 9

Features.... 9

Other changes.... 10

Chapter 3: Network design fundamentals.... 11

Chapter 4: Hardware fundamentals and guidelines.... 13

Chassis considerations.... 13

Modules.... 14

Optical device guidelines.... 15

1000BASE-X and 10GBASE-X reach.... 16

Dispersion considerations for long reach.... 16

10/100BASE-X and 1000BASE-TX reach.... 17

10/100/1000BASE-TX Auto-Negotiation recommendations.... 18

Auto MDIX.... 18

CANA.... 19

Chapter 5: Platform redundancy.... 21

Power redundancy.... 21

Input/output port redundancy.... 21

Control plane redundancy.... 22

Switch Fabric redundancy.... 22

Configuration redundancy.... 23

Link redundancy.... 23

Switch redundancy.... 23

High Availability mode.... 23

Chapter 6: Link redundancy.... 27

Physical layer redundancy.... 27

Multilink Trunking.... 31

802.3ad-based link aggregation.... 32

Chapter 7: Redundant network design.... 35

Network Load Balancing.... 35

Chapter 8: Lossless networks.... 39

Chapter 9: Layer 2 switch clustering and SMLT.... 49

Modular design for redundant networks.... 49

Network edge redundancy.... 53

Split MultiLink Trunk configuration.... 54

SMLT full-mesh recommendations with OSPF.... 63

Chapter 10: Layer 3 switch clustering and RSMLT.... 65

Routed SMLT.... 65

Switch clustering topologies and interoperability with other products.... 72

Chapter 11: Layer 3 switch clustering and multicast SMLT.... 75

General guidelines.... 75

Multicast triangle topology.... 77

Square and full-mesh topology multicast guidelines.... 79

SMLT and multicast traffic issues.... 82

Chapter 12: Layer 2 loop prevention.... 85
Loop prevention and detection.... 85
CPU protection and loop prevention compatibility.... 90
Chapter 13: Spanning tree.... 91
Spanning tree and protection against isolated VLANs.... 91
MSTP and RSTP considerations.... 92
Chapter 14: Layer 3 network design.... 95
VRF Lite.... 95
Subnet-based VLAN guidelines.... 107
Open Shortest Path First.... 107
Border Gateway Protocol.... 112
IPv6 BGP+.... 118
IP routed interface scaling.... 118
IPv6.... 119
Chapter 15: IP multicast network design.... 123
Multicast and MultiLink Trunking considerations.... 123
Multicast scalability design rules.... 123
IP multicast address range restrictions.... 125
Multicast MAC address mapping considerations.... 126
Dynamic multicast configuration changes.... 128
IGMPv3 backward compatibility.... 128
TTL in IP multicast packets.... 129
Multicast MAC filtering.... 129
Guidelines for multicast access policies.... 130
Split-subnet and multicast.... 130
Protocol Independent Multicast-Sparse Mode guidelines.... 131
Protocol Independent Multicast-Source Specific Multicast guidelines.... 143
IGMP and PIM-SM interaction.... 144
Layer 3 multicast extended VLANs.... 144
Multicast for multimedia.... 147
Chapter 16: System and network stability and security.... 149
Control plane rate limit (CP-Limit).... 149
DoS protection mechanisms.... 150
Damage prevention.... 151
Security and redundancy.... 153
Data plane security.... 153
Control plane security.... 157
Additional information.... 163
Chapter 17: QoS design guidelines.... 165
QoS mechanisms.... 165
QoS interface considerations.... 168
Network congestion and QoS design.... 170
QoS examples and recommendations.... 171
Chapter 18: Layer 1, 2, and 3 design examples.... 175
Layer 1 examples.... 175
Layer 2 examples.... 178
Layer 3 examples.... 182

RSMLT redundant network with bridged and routed VLANs in the core 186

Chapter 19: Optical routing design 191

Optical routing system components 191

Chapter 20: Software and hardware scaling capabilities 193

Hardware scaling capabilities 193

Software scaling capabilities 194

Chapter 21: Supported standards, request for comments, and Management Information Bases 199

Supported standards 199

Supported RFCs 200

Quality of service 203

Network management 203

MIBs 205

Standard MIBs 206

Proprietary MIBs 209

Chapter 22: Customer service 211

Getting technical documentation 211

Getting product training 211

Getting help from a distributor or reseller 211

Getting technical support from the Avaya Web site 211

Glossary 213

Chapter 1: Purpose of this document

Planning and Engineering — Network Design provides information to help you build robust, efficient Virtual Services Platform 9000 networks

You can use the Network Design document to achieve redundancy and provide security and Quality of Service (QoS) in your network, and you can use the examples and important design guidelines for many features and protocols.

Chapter 2: New in this release

The following sections detail what is new in *Avaya Virtual Services Platform 9000 Planning and Engineering — Network Design* , NN46250–200, for Release 3.2:

Features

See the following section for information on feature-related changes.

IGMPv3

Release 3.2 enhances IGMPv3 to fully support RFC3376. For more information, see the following sections:

- IGMPv3 backward compatibility on page 128
- Protocol Independent Multicast-Sparse Mode guidelines on page 131
- Protocol Independent Multicast-Source Specific Multicast guidelines on page 143
- Multicast for multimedia on page 147

IPv6

Release 3.2 adds IPv6 routing support. For more information about IPv6, see the following sections:

- IPv6 BGP+ on page 118
- IPv6 on page 119
- Routed SMLT on page 65

Layer 3 multicast extended VLANs

Layer 3 multicast extended VLANs on page 144 is added to address issues with this network design.

Lossless Ethernet

Release 3.2 adds the Lossless Ethernet feature for 10 GbE ports. For more information, see Lossless networks on page 39. For configuration procedures, see *Avaya Virtual Services Platform 9000 Configuration — Ethernet Modules*, NN46250–508.

Other changes

See the following section for information about changes that are not feature-related.

Glossary

A glossary of related terms is added to the end of this document.

Purpose of this document

To improve documentation usability, a brief description of the purpose of this document is now the first chapter.

RFCs

Supported RFCs on page 200 is updated for new support.

Chapter 3: Network design fundamentals

To efficiently and cost-effectively use Avaya Virtual Services Platform 9000, you must properly design your network. Use the information in this section to help you properly design the network. To design networks, you must consider

- reliability and availability
- platform redundancy
- desired level of redundancy

A robust network depends on the interaction between system hardware and software. System software can be divided into different functions as shown in the following figure.

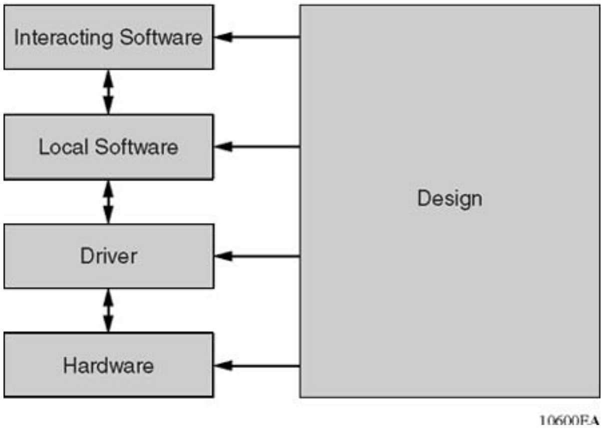

Figure 1: Hardware and software interaction

These levels are based on the software function. A driver is the lowest level of software that actually performs a function. Drivers reside on a single module and do not interact with other modules or external devices. Drivers are very stable.

Statically configured MultiLink Trunking (MLT) is a prime example of local software because it interacts with several modules within in the same device. No external interaction is needed, so you can easily test the function.

Interacting software is the most complex level of software because it depends on interaction with external devices. The Open Shortest Path First (OSPF) protocol is a good example of this software level. Interaction can occur between devices of the same type or with devices of other vendors than run a completely different implementation.

Based on network problem-tracking statistics, the following list is an approximate stability estimation model of a system that uses these components:

- Hardware and drivers represent a small portion of network problems.
- Local software represents a more significant share.
- Interacting software represents the vast majority of the reported issues.

Based on this model, network design attempts to off-load the interacting software level as much as possible to the other levels, especially to the hardware level. Avaya recommends that you follow these generic rules when you design networks:

1. Design networks as simply as possible.
2. Provide redundancy, but do not over-engineer your network.
3. Use a toolbox to design your network.
4. Design according to the product capabilities described in the latest release notes.
5. Follow the design rules provided in this document and also in the various configuration documents for the device.

Chapter 4: Hardware fundamentals and guidelines

This section provides general hardware guidelines to use this product in a network. Use the information in this section to help you during the hardware design and planning phase.

Chassis considerations

This section provides chassis power and cooling considerations. You must properly power and cool your chassis, or nonoptimal operation can result.

Chassis power considerations

Avaya Virtual Services Platform 9000 chassis supports up to six AC power supplies. You must install at least one power supply in each chassis. You can install more than one based on additional power requirements or to provide power redundancy.

The nominal input voltage range is 100–120 VAC and 200–240 VAC; however, the output power is limited to 1200 W maximum at 100–120 VAC nominal input voltage conditions. To obtain full output power of 2000 W, you must connect the 9006AC power supply to a 200–240 VAC nominal input voltage source.

❗ **Important:**

Avaya recommends that power supplies use the same input AC voltage. Do not operate the chassis with power supplies under different input AC voltage conditions. The product functions but Avaya does not support this configuration.

The following table describes the power requirements for chassis components.

Table 1: Component power draw

Component	Power required (Watts)
9080CP Control Processor module	80 W
9090SF Switch Fabric module	70 W (slots SF1 and SF4) 50 W (slots SF2, SF3, SF5, and SF6)
9012FC IO cooling module	65 W
9012SC SF cooling module	150 W
9048GT 48-port 10/100/1000BASE-T module	350 W

Component	Power required (Watts)
9048GB 48-port 1000BASE-X SFP module	340 W, with short range SFP
9024XL 24-port 10GBASE-X SFP+ module	575 W, with short range SFP+

Chassis cooling

You must install four cooling modules in the chassis, two at the front and two at the back.

The 9012FC cooling modules provide cooling for the interface modules and the Control Processor (CP) modules. Each cooling module includes eight fans.

You must install two 9012RC SF cooling modules in the back of the chassis. Each cooling module includes two fans.

If you remove an interface module from the chassis, you must replace it with another interface module or a filler module. Never leave a module slot empty; an empty slot affects the chassis cooling. If you leave a module slot empty, the interface modules can overheat and shutdown. An empty slot can also affect the mean time between failure (MTBF) of the interface modules.

Caution:

Risk of electromagnetic interference

Do not operate Virtual Services Platform 9000 with an empty module slot. If you need to replace a failed module and you do not have a replacement module, leave the failed module installed or install a filler module.

Operating the switch with an empty module slot can cause electromagnetic interference to other equipment in the area. Improper equipment cooling can also result.

Modules

Use modules to interface the device to the network. This section discusses design guidelines and considerations for Virtual Services Platform 9000 modules.

SF modules

The Switch Fabric (SF) module performs intelligent switching. You can install a maximum of six SF modules in each chassis in an N + 1 configuration with redundancy.

The slot location determines the module function. Slots SF1 and SF4 provide the arbitration and scheduling for traffic (and therefore, bandwidth management) from the interface modules and provide redundancy if you populate both slots. The remaining slots provide additional bandwidth. Each chassis has slots for five operational SF modules plus one hot backup. Install six SF modules to achieve full line rate and redundancy.

Important:

You must install a minimum of three SF modules in the chassis. Install an SF module in both slots SF1 and SF4. Install a third SF module in one of the remaining slots.

CP modules

The CP module performs routing and manages the SF modules. You can operate the chassis with only one CP module; you can install a second CP module for redundancy.

For more information about how to protect the CPU from DOS attacks, see *Avaya Virtual Services Platform 9000 Administration*, NN46250-600.

Interface modules

You can install a maximum of 10 interface modules in the chassis. Interface modules provide support for a variety of technologies, interfaces, and feature sets and provide up to 1 and 10 Gb/s port rates. Virtual Services Platform 9000 supports the following interface modules:

- 9048GT: 48-port 10/100/1000BASE-T
- 9048GB: 48-port 1000BASE-X SFP
- 9024XL: 24-port 10GBASE-X SFP+

  When the chassis is equipped with 5 SF modules, the 9024XL module has a 3.5:1 oversubscribed line rate over 24 ports; 6 ports can provide full line rate if you do not use the remaining ports. Each continuous physical group of 4 ports supports a combined bandwidth of 10G. Use only a single port for each grouping to ensure no oversubscription. As a helpful guide the last port in each group has a black mark on the faceplate.

For more information about SFP and SFP+ specifications, see *Avaya Virtual Services Platform 9000 Installation — SFP Hardware Components*, NN46250-305.

To decide which modules you need to use, consider the feature and scaling each module supports. For more information about features and scaling, see Software and hardware scaling capabilities on page 193. For the most recent scaling information, always consult the latest version of release notes.

Optical device guidelines

Use optical devices to achieve high bit-rate communications and long transmission distances. Use the information in this section to properly use optical devices in a network.

Optical power considerations

When you connect the device to collocated equipment, ensure that enough optical attenuation exists to avoid overloading the receivers of each device. You must consider the minimum attenuation requirement based on the specifications of third-party equipment. For more information about minimum insertion losses for Avaya optical products, see *Avaya Virtual Services Platform 9000 Installation — SFP Hardware Components*, NN46250-305.

1000BASE-X and 10GBASE-X reach

You can use various SFP (1Gb/s) and SFP+ (10Gb/s) to attain different line rates and reaches. The following tables show typical reach attainable with optical devices.

For more information about these devices, including compatible fiber type, see *Avaya Virtual Services Platform 9000 Installation — SFP Hardware Components*, NN46250-305.

Table 2: SFP optical devices and maximum reach

SFP	Maximum reach
1000BASE-SX	Up to 550 m
1000BASE-LX	Up to 10 km over one fiber; up to 550 m over mulitmode fiber
1000BASE-BX	Up to 10 km over one fiber
1000BASE-XD	Up to 40 km
1000BASE-ZX	Up to 70 km
1000BASE-EX	Up to 120 km

Table 3: SFP+ optical devices and maximum reach

SFP+	Maximum reach
10GBASE-LRM	Up to 220 m
10GBASE-SR/SW	Up to 300 m
10GBASE-LR/LW	Up to 10 km
10GBASE-ER/EW	Up to 40 km

Dispersion considerations for long reach

Precise engineering of transmission links is difficult; specifications and performance are often unknown, undocumented, or impractical to measure before equipment installation. Moreover, the skills required to perform rigorous link budget analysis are extensive. Fortunately, a simple, straightforward approach can assure robust link performance for most optical fiber systems in which you use Avaya switches and routers.

This method uses an optical power budget, the difference between transmitter power and receiver sensitivity, to determine whether the installed link can operate with low bit error ratio for extended periods. The power budget must accommodate the sum of link loss (that is, attenuation), dispersion, and system margin, described in the following paragraphs.

Link losses are the sum of cabled fiber loss, splices, and connectors, often with an allocation for additional connectors. Cabled fiber loss is wavelength and installation-dependant, and is typically in the range of 0.20 to 0.5 dB/km. See the cable plant owner or operator for specifications of the cable you use, particularly if the available system margin is unsatisfactory. Engineered links require precise knowledge of the cable plant.

For long, high bitrate systems, pulse distortion, caused by the transmitter laser spectrum interaction with fiber chromatic dispersion, reduces receiver sensitivity. Transceivers for long reach single mode fiber systems have an associated maximum dispersion power penalty (DPP_{max}) specification, which applies to G.652 (dispersion unshifted) single mode fiber and the rated transceiver reach. The actual power penalty that you must use is

DPP_{budget} = [link length(km) / transceiver max reach (km)] * DPP_{max}

For example, if an 80 km transceiver is specified as having DPP < 3 dB, and if the actual link length will be 40 km, DPP_{budget} is one-half the maximum, or 1.5 dB.

Link operating margins are sometimes allocated for impairments such as aging, thermal, or other environmental effects. Due to the potentially large number of factors that can degrade performance, you can usually rely on statistics to represent these factors as a single margin value, in dB, to cover all effects. Margin is life and design-dependent, but is typically 3.5 to 4.5 dB, minimum. Whether you require additional margin depends on the details, such as whether actual or specified transmitter power and receiver sensitivity are used. Avaya specifications represent worst-case values.

The sum of margin, dispersion power penalty, and passive cable plant losses must be less than the available power budget. Alternatively, if you calculate available power margin as the difference between available budget and the sum of losses and dispersion, the margin can be more or less than required, which determines whether additional consideration is needed. If the power budget is exceeded or margin is insufficient, you can either use a transceiver rated for longer distance operation, or calculate budget and losses using actual values rather than specified limit values. Either method can improve link budget by 4 to 5 dB or more.

10/100BASE-X and 1000BASE-TX reach

The following table lists maximum transmission distances for 10/100BASE-X and 1000BASE-TX Ethernet cables.

Table 4: Maximum cable distances

	10BASE-T	100BASE-TX	1000BASE-TX
IEEE standard	802.3 Clause 14	802.3 Clause 21	802.3 Clause 40

	10BASE-T	100BASE-TX	1000BASE-TX
Date rate	10 Mb/s	100 Mb/s	1000 Mb/s
Cat 5 UTP distance	100 m	100 m	100 Ω, 4 pair: 100 m

10/100/1000BASE-TX Auto-Negotiation recommendations

Auto-Negotiation lets devices share a link and automatically configures both devices so that they take maximum advantage of their abilities. Auto-Negotiation uses a modified 10BASE-T link integrity test pulse sequence to determine device ability.

The Auto-Negotiation feature allows the devices to switch between the various operational modes in an ordered fashion and allows management to select a specific operational mode. The Auto-Negotiation feature also provides a parallel detection (also called autosensing) function to allow the recognition of 10BASE-T, 100BASE-TX, 100BASE-T4, and 1000BASE-TX compatible devices, even if they do not support Auto-Negotiation. In this case, only the link speed is sensed; not the duplex mode. Avaya recommends the Auto-Negotiation configuration as shown in the following table, where A and B are two Ethernet devices.

Table 5: Recommended Auto-Negotiation configuration on 10/100/1000BASE-TX ports

Port on A	Port on B	Remarks	Recommendations
Auto-Negotiation enabled	Auto-Negotiation enabled	Ports negotiate on highest supported mode on both sides.	Avaya recommends that you use this configuration if both ports support Auto-Negotiation mode.
Full-duplex	Full-duplex	Both sides require the same mode.	Avaya recommends that you use this configuration if you require full-duplex, but the configuration does not support Auto-Negotiation.

Auto-Negotiation cannot detect the identities of neighbors or shut down misconnected ports. Upper-layer protocols perform these functions.

Auto MDIX

Automatic medium dependent interface crossover (Auto-MDIX) automatically detects the need for a straight-through or crossover cable connection and configures the connection appropriately. This removes the need for crossover cables to interconnect switches and ensures either type of cable can be used. The speed and duplex setting of an interface must be set to auto for Auto-MDIX to operate correctly.

CANA

Use Custom Auto-Negotiation Advertisement (CANA) to control the speed and duplex settings that the interface modules advertise during Auto-Negotiation sessions between Ethernet devices. Modules can only establish links using these advertised settings, rather than at the highest common supported operating mode and data rate.

Use CANA to provide smooth migration from 10/100 Mb/s to 1000 Mb/s on host and server connections. Using Auto-Negotiation only, the switch always uses the fastest possible data rates. In limited-uplink-bandwidth scenarios, CANA provides control over negotiated access speeds, and improves control over traffic load patterns.

You can use CANA only on 10/100/1000 Mb/s RJ-45 ports. To use CANA, you must enable Auto-Negotiation.

Important:

If a port belongs to a Multilink Trunking (MLT) group and you configure CANA on the port (that is, you configure an advertisement other than the default), then you must apply the same configuration to all other ports of the MLT group (if they support CANA).

If a 10/100/1000 Mbit/s port that supports CANA is in a MLT group that has 10/100BASE-TX ports, or any other port type that does not support CANA, then use CANA only if it does not conflict with MLT abilities.

Chapter 5: Platform redundancy

This section includes recommendations to provide a fault tolerant platform.

Power redundancy

Virtual Services Platform 9000 provides n + 1 redundancy. Employ n + 1 power supply redundancy, where n is the number of required power supplies to power the chassis and modules. Connect the power supplies to an additional power supply line to protect against supply problems.

The nominal input voltage range is 100–120 VAC and 200–240 VAC; however, the output power is limited to 1200 W maximum at 100–120 VAC nominal input voltage conditions. To obtain full output power of 2000 W, you must connect the 9006AC power supply to a 200–240 VAC nominal input voltage source.

> ❗ **Important:**
>
> Avaya recommends that power supplies use the same input AC voltage. Do not operate the chassis with power supplies under different input AC voltage conditions. The product can function but Avaya does not support this configuration.

For more information about power supplies, see *Avaya Virtual Services Platform 9000 Installation — AC Power Supply*, NN46250-303.

Input/output port redundancy

You can protect I/O ports using a link aggregation mechanism. MultiLink Trunking (MLT), which is compatible with 802.3ad static, provides a load sharing and failover mechanism to protect against module, port, fiber, or complete link failures.

You can use MLT with Link Access Control Protocol (LACP) disabled or use LACP enabled by itself.

Control plane redundancy

The Control Processor (CP) module is the control plane of the platform. The CP module controls all learning, calculates routes, and maintains port states. If the last CP module in a system fails, the switch restarts the I/O cards after a heartbeat timeout period of 3 seconds.

When the last CP module fails, or is removed from the chassis, all I/O and S/F modules reboot immediately. All I/O ports immediately become disabled. The I/O and S/F modules then reset, reboot, and wait for a CP module to be inserted. The modules reset and reboot intermittently while waiting for a new CP module to be inserted.

For more information about how to configure HA-CPU mode and how to protect the CPU from DOS attacks, see *Avaya Virtual Services Platform 9000 Administration*, NN46250-600.

Switch Fabric redundancy

Avaya recommends that you use three Switch Fabric (SF) modules to protect against switch fabric failures. Avaya recommends that you install SF modules in both SF1 and SF4. Install an SF module in one of the additional slots for bandwidth.

The slot location determines the module function. Slots SF1 and SF4 provide the arbitration and scheduling for traffic (and therefore, bandwidth management) from the interface modules and provide redundancy when both slots are populated. You must install an SF module in one of these slots. Avaya recommends that you install SF modules in both slots to provide redundancy.

If you do not install an SF module in either SF1 or SF4, or if the modules are not operational, all ports on the interface modules shut down. No data can pass through the system without an SF module in one of these two slots. You can access the system only through the console port on the CP module. After you bring a nonoperational SF module online in either SF1 or SF4, the system automatically restarts to ensure it initializes correctly and to provide predictable behavior.

Slots SF2, SF3, SF5, and SF6 provide additional bandwidth. You can install more SF modules than you require to provide bandwidth redundancy.

For more information about the SF modules, see *Avaya Virtual Services Platform 9000 Installation — Modules*, NN46250-301.

Configuration redundancy

You can define primary and backup configuration file paths. This configuration protects against system failures. For example, the primary path can point to system flash memory and the backup path to the external Compact Flash card.

If you enable the system flag **`boot config flags savetostandby`**, it ensures that configuration changes are always saved to both CPUs.

Link redundancy

Provide physical and link layer redundancy to eliminate a single point of failure in the network. For more information, see Link redundancy on page 27.

Switch redundancy

When you use Split MultiLink Trunking (SMLT) to provide switch redundancy, Avaya recommends that you use Virtual Link Aggregation Control Protocol (VLACP) to avoid packet forwarding to a failed switch that cannot process the packets.

Provide network redundancy so that a faulty switch does not interrupt service. You can configure mechanisms that direct traffic around a malfunctioning switch. For more information, see Layer 2 switch clustering and SMLT on page 49, Layer 3 switch clustering and RSMLT on page 65, and Layer 3 switch clustering and multicast SMLT on page 75

High Availability mode

High Availability (HA) mode, also called HA-CPU, activates two CPUs simultaneously. These CPUs exchange topology data so that, if a failure occurs, either CPU can take over the operations of the other.

In HA-CPU mode, the two CPUs are active and exchange topology data through an internal dedicated bus. This configuration allows for a complete separation of traffic. To guarantee total security, users cannot access this bus.

In HA-CPU mode, also called Hot Standby, the two CPUs are synchronized. In non HA-CPU mode, also called Warm Standby, the two CPUs are not synchronized.

The following tables lists feature support and synchronization information for HA-CPU.

Table 6: Feature support for HA-CPU

Feature	Release 3.2
Modules	Yes
Platform	Yes
Layer 2	Yes
Layer 3	Yes; partial-HA for Border Gateway Protocol and IPv6
Multicast	Partial HA
Security	Yes
Applications	Partial HA

Table 7: Synchronization capabilities in HA-CPU mode

Synchronization of	Release 3.2
Layer 1	
Port configuration parameters	Yes
Layer 2	
Multiple Spanning Tree Protocol parameters	Yes
Quality of Service (QoS) parameters	Yes
Rapid Spanning Tree Protocol parameters	Yes
SMLT parameters	Yes
VLAN parameters	Yes
Layer 3	
ARP entries	Yes
Border Gateway Protocol (BGP)	Partial (configuration only)
Dynamic Host Configuration Protocol (DHCP) Relay	Partial (configuration only)
Internet Group Management Protocol	Partial (configuration only)
Internet Group Management (IGMP) Snooping	Yes
IP Filters	Yes
IPv6	Partial (configuration only)
Layer 3 Filters: access control entries, access control lists	Yes

Synchronization of	Release 3.2
Open Shortest Path First (OSPF)	Yes
Packet Capture (PCAP) tool	Yes
PIM	Partial (configuration only)
Prefix lists and route policies	Yes
Routing Information Protocol	Yes
Router Discovery	Yes
Routed Split Multi-Link Trunking (RSMLT)	Yes
RSMLT edge support	Yes
Static and default routes	Yes
Virtual IP (VLANs)	Yes
Virtual Router Redundancy Protocol	Yes
VRF Lite	Yes
Transport Layer	
Network Load Balancing (NLB)	Yes
Remote Access Dial-In User Services (RADIUS)	Yes
UDP forwarding	Yes
Applications	
VSP Talk	Partial (configuration only). After a CPU switchover, you must re-enable event notification.

For more information about how to configure HA-CPU, see *Avaya Virtual Services Platform 9000 Administration*, NN46250-600.

A few applications in HA-mode, or Hot Standby mode, have partial HA implementation. This means that the system synchronizes user configuration data (including interfaces, IPv6 addresses and static routes) between the master CPU and standby CPU. However, for applications in HA-mode with partial HA implementation, the platform does not synchronize dynamic data learned by protocols. As a result, after failure those applications need to restart and rebuild their tables. This operation causes an interruption to traffic that is dependent on a protocol or application with Partial HA support. The following applications support Partial High Availability:

- Border Gateway Protocol (BGP)
- Dynamic Host Configuration Protocol (DHCP) Relay
- Internet Group Management Protocol (IGMP)
- IPv6

- Protocol Independent Multicast-Sparse Mode (PIM-SM)
- Protocol Independent Multicast-Source Specific Mode (PIM-SSM)
- Border Gateway Protocol (BGP)
- VSP Talk

HA-CPU limitations and considerations

You must take the following limitations and considerations into account when you use the HA-CPU feature:

- Activating or deactivating HA-CPU mode causes the standby CP to reset. The active CP continues to operate normally.
- In HA-CPU mode, Avaya recommends that you do not configure the Open Shortest Path First (OSPF) dead router interval for less than 15 seconds.

Chapter 6: Link redundancy

You can build link redundancy into your network to

- help eliminate a single point of failure in your network (provide physical and link layer redundancy)
- prevent a service interruption caused by a faulty link (provide link layer redundancy)

This chapter explains the following design options that you can use to achieve link redundancy (provide alternate data paths) :

- physical layer redundancy
- Multilink Trunking
- 802.1ad-based link aggregation

Physical layer redundancy

To ensure that a faulty link does not cause a service interruption, you can provide physical layer redundancy in your network.

You can also configure the platform to detect link failures with, for example:

- Remote fault indication
- VLACP

Gigabit Ethernet and remote fault indication

The 802.3z Gigabit Ethernet (GbE) standard defines remote fault indication (RFI) as part of the Auto-Negotiation function.

Because RFI is part of the Auto-Negotiation function, if you disable Auto-Negotiation, you automatically disable RFI.

The stations on both ends of a fiber pair use RFI to inform one another after a problem occurs on one of the fibers.

Avaya recommends that you enable Auto-Negotiation on GbE links when the devices on both ends of a fiber link support Auto-Negotiation because, without RFI support, if one of two unidirectional fibers that form the connection between the two platforms fails, the transmitting side cannot determine that the link is broken in one direction (see the following figure).

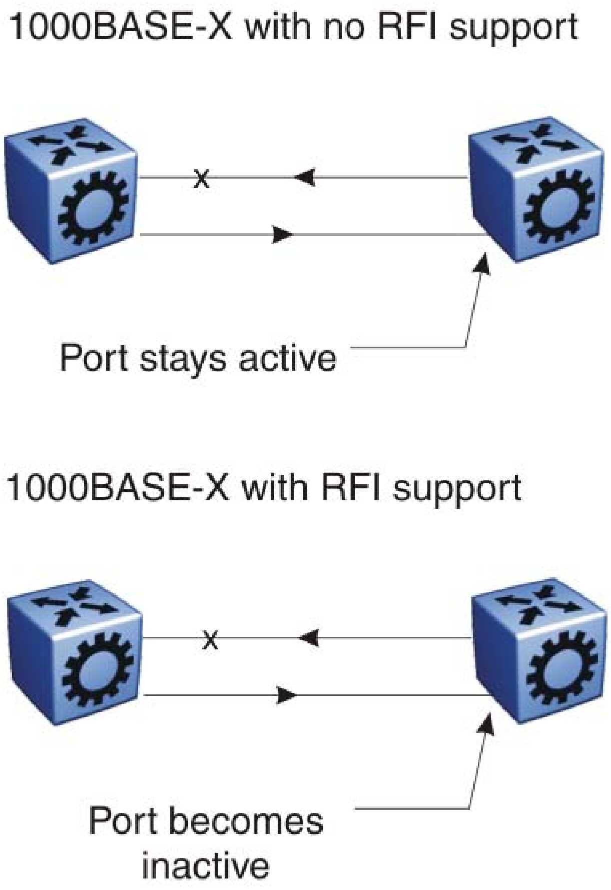

Figure 2: 1000BASE-X RFI

End-to-end fault detection and VLACP

Because remote fault indication (RFI) terminates at the next Ethernet hop, the device that uses only RFI cannot determine failures on an end-to-end basis over multiple hops.

However, you can use Virtual Link Aggregation Control Protocol (VLACP) to provide an end-to-end failure detection mechanism. You can configure VLACP on a port and enable it over single links or Multilink trunks (MLT).

You can use VLACP with MLT to enhance its capabilities and provide quick failure detection. With VLACP, the device can detect far-end failures, which permits MLT to failover properly when end-to-end connectivity is not guaranteed for some links in an aggregation group.

Avaya recommends that you use VLACP for all Split MultiLink Trunk (SMLT) access links, when the links are configured as MLT, to ensure both end devices can communicate.

By using VLACP over SMLT, you extend enhanced failure detection beyond the limits of the number of SMLT or LACP instances that you can create on an Avaya device.

To minimize network outages, you can also use VLACP to switch traffic around entire network devices before Layer 3 protocols detect a network failure.

VLACP is an extension of the Link Aggregation Control Protocol (LACP) but LACP and VLACP are independent features.

VLACP does not perform link aggregation, it detects end-to-end link failures.

VLACP periodically checks the end-to-end health of a point-to-point connection and it uses the hello mechanism of LACP to periodically send hello packets to ensure end-to-end communication.

If VLACP does not receive hello packets it transitions to a failure state, which indicates a service provider failure, and the port is disabled. The system sends VLACP trap messages to the management stations if the VLACP state changes. If the failure is local, the system generates only port linkdown or port linkup traps.

VLACP works for port-to-port communications only where a guarantee exists for a logical port-to-port match through the service provider.

VLACP does not work for port-to-multiport communications where no guarantee exists for a point-to-point match through the service provider.

Example:

When the enterprise networks connect the aggregated Ethernet trunk groups through a service provider network connection, far-end failures cannot be signaled with Ethernet-based functions that operate end-to-end through the service provider network. The Multilink trunk (between enterprise switches S1 and S2) extends through the service provider network.

The following figure shows an MLT that operates with VLACP. VLACP can operate end-to-end, but you can also use it in a point-to-point link.

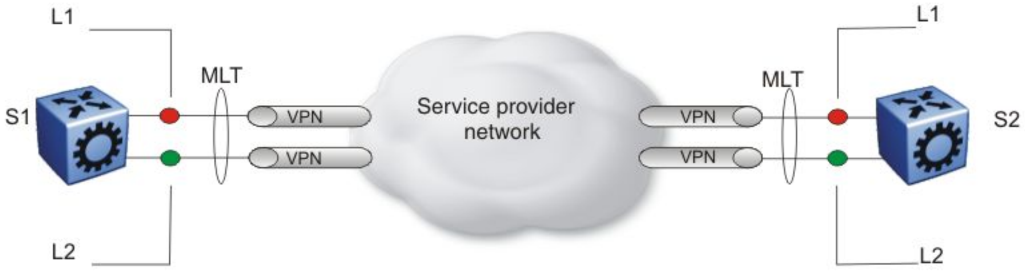

Figure 3: Problem description (1 of 2)

In the following figure, if the L2 link on S1 (S1/L2) fails, the link-down failure is not propagated over the service provider network to S2 and S2 continues to send traffic over the failed S2/ L2 link.

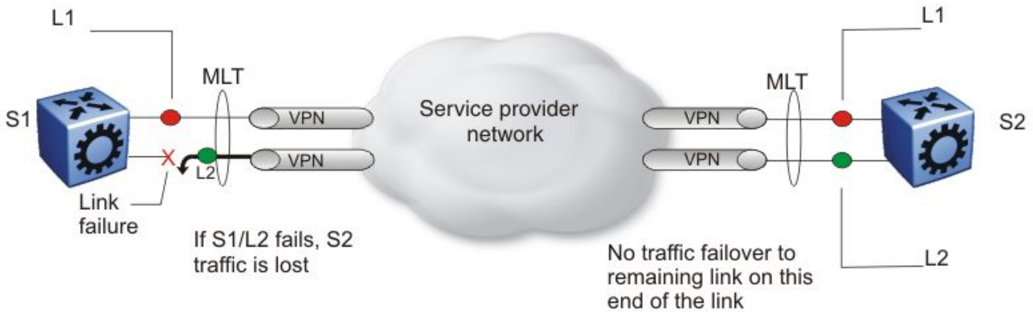
L1
MLT
S1
VPN
VPN
L2
X
Link failure
If S1/L2 fails, S2 traffic is lost
Service provider network
MLT
VPN
VPN
L1
S2
No traffic failover to remaining link on this end of the link
L2

Figure 4: Problem description (2 of 2)

However, if you use VLACP to detect far-end failures and allow MLT to failover when end-to-end connectivity is not guaranteed for links in an aggregation group, then VLACP prevents the failure scenario in the preceding figure.

Avaya recommends that you use the following guidelines for VLACP implementation:

- Do not use VLACP on configured LACP MLTs because LACP provides the same functionality as VLACP for link failure. Virtual Services Platform 9000 does not support VLACP and LACP on the same link.
- Use the following best practice standard settings for VLACP :
    - a short timer—no less than 500 milliseconds (ms)
    - a time-out scale of 5

  Tip:

  Avaya Virtual Services Platform 9000 supports both faster timers and lower time-out scales, but if VLACP flapping occurs, increase the short timer and the time-out scale to their recommended values: 500 and 5, respectively. Although the software configuration supports VLACP short timers of less than 30 ms, the platform does not support using values less than 30 ms in practice. The shortest (fastest) supported VLACP timer is 30 ms with a timeout of 3, which achieves sub-100 ms failover. The platform does not support 30 ms timers in High Availability (HA) mode, and may not be stable in scaled networks.
- Use only long timers for .Interswitch trunk (IST) links. IST links do not support VLACP with short timers. For IST MLTs, Avaya recommends that you do not configure the VLACP long periodic timer to less than 30 seconds.
- Do not configure VLACP timers to less than 100 ms if you plan to use a Layer 3 core with Equal Cost Multipath (ECMP).

  Note:

  This recommendation assumes a combination of basic Layer 2 and Layer 3 with Open Shortest Path First (OSPF). If you have more complex configurations, you can require higher timer values.
- Ensure that the VLACP configuration at the port level is consistent, that both sides of the point-to-point connection are either enabled or disabled. If a VLACP-enabled port does not receive a VLACP protocol data unit (PDU), it enters the disabled state. However, occasions exist when a VLACP-enabled port does not receive a VLACP PDU but remains

in the forwarding state. You can avoid this situation with consistent port level VLACP configuration.

- Configure VLACP on an individual port basis.

  The port can be either an individual port or an MLT member. Each VLACP-enabled port periodically sends VLACP PDUs. This action allows the exchange of VLACP PDUs from an end-to-end perspective. If a particular link does not receive VLACP PDUs, the platform shuts the link down after the expiry timeout occurs (timeout scale x periodic time). This action implies that unless you enable VLACP on the IST peer, the ports stay in a disabled state. When you enable VLACP at the IST peer, the VLACP PDU is received and the ports are re-enabled. You can replicate this behavior despite the IST connectivity between the end-to-end peers. When you enable VLACP on the IST ports at one end of the IST, the ports are taken down along with the IST. However, the IST at the other end remains active until the expiry timeout occurs on the other end. As soon you enable VLACP at the other end, the VLACP PDU is received by the peer and the ports are brought up at the software level.

Multilink Trunking

Use MLT to provide link layer redundancy. You can use MLT to provide alternate paths around failed links. When you configure MLT links, consider the following information:

- The device supports 512 MLT aggregation groups.
- Up to 16 ports can belong to a single MLT group.

MLT and LACP groups and port speed

Ensure that all ports that belong to the same MLT or LACP group use the same port speed, for example, 1 Gb/s, even if you use Auto-Negotiation. The software does not enforce this requirement. Avaya recommends that you use Custom Auto-Negotiation Advertisement (CANA) to ensure proper speed negotiation in mixed-port type scenarios.

To maintain Link Aggregation Group (LAG) stability during failover, use CANA: configure the advertised speed to be the same for all LACP links. For 10/100/1000 ports, ensure that CANA uses one particular setting, for example, 1000-full or 100-full. Otherwise, a remote device can restart Auto-Negotiation and the link can use a different capability.

Each port must use only one speed and duplex mode; all links in Up state are guaranteed to have the same capabilities. If you do not use Auto-Negotiation and CANA, you must use the same speed and duplex mode settings on all ports of the MLT.

Platform-to-platform MLT link recommendations

Avaya recommends that you connect physical connections in platform-to-platform MLT and link aggregation links in a specific order. To connect an MLT link between two platforms, connect the lower number port on one platform with the lower number port on the other platform. For example, to establish an MLT platform-to-platform link between ports 3/1 and 4/8 on platform A with ports 7/4 and 8/1 on platform B, do the following:

- Connect port 3/1 on platform A to port 7/4 on platform B
- Connect port 4/8 on platform A to port 8/1 on platform B

In Virtual Services Platform 9000, brouter ports do not support MLT. You cannot use brouter ports to connect two platforms with an MLT. An alternative is to use a VLAN. This configuration option provides a routed VLAN with a single logical port or MLT.

MLT and spanning tree protocols

The implementation of 802.1w (Rapid Spanning Tree Protocol—RSTP) and 802.1s (Multiple Spanning Tree Protocol—MSTP), provides a path cost calculation method. The following table provides the path costs associated with each interface type:

Table 8: Path cost for RSTP or MSTP mode

Link speed	Recommended path cost
Less than or equal 100 Kb/s	200 000 000
1 Mb/s	20 000 000
10 Mb/s	2 000 000
100 Mb/s	200 000
1 Gb/s	20 000
10 Gb/s	2000
100 Gb/s	200
1 Tb/s	20
10 Tb/s	2

802.3ad-based link aggregation

Link aggregation provides link layer redundancy. Use IEEE 802.3ad-based link aggregation (IEEE 802.3 2002 clause 43) to aggregate one or more links together to form LAGs to allow a MAC client to treat the LAG as if it were a single link. Use link aggregation to increase aggregate throughput of the interconnection between devices and provide link redundancy. LACP can dynamically add or remove LAG ports, depending on their availability and states.

Although IEEE 802.3ad-based link aggregation and MLT provide similar services, MLT is statically defined. By contrast, IEEE 802.3ad-based link aggregation is dynamic and provides additional functionality.

LACP and MLT

When you configure standards-based link aggregation, you must enable the aggregatable parameter. This configuration creates a one-to-one mapping between the LACP aggregator and the specified MLT.

A newly-created MLT or LAG adopts the VLAN membership of its member ports after the first port attaches to the aggregator associated with this LAG. After a port detaches from an aggregator, the port is deleted from the associated LAG port member list. After the last port member is deleted from the LAG, the LAG is deleted from all VLANs.

After you configure the MLT as aggregatable, you cannot add or delete ports or VLANs manually.

To enable tagging on ports that belong to a LAG, first disable LACP on the port, enable tagging on the port, and then enable LACP.

LACP and spanning tree interaction

Only the physical link state or the LACP peer status affects the operation of LACP. After a link goes up and down, the LACP module receives notification. The spanning tree forwarding state does not affect the operation of the LACP module. LACP data units (LACPDU) can be sent even if the port is in spanning tree blocking state.

Configuration changes (such as speed, duplex mode, and so on) made to a LAG member port do not apply to all the member ports of the MLT. Instead, the changed port is removed from the LAG, and the corresponding aggregator and user is alerted.

In contrast to MLT, IEEE 802.3ad-based link aggregation does not require the system to replicate BPDUs over all ports in the trunk group.

LACP and minimum link

The minimum link function defines the minimum number of active links required for a LAG to remain in the forwarding state. You cannot configure the minimum link on Virtual Services Platform 9000. The minimum link value is always 1.

If the number of active links in a LAG is 0, the entire LAG is declared down and Virtual Services Platform 9000 informs the remote end of the LAG state by using an LACPDU.

Link aggregation group rules

Link aggregation is compatible with RSTP and MSTP. LAGs operate using the following rules:

- All ports in a LAG must operate in full-duplex mode.
- All ports in a LAG must use the same data rate.
- All ports in a LAG must be in the same VLANs.
- Ports in a LAG can exist on different modules.
- LAGs form using LACP.
- The platform supports a maximum of 128 LAGs.
- Each LAG supports a maximum of eight active links.

For LACP fundamentals and configuration procedures, see *Avaya Virtual Services Platform 9000 Configuration — Link Aggregation, MLT, and SMLT*, NN46250-503.

Chapter 7: Redundant network design

This section provides guidelines that help you design redundant networks. Redundancy eliminates a single point of failure in your network.

Network Load Balancing

Network Load Balancing is a clustering technology available with Microsoft Windows 2000 and Windows 2003 Server family of operating systems. Network Load Balancing uses a distributed algorithm to load balance TCP/IP network traffic across a number of hosts, enhancing the scalability and availability of mission critical, IP based services, for example, Web and streaming media. Network Load Balancing also provides high availability by detecting host failures and automatically redistributing traffic to remaining operational hosts.

Each host runs separate copies of the desired server applications, for example, a Web server or a File Transfer Protocol (FTP) server. Network Load Balancing distributes incoming client requests to the hosts in the cluster group. You can configure the load weight that each host handles and you can add or remove hosts dynamically from the cluster as necessary. Network Load Balancing can direct all traffic to a designated single host, called the default host.

No restrictions exist on the number of network adapters that can be bound to network load balancing on a host computer. Each host can have a different number of adapters as long as the adapters are part of different clusters. A host cannot have more than one adapter that is in the same cluster.

Important:

Network Load Balancing does not support a mixed unicast and multicast environment within a single cluster. Within each cluster, all network adapters in that cluster must be either multicast or unicast; otherwise, the cluster does not function properly.

Virtual Services Platform 9000 supports the Network Load Balancing topologies in the following figures. Use the topology in the following figure to deploy Network Load Balancing clusters in unicast, multicast, and Internet Group Management Protocol (IGMP)-multicast modes without routing. Do not enable IP routing on the switching platform.

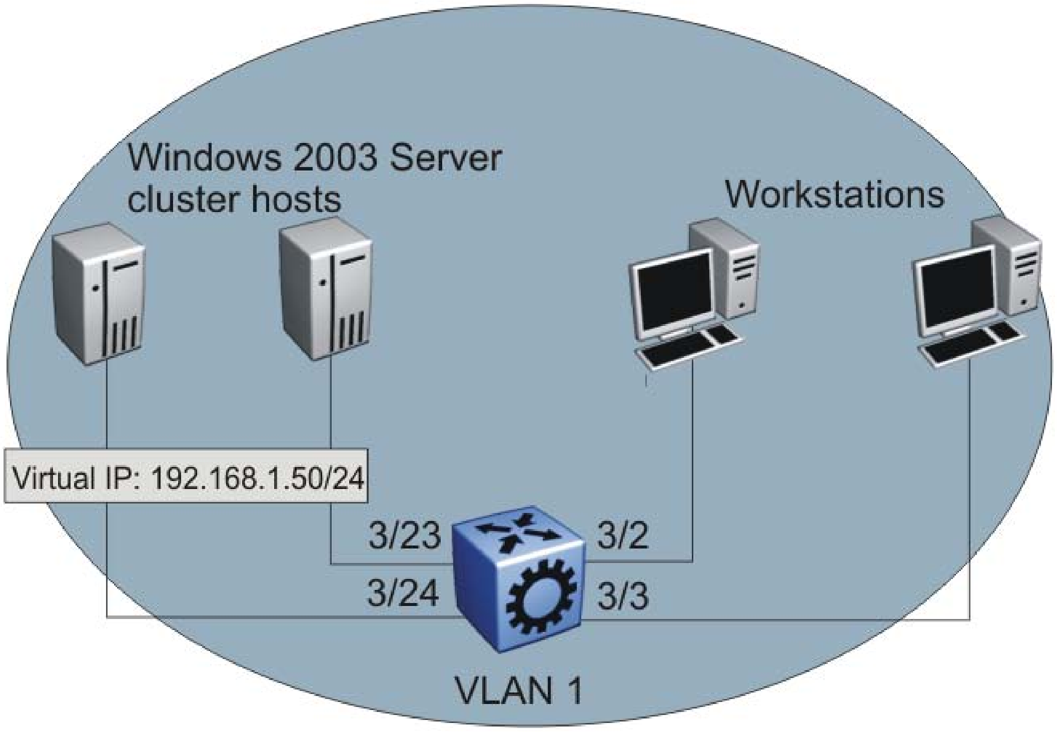
Windows 2003 Server cluster hosts
Workstations
Virtual IP: 192.168.1.50/24
3/23
3/24
3/2
3/3
VLAN 1

Figure 5: Network server load balancing supported topology 1

In the preceding figure, Virtual Services Platform 9000 floods Network Load Balancing cluster traffic by default. If you deploy the clusters using multicast or IGMP-multicast mode, you can optionally enable IGMP snooping and proxy to eliminate the flooding of cluster traffic to noncluster hosts.

In the following figure, Virtual Services Platform 9000 performs routing between server and client VLANs and both the cluster hosts and clients connect directly to the platform. Enable IP routing. This topology supports Network Load Balancing clusters in unicast, multicast, and IGMP-multicast modes.

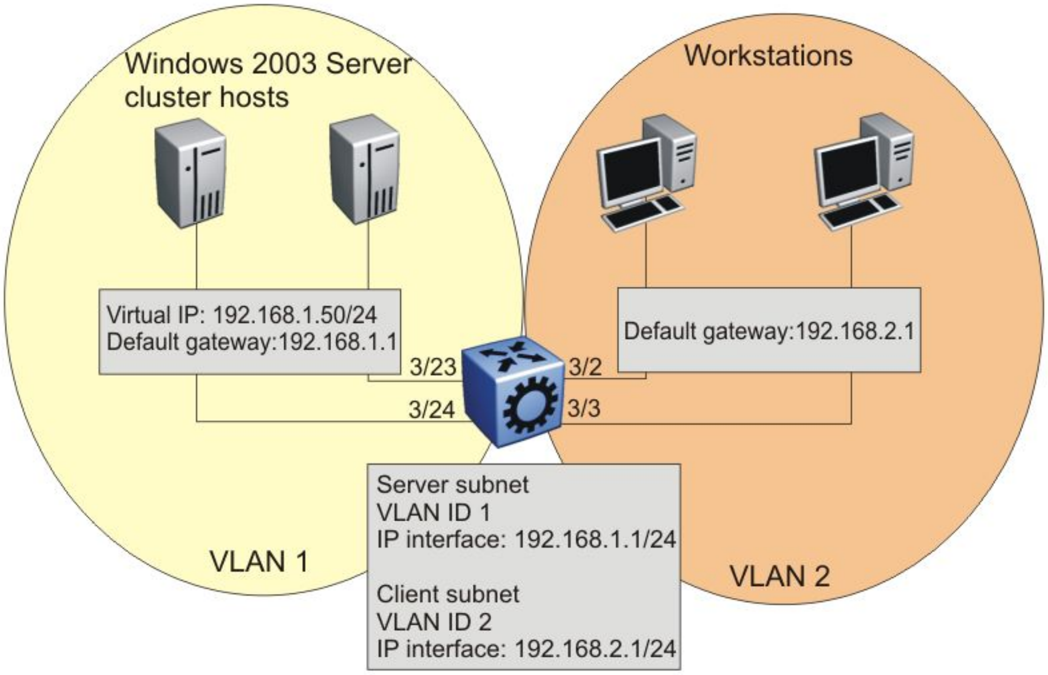
Windows 2003 Server cluster hosts
Workstations
Virtual IP: 192.168.1.50/24
Default gateway:192.168.1.1
Default gateway:192.168.2.1
3/23
3/24
3/2
3/3
Server subnet
VLAN ID 1
IP interface: 192.168.1.1/24
Client subnet
VLAN ID 2
IP interface: 192.168.2.1/24
VLAN 1
VLAN 2

Figure 6: Network server load balancing supported topology 2

In the preceding topology example, you must perform the following configurations:

- unicast mode: individual NLB unicast support
- multicast mode: individual NLB multicast support
- IGMP-multicast mode: individual NLB IGMP-multicast support

For more information about Network Load Balancing, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500, and *Technical Configuration Guide for Microsoft Network Load Balancing*, NN48500–593.

Chapter 8: Lossless networks

This section includes recommendations to configure a Lossless environment to guarantee the switch does not drop certain traffic types.

Introduction

On Virtual Services Platform 9000, you can configure 10 GbE ports to be Lossless. You need to purchase and install a Premier license in order to configure Lossless.

You can configure all unicast traffic on the port to be Lossless, or you can configure only unicast traffic with a specific Lossless 802.1p value to be Lossless.

Note:

Multicast is not supported on 10GbE ports configured as Lossless.

To configure the Lossless Ethernet feature, you enable it on a port and configure the pause type. The pause type can be one of two values:

- pause (FC)

  All unicast traffic on the port is Lossless. FC is the default pause type.

- priority-based flow control (PFC)

  All unicast tagged-traffic on the port with an 802.1p value that matches the Lossless-802.1p value is Lossless. The switch treats traffic that does not meet this requirement as Lossy traffic. The switch can drop the Lossy traffic.

  You can configure the Lossless-802.1p value; the default value is 3. If you change the Lossless-802.1p value, the change affects all PFC-configured ports. You cannot change this value for individual Lossless ports.

  For Lossless-PFC, you can also configure the guaranteed Lossless bandwidth for PFC-configured ports. The default Lossless bandwidth is 50% of the port bandwidth. You can configure the guaranteed Lossless bandwidth in steps of 1Gb within the range of 2Gb to 8Gb.

The Lossless Ethernet feature applies to both directions on a Lossless port. If the port becomes congested, the sender is flow controlled by sending FC or PFC frame based on the pause-type setting.

If traffic enters the device on a PFC-enabled port and exits the device on an FC-enabled port, only packets with the matching Lossless-802.1p value are Lossless. The following table shows which packets are Lossless in specific ingress and egress situations.

Table 9: Lossless treatment based on pause type

Ingress port	Egress port	Lossless guaranteed
PFC	PFC	802.1p only
PFC	FC	802.1p only

Ingress port	Egress port	Lossless guaranteed
FC	PFC	802.1p only
FC	FC	all

General configuration limitations

The following list identifies configuration limitations for Lossless Ethernet:

- If you configure a port as Lossless-PFC, it does not generate nor react to FC frames.
- If you configure a port as Lossless-FC, it does not generate nor react to PFC frames.
- You cannot configure an interface policer and Lossless Ethernet on the same port.
- If you configure port mirroring or packet capture (PCAP) on a Lossless port, the system does not guarantee Lossless behavior.
- The chassis must use at least five Switch Fabric modules to support the Lossless Ethernet feature.

Interface module configuration limitations

The following list identifies configuration limitations for interface module support:

- You can configure Lossless Ethernet only on 10 GbE ports.

  You cannot configure Lossless Ethernet on the 9048GB or 9048GT modules. You cannot configure Lossless Ethernet on a 10GbE port if you install a 1 GbE SFP.

  Only a 9024XL module with a MAC version of BCM56820-B0 support Lossless-PFC mode. Use the command **`show sys-info asic`** to view the MAC version of the module.
- The ports on the 9024XL module divide into three 8–port clusters:

  - ports 1–8
  - ports 9–16
  - ports 17–24

  You can enable Lossless Ethernet on a maximum of two ports per cluster. Within a cluster, one port must be in the top row of the cluster, and one port must be in the bottom row of the cluster. You cannot enable Lossless Ethernet on two ports in the same row of a cluster. The following figure identifies the clusters for the module.

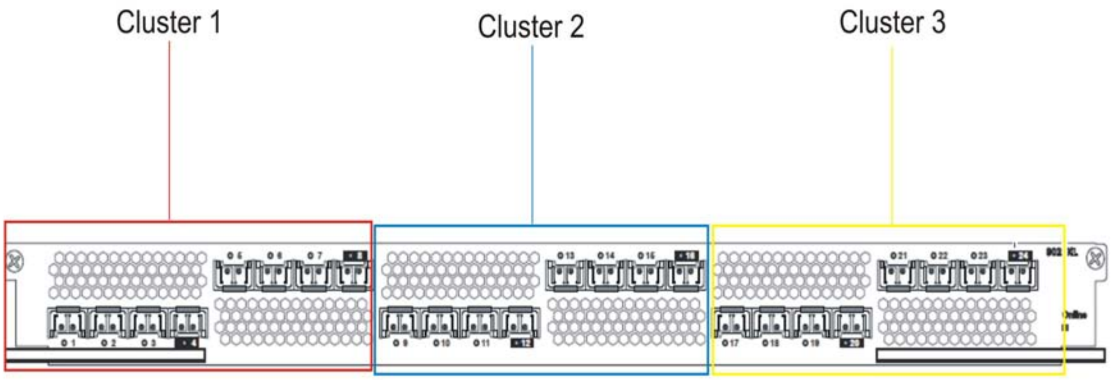

Figure 7: 9024XL Lossless Ethernet clusters

- A cluster is considered Lossless if you enable Lossless Ethernet on at least one port in the cluster.
- The pause type of two Lossless ports in the same cluster must be the same. If you configure the second port in a cluster with a different pause type, the system prompts you to apply that change to the first Lossless port in the cluster.
- A maximum of two ports can be operational in a Lossless cluster: either two Lossless ports, or one Lossless port and one Lossy port. If a port in a cluster half is Lossless enabled, the other three ports in that half will be held operationally down.
- If you do not enable Lossless Ethernet on any ports in a Lossless cluster half, only the lowest numerical port, which is administratively enabled, is operational. The remaining three ports in the cluster half are operationally down.
- After you disable Lossless Ethernet on one half of a cluster, the system allows one of the four ports in that half to be up operationally in Lossy mode.
- After you disable Lossless Ethernet on the last port in a cluster, all eight ports return to the operational state prior to Lossless Ethernet configuration.
- Avaya recommends that you configure two Lossless ports in a cluster only if you expect bursts of egress oversubscription that span a period of a few seconds. If you expect sustained or prolonged egress oversubscription, Avaya recommends that you configure one port in the cluster as Lossless and one port as Lossy.

You cannot enable Lossless-PFC on a port if the port QoS value equals the Lossless 802.1p value. On a Lossless-PFC enabled port, you cannot set the port QoS and the Lossless 802.1p to the same value.

PFC configuration limitations

You cannot configure Layer 2 untrusted on a Lossless port. Layer 2 is always trusted for a Lossless port. If you enable Lossless on a Layer 2-untrusted port, the port will be forced to be Layer 2 trusted. If you then disable Lossless on that port, Layer 2 will remain trusted on that port.

The following list identifies configuration limitations for the Lossless-802.1p value:

- You must configure the internal QoS level that corresponds to the Lossless-802.1p value to 3. Use the **`qos ingressmap 1p`** command to change the internal QoS level. Avaya recommends that you do not use filters to remark the internal QoS.
- If you change the Lossless-802.1p value, the change does not apply if Lossless-802.1p maps to an internal QoS level other than 3, or if the internal QoS level 3 maps to another 802.1p value.
- Even if you use the default Lossless-802.1p value, a port does not become Lossless-PFC if it maps to an internal QoS level other than 3, or if the internal QoS level 3 maps to another 802.1p value.
- The system displays a warning message if at least one cluster is in Lossless-PFC mode, and you attempt to map a non-Lossless 802.1p value to internal QoS level 3, or map the Lossless 802.1p value to an internal QoS level other than 3.
- You cannot change the Lossless 802.1p value to equal the configured port QoS value on any of the Lossless-PFC enabled ports in the system.

You cannot configure Layer 3 trusted on a Lossless port. Layer 3 is always untrusted for a Lossless port. If you enable Lossless on a Layer 3 trusted port, the system forces the port to be Layer 3 untrusted. If you disable Lossless on that port, Layer 3 will become trusted.

Lossless is not guaranteed on IP packets that have the DSCP value set to one of the AVAYA-on-AVAYA values(47, 41, 35 and 27). AVAYA-ON-AVAYA was known as NT-on-NT.

In a Lossless-PFC (802.1Qbb) domain, the Lossless behavior is guaranteed as long as the Lossless 802.1p, ingress "1p to QoS" map and the egress "QoS to 1p" maps are consistent. You must configure the egress "QoS to 1p" map correctly when you change the Lossless 802.1p and Ingress "1P to QoS" maps.

MLT configuration limitations

The following list identifies configuration limitations if you use Lossless ports in an MLT:

- You cannot add a port that operates at 1Gb speed to an MLT with Lossless port members.
- If an MLT has Lossless port members, you can add a regular port to the MLT only if you can enable the Lossless Ethernet feature on that port.

  Conversely, if an MLT has regular ports, you can enable Lossless Ethernet on a port member only if you can enable the Lossless Ethernet feature on all MLT ports. Enabling or disabling Lossless applies to all the ports of the MLT. If you change the pause type or PFC-bandwidth on an MLT port, the change does not apply to other ports in the MLT.
- If you add a Lossless port to an MLT with regular port members, the system disables Lossless Ethernet on the port.
- Avaya recommends that all Lossless ports in an MLT use the same pause type. The system does not guarantee Lossless behavior if Lossless ports in an MLT use different pause types.
- If you remove a Lossless port from an MLT, Lossless Ethernet remains enabled on the port.
- You must configure MLT ports that use the same LACP key with the same Lossless configuration.

  The LACP key does not apply to a port if the Lossless configuration does not match that of other ports with the same LACP key.

- You cannot change the Lossless configuration on an MLT port if it has an LACP key.
- If you insert an SFP in a Lossless port, the system shows the Lossless Ethernet feature as enabled on the port, and the MLT port membership remains, but the port is down.

Lossless deployments

The following figure shows pause-frame support over a single hop. All links are 10 GbE. The egress port can become over subscribed. Depending on the traffic rates on the ingress ports, over subscription duration, and the location of ingress ports (whether on the same 8-port-cluster or not), the pause frames will be generated accordingly.

The orange and red data paths cause congestion on the link to server A. The Lossless configuration sends a pause to the senders that congest the port, and does not affect the blue data path. The benefit of this deployment is that only offenders are paused, and no head-of-line blocking occurs.

Avaya recommends that you configure two Lossless ports in a cluster only if you expect bursts of egress oversubscription that span a period of a few seconds. If you expect sustained or prolonged egress oversubscription, Avaya recommends that you configure one port in the cluster as Lossless and one port as Lossy.

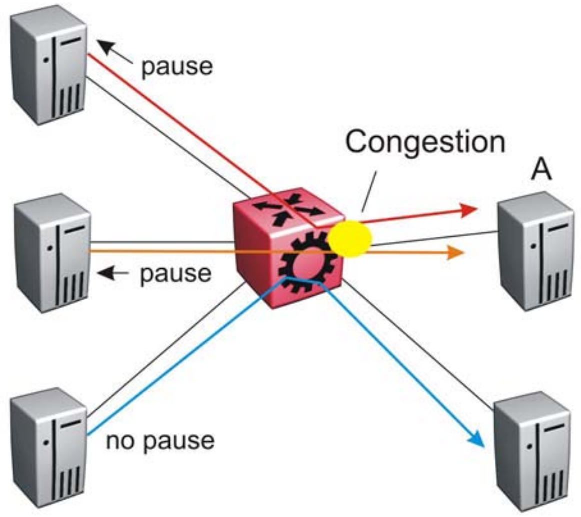

Figure 8: Pause frame over a single hop

The following figure shows pause-frame support over two hops. All links are 10 GbE. An MLT exists on the two links that connect the VSP 9000 to the Layer 3 switch. The egress port can become over subscribed. Depending on the traffic rates on the ingress ports, over subscription duration, and the location of ingress ports (whether on the same 8-port-cluster or not), the pause frames will be generated accordingly.

The orange and red data paths cause congestion on the link to the Layer 3 switch, and on the link from the Layer 3 switch to server A. The Lossless configuration sends a pause to the VSP 9000 and the senders, and does not affect the blue data path.

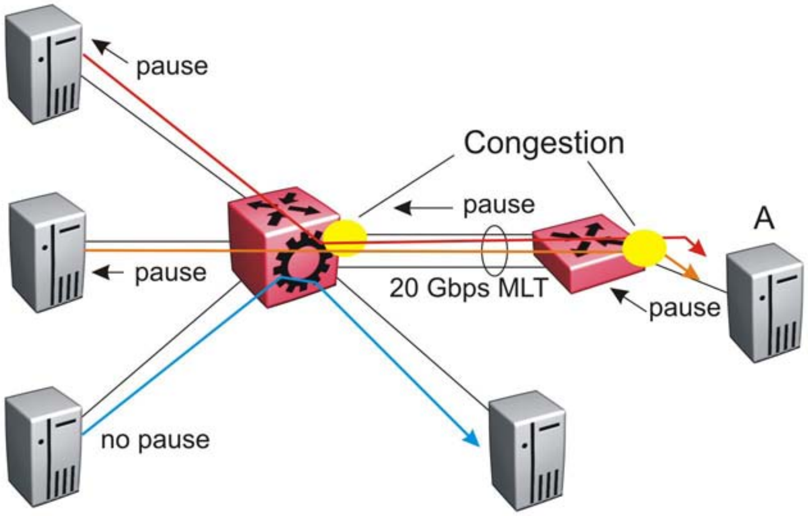

Figure 9: Pause frame over two hops

The following figure shows PFC-frame support over a single hop. All links are 10 GbE. The egress port can become over subscribed. Depending on the traffic rates on the ingress ports, over subscription duration, and the location of ingress ports (whether on the same 8-port-cluster or not), the PFC frames will be generated accordingly. Lossy traffic can potentially get dropped but the ingress rate will not be affected by the PFC frames sent by the system.

The orange and red data paths cause congestion on the link to server A. The red and orange data paths flow through the PFC queue. The blue data paths do not flow through the PFC queue. The Lossless configuration sends a PFC frame to the senders of packets that match the Lossless-802.1p value, drops packets that do not match the Lossless-802.1p value, and does not affect the blue data paths.

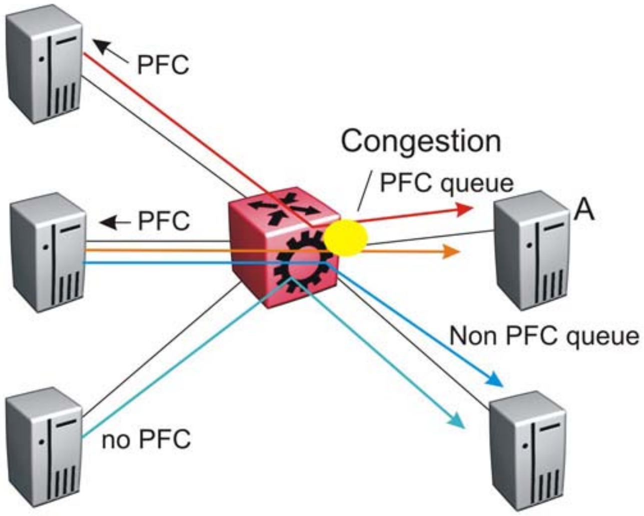

Figure 10: PFC frame over single hop

The following figure shows PFC-frame support over two hops. All links are 10 GbE. An MLT exists on the two links that connect VSP-A to VSP-B. The egress port can become over subscribed. Depending on the traffic rates on the ingress ports, over subscription duration, and the location of ingress ports (whether on the same 8-port-cluster or not), the PFC frames will be generated accordingly. Lossy traffic can potentially get dropped but the ingress rate will not be affected by the PFC frames sent by the system.

The orange and red data paths cause congestion on the link to VSP-B, and on the link from VSP-B to server A. The red and orange data paths flow through the PFC queue. The blue data paths do not flow through the PFC queue. The Lossless configuration sends a PFC frame to the senders of packets that match the Lossless-802.1p value, drops packets that do not match the Lossless-802.1p value, and does not affect the blue data paths.

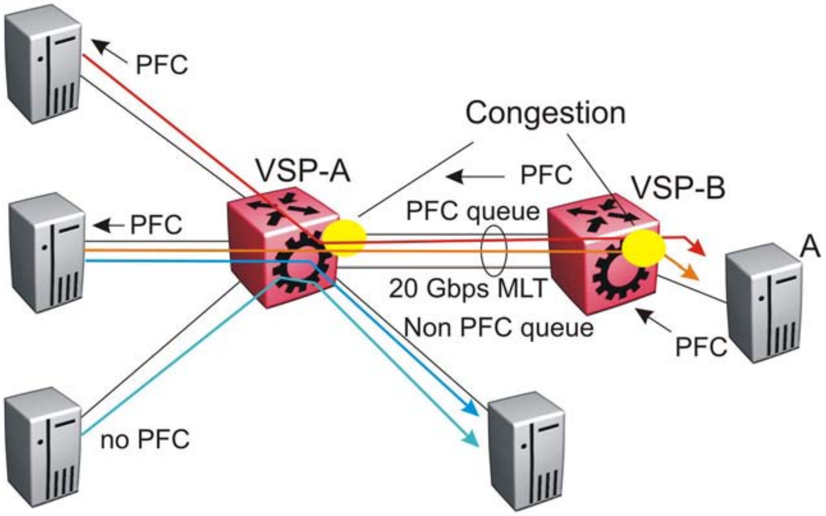

Figure 11: PFC frame over two hops

The following figures show two data-center deployments: a one-switch core and a two-switch core. All links use Lossless mode (802.3X) pause frames. This deployment provides multihop Lossless within the core and at the edge.

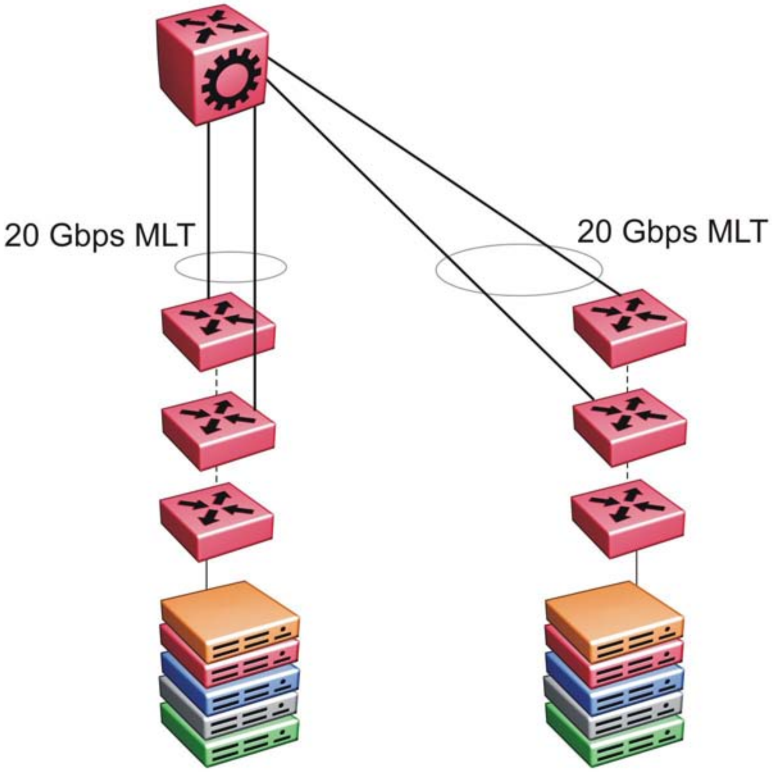

Figure 12: Data-center deployment with one-switch core

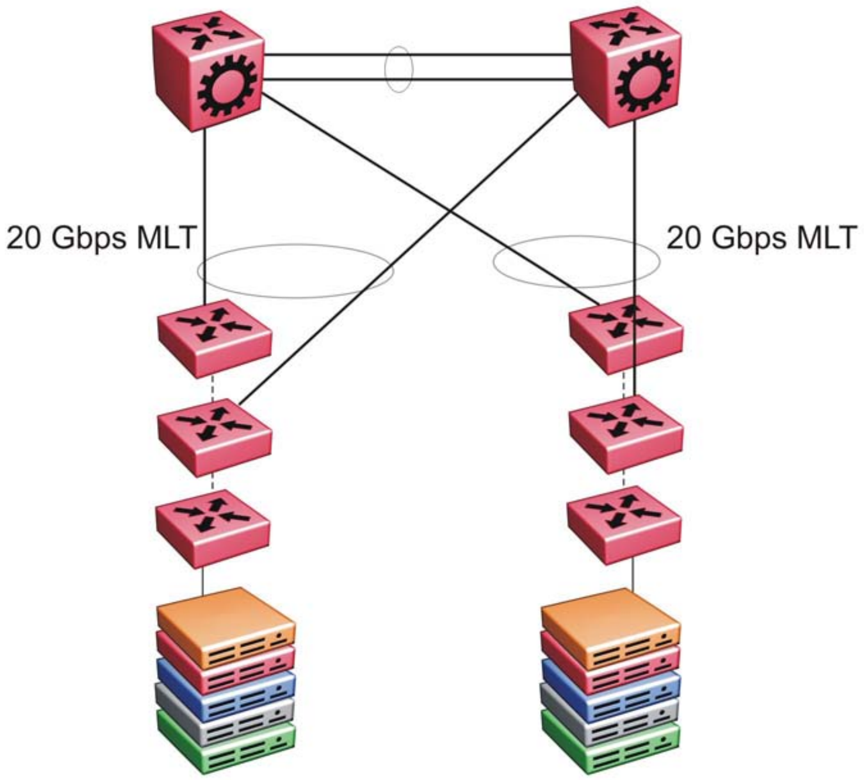

Figure 13: Data-center deployment with two-switch core

Chapter 9: Layer 2 switch clustering and SMLT

Split MultiLink Trunking (SMLT) enables node redundancy by allowing aggregated link groups to be dual-homed across a pair of aggregating devices. This introduces an extra level of redundancy and failure protection. SMLT is introduced into existing subnetworks to provide this redundancy without the need to upgrade installed equipment. Bandwidth availability and network resiliency are improved by allowing all aggregation paths in a dual-homed configuration to be active and to forward traffic. In the event of a link failure, traffic failover is fast. An SMLT aggregation device pair uses an interswitch trunk (IST) to exchange information and appear as a single, logical path aggregation end point to dual-homed devices. IST signalling protects against single points of failure such as link outages by detecting and modifying information about forwarding data paths.

The following sections describe SMLT and its implementation.

Modular design for redundant networks

Network designs typically depend on the physical location and fiber and copper cable layout of an area. Avaya recommends approaching network design from a modular approach. Modular network design entails that you break a design in different sections that can be replicated as necessary; several functional layers or tiers. When you design functional tiers, consider campus architectures separately from data center architectures.

Campus architecture

A three tier campus architecture consists of an edge, distribution, and core layer.

- **Edge Layer**

  The edge layer provides direct connections to end user devices. These devices are normally the wiring closet switches that connect devices such as PCs, IP Deskphones, and printers.

- **Distribution Layer**

  The distribution layer provides connections between edge and core layer devices.

- **Core Layer**

  The core layer is the center of the network. In a three-tier architecture, all distribution layer switches terminate in the core. In a two-tier architecture, the edge layer terminates directly in the core and no distribution layer is required.

Important:

Avaya recommends against directly connecting servers and clients in core switches. If one IST switch fails, connectivity to the server is lost.

Data center architecture

The modular approach also applies to data center architectures. In this case, the core and distribution layers provide similar functions to those in a campus architecture while a server access layer replaces the edge layer.

- **Server Access Layer**

  The server access layer provides direct connections to servers.
- **Distribution Layer**

  The distribution layer provides connections between the server access and core layers.
- **Core Layer**

  The core layer is the center of the network. In a three-tier architecture, all distribution layer switches terminate in the core. In a two-tier architecture, the server access layer terminates directly in the core and no distribution layer is required.

Example network layouts

The following figure displays a sample network that incorporates the elements of campus and data center architectures. When designing networks, keep in mind most network topologies have overlap between these architectures.

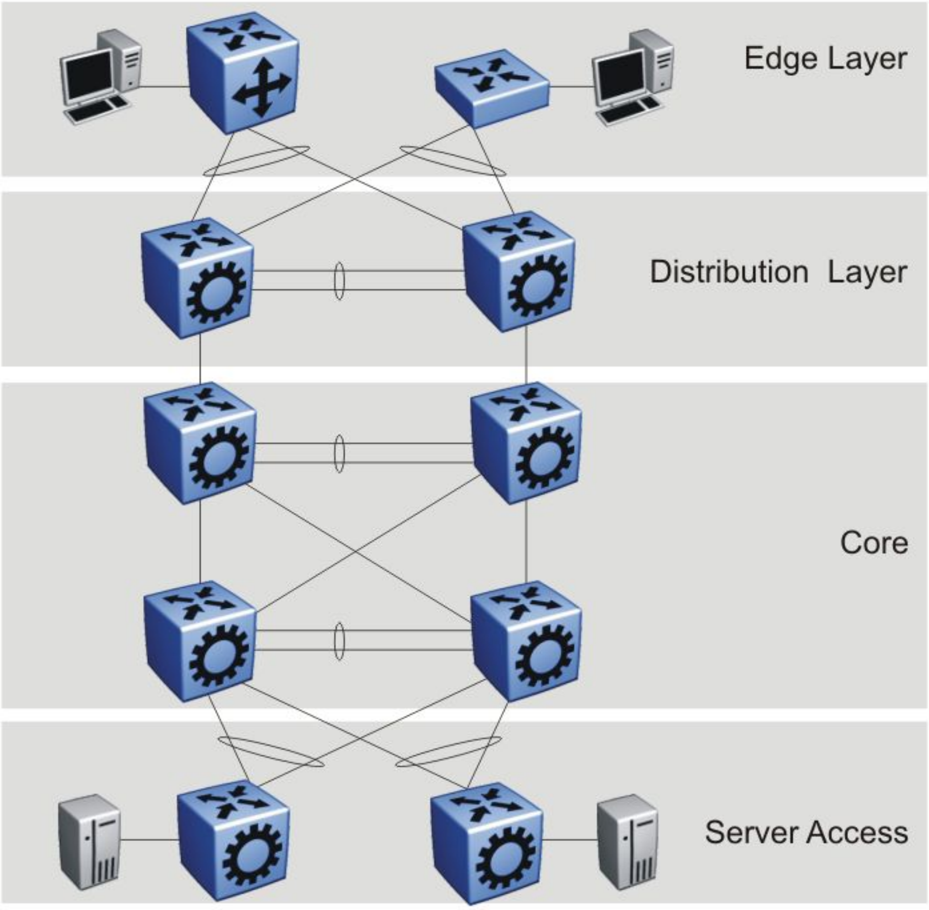

Figure 14: Three-tiered architecture plus data center

In many cases, the distribution layer can be removed from the campus network layout. This configuration maintains functionality but decreases cost, complexity, and network latency. The following figure shows an architecture where the edge layer connects directly into the core.

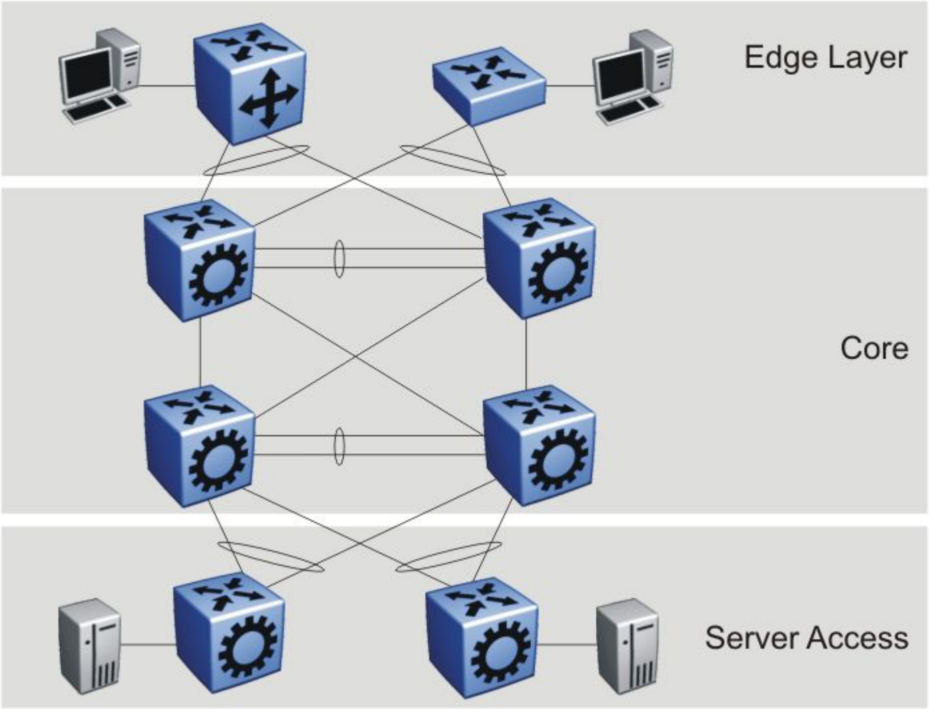

Figure 15: Two-tiered architecture with four-switch core plus data center

The following figure shows a two-tiered architecture with a two-switch core.

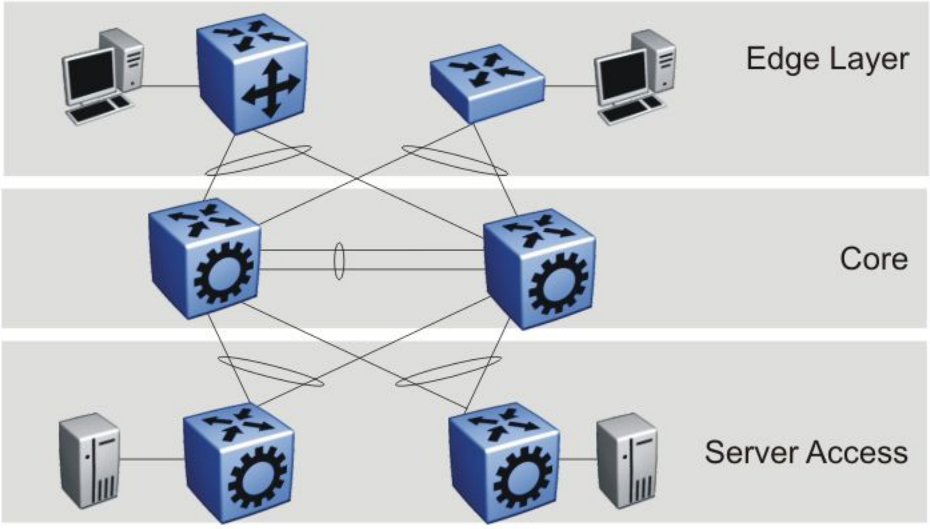

Figure 16: Two-tiered architecture with two-switch core plus data center

For more information about specific design and configuration parameters refer to *The Large Campus Technical Solution Guide*, NN48500-575 and *Switch Clustering using Split Multilink Trunking (SMLT) Technical Configuration Guide*, NN48500-518.

Network edge redundancy

The following figure depicts a switch pair at the distribution layer that provides riser links to wiring closets. If one edge layer switch fails, the other can maintain user services.

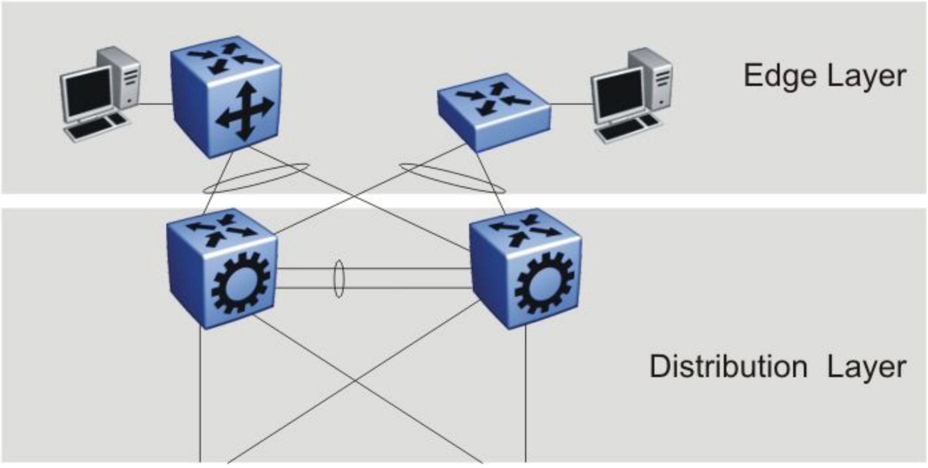

Figure 17: Redundant network edge diagram

Avaya recommends the following network edge design. This configuration is easy to implement and maintain and provides redundancy if one of the edge or distribution layer switches fails.

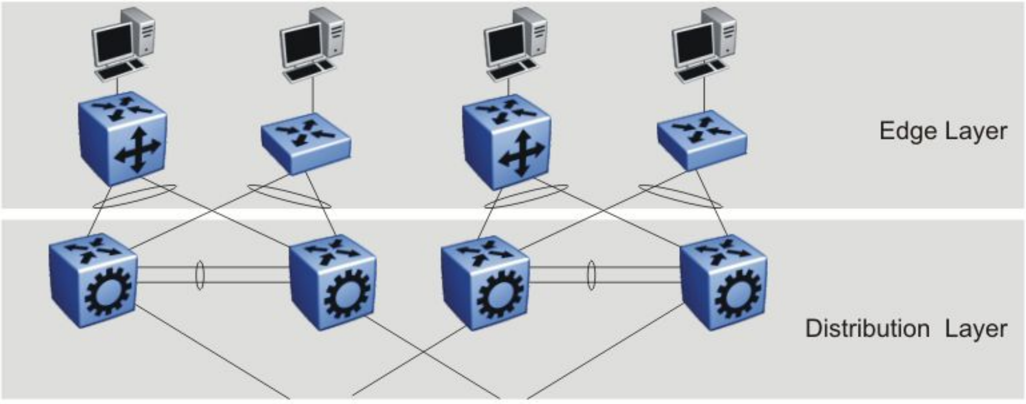

Figure 18: Recommended network edge design

Split MultiLink Trunk configuration

SMLT improves Layer 2 resiliency by providing switch failure redundancy with subsecond failover in addition to standard MLT link failure protection and flexible bandwidth scaling functionality. Use SMLT to connect a device that supports link aggregation to two distinct SMLT endpoints to form a triangle. These SMLT switches form a switch cluster and are referred to as an IST core switch pair.

Switch clusters are always formed as a pair but you can combine pairs of clusters in either a square or full-mesh fashion to increase the size and port density of the switch cluster.

SMLT redundancy

The following figure demonstrates an SMLT triangle configuration with two Avaya Virtual Services Platform 9000 devices acting as aggregation switches (E and F). Four MLT compatible wiring closet switches labeled A, B, C, and D.

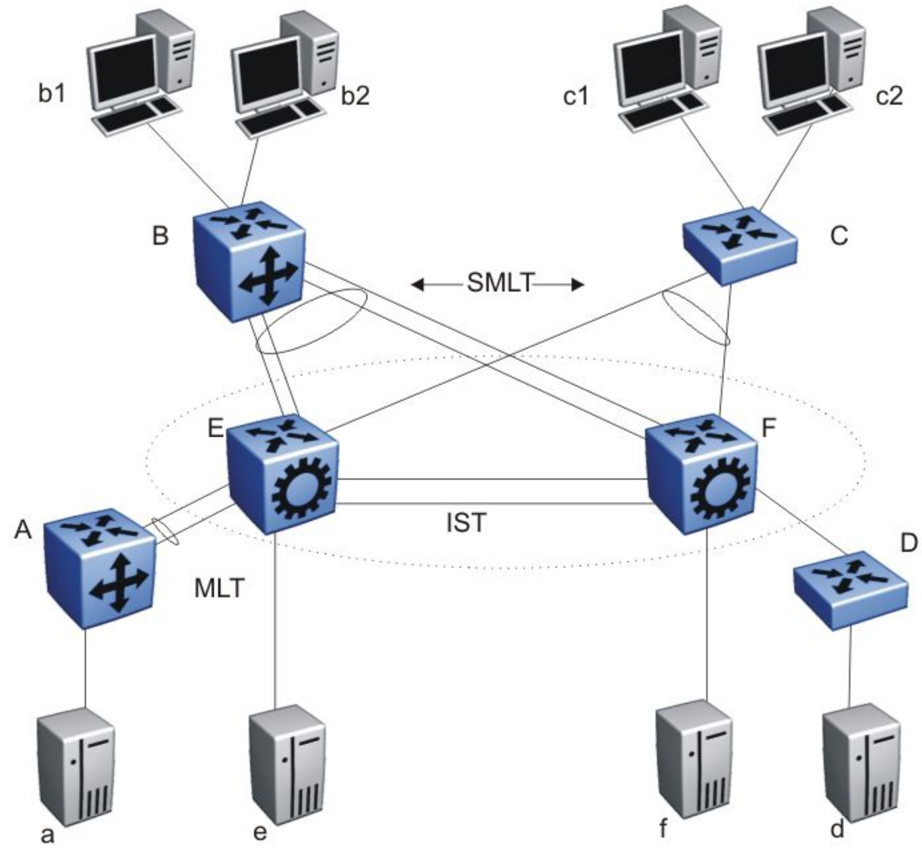

Figure 19: SMLT triangle configuration

B and C connect to the aggregation switches through multilink trunks split between two VSP 9000 devices. SMLT edge switch B can use two parallel links for its connection to E and two additional parallel links for its connection to F. This configuration provides redundancy.

SMLT edge switch C has only a single link to both E and F. Switch A is configured for MLT but the MLT terminates on only one switch in the network core. Switch D has a single connection to the core. Although you can configure both switch A and switch D to terminate across both of the aggregation switches using SMLT, neither switch benefits from SMLT in this network configuration.

The SMLT edge switches are dual-homed to the aggregation switches yet they require no knowledge of whether they connect to a single switch or to two switches. You need SMLT only on the aggregation switches. Logically, E and F appear as a single switch to the edge switches. Therefore, the SMLT edge switches only require an MLT configuration. The connection between the SMLT aggregation switches and the SMLT edge switches are the SMLT links.

Virtual Services Platform 9000 supports all interfaces as operational SMLT links.

SMLT and VLACP

Avaya recommends the use of Virtual Link Aggregation Control Protocol (VLACP) for all SMLT access links configured as MultiLink Trunks to ensure both end devices can communicate.

Virtual Services Platform 9000 does not support LACP and VLACP on the same links simultaneously.

VLACP for SMLT also protects against CPU failures by causing traffic to switch or reroute to the SMLT peer if the CPU fails or stops responding.

The following table provides the recommended values for VLACP in an SMLT environment:

Table 10: Recommended VLACP values

Parameter	Value
SMLT access	
Timeout	Short
Timer	500ms
Timeout scale	5
VLACP MAC	01:80:C2:00:00:0F
SMLT core	
Timeout	Short
Timer	500ms
Timeout scale	5
VLACP MAC	01:80:C2:00:00:0F
IST	
Timeout	Long
Timer	10000
Timeout scale	3
VLACP MAC	01:80:C2:00:00:0F

SMLT and loop prevention

SMLT-based network designs form physical loops for redundancy that logically do not function as loops. Under certain adverse conditions, loops can form, for example, if you use incorrect configurations or cabling.

The two solutions to detect loops are Loop Detect and Simple Loop Prevention Protocol (SLPP). Loop Detect and SLPP detect a loop and automatically stop the loop. Both solutions determine on which port the loop occurs, and shut down that port.

Interswitch Trunking recommendations

Figure 19: SMLT triangle configuration on page 55 shows that SMLT requires only two SMLT capable aggregation switches connected by an interswitch trunk. The aggregation switches use the interswitch trunk to perform the following functions:

- Confirm that each switch is alive and exchanges MAC address information. The link must be reliable and must not exhibit a single point of failure in itself.
- Forward flooded packets or packets destined for non-SMLT connected switches, or for servers that physically connect to the other aggregation switch.
- Forward traffic in a failover scenario so traffic can pass through the interswitch trunk.

To ensure the proper and optimal operation of an IST, Avaya recommends the following items:

- The amount of traffic that requires forwarding across the interswitch trunk from a single SMLT wiring-closet switch is usually small. However, if the aggregation switches terminate connections to single-homed devices, or if uplink SMLT failures occur, the interswitch trunk traffic volume can be significant. To ensure that no single point of failure exists in the interswitch trunk, configure it to have multiple links with connections spread across different modules on both aggregation switches.
- The SMLT aggregation switches establish an IST session based on common VLAN membership and knowledge of the peer switch IP address. Use a dedicated VLAN for this IST peer session. Ensure the VLAN chosen for the peer session is dedicated to the purpose by only including interswitch trunk ports in its membership.
- Do not enable dynamic routing protocols on the IST VLAN. The IST VLAN is meant to support adjacent switches. In most cases, you must not use the IST VLAN as a next hop route for non-IST traffic or to route traffic. The only exception to this rule is multicast broadcasts, which use Protocol Independent Multicast — Sparse Mode (PIM-SM). In this instance, enable PIM-SM support on the VLAN.
- Use at least two physical ports, on different interface modules, in an IST. This is not mandatory but it does increase the bandwidth available to the IST, as well as its resiliency and redundancy.

SMLT and client and server applications

Do not use unbalanced client-server configurations where core switches are directly connected to servers or clients. Loss of one of the IST pair switches in such a configuration causes connectivity to the server to be lost.

SMLT and Layer 2 traffic load sharing

SMLT achieves load sharing on the edge switch using the MLT path selection algorithm. For more information about the algorithm, see *Avaya Virtual Services Platform 9000 Configuration — Link Aggregation, MLT, and SMLT*, NN46250-503. The algorithm typically operates on a source or destination IP or MAC address basis.

SMLT achieves load sharing on the aggregation switch by sending all traffic destined for the SMLT edge switch directly to it and not over the IST trunk. The IST trunk is never used to cross traffic to and from an SMLT dual-homed wiring closet. Traffic received on the IST by an aggregation switch is not forwarded to SMLT links (the other aggregation switch does this), thus eliminating the possibility of a network loop.

SMLT and Layer 3 traffic redundancy (VRRP and RSMLT)

VLANs that are part of an SMLT network can be routed on SMLT aggregation switches. Routing VLANs enables the SMLT edge network to connect to other Layer 3 networks. Virtual Router Redundancy Protocol (VRRP), which provides redundant default gateway configurations, additionally has BackupMaster capability. BackupMaster improves the Layer 3 capabilities of VRRP operating in conjunction with SMLT. Use a VRRP BackupMaster configuration with an SMLT configuration that currently uses VRRP.

Important:

Avaya strongly recommends using Routed SMLT (RSMLT) Layer 2 Edge configuration as a better alternative to SMLT with VRRP BackupMaster. Unless it is specifically required, use an RSMLT configuration.

RSMLT Layer 2 Edge configurations provide:

- Greater scalability — RSMLT scales to the maximum number of VLANs, while VRRP scales to 255 for each VRF and 512 for each system. VRRP IDs 1-255 are unique to each VLAN.
- Simpler configuration — A Routed SMLT Layer 2 Edge configuration only requires enabling RSMLT on a VLAN. VRRP requires virtual IP configuration along with other parameters.

For connections in pure Layer 3 configurations using a static or dynamic routing protocol, use a Layer 3 RSMLT configuration instead of SMLT with VRRP. RSMLT configuration provides faster failover than VRRP.

Important:

In an SMLT-VRRP environment that uses VRRP critical IP within both IST core switches, routing between directly connected subnets ceases to work when connections from each of the switches to the exit router (the critical IP) fail. Do not configure VRRP critical IPs within SMLT or RSMLT environments because SMLT operation automatically provides the same level of redundancy.

Do not use VRRP BackupMaster and critical IP at the same time; use one or the other. Do not use VRRP in RSMLT environments.

The VRRP Master typically forwards traffic for a given subnet. Use BackupMaster on the SMLT aggregation switch with a destination routing table entry and the Backup VRRP switch also routes traffic. The VRRP BackupMaster uses the VRRP standardized backup switch state machine. This makes the VRRP BackupMaster compatible with standard VRRP. This capability prevents the traffic from edge switches from unnecessarily utilizing the IST to deliver frames destined for a default gateway. In a traditional VRRP implementation, this operates only on one of the aggregation switches.

The BackupMaster switch routes all traffic received on the BackupMaster IP interface according to the switch routing table. The BackupMaster switch does not perform Layer 2 switching for the traffic to the VRRP Master.

Ensure that both SMLT aggregation switches can reach the same destinations using a given routing protocol. Configure individual VLAN IP addresses on both SMLT aggregation switches for routing purposes. Introduce an additional subnet on the IST that has a shortest-route path to avoid issuing Internet Control Message Protocol (ICMP) redirect messages on the VRRP subnets. To reach the destination, ICMP redirect messages are issued if the router sends a packet back out through the same subnet on which it is received.

SMLT failure and recovery

Traffic can cease if an SMLT link is lost. In such an instance, the SMLT edge switch detects the loss and sends traffic on the other SMLT links, as it does with standard MLT. If the link is not the only one between the SMLT client and the aggregation switch, the aggregation switch also uses standard MLT detection and rerouting to move traffic to the remaining links. If the link is the only route to the aggregation switch, the switch informs the other aggregation switch of the SMLT trunk failure. The other aggregation switch then treats the SMLT trunk as a regular multilink trunk. In this case, the MLT operational type changes from SMLT to NORMAL. If the link is reestablished, the aggregation switches detect it and resumes regular SMLT operations and the operational type returns to SMLT.

Traffic can also cease if an aggregation switch fails. If this happens, the SMLT edge switch detects the failure and sends traffic out on other SMLT links. The operational aggregation switch detects the loss of the partner IST. The SMLT trunks operate as normal MLT. If the partner switch returns, the operational aggregation switch detects it. The IST again becomes active, and after full connectivity establishes, the trunks are moved back to regular SMLT.

If an IST link fails, the SMLT edge switches do not detect a failure and continue to communicate as usual. Normally, more than one link in the IST is available since the interswitch trunk is itself a distributed MLT. IST traffic resumes over the remaining links in the IST.

Finally, if all IST links are lost between an aggregation switch pair, the aggregation switches cannot communicate with each other. Both switches assume that the other switch failed. Generally, a complete IST link failure causes no ill effects in a network if all SMLT edge switches are dual-homed to the SMLT aggregation switches. However, traffic that comes from single attached switches or devices no longer predictably reaches the destination. IP forwarding can cease because both switches try to become the VRRP Master. Because the wiring closets switches do not know about the interswitch trunk failure, the network provides intermittent connectivity for devices that attach to only one aggregation switch. Data forwarding, while functional, is not optimal because the aggregation switches can not learn all MAC addresses, and the aggregation switches can flood traffic that does not normally flood.

SMLT and IEEE 802.3ad interaction

Virtual Services Platform 9000 fully supports IEEE 802.3ad LACP on MLT and distributed MLT links. On a pair of SMLT switches:

- MLT peer and SMLT client devices can be network switches, a server, or a workstation that supports link bundling through IEEE 802.3ad.
- Multilink SMLT solutions support dual-homed connectivity for more than 350 attached devices, which allow dual-homed server farm solutions.

Only dual-homed devices benefit from LACP and SMLT interactivity.

SMLT and IEEE link aggregation supports all known SMLT scenarios where an IEEE 802.3ad SMLT pair can connect to SMLT clients or where two IEEE 802.3ad SMLT pairs can connect to each other in a square or full-mesh topology.

Known SMLT and LACP failure scenarios include

- wrong ports connected
- LACP is disabled on the SMLT edge switch

  SMLT aggregation switches detect that aggregation is disabled on the SMLT client, thus no automatic link aggregation establishes until the configuration is resolved.
- Single CPU failure

  In this case, LACP on other switches detects the remote failure, and all links that connect to the failed system are removed from the link aggregation group. This process allows failure recovery to a different network path.
- LACP and VLACP cannot run on the same interfaces simultaneously.

SMLT and LACP System ID

The LACP SMLT System ID used by SMLT core aggregation switches is configurable. Configure the LACP SMLT system ID to be the base MAC address of one of the aggregate switches and include the SMLT-ID. Ensure that the same System ID is configured on both of the SMLT core aggregation switches.

The LACP System ID is the base MAC address of the switch, which is carried in Link Aggregation Control Protocol Data Units (LACPDU). When two links interconnect two switches running LACP, each switch is aware both links connect to the same remote device because the LACPDUs originate from the same System ID. If the links are enabled for aggregation using the same key, LACP can dynamically aggregate them into a LAG (MLT).

When SMLT is used between the two switches, they act as one logical device. Both aggregation switches must use the same LACP System ID over the SMLT links. This ensures the edge switch sees one logical LACP peer, and can aggregate uplinks towards the SMLT aggregation switches. This process automatically occurs over the IST connection, where the base MAC address of one of the SMLT aggregation switches is chosen and used by both SMLT aggregation switches.

If the switch that owns that Base MAC address reboots, the IST is no longer operational and the other switch reverts to using its own Base MAC address as the LACP System ID. This action causes all edge switches that run LACP to think their links are connected to a different switch. The edge switches stop forwarding traffic on their remaining uplinks until the aggregation can reform. Aggregation reformation can take several seconds. When the rebooted switch comes back online, the same actions occur and disrupt traffic twice. The solution to this situation is to statically configure the same SMLT System ID MAC address on both aggregation switches.

For more information about how to configure the LACP SMLT system ID, see *Avaya Virtual Services Platform 9000 Configuration — Link Aggregation, MLT, and SMLT*, NN46250-503.

SMLT scalability

Virtual Services Platform 9000 does not have VLAN limitations for SMLT-enabled VLANs. All VLANs (minus 1 for the IST VLAN) can be used for SMLT. Additionally, all 512 MLT groups (minus 1 for the IST) can be used for SMLT. Up to 128,000 MAC addresses are supported for SMLT.

For more information about VLAN scalability, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500.

SMLT topologies

SMLT supports three topologies with switch clustering; the SMLT triangle, the SMLT square, and the SMLT full mesh.

A triangle design is an SMLT configuration where edge switches or SMLT clients connect to two aggregation switches. The aggregation switches connect with an interswitch trunk that carries all the SMLTs configured on the switches. Each switch pair can have up to 31 SMLT edge switch connections.

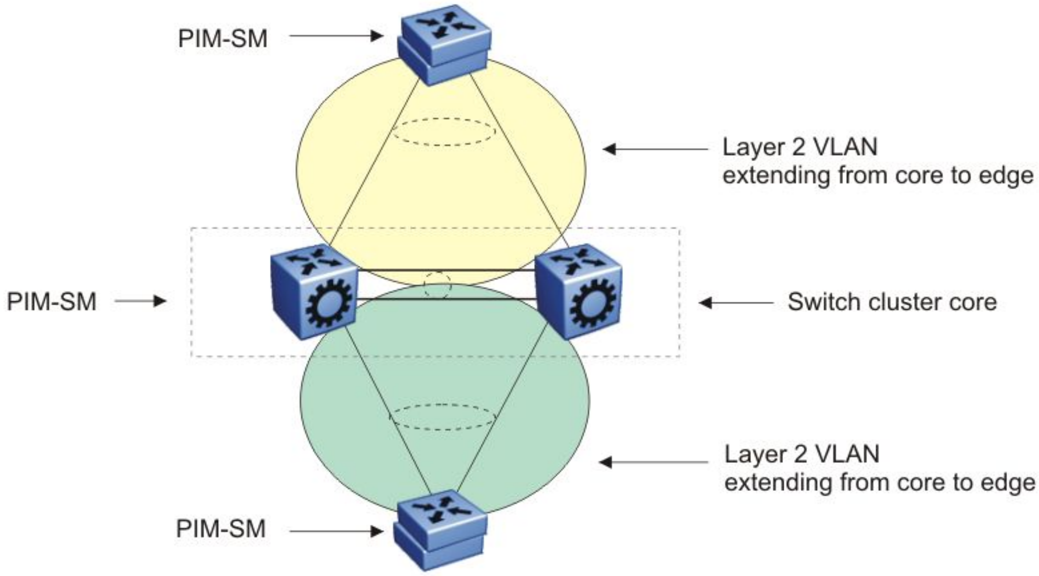

Figure 20: SMLT triangle configuration

In a square design, a pair of aggregation switches (a switch cluster) connect to another pair of aggregation switches (another switch cluster). You can scale a square topology by adding pairs of switch clusters. The following figure shows a square configuration.

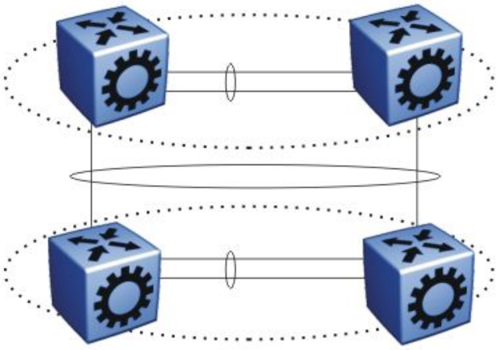

Figure 21: SMLT square configuration

The full mesh expands on the square topology by adding additional connections between the pairs. This ensures each switch has at least one connection to every other switch in the square. You can scale a full-mesh topology with additional pairs of switch clusters. Configure an SMLT full-mesh configuration as shown in the following figure. The vertical and diagonal links emanating from a switch are part of an MLT.

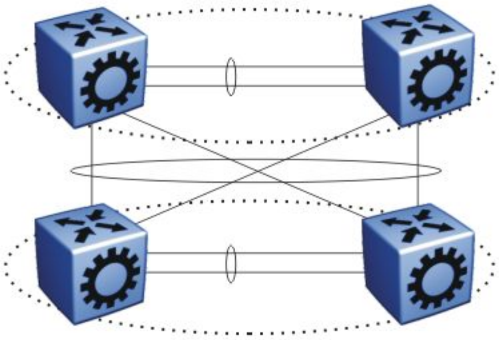

Figure 22: SMLT full-mesh configuration

Virtual Services Platform 9000 supports up to 512 MLT groups of 16 ports. Configure SMLT groups as shown in Figure 23: SMLT scaling on page 63 within the network core. Configure both sides of the links for SMLT. No state information passes across the MLT link; both ends believe that the other is a single switch. The result is that no loop is introduced into the network. One of the core switches or the connecting links between them can fail, but the network recovers rapidly.

Scale SMLT groups to achieve hierarchical network designs by connecting SMLT groups together. This allows redundant loop-free Layer 2 domains that fully use all network links without using an additional protocol. The following figure shows the connection of multiple SMLT switch clusters to create a larger, scaled network.

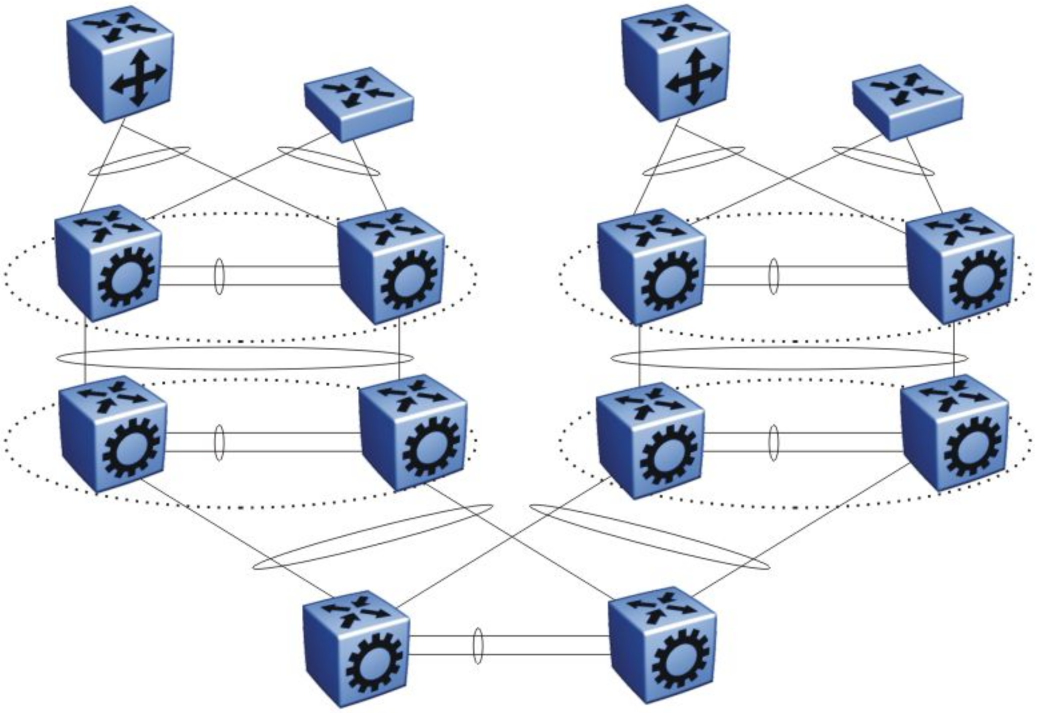

Figure 23: SMLT scaling

For more information about the SMLT triangle, square, and full-mesh designs, see *Avaya Virtual Services Platform 9000 Configuration — Link Aggregation, MLT, and SMLT*, NN46250-503 and *The Large Campus Technical Solution Guide*, NN48500-575.

SMLT full-mesh recommendations with OSPF

Place the MLT ports that form the square leg of the mesh, not the cross connections, on lower numbered slots and ports in a full-mesh SMLT configuration that runs OSPF. This is typically an RSMLT configuration. CP-generated traffic is always transmitted on lower numbered MLT ports when active. This configuration keeps some OSPF adjacencies up if the IST on one cluster fails. In the absence of such a configuration an operational switch in this scenario can lose complete OSPF adjacency to both switches in the other cluster and become isolated.

Chapter 10: Layer 3 switch clustering and RSMLT

This section describes designs for achieving network redundancy. Network redundancy minimizes failure and ensures a faulty switch does not interrupt service.

Routed SMLT

Core network convergence time usually depends on the length of time a routing protocol requires to successfully converge. This convergence time can cause network interruptions that range from seconds to minutes depending on the specific routing protocol. Routed Split Multilink Trunking (RSMLT) allows rapid failover for core topologies by providing an active-active router concept to core SMLT networks. Virtual Services Platform 9000 supports RSMLT on SMLT triangles, squares, and SMLT full-mesh topologies that have routing enabled on the core VLANs. RSMLT provides redundancy as well. If a core router fails, RSMLT provides packet forwarding. This eliminates dropped packets during convergence.

Virtual Services Platform 9000 can use one of the following routing protocols to provide convergence:

- IP or IPv6 unicast static routes
- Routing Information Protocol version 1 (RIPv1) or version 2 (RIPv2) (IPv4)
- Open Shortest Path First (OSPF) and OSPFv3
- Border Gateway Protocol (BGP) (IPv4) and BGP+

SMLT and RSMLT operation

Figure 24: SMLT and RSMLT in Layer 2 and 3 environments on page 66 shows a typical redundant network with user aggregation, core, and server access layers. To minimize the creation of many IP subnets, one VLAN (VLAN 1, IP subnet A) spans all wiring closets. SMLT provides loop prevention and enables all links to forward to VLAN 1, IP Subnet A. RSMLT runs on the core.

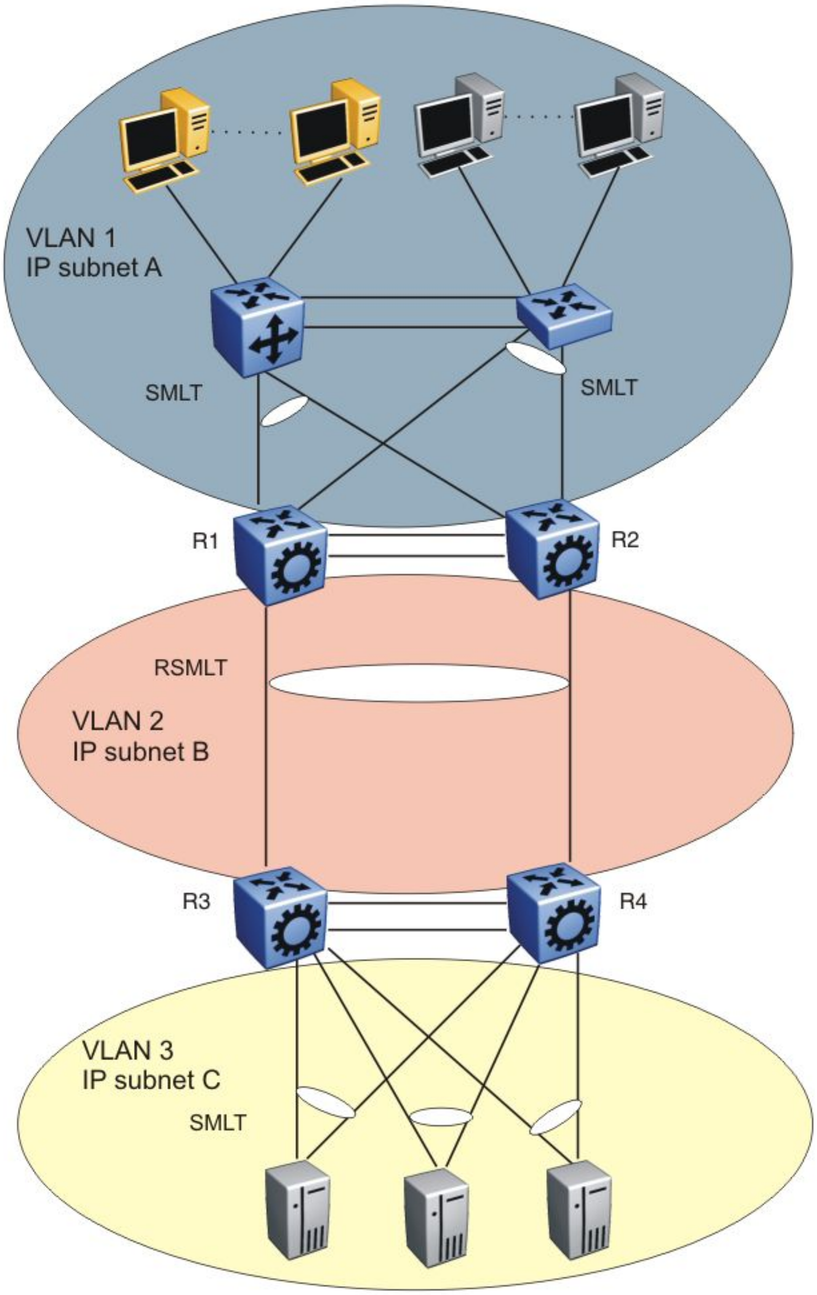

Figure 24: SMLT and RSMLT in Layer 2 and 3 environments

The aggregation layer switches are routing-enabled and provide active-active default gateway functions through RSMLT. Routers R1 and R2 forward traffic for IP subnet A. RSMLT provides both router and link failover. If the SMLT link between R2 and R4 breaks, the traffic fails over to R1.

For IP subnet A, Virtual Router Redundancy Protocol (VRRP) Backup-Master can provide the same functions as RSMLT, as long as an additional router is not connected to IP subnet A. In large scale environments, for example, more than 64 VRRP instances, Avaya recommends that you use RSMLT with RSMLT edge instead of VRRP. For more information, see RSMLT redundant network with bridged and routed VLANs in the core on page 186.

RSMLT provides superior router redundancy in core networks (for example, IP subnet B) in which OSPF is used. Routers R1 and R2 provide router backup for each other—not only for the edge IP subnet A but also for the core IP subnet B. Similarly, routers R3 and R4 provide router redundancy for IP subnet C and also for core IP subnet B.

RSMLT router failure and recovery

This section describes the failure and recovery of router R1 in Figure 24: SMLT and RSMLT in Layer 2 and 3 environments on page 66.

R3 and R4 both use R1 as their next-hop to reach IP subnet A. Even though R4 sends packets to R2, these packets are routed directly to subnet A at R2. R3 sends packets towards R1; these packets are also sent directly to subnet A. After R1 fails, with the help of SMLT, all packets are directed to R2. R2 provides routing for R2 and R1.

After OSPF converges, R3 and R4 change their next-hop to R2 to reach IP subnet A. You can configure the hold-up timer (that is, the amount of time R2 routes for R1 in the event of failure) to a time period greater than the routing protocol convergence or to indefinite (that is, the pair always routes for each other). Avaya recommends that you configure the hold up and hold down timer to 1.5 times the convergence time of the network.

In an application where you use RSMLT at the edge instead of VRRP, Avaya recommends that you configure the hold-up timer value to indefinite.

After R1 restarts after a failure, it first becomes active as a VLAN bridge. Using the bridge forwarding table, packets destined to R1 are switched to R2 for as long as the hold-down timer value. These packets are routed at R2 for R1. Like VRRP, to converge routing tables, the hold-down timer value needs to be greater than the one required by the routing protocol.

After the hold-down time expires and the routing tables have converged, R1 starts routing packets for itself and also for R2. Therefore, it does not matter which one of the two routers is used as the next-hop from R3 and R4 to reach IP subnet A.

If you configure single-homed IP subnets on R1 or R2, Avaya recommends that you add another routed VLAN to the interswitch trunks (IST). As a traversal VLAN or subnet, this additional routed VLAN needs lower routing protocol metrics to avoid unnecessary Internet Control Message Protocol (ICMP) redirect generation messages. This recommendation also applies to VRRP implementations.

RSMLT guidelines

Use the following guidelines when you create RSMLT configurations:

- RSMLT is based on SMLT so all SMLT configuration rules apply. Enable RSMLT on the SMLT aggregation switches on an individual VLAN basis. The VLAN must be a member of SMLT links and the IST trunk.
- The VLAN must be routable (IP address configured). On all four routers in a square or full-mesh topology, configure an Interior Routing Protocol, such as OSPF, although the protocol is independent from RSMLT.

- Routing protocols and static routes can be used with RSMLT.
- RSMLT pair switches provide backup for each other. As long as one of the two routers in an IST pair is active, traffic forwarding is available for both next-hops.

For design examples using RSMLT, see the following sections and RSMLT redundant network with bridged and routed VLANs in the core on page 186.

RSMLT timer tuning

RSMLT enables a participating peer switch to act as a router for its peer by MAC address. This doubles router capacity and enables fast failover in the event of a peer switch failure. RSMLT provides hold-up and hold-down timer parameters to aid these functions.

The hold-up timer defines the length of time the RSMLT-peer switch routes for its peer after a peer switch failure. Configure the hold-up timer to at least 1.5 times greater than the routing protocol convergence time.

The RSMLT hold-down timer defines the length of time that the recovering switch remains in a non-Layer 3 forwarding mode for the MAC address of its peer. Configure the hold-down timer to at least 1.5 times greater than the routing protocol convergence time. The configuration of the hold-down timer gives RIP, OSPF or BGP time to build up the routing table before Layer 3 forwarding for the peer router MAC address begins again.

Important:

When using a Layer 3 SMLT client switch without a routing protocol, configure two static routes to point to both RSMLT switches or configure one static route. Configure the RSMLT hold-up timer to 9999 (infinity). Also configure the RSMLT hold-up timer to 9999 (infinity) for RSMLT Edge (Layer 2 RSMLT).

IPv6 differences

The following list identifies ways in which the IPv6 implementation of RSMLT differs from the IPv4 implementation of RSMLT.

- After the switch begins to forward traffic on behalf of the peer, duplicate address detection (DAD) is not executed for the IPv6 address of the peer. The implementation assumes that the peer IPv6 address is already known to be unique.
- An RSMLT switch installs a neighbor entry for the peer IPv6 address immediately after the peer disappearance is detected, possibly while a route for the peer still exists. This action can result in packets destined to the peer IPv6 address being delivered to the CP for a short period of time. This limitation is caused by the fact that route reference count is not maintained for IPv6 neighbor entries.
- You can not configure an IST with IPv6 peer address
- In a dual-stack VLAN, adding or deleting IPv4 or IPv6 does not affect the RSMLT functionality of one another. If you add IPv4 or IPv6 to an existing IPv6 or IPv4 RSMLT VLAN, the RSMLT state for the protocol you add second will be the same as the previous RSMLT state.

Example: RSMLT redundant network with bridged and routed edge VLANs

Many Enterprise networks require the support of VLANs that span multiple wiring closets. VLANs are often local to wiring closets and routed towards the core. The following figure shows

VLAN-10, which has all IP Deskphones as members and resides everywhere, while at the same time VLANs 20 and 30 are user VLANs that are routed through VLAN-40.

A combination of SMLT and RSMLT provide sub-second failover for all VLANs bridged or routed. VLAN-40 is RSMLT enabled that provides for the required redundancy. You can use unicast routing protocols—such as RIP, OSPF, or BGP—and routing convergence times do not impact the network convergence time provided by RSMLT.

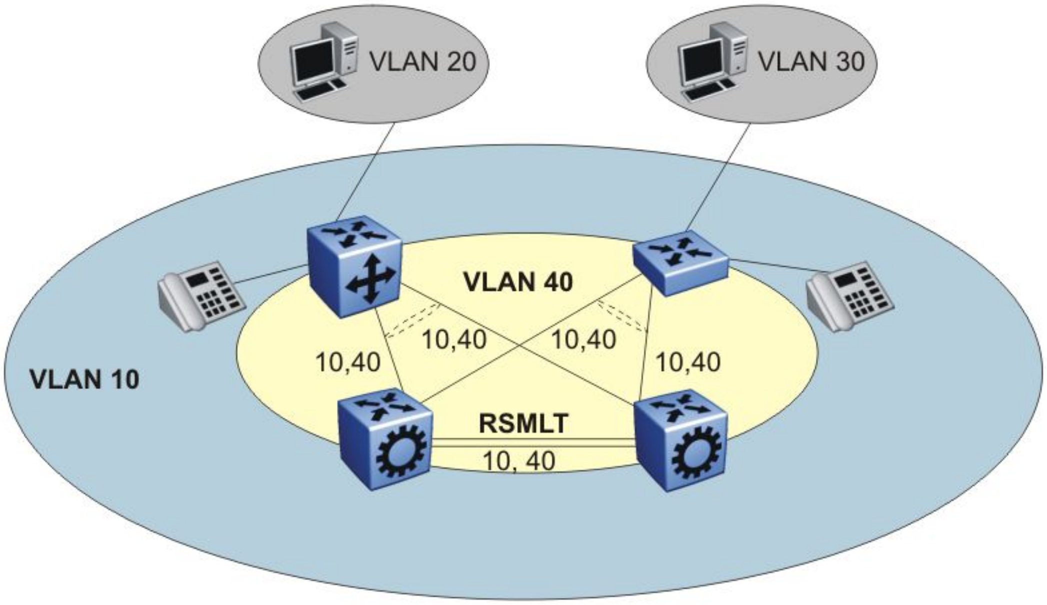

Figure 25: VLAN with all IP Deskphones as members

Example: RSMLT network with static routes at the access layer

Use default routes that point towards the RSMLT IP interfaces of the aggregation layer to achieve a robust redundant edge design, as shown in the following figure.

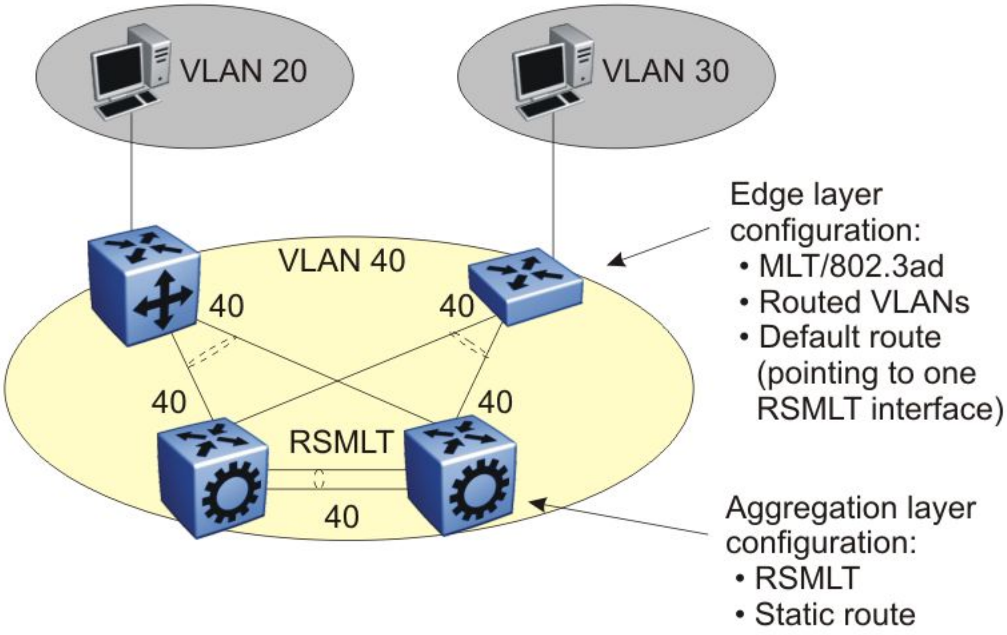
VLAN 20
VLAN 30
VLAN 40
40
40
40
40
RSMLT
40
Edge layer configuration:
• MLT/802.3ad
• Routed VLANs
• Default route (pointing to one RSMLT interface)
Aggregation layer configuration:
• RSMLT
• Static route

Figure 26: VLAN edge configuration

Example: RSMLT IPv6 network topology

The following figure shows a sample IPv6 RSMLT topology. The figure shows a typical redundant network example with user aggregation, core, and server access layers. To minimize the creation of many IPv6 prefixes, one VLAN (VLAN 1, IP prefix A) spans all wiring closets. RSMLT provides the loop-free topology. The aggregation layer switches are configured with routing enabled and provide active-active default gateway functionality through RSMLT.

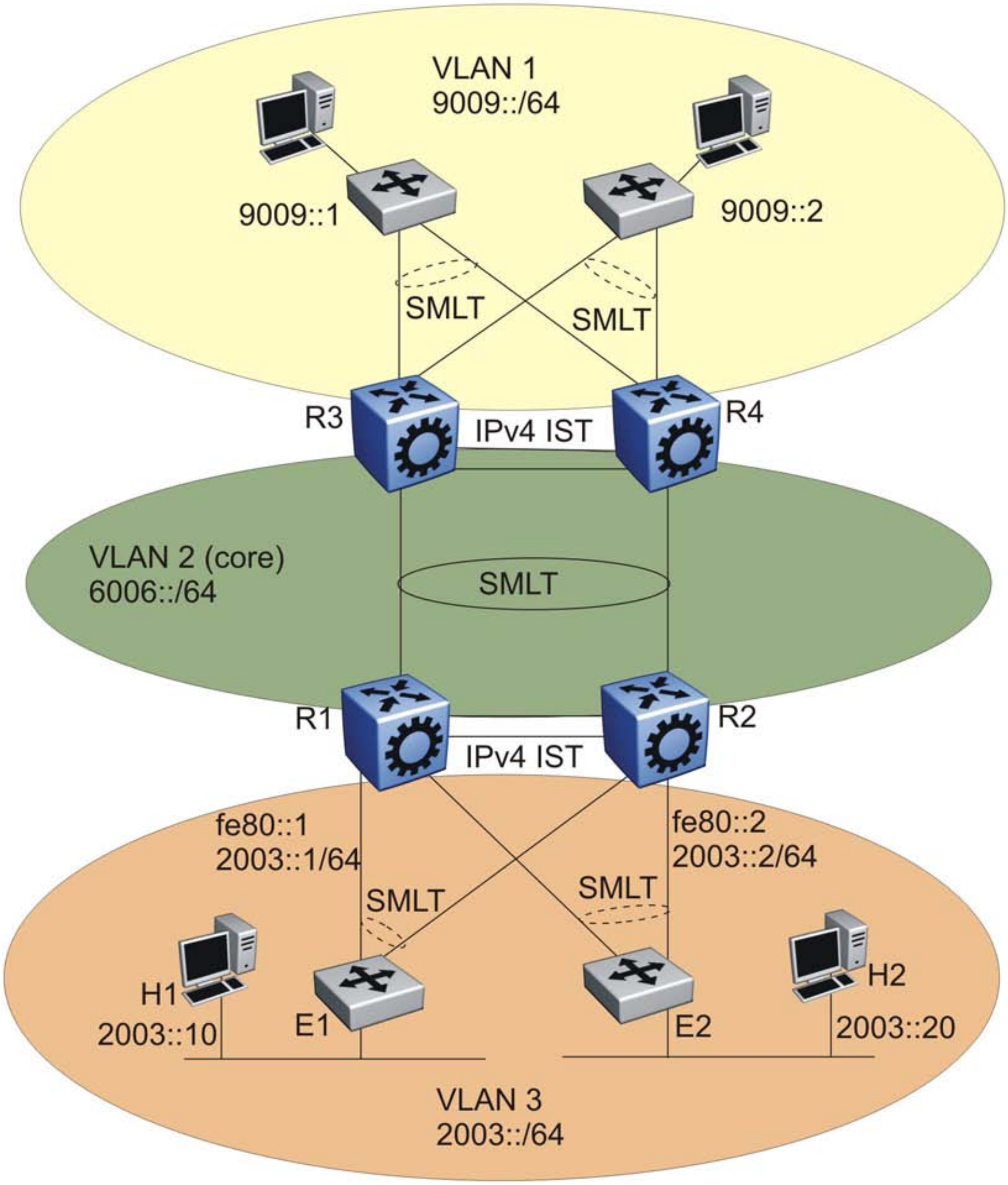

Figure 27: IPv6 RSMLT topology

In VLAN 3 of the preceding figure, routers R1 and R2 provide RSMLT-enabled IPv6 service to hosts H1 and H2. Router R1 is the default IPv6 router for H1 and R2 is the default router for H2. R1 uses the following configuration:

- link-local address of fe80::1
- global unicast address 2003::1
- routing prefix of 2003::/64

As a shorthand, the last two items in the preceding list are referred to as 2003::1/64.

R2 uses the following configuration:

- link-local address of fe80::1
- global unicast address and routing prefix if routing prefix of 2003::2/64.

Host H1 sends IPv6 traffic destined to VLAN 1 to the MAC address for R1. H2 sends traffic to the MAC address for R2. After an IPv6 packet destined to the MAC address of R1 is received at R2 on its SMLT links (the expected MLT behavior), R2 performs IPv6 forwarding on the packet and does not bridge it over the IST. The same behavior occurs on R1.

At startup, R1 and R2 use the IST link to exchange full configuration information that includes the MAC address for the IPv6 interfaces that reside on SMLT VLAN 3.

After R2 detects that the RSMLT in R1 transitions to the down state (for example, if R1 itself is down, the SMLT links are down, or the IST link is down) R2 takes over IPv6 termination and IPv6 neighbor discovery on behalf of the IPv6 SMLT interface on R1. The following list shows the order of action in this situation:

1. After R2 detects the event, it transmits an unsolicited IPv6 neighbor advertisement for each IPv6 address configured on the SMLT link of R1 using the R1 MAC address (fe80::1 and 2003::1 in this example).
2. R2 transmits an unsolicited router advertisement for each of the R1 routing prefixes (unless the prefixes are configured as not advertised).
3. R2 responds to neighbor solicitations and, if the configuration allows, router advertisements on behalf of R1.
4. R2 terminates IPv6 traffic (such as ping requests) destined to the R1 SMLT IPv6 addresses.

After R1 RSMLT transitions back into the up state and the hold-down timer expires, R1 resumes IPv6 forwarding and R2 ceases to terminate IPv6 traffic on behalf of R1.

IPv6 provides a rich set of configuration options to advertise IPv6 routing prefixes (equivalent to IPv4 subnets) and to configure hosts on a link. You can configure a prefix to be or not be advertised, and to carry various flags or lifetime values. These parameters affect how hosts can automatically configure their IPv6 addresses and select their default routers. Most relevant from the RSMLT perspective is that an RSMLT node fully impersonates the peer IPv6 configuration and behavior on the SMLT link. The preceding network example illustrates one of the many possible deployment scenarios for IPv6 routers and hosts on a VLAN.

RSMLT provides both router failover and link failover. For example, if the SMLT link between R2 and R4 is broken, the traffic fails over to R1 as well.

Switch clustering topologies and interoperability with other products

The switch clustering, unicast routing, and multicast routing configurations vary with switch type when using Ethernet Routing Switch products with Avaya Virtual Services Platform 9000.

Use the supported topologies and features when you perform inter-product switch clustering. For more information see *Switch Clustering (SMLT/SLT/RSMLT/MSMLT) Supported Topologies and Interoperability with ERS 8800 / 5500 / 8300 / 1600*, NN48500-555. For specific design and configuration parameters see *The Large Campus Technical Solution Guide*, NN48500-575 and *Switch Clustering using Split-Multilink Trunking (SMLT) Technical Configuration Guide*, NN48500-518.

Chapter 11: Layer 3 switch clustering and multicast SMLT

Switch clustering is the logical aggregation of two nodes to form one logical entity known as the switch cluster. The two peer nodes in a switch cluster connect using an interswitch trunk (IST). The IST exchanges forwarding and routing information between the two peer nodes in the cluster. This section provides guidelines for switch clusters that use multicast and Split Multilink Trunking (SMLT).

General guidelines

The following list identifies general guidelines to follow if you use multicast and switch clustering:

- Enable Protocol Independent Multicast - Sparse Mode (PIM-SM) on the IST VLAN for fast recovery of multicast. A unicast routing protocol is not required.
- Enable Internet Group Management Protocol (IGMP) snooping and proxy on the edge switches.

The following figure shows multicast behavior in an SMLT environment. The configuration in the following figure provides fast failover if the switch or rendezvous point (RP) fails.

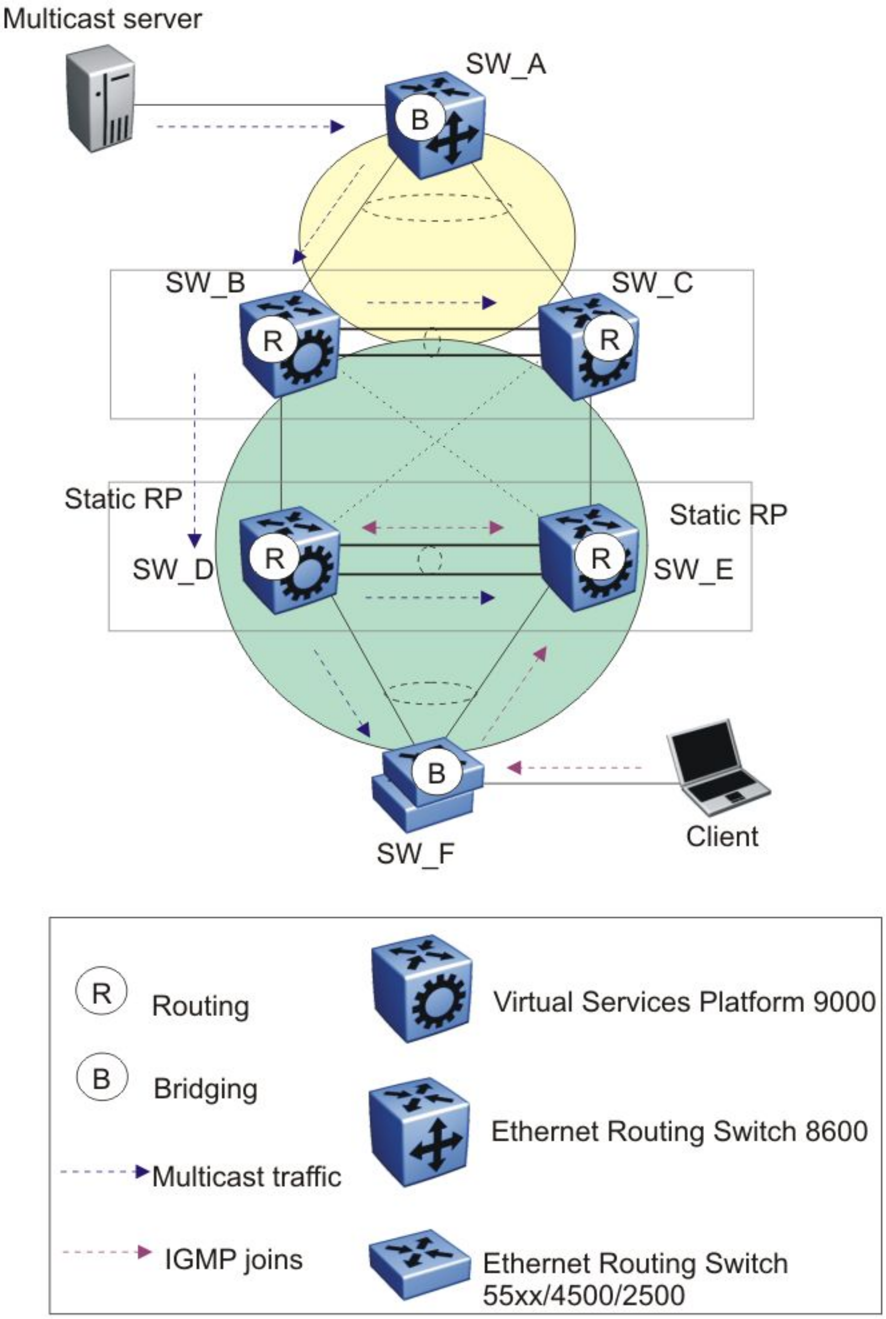

Figure 28: Multicast behavior in SMLT environment

In Figure 28: Multicast behavior in SMLT environment on page 76, the following actions occur:

1. The multicast server sends multicast data towards the source designated router (DR).
2. The source DR sends register messages with encapsulated multicast data towards the RP.

3. After the client sends IGMP membership reports towards the multicast router, the router creates a (*,G) entry.
4. The RP sends join messages towards the source DR on the reverse path.
5. After the source DR receives the join messages, it sends native multicast traffic.
6. After SW_B or SW_D receives multicast traffic from upstream, it forwards the traffic on the IST as well as on the SMLT link. Other aggregation switches drop multicast traffic received over the IST at egress. This action provides fast failover for multicast traffic. Both SW_D and SW_E (Aggregation switches) have similar (S,G) records.
7. In case of SW_D or RP failure, SW_B changes only the next-hop interface towards SW_E. Because the circuitless IP (CLIP) RP address is the same, SW_B does not flush (S,G) entries and achieves fast failover.

Multicast triangle topology

A triangle design is an SMLT configuration that connects edge switches or SMLT clients to two aggregation switches. Connect the aggregation switches together with an IST that carries all the SMLT trunks configured on the switches.

Virtual Services Platform 9000 supports the following triangle configurations:

- a configuration with Layer 3 PIM-SM routing on both the edge and aggregation switches
- a configuration with Layer 2 snooping on the client switches and Layer 3 routing with PIM-SM on the aggregation switches

To avoid using an external query device to provide correct handling and routing of multicast traffic to the rest of the network, use the triangle design with IGMP Snoop at the client switches. Use multicast routing at the aggregation switches as shown in the following figure.

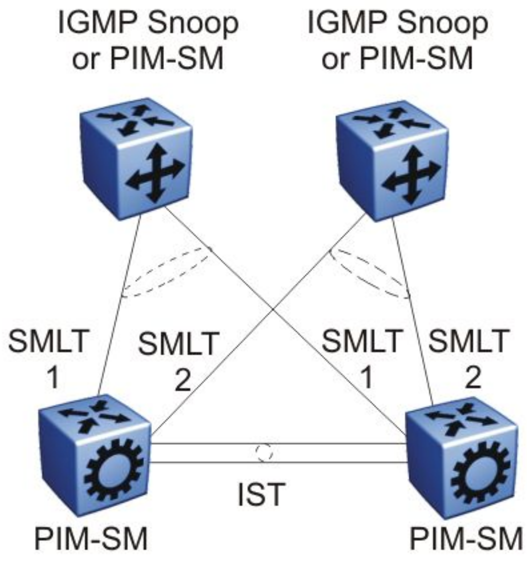

Figure 29: Multicast routing using PIM-SM

Client switches run IGMP Snoop or PIM-SM, and the aggregation switches run PIM-SM. This design is simple and, for the rest of the network, PIM-SM performs IP multicast routing. The aggregation switches are the query devices for IGMP, so an external query device is not required to activate IGMP membership. These switches also act as redundant switches for IP multicast.

Multicast data flows through the IST link when receivers are learned on the client switch and senders are located on the aggregation switches, or when sourced data comes through the aggregation switches. This data is destined for potential receivers attached to the other side of the IST. The data does not reach the client switches through the two aggregation switches because only the originating switch forwards the data to the client switch receivers.

Note:

Always place multicast receivers and senders on the core switches on VLANs different from those that span the IST.

The following figure shows a switch clustering configuration with a single switch cluster core and dual-connected edge devices. This topology represents different VLANs spanning from each edge device and those VLANs routed at the switch cluster core. You can configure multiple VLANs on the edge devices, 802.1Q tagged to the switch cluster core.

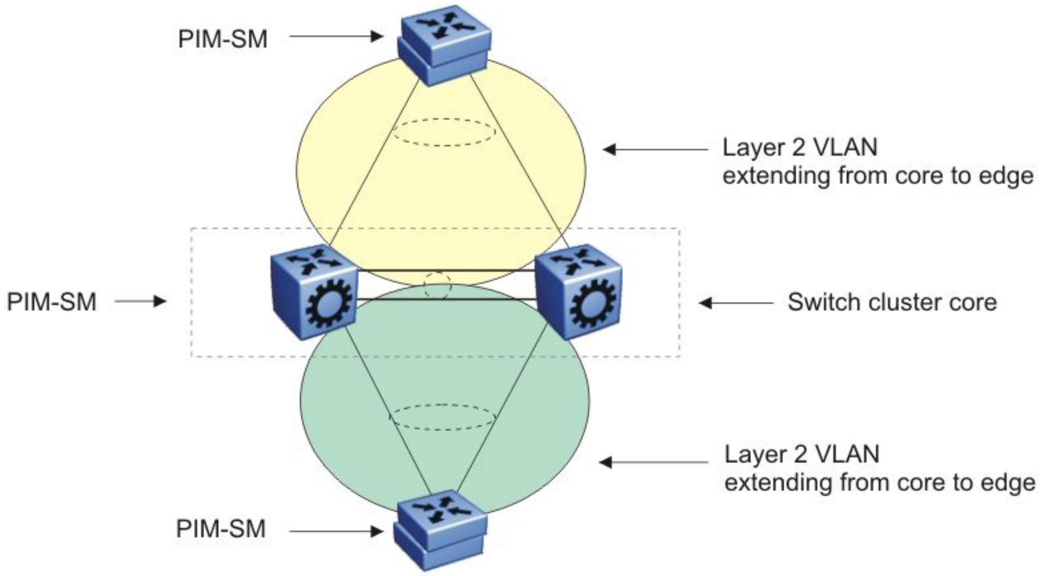

Figure 30: Multicast SMLT triangle

Use an edge device that supports a form of link aggregation. Disable spanning tree on the link aggregation group on the edge devices. Enable either Virtual Router Redundancy Protocol (VRRP) BackupMaster or Routed SMLT (RSMLT) Layer 2 Edge on the switch cluster core.

Square and full-mesh topology multicast guidelines

A square design connects a pair of aggregation switches to another pair of aggregation switches. A square design becomes a full-mesh design if the aggregation switches are connected in a full-mesh. Virtual Services Platform 9000 supports Layer 3 IP multicast (PIM-SM only) over a full-mesh SMLT or RSMLT configuration.

In a square design, configure all switches with PIM-SM. Place the bootstrap router (BSR) and RP in one of the four core switches; Avaya recommends that you place the RP closest to the source. If using PIM-SM over a square or full-mesh configuration, enable the **`multicast smlt-square`** flag.

The following three figures show switch clustering configurations with two-switch cluster cores and dual-connected edge devices.

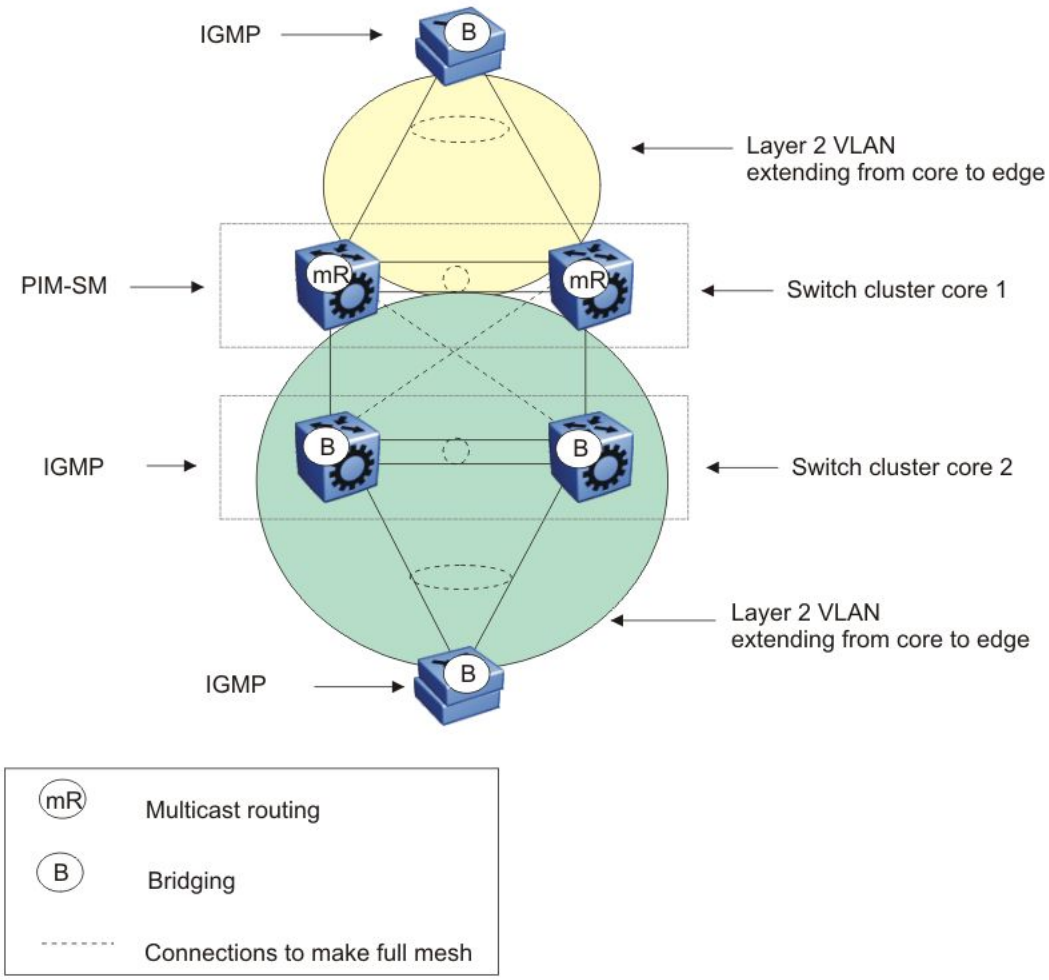

Figure 31: Multicast SMLT square 1

In the preceding figure, only one of the switch cluster cores performs Layer 3 multicast routing while the other is strictly Layer 2. Configure multiple VLANs on the edge devices, 802.1Q tagged to the switch cluster cores.

Use an edge device that supports a form of link aggregation. Disable spanning tree on the link aggregation group on the edge devices. Enable either the VRRP BackupMaster or RSMLT Layer 2 Edge on the switch cluster core.

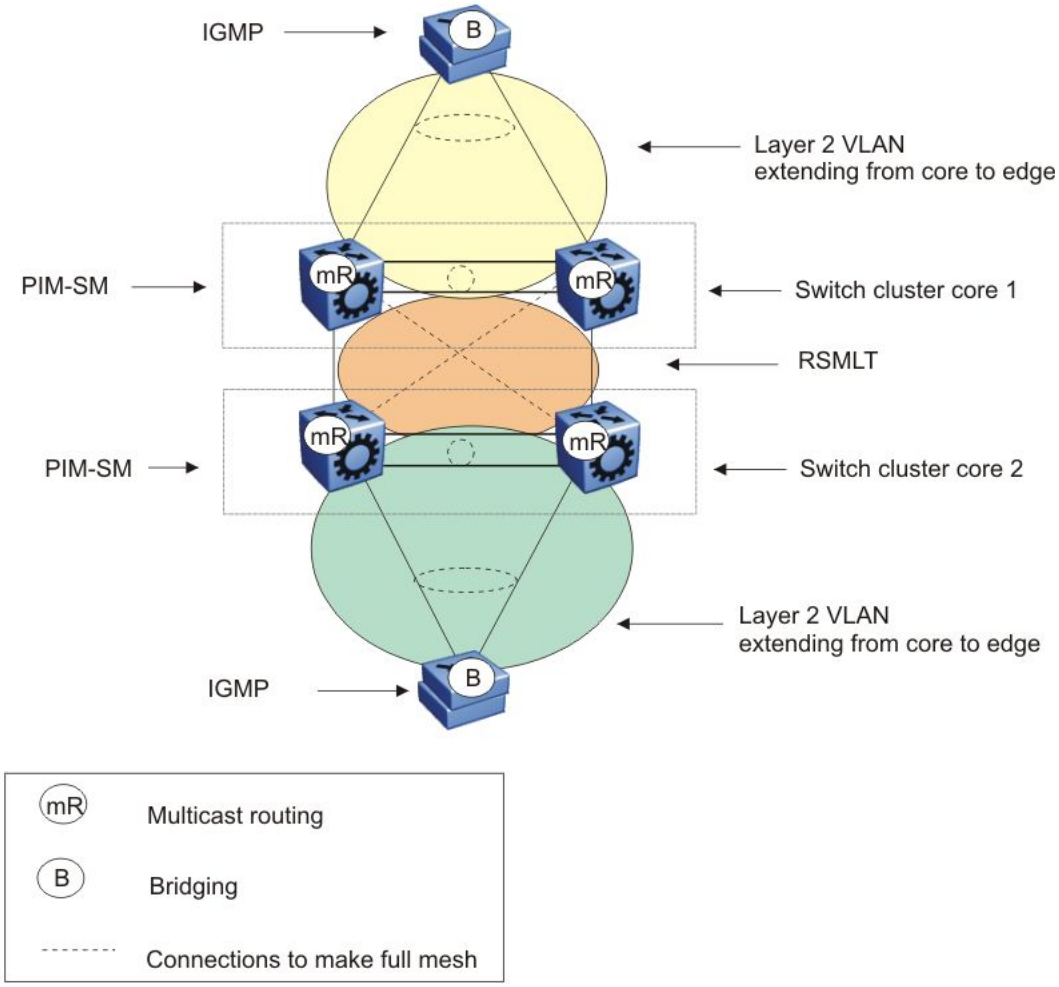

Figure 32: Multicast SMLT square 2

In the preceding figure, both of the switch cluster cores performs Layer 3 multicast routing, while the edge devices are Layer 2 IGMP.

Use an edge device that supports a form of link aggregation. Disable spanning tree on the link aggregation group on the edge devices. Enable either the VRRP BackupMaster or RSMLT Layer 2 Edge on the switch cluster cores. Do not enable VRRP on the RSMLT VLAN between switch cluster cores.

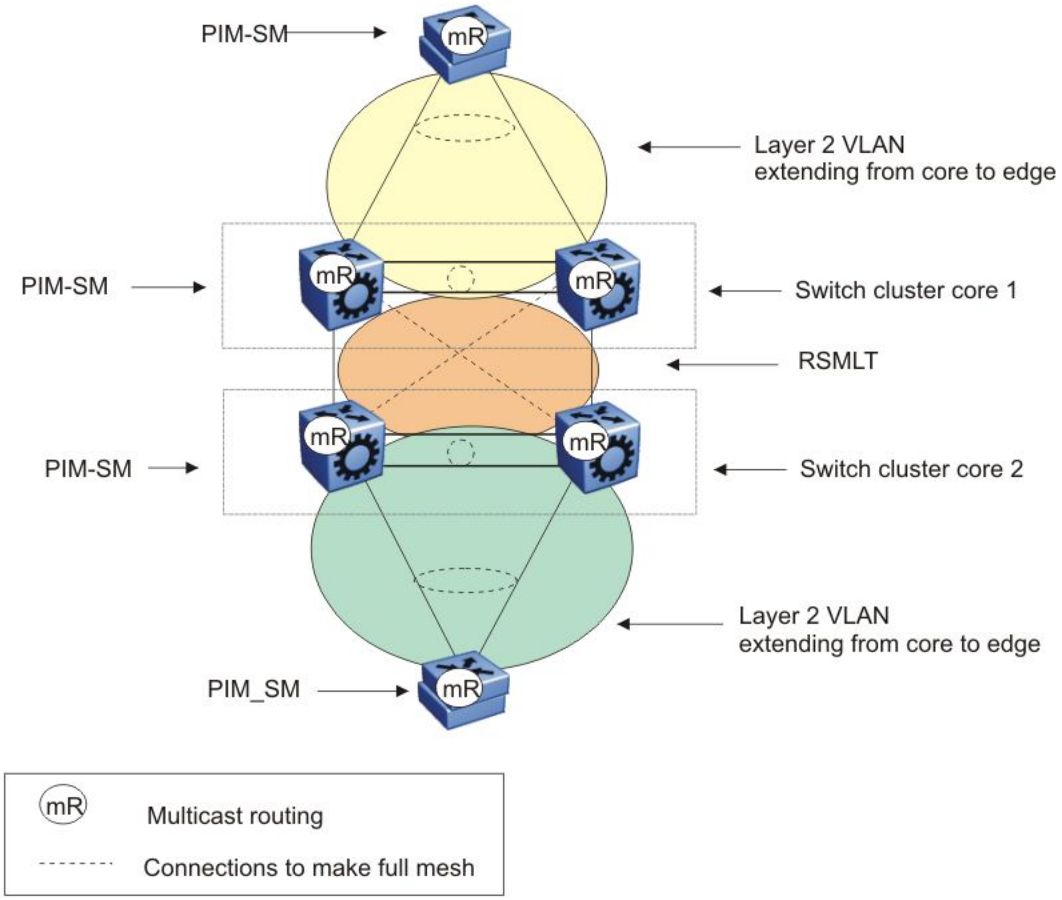

Figure 33: Multicast SMLT square 3

In the preceding figure, both of the switch cluster cores and the edge devices perform Layer 3 multicast routing.

Use an edge device that supports a form of link aggregation. Disable spanning tree on the link aggregation group on the edge devices. Enable either the VRRP BackupMaster or RSMLT Layer 2 Edge on the switch cluster cores. Do not enable VRRP on the RSMLT VLAN between switch cluster cores.

SMLT and multicast traffic issues

If PIM-SM or other multicast protocols are used in an SMLT environment, enable the protocol on the IST. Routing protocols in general are not run over an IST but multicast routing protocols are an exception. When using PIM-SM and a unicast routing protocol, ensure the unicast route to the BSR and RP has PIM-SM active and enabled. If multiple OSPF paths exist and PIM-SM is not active on each pair, the BSR is learned on a path that does not have PIM-SM active. The following figure demonstrates this issue.SMLTand multicast traffic duplication

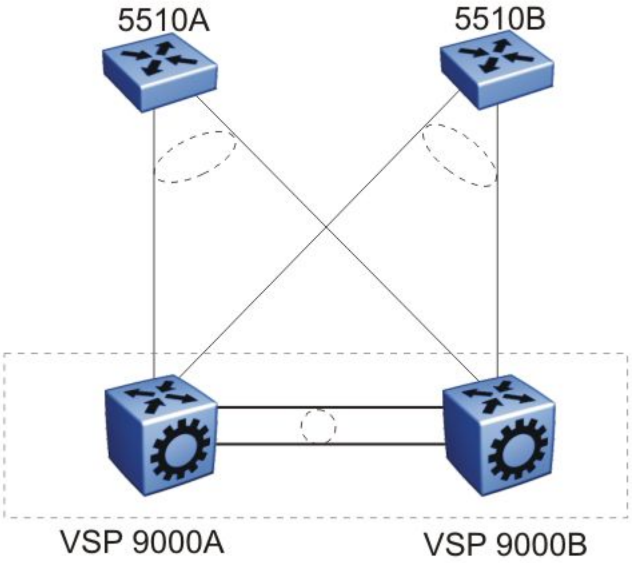

Figure 34: Unicast route example

The network configuration in the preceding figure is as follows:

- 5510A is on VLAN 101.
- 5510B is on VLAN 102.
- VSP 9000B is the BSR.
- VSP 9000A and VSP 9000B have OSPF enabled.
- PIM is enabled and active on VLAN 101.
- PIM is either disabled or passive on VLAN 102.

In this example, the unicast route table on VSP 9000A learns the BSR on VSP 9000B through VLAN 102 using OSPF. The BSR is either not learned or does not provide the RP to VSP 9000A.

Dropped multicast traffic during startup

The egress-access SMLT port makes the drop decisions for unwanted IP multicast traffic from core SMLT ports and IST ports. All ingress IP multicast traffic is replicated to all access SMLT and IST ports. If you configure a high number of SMLT and IST ports, or if you configure SMLT and IST ports on the same lane, and before unicast traffic is learned, the unicast traffic is flooded to all port members of the VLAN. In this situation, the switch fabric bandwidth can become overloaded and drop multicast traffic. After the unicast traffic is learned, multicast traffic returns to normal. To avoid dropped multicast traffic during startup, reduce the number of SMLT and IST ports or configure them on a different lane to reduce the amount of multicast traffic flooding to all lanes.

Chapter 12: Layer 2 loop prevention

To use bandwidth and network resources efficiently, prevent Layer 2 data loops. Use the information in this section to help you use loop prevention mechanisms.

Loop prevention and detection

Split MultiLink Trunking (SMLT) based network designs form physical loops for redundancy that logically do not function as a loop. Under certain adverse conditions, loops can form, for example, if you have incorrect configurations or cabling.

The two solutions to detect loops are Loop Detect and Simple Loop Prevention Protocol (SLPP). Loop Detect and SLPP detect a loop and automatically stop the loop. Both solutions determine on which port the loop is occurring and shut down that port.

Avaya recommends the following loop prevention and recovery features in order of preference:

- SLPP
- Loop Detect with ARP-Detect activated

For information about how to configure CP-Limit, see *Avaya Virtual Services Platform 9000 Administration*, NN46250-600. For more information about SLPP and loop detection, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500.

SLPP

Avaya recommends that you use SLPP to protect the network against Layer 2 loops. If you configure and enable SLPP, the switch sends a test packet to the VLAN. A loop is detected if the switch or a peer aggregation switch on the same VLAN receives the original packet. If the switch detects a loop, the switch disables the port. To enable the port requires manual intervention. As an alternative, you can use port auto-enable to re-enable the port after a predefined interval.

Use SLPP to prevent loops in an SMLT network.

Loops can be introduced into the network in many ways. One way is through the loss of a multilink trunk configuration caused by user error or malfunction. This scenario does not introduce a broadcast storm, but because all MAC addresses are learned through the looping ports, Layer 2 MAC learning is significantly impacted. Spanning tree protocols cannot always detect such a configuration issue, whereas SLPP reacts and disables the malfunctioning links, which minimizes the impact on the network.

SLPP and SMLT examples

The following configurations show how to configure SLPP so that the protocol detects VLAN-based network loops for untagged and tagged IEEE 802.1Q VLAN link configurations.

The following figure shows the network configuration. A and B exchange untagged packets over SMLT.

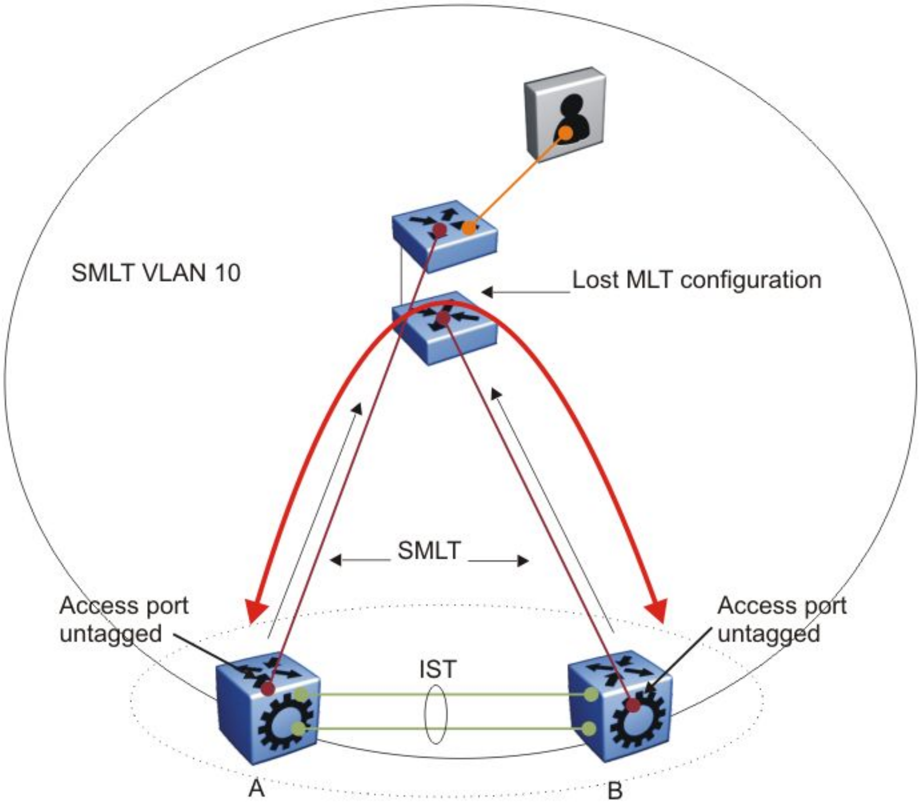

Figure 35: Untagged SMLT links

For the network shown in the preceding figure, the configuration consists of the following:

- SLPP-Tx is enabled on SMLT VLAN-10.
- On switches A and B, SLPP-Rx is enabled on untagged access SMLT links.
- On switch A, the SLPP-Rx threshold is 5.
- In case of a network failure, to avoid edge isolation, the SLPP rx-threshold is 50 on SMLT switch B.

The configuration in Figure 35: Untagged SMLT links on page 86 detects loops and avoids edge isolation. For tagged data, consider the configuration in the following figure.

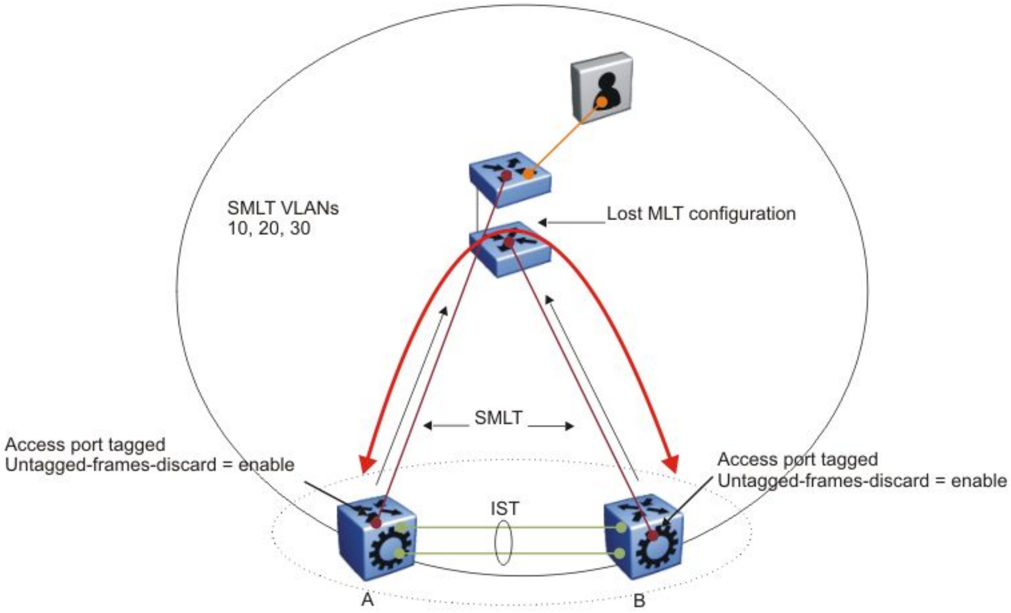

Figure 36: Tagged SMLT links

The configuration changes to

- SLPP-Tx is enabled on SMLT VLANs 10, 20, and 30. A loop in one of these VLANs triggers an event and resolves the loop.
- On switches A and B, SLPP-Rx is enabled on tagged SMLT access links.
- On switch A, the SLPP Rx threshold is 5.
- On SMLT switch B, the SLPP Rx threshold is 50 to avoid edge isolation in the case of a network failure.

In this scenario, Avaya recommends that you enable the untagged-frames-discard parameter on the SMLT uplink ports.

For square and full mesh configurations that use a bridged core (Layer 2 VLANs extend from the edge through all switches in the core), Avaya recommends that you enable SLPP on the primary switches. Enable SLPP on half the core to prevent possible loops and not allow a loop at the edge of the network to shut down the entire core.

For square and full mesh configurations that use a routed core, Avaya recommends that you use or create a separate core VLAN and enable SLPP on that VLAN and the square or full mesh links between the switch clusters. This configuration catches loops created in the core and loops at the edge do not affect core ports. If you use Routed SMLT (RSMLT) between the switch clusters, enable SLPP on the RSMLT VLAN.

SLPP configuration considerations and recommendations

SLPP uses an individual VLAN hello packet mechanism to detect network loops. Sending hello packets on an individual VLAN basis allows SLPP to detect VLAN-based network loops for untagged and tagged IEEE 802.1Q VLAN link configurations. You can decide to which VLANs a switch sends SLPP test packets. All port members of the SLPP-enabled VLAN replicate the packets.

Use the information in this section to understand the considerations and recommendations when you configure SLPP in an SMLT network:

- You must enable SLPP packet receive on each port to detect a loop.
- Vary the SLPP packet receive threshold between the two core SMLT switches so that if a loop is detected, the access ports on both switches do not go down, and to avoid SMLT client isolation.
- SLPP test packets (SLPP-PDU) are forwarded for each VLAN.
- SLPP-PDUs are automatically forwarded VLAN ports configured for SLPP.
- The SLPP-PDU destination MAC address is the switch MAC address (with the multicast bit set) and the source MAC address is the switch MAC address.
- The SLPP-PDU is sent out as a multicast packet and is constrained to the VLAN on which it is sent.
- If an MLT port receives an SLPP-PDU, the port goes down.
- The originating CP or the peer SMLT CP can receive the SLPP-PDU. All other switches treat the SLPP-PDU as a normal multicast packet, and forward it to the VLAN.
- SLPP is port-based, so a port is disabled if it receives SLPP-PDU on one or more VLANs on a tagged port. For example, if the SLPP packet receive threshold is 5, a port is shut down if it receives 5 SLPP-PDU from one or more VLANs on a tagged port.
- The switch does not act on SLPP packets other than those that it transmits.
- In an SMLT cluster, Avaya recommends that you configure an SLPP packet-Rx threshold of 5 on the primary switch and 50 on the secondary switch .
- For square and full-mesh configurations that use a routed core, Avaya recommends that you create a separate core VLAN. Enable SLPP on the core VLAN and the square or full-mesh links between switch clusters. This configuration detects loops created in the core and loops at the edge do not affect core ports.

  If you use RSMLT between the switch clusters, enable SLPP on the RSMLT VLAN. Because you enable SLPP only on one or two VLANs in the core, changing the RX-threshold values will not be necessary.
- You can tune network failure behavior by choosing how many SLPP packets must be received before a switch takes action.
- SLPP-Tx operationally disables ports that receive their own SLPP packet.

Avaya recommends the values in the following table.

Table 11: SLPP recommended values

	Configuration
Enable SLPP	
Access SMLT	Yes
Core SMLT	No

	Configuration
Enable SLPP	
IST	No
Primary switch	
Packet Rx threshold	5
Transmission interval	500 milliseconds (ms) (default)
Secondary switch	
Packet Rx threshold	50
Transmission interval	500 ms (default)

Loop Detect

Use the Loop Detect feature at the edge of a network to prevent loops. This feature detects whether the same MAC address appears on different ports. This feature can disable a VLAN or a port. The Loop Detect feature can also disable a group of ports if it detects the same MAC address on two different ports five times in a configurable amount of time.

On a individual port basis, the Loop Detect feature detects MAC addresses that are looping from one port to other ports. After the switch detects a loop, it disables the port on which the MAC addresses were learned. Additionally, if Loop Detect finds a MAC address loop, it disables the MAC address for that VLAN.

ARP Detect

The ARP-Detect feature is an enhancement over Loop Detect to account for ARP packets on IP configured interfaces. For network loops that involve ARP frames on routed interfaces, Loop-Detect does not detect the network loop condition due to how ARP frames are copied to the CPU. Use ARP-Detect on Layer 3 interfaces. The ARP-Detect feature supports only the vlan-block and port-down options.

VLACP

This feature provides an end-to-end failure detection mechanism, which prevents potential problems caused by misconfigurations in a switch cluster design.

Configure VLACP on an individual port basis. The system forwards traffic only across the uplinks when VLACP is up and running correctly. You must configure the ports on each end of the link for VLACP. VLACP takes the point-to-point hello mechanism of LACP and uses it to periodically send PDU packets to ensure end-to-end reachability and provide failure detection (across a Layer 2 domain). If one end of the link does not receive the VLACP PDUs, it logically disables that port and no traffic passes. This action insures that even if no link exists on the port at the other end, if it is not processing VLACP PDUs correctly, no traffic is sent. This function alleviates potential black hole situations by only sending traffic to ports that are functioning properly.

You can reduce VLACP timers to 400 milliseconds between a pair of Avaya Virtual Services Platform 9000. This timer provides approximately one second failure detection and switchover. When you configure VLACP, you must configure both ends of the link with the same multicast MAC address and timers. Most products in the Avaya Ethernet switch and Ethernet routing switch line use the same timers, with the exception of the FastPeriodicTimer, which is 200ms on the Ethernet Routing Switch 8800 and 500ms on all other switches.

You can use VLACP as a loop prevention mechanism in SMLT configurations when you configure the IST. VLACP also protects against CPU failures by switching or rerouting traffic to the SMLT peer in the case the CPU fails or stops responding. For more information about VLACP in SMLT networks, see SMLT and VLACP on page 55

Loop prevention recommendations

VSP 9000 supports Loop Detect and SLPP. Do not enable SLPP on IST links.

CPU protection and loop prevention compatibility

Avaya recommends several best-practice methods for loop prevention, especially in a switch cluster environment. For more information about loop detection and compatibility for each software release, see *The Large Campus Technical Solution Guide*, NN48500-575.

Chapter 13: Spanning tree

Spanning tree prevents loops in switched networks. Virtual Services Platform 9000 supports Rapid Spanning Tree Protocol (RSTP) and Multiple Spanning Tree Protocol (MSTP). This section describes issues to consider when you configure spanning tree protocols.

For more information about spanning tree protocols, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500.

Spanning tree and protection against isolated VLANs

Virtual Local Area Network (VLAN) isolation disrupts packet forwarding. Figure 37: VLAN isolation on page 92 shows the problem. Two VLANs (V1 and V2) connect four devices, and both VLANs are in the same STG. V2 includes three of the four devices, whereas V1 includes all four devices. After a spanning tree protocol detects a loop, it blocks the link with the highest link cost. In this case, the 100 Mbit/s link is blocked, which isolates a device in V2. To avoid this problem, either configure V2 on all four devices or use MSTP with a different Multiple Spanning Tree Instance (MSTI) for each VLAN.

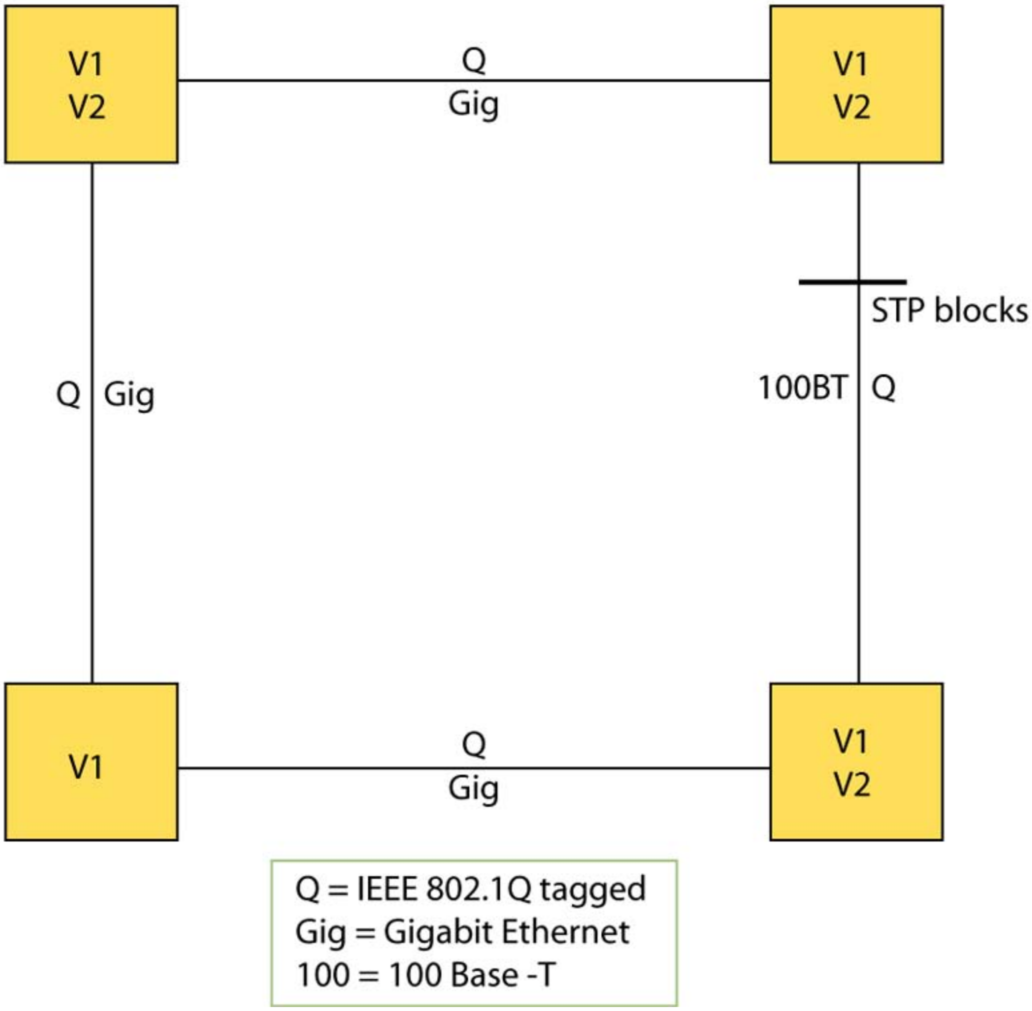

Figure 37: VLAN isolation

MSTP and RSTP considerations

The Spanning Tree Protocol (STP) provides loop protection and recovery, but it is slow to respond to a topology change in the network (for example, a dysfunctional link in a network). The RSTP (IEEE 802.1w) and MSTP (IEEE 802.1s) protocols reduce the recovery time after a network failure. RSTP and MSTP also maintain a backward compatibility with IEEE 802.1D. Typically, the recovery time of RSTP and MSTP is less than 1 second. RSTP and MSTP also reduce the amount of flooding in the network by enhancing the way that Topology Change Notification (TCN) packets are generated.

Use MSTP to configure MSTIs on the same switch. Each MSTI can include one or more VLANs.

In MSTP mode you can configure up to 64 instances. Instance 0 or Common and Internal Spanning Tree (CIST) is the default group, which includes default VLAN 1. Instances 1 to 63 are MSTIs.

RSTP and MSTP provide a global spanning tree parameter, called **`version`**, for backward compatibility with legacy STP. You can configure **`version`** to either STP-compatible mode, RSTP mode, or MSTP mode:

- An STP-compatible port transmits and receives only STP Bridge Protocol Data Units (BPDU). An RSTP or MSTP BPDU that the port receives in this mode is discarded.
- An RSTP or MSTP port transmits and receives only RSTP or MSTP BPDUs. If an RSTP or MSTP port receives an STP BPDU, it becomes an STP port. You must manually intervene to configure this port for RSTP or MSTP mode again. This process is called Port Protocol Migration.

You must be aware of the following recommendations before you implement MSTP or RSTP:

- The default mode is MSTP. A special boot configuration flag identifies the mode.
- You can lose your configuration if you change the spanning tree mode from MSTP to RSTP and the configuration file contains VLANs configured with MSTI greater than 0. RSTP only supports VLANs configured with the default instance 0.
- For best interoperability results, contact your Avaya representative.

Chapter 14: Layer 3 network design

This section describes Layer 3 design considerations that you need to understand to properly design an efficient and robust network.

VRF Lite

Virtual Services Platform 9000 supports the Virtual Router Forwarding (VRF) Lite feature, which supports many virtual routers, each with its own routing domain. VRF Lite virtualizes the routing tables to form independent routing domains, which eliminates the need for multiple physical routers.

To use VRF Lite, you must use the Premier Software License.

VRF Lite supports the High Availability feature. Dynamic tables built by VRF Lite are synchronized. If failover occurs after you enable HA, VRF Lite does not experience an interruption.

Virtual Services Platform 9000 provides the MgmtRouter VRF by default. Use this VRF to configure the management port for out-of-band (OOB) management. You cannot delete this VRF.

For more information about VRF Lite, see *Avaya Virtual Services Platform 9000 Configuration — IP Routing*, NN46250-505.

VRF Lite route redistribution

Using VRF Lite, Virtual Services Platform 9000 can function as many routers; each VRF routing engine works independently. Normally, no route leak occurs between different VRFs. Use the route redistribution option to facilitate the redistribution of routes. VRFs can redistribute Open Shortest Path First (OSPF), Routing Information Protocol (RIP), Border Gateway Protocol (BGP), direct, and static routes.

If you enable route redistribution between two VRFs, ensure that the IP addresses do not overlap. The software does not enforce this requirement.

VRF Lite capability and functionality

On a VRF instance, VRF Lite supports the following protocols: IP, Internet Control Message Protocol (ICMP), Address Resolution Protocol (ARP), static routes, default routes, RIP, OSPF, external BGP (eBGP), route policies, Virtual Router Redundancy Protocol (VRRP), and the Dynamic Host Configuration Protocol/BootStrap Protocol relay agent.

The device uses VRF Lite to perform the following actions:

- partition traffic and data, and represent an independent router in the network
- provide virtual routers that are transparent to end-users
- support overlapping IP address spaces in separate VRFs
- support addresses that are not restricted to the assigned address space given by host Internet Service Providers (ISP)
- support Split MultiLink Trunking (SMLT) and Routed SMLT (RSMLT)
- support eBGP

VRF Lite architecture examples

VRF Lite enables a router to act as many routers. This provides virtual traffic separation for each user and provides security. For example, you can use VRF Lite to

- provide different departments within a company with site-to-site connectivity as well as Internet access
- provide centralized and shared access to data centers.

The following figure shows how VRF Lite can emulate VPNs.

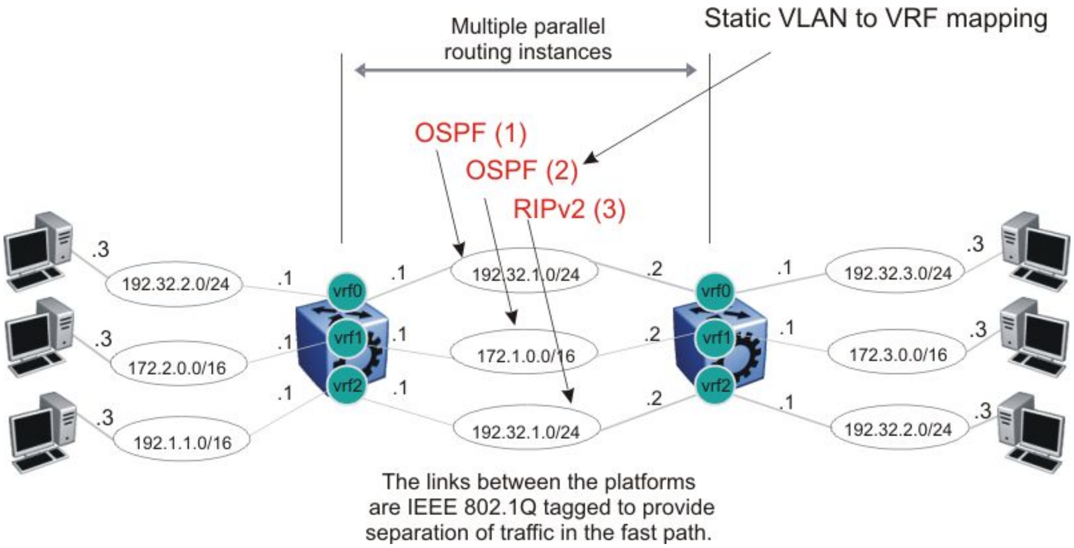

Figure 38: VRF Lite example

The following figure shows how you can use VRF Lite in an SMLT topology. VRRP runs between the two bottom routers.

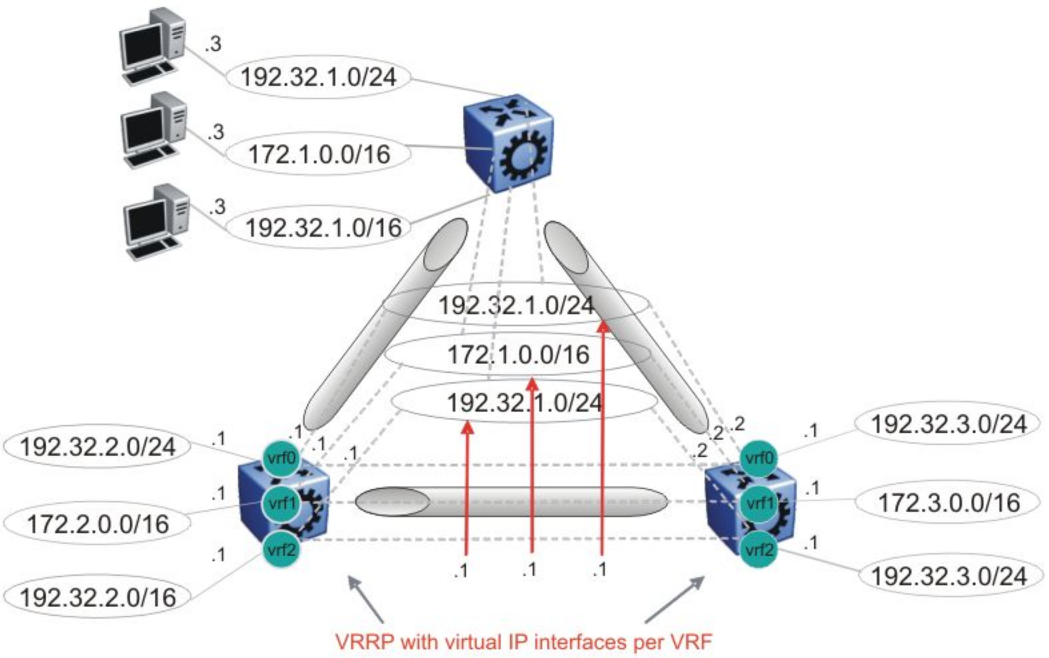

Figure 39: VRRP and VRF in SMLT topology

The following figure shows how you can use VRF Lite in an RSMLT topology.

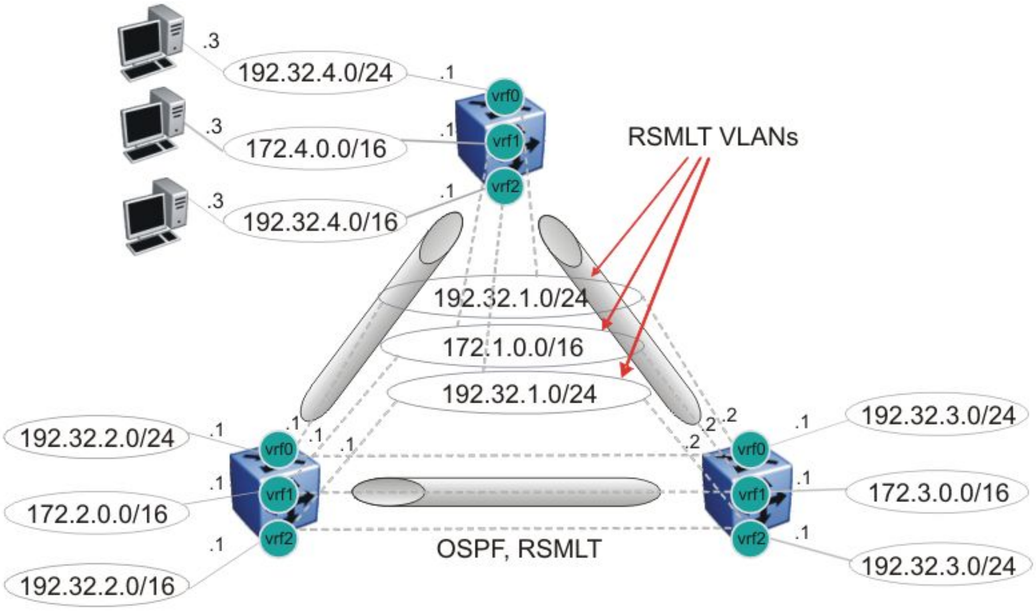

Figure 40: Router redundancy for multiple routing instances (using RSMLT)

The following figure shows how VRFs can interconnect through an external firewall.

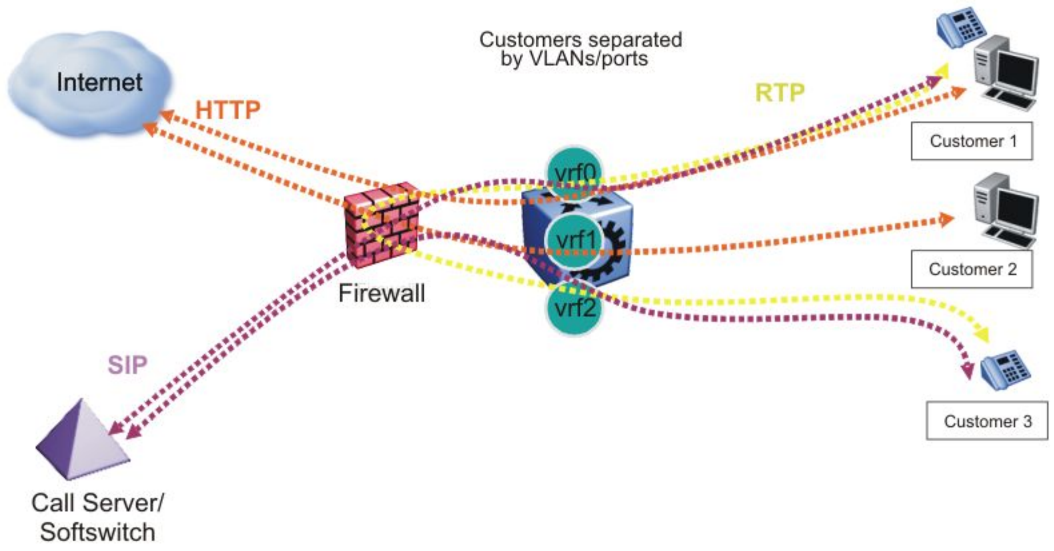

Figure 41: Inter-VRF forwarding based on external firewall

Although customer data separation into Layer 3 virtual routing domains is usually a requirement, sometimes customers must access a common network infrastructure. For example, they want to access the Internet, data storage, VoIP-PSTN, or call signaling services. To interconnect VRF instances, you can use an external firewall that supports virtualization, or use inter-VRF forwarding for specific services. Using the inter-VRF solution, you can use routing policies and static routes to inject IP subnets from one VRF instance to another, and filters to restrict access to certain protocols.

The following figure shows inter-VRF forwarding. In this solution, you can use routing policies to leak IP subnets from one VRF to another. You can use filters to restrict access to certain protocols. This configuration enables hub-and-spoke network designs, for example, for VoIP gateways.

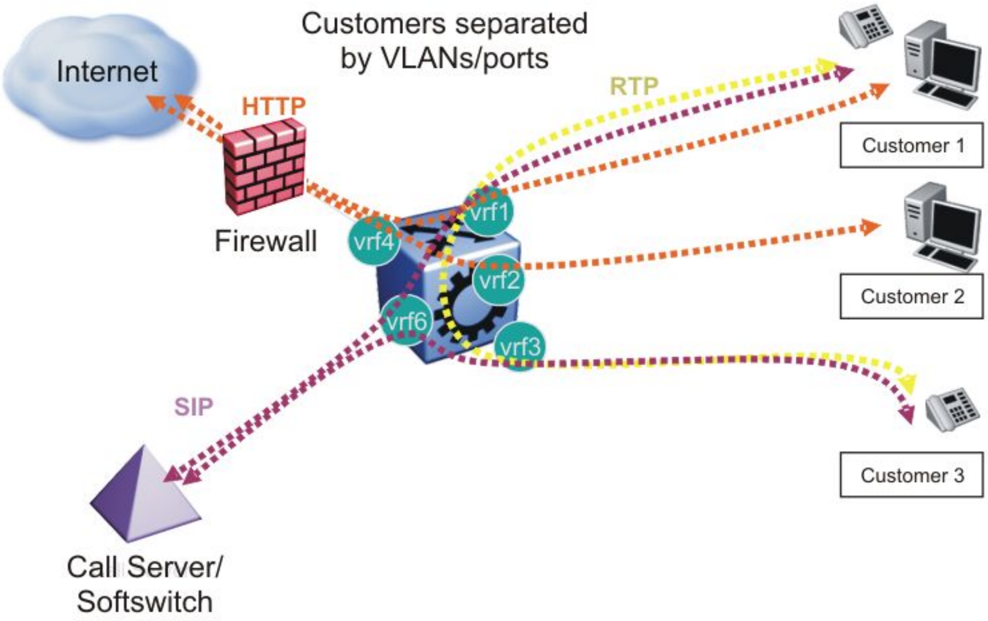

Figure 42: Inter VRF communication, internal inter-VRF forwarding

Virtual Router Redundancy Protocol

The Virtual Router Redundancy Protocol provides a backup router that takes over if a router fails, which is important if you must provide redundancy mechanisms.

VRRP guidelines

VRRP provides another layer of resiliency to your network design by providing default gateway redundancy for end users. If a VRRP-enabled router that connects to the default gateway fails, failover to the VRRP backup router ensures no interruption for end users who attempt to route from their local subnet.

Typically in an SMLT network, only the VRRP Master router forwards traffic for a given subnet. The backup VRRP router does not route traffic destined for the default gateway. Instead, the backup router employs Layer 2 switching on the IST to deliver traffic to the VRRP Master for routing.

To allow both VRRP switches to route traffic, Virtual Services Platform 9000 has an extension to VRRP, the BackupMaster, that creates an active-active environment for routing. If you enable BackupMaster on the backup router, the backup router no longer switches traffic to the VRRP Master. Instead the BackupMaster routes all traffic received on the BackupMaster IP interface according to the switch routing table. This configuration prevents the edge switch traffic from unnecessarily utilizing the IST to reach the default gateway.

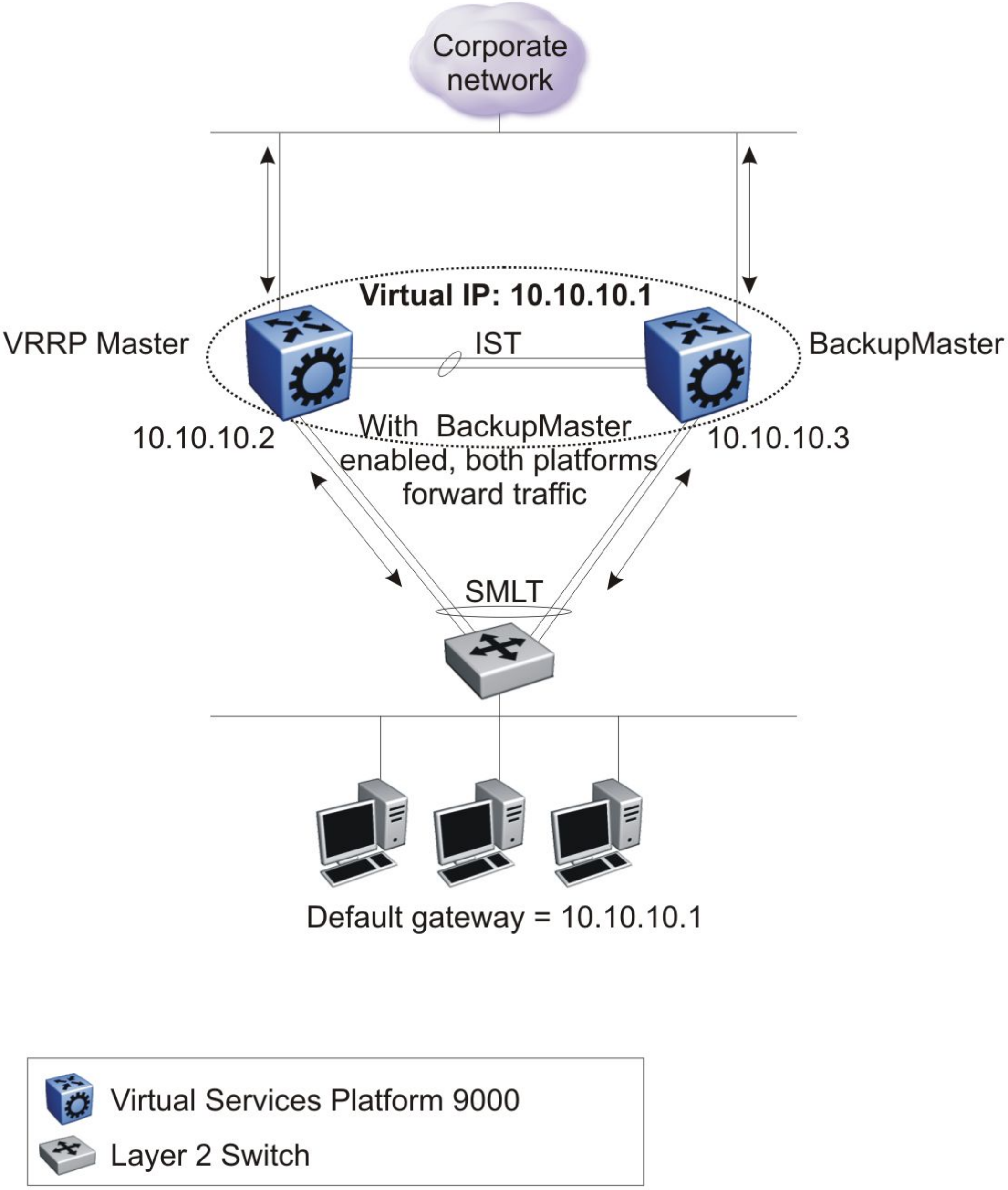
Corporate network
Virtual IP: 10.10.10.1
VRRP Master
IST
BackupMaster
10.10.10.2
With BackupMaster enabled, both platforms forward traffic
10.10.10.3
SMLT
Default gateway = 10.10.10.1
Virtual Services Platform 9000
Layer 2 Switch

Figure 43: VRRP with BackupMaster

Avaya recommends that you stagger VRRP instances on a network or subnet basis. The following figure shows the VRRP Masters and BackupMasters for two subnets. For more information about how to configure VRRP using ACLI and Enterprise Device Manager (EDM), see *Avaya Virtual Services Platform 9000 Configuration — IP Routing*, NN46250-505.

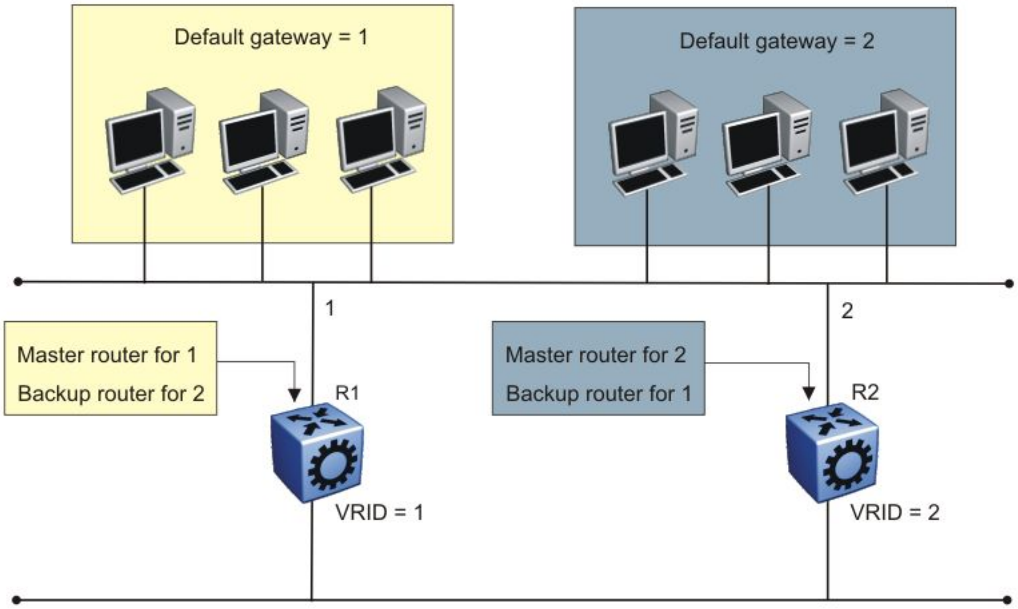

Figure 44: VRRP network configuration

Avaya recommends that you use a VRRP BackupMaster configuration with an SMLT configuration that has an existing VRRP configuration.

The VRRP BackupMaster uses the VRRP standardized backup switch state machine. Thus, VRRP BackupMaster is compatible with standard VRRP.

Avaya recommends that you use the following best practices to implement VRRP:

- Do not configure the virtual address as a physical interface that is used on the routing switches. Instead, use a third address, for example:
    - Interface IP address of VLAN A on Switch 1 = x.x.x.2
    - Interface IP address of VLAN A on Switch 2 = x.x.x.3
    - Virtual IP address of VLAN A = x.x.x.1
- Configure the VRRP hold down timer long enough that the Interior Gateway Protocol (IGP) routing protocol has time to converge and update the routing table. In some cases, configuring the VRRP hold down timer to a minimum of 1.5 times the IGP convergence time is sufficient. For OSPF, Avaya recommends that you use a value of 90 seconds if you use the default OSPF timers.
- Implement VRRP BackupMaster for an active-active configuration (BackupMaster works across multiple switches that participate in the same VRRP domain.).
- Configure VRRP priority as 200 to configure VRRP Master.

- Stagger VRRP Masters between switches in the core to balance the load between switches.
- If you use multiple VLANs with VRRP enabled, Avaya recommends that you stagger the VRRP Master such that both SMLT cluster switches are VRRP Master for half the VLANs.
- If you implement VRRP Fast, you create additional control traffic on the network and also create a greater load on the CPU. To reduce the convergence time of VRRP, the VRRP Fast feature allows the modification of VRRP timers to achieve subsecond failover of VRRP. Without VRRP Fast, normal convergence time is approximately 3 seconds.
- Ensure that both SMLT aggregation switches can reach the same destinations using a routing protocol. For routing purposes, configure individual VLAN addresses on both SMLT aggregation switches.
- Introduce an additional subnet on the IST that has a shortest-route-path to avoid issuing ICMP redirect messages on the VRRP subnets. (To reach the destination, ICMP redirect messages are issued if the router sends a packet back out through the same subnet on which it is received).
- Do not use VRRP BackupMaster and critical IP at the same time. Use one or the other.
- When you implement VRRP on multiple VLANs between the same switches, Avaya recommends that you configure a unique VRID on each VLAN.

VRRP and spanning tree

Virtual Services Platform 9000 can use one of two spanning tree protocols. These include the Rapid Spanning Tree Protocol (RSTP) and the Multiple Spanning Tree Protocol (MSTP).

VRRP protects clients and servers from link or aggregation switch failures. Configure the network to limit the amount of time a link is down during VRRP convergence. The following figure shows two possible configurations of VRRP and spanning tree; configuration A is optimal and configuration B is not.

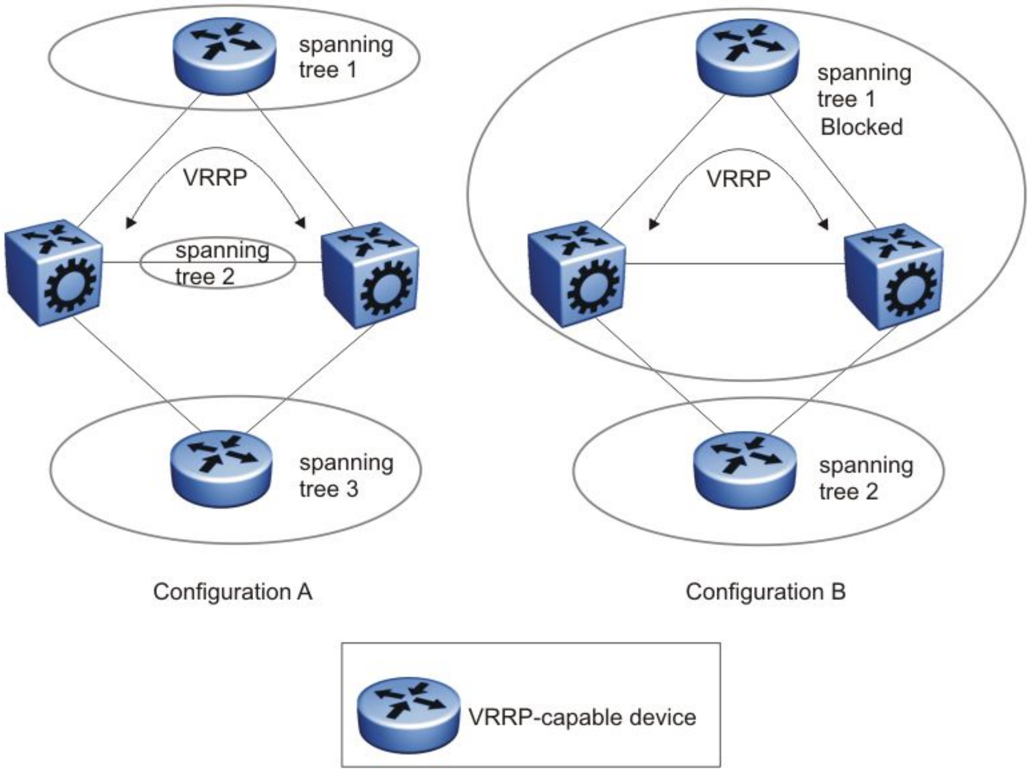

Figure 45: VRRP and STG configurations

In this figure, configuration A is optimal because VRRP convergence occurs within 2 to 3 seconds. In configuration A, three spanning tree instances exist and VRRP runs on the link between the two routers. Spanning tree instance 2 exists on the link between the two routers, which separates the link between the two routers from the spanning tree instances found on the other devices. All uplinks are active.

In configuration B, VRRP convergence takes between 30 and 45 seconds because it depends on spanning tree convergence. After initial convergence, spanning tree blocks one link (an uplink), so only one uplink is used. If an error occurs on the uplink, spanning tree reconverges, which can take up to 45 seconds. After spanning tree reconvergence, VRRP can take a few more seconds to failover.

Avaya recommends that you enable SMLT with VRRP to simplify the network configuration and reduce the failover time. SMLT turns off spanning tree for SMLT ports. For VRRP and SMLT information, see SMLT and Layer 3 traffic redundancy (VRRP and RSMLT) on page 58.

VRRP and ICMP redirect messages

You can use VRRP and ICMP together. However, doing so can provide nonoptimal network performance.

Consider the network shown in the following figure. Traffic from the client on subnet 30.30.30.0, destined for the 10.10.10.0 subnet, is sent to routing switch 1 (VRRP Master). Routing switch 1 forwards this traffic on the same subnet to routing switch 2 where it is routed to the destination.

With ICMP redirect enabled, for each packet received, routing switch 1 sends an ICMP redirect message to the client to inform it of a shorter path to the destination through routing switch 2.

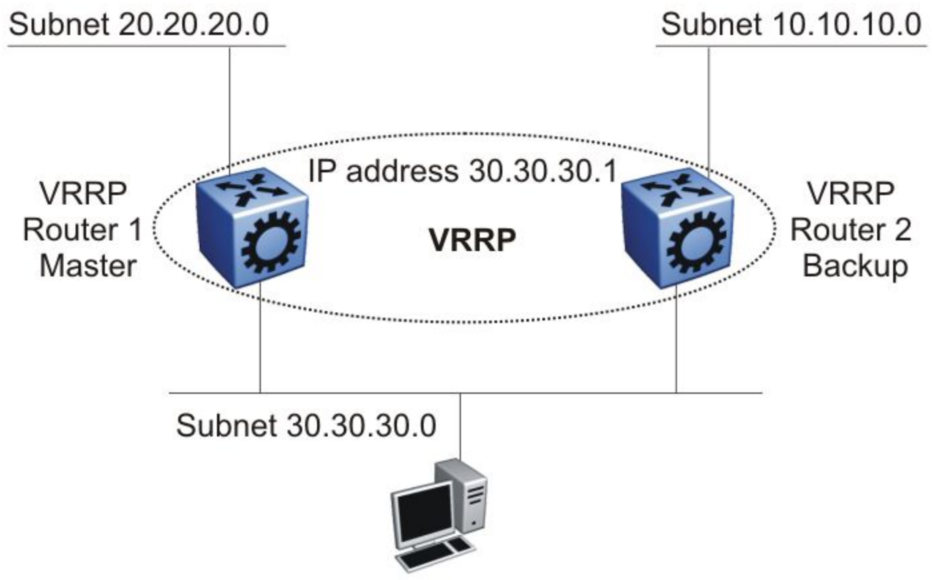

Figure 46: ICMP redirect messages

If network clients do not recognize ICMP redirect messages, disable ICMP redirect messages on Avaya Virtual Services Platform 9000 to avoid excessive ICMP redirect messages. Avaya recommends the network designs shown in the following figures.

Ensure that the routing path to the destination through both routing switches has the same metric to the destination. One hop goes from 30.30.30.0 to 10.10.10.0 through routing switch 1 and routing switch 2. Do this by building symmetrical networks based on the network design examples in Modular design for redundant networks on page 49.

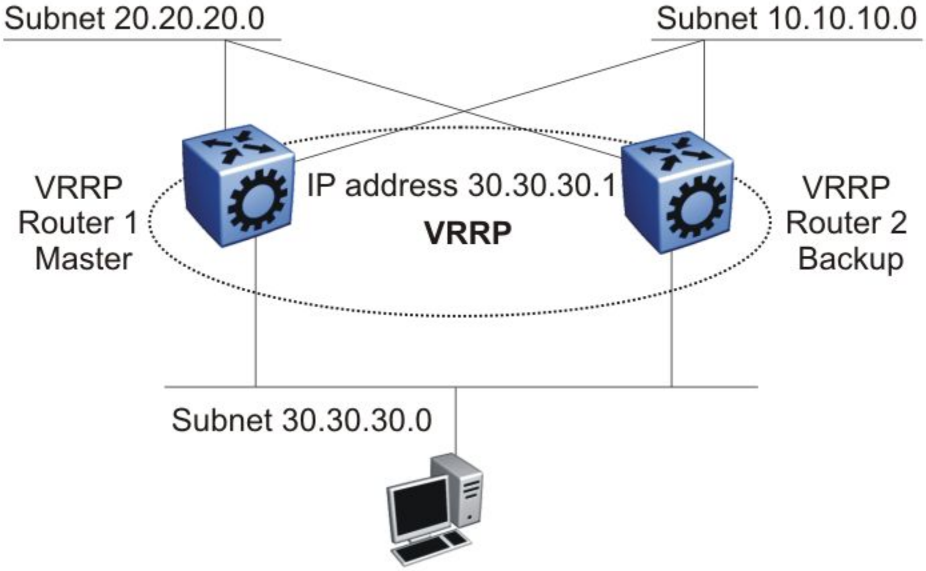

Figure 47: Avoiding excessive ICMP redirect messages without SMLT

Alternatively, in an SMLT environment, you can create a VLAN on the IST between the two routing switches that uses a lower cost than the 30.30.30.0 subnet. With ICMP redirect disabled, the network does not generate ICMP redirect messages and traffic from the client passes directly to router 1 and router 1 sends the message without ICMP messages.

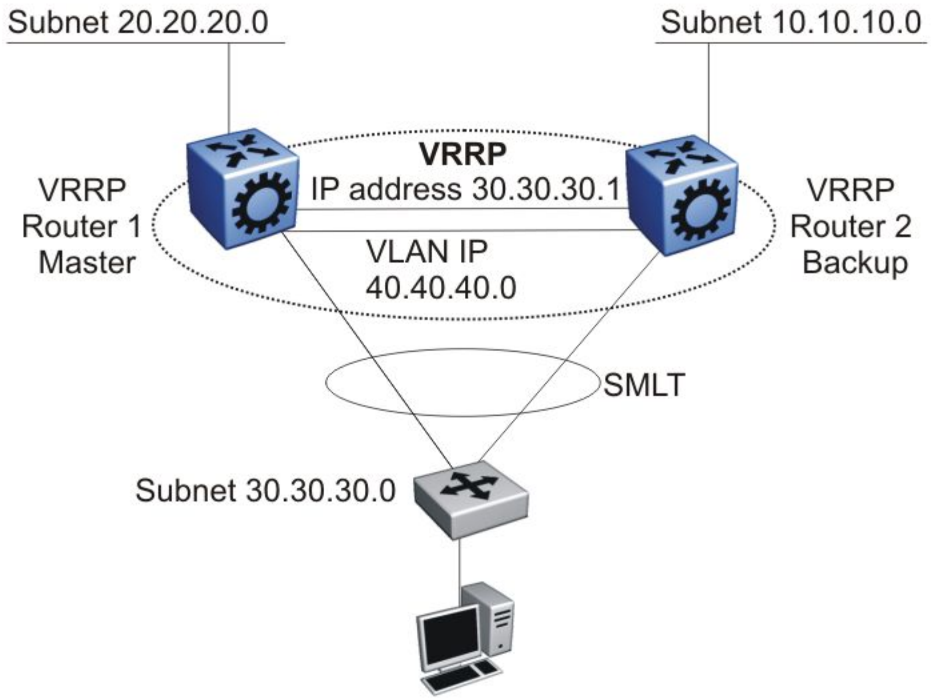

Figure 48: Avoiding excessive ICMP redirect messages with SMLT

VRRP versus RSMLT for default gateway resiliency

A better alternative than VRRP is to use RSMLT Layer 2 Edge. Avaya recommends that you use an RSMLT Layer 2 Edge configuration, rather than VRRP, for those products that support RSMLT Layer 2 Edge.

RSMLT Layer Edge provides the following advantages:

- Greater scalability—VRRP scales to 255 instances, while RSMLT scales to the maximum number of VLANs.
- Simpler configuration—Enable RSMLT on a VLAN; VRRP requires virtual IP configuration, along with other parameters.

For connections in pure Layer 3 configurations (using a static or dynamic routing protocol), Avaya recommends that you use a Layer 3 RSMLT configuration over VRRP. In these instances, an RSMLT configuration provides faster failover than one with VRRP because the connection is a Layer 3 connection, not just a Layer 2 connection for default gateway redundancy.

Both VRRP and RSMLT can provide resiliency for the default gateway of an end station. The configurations of these features are different, but both provide the same end result and are transparent to the end station.

For more information about RSMLT, see Routed SMLT on page 65.

Subnet-based VLAN guidelines

You can use subnet-based VLANs to classify end-users in a VLAN based on the end-user source IP addresses. For each packet, the switch performs a lookup, and, based on the source IP address and mask, determines to which VLAN the traffic belongs. To provide security, use subnet-based VLANs to allow only users on the appropriate IP subnet access to the network.

You cannot classify non-IP traffic using a subnet-based VLAN.

You can enable routing in each subnet-based VLAN by assigning an IP address to the subnet-based VLAN. If you do not configure an IP address, the subnet-based VLAN is in Layer 2 switch mode only.

You can enable VRRP for subnet-based VLANs. Hardware forwards the traffic routed by the VRRP Master interface. Therefore, no throughput impact occurs.

You can use subnet-based VLANs to achieve multinetting functionality; however, multiple subnet-based VLANs on a port can only classify traffic based on the sender IP source address. You cannot multinet by using multiple subnet-based VLANs between routers (Layer 3 devices). All end-user-facing ports support multinetting.

You cannot classify Dynamic Host Configuration Protocol (DHCP) traffic into subnet-based VLANs because DHCP requests do not carry a specific source IP address; instead, they use an all broadcast address. To support DHCP to classify subnet-based VLAN members, create an overlay port-based VLAN to collect the bootp and DHCP traffic and forward it to the appropriate DHCP server. After the DHCP response is forwarded to the DHCP client and it learns the source IP address, the end-user traffic is appropriately classified into the subnet-based VLAN.

The switch supports a maximum of 256 subnet-based VLANs.

Open Shortest Path First

Use OSPF to ensure that the switch can communicate with other OSPF-speaking routers. This section describes some general design considerations and presents a number of design scenarios for OSPF.

For more information about OSPF concepts and configuration, see *Avaya Virtual Services Platform 9000 Configuration — OSPF and RIP*, NN46250-506.

OSPF LSA limits

To determine OSPF link state advertisement (LSA) limits:

1. Use the command **`show ip ospf area`** to determine the LSA_CNT and to obtain the number of LSAs for a given area.
2. Use the following formula to determine the number of areas. Ensure the total is less than 40 000 (40K):

   $$\sum \text{Adj}_N * \text{LSA\_CNT}_N < 40\text{k}$$

   N = 1 to the number of areas for each switch

   Adj_N = number of adjacencies for each Area N

   LSA_CNT_N = number of LSAs for each Area N

For example, assume that a switch has a configuration of three areas with a total of 18 adjacencies and 1000 routes. This includes:

- 3 adjacencies with an LSA_CNT of 500 (Area 1)
- 10 adjacencies with an LSA_CNT of 1000 (Area 2)
- 5 adjacencies with an LSA_CNT of 200 (Area 3)

Calculate the number as follows:

3*500+10*1000+5*200=12.5K < 40K

This configuration ensures that the switch operates within accepted scalability limits.

OSPF design guidelines

Follow these additional OSPF guidelines:

- OSPF timers must be consistent across the entire network.
- Use OSPF area summarization to reduce routing table sizes.
- Use OSPF passive interfaces to reduce the number of active neighbor adjacencies.
- Use OSPF active interfaces only on intended route paths.

  Configure wiring closet subnets as OSPF passive interfaces unless they form a legitimate routing path for other routes.
- Minimize the number of OSPF areas for each switch to avoid excessive shortest path calculations.

  The switch executes the Djikstra algorithm for each area separately.
- Ensure that the OSPF dead interval is at least four times the OSPF hello interval
- Use MD5 authentication on untrusted OSPF links.
- Use stub or NSSA areas as much as possible to reduce CPU overhead.

OSPF and CPU utilization

After you create an OSPF area route summary on an area border router (ABR), the summary route can attract traffic to the ABR for which the router does not have a specific destination route. Enabling ICMP unreachable message generation on the switch can result in a high CPU utilization rate.

To avoid high CPU utilization, Avaya recommends that you use a black hole static route configuration. The black hole static route is a route (equal to the OSPF summary route) with a next-hop of 255.255.255.255. This configuration ensures that all traffic that does not have a specific next-hop destination route is dropped.

OSPF network design examples

You can use OSPF routing in the core of an RSMLT network. For more information, see Layer 1, 2, and 3 design examples on page 175.

The following figure describes a simple implementation of an OSPF network: enabling OSPF on two switches (S1 and S2) that are in the same subnet in one OSPF area.

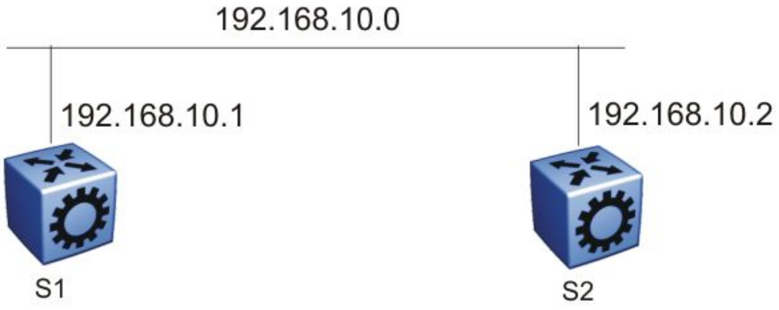

Figure 49: Example 1: OSPF on one subnet in one area

The routers in the preceding figure use the following configuration:

- S1 has an OSPF router ID of 1.1.1.1, and the OSPF port uses an IP address of 192.168.10.1.
- S2 has an OSPF router ID of 1.1.1.2, and the OSPF port uses an IP address of 192.168.10.2.

The general method to configure OSPF on each routing switch is:

1. Enable OSPF globally.
2. Enable IP forwarding on the switch.
3. Configure the IP address, subnet mask, and VLAN ID for the port.
4. Disable RIP on the port, if you do not need it.
5. Enable OSPF for the port.

After you configure S2, the two switches elect a designated router (DR) and a backup designated router (BDR). They exchange hello packets to synchronize their link state databases (LSDB).

The following figure shows a configuration in which OSPF operates on three switches. OSPF performs routing on two subnets in one OSPF area. In this example, S1 directly connects to S2, and S3 directly connects to S2, but traffic between S1 and S3 is indirect, and passes through S2.

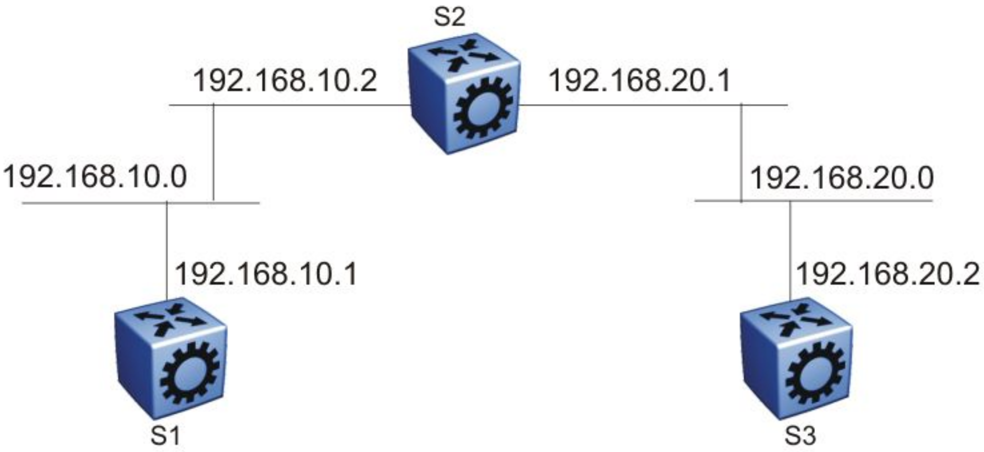

Figure 50: Example 2: OSPF on two subnets in one area

The routers in example 2 use the following configuration:

- S1 has an OSPF router ID of 1.1.1.1, and the OSPF port uses an IP address of 192.168.10.1.
- S2 has an OSPF router ID of 1.1.1.2, and two OSPF ports use IP addresses of 192.168.10.2 and 192.168.20.1.
- S3 has an OSPF router ID of 1.1.1.3, and the OSPF port uses an IP address of 192.168.20.2.

The general method to configure OSPF on each routing switch is:

1. Enable OSPF globally.
2. Insert IP addresses, subnet masks, and VLAN IDs for the OSPF ports on S1 and S3, and for the two OSPF ports on S2. The two ports on S2 enable routing and establish the IP addresses related to the two networks.
3. Enable OSPF for each OSPF port allocated with an IP address.

After you configure all three switches for OSPF, they elect a DR and BDR for each subnet and exchange hello packets to synchronize their LSDBs.

The following figure shows an example where OSPF operates on two subnets in two OSPF areas. S2 becomes the ABR for both networks.

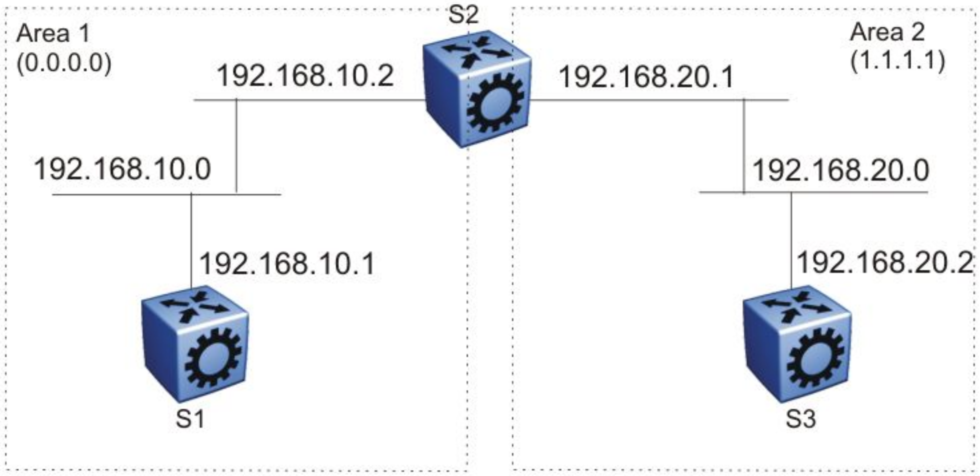

Figure 51: Example 3: OSPF on two subnets in two areas

The routers in scenario 3 use the following configuration:

- S1 has an OSPF router ID of 1.1.1.1. The OSPF port uses an IP address of 192.168.10.1, which is in OSPF area 1.
- S2 has an OSPF router ID of 1.1.1.2. One port uses an IP address of 192.168.10.2, which is in OSPF area 1. The second OSPF port on S2 uses an IP address of 192.168.20.1, which is in OSPF area 2.
- S3 has an OSPF router ID of 1.1.1.3. The OSPF port uses an IP address of 192.168.20.2, which is in OSPF area 2.

The general method to configure OSPF for this three-switch network is:

1. On all three switches, enable OSPF globally.
2. Configure OSPF on one network.

   On S1, insert the IP address, subnet mask, and VLAN ID for the OSPF port. Enable OSPF on the port. On S2, insert the IP address, subnet mask, and VLAN ID for the OSPF port in area 1, and enable OSPF on the port. Both routable ports belong to the same network. Therefore, by default, both ports are in the same area.
3. Configure three OSPF areas for the network.
4. Configure OSPF on two additional ports in a second subnet.

   Configure additional ports and verify that IP forwarding is enabled for each switch to ensure that routing can occur. On S2, insert the IP address, subnet mask, and VLAN ID for the OSPF port in area 2, and enable OSPF on the port. On S3, insert the IP address, subnet mask, and VLAN ID for the OSPF port, and enable OSPF on the port.

The three switches exchange hello packets.

In an environment with a mix of Cisco and Avaya switches and routers, you may need to manually modify the OSPF parameter RtrDeadInterval to 40 seconds.

Border Gateway Protocol

Use the Border Gateway Protocol (BGP) to ensure that the switch can communicate with other BGP-speaking routers on the Internet backbone. BGP is an exterior gateway protocol that exchanges network reachability information with other BGP systems in the same or other autonomous systems (AS). This network reachability information includes information about the AS list that the reachability information traverses. By using this information, you can prune routing loops and enforce policy decisions at the AS level.

BGP performs routing between two sets of routers that operate in different autonomous systems. An AS can use two kinds of BGP: Interior BGP (IBGP), which refers to the protocol that BGP routers use within an autonomous system, and Exterior BGP (EBGP), which refers to the protocol that BGP routers use across two different autonomous systems. BGP information is redistributed to Interior Gateway Protocols (IGP) that run in the autonomous system.

BGP version 4 (BGPv4) supports classless inter-domain routing (CIDR). BGPv4 advertises the IP prefix and eliminates the concept of network class within BGP. BGP4 can aggregate routes and AS paths. BGP aggregation does not occur when routes have different Multi-Exit Discriminators (MED) or next-hops.

BGP Equal-Cost Multipath (ECMP) allows a BGP speaker to perform route balancing within an AS by using multiple equal-cost routes submitted to the routing table by OSPF or RIP. BGP performs load balancing on an individual packet basis.

To control route propagation and filtering, RFC1772 and RFC2270 recommends that multihomed, nontransit Autonomous Systems not run BGPv4. To address the load sharing and reliability requirements of a multihomed user, use BGP between them.

For more information about BGP concepts and configuration, see *Avaya Virtual Services Platform 9000 Configuration — BGP Services*, NN46250-507.

BGP implementation guidelines

To successfully implement BGP in a Virtual Services Platform 9000 network, follow these guidelines:

- BGP does not operate with an IP router in nonforwarding (host-only) mode. Ensure that the routers with which you want BGP to operate are in forwarding mode.
- If you use BGP for a multi-homed AS (one that contains more than a single exit point), Avaya recommends that you use OSPF for the IGP, and BGP for the sole exterior gateway protocol. Otherwise, use intra-AS IBGP routing.
- If OSPF is the IGP, use the default OSPF tag construction. The use of EGP or the modification of the OSPF tags makes network administration and proper configuration of BGP path attributes difficult.
- For routers that support both BGP and OSPF, you must configure the OSPF router ID and the BGP identifier to the same IP address. The BGP router ID automatically uses the OSPF router ID.

- In configurations where BGP speakers reside on routers that have multiple network connections over multiple IP interfaces (the typical case for IBGP speakers), consider using the address of the circuitless (virtual) IP interface as the local peer address. In this configuration, you ensure that BGP is reachable as long as an active circuit exists on the router.
- By default, BGP speakers do not advertise or inject routes into their IGP. You must configure route policies to enable route advertisement.
- Coordinate routing policies among all BGP speakers within an AS so that every BGP border router within an AS constructs the same path attributes for an external path.
- Configure accept and announce policies on all IBGP connections to accept and propagate all routes. Make consistent routing policy decisions on external BGP connections.
- Use the max-prefix parameter to limit the number of routes BGP imports from a peer. Use a configuration of 0 to accept an unlimited number of prefixes.
- You cannot enable or disable the MED selection process. BGP aggregation does not occur when routes have different MEDs or next-hops.

BGP and OSPF interaction

RFC1745 defines the interaction between BGP and OSPF when OSPF is the IGP within an autonomous system. For routers that run both protocols, the OSPF router ID and the BGP ID must be the same IP address. You must configure a BGP route policy to allow BGP advertisement of OSPF routes.

Interaction between BGPv4 and OSPF includes the ability to advertise supernets to support CIDR. BGPv4 supports interdomain supernet advertisements; OSPF can carry supernet advertisements within a routing domain.

BGP and other vendor interoperability

BGP interoperability is compatible between Virtual Services Platform 9000, Cisco 6500 Software Release IOS 11.3, and Juniper M20 Software Release 5.3R2.4.

For more information about BGP, see *Avaya Virtual Services Platform 9000 Configuration — BGP Services*, NN46250-507.

BGP and Internet peering

By using BGP, you can perform Internet peering directly between Virtual Services Platform 9000 and another edge router. In such a scenario, you can use each Virtual Services Platform 9000 for aggregation and peer it with a Layer 3 edge router, as shown in the following figure.

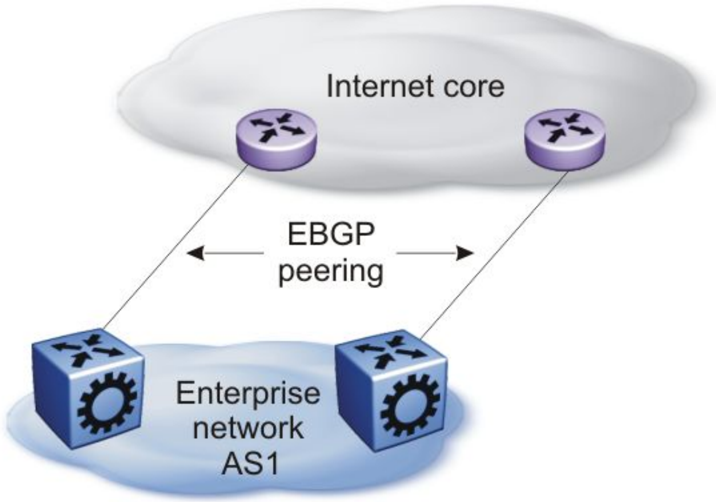

Figure 52: BGP and Internet peering

In cases where the Internet connection is single-homed, to reduce the size of the routing table, Avaya recommends that you advertise Internet routes as the default route to the IGP. Virtual Services Platform 9000 supports three full Internet pairs and 1.5M routes.

Routing domain interconnection with BGP

You can implement BGP so that autonomous routing domains, such as OSPF routing domains, connect. This connection allows the two different networks to begin communicating quickly over a common infrastructure, thus providing additional time to plan the IGP merger. Such a scenario is particularly effective when you need to merge two OSPF area 0.0.0.0s (see the following figure).

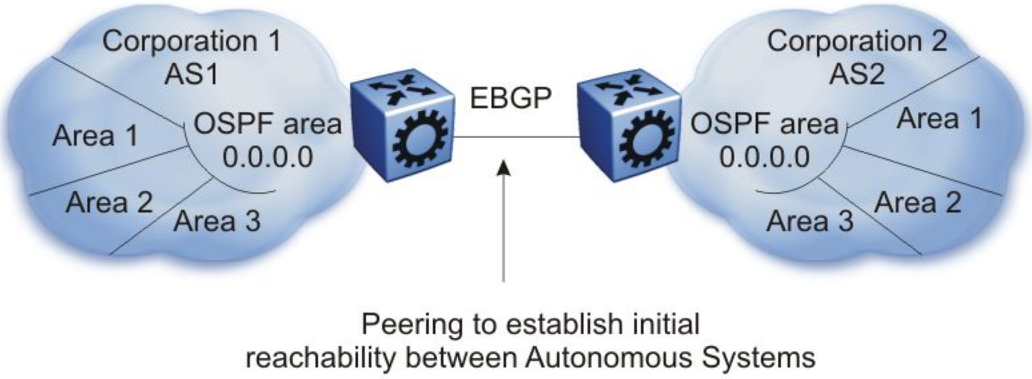

Figure 53: Routing domain interconnection with BGP

BGP and edge aggregation

You can perform edge aggregation with multiple point of presence or edge concentrations. Virtual Services Platform 9000 supports 512 pairs (peering services). You can use BGP to inject dynamic routes rather than using static routes or RIP (see the following figure).

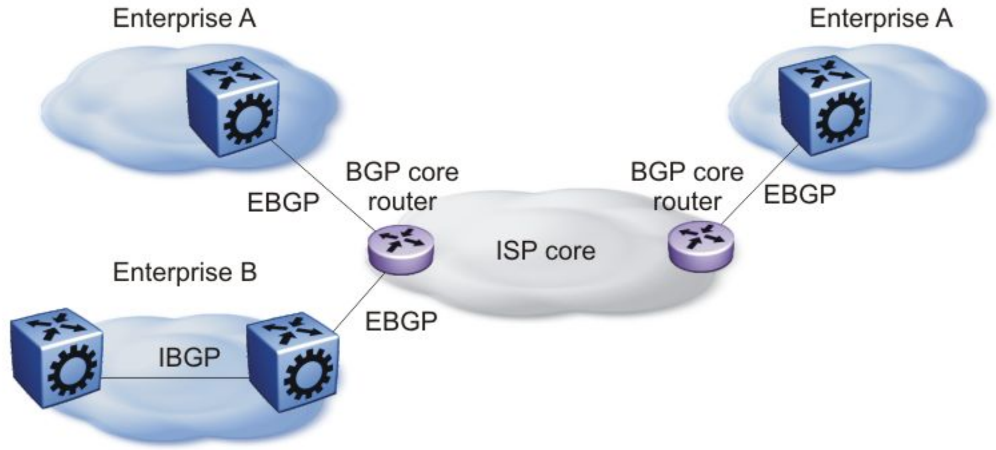

Figure 54: BGP and edge aggregation

BGP and ISP segmentation

You can use the platform as a peering point between different regions or ASs that belong to the same ISP. In such cases, you can define a region as an OSPF area, an AS, or a part of an AS.

You can divide the AS into multiple regions that each run different IGPs. Interconnect regions logically by using a full IBGP mesh. Each region then injects its IGP routes into IBGP and also injects a default route inside the region. For destinations that do not belong to the region, each region defaults to the BGP border router.

Use the community parameter to differentiate between regions. To provide Internet connectivity, this scenario requires you to make your Internet connections part of the central IBGP mesh (see the following figure).

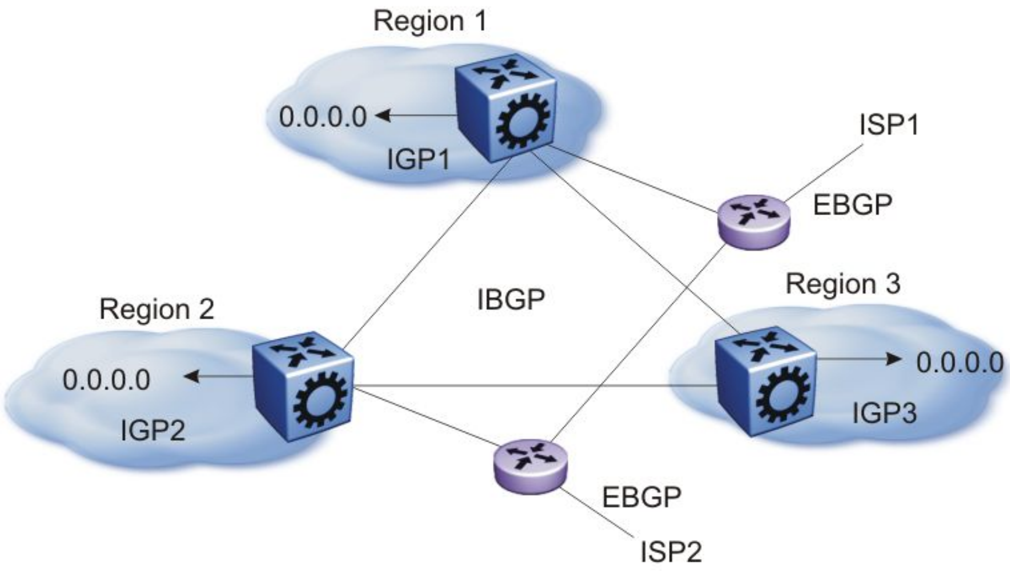

Figure 55: Multiple regions separated by IBGP

In the preceding figure, consider the following

- The AS is divided into three regions that each run different and independent IGPs.
- Regions logically interconnect by using a full-mesh IBGP, which also provides Internet connectivity.
- Internal nonBGP routers in each region default to the BGP border router, which contains all routes.
- If the destination belongs to another region, the traffic is directed to that region; otherwise, the traffic is sent to the Internet connections according to BGP policies.

To configure multiple policies between regions, represent each region as a separate AS. Implement EBGP between ASs, and implement IBGP within each AS. In such instances, each AS injects its IGP routes into BGP where they are propagated to all other regions and the Internet.

The following figure shows the use of EBGP to join several ASs.

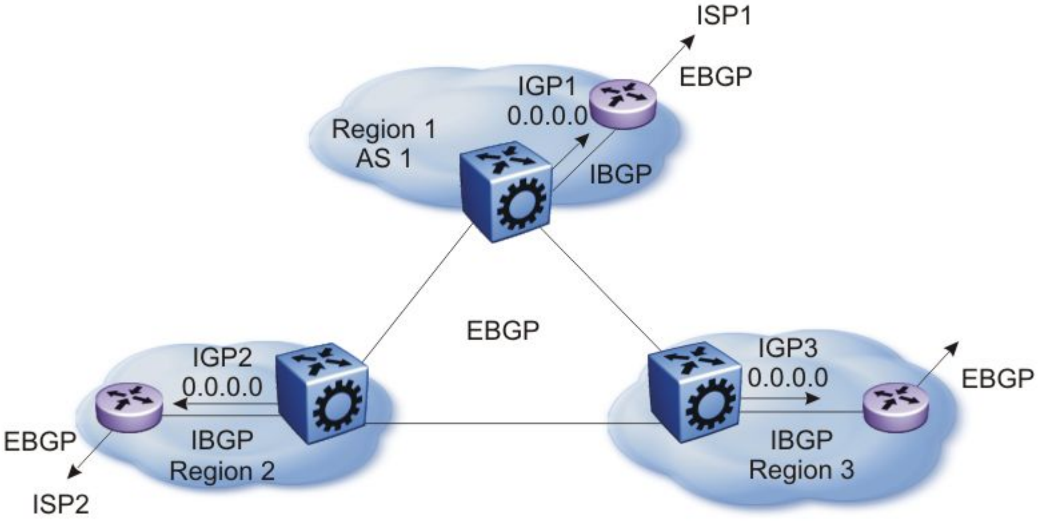

Figure 56: Multiple regions separated by EBGP

You can obtain AS numbers from the Inter-Network Information Center (NIC) or use private AS numbers. If you use private AS numbers, be sure to design your Internet connectivity carefully. For example, you can introduce a central, well-known AS to provide interconnections between all private ASs and the Internet. Before it propagates the BGP updates, this central AS strips the private AS numbers to prevent them from leaking to providers.

The following figure illustrates a design scenario in which you use multiple OSPF regions to peer with the Internet.

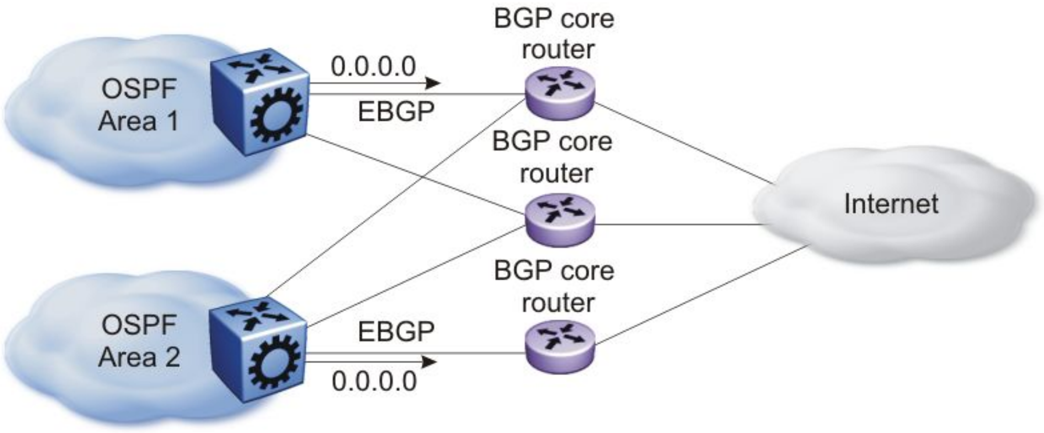

Figure 57: Multiple OSPF regions peering with the Internet

IPv6 BGP+

Virtual Services Platform 9000 extends the BGPv4 process to support the exchange of IPv6 routes using BGPv4 peering.

Release 3.2 does not support native BGPv6 peering over the IPv6 transport layer.

Virtual Services Platform 9000 supports the following BGP extensions:

- RFC2545—BGPv4 multiprotocol extensions for IPv6
- RFC4760—multiprotocol extensions to BGPv4

If you configure BGP on a router that is only enabled for IPv6, you must manually configure the BGP router ID.

BGP+ supports the redistribution of OSPFv3, static IPv6, and directly connected IPv6 routes.

Virtual Services Platform 9000 supports partial HA for BGP+.

IPv6 BGP+ requires that IPv6 peering occurs over the IPv6 tunnel.

So that IPv6 routes learned through the IPv6 peer will point through the IPv6 tunnel, you must configure a static route on each node and these static routes must point toward the remote IPv6 peer through the configured IPv6 tunnel.

IP routed interface scaling

Virtual Services Platform 9000 supports up to 4000 IP routed interfaces.

When you configure a large number of IP routed interfaces, use the following guidelines:

- Use passive interfaces on most of the configured interfaces. You can make very few interfaces active.
- When you use Protocol Independent Multicast (PIM), configure a maximum of 10 PIM active interfaces. The remainder can be passive interfaces. Avaya recommends that you use IP routing policies with one or two unicast IP active interfaces.

IPv6

IPv6 provides high-performance, scalable Internet communications. Use the information in this section to help deploy IPv6 in your network. For configuration information, see *Avaya Virtual Services Platform 9000 Configuration — IPv6 Routing*, NN46250–509.

Design recommendations

Avaya Layer 2 and Layer 3 Ethernet switches support protocol-based IPv6 VLANs. To simplify network configuration with IPv6, Avaya recommends that you use protocol-based IPv6 VLANs from edge Layer 2 switches if you want to segregate all IPv6 traffic into a single VLAN at one switch or across a set of Layer 2 access switches. The core switch performs hardware-based IPv6 line-rate routing.

Transition mechanisms

Virtual Services Platform 9000 uses the following functions to help you transition your network from IPv4 to IPv6:

- Dual stack mechanism, where the IPv4 and IPv6 protocol stacks can communicate with both IPv6 and IPv4 devices
- Tunneling, which involves the encapsulation of IPv6 packets to traverse IPv4 networks, and the decapsulation of IPv4 packets to traverse IPv6 networks

Tunneling

Virtual Services Platform 9000 supports manually configured IPv6-in-IPv4 tunnels per RFC4213.

A manually-configured tunnel is equivalent to a permanent link between two IPv6 domains over an IPv4 backbone.

Use tunnels to provide stable, secure communications between two edge routers, an end system and an edge router, or to provide a connection to remote IPv6 networks.

Configure an IPv6 address on the tunnel interface.

Configure an IPv4 address on the tunnel source and destination.

Note:

The host or router at each end of the tunnel must be a dual-stack device.

The following figure demonstrates an IPv6 over IPv4 tunnel.

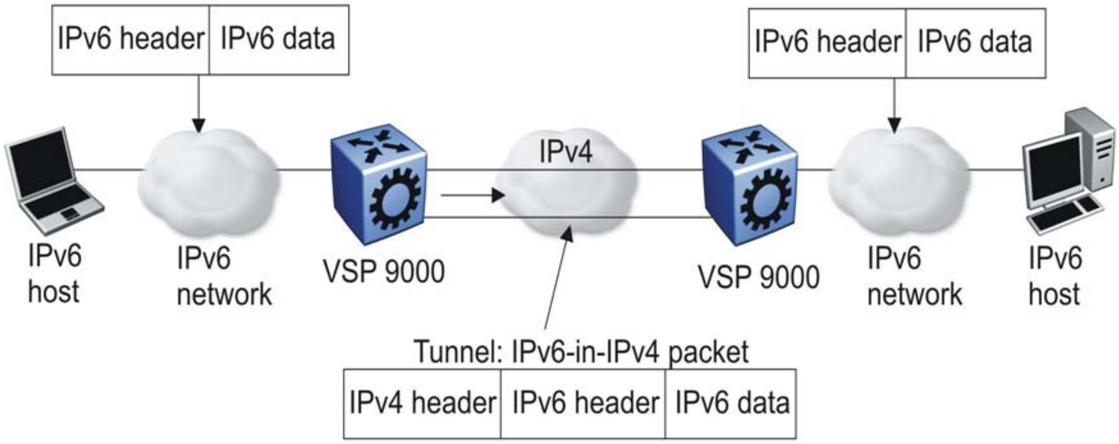

Figure 58: IPv6 over IPv4 tunneling

The tunnel in the preceding figure exists only between two VSP 9000 devices.

If you add new devices, you must configure additional tunnels.

VRRP

For IPv6 hosts on a LAN to learn about one or more default routers, IPv6-enabled routers send router advertisements using the IPv6 Neighbor Discovery (ND) protocol. The routers multicast these router advertisements every few minutes.

The ND protocol includes a mechanism called Neighbor Unreachability Detection to detect the failure of a neighbor node (router or host) or the failure of the forwarding path to a neighbor. Nodes can monitor the health of a forwarding path by sending unicast ND neighbor solicitation messages to the neighbor node. To reduce traffic, nodes only send neighbor solicitations to neighbors to which they are actively sending traffic, and only after the node receives no positive indication that the neighbors are up for a period of time. Using the default ND parameters, it takes a host approximately 38 seconds to learn that a router is unreachable before it switches to another default router. This delay is very noticeable and causes some transport protocol implementations to timeout.

While you can decrease the ND unreachability detection period by modifying the ND parameters, the lowest limit that can be achieved is 5 seconds, with the added downside of significantly increasing ND traffic, especially when many hosts try to determine the reachability of one of more routers.

To provide fast failover of a default router for IPv6 LAN hosts, Virtual Services Platform 9000 supports the Virtual Router Redundancy Protocol (VRRP) for IPv6

VRRP for IPv6 provides a faster switchover to an alternate default router than is possible using the ND protocol. With VRRP for IPv6, a backup router can take over for a failed default router in approximately 3 seconds (using default parameters). This failover is accomplished without host interaction and with a minimum amount of VRRP traffic.

IPv6 VRRP operation is similar to the IPv4 VRRP operation, including support for the holddown timer, critical IP, fast advertisements, and backup master. With backup master enabled, the backup switch routes all traffic according to its routing table. The backup master switch does not perform Layer 2 switching for the traffic to the VRRP master.

New to the IPv6 implementation of VRRP from the IPv4 implementation, you must specify a link-local address to associate with the virtual router. Optionally, you can also assign global unicast IPv6 addresses to associate with the virtual router. Network prefixes for the virtual router are derived from the global IPv6 addresses assigned to the virtual router.

VRRP backup master with triangular SMLT:

The standard implementation of VRRP supports one active master switch for each IPv6 subnet. All other VRRP interfaces in a network are in backup mode.

A deficiency occurs when VRRP-enabled switches use SMLT. If VRRP switches are aggregated into two SMLT switches, the end host traffic is load-shared on all uplinks to the aggregation switches (based on the Multilink Trunk [MLT] traffic distribution algorithm).

However, VRRP usually has only one active routing interface enabled. All other VRRP routers are in backup mode. Therefore, all traffic that reaches the backup VRRP router is forwarded over the interswitch trunk (IST) link toward the master VRRP router. In this case, the IST link potentially does not have enough bandwidth to carry all the aggregated traffic.

To resolve this issue, assign the backup router as the backup master router. The backup master router can actively load-share the routing traffic with a master router.

Because the two VRRP peer nodes exchange MAC address tables, the VRRP backup master can forward traffic directly, on behalf of the master router. The switch in the backup master state routes all traffic received on the backup master IP interface according to its routing table. The backup master switch does not perform Layer 2 switching for the traffic to the VRRP master.

If you enable SMLT on the backup master router, the incoming host traffic is forwarded over the SMLT links as usual.

The backup master feature only supports the triangular SMLT topology.

Important:

Do not use VRRP backup master and critical IP at the same time. Use one or the other. Use VRRP backup master only in SMLT configurations.

IPv6 VRRP and ICMP redirects:

In IPv6 networks, do not enable ICMP redirects on VRRP VLANs. If you enable this option (using the **`ipv6 icmp redirect-msg`** command), VRRP cannot function. The option is disabled by default.

Chapter 15: IP multicast network design

Use multicast routing protocols to efficiently distribute a single data source among multiple users in the network. This section provides information about how to design networks that support IP multicast routing.

For more information about multicast routing, see *Avaya Virtual Services Platform 9000 Configuration — IP Multicast Routing Protocols*, NN46250-504.

Multicast and MultiLink Trunking considerations

Multicast traffic distribution is important because the bandwidth requirements can be substantial when a large number of streams are employed. Virtual Services Platform 9000 can distribute IP multicast streams over links of a multilink trunk using one of the methods in this section.

PIM route tuning to load share streams

You can use Protocol Independent Multicast (PIM) routing to distribute multicast traffic. With this method, you must distribute sources of multicast traffic on different IP subnets and configure routing metrics so that traffic from different sources flows on different paths to the destination groups.

Multicast flow distribution over MLT

MultiLink Trunking (MLT) distributes multicast streams over a multilink trunk based on the source-subnet and group addresses of the packets. As a result, the load is distributed on different ports of the MLT more evenly. You cannot configure this feature.

To determine the egress port for a particular multicast stream, the hash calculator can be used in ACLI config mode. The command is `hash-calc getmltindex traffic-type` and enter desired parameters.

Multicast scalability design rules

To increase multicast route scaling, follow these eight design rules:

1. Whenever possible, use simple network designs that do not use VLANs that span several switches. Instead, use routed links to connect switches.
2. Whenever possible, group sources sending to the same group in the same subnet. Virtual Services Platform 9000 uses a single egress forwarding pointer for all

sources in the same subnet sending to the same group. Be aware that these streams have separate hardware forwarding records on the ingress side.

3. Do not configure multicast routing on edge switch interfaces that do not contain multicast senders or receivers. By following this rule, you:

    - Provide secure control over multicast traffic that enters or exits the interface.
    - Reduce the load on the switch, as well as the number of routes. This improves overall performance and scalability.

4. Avoid initializing many (several hundred) multicast streams simultaneously. Initial stream setup is a resource-intensive task, and initializing a large number can increase the setup time. In some cases, this delay can result in stream loss.

5. Whenever possible, do not connect IP multicast sources and receivers by using VLANs that interconnect switches (see the following figure). In some cases, this can result in excessive hardware record use. By placing the source on the interconnected VLAN, traffic takes two paths to the destination, depending on the RPF checks and the shortest path to the source.

    For example, if a receiver is on VLAN 1 on switch S1 and another receiver is on VLAN 2 on switch S1, traffic can be received from two different paths to the two receivers, which results in the use of two forwarding records. If the source on switch S2 is on a different VLAN than VLAN 3, traffic takes a single path to switch S1 where the receivers are located.

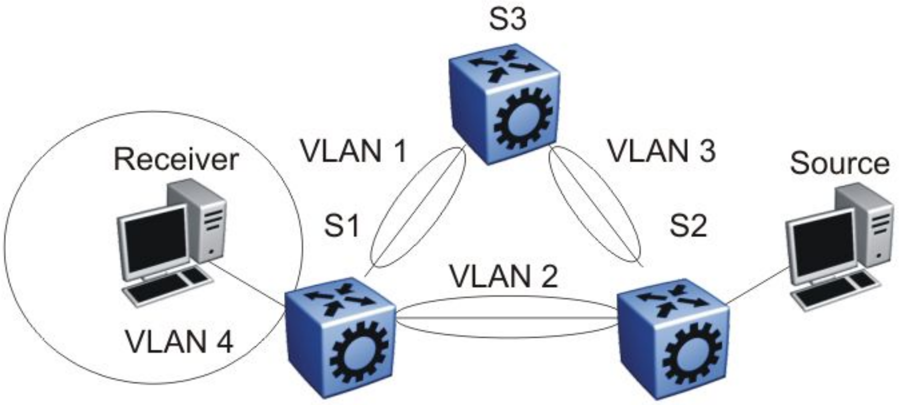

Figure 59: IP multicast sources and receivers on interconnected VLANs

6. Use default timer values for PIM. After you decrease timers for faster convergence, they usually adversely affect scalability because control messages are sent more frequently. If you need faster network convergence, configure the timers with the same values on all switches in the network. Also, in most cases, you must perform baseline testing to achieve optimal values for timers versus required convergence times and scalability.

7. For faster convergence, configure the bootstrap and rendezvous point (RP) routers on a circuitless IP (CLIP). See Circuitless IP for PIM-SM on page 134.
8. Avaya recommends the use of Static group-range-to-RP mappings in an SMLT topology as opposed to RP set learning via the Bootstrap Router (BSR) mechanism. Static RP allows for faster convergence in box failure, reset, and HA failover scenarios; whereas there are inherent delays in the BSR mechanism as follows:
    - When a router comes back up after a failover or reset, to accept and propagate (*,g) Join requests from surrounding routers (either PIM Join messages or local IGMP membership reports) to the RP, a PIM router must determine the address of the RP for each group for which they desire (*,g) state. The PIM router needs to know the unicast route to the RP address. The route to the RP address is learned by using a unicast routing protocol such as OSPF, and the RP address is either statically configured or dynamically learned using the BSR mechanism.
    - When a system comes up after a reset or the standby CP becomes master after an HA failover, if the RP is not statically configured, it must wait for the BSR to select the RP from candidate RP routers, and then propagate the RP set hop-by-hop to all PIM routers. This must be done before a Join message can be processed. If the PIM router receives a Join message before it learns the RP set, the router drops the Join message and waits for another Join or Prune message to arrive before it creates the multicast router, and propagates the Join messages to the RP. The default Join/Prune timer is 60 seconds, and because of this and the delays inherent in BSR RP-set learning, significant multicast traffic interruptions can occur. If the RP is statically configured, the only delay is in the unicast routing table convergence and the arrival of the Join/Prune messages from surrounding boxes.

IP multicast address range restrictions

IP multicast routers use D class addresses, which range from 224.0.0.0 to 239.255.255.255. Although you can use subnet masks to configure IP multicast address ranges, the concept of subnets does not exist for multicast group addresses. Consequently, the usual unicast conventions—where you reserve the all 0s subnets, all 1s subnets, all 0s host addresses, and all 1s host addresses—do not apply.

Internet Assigned Numbers Authority (IANA) reserves addresses from 224.0.0.0 through 224.0.0.255 for link-local network applications. Multicast-capable routers do not forward packets with an address in this range. For example, Open Shortest Path First (OSPF) uses 224.0.0.5 and 224.0.0.6, and Virtual Router Redundancy Protocol (VRRP) uses 224.0.0.18 to communicate across local broadcast network segments.

IANA also reserves the range of 224.0.1.0 through 224.0.1.255 for well-known applications. IANA assigns these addresses to specific network applications. For example, the Network

Time Protocol (NTP) uses 224.0.1.1, and Mtrace uses 224.0.1.32. RFC1700 contains a complete list of these reserved addresses.

Multicast addresses in the 232.0.0.0/8 (232.0.0.0 to 232.255.255.255) range are reserved only for source-specific multicast (SSM) applications, such as one-to-many applications. While this range is the publicly reserved range for SSM applications, private networks can use other address ranges for SSM.

Finally, addresses in the range 239.0.0.0/8 (239.0.0.0 to 239.255.255.255) are administratively scoped addresses; they are reserved for use in private domains. Do not advertise these addresses outside the private domain. This multicast range is analogous to the 10.0.0.0/8, 172.16.0.0/20, and 192.168.0.0/16 private address ranges in the unicast IP space.

In a private network, only assign multicast addresses from 224.0.2.0 through 238.255.255.255 to applications that are publicly accessible on the Internet. Assign addresses in the 239.0.0.0/8 range to multicast applications that are not publicly accessible.

Although you can use a multicast address you choose on your own private network, it is generally not good design practice to allocate public addresses to private network entities. Do not use public addresses for unicast host or multicast group addresses on private networks.

Multicast MAC address mapping considerations

Like IP, Ethernet has a range of multicast MAC addresses that natively support Layer 2 multicast capabilities. While IP has a total of 28 addressing bits available for multicast addresses, Ethernet has only 23 addressing bits assigned to IP multicast. The Ethernet multicast MAC address space is much larger than 23 bits, but only a subrange of that larger space is allocated to IP multicast. Because of this difference, 32 IP multicast addresses map to one Ethernet multicast MAC address.

IP multicast addresses map to Ethernet multicast MAC addresses by placing the low-order 23 bits of the IP address into the low-order 23 bits of the Ethernet multicast address 01:00:5E: 00:00:00. Thus, more than one multicast address maps to the same Ethernet address (see the following figure). For example, all 32 addresses 224.1.1.1, 224.129.1.1, 225.1.1.1, 225.129.1.1, 239.1.1.1, 239.129.1.1 map to the same 01:00:5E:01:01:01 multicast MAC address.

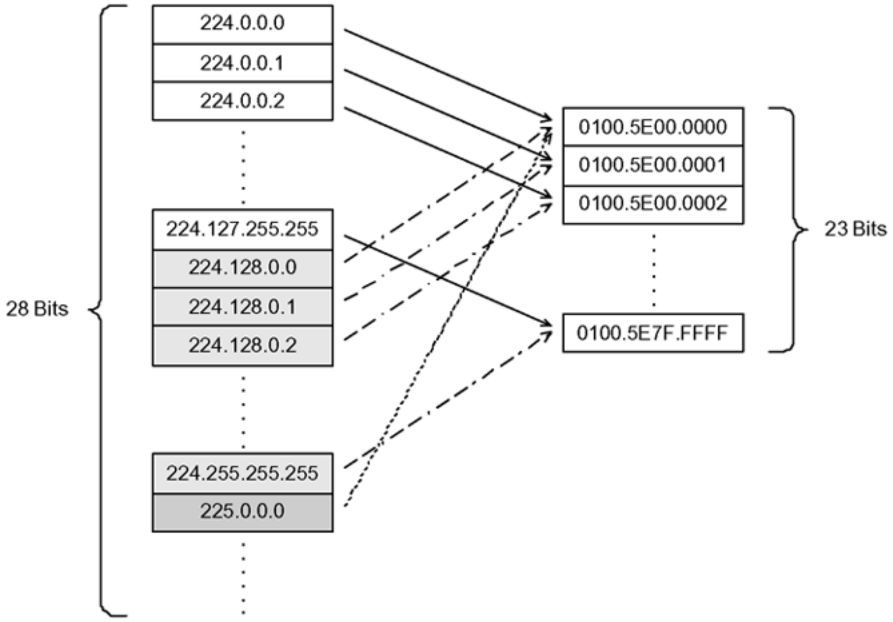

Figure 60: Multicast IP address to MAC address mapping

Most Ethernet switches handle Ethernet multicast by mapping a multicast MAC address to multiple switch ports in the MAC address table. Therefore, when you design the group addresses for multicast applications, take care to efficiently distribute streams only to hosts that are receivers. Virtual Services Platform 9000 switches IP multicast data based on the IP multicast address, not the MAC address, and thus, does not have this issue.

As an example, consider two active multicast streams using addresses 239.1.1.1 and 239.129.1.1. Suppose that two Ethernet hosts, receiver A and receiver B, connect to ports on the same switch and only want the stream addressed to 239.1.1.1. Suppose also that two other Ethernet hosts, receiver C and receiver D, also connect to the ports on the same switch as receiver A and B, and want to receive the stream addressed to 239.129.1.1. If the switch uses the Ethernet multicast MAC address to make forwarding decisions, then all four receivers receive both streams—even though each host only wants one stream. This transmission increases the load on both the hosts and the switch. To avoid this extra load, Avaya recommends that you manage the IP multicast group addresses used on the network.

Virtual Services Platform 9000 does not forward IP multicast packets based on multicast MAC addresses—even when bridging VLANs at Layer 2. Thus, the platform does not encounter this problem. Instead, the platform internally maps IP multicast group addresses to the ports that contain group members.

When an IP multicast packet is received, the lookup is based on the IP group address, regardless of whether the VLAN is bridged or routed. While Virtual Services Platform 9000 does not suffer from the problem described in the previous example, other switches in the network can. This problem is particularly true of pure Layer 2 switches.

In a network that includes non Virtual Services Platform 9000 equipment, the easiest way to ensure that this issue does not arise is to use only a consecutive range of IP multicast addresses that correspond to the lower order 23 bits of that range. For example, use an address range from 239.0.0.0 through 239.127.255.255. A group address range of this size can still easily accommodate the needs of even the largest private enterprise.

Dynamic multicast configuration changes

Avaya recommends that you do not perform dynamic multicast configuration changes when multicast streams flow in a network. For example, do not change the routing protocol that runs on an interface, or the IP address, or the subnet mask for an interface until multicast traffic ceases.

For such changes, Avaya recommends that you temporarily stop all multicast traffic. If the changes are necessary and you have no control over the applications that send multicast data, you can disable the multicast routing protocols before you perform the change. For example, consider disabling multicast routing before making interface address changes. In all cases, these changes result in traffic interruptions because they impact neighbor state machines and stream state machines.

In addition, Avaya recommends that when removing port members of an MLT group that you first disable the ports before removing them from that MLT group. Changing the group set without first shutting the ports down can result in high CPU utilization and processing in a scaled multicast environment due to the necessary hardware reprogramming on the multicast records.

IGMPv3 backward compatibility

IGMPv3 for PIM is backward compatible with IGMPv1/v2. According to RFC3376, the multicast router with IGMPv3 can use one of two methods to handle older query messages:

- If an older version of IGMP is present on the router, the querier must use the lowest version of IGMP present on the network.
- If a router that is not explicitly configured to use IGMPv1 or IGMPv2, hears an IGMPv1 query or IGMPv2 general query, it logs a rate-limited warning.

You can configure the IGMP version of an interface to version 3 regardless of the PIM or snooping mode.

You can configure whether the switch downgrades the version of IGMP to handle older query messages. If the switch downgrades, the host with IGMPv3 only capability does not work. If you do not configure the switch to downgrade the version of IGMP, the switch logs a warning.

Note:

If you enable the explicit host tracking option on an IGMPv3 interface, you cannot downgrade to IGMPv1 or IGMPv2. You must disable explicit host tracking to downgrade the IGMP version.

TTL in IP multicast packets

Virtual Services Platform 9000 treats multicast data packets with a time-to-live (TTL) of 1 as expired packets and sends them to the CPU before dropping them. To avoid this, ensure that the originating application uses a hop count large enough to enable the multicast stream to traverse the network and reach all destinations without reaching a TTL of 1. Avaya recommends that you use a TTL value of 33 or 34 to minimize the effect of looping in an unstable network.

Multicast MAC filtering

Certain network applications, such as the Microsoft Network Load Balancing solution, require multiple hosts to share a multicast MAC address. Instead of flooding all ports in the VLAN with this multicast traffic, you can use Multicast MAC filtering to forward traffic to a configured subset of the ports in the VLAN. This multicast MAC address is not an IP multicast MAC address.

At a minimum, map the multicast MAC address to a set of ports within the VLAN. In addition, if traffic is routed on the local Virtual Services Platform 9000, you must configure an Address Resolution Protocol (ARP) entry to map the shared unicast IP address to the shared multicast MAC address. You must configure an ARP entry because the hosts can also share a virtual IP address, and packets addressed to the virtual IP address need to reach each host.

Avaya recommends that you limit the number of such configured multicast MAC addresses to a maximum of 100. This number is related to the maximum number of possible VLANs you can configure because for every multicast MAC filter that you configure the maximum number of configurable VLANs reduces by one. Similarly, configuring large numbers of VLANs reduces the maximum number of configurable multicast MAC filters downwards from 100.

Although you can configure addresses starting with 01.00.5E, which are reserved for IP multicast address mapping, do not enable IP multicast with streams that match the configured addresses. This can result in incorrect IP multicast forwarding and incorrect multicast MAC filtering.

Guidelines for multicast access policies

Use the following guidelines when you configure multicast access policies:

- Use masks to specify a range of hosts. For example, 10.177.10.8 with a mask of 255.255.255.248 matches hosts addresses 10.177.10.8 through 10.177.10.15. The host subnet address and the host mask must be equal to the host subnet address. An easy way to determine this is to ensure that the mask has an equal or fewer number of trailing zeros than the host subnet address. For example, 3.3.0.0/255.255.0.0 and 3.3.0.0/255.255.255.0 are valid. However, 3.3.0.0/255.0.0.0 is not.
- Apply receive access policies to all eligible receivers on a segment. Otherwise, one host joining a group makes that multicast stream available to all.
- Receive access policies are initiated after the switch receives reports with addresses that match the filter criteria.
- Transmit access policies apply after the switch receives the first packet of a multicast stream.

Multicast access policies can apply to a routed PIM interface if Internet Group Management Protocol (IGMP) reports the reception of multicast traffic.

The following rules and limitations apply to IGMP access policy parameters when you use them with IGMP instead of PIM:

- The static member parameter applies to IGMP snooping and PIM on both interconnected links and edge ports.
- The Static Not Allowed to Join parameter applies to IGMP snooping and PIM on both interconnected links and edge ports.
- For multicast access control, the denyRx parameter applies to IGMP snooping and PIM. The DenyTx and DenyBoth parameters apply only to IGMP snooping.

Split-subnet and multicast

The split-subnet issue arises when you divide a subnet into two unconnected sections in a network. This division results in the production of erroneous routing information about how to reach the hosts on that subnet. The split-subnet problem applies to all types of traffic, but it has a larger impact on a PIM-SM network.

To avoid the split-subnet problem in PIM networks, ensure that the RP router is not in a subnet that can become a split subnet. Also, avoid having receivers on this subnet. Because the RP is an entity that must be reached by all PIM-enabled switches with receivers in a network, placing the RP on a split-subnet can impact the whole multicast traffic flow. Traffic can be affected even for receivers and senders that are not part of the split-subnet.

Protocol Independent Multicast-Sparse Mode guidelines

Protocol Independent Multicast-Sparse Mode (PIM-SM) uses an underlying unicast routing information base to perform multicast routing. PIM-SM builds unidirectional shared trees rooted at a RP router for each group and can also create shortest-path trees for each source.

PIM-SM and PIM-SSM scalability

VSP 9000 supports up to 4084 PIM interfaces, 512 active and the rest passive.

Interfaces that run PIM must also use a unicast routing protocol (PIM uses the unicast routing table), which puts stringent requirements on the system. As a result, 1500 interfaces cannot be supported in some scenarios, especially if the number of routes and neighbors is high. With a high number of interfaces, take special care to reduce the load on the system.

Use few active IP routed interfaces. You can use IP forwarding without a routing protocol enabled on the interfaces, and enable only one or two with a routing protocol. You can configure proper routing by using IP routing policies to announce and accept routes on the switch. Use PIM passive interfaces on the majority of interfaces.

Important:

Avaya Virtual Services Platform 9000 supports 4084 PIM interfaces. You can configure 512 active interfaces and the remainder must be passive.

When you use PIM-SM, the number of routes can scale up to the unicast route limit because PIM uses the unicast routing table to make forwarding decisions. For higher route scaling, Avaya recommends that you use OSPF rather than Routing Information Protocol (RIP).

As a general rule, a well-designed network does not have many routes in the routing table. For PIM to work properly, ensure that all subnets configured with PIM are reachable and that PIM uses the information in the unicast routing table. For the RPF check, to correctly reach the source of any multicast traffic, PIM requires the unicast routing table. For more information, see PIM network with non-PIM interfaces on page 142.

PIM general requirements

Avaya recommends that you design simple PIM networks where VLANs do not span several switches.

PIM relies on unicast routing protocols to perform its multicast forwarding. As a result, include in your PIM network design, a unicast design where the unicast routing table has a route to every source and receiver of multicast traffic, as well as a route to the RP router and Bootstrap router (BSR) in the network. Ensure that the path between a sender and receiver contains PIM-enabled interfaces. Receiver subnets are not always required in the routing table.

Avaya recommends that you follow these guidelines:

- Ensure that every PIM-SM domain is configured with an RP, either by static definition or via BSR.
- Ensure that every group address used in multicast applications has an RP in the network.
- As a redundancy option, you can configure several RPs for the same group in a PIM domain.
- As a load sharing option, you can have several RPs in a PIM-SM domain map to different groups.
- In order to configure an RP to cover the entire multicast range, configure an RP to use the IP address of 224.0.0.0 and the mask of 240.0.0.0.
- Configure an RP to handle a range of multicast groups by using the mask parameter. For example, an entry for group value of 224.1.1.0 with a mask of 255.255.255.192 covers groups 224.1.1.0 to 224.1.1.63.
- In a PIM domain with both static and dynamic RP switches, you cannot configure one of the (local) interfaces for the static RP switches as the RP. For example, in the following scenario:

  (static RP switch) Sw1 ------ Sw2 (BSR/Cand-RP1) -----Sw3

  You cannot configure one of the interfaces on switch Sw1 as static RP because the BSR cannot learn this information and propagate it to Sw2 and Sw3. PIM requires that you consistently configure RP on all the routers of the PIM domain, so you can only add the remote interface Candidate-RP1 (Cand-RP) to the static RP table on Sw1.
- If a switch needs to learn an RP-set, and has a unicast route to reach the BSR through this switch, you cannot enable or configure static RP on a switch in a mixed mode of candidate RP and static RP switches. For examples, see the following two figures.

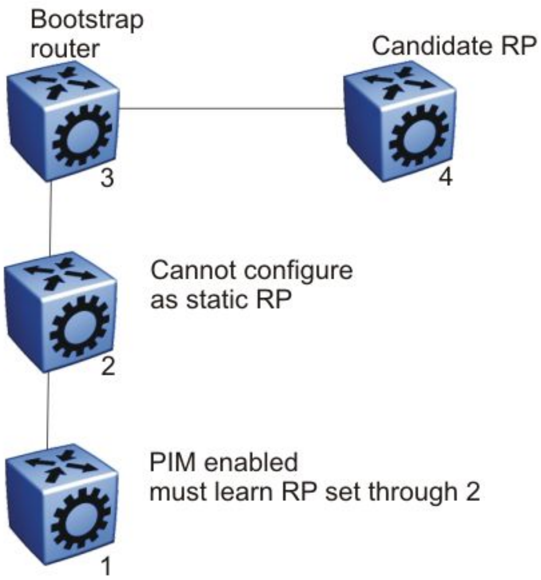
Bootstrap router
Candidate RP
3
4
Cannot configure as static RP
2
PIM enabled must learn RP set through 2
1

Figure 61: Example 1

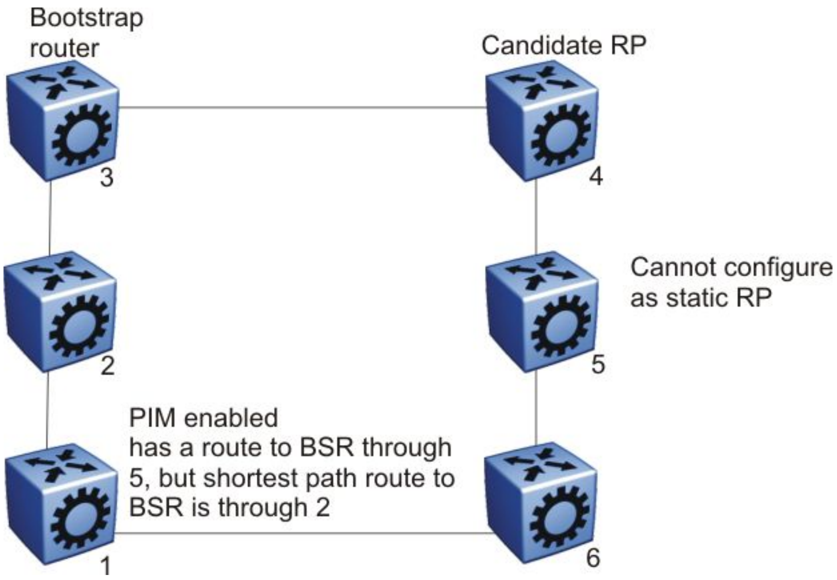
Bootstrap router
Candidate RP
3
4
Cannot configure as static RP
2
5
PIM enabled has a route to BSR through 5, but shortest path route to BSR is through 2
1
6

Figure 62: Example 2

PIM and shortest path tree switchover

When an IGMP receiver joins a multicast group, PIM on the leaf router first joins the shared tree. After the first packet is received on the shared tree, the router uses the source address information in the packet to immediately switch over to the shortest path tree (SPT).

To guarantee a simple, yet high-performance implementation of PIM-SM, the switch does not support a threshold bit rate in relation to SPT switchover. Intermediate routers (that is, not directly connected IGMP hosts) do not switch over to the SPT until directed to do so by the leaf routers.

Other vendors can offer a configurable threshold, such as a certain bit rate at which the SPT switch-over occurs. Regardless of their implementation, no interoperability issues with Virtual Services Platform 9000 result. Switching to and from the shared and shortest path trees is independently controlled by each downstream router. Upstream routers relay joins and prunes upstream hop-by-hop, building the desired tree as they go. Because a PIM-SM compatible router already supports shared and shortest path trees, no compatibility issues arise from the implementation of configurable switchover thresholds.

PIM traffic delay and SMLT peer reboot

PIM uses a designated router (DR) to forward data to receivers on the DR VLAN. The DR is the router with the highest IP address on a LAN. If this router is down, the router with the next highest IP address becomes the DR. However, if the VLAN is an SMLT VLAN, the DR is not a factor in determining which switch forwards the data down to the receiver. Either aggregate switch can forward data to the receiver, because the switches act as one. The switch that forwards depends on where the source is located (on another SMLT/IST link or on a non-SMLT/ non-IST link) and whether either side of the receiver SMLT link is up or down. If the forwarder switch is rebooted, traffic loss occurs until protocol convergence is completed.

Consider the following cases:

- If the source is on an SMLT link that is not the receiver SMLT, the switch that directly received the data on its side of the source SMLT link forwards it down to the receiver on the receiver SMLT regardless of which switch is the DR for the receiver VLAN. The forwarding switch sends a copy of the data over the IST link to the peer switch, which drops the data because it knows that the remote SMLT is up and therefore the remote peer has already forwarded the data. If the forwarding switch goes down, the other switch receives the data directly over its source SMLT link and takes over forwarding to the receivers. After the original switch comes back up, the original switch again receives the data directly over its source SMLT. The original switch may not be ready to forward the data because of the protocol reconvergence, so the original switch loses traffic until reconvergence is complete.
- If the source is not learned on another SMLT link or the IST link on each aggregate switch; they have a route to the source which is not on an SMLT or across the IST. The switches must choose which one forwards the data down the receiver SMLT link; which one is the designated forwarder, so that duplicate data does not occur. The highest IP address is the designated forwarder. If the designated forwarder becomes disabled, the other takes over. When it is reenabled, the other switch sees that it is no longer the highest IP address and it sees that the remote SMLT link comes up. The other switch then assumes that the IST peer is capable of being the designated forwarder and it stops forwarding down to the receivers. If the original switch is not ready to forward the data due to reconvergence, traffic loss occurs.

In either case, configuring a static RP helps the situation. To avoid this traffic delay, a workaround is to configure a static RP on the peer SMLT switches. This configuration avoids the process of selecting an active RP router from the list of candidate RPs, and also of dynamically learning about RPs through the BSR mechanism. Then, when the DR comes back, traffic resumes as soon as OSPF converges. This workaround reduces the traffic delay.

Circuitless IP for PIM-SM

Use CLIP to configure a resilient RP and BSR for a PIM network. When you configure an RP or BSR on a regular interface, if it becomes nonoperational, the RP and BSR also become nonoperational. This status results in the election of other redundant RPs and BSRs, and can disrupt IP multicast traffic flow in the network. As a best practice for multicast networks design, always configure the RP and BSR on a CLIP interface to prevent a single interface failure from causing these entities to fail.

Avaya also recommends that you configure redundant RPs and BSRs on different switches and that these entities be on CLIP interfaces. For the successful setup of multicast streams, ensure that a unicast route exists to all CLIP interfaces from all locations in the network. A unicast route is mandatory because, for proper RP learning and stream setup on the shared RP tree, every switch in the network needs to reach the RP and BSR. You can use PIM-SM CLIP interfaces only for RP and BSR configurations, and are not intended for other purposes.

It is not recommended to have non-SMLT IGMP leaf ports on a VSP router configured to be one of the redundant RP CLIP devices. It is possible that these IGMP hosts can become isolated from the multicast data stream(s).

If you configure dual-redundant RPs (IST peers with the same CLIP interface IP address used for the RP), the topology in Figure 30: Multicast SMLT triangle on page 79 does not work in link-failure scenarios. Use caution if you design a network with this topology where the IST peers are PIM enabled, and the source and receiver edges are Layer 2.

Consider an example where one of the peers, IST-A, is the PIM DR for the source VLAN, and the source data is hashed to IST-A from the Layer 2 source edge. IST-A forwards traffic to the receiver edge using the SMLT link from IST-A to the receiver edge. If the SMLT link fails, IST-A does not forward traffic over the IST link to IST-B, and the receiver edge does receive the data.

In this topology, the receiver edge sends an IGMP membership report for a group, which is recorded on both IST peers as an IGMP LEAF on the receiver SMLT port on the receiver VLAN.

Because both of the IST peers are the RP for the group, they do not send a (*,g) PIM JOIN message toward the other RP. The (*,g) PIM mroute does not record the IST port as a JOIN port on either IST device. The PIM (*,g) mroute records only a LEAF on the SMLT receiver port.

Because the source is local (Layer 2 edge), there is no PIM (s,g) JOIN message toward the source and the (s,g) PIM mroute does not record the IST port as a JOIN port on either IST device. The PIM (s,g) mroute records only a LEAF on the SMLT receiver port.

If the source is hashed to IST-A, the PIM DR for the incoming VLAN, traffic is forwarded to the receiver correctly. IST-A does not forward traffic over the IST to IST-B, because no JOIN exists on the IST port. If the receiver SMLT link from the IST-A peer is down, the traffic is not forwarded to IST-B, and is not received by the receiver edge. Traffic resumes after the link is restored. If the source data hashes to the non-DR peer, IST-B, no problem occurs because the non-DR always forwards traffic to the DR.

A similar situation exists in this topology when IST-A is both the RP and the DR for the Layer 2 receiver edge. The IST port is not in the outgoing port list because there is no JOIN message from the peer toward the source (which is not PIM enabled). Therefore, if the SMLT link from IST-A to the receiver edge is down, the system does not forward traffic to the peer IST-B and down to the receiver.

You can avoid the preceding problems with this topology by performing one of the following actions:

- Enable PIM on the source edge.

  The IST peers send PIM joins toward the source and the JOIN is recorded on the IST port for the (s,g). Data is forwarded to the peer.
- Do not configure dual redundant RPs.

  One IST peer is the RP for a group.
- Do not configure one IST peer as both the DR for the source VLAN and the RP for the receiver group.

  The system forwards the traffic to the RP or to the DR, depending on which peer receives the source, and, if the SMLT link to the receiver goes down there will be no data loss.

PIM-SM and static RP

Use static RP to provide security, interoperability, and redundancy for PIM-SM multicast networks. Consider if the administrative ease derived from using dynamic RP assignment is worth the security risks involved. For example, if an unauthorized user connects a PIM-SM router that advertises itself as a candidate RP (C-RP), it can possibly take over new multicast streams that otherwise distribute through an authorized RP. If security is important, use static RP assignment.

You can use the static RP feature in a PIM environment with devices that run legacy PIM-SMv1 and auto-RP (a proprietary protocol that Virtual Services Platform 9000 does not support). For faster convergence, you can also use static RP in a PIM-SMv2 environment. If you configure static RP with PIM-SMv2, the BSR is not active.

Static RP and auto-RP

Some legacy PIM-SMv1 networks use the auto-RP protocol. Auto-RP is a Cisco proprietary protocol that provides equivalent functionality to the standard Virtual Services Platform 9000 PIM-SM RP and BSR. You can use the static RP feature to interoperate in this environment. For example, in a mixed-vendor network, you can use auto-RP among routers that support the protocol, while other routers use static RP. In such a network, ensure that the static RP configuration mimics the information that is dynamically distributed to guarantee that multicast traffic is delivered to all parts of the network.

In a mixed auto-RP and static RP network, ensure that Virtual Services Platform 9000 does not serve as an RP because it does not support the auto-RP protocol. In this type of network, the RP must support the auto-RP protocol.

Static RP and RP redundancy

You can provide RP redundancy through static RPs. To ensure consistency of RP selection, implement the same static RP configuration on all PIM-SM routers in the network. In a mixed vendor network, ensure that the same RP selection criteria is used among all routers. For example, to select the active RP for each group address, the switch uses a hash algorithm defined in the PIM-SMv2 standard. If a router from another vendor selects the active RP based on the lowest IP address, then the inconsistency prevents stream delivery to certain routers in the network.

If a group address-to-RP discrepancy occurs among PIM-SM routers, network outages occur. Routers that are unaware of the true RP cannot join the shared tree and cannot receive the multicast stream.

Failure detection of the active RP is determined by the unicast routing table. As long as the RP is considered reachable from a unicast routing perspective, the local router assumes that the RP is fully functional and attempts to join the shared tree of that RP.

The following figure shows a hierarchical OSPF network where a receiver is in a totally stubby area. If RP B fails, PIM-SM router A does not switch over to RP C because the injected default route in the unicast routing table indicates that RP B is still reachable.

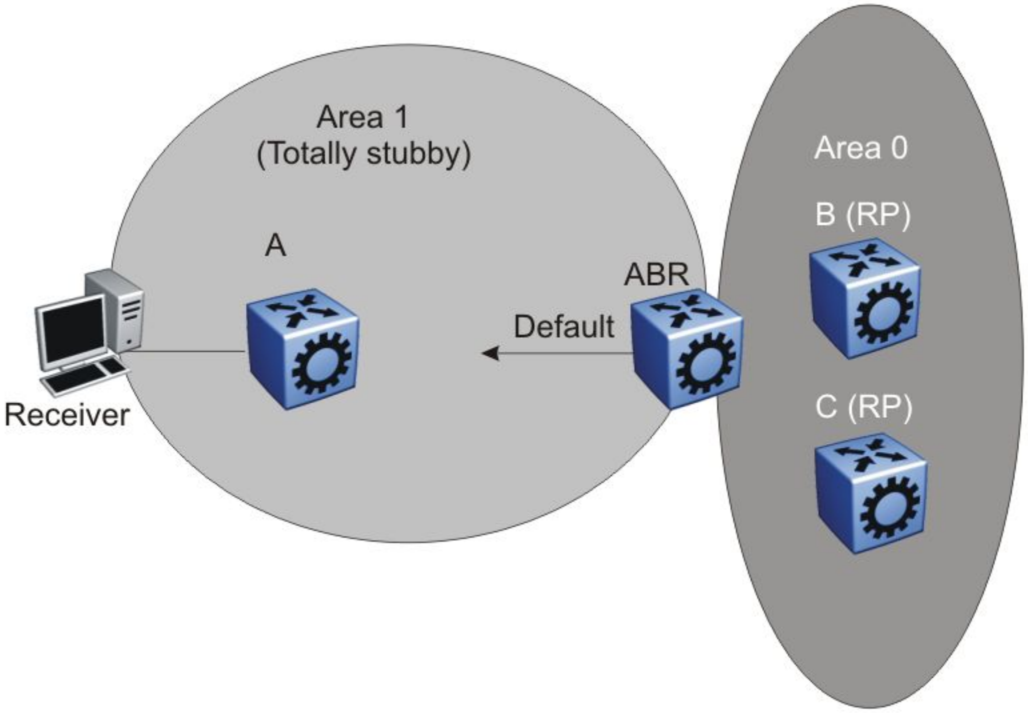

Figure 63: RP failover with default unicast routes

Because failover is determined by unicast routing behavior, carefully consider the unicast routing design, as well as the IP address you select for the RP. Static RP failover performance depends on the convergence time of the unicast routing protocol. For quick convergence, Avaya recommends that you use a link state protocol, such as OSPF. For example, if you use RIP as the routing protocol, an RP failure can take minutes to detect. Depending on the application, this situation can be unacceptable.

Static RP failover time does not affect routers that have already switched over to the SPT; failover time only affects newly-joining routers.

Unsupported static RP configurations

If you use static RP, you disable dynamic RP learning. The following figure shows an unsupported configuration for static RP. In this example because of interoperation between static RP and dynamic RP, no RP exists at switch 2. However, (S,G) creation and deletion occurs every 210 seconds at switch 16.

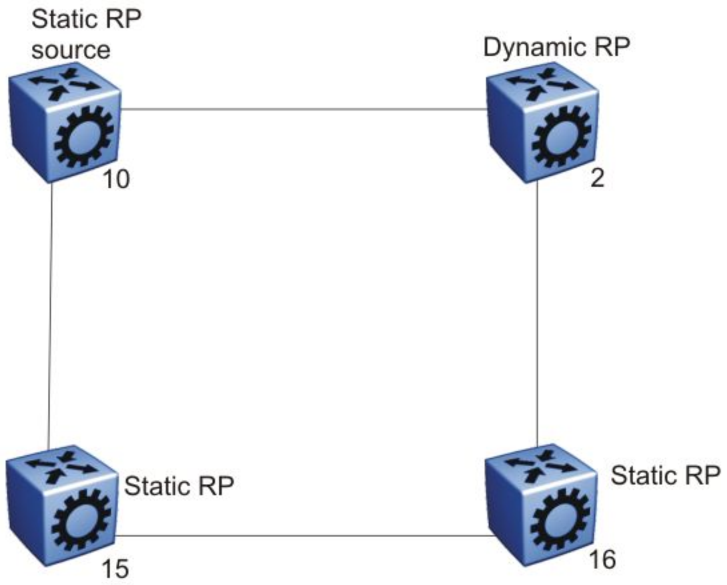

Figure 64: Unsupported static RP configuration

Switches 10, 15, and 16 use static RP, whereas switch 2 uses dynamic RP. The source is at switch 10, and the receivers are switches 15 and 16. The RP is at switch 15 locally. The receiver on switch 16 cannot receive packets because its SPT goes through switch 2.

Switch 2 is in a dynamic RP domain, so it cannot learn about the RP on switch 15. However, (S, G) records are created and deleted on switch 16 every 210 seconds.

Rendezvous point router considerations

You can place an RP on a switch when VLANs extend over several switches. However, when you use PIM-SM, Avaya recommends that you not span VLANs on more than two switches.

Avaya recommends the use of Static group-range-to-RP mappings in an SMLT topology as opposed to RP set learning via the Bootstrap Router (BSR) mechanism. Static RP allows for faster convergence in box failure, reset and HA failover scenarios, whereas there are inherent delays in the BSR mechanism as follows:

- When a router comes back up after a failover or reset, to accept and propagate (*,g) join requests from surrounding routers (either PIM join messages or local IGMP membership reports) to the RP, a PIM router must determine the address of the RP for each group for which they desire (*,g) state. The PIM router must know the unicast route to the RP address. The route to the RP address is learned by using a unicast routing protocol such as OSPF, and the RP address is either statically configured or dynamically learned using the BSR mechanism.
- When a box comes up after a reset or the standby CP becomes master after an HA failover, if the RP is not statically configured, it must wait for the BSR to select the RP from candidate RP routers, and then propagate the RP set hop-by-hop to all PIM routers. This must be done before a join message can be processed. If the PIM router receives a join message before it learns the RP set, it drops the join message, and the router waits

for another join or prune message to arrive before it creates the multicast route and propagates the join message to the RP. The default Join/Prune timer is 60 seconds, and because of this and the delays inherent in BSR RP-set learning, significant multicast traffic interruptions can occur. If the RP is statically configured, the only delay is in the unicast routing table convergence and the arrival of the Join/Prune messages from surrounding boxes.

PIM-SM design and the BSR hash algorithm

To optimize the flow of traffic down the shared trees in a network that uses a BSR to dynamically advertise candidate RPs, consider the hash function. The BSR uses the hash function to assign multicast group addresses to each C-RP.

The BSR distributes the hash mask used to compute the RP assignment. For example, if two RPs are candidates for the range 239.0.0.0 through 239.0.0.127, and the hash mask is 255.255.255.252, that range of addresses is divided into groups of four consecutive addresses and assigned to one or the other C-RP.

The following figure illustrates a suboptimal design where Router A sends traffic to a group address assigned to RP D. Router B sends traffic assigned to RP C. RP C and RP D serve as backups for each other for those group addresses. To distribute traffic, it is desirable that traffic from Router A use RP C and that traffic from Router B use RP D.

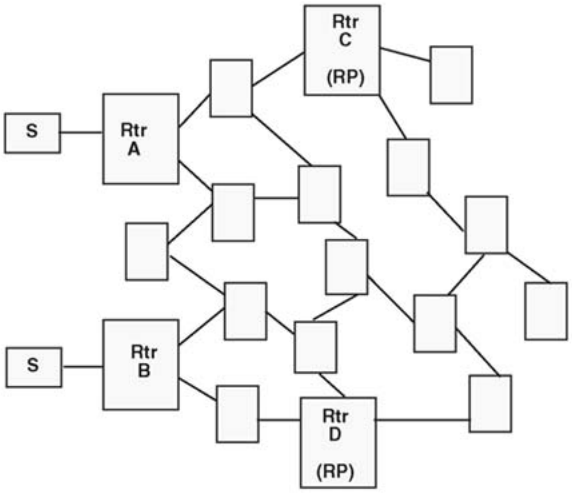

Figure 65: Example multicast network

While still providing redundancy in the case of an RP failure, you can ensure that the optimal shared tree is used by using the following methods.

1. Use the hash algorithm to proactively plan the group-address-to-RP assignment.

   Use this information to select the multicast group address for each multicast sender on the network and to ensure optimal traffic flows. This method is helpful for modeling more complex redundancy and failure scenarios, where each group address has three or more C-RPs.

2. Allow the hash algorithm to assign the blocks of addresses on the network, and then view the results using the command **`show ip pim active-rp`** .

Use the command output to assign multicast group addresses to senders that are located near the indicated RP. The limitation to this approach is that while you can easily determine the current RP for a group address, the backup RP is not shown. If more than one backup for a group address exists, the secondary RP is not obvious. In this case, use the hash algorithm to reveal which of the remaining C-RPs take over for a particular group address in the event of primary RP failure.

The hash algorithm works as follows:

1. For each C-RP router with matching group address ranges, a hash value is calculated according to the formula:

   Hash value [G, M, C(i)] = {1 103 515 245 * [(1 103 515245 * (G&M) +12 345) XOR C(i)] + 12 345} mod 2^31

   The hash value is a function of the group address (G), the hash mask (M), and the IP address of the C-RP C(i). The expression (G&M) guarantees that blocks of group addresses hash to the same value for each C-RP, and that the size of the block is determined by the hash mask.

   For example, if the hash mask is 255.255.255.248, the group addresses 239.0.0.0 through 239.0.0.7 yield the same hash value for a given C-RP. Thus, the block of eight addresses are assigned to the same RP.
2. The C-RP with the highest resulting hash value is chosen as the RP for the group. In the event of a tie, the C-RP with the highest IP address is chosen.

   This algorithm runs independently on all PIM-SM routers so that every router has a consistent view of the group-to-RP mappings.

Candidate RP considerations

The C-RP priority parameter determines an active RP for a group. The hash values for different RPs are only compared for RPs with the highest priority. Among the RPs with the highest priority value and the same hash value, the C-RP with the highest RP IP address is chosen as the active RP.

You cannot configure the C-RP priority. Each RP has a default C-RP priority value of 0, and the algorithm uses the RP if the group address maps to the grp-prefix that you configure for that RP. If a different router in the network has a C-RP priority value greater than 0, the switch uses this part of the algorithm in the RP election process.

Currently, you cannot configure the hash mask used in the hash algorithm. Unless you configure a different PIM BSR in the network with a nondefault hash mask value, the default hash mask of 255.255.255.252 is used. Static RP configurations do not use the BSR hash mask; they use the default hash mask.

For example:

RP1 = 128.10.0.54 and RP2 = 128.10.0.56. The group prefix for both RPs is 238.0.0.0/255.0.0.0. Hash mask = 255.255.255.252.

The hash function assigns the groups to RPs in the following manner:

The group range 238.1.1.40 to 238.1.1.51 (12 consecutive groups) maps to 128.10.0.56. The group range 238.1.1.52 to 238.1.1.55 (4 consecutive groups) maps to 128.10.0.54. The group range 238.1.1.56 to 238.1.1.63 (8 consecutive groups) maps to 128.10.0.56.

PIM-SM receivers and VLANs

Some designs cause unnecessary traffic flow on links in a PIM-SM domain. In these cases, traffic is not duplicated to the receivers, but wastes bandwidth.

The following figure shows such a situation. Switch B is the DR between switches A and B. Switch C is the RP. A receiver R is on the VLAN (V1) that connects switches A and B. A source sends multicast data to the receiver.

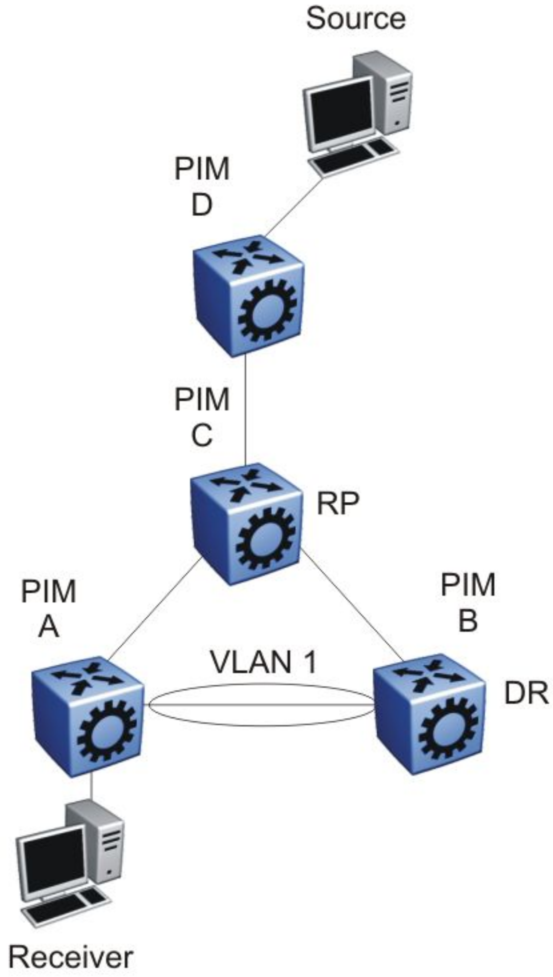

Figure 66: Receivers on interconnected VLANs

IGMP reports that the messages that the receiver sends are forwarded to the DR, and both A and B create (*,G) records. Switch A receives duplicate data through the path from C to A, and through the second path from C to B to A. Switch A discards the data on the second path (assuming the upstream source is A to C).

To avoid this waste of resources, Avaya recommends that you do not place receivers on V1. This configuration guarantees that no traffic flows between B and A for receivers attached to A. In this case, the existence of the receivers is only learned through PIM join messages to the RP [for (*,G)] and of the source through SPT joins.

PIM network with non-PIM interfaces

For proper multicast traffic flow in a PIM-SM domain, as a general rule, enable PIM-SM on all interfaces in the network (even if paths exist between all PIM interfaces). Enable PIM on all interfaces because PIM-SM relies on the unicast routing table to determine the path to the RP, BSR, and multicast sources. Ensure that all routers on these paths have PIM-SM enabled interfaces.

Figure 67: PIM network with non-PIM interfaces on page 142 provides an example of this situation. If A is the RP, then initially the receiver receives data from the shared tree path (that is, through switch A).

If the shortest path from C to the source is through switch B, and the interface between C and B does not have PIM-SM enabled, then C cannot switch to the SPT. C discards data that comes through the shared path tree (that is, through A). The simple workaround is to enable PIM on VLAN1 between C and B.

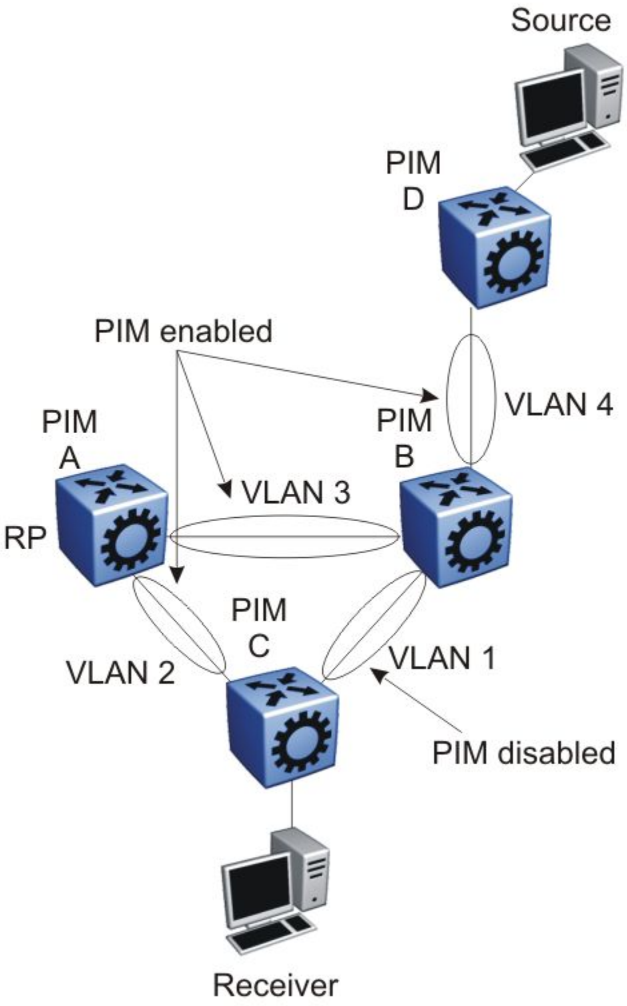

Figure 67: PIM network with non-PIM interfaces

Source filtering

The system can report interest in receiving packets from *only* a specific source address (INCLUDE), from all *but* specific source addresses (EXCLUDE), or sent to specific multicast addresses. IGMPv3 interacts with PIM-SM, PIM-SSM, and snooping to provide source filtering. For more information, see *Avaya Virtual Services Platform 9000 Configuration — IP Multicast Routing Protocols*, NN46250–504.

Protocol Independent Multicast-Source Specific Multicast guidelines

PIM-Source Specific Multicast (SSM) is a one-to-many model that uses a subset of the PIM-SM features. In this model, members of an SSM group can only receive multicast traffic from a single source, which is more efficient and puts less load on multicast routing devices.

IGMPv3 supports PIM-SSM by enabling a host to selectively request or filter traffic from individual sources within a multicast group. The system can report interest in receiving packets from *only* a specific source address (INCLUDE), from all *but* specific source addresses (EXCLUDE), or sent to specific multicast addresses. IGMPv3 interacts with PIM-SM, PIM-SSM, and snooping to provide source filtering.

IGMPv3 and PIM-SSM operation

Virtual Services Platform 9000 processes messages according to the following rules:

- After IGMPv3 receives an IGMPv2 report in the SSM range, the system translates the report to an IGMPv3 report message.
- After an IGMPv2 router sends queries on an IGMPv3 interface, the switch downgrades this interface to IGMPv2 (backward compatibility).

  This can cause traffic interruption, but the switch recovers quickly.
- After an interface receives an IGMPv3 report for a group with a different source than the one in the SSM channel table, the switch drops the report.

PIM-SSM design considerations

Use the following information when you design an SSM network:

- If you configure SSM, it affects SSM groups only. The switch handles other groups in sparse mode (SM) if a valid RP exists on the network.
- You can configure PIM-SSM only on switches at the edge of the network. Core switches use PIM-SM if they do not have receivers for SSM groups.
- For networks where group addresses are already in use, you can change the SSM range to match the groups.
- One switch has a single SSM range.
- You can have different SSM ranges on different switches.

Configure the core switches that relay multicast traffic so that they cover all of these groups in their SSM range, or use PIM-SM.

- One group in the SSM range can have multiple sources for a given SSM group.

For more information about multicast concepts and configuration, see *Avaya Virtual Services Platform 9000 Configuration — IP Multicast Routing Protocols*, NN46250–504.

IGMP and PIM-SM interaction

This section describes a possible problem that can arise if IGMP Snoop and PIM-SM interact. Consider the following network as an example: switches A and B run PIM-SM, and switch C runs IGMP Snoop. A and B interconnect with VLAN 1, and C connects A and B with VLAN 2.

If a receiver is placed in VLAN 2 on switch C, it does not receive data. PIM chooses the router with the higher IP address as the DR, whereas IGMP chooses the router with the lower IP address as the querier. Thus, if B becomes the DR, A becomes the querier on VLAN 2. IGMP reports are forwarded only to A on the mrouter port P1. A does not create a leaf because reports are received on the interface towards the DR.

You can avoid this problem in two ways:

- Configure ports P1 and P2 as mrouter ports on the IGMP Snoop VLAN.
- Configure switches A, B, and C to run the Multicast Router Discovery (MRDISC) protocol on their common VLANs.

A Layer 2 switch uses MRDISC to dynamically learn the location of switches A and B and thus, add them as mrouter ports.

Layer 3 multicast extended VLANs

Avoid using a Layer 3 multicast extended VLAN topology without SMLT.

Do not connect non-SMLT PIM routers in a linear fashion on the same VLAN. This topology is called an extended VLAN. Unlike a shared VLAN topology where all routers on the same VLAN are physically one hop away from each other, a VLAN router at one end of the extended VLAN has one or more routers in between it and the router at the far end of the extended VLAN. The following figure shows an extended VLAN.

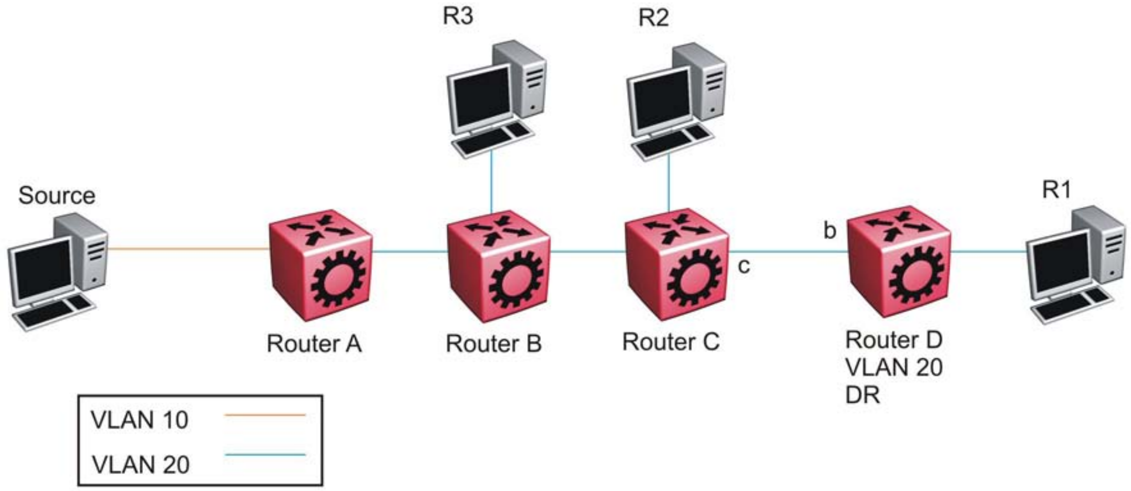

Figure 68: Extended VLAN topology

In the preceding figure, all routers use PIM-SSM. The source connects to Router A on VLAN 10. All routers and receiver hosts connect on the same extended VLAN, VLAN 20. All routers have a receiver in VLAN 20. Router D is the PIM DR for VLAN 20 and the source host is not on VLAN 20. PIM-SSM does not require a Rendezvous Point (RP).

In this topology, each router receives an IGMP membership report from its local receiver host, and then sends a PIM SG join message towards the source on VLAN 20. VLAN flooding propagates the PIM SG join message through to Router A, the PIM DR for the source VLAN 10. Each router from Router D to Router A records a PIM join on the port on which the join message was received, and then sends out its own join message toward the source. Data then flows from the source to the receiver, as long as a join exists on those ports.

Because all routers are in the same VLAN 20, they receive joins from one another due to flooding in the VLAN. For example, Router D receives join messages from Router C on its port ‘b’, and Router C receives join messages from Router B on its port toward Router B. On VSP 9000 routers, PIM join suppression is always enabled. In accordance with the PIM protocol rules, suppression causes Router D to stop sending a join towards the source because it receives a join for the same group and same RP on the port (port b) of the upstream neighbor (the router towards the source). Router D does not need to send a redundant join on the same VLAN. Router D stops sending a join, and the join that is recorded on port c of Router C eventually times out and is removed from the egress list of the (s,g) multicast route entry on Router C. This removal causes Router C to stop forwarding multicast traffic to Router D, and to the receiver (R1).

The purpose of join suppression is to suppress joins on a shared VLAN, such that if all routers on the shared VLAN want to receive data from the same RP and group, then only one of them needs to send the join on the VLAN. One join is enough to pull the data from the source router to the shared VLAN for all routers to receive. The other routers can suppress sending their own joins when they see such a join on the port toward the upstream router. In this way, less protocol message congestion exists in the shared VLAN. In the following figure, Router D sends the initial join message, which is seen by Router B and Router C. Router B and Router C suppress their own join messages. Router A (the PIM DR for the source VLAN 10) sends the data to VLAN 20, which is received by Routers B, C, and D due to the shared (non-extended) VLAN topology, and traffic is forwarded to all receiver hosts.

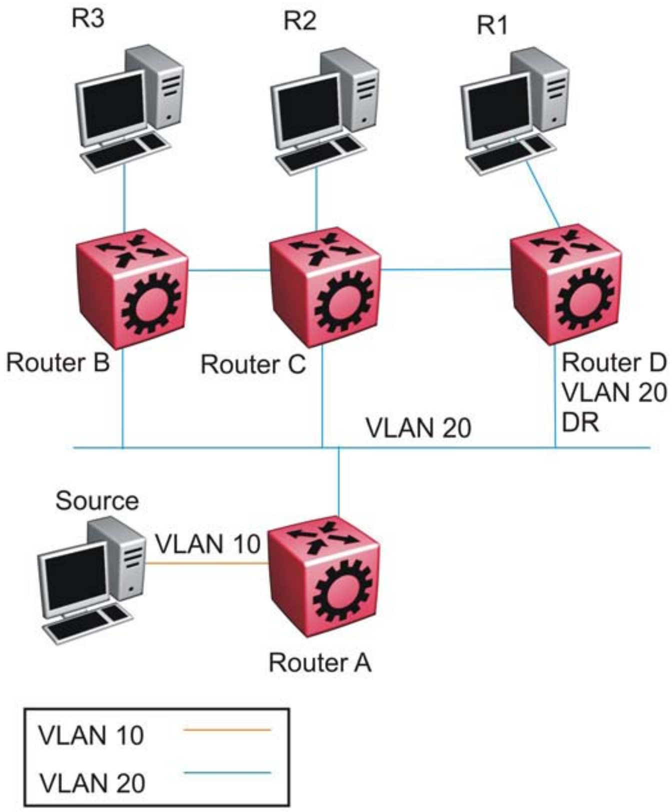

Figure 69: Shared VLAN topology

The extended VLAN topology looks exactly like the non-extended shared VLAN topology to the router, which cannot distinguish between the two.

In the current release, you cannot disable join suppression on a router. This enhancement will be added in a future release. Until this enhancement is included, you can perform the following actions:

1. Avoid this type of extended VLAN topology, and instead use Layer 3 routing between the routers. Do not extend VLAN 20 throughout, but rather, create a different VLAN between each router.
2. Configure the PIM DR for VLAN 20 to be the router closer to the source (Router B) so that any join received on the VLAN 20 DR (Router B) will be recorded as an IGMP local leaf on VLAN 20 as opposed to a PIM join, which does not time out until the receiver host stops sending IGMP membership reports.

Multicast for multimedia

Virtual Services Platform 9000 provides a flexible and scalable multicast implementation for multimedia applications. Several features are dedicated to multimedia applications and in particular, to television distribution.

Join and leave performance

For TV applications, you can attach several TVs directly, or through Ethernet Routing Switch 5600, to the Virtual Services Platform 9000. Base this implementation on IGMP; the set-top boxes use IGMP reports to join a TV channel and IGMP leaves to exit the channel. After a viewer changes channels, an IGMPv2 leave for the old channel (multicast group) is issued, and a membership report for the new channel is sent. If viewers change channels continuously, the number of joins and leaves can become large, particularly if many viewers attach to the switch.

Virtual Services Platform 9000 supports more than a thousand joins and leaves per second, which is well adapted to TV applications.

Important:

For IGMPv3, Avaya recommends that you ensure a join rate of 1000 per second or less. This ensures the timely processing of join requests.

If you use the IGMP proxy functionality at the receiver edge, you reduce the number of IGMP reports received by Virtual Services Platform 9000. This provides better overall performance and scalability.

Fast Leave

IGMP Fast Leave supports two modes of operation: single user mode and multiple user mode.

In single user mode, if more than one member of a group is on the port and one of the group members leaves the group, everyone stops receiving traffic for this group. A group-specific-query is not sent before the effective leave takes place.

Multiple user mode allows several users on the same port or VLAN. If one user leaves the group and other receivers exist for the same stream, the stream continues. The switch tracks the number of receivers that join a given group. For multiple user mode to operate properly, do not suppress reports. This ensures that the switch properly tracks the correct number of receivers on an interface.

The Fast Leave feature is particularly useful in IGMP-based TV distribution where only one receiver of a TV channel connects to a port. In the event that a viewer changes channels quickly, you create considerable bandwidth savings if you use Fast Leave.

You can implement Fast Leave on a VLAN and port combination; a port that belongs to two different VLANs can have Fast Leave enabled on one VLAN (but not on the other). Thus, with the Fast Leave feature enabled, you can connect several devices on different VLANs to the same port. This strategy does not impact traffic after one device leaves a group to which another device subscribes. For example, you can use this feature when two TVs connect to a port through two set-top boxes, even if you use the single user mode.

To use Fast Leave, you must first enable explicit host tracking. IGMP uses explicit host tracking to track all source and group members. Explicit host tracking is disabled by default. For configuration information, see *Avaya Virtual Services Platform 9000 Configuration — IP Multicast Routing Protocols*, NN46250–504.

Last member query interval tuning

If an IGMPv2 host leaves a group, it notifies the router by using a leave message. Because of the IGMPv2 report suppression mechanism, the router is unaware of other hosts that require the stream. Thus, the router broadcasts a group-specific query message with a maximum response time equal to the last member query interval (LMQI).

Because this timer affects the latency between the time that the last member leaves and when the stream actually stops, you must properly tune this parameter. This timer can especially affect TV delivery or other large-scale, high-bandwidth multimedia applications. For instance, if you assign a value that is too low, this can lead to a storm of membership reports if a large number of hosts are subscribed. Similarly, assigning a value that is too high can cause unwanted high-bandwidth stream propagation across the network if users change channels rapidly. Leave latency also depends on the robustness value, so a value of two equates to a leave latency of twice the LMQI.

Determine the proper LMQI value for your particular network through testing. If a very large number of users connect to a port, assigning a value of three can lead to a storm of report messages after a group-specific query is sent. Conversely, if streams frequently start and stop in short intervals, as in a TV delivery network, assigning a value of ten can lead to frequent congestion in the core network.

Another performance-affecting factor that you need to be aware of is the error rate of the physical medium. For links that have high packet loss, you can find it necessary to adjust the robustness variable to a higher value to compensate for the possible loss of IGMP queries and reports.

In such cases, leave latency is adversely impacted as numerous group-specific queries are unanswered before the stream is pruned. The number of unanswered queries is equal to the robustness variable (default two). The assignment of a lower LMQI can counterbalance this effect. However, if you configure the LMQI too low, it can actually exacerbate the problem by inducing storms of reports on the network. LMQI values of three and ten, with a robustness value of two, translate to leave latencies of six tenths of a second and two seconds, respectively.

When you choose an LMQI, consider all of these factors to determine the best configuration for the given application and network. Test that value to ensure that it provides the best performance.

Important:

In networks that have only one user connected to each port, Avaya recommends that you use the Fast Leave feature instead of LMQI, because no wait is required before the stream stops. Similarly, the robustness variable does not impact the Fast Leave feature, which is an additional benefit for links with high loss.

Chapter 16: System and network stability and security

Use the information in this section to design and implement a secure network.

You must provide security mechanisms to prevent your network from attack. If links become congested due to attacks, you can immediately halt end-user services. During the design phase, study availability issues for each layer. Without redundancy, all services can become unstable or unavailable. For more information about redundancy, see Redundant network design on page 35.

To provide additional network security, you can use the Avaya VPN Router product suite, or the Ethernet Routing Switch 8800. These products offer differing levels of protection against Denial of Service (DoS) attacks through either third party IDS partners, or through their own high-performance stateful firewalls.

Control plane rate limit (CP-Limit)

Use port and MultiLink Trunking (MLT) meters to configure the limit on the number of control and data exception packets that can enter on each port or MLT interface. You can optionally configure port and MLT meters to shut down the port or all ports in the case of MLT. If the number of packets exceeds the configured limit, the system generates a message in the log file. If shutdown is enabled, the system shuts down the port or all ports in the case of MLT and raises an alarm. You can disable the port to clear the alarm.

Be careful with shutdown. Sometimes a larger number of packets come in a port or MLT for legitimate reasons. For instance, when the chassis boots, if many multicast flows exist, the Control Processor is expected to receive a large amount of multicast data and control packets, which can exceed the configured meter value. This situation is normal behavior and you may not want to shut down the port (or all ports of the MLT) while the CP is learning multicast information.

The default value is 8000 packets per second with no shutdown. This feature prevents a single, unstable port from flooding the Control Processor with traffic. This feature differs from normal port rate limiting, which limits non-control traffic on the physical port that is not sent to the Control Processor (for example, IP subnet broadcast). Configure the CP-Limit feature on an individual port basis within the chassis.

CP-Limit cannot be enabled on IST ports because these are a critical for Split MLT (SMLT) configurations.

DoS protection mechanisms

Several internal mechanisms and features protect Virtual Services Platform 9000 against Denial-of-Service (DoS) attacks.

Broadcast and multicast rate limiting

To protect the switch and other devices from excessive broadcast traffic, you can use broadcast and multicast rate limiting on an individual port basis.

For more information about how to configure the rate limits for broadcast or multicast packets on a port, see *Avaya Virtual Services Platform 9000 Configuration — QoS and IP Filtering*, NN46250-502.

Directed broadcast suppression

You can enable or disable forwarding for directed broadcast traffic on an IP-interface basis. A directed broadcast is a frame sent to the subnet broadcast address on a remote IP subnet. By disabling or suppressing directed broadcasts on an interface, you cause all frames sent to the subnet broadcast address for a local router interface to be dropped. Directed broadcast suppression protects hosts from possible DoS attacks.

To prevent the flooding of other networks with DoS attacks, such as the Smurf attack, Virtual Services Platform 9000 is protected by directed broadcast suppression. This feature is enabled by default. Avaya recommends that you not disable it.

For more information about directed broadcast suppression, see *Avaya Virtual Services Platform 9000 Security*, NN46250-601.

Prioritization of control traffic

Virtual Services Platform 9000 uses a sophisticated prioritization scheme to schedule control packets on physical ports. This scheme involves two levels with both hardware and software queues to guarantee proper handling of control packets regardless of the switch load. In turn, this guarantees the stability of the network. Prioritization also guarantees that applications that use many broadcasts are handled with lower priority.

You cannot view, configure, or modify control traffic queues.

ARP request threshold recommendations

The Address Resolution Protocol (ARP) request threshold limits the ability of Virtual Services Platform 9000 to source ARP requests for workstation IP addresses the VSP 9000 does not learn within its ARP table. The default value for this function is 500 ARP requests per second. To avoid excessive amounts of subnet scanning caused by a virus, Avaya recommends that you change the ARP request threshold to a value between 100 to 50. This configuration protects the CPU from causing excessive ARP requests, protects the network, and lessens the spread of the virus to other PCs. The following list provides further recommended ARP threshold values:

- default: 500
- severe conditions: 50

- continuous scanning conditions: 100
- moderate: 200
- relaxed: 500

For more information about how to configure the ARP threshold, see *Avaya Virtual Services Platform 9000 Configuration — IP Routing* , NN46250-505.

Multicast Learning Limitation

The Multicast Learning Limitation feature protects the CPU from multicast data packet bursts generated by malicious applications. If more than a certain number of multicast streams enter the CPU through a port during a sampling interval, the port is shut down until the user or administrator takes the appropriate action.

For more information, see *Avaya Virtual Services Platform 9000 Configuration — IP Multicast Routing Protocols*, NN46250-504.

Damage prevention

To further reduce the chance that unauthorized users can use your network to damage other existing networks, take the following actions:

1. Prevent IP spoofing.

   You can use the spoof-detect feature.
2. Prevent the use of the network as a broadcast amplification site.
3. To block illegal IP addresses, enable the **`hsecure`** flag (High Secure mode).

   For more information, see *Avaya Virtual Services Platform 9000 Security*, NN46250-601.

Packet spoofing

You can stop spoofed IP packets by configuring the switch to only forward IP packets that contain the correct source IP address of your network. By denying all invalid source IP addresses, you minimize the chance that your network is the source of a spoofed DoS attack.

A spoofed packet is one that comes from the Internet into your network with a source address equal to one of the subnet addresses on your network. The source address belongs to one of the address blocks or subnets on your network. To provide spoofing protection, you can use a filter that examines the source address of all outside packets. If that address belongs to an internal network or a firewall, the packet is dropped.

To prevent DoS attack packets that come from your network with valid source addresses, you need to know the IP network blocks in use. You can create a generic filter that:

- permits valid source addresses
- denies all other source addresses

To do so, configure an ingress filter that drops all traffic based on the source address that belongs to your network.

If you do not know the address space completely, it is important that you at least deny private (see RFC1918) and reserved source IP addresses. The following table lists the source addresses to filter.

Table 12: Source addresses to filter

Address	Description
0.0.0.0/8	Historical broadcast. High Secure mode blocks addresses 0.0.0.0/8 and 255.255.255.255/16. If you enable this mode, you do not need to filter these addresses.
10.0.0.0/8	RFC1918 private network
127.0.0.0/8	Loopback
169.254.0.0/16	Link local networks
172.16.0.0/12	RFC1918 private network
192.0.2.0/24	TEST-NET
192.168.0.0/16	RFC1918 private network
224.0.0.0/4	Class D multicast
240.0.0.0/5	Class E reserved
248.0.0.0/5	Unallocated
255.255.255.255/32	Broadcast1

You can also enable the spoof-detect feature on a port.

For more information about the spoof-detect feature, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500.

High Secure mode

To ensure that Virtual Services Platform 9000 does not route packets with an illegal source address of 255.255.255.255 (RFC1812 Section 4.2.2.11 and RFC971 Section 3.2), you can enable High Secure mode.

By default, this feature is disabled. After you enable this flag, the feature applies to all ports.

For more information about hsecure, see *Avaya Virtual Services Platform 9000 Security*, NN46250-601.

Security and redundancy

Redundancy in hardware and software is one of the key security features of Virtual Services Platform 9000. High availability is achieved by eliminating single points of failure in the network and by using the unique features of Virtual Services Platform 9000 including:

- a complete, redundant hardware architecture (switching fabrics in load sharing, CPU in redundant mode or High Availability [HA] mode, redundant power supplies)
- hot swapping of all elements (I/O modules, Switch Fabric (SF) modules, CPUs, power supplies)
- flash cards to save multiple configuration files
- a list of software features that allow high availability including:
    - link aggregation: MultLink Trunking (MLT), distributed MLT, and 802.3ad
    - dual-homing of edge switches to two core switches: Split MLT (SMLT) and Routed SMLT (RSMLT)
    - unicast dynamic routing protocols: Routing Information Protocol (RIP) versions 1 and 2, Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP) version 4
    - multicast dynamic routing protocols: partial support for Protocol Independent Multicast-Sparse Mode (PIM-SM) and Protocol Independent Multicast-Source Specific Mode (PIM-SSM)
    - distribution of routing traffic along multiple paths: Equal Cost Multipath (ECMP)
    - router redundancy: Virtual Router Redundancy Protocol (VRRP)

Data plane security

Data plane security mechanisms include the Extended Authentication Protocol (EAP) 802.1x, VLANs, filters, routing policies, and routing protocol protection.

To protect the network from inside threats, the switch supports the 802.1x standard. EAP separates user authentication from device authentication. If you enable EAP, end-users must securely logon to the network before obtaining access to a resource.

Interaction between 802.1x and Optivity Policy Server v4.0

User-based networking links EAP authorization to individual user-based security policies based on individual policies. As a result, network managers can define corporate policies and configure them on an individual port basis. This configuration provides additional security based on a logon and password.

The Avaya Optivity Policy Server supports 802.1x EAP authentication against Remote Authentication Dial-in User Service (RADIUS) and other authentication, authorization, and accounting (AAA) repositories. This support authenticates the user, grants access to specific applications, and provides real time policy provisioning capabilities to mitigate the penetration of unsecured devices.

The following figure shows the interaction between 802.1x and Optivity Policy Server. First, the user initiates a logon from a user access point and receives a request/identify request from the switch (EAP access point). The user receives a network logon. Prior to Dynamic Host Configuration Protocol (DHCP), the user does not have network access because the EAP access point port is in EAP blocking mode. The user provides logon credentials to the EAP access point using the Extensible Authentication Protocol Over LAN (EAPoL). The client PC is both a RADIUS peer user and an EAP supplicant.

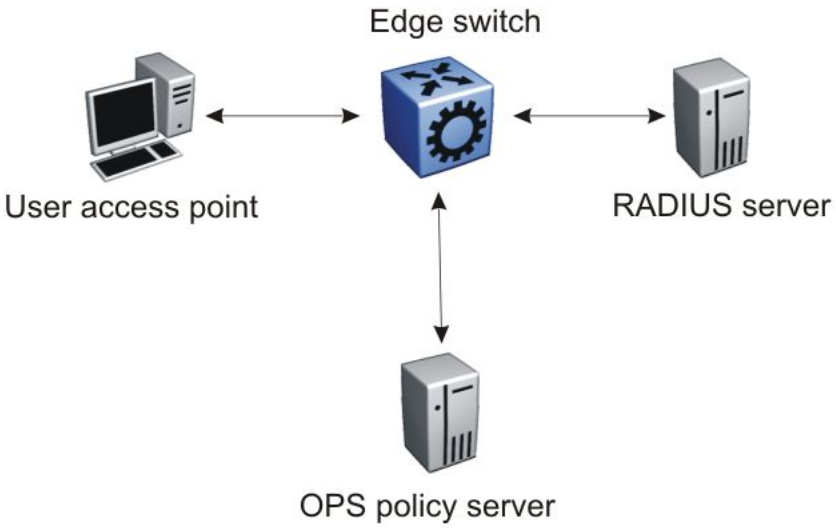

Figure 70: 802.1x and OPS interaction

Virtual Services Platform 9000 includes software support for the Preside (Funk) and Microsoft IAS RADIUS servers. Additional RADIUS servers that support the EAP standard are also compatible with Virtual Services Platform 9000. For more information, contact your Avaya representative.

802.1x and the LAN Enforcer or Avaya Health Agent

The Sygate LAN Enforcer or the Avaya Health Agent enables Virtual Services Platform 9000 to use the 802.1x standard to ensure that a user who connects from inside a corporate network is legitimate. The LAN Enforcer or Health Agent also checks the endpoint security posture, including anti-virus, firewall definitions, Windows registry content, and specific file content (plus date and size). Noncompliant systems that attempt to obtain switch authentication can be placed in a remediation VLAN, where updates can be pushed to the internal user, and users can subsequently attempt to join the network again.

VLANs and traffic isolation

You can use Virtual Services Platform 9000 to build secure VLANs. If you configure port-based VLANs, each VLAN is completely separate from the others. Virtual Services Platform 9000 supports the IEEE 802.1Q specification for tagging frames and coordinating VLANs across multiple switches.

Virtual Services Platform 9000 analyzes each packet independently of preceding packets. This mode, as opposed to the cache mode that other vendors use, allows complete traffic isolation.

For more information about VLANs, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500.

Security at layer 2

At Layer 2, Virtual Services Platform 9000 provides the following security mechanisms:

- access policies

  If you enable access policies globally, the system creates a default policy (1) that allows File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP), Telnet, and Secure Shell (SSH). If you enable access policies globally but disable the default policy, the system denies FTP, HTTP, rlogin, SSH, Simple Network Management Protocol (SNMP), Telnet, and Trivial FTP (TFTP).

  The **`access-strict`** parameter ties to the **`accesslevel`** parameter. If you enable **`access-strict`**, the access policy looks at the **`accesslevel`** parameter, and only applies to that access level. Use the following configuration as an example:

  ```
  VSP-9012:1(config)#show access-policy

    AccessPolicyEnable: off

                    Id: 1
                  Name: default
          PolicyEnable: false
                  Mode: allow
               Service: ftp|http|telnet|ssh
            Precedence: 128
           NetAddrType: any
               NetAddr: N/A
               NetMask: N/A
       TrustedHostAddr: N/A
   TrustedHostUserName: none
           AccessLevel: readOnly
          AccessStrict: false
                 Usage: 0
  ```

  If you disable **`access-strict`** (false), the policy looks at the value for **`accesslevel`**, and then the system applies the policy to anyone with equivalent rights or higher. In this example, all levels include readonly so the default policy applies to l1, l2, l3, rw, ro, and rwa. If you enable **`access-strict`**, the system applies the policy only to ro.

  For SNMP and access policies, you must apply the service to the access policy - the only choice is snmpv3 but this parameter applies to all versions of SNMP. The additional command **`access-policy <1–65535> snmp-group WORD<1–32> <snmpv1|snmpv2|usm>`** applies the policy to the SNMP community or the SNMP group.

- Loop detect

  Use loop detect to detect the MAC addresses that loop from one port to another port. After the system detects a loop, you can configure the action taken for the port on which the MAC addresses are learned. Enable loop detect on the interface and select one of three actions - mac-discard, port-down, or vlan-block.

Loop detect examines the source MAC addresses that enter the device and after the system detects a loop, it blocks source or destination addresses that match that MAC address.

- filters

  ACL filters are used by individual VLANs to filter out packets based on source MAC, destination MAC and other criteria.

  For more information about these filters, see *Avaya Virtual Services Platform 9000 Configuration — QoS and IP Filtering*, NN46250-502.

- MAC security

  This feature eliminates the need for you to configure multiple individual VLAN filter records for the same MAC address. By using a global MAC filter, you can discard frames whose source or destination MAC addresses match a global list stored in the switch. You can also apply global MAC filtering to multicast MAC address. However, you cannot apply it to local, broadcast, Bridge Protocol Data Unit (BPDU) MAC, TDP MAC, or All-Zeroes MAC addresses. After you add a MAC address to this global list, you cannot configure it statically on a VLAN and it cannot be learned on a VLAN. In addition, the switch does not perform bridging or routing on packets to or from this MAC address on a VLAN.

  MAC security applies only to bridged packets. Use ACL-based filters for routed packets.

  For more information about how to configure MAC security, see *Avaya Virtual Services Platform 9000 Configuration — VLANs and Spanning Tree*, NN46250-500.

- unknown MAC discard for Layer 2 MAC security

  The unknown MAC discard feature secures the network by restricting MAC learning by configuring a set of allowed MACs per port. In addition, users can limit the number of MACS learned on a port. The switch locks these learned MAC addresses in the forwarding database (FDB) and does not accept new MAC addresses on the port.

- limited MAC learning

  This feature limits the number of FDB-entries learned on a particular port to a user-specified value. After the number of learned FDB-entries reaches the maximum limit, the switch drops packets with unknown source MAC addresses. If the count drops below a configured minimum value due to FDB aging, learning is reenabled on the port.

  You can configure various actions like logging, sending traps, and disabling the port after the number of FDB entries reaches the configured maximum limit.

Security at Layer 3: filtering

At Layer 3 and higher, Virtual Services Platform 9000 provides enhanced filtering capabilities as part of its security strategy to protect the network from different attacks.

Virtual Services Platform 9000 supports advanced filters based on Access Control Lists (ACL).

Customer Support Bulletins (CSBs) are available on the Avaya Technical Support Web site to provide information and configuration examples about how to block some attacks.

Security at Layer 3: announce and accept policies

You can use route policies to selectively accept or announce some networks and to block the propagation of some routes. Route policies enhance the security in a network by hiding the visibility of some networks (subnets) from other parts of the network.

You can apply one policy for one purpose. For example, you can apply a RIP announce policy on a given RIP interface. In such cases, all sequence numbers under the given policy apply to that filter. A sequence number also acts as an implicit preference (that is, a lower sequence number is preferred).

Routing protocol security

You can protect OSPF and BGP updates with a Message Digest 5 (MD5) key on each interface. At most, you can configure two MD5 keys for each interface. You can also use multiple MD5 key configurations for MD5 transitions without bringing down an interface.

For more information, see *Avaya Virtual Services Platform 9000 Configuration — OSPF and RIP*, NN46250-506 and *Avaya Virtual Services Platform 9000 Configuration — BGP Services*, NN46250-507.

Control plane security

The control plane physically separates management traffic using the out of band (OOB) interface. The control plane facilitates High Secure mode, management access control, access policies, authentication, SSH and Secure Copy, and SNMP.

Management port

Virtual Services Platform 9000 provides an isolated management port on the Control Processor (CP) module. This port separates user traffic from management traffic in highly sensitive environments, such as brokerages and insurance agencies. By using this dedicated network (see Figure 71: Dedicated Ethernet management link on page 158) to manage the switch, and by configuring access policies (if you enable routing), you can manage the switch in a secure fashion. You can also use terminal servers to access the console port on the CP module (see Figure 72: Terminal server access on page 158).

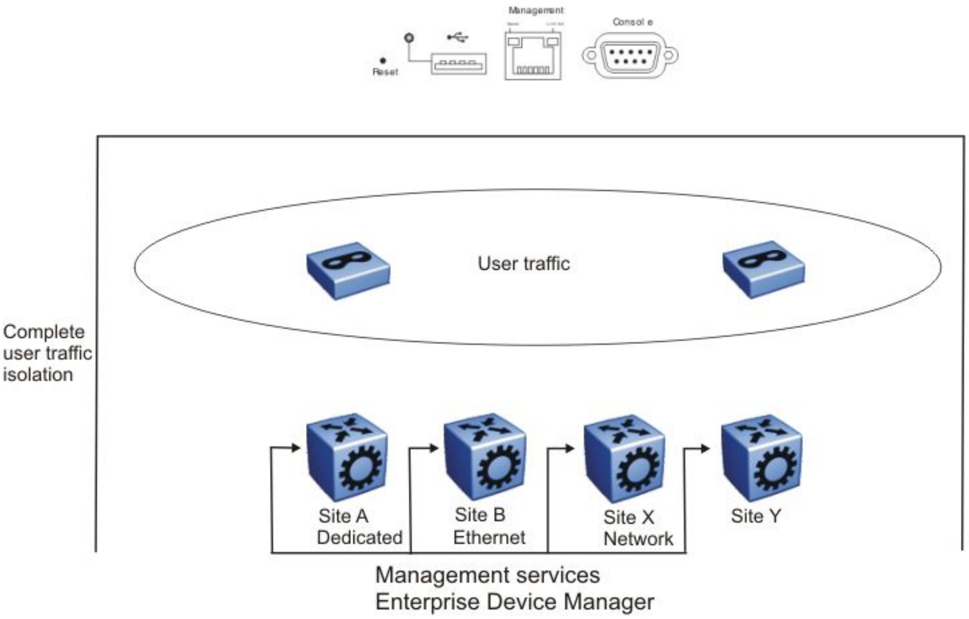

Figure 71: Dedicated Ethernet management link

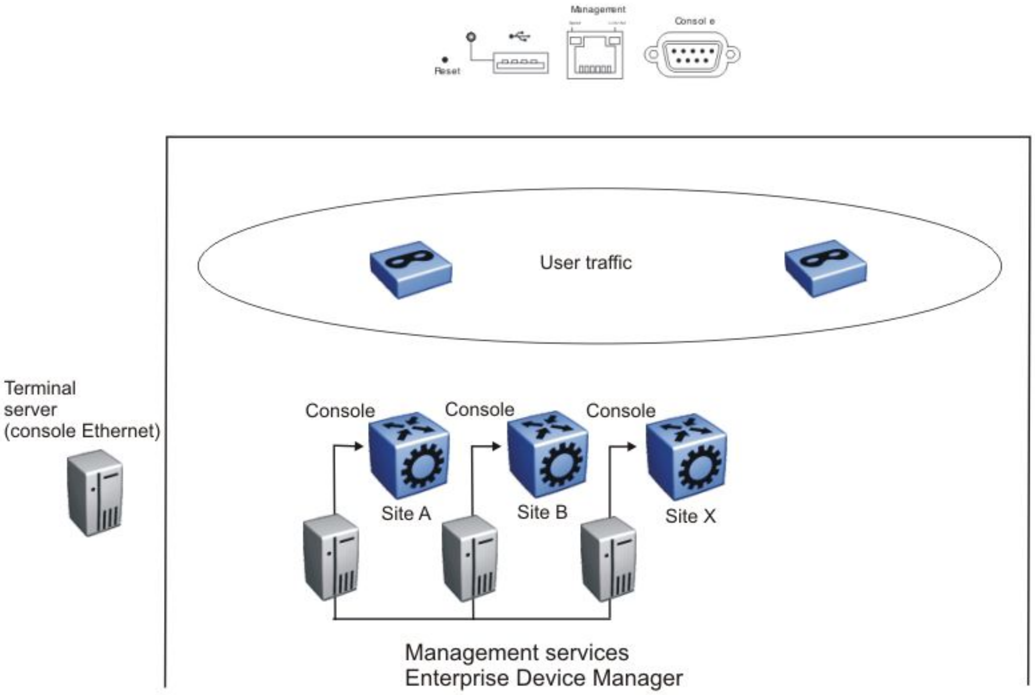

Figure 72: Terminal server access

If you must access the switch, Avaya recommends that you use the console port. The switch is always reachable, even if an issue occurs with the in-band network management interface.

Management access control

The following table shows management access levels. For more information, see *Avaya Virtual Services Platform 9000 Security*, NN46250-601.

Table 13: Management access levels

Access level	Description
Read only	Use this level to view the device configuration. You cannot change the configuration.
Layer 1 Read Write	Use this level to view switch configuration and status information and change only physical port parameters.
Layer 2 Read Write	Use this level to view and edit device configuration related to Layer 2 (bridging) functionality. The Layer 3 configuration, for example, OSPF, DHCP, are not accessible. You cannot change the security and password configuration.
Layer 3 Read Write	Use this level to view and edit device configuration related to Layer 2 (bridging) and Layer 3 (routing). You cannot change the security and password configuration.
Read Write	Use this level to view and edit most device configuration. You cannot change the security and password configuration.
Read Write All	Use this level to do everything. You have all the privileges of read-write access and the ability to change the security configuration. The security configuration include access passwords and the Web-based management user names and passwords. Read-Write-All (RWA) is the only level from which you can modify user-names, passwords, and SNMP community strings, with the exception of the RWA community string, which cannot be changed.

High Secure mode

Use High Secure to disable all unsecured applications and daemons, for example, FTP, TFTP, and rlogin. Avaya strongly recommends that you do not use unsecured protocols. See also High Secure mode on page 152.

Use Secure Copy (SCP) rather than FTP or TFTP.

Security and access policies

Access policies permit secure switch access by specifying a list of IP addresses or subnets that can manage the switch for a specific daemon, such as Telnet, SNMP, HTTP, SSH, TFTP,

FTP, RSH, and rlogin. Rather than using a management VLAN that is spread out among all of the switches in the network, you can build a full Layer 3 routed network and securely manage the switch with one of the in-band IP addresses attached to one of the VLANs (see the following figure).

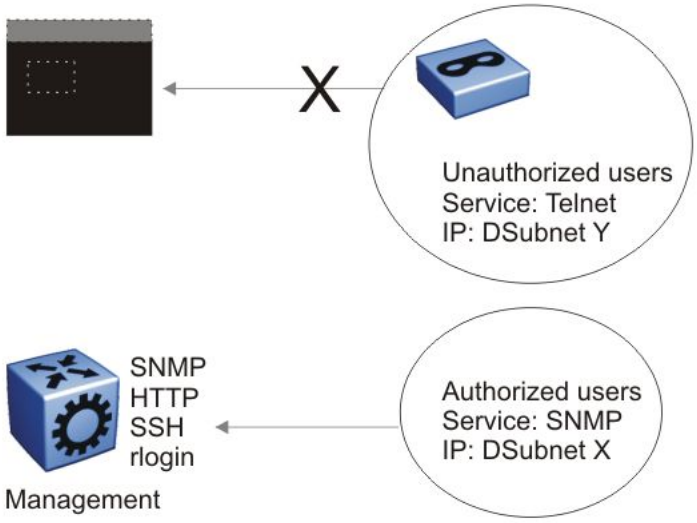

Figure 73: Access levels

Avaya recommends that you use access policies for in-band management to secure access to the switch. By default, all services are denied. You must enable the default policy or enable a custom policy to provide access. A lower precedence takes higher priority if you use multiple policies. Preference 120 has priority over preference 128.

RADIUS authentication

You can enforce access control by using RADIUS. RADIUS provides a high degree of security against unauthorized access and centralizes the knowledge of security access based on a client and server architecture. The database within the RADIUS server stores a list of pertinent information about client information, user information, password, and access privileges including the use of the shared secret.

When the switch acts as a Network Access Server, it operates as a RADIUS client. The switch is responsible for passing user information to the designated RADIUS servers. Because the switch operates in a LAN environment, it allows user access through Telnet, rlogin, and console logon.

You can configure a list of up to 10 RADIUS servers on the switch. If the first server is unavailable, Virtual Services Platform 9000 tries the second, and so on, until it establishes a successful connection.

RADIUS authentication supports: WEB, CLI, SNMP, or Extensible Authentication Protocol over LAN (EAPoL). You can configure a list of up to 10 RADIUS servers for all four methods combined. If you configure six servers for EAPoL, you can configure four servers for the other methods.

You can use the RADIUS server as a proxy for stronger authentication (see the following figure), such as:

- SecurID cards
- Kerberos
- other systems like Terminal Access Controller Access-Control System Plus (TACACS+)

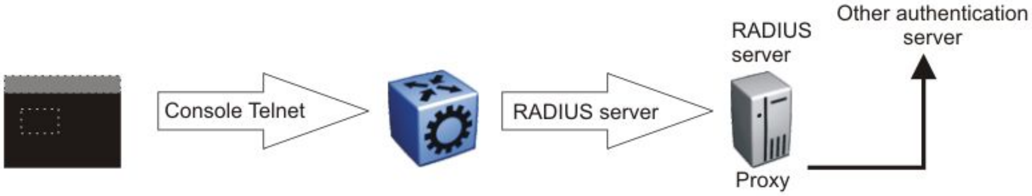

Figure 74: RADIUS server as proxy for stronger authentication

You must configure each RADIUS client to contact the RADIUS server. When you configure a client to work with a RADIUS server, complete the following configurations:

- Enable RADIUS.
- Provide the IP address of the RADIUS server.
- Ensure the shared secret matches what is defined in the RADIUS server.
- Provide the attribute value.
- Provide the use by value.

  The use by value can be CLI, SNMP, IGAP, or EAPoL.
- Indicate the order of priority in which the RADIUS server is used. (Order is essential when more than one RADIUS server exists in the network.)
- Specify the User Datagram Protocol (UDP) port that the client and server use during the authentication process. The UDP port between the client and the server must have the same or equal value. For example, if you configure the server with UDP 1812, the client must use the same UDP port value.

Other customizable RADIUS parameters require careful planning and consideration, for example, switch timeout and retry. Use the switch timeout to define the number of seconds before the authentication request expires. Use the retry parameter to indicate the number of retries the server accepts before sending an authentication request failure.

Avaya recommends that you use the default value in the attribute-identifier field. If you change the default value, you must alter the dictionary on the RADIUS server with the new value. To configure the RADIUS feature, you require Read-Write-All access to the switch.

For more information about RADIUS, see *Avaya Virtual Services Platform 9000 Security*, NN46250-601.

Encryption of control plane traffic

Control plane traffic encryption involves SSHv1/v2, SCP, and SNMPv3.

Use SSH to conduct secure communications over a network between a server and a client. The switch supports only the server mode (supply an external client to establish communication). The server mode supports SSHv1 and SSHv2.

The SSH protocol offers

- Authentication

  SSH determines identities. During the logon process, the SSH client asks for digital proof of the identity of the user.

- Encryption

  SSH uses encryption algorithms to scramble data. This data is rendered unintelligible except to the intended receiver.

- Integrity

  SSH guarantees that data is transmitted from the sender to the receiver without alteration. If a third party captures and modifies the traffic, SSH detects this alteration.

Virtual Services Platform 9000 supports

- SSH version 1, with password and Rivest, Shamir, Adleman (RSA) authentication
- SSH version 2 with password and Digital Signature Algorithm (DSA) authentication
- Digital Encryption Standard (DES)
- Triple DES (3DES)
- Advanced Encryption Standard (AES)

You must load the encryption module before you can enable it. For more information about how to load encryption modules, see *Avaya Virtual Services Platform 9000 Security*, NN46250-601.

SNMP header network address

You can direct an IP header to have the same source address as the management virtual IP address for self-generated UDP packets. If you configure a management virtual IP address and enable the **`udpsrc-by-vip`** flag, the network address in the SNMP header is always the management virtual IP address. This configuration is true for all traps routed out on the I/O ports or on the out-of-band management Ethernet port.

SNMPv3 support

SNMP version 1 and version 2 are not secure because communities are not encrypted.

Avaya strongly recommends that you use SNMP version 3. SNMPv3 provides stronger authentication services and the encryption of data traffic for network management.

Other security equipment

Avaya offers other devices that increase the security of your network.

For sophisticated state-aware packet filtering (real stateful inspection), you can add an external firewall to the architecture. State-aware firewalls can recognize and track application flows that use not only static TCP and UDP ports, like Telnet or HTTP, but also applications that create and use dynamic ports, such as FTP, and audio and video streaming. For every packet, the state-aware firewall finds a matching flow and conversation.

The following figure shows a typical configuration used in firewall load balancing.

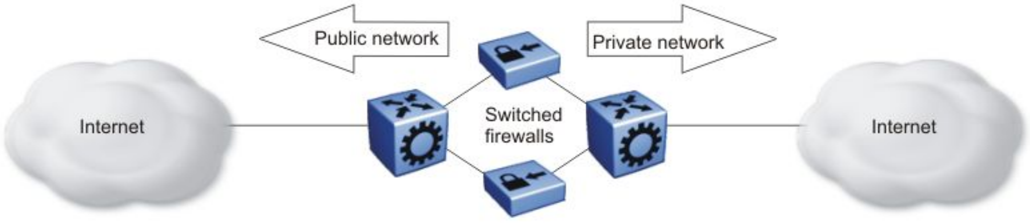

Figure 75: Firewall load balancing configuration

Use this configuration to redirect incoming and outgoing traffic to a group of firewalls and to automatically load balance across multiple firewalls. The benefits of such a configuration are

- increased firewall performance
- reduced response time
- redundant firewalls ensure Internet access

Virtual private networks (VPN) replace the physical connection between the remote client and access server with an encrypted tunnel over a public network. VPN technology employs IP security (IPsec) and Secure Sockets Layer (SSL) services.

Several Avaya products support IPSec and SSL, including Avaya VPN Gateway and Secure Router.

Additional information

The following organizations provide the most up-to-date information about network security attacks and recommendations about good practices:

- The Center of Internet Security Expertise (CERT)
- The Research and Education Organization for Network Administrators and Security Professionals (SANS)
- The Computer Security Institute (CSI)

Chapter 17: QoS design guidelines

This section provides design guidelines to provide Quality of Service (QoS) to user traffic on the network.

For more information about fundamental QoS mechanisms, and how to configure QoS, see *Avaya Virtual Services Platform 9000 Configuration — QoS and IP Filtering*, NN46250-502.

QoS mechanisms

Virtual Services Platform 9000 has a solid, well-defined architecture to handle QoS in an efficient and effective manner. The following sections briefly describe several QoS mechanisms that the platform uses.

QoS classification and mapping

Virtual Services Platform 9000 provides a hardware-based QoS platform through hardware packet classification. Packet classification is based on the examination of the QoS fields within the Ethernet packet, primarily the DiffServ Codepoint (DSCP) and the 802.1p fields.

You can configure ingress interfaces in one of two ways. In the first type of configuration, the interface does not classify traffic, but it forwards the traffic based on the packet markings. This mode of operation applies to trusted interfaces (core port mode) because the DSCP or 802.1p field is trusted to be correct, and the edge switch performs the mapping without classification.

In the second type of configuration, the interface classifies traffic as it enters the port, and marks the packet for further treatment as it traverses Virtual Services Platform 9000 network. This mode of operation applies to untrusted interfaces (access port mode) because the DSCP or 802.1p field is not trusted to be correct.

Virtual Services Platform 9000 assigns an internal QoS level to each packet that enters a port.

The Avaya QoS strategy simplifies QoS implementation by providing a mapping of various traffic types and categories to a Class of Service. These service classes are termed Avaya Service Classes (ASC). The following table provides a summary of the mappings and their typical traffic types.

Table 14: Traffic categories and ASC mappings

Traffic category	Application example	ASC
Network Control	Alarms and heartbeats	Critical
	Routing table updates	Network

Traffic category		Application example	ASC
Real-Time, Delay Intolerant		IP telephony; interhuman communication	Premium
Real-Time, Delay Tolerant		Video conferencing; interhuman communication.	Platinum
		Audio and video on demand; human-host communication	Gold
NonReal-Time Mission Critical	Interactive	eBusiness (B2B, B2C) transaction processing	Silver
	NonInteractive	Email; store and forward	Bronze
NonReal Time, NonMission Critical		FTP; best effort	Standard
		PointCast; Background/standby	Custom/ best effort

QoS and filters

Filters help you provide QoS by permitting or dropping traffic based on the parameters you configure. You can use filters to mark packets for specific treatment.

Typically, filters act as firewalls or are used for Layer 3 redirection. In more advanced cases, traffic filters can identify Layer 3 and Layer 4 traffic streams. The filters cause the streams to be re-marked and classified to attain a specific QoS level at both Layer 2 (802.1p) and Layer 3 (DSCP).

Traffic filtering is a key QoS feature. Virtual Services Platform 9000, by default, determines incoming packet 802.1p or DiffServ markings, and forwards traffic based on their assigned QoS levels. However, situations exist where the markings are incorrect, or the originating user application does not have 802.1p or DiffServ marking capabilities. Also, you can give a higher priority to select users (executive class). In these situations, use filters to prioritize specific traffic streams.

You can use filters to assign QoS levels to devices and applications. To help you decide whether or not to use a filter, key questions include:

1. Does the user or application have the ability to mark QoS information on data packets?
2. Is the traffic source trusted? Are the QoS levels configured appropriately for each data source?

   Users can maliciously configure QoS levels on their devices to take advantage of higher priority levels.
3. Do you want to prioritize traffic streams?

This decision-making process is outlined in the following figure.

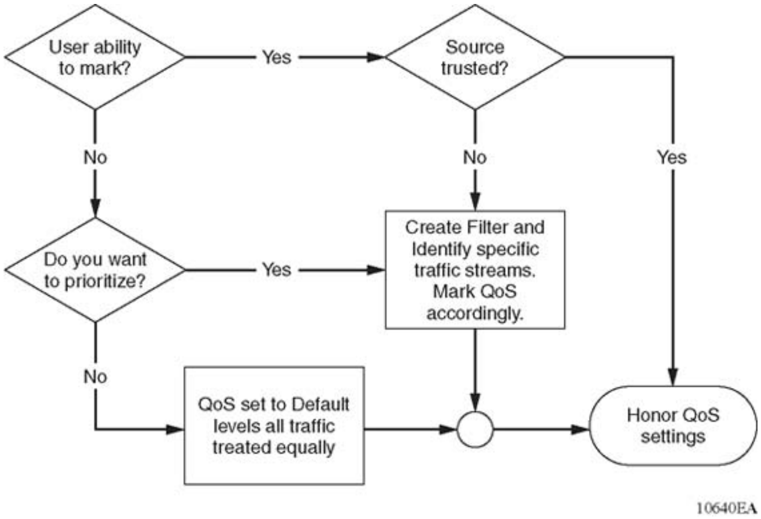
User ability to mark?
Yes
Source trusted?
No
No
Yes
Do you want to prioritize?
Yes
Create Filter and Identify specific traffic streams. Mark QoS accordingly.
No
QoS set to Default levels all traffic treated equally
Honor QoS settings
10640EA

Figure 76: Filter decision-making process

Configure filters through the use of access control lists (ACL) and access control entries (ACE), which are implemented in software. An ACL can include both security and QoS type ACEs. The platform supports 2048 ACLs and 1000 ACEs for each ACL to a maximum of 16 000 ACEs for each plaform.

The following steps summarize the filter configuration process:

1. Determine your desired match fields.
2. Create an ACL.
3. Create an ACE within the ACL.
4. Configure the desired precedence, traffic type, and action.

   You determine the traffic type by creating an ingress or egress ACL.
5. Modify the parameters for the ACE.

Policing and shaping

As part of the filtering process, you can police ingress traffic. Policing is performed according to the traffic filter profile assigned to the traffic flow. For enterprise networks, policing ensures that traffic flows conform to the criteria assigned by network managers.

Traffic policers identify traffic using a traffic policy. Traffic that conforms to this policy is guaranteed for transmission, whereas nonconforming traffic is considered to be in violation. Traffic policers drop packets if traffic is excessive, or remark the DSCP or 802.1p markings by using filter actions. With Virtual Services Platform 9000, you can define multiple actions in case of traffic violation.

For service providers, policing at the network edge provides different bandwidth options as part of a Service Level Agreement (SLA). For example, in an enterprise network, you can police the traffic rate from one department to give critical traffic unlimited access to the network. In a service provider network, you can control the amount of traffic customers send to ensure that they comply with their SLA. Policing ensures that users do not exceed their traffic contract for a QoS level.

VSP 9000 supports two rate, three color marking for policers as described in RFC 2698. Policers mark packets as Green, Yellow, or Red. Red packets are dropped automatically. Out of profile packets cannot be remarked to a lower QoS level.

The system can perform rate metering only on a Layer 3 basis.

Traffic shapers buffer and delay violating traffic. These operations occur at the egress level. Virtual Services Platform 9000 supports traffic shaping at the port level.

QoS interface considerations

Four QoS interface types are explained in detail in the following sections. You can configure an interface as trusted or untrusted, and for bridging or routing operations. Use these parameters to properly apply QoS to network traffic.

Trusted and untrusted interfaces

You can configure an interface as trusted (core) or untrusted (access). The default is trusted (core).

Use trusted interfaces (core) to mark traffic in a specific way, and to ensure that packets are treated according to the service level of those markings. Use a core interface if you need control over network traffic prioritization. For example, use 802.1p-bits to apply desired class of service (CoS) attributes to the packets before they are forwarded to the access node. You can also classify other protocol types ahead of IP packets.

A core port preserves the DSCP and 802.1p-bits markings. The device uses these values to assign a corresponding QoS level to the packets.

Use an access port to control the classification and mapping of traffic for delivery through the network. Untrusted interfaces require you to configure filter sets to classify and re-mark ingress traffic. For untrusted interfaces in the packet forwarding path, the DSCP is mapped to an IEEE 802.1p user priority field in the IEEE 802.1Q frame, and both of these fields are mapped to an IP Layer 2 drop precedence value that determines the forwarding treatment at each network node along the path. Traffic that enters an access port is re-marked with the appropriate DSCP and 802.1p markings, and given an internal QoS level. The switch performs this re-marking based on the filters and traffic policies that you configure.

The following logical table shows how the system performs ingress mappings for data packets and for control packets not destined for the Control Processor (CP).

Table 15: Data packet ingress mapping

Enable DiffServ	Access DiffServ	802.1p Override	Routed Packet	Tagged Ingress Packet	Internal QoS Derived From	Egress Packet DSCP Derived from	Egress Packet 802.1p Derived from
1	0, L3T=1	0, L2T=1	1	1	DSCP	Stays untouched	iQoS
1	0, L3T=1	0, L2T=1	0	1	.1p	Stays untouched	iQoS
1	0, L3T=1	0, L2T=1	X	0	DCSP	Stays untouched	iQoS
1	1, L3T=0	0, L2T=1	X	1	.1p	iQoS	iQoS
1	1, L3T=0	0, L2T=1	X	0	Port QoS	iQoS	iQoS
0	X, L3T=0	0, L2T=1	X	1	.1p	Stays untouched	iQoS
0	X, L3T=0	0, L2T=1	X	0	Port QoS	Stays untouched	iQoS
1	0, L3T=1	1, L2T=0	X	X	DSCP	Stays untouched	iQoS
1	1, L3T=0	1, L2T=0	X	X	Port QoS	iQoS	iQoS

Bridged and routed traffic

In a service provider network, access nodes use Virtual Services Platform 9000 for bridging. In this case, Virtual Services Platform 9000 uses DiffServ to manage network traffic and resources, but some QoS features are unavailable in the bridging mode of operation.

In an enterprise network, access nodes use Virtual Services Platform 9000 for bridging, and core nodes use it for routing. For bridging, ingress traffic is mapped from the 802.1p-bit marking to a QoS level. For routing, ingress traffic is mapped from the DSCP marking to the appropriate QoS level.

802.1p and 802.1Q recommendations

In a network, to map the 802.1p user priority bits, use 802.1Q-tagged encapsulation on customer premises equipment (CPE). You require encapsulation because Virtual Services Platform 9000 does not provide classification when it operates in bridging mode.

To ensure consistent Layer 2 QoS boundaries within the service provider network, you must use 802.1Q encapsulation to connect a CPE directly to Virtual Services Platform 9000 access node. If you do not require packet classification, use Ethernet Routing Switch 5600 to connect to the access node. In this case, configure the traffic classification functions in the Ethernet Routing Switch 5600.

At the egress access node, packets are examined to determine if their IEEE 802.1p or DSCP values must be re-marked before leaving the network. Upon examination, if the packet is a tagged packet, the IEEE 802.1p tag is configured based on the QoS level-to-IEEE 802.1p-bit mapping. For bridged packets, the DSCP is re-marked based on the QoS level.

Network congestion and QoS design

When you provide QoS in a network, one of the major elements you must consider is congestion, and the traffic management behavior during congestion. Congestion in a network is caused by many different conditions and events, including node failures, link outages, broadcast storms, and user traffic bursts.

At a high level, three main types or stages of congestion exist:

1. no congestion
2. bursty congestion
3. severe congestion

In a noncongested network, QoS actions ensure that delay-sensitive applications, such as real-time voice and video traffic, are sent before lower-priority traffic. The prioritization of delay-sensitive traffic is essential to minimize delay and reduce or eliminate jitter, which has a detrimental impact on these applications.

A network can experience momentary bursts of congestion for various reasons, such as network failures, rerouting, and broadcast storms. Virtual Services Platform 9000 has sufficient capacity to handle bursts of congestion in a seamless and transparent manner. If the burst is not sustained, the traffic management and buffering process on the switch allows all the traffic to pass without loss.

Severe congestion is defined as a condition where the network or certain elements of the network experience a prolonged period of sustained congestion. Under such congestion conditions, congestion thresholds are reached, buffers overflow, and a substantial amount of traffic is lost.

After the switch detects severe congestion, Virtual Services Platform 9000 discards traffic based on drop precedence values. This mode of operation ensures that high-priority traffic is not discarded before lower-priority traffic.

When you perform traffic engineering and link capacity analysis for a network, the standard design rule is to design the network links and trunks for a maximum average-peak utilization of no more than 80%. This value means that the network peaks to up to 100% capacity, but the average-peak utilization does not exceed 80%. The network is expected to handle momentary peaks above 100% capacity.

QoS examples and recommendations

The sections that follow present QoS network scenarios for bridged and routed traffic over the core network.

Bridged traffic

If you bridge traffic over the core network, you keep customer VLANs separate (similar to a Virtual Private Network). Normally, a service provider implements VLAN bridging (Layer 2) and no routing. In this case, the 802.1p-bit marking determines the QoS level assigned to each packet. If DiffServ is active on core ports, the level of service received is based on the highest of the DiffServ or 802.1p settings.

The following cases provide sample QoS design guidelines you can use to provide and maintain high service quality in a network.

If you configure a core port, you assume that, for all incoming traffic, the QoS value is properly marked. All core switch ports simply read and forward packets; they are not re-marked or reclassifiied. All initial QoS markings are performed at the customer device or on the edge devices.

The following figure illustrates the actions performed on three different bridged traffic flows (that is VoIP, video conference, and e-mail) at access and core ports throughout the network.

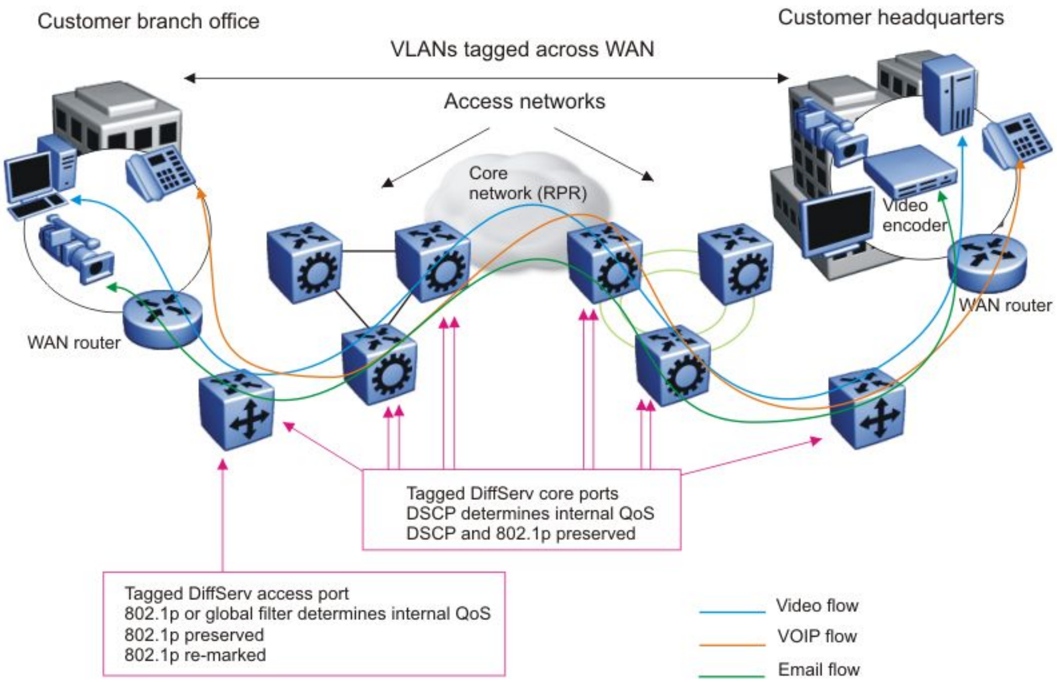

Figure 77: Trusted bridged traffic

For bridged, untrusted traffic, if you configure the port to access, mark and prioritize traffic on the access node using global filters. Reclassify the traffic to ensure it complies with the class of service specified in the SLA.

For Resilient Packet Ring (RPR) interworking, you can assume that, for all incoming traffic, the QoS configuration is properly marked by the access nodes. The core switch ports, configured as core or trunk ports, perform the RPR interworking. These ports preserve the DSCP marking and re-mark the 802.1p bit to match the 802.1p bit of the RPR. The following figure shows the actions performed on three different traffic flows (VoIP, video conference, and e-mail) over an RPR core network.

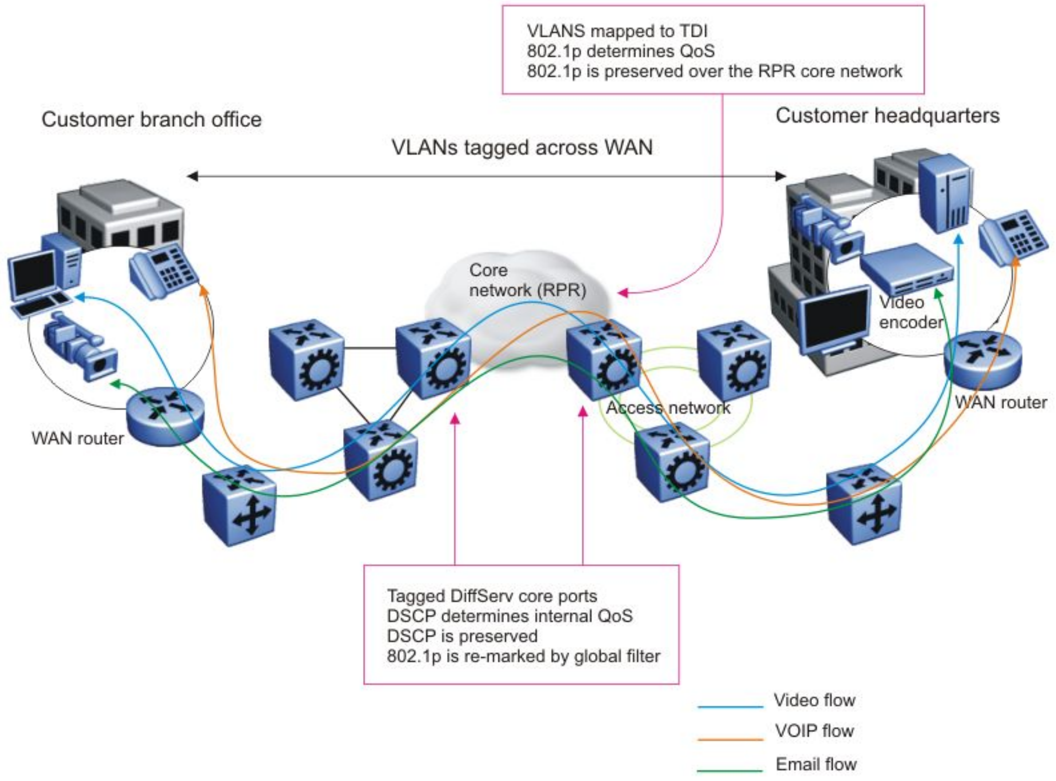

Figure 78: RPR QoS internetworking

Routed traffic

If you route traffic over the core network, VLANs are not kept separate.

If you configure the port to core, you assume that, for all incoming traffic, the QoS configuration is properly marked. All core switch ports simply read and forward packets. The switch does not re-mark or classify the packets. The customer device or the edge devices perform all initial QoS markings.

The following figure shows the actions performed on three different routed traffic flows (that is VoIP, video conference, and e-mail) at access and core ports throughout the network.

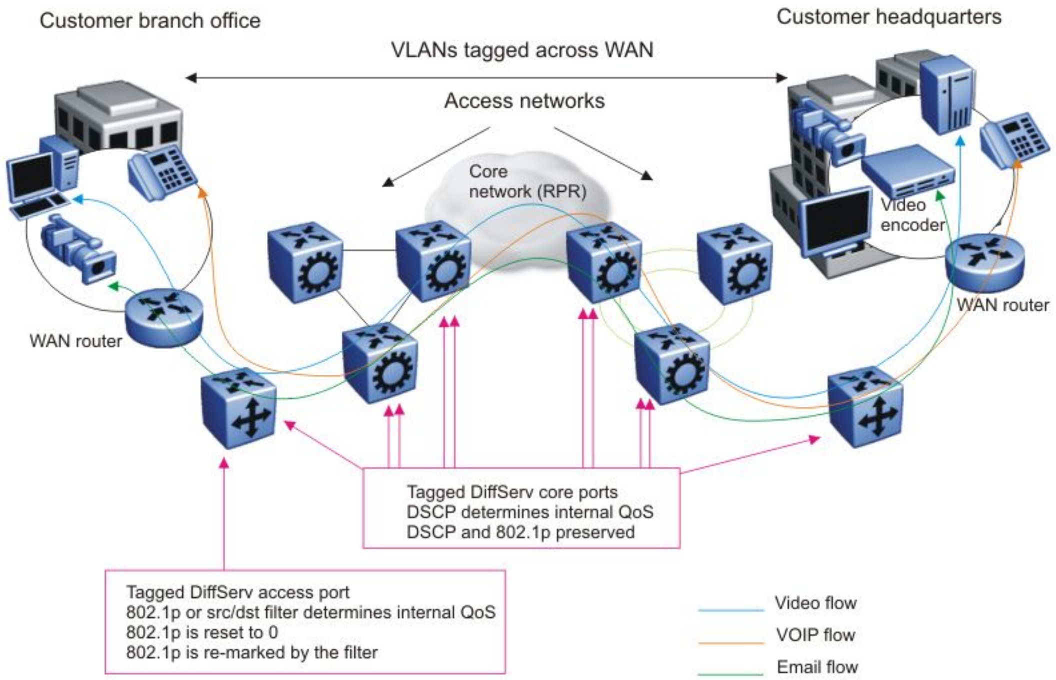
Customer branch office
Customer headquarters
VLANs tagged across WAN
Access networks
Core network (RPR)
Video encoder
WAN router
WAN router
Tagged DiffServ core ports
DSCP determines internal QoS
DSCP and 802.1p preserved
Tagged DiffServ access port
802.1p or src/dst filter determines internal QoS
802.1p is reset to 0
802.1p is re-marked by the filter
Video flow
VOIP flow
Email flow

Figure 79: Trusted routed traffic

For routed, untrusted traffic, in an access node, packets that enter through a tagged or untagged access port exit through a tagged or untagged core port.

Chapter 18: Layer 1, 2, and 3 design examples

This section provides examples to help design your network. Layer 1 examples deal with the physical network layouts. Layer 2 examples map Virtual Local Area Networks (VLAN) on top of the physical layouts. Layer 3 examples show the routing instances that Avaya recommends to optimize IP for network redundancy.

Layer 1 examples

The following figures are a series of Layer 1 examples that focus primarily on the physical network layout.

The following figure uses double physical links and distributed MultiLink Trunking (DMLT) to provide a redundant network.

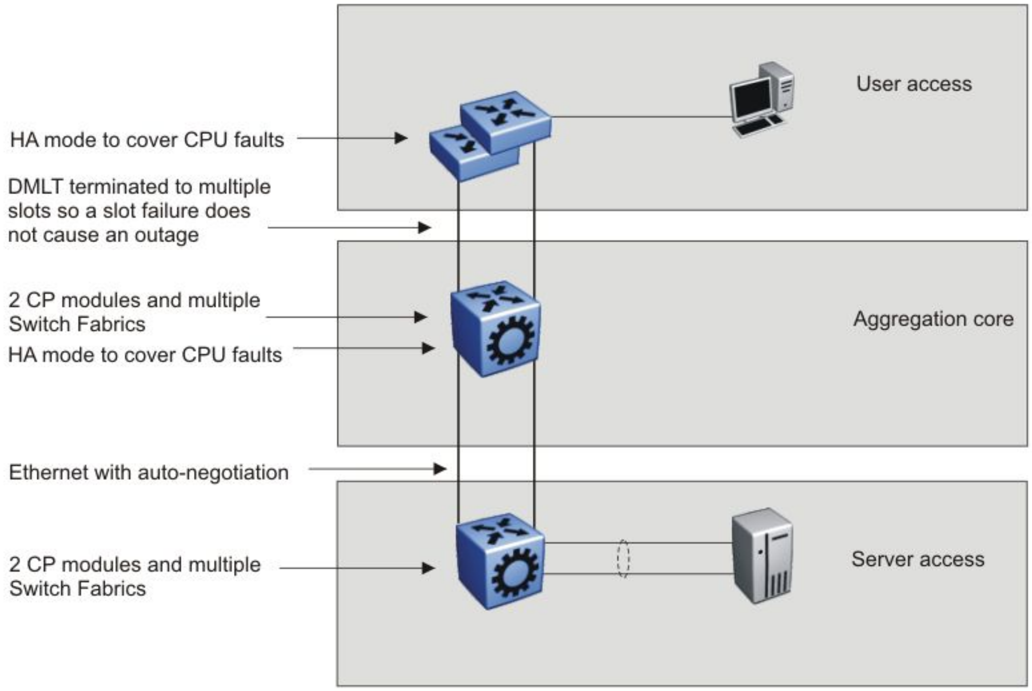
HA mode to cover CPU faults
DMLT terminated to multiple slots so a slot failure does not cause an outage
2 CP modules and multiple Switch Fabrics
HA mode to cover CPU faults
Ethernet with auto-negotiation
2 CP modules and multiple Switch Fabrics
User access
Aggregation core
Server access

Figure 80: Layer 1 design example 1

The following figure uses Split MultiLink Trunking (SMLT) to provide switch redundancy and DMLT to provide module redundancy.

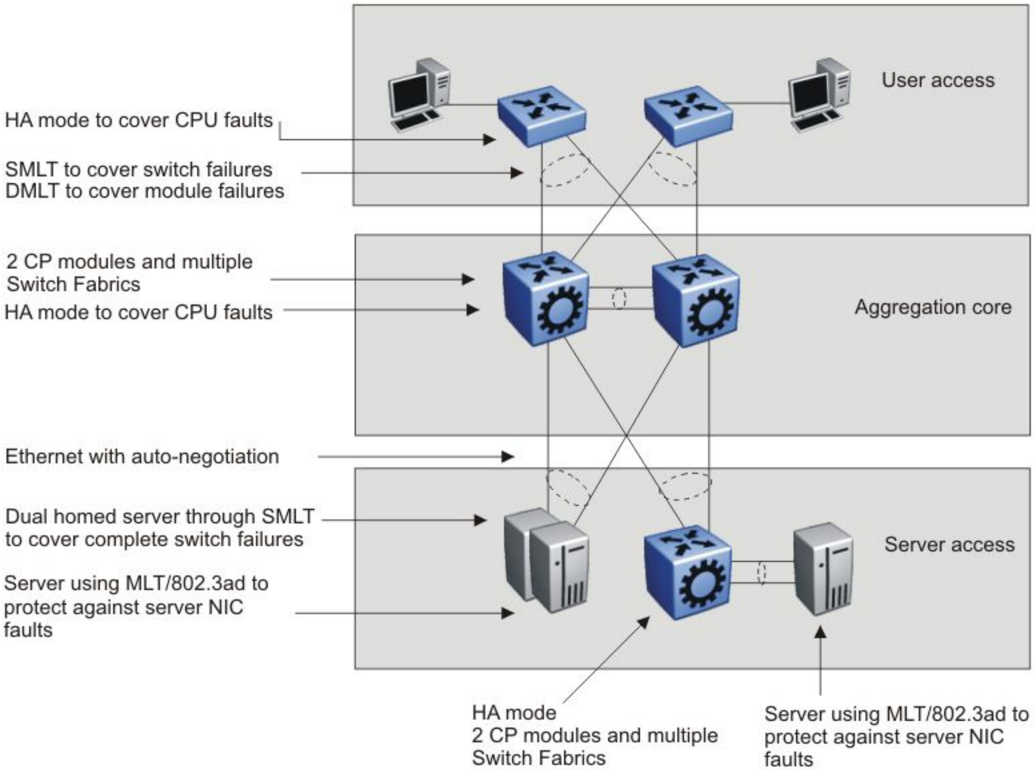
User access
HA mode to cover CPU faults
SMLT to cover switch failures
DMLT to cover module failures
2 CP modules and multiple
Switch Fabrics
HA mode to cover CPU faults
Aggregation core
Ethernet with auto-negotiation
Dual homed server through SMLT
to cover complete switch failures
Server using MLT/802.3ad to
protect against server NIC
faults
Server access
HA mode
2 CP modules and multiple
Switch Fabrics
Server using MLT/802.3ad to
protect against server NIC
faults

Figure 81: Layer 1 design example 2

The following figure is an example of a four-tiered topology. This example adds redundancy in the aggregation core.

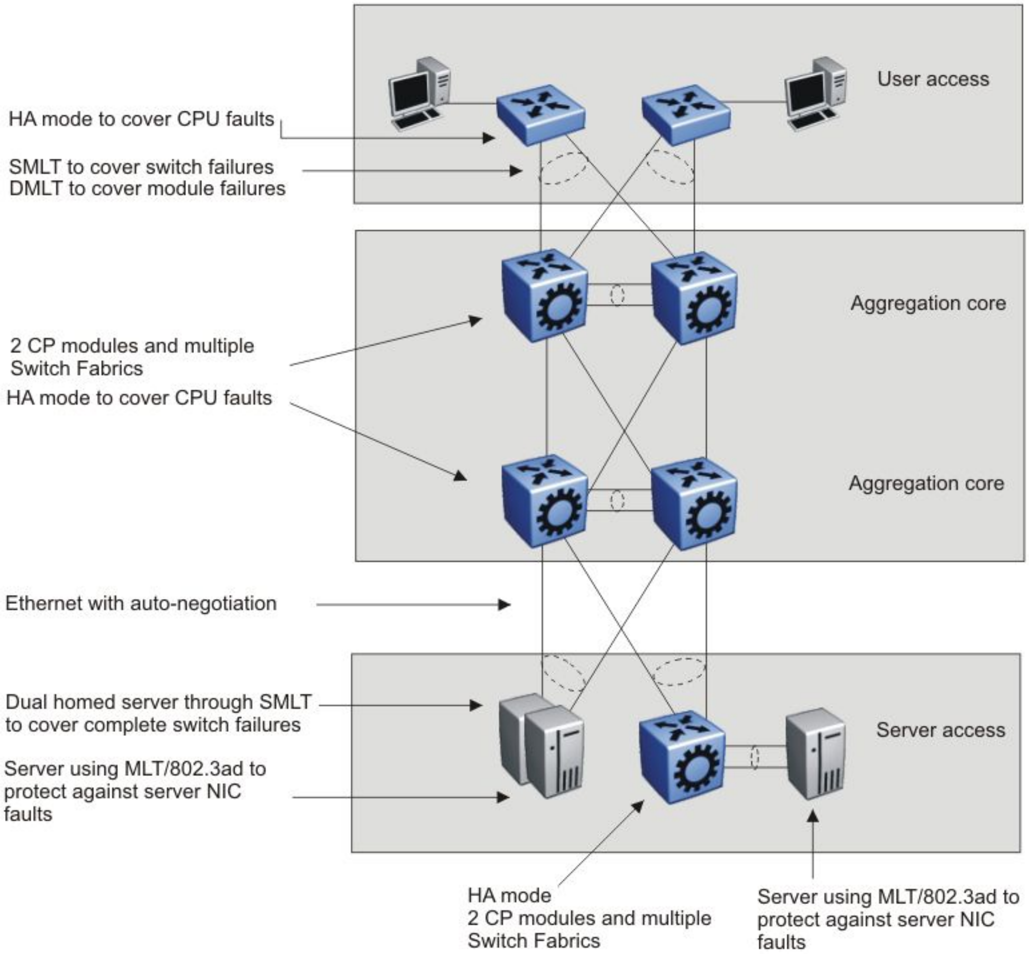
HA mode to cover CPU faults
SMLT to cover switch failures
DMLT to cover module failures
User access
2 CP modules and multiple Switch Fabrics
HA mode to cover CPU faults
Aggregation core
Aggregation core
Ethernet with auto-negotiation
Dual homed server through SMLT to cover complete switch failures
Server using MLT/802.3ad to protect against server NIC faults
Server access
HA mode
2 CP modules and multiple Switch Fabrics
Server using MLT/802.3ad to protect against server NIC faults

Figure 82: Layer 1 design example 3

Layer 2 examples

The following figures are a series of Layer 2 network design examples that map VLANs over the physical network layout.

Example 1 shows a redundant device network that uses one VLAN for all switches. To support multiple VLANs, you need 802.1Q tagging on the links with trunks.

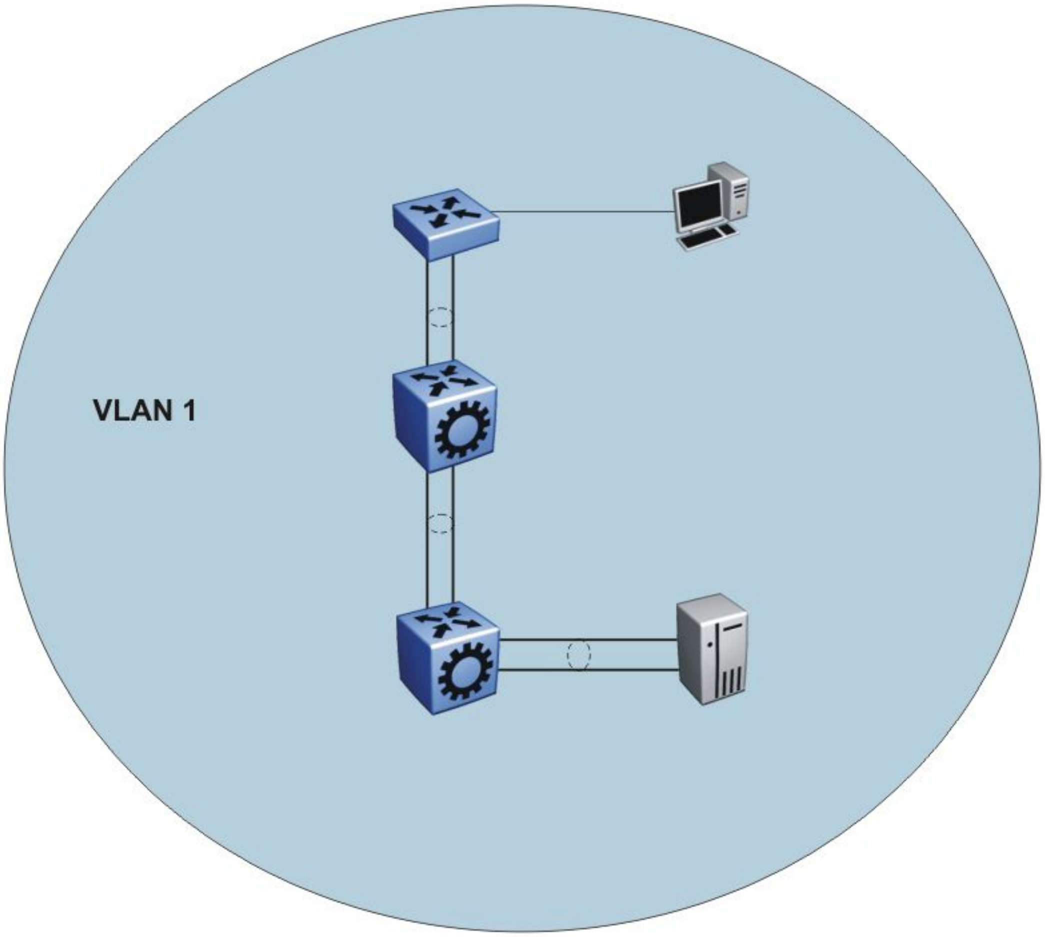

Figure 83: Layer 2 design example 1

Figure 84: Layer 2 design example 2 on page 180 depicts a redundant network that uses SMLT. This layout does not require the use of a spanning tree protocol. SMLT prevents loops and ensures that all paths are actively used. Each MLT trunk can have up to 16 links to the core. This SMLT configuration example is based on a three-stage network.

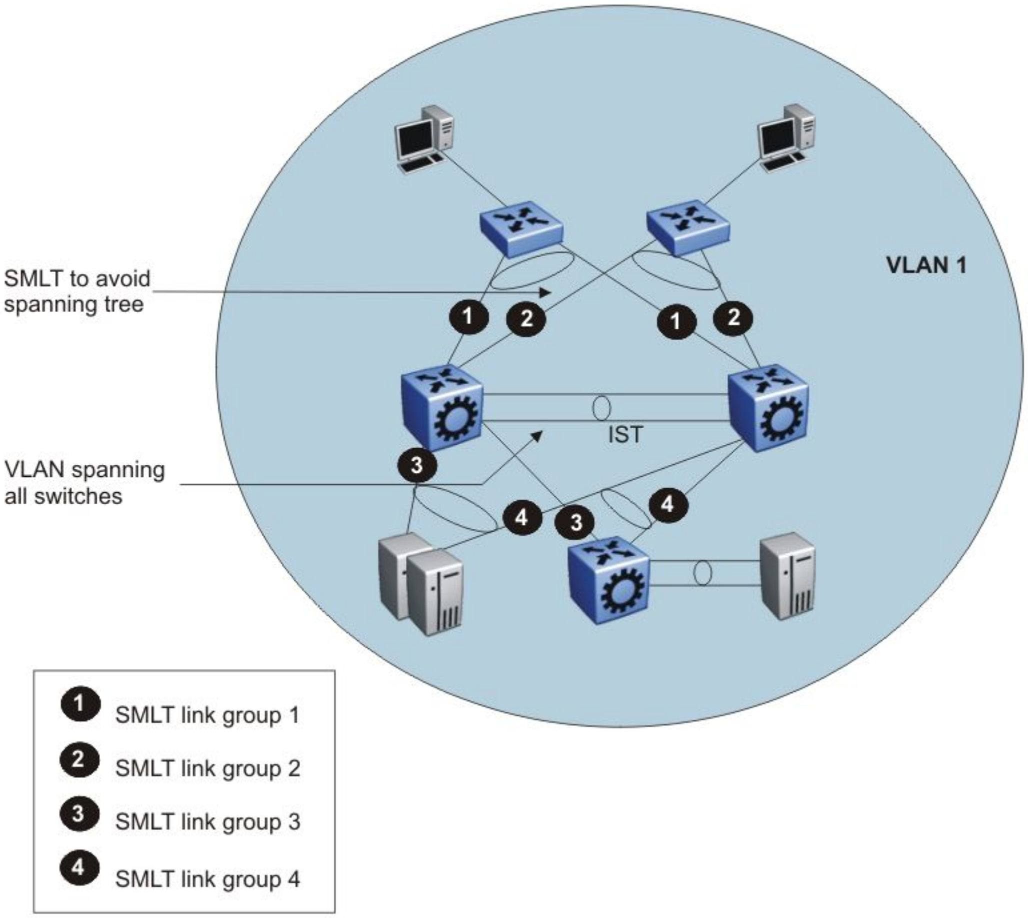

Figure 84: Layer 2 design example 2

The following figure depicts a redundant network that uses SMLT in both a triangle and full-mesh configuration. This layout does not require the use of a spanning tree protocol. SMLT prevents loops and ensures that all paths are actively used.

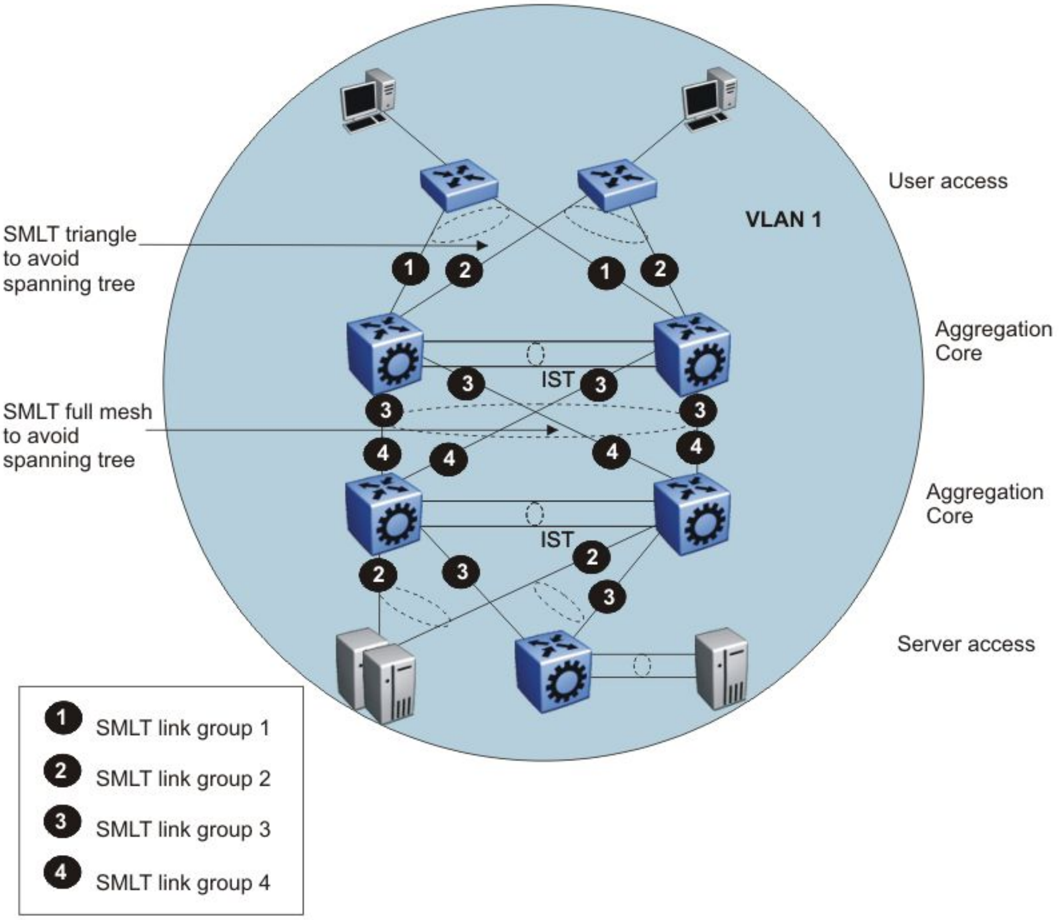

Figure 85: Layer 2 design example 3

The following figure shows a typical SMLT configuration with one aggregation pair connected. You can connect multiple aggregation-layer and access switches to create a very large, scalable network. This example uses one VLAN for all switches. To support multiple VLANs, you need 802.1Q tagging on the links with trunks.

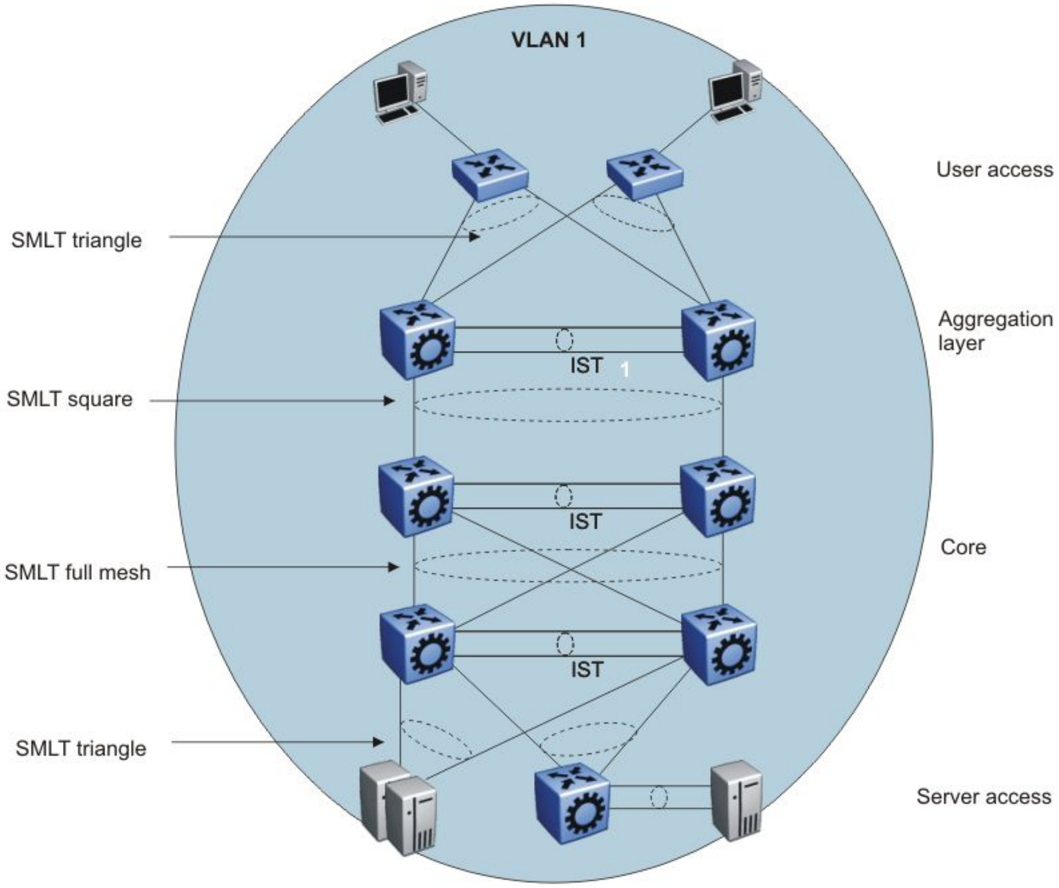

Figure 86: Layer 2 design example 4

Layer 3 examples

The following figures are a series of Layer 3 network design examples that show the routing instances that Avaya recommends you use to optimize IP for network redundancy.

Figure 87: Layer 3 design example 1 on page 183 uses redundant links.

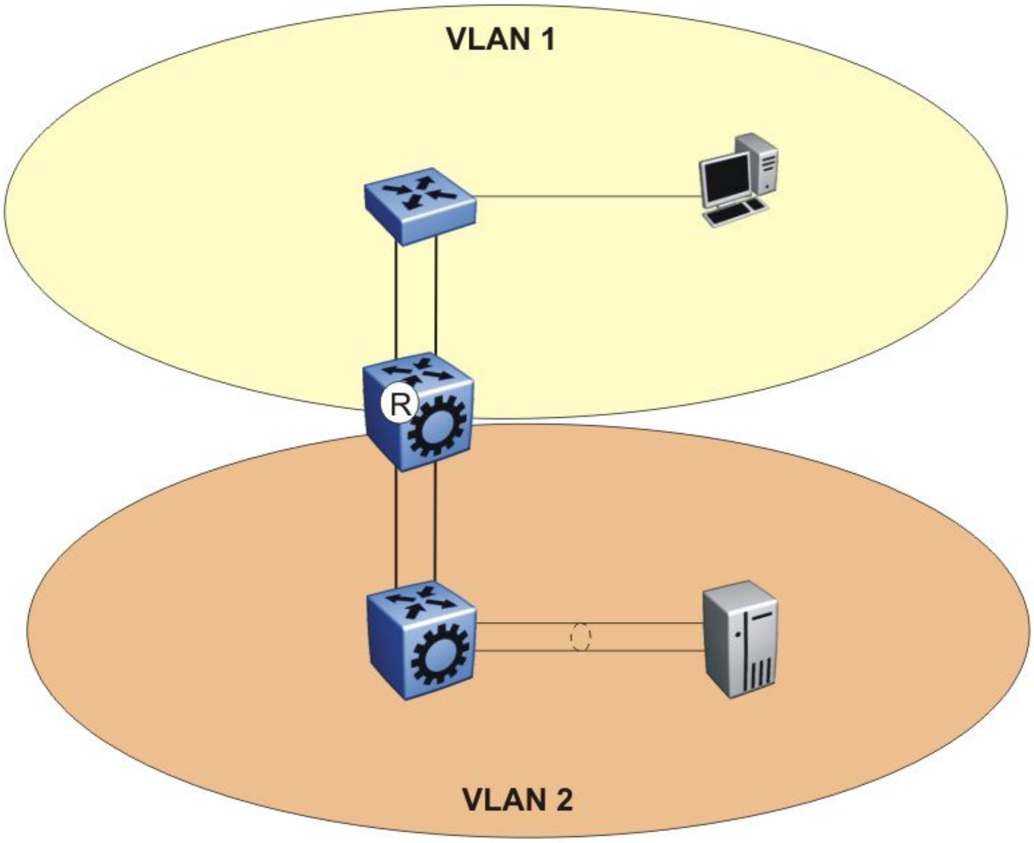

Figure 87: Layer 3 design example 1

Figure 88: Layer 3 design example 2 on page 184 uses the Virtual Router Redundancy Protocol (VRRP) to provide redundancy between the two switches.

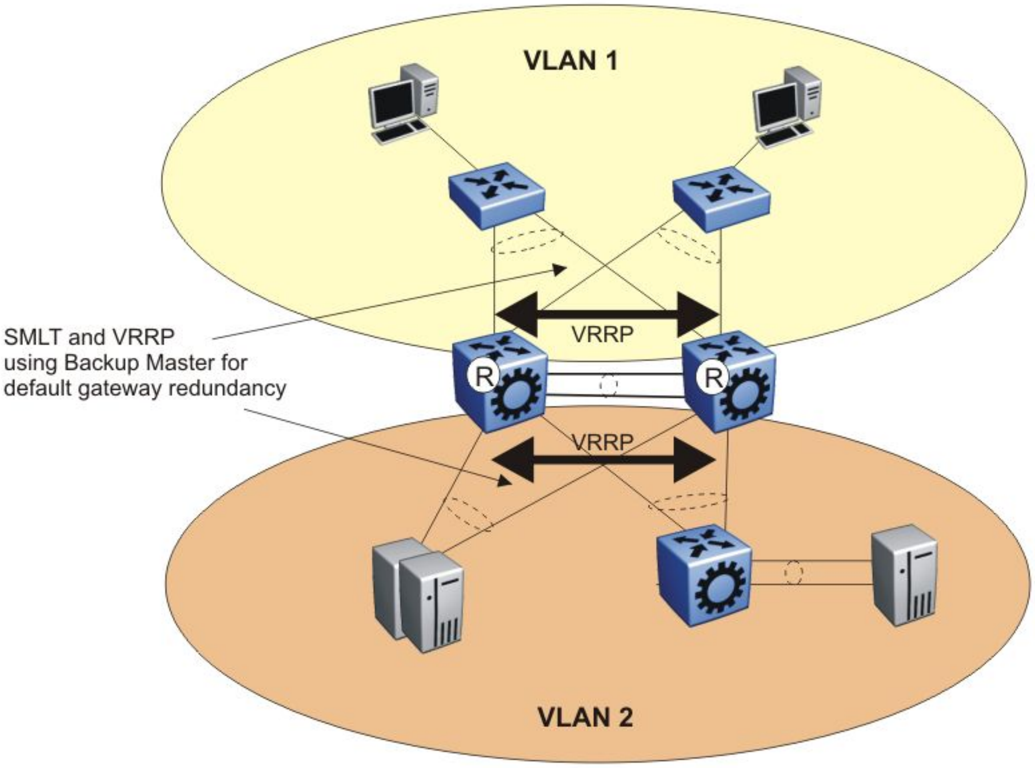
VLAN 1
VRRP
SMLT and VRRP using Backup Master for default gateway redundancy
R
R
VRRP
VLAN 2

Figure 88: Layer 3 design example 2

Figure 89: Layer 3 design example 3 on page 185 uses VRRP with Backup Master with Open Shortest Path First (OSPF) and Equal Cost Multipath (ECMP). In large scale environments, for example, more than 64 VRRP instances, Avaya recommends that you use Routed Split MultiLink Trunking (RSMLT) with RSMLT edge instead of VRRP.

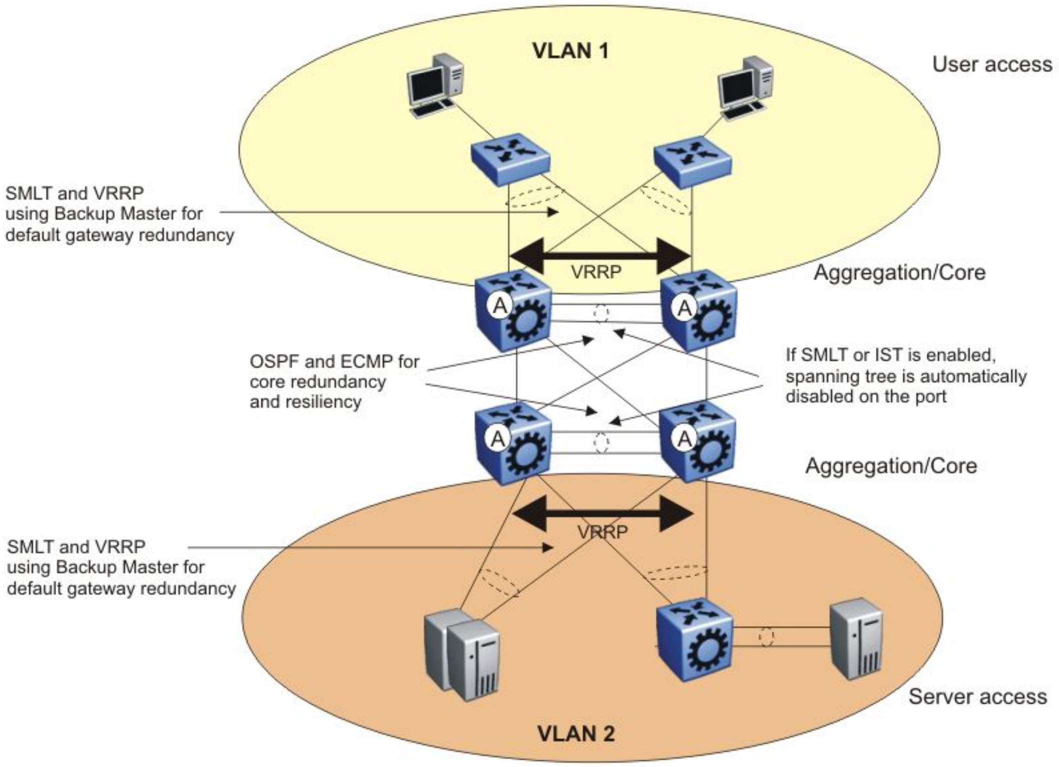

Figure 89: Layer 3 design example 3

Figure 90: Layer 3 design example 4 on page 186 uses one aggregation pair. You can connect multiple aggregation-layer and access switches to create a very large, scalable network.

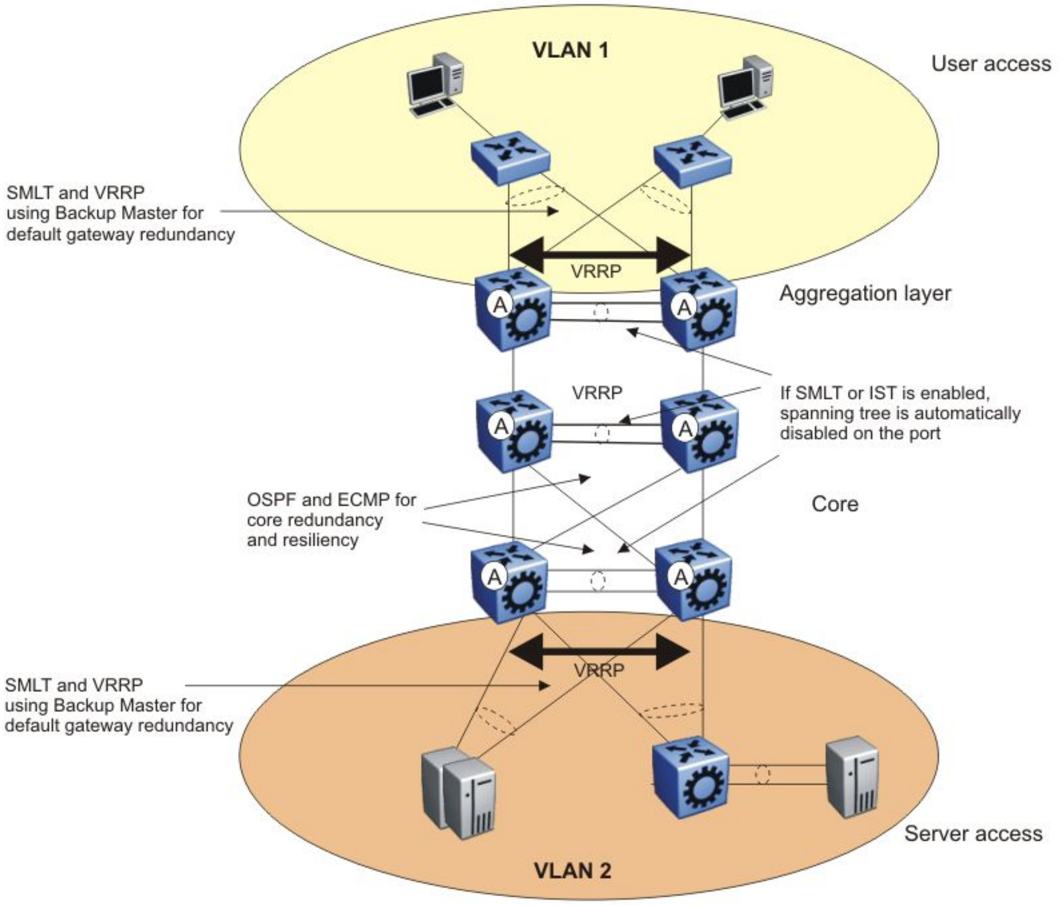

Figure 90: Layer 3 design example 4

RSMLT redundant network with bridged and routed VLANs in the core

In some networks, you need a VLAN to span through the core of a network, for example, a VoIP VLAN or guest VLAN, while routing other VLANs to reduce the amount of broadcasts or to provide separation.Figure 91: Redundant network design on page 187 shows a redundant network design that can perform these functions.

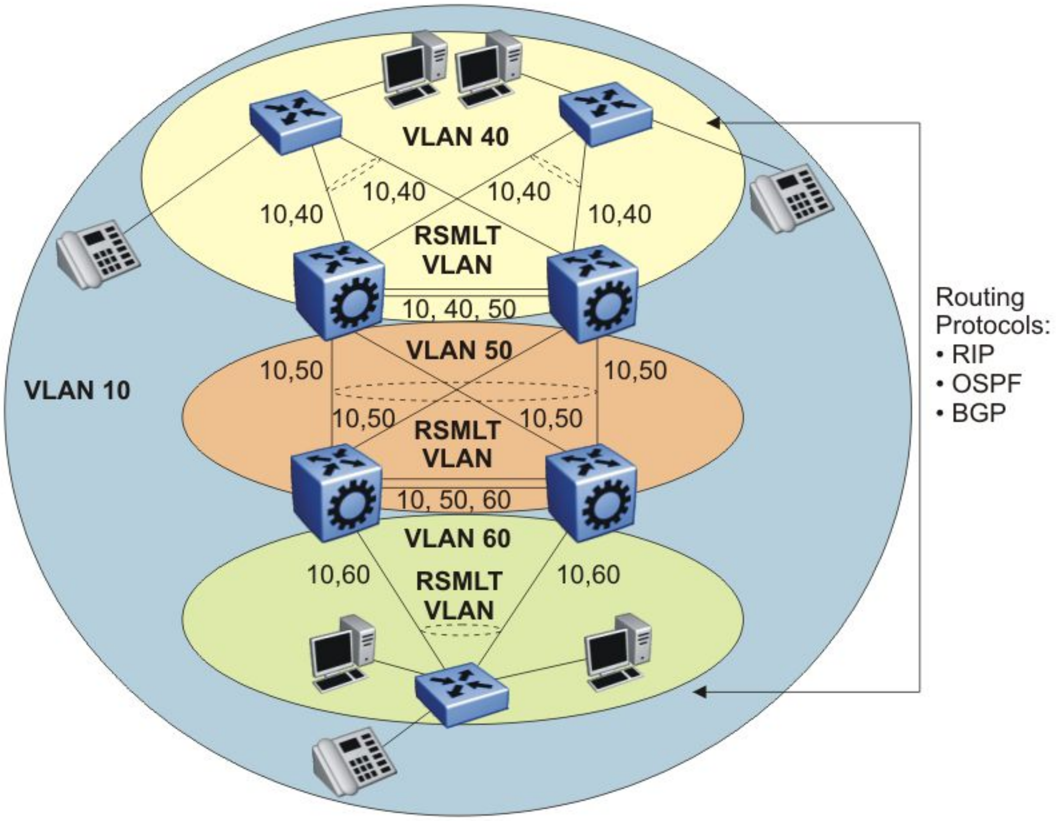

Figure 91: Redundant network design

In Figure 91: Redundant network design on page 187, VLAN-10 spans the complete campus network. VLANs 40, 50, and 60 are core VLANs with RSMLT enabled. VLANs 40, 50, and 60 and their IP subnets provide subsecond failover for the routed edge VLANs. You can use Routing Information Protocol (RIP), OSPF, or Border Gateway Protocol (BGP) to exchange routing information. RSMLT and its protection mechanisms prevent the routing protocol convergence time from impacting network convergence time.

All client stations that are members of a VLAN receive every broadcast packet. Each station analyzes each broadcast packet to decide whether the packets are destined for it or for another node in the VLAN. Typical broadcast packets are Address Resolution Protocol (ARP) requests, RIP updates, NetBios broadcasts, or Dynamic Host Control Protocol (DHCP) requests. Broadcasts increase the CPU load of devices in the VLAN.

To reduce this load, and to lower the impact of a broadcast storm (potentially introduced through a network loop), keep the number of VLAN members below 512 in a VLAN or IP subnet (you can use more clients for each VLAN or IP subnet). Use Layer 3 routing to connect the VLANs and IP subnets.

You can enable IP routing at the wiring-closet access layer in networks where many users connect to wiring-closets. Most late-model high-end access switches support Layer 3 routing in hardware.

To reduce network convergence time in case of a failure in a network with multiple IP client stations, Avaya recommends that you distribute the ARP request/second load to multiple IP routers or switches. Enabling routing at the access layer distributes the ARP load, which reduces the IP subnet sizes.Redundant network design on page 35 shows how to enable routing at the access layer while keeping the routing protocol design robust and simple.

In large scale environments, for example, more than 64 VRRP instances, Avaya recommends that you use RSMLT with RSMLT edge instead of VRRP. RSMLT provides a simple, more scalable solution than VRRP. The following figure shows a network with RSMLT edge. For more information, see SMLT and Layer 3 traffic redundancy (VRRP and RSMLT) on page 58.

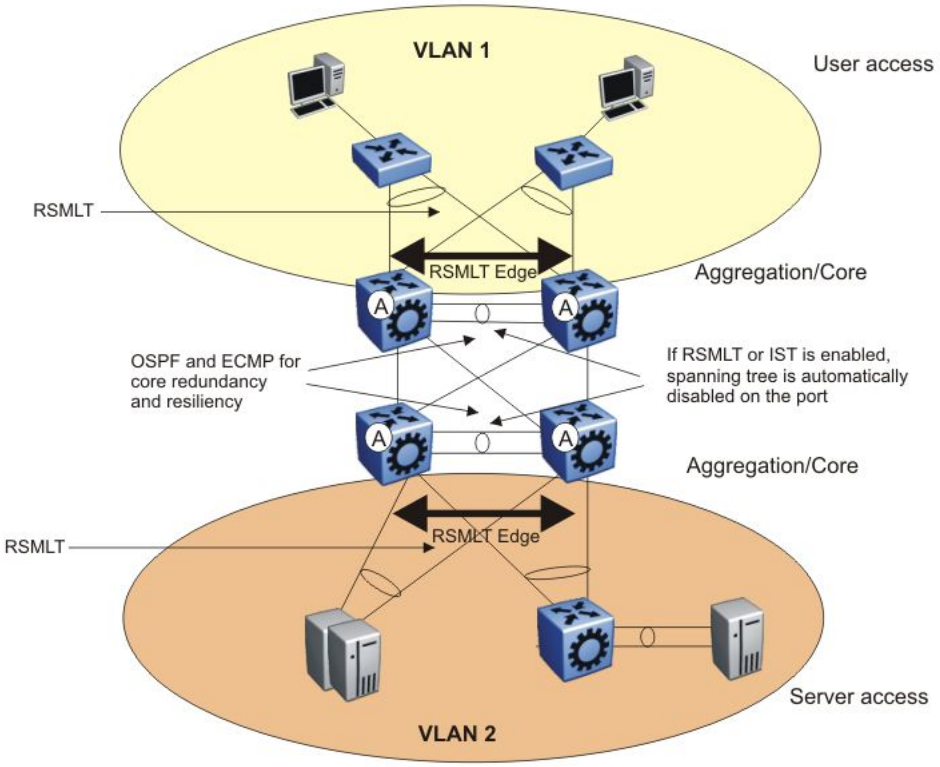
VLAN 1
User access
RSMLT
RSMLT Edge
Aggregation/Core
A
A
OSPF and ECMP for core redundancy and resiliency
If RSMLT or IST is enabled, spanning tree is automatically disabled on the port
A
A
Aggregation/Core
RSMLT
RSMLT Edge
Server access
VLAN 2

Figure 92: RSMLT Layer 2 Edge

The following figure uses RSMLT in a configuration with dual core VLANs to minimize traffic interruption in the event of losing the OSPF designated router (DR). This configuration creates a second OSPF core VLAN, forcing different nodes to become the DR for each VLAN. Each OSPF core VLAN has a DR (priority of 100) and no backup DRs (BDR). This configuration does not require a BDR because the two VLANs provide backup for each other from a routing perspective.

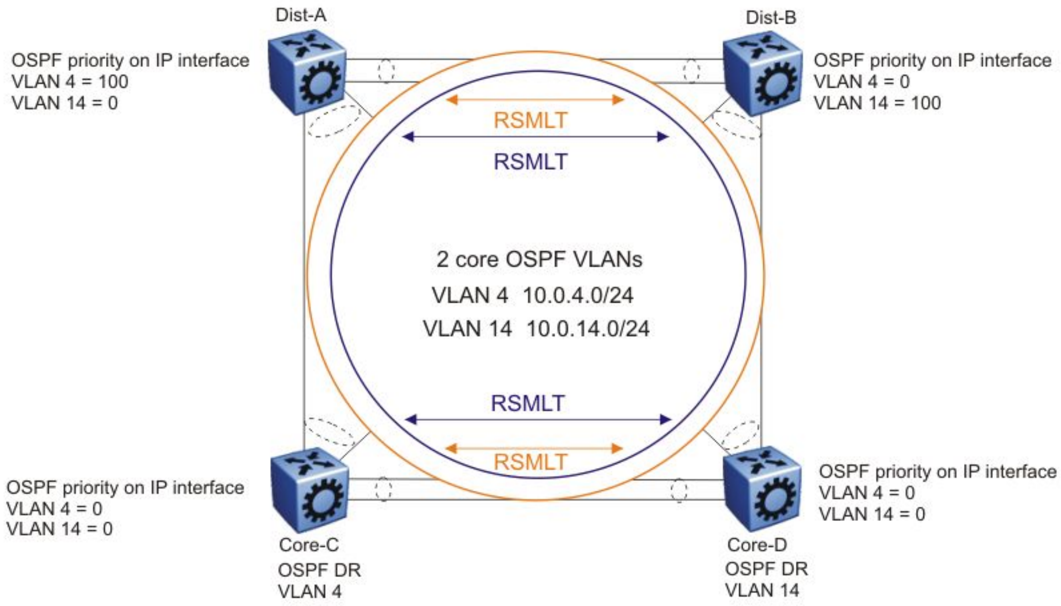

Figure 93: RSMLT with dual core VLANs

Chapter 19: Optical routing design

The Avaya optical routing system uses coarse wavelength division multiplexing (CWDM) in a grid of eight optical wavelengths. Use the Avaya optical routing system to maximize bandwidth on a single optical fiber. This section provides optical routing system information that you can use to help design your network.

Optical routing system components

Small form factor pluggable (SFP) transceivers transmit optical signals from Gigabit Ethernet (GbE) ports to multiplexers in a passive optical shelf.

Multiplexers combine multiple wavelengths traveling on different fibers onto a single fiber. At the receiver end of the link, demultiplexers separate the wavelengths and route them to different fibers, which terminate at separate CWDM devices. The following figure shows multiplexer and demultiplexer operations.

Important:

For clarity, the following figure shows a single fiber link with signals traveling in one direction only. A duplex connection requires communication in the reverse direction as well.

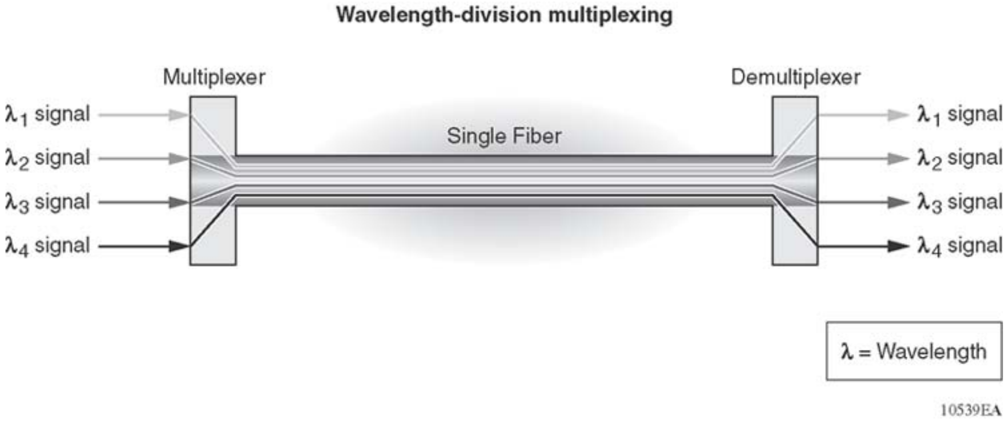

Figure 94: Wavelength division multiplexing

The Avaya optical routing system supports both ring and point-to-point configurations. The optical routing system includes the following parts:

- CWDM SFPs
- Optical add/drop multiplexers (OADM)

- Optical multiplexer/demultiplexers (OMUX)
- Optical shelf to house the multiplexers

OADMs drop or add a single wavelength from or to an optical fiber.

VSP 9000 supports the following CWDM SFPs:

- SFP 1-port 1000BaseCWDM SFP-LC 1470nm, 40km (17dB) AA1419053-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1490nm, 40km (17dB) AA1419054-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1510nm, 40km (17dB) AA1419055-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1530nm, 40km (17dB) AA1419056-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1550nm, 40km (17dB) AA1419057-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1570nm, 40km (17dB) AA1419058-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1590nm, 40km (17dB) AA1419059-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1610nm, 40km (17dB) AA1419060-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1470nm, 70km (24dB) AA1419061-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1490nm, 70km (24dB) AA1419062-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1510nm, 70km (24dB) AA1419063-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1530nm, 70km (24dB) AA1419064-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1550nm, 70km (24dB) AA1419065-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1570nm, 70km (24dB) AA1419066-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1590nm, 70km (24dB) AA1419067-E6
- SFP 1-port 1000BaseCWDM SFP-LC 1610nm, 70km (24dB) AA1419068-E6

For more information about SFPs, including technical specifications and installation instructions, see *Avaya Virtual Services Platform 9000 Installation — SFP Hardware Components*, NN46250-305.

Chapter 20: Software and hardware scaling capabilities

This chapter details the software and hardware scaling capabilities of Avaya Virtual Services Platform 9000. The information in *Avaya Virtual Services Platform 9000 Release Notes*, NN46250–401 takes precedence over information in other documents.

Hardware scaling capabilities

This section lists hardware scaling capabilities of Avaya Virtual Services Platform 9000.

Table 16: Hardware scaling capabilities

	Maximum number supported
9024XL I/O module	
10GbE fiber connections	240 (10 x 24)
Processor	1 GHz
9048GB I/O module	
GbE fiber connections	480 (10 x 48)
Processor	1 GHz
9048GT I/O module	
10/100/1000 copper connections	480 (10 x 48)
Processor	1 GHz
9080CP CP module	
Processor	1.33 GHz
Console port	1 D-subminiature 25-pin shell 9 pin connector (DB9) per CP module
Ethernet management	1 Registered Jack (RJ) 45 per CP module
USB port	1 Universal Serial Bus (USB) Type A (Master) per CP module
External Compact Flash	1 per CP module
9012 Chassis	

	Maximum number supported
Control Processor (CP) modules	2
Interface modules	10
Switch Fabric (SF) modules	6 You must install a minimum of 3 SF modules in the chassis.
Auxiliary slots	2
Power supplies	6
Total power capacity	• 10 kW in 220 V AC mode • 6 kW in 110 V AC mode
Jumbo packets	9600 bytes for IPv4 9500 bytes for IPv6

Software scaling capabilities

This section lists software scaling capabilities of Avaya Virtual Services Platform 9000.

Table 17: Software scaling capabilities

	Maximum number supported
Layer 2	
IEEE/Port-based VLANs	4,084
Inter-Switch Trunk (IST)	1 group
Internet Protocol (IP) Subnet-based VLANs	256
LACP	512 aggregators
LACP ports per aggregator	8 active and 8 standby
Lossless Ethernet	2 ports for each 8–port cluster 6 ports for each 9024XL module
MACs in forwarding database (FDB)	128K
Multi-Link Trunking (MLT)	512 groups
Multiple Spanning Tree Protocol (MSTP)	64 instances
Protocol-based VLANs	16
Rapid Spanning Tree Protocol (RSTP)	1 instance
SLPP	500 VLANs

	Maximum number supported
Source MAC-based VLANs	100
Split Multi-Link Trunking (SMLT)	511 groups
SMLT ports per group	16
VLACP Interfaces	128
Layer 3	
Address Resolution Protocol (ARP) for each port, VRF, or VLAN	64,000 entries total
BGP Internet peers (full)	3
BGP peers	256
BGP routes	1.5 million
BGP+ routes	128,000
Circuitless IP interfaces	256
ECMP routes	64,000
ECMP routes (fastpath)	8
FIB IPv4 routes	500,000
FIB IPv6 routes	128,000
The fastpath forwarding table uses a common table for IPv4 and IPv6 forwarding records. IPv6 records are approximately four times the size of IPv4 records. The maximum number of 500,000 IPv4 routes is possible when no IPv6 routes are configured, and the maximum number of IPv6 routes is 128,000 when no IPv4 routes are configured.	
IPv4 interfaces	4,343
IP interfaces (Brouter)	480
IP prefix entries	25 000
IPv4 prefix list	500
IP routing policies	500 for each VRF 5,000 for each system
IPFIX flows	96,000 for each interface module 960,000 for each chassis
IPv4 or IPv6 FTP sessions	4 each, 8 total
IPv4 or IPv6 Rlogin sessions	8 each, 16 total
IPv4 or IPv6 SSH sessions	8 total (any combination of IPv4 and IPv6 up to 8)
IPv4 or IPv6 Telnet sessions	8 each, 16 total
IPv4 VRF instances	512

	Maximum number supported
IPv6 dynamic neighbors/interface	64K
IPv6 interfaces	4,087 (4,084 VLAN and 3 management [1/1, 2/1, virtual IP])
IPv6 routes (fastpath)	128,000
IPv6 static neighbors	1,000
IPv6 static routes	10,000
IPv6 tunnels	2,000
Multicast IGMP interfaces	4,084
Multicast source and group (S, G)	6,000
NLB Clusters — Multicast, with multicast MAC flooding disabled	1 for each VLAN 2,000 for each system
NLB Clusters — Multicast, with multicast MAC flooding enabled	128 for each VLAN 2,000 for each system
NLB Clusters — Unicast	128 for each VLAN 2,000 for each system
OSPF adjacencies	512
OSPF areas	12 for each OSPF instance 80 for each system
OSPF instances	64 (one per VRF)
OSPF interfaces	512 active, 2000 passive
OSPF LSA packet size	Jumbo packets
OSPF routes	64,000
OSPFv3 adjacencies	512
OSPFv3 adjacencies per interface	256
OSPFv3 areas	64
OSPFv3 passive interfaces	1,000
OSPFv3 routers per area	250
OSPFv3 routes	64,000
PIM interfaces	512 active; 4084 passive
RIB IPv4 routes	3 * fastpath routes
RIP instances	64 (one for each VRF)
RIP interfaces	200
RIP routes	2,500 for each VRF

	Maximum number supported
	10,000 for each system
RSMLT interfaces (IPv4/IPv6)	4,000 over 512 SMLT interfaces
Static ARP entries	2,048 for each VRF 10,000 for each system
Static routes (IPv4)	2,000 for each VRF 10,000 total across VRFs
UDP/DHCP forwarding entries	512 for each VRF 1,024 for each system
VRRP interfaces (IPv4)	255 for a VRF 512 for a system
VRRP interfaces (IPv6)	255 for a system
VRRP interfaces fast timers (200ms)	24
Diagnostics	
Mirrored ports	479
Remote Mirroring Termination (RMT) ports	32
Filters and QoS	
Flow-based policers (IPv4 and IPv6)	16,000
Port shapers (IPv4 and IPv6)	480
Access control lists (ACL) for each chassis	2,048 The current release does not support IPv6 filters
Access control entries (ACE) for each chassis (IPv4)	16,000
ACEs per ACL (a combination of Security and QoS ACEs)	1,000
Unique redirect next hop values for ACE Actions (IPv4)	2,000

Chapter 21: Supported standards, request for comments, and Management Information Bases

This chapter details the standards, request for comments (RFC), and Management Information Bases (MIB) that Avaya Virtual Services Platform 9000 supports.

Supported standards

The following table details the standards that Avaya Virtual Services Platform 9000 supports.

Table 18: Supported standards

Standard	Description
802.1AX	Link Aggregation Control Protocol (LACP)
802.1p	VLAN prioritization
802.1Q	Virtual Local Area Network (VLAN) tagging
802.1Qbb	Virtual Bridged Local Area Networks - Amendment: Priority-based Flow Control
802.1s	Multiple Spanning Tree Protocol
802.1t	802.1D maintenance
802.1w-2001	Rapid Spanning Tree protocol (RSTP)
802.1X	Extended Authentication Protocol (EAP), and EAP over LAN (EAPoL)
802.1X-2004	Port Based Network Access Control
802.3 CSMA/CD Ethernet ISO/IEC 8802	International Organization for Standardization (ISO) /International Eletrotechnical Commission (IEC) 8802-3
802.3ab	Gigabit Ethernet 1000BaseT 4 pair Category 5 (Cat5) Unshieled Twisted Pair (UTP)

Standard	Description
802.3ae	10 Gigabit Ethernet
802.3an	10 Gigabit Copper
802.3i	10BaseT
802.3u	100BaseT
802.3x	flow control
802.3z	Gigabit Ethernet

Supported RFCs

The following table and sections list the RFCs that Avaya Virtual Services Platform 9000 supports.

Table 19: Supported request for comments

Request for comment	Description
RFC768	UDP Protocol
RFC783	Trivial File Transfer Protocol (TFTP)
RFC791	Internet Protocol (IP)
RFC792	Internet Control Message Protocol (ICMP)
RFC793	Transmission Control Protocol (TCP)
RFC826	Address Resolution Protocol (ARP)
RFC854	Telnet protocol
RFC894	A standard for the Transmission of IP Datagrams over Ethernet Networks
RFC896	Congestion control in IP/TCP internetworks
RFC903	Reverse ARP Protocol
RFC906	Bootstrap loading using TFTP
RFC950	Internet Standard Subnetting Procedure
RFC951	BootP
RFC959, RFC1350, and RFC2428	IPv6 FTP and TFTP client and server
RFC1027	Using ARP to implement transparent subnet gateways/Nortel Subnet based VLAN
RFC1058	RIPv1 Protocol

Request for comment	Description
RFC1112	IGMPv1
RFC1122	Requirements for Internet Hosts
RFC1253	OSPF
RFC1256	ICMP Router Discovery
RFC1258	IPv6 Rlogin server
RFC1305	Network Time Protocol v3 Specification, Implementation and Analysis
RFC1340	Assigned Numbers
RFC1519	Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy
RFC1541	Dynamic Host Configuration Protocol1
RFC1542	Clarifications and Extensions for the Bootstrap Protocol
RFC1583	OSPFv2
RFC1587	The OSPF NSSA Option
RFC1591	DNS Client
RFC1723	RIP v2 – Carrying Additional Information
RFC1745	BGP / OSPF Interaction
RFC1771 and RFC1772	BGP-4
RFC1812	Router requirements
RFC1866	HyperText Markup Language version 2 (HTMLv2) protocol
RFC1965	BGP-4 Confederations
RFC1966	BGP-4 Route Reflectors
RFC1981	Path MTU discovery
RFC1997	BGP-4 Community Attributes
RFC1998	An Application of the BGP Community Attribute in Multi-home Routing
RFC2068	Hypertext Transfer Protocol
RFC2131	Dynamic Host Control Protocol (DHCP)
RFC2138	RADIUS Authentication
RFC2139	RADIUS Accounting

Request for comment	Description
RFC2178	OSPF MD5 cryptographic authentication / OSPFv2
RFC2236	IGMPv2 for snooping
RFC2270	BGP-4 Dedicated AS for sites/single provide
RFC2328	OSPFv2
RFC2338	VRRP: Virtual Redundancy Router Protocol
RFC2362	PIM-SM
RFC2385	BGP-4 MD5 authentication
RFC2439	BGP-4 Route Flap Dampening
RFC2453	RIPv2 Protocol
RFC2460	IPv6 base stack
RFC2464	Transmission of IPv6 packets over Ethernet networks
RFC2545 and RFC4710	IPv6 capable BGPv4+
RFC2616	IPv6 HTTP server
RFC2710 and RFC3810	MLD (host-mode only)
RFC2740	OSPFv3
RFC2796	BGP Route Reflection – An Alternative to Full Mesh IBGP
RFC2819	RMON
RFC2918	Route Refresh Capability for BGP-4
RFC2992	Analysis of an Equal-Cost Multi-Path Algorithm
RFC3046	DHCP Option 82
RFC3065	Autonomous System Confederations for BGP
RFC3162	IPv6 RADIUS client
RFC3315	IPv6 DHCP Relay
RFC3376	Internet Group Management Protocol, v3
RFC3411 and RFC2418	SNMP over IPv6 networks
RFC3513	Internet Protocol Version 6 (IPv6) Addressing Architecture
RFC3569	An overview of Source-Specific Multicast (SSM)

Request for comment	Description
RFC3587	IPv6 Global Unicast Address Format
RFC3768 and draft-ietf-vrrp-ipv6-spec-08.txt	IPv6 capable VRRP
RFC4213	IPv6 configured tunnel
RFC4250–RFC4256	SSH server and client support
RFC4443	Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
RFC4861	IPv6 Neighbor discovery
RFC4862	IPv6 stateless address autoconfiguration
RFC4893	BGP support for Four-octet AS Number Space

Quality of service

Table 20: Supported request for comments

Request for comment	Description
RFC2474 and RFC2475	DiffServ Support
RFC2597	Assured Forwarding PHB Group
RFC2598	An Expedited Forwarding PHB

Network management

Table 21: Supported request for comments

Request for comment	Description
RFC1155	SMI
RFC1157	SNMP
RFC1215	Convention for defining traps for use with the SNMP
RFC1269	Definitions of Managed Objects for the Border Gateway Protocol: v3
RFC1271	Remote Network Monitoring Management Information Base

Request for comment	Description
RFC1305	Network Time Protocol v3 Specification, Implementation and Analysis3
RFC1350	The TFTP Protocol (Revision 2)
RFC1354	IP Forwarding Table MIB
RFC1389	RIP v2 MIB Extensions
RFC1757	Remote Network Monitoring Management Information Base
RFC1907	Management Information Base for Version 2 of the Simple Network Management Protocol (SNMPv2)
RFC1908	Coexistence between v1 & v2 of the Internet-standard Network Management Framework
RFC1930	Guidelines for creation, selection, and registration of an Autonomous System (AS)
RFC2541	Secure Shell Protocol Architecture
RFC2571	An Architecture for Describing SNMP Management Frameworks
RFC2572	Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)
RFC2573	SNMP Applications
RFC2574	User-based Security Model (USM) for v3 of the Simple Network Management Protocol (SNMPv3)
RFC2575	View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)
RFC2576	Coexistence between v1, v2, & v3 of the Internet standard Network Management Framework
RFC2819	Remote Network Monitoring Management Information Base

MIBs

Table 22: Supported request for comments

Request for comment	Description
RFC1156	MIB for network management of TCP/IP
RFC1212	Concise MIB definitions
RFC1213	TCP/IP Management Information Base
RFC1354	IP Forwarding Table MIB
RFC1389	RIP v2 MIB Extensions
RFC1398	Ethernet MIB
RFC1442	Structure of Management Information for version 2 of the Simple Network Management Protocol (SNMPv2)
RFC1450	Management Information Base for v2 of the Simple Network Management Protocol (SNMPv2)
RFC1573	Interface MIB
RFC1650	Definitions of Managed Objects for the Ethernet-like Interface Types
RFC1657	BGP-4 MIB using SMIv2
RFC1724	RIPv2 MIB extensions
RFC1850	OSPF MIB
RFC2021	RMON MIB using SMIv2
RFC2096	IP Forwarding Table MIB
RFC2452	IPv6 MIB: TCP MIB
RFC2454	IPv6 MIB: UDP MIB
RFC2466	IPv6 MIB: ICMPv6 Group
RFC2578	Structure of Management Information v2 (SMIv2)
RFC2674	Bridges with Traffic MIB
RFC2787	Definitions of Managed Objects for the Virtual Router Redundancy Protocol
RFC2863	Interface Group MIB

Request for comment	Description
RFC2925	Remote Ping, Traceroute & Lookup Operations MIB
RFC2932	IPv4 Multicast Routing MIB
RFC2933	IGMP MIB
RFC2934	PIM MIB
RFC3416	v2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)
RFC4022	Management Information Base for the Transmission Control Protocol (TCP)
RFC4113	Management Information Base for the User Datagram Protocol (UDP)

Standard MIBs

The following table details the standard MIBs that Avaya Virtual Services Platform 9000 supports.

Table 23: Supported MIBs

Standard MIB name	Institute of Electrical and Electronics Engineers/ Request for Comments (IEEE/RFC)	File name
STDMIB2— Link Aggregation Control Protocol (LACP) (802.3ad)	802.3ad	ieee802-lag.mib
STDMIB3—Exensible Authentication Protocol Over Local Area Networks (EAPoL) (802.1x)	802.1x	ieee8021x.mib
STDMIB4—Internet Assigned Numbers Authority (IANA) Interface Type	—	iana_if_type.mib
STDMIB5—Structure of Management Information (SMI)	RFC1155	rfc1155.mib
STDMIB6—Simple Network Management Protocol (SNMP)	RFC1157	rfc1157.mib

Standard MIB name	Institute of Electrical and Electronics Engineers/ Request for Comments (IEEE/RFC)	File name
STDMIB7—MIB for network management of Transfer Control Protocol/Internet Protocol (TCP/IP) based Internet MIB2	RFC1213	rfc1213.mib
STDMIB8—A convention for defining traps for use with SNMP	RFC1215	rfc1215.mib
STDMIB9—Routing Information Protocol (RIP) version 2 MIB extensions	RFC1389	rfc1389.mib
STDMIB10—Definitions of Managed Objects for Bridges	RFC1493	rfc1493.mib
STDMIB11—Evolution of the Interface Groups for MIB2	RFC2863	rfc2863.mib
STDMIB12—Definitions of Managed Objects for the Ethernet-like Interface Types	RFC1643	rfc1643.mib
STDMIB13—Definitions of Managed Objects for the Fourth Version of the Border Gateway Protocol (BGP-4) using SMIv2	RFC1657	rfc1657.mib
STDMIB14—RIP version 2 MIB extensions	RFC1724	rfc1724.mib
STDMIB15—Remote Network Monitoring (RMON)	RFC2819	rfc2819.mib
STDMIB16—Open Shortest Path First (OSPF) Version 2	RFC1850	rfc1850.mib
STDMIB17—Management Information Base of the Simple Network Management Protocol version 2 (SNMPv2)	RFC1907	rfc1907.mib
STDMIB21—Interfaces Group MIB using SMIv2	RFC2233	rfc2233.mib
STDMIB26a—An Architecture for Describing	RFC2571	rfc2571.mib

Standard MIB name	Institute of Electrical and Electronics Engineers/ Request for Comments (IEEE/RFC)	File name
SNMP Management Frameworks		
STDMIB26b—Message Processing and Dispatching for the SNMP	RFC2572	rfc2572.mib
STDMIB26c—SNMP Applications	RFC2573	rfc2573.mib
STDMIB26d—User-based Security Model (USM) for version 3 of the SNMP	RFC2574	rfc2574.mib
STDMIB26e—View-based Access Control Model (VACM) for the SNMP	RFC2575	rfc2575.mib
STDMIB26f —Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework	RFC2576	rfc2576.mib
STDMIB29—Definitions of Managed Objects for the Virtual Router Redundancy Protocol	RFC2787	rfc2787.mib
STDMIB31—Textual Conventions for Internet Network Addresses	RFC2851	rfc2851.mib
STDMIB32—The Interface Group MIB	RFC2863	rfc2863.mib
STDMIB33—Definitions of Managed Objects for Remote Ping, Traceroute, and Lookup Operations	RFC2925	rfc2925.mib
STDMIB34—IPv4 Multicast Routing MIB	RFC2932	rfc2932.mib
STDMIB35—Internet Group Management Protocol MIB	RFC2933	rfc2933.mib
STDMIB36—Protocol Independent Multicast MIB for IPv4	RFC2934, RFC2936	rfc2934.mib, rfc2936.mib

Standard MIB name	Institute of Electrical and Electronics Engineers/ Request for Comments (IEEE/RFC)	File name
STDMIB38—SNMPv3 These Request For Comments (RFC) make some previously named RFCs obsolete	RFC3411, RFC3412, RFC3413, RFC3414, RFC3415	rfc2571.mib, rfc2572.mib, rfc2573.mib, rfc2574.mib, rfc2575.mib
STDMIB39—Entity Sensor Management Information Base	RFC3433	
STDMIB40—The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model	RFC3826	rfc3826.mib
STDMIB41—Management Information Base for the Transmission Control protocol (TCP)	RFC4022	rfc4022.mib
STDMIB43—Management Information Base for the User Datagram Protocol (UDP)	RFC4113	rfc4113.mib
STDMIB44—Entity MIB	RFC4133	rfc4133.mib
STDMIB46—Definitions of Managed Objects for BGP-4	RFC4273	rfc4273.mib

Proprietary MIBs

The following table details the proprietary MIBs that Avaya Virtual Services Platform 9000 supports.

Table 24: Proprietary MIBs

Proprietary MIB name	File name
PROMIB1 - Rapid City MIB	rapid_city.mib
PROMIB 2 - SynOptics Root MIB	synro.mib
PROMIB3 - Other SynOptics definitions	s5114roo.mib
PROMIB4 - Other SynOptics definitions	s5tcs112.mib

Proprietary MIB name	File name
PROMIB5 - Other SynOptics definitions	s5emt103.mib
PROMIB6 - Avaya RSTP/MSTP proprietary MIBs	nnrst000.mib, nnmst000.mib
PROMIB7 - Avaya IGMP MIB	rfc_igmp.mib
PROMIB8 - MIAvayal IP Multicast MIB	ipmroute_rcc.mib
PROMIB9 - Avaya PIM MIB	pim-rcc.mib
PROMIB11 - Avaya MIB definitions	wf_com.mib

Chapter 22: Customer service

Visit the Avaya Web site to access the complete range of services and support that Avaya provides. Go to www.avaya.com or go to one of the pages listed in the following sections.

Getting technical documentation

To download and print selected technical publications and release notes directly from the Internet, go to www.avaya.com/support.

Getting product training

Ongoing product training is available. For more information or to register, you can access the Web site at www.avaya.com/support. From this Web site, you can locate the Training contacts link on the left-hand navigation pane.

Getting help from a distributor or reseller

If you purchased a service contract for your Avaya product from a distributor or authorized reseller, contact the technical support staff for that distributor or reseller for assistance.

Getting technical support from the Avaya Web site

The easiest and most effective way to get technical support for Avaya products is from the Avaya Technical Support Web site at www.avaya.com/support.

Glossary

bit error rate (BER) The ratio of the number of bit errors to the total number of bits transmitted in a given time interval.

Control Processor Unit High Availability (CPU-HA) Activates two CP modules simultaneously. The CP modules exchange topology data so, if a failure occurs, either CP module can take precedence in less than 1 second with the most recent topology data.

coarse wavelength division multiplexing (CWDM) A technology that uses multiple optical signals with different wavelengths to simultaneously transmit in the same direction over one fiber, and then separates by wavelength at the distant end.

dense wavelength division multiplexing (DWDM) A technology that uses many optical signals (16 or more) with different wavelengths to simultaneously transmit in the same direction across one fiber, and then separates by wavelength at the distant end.

Distributed MultiLink Trunking (DMLT) A point-to-point connection that aggregates similar ports from different modules to logically act like a single port, but with the aggregated bandwidth.

Global routing engine (GRE) The base router or routing instance 0 in the Virtual Routing and Forwarding (VRF).

internal spanning tree (IST) Runs in a given multiple spanning tree (MST) region. Within an MST region, you can configure multiple spanning instances. Instance 0 within a region is the IST.

Internet Protocol security (IPsec) A secure version of the Internet Protocol (IP) that provides optional authentication and encryption at the packet level.

jitter The delay variance between received packets. Packets may not arrive at the destination address in consecutive order, or on a timely basis, and the signal can vary from its original reference timing. This distortion damages multimedia traffic.

last member query interval (LMQI) The time between when the last IGMP member leaves the group and the stream stops.

latency The time between when a node sends a message and receipt of the message by another node; also referred to as propagation delay.

Layer 1

The Physical Layer of the Open System Interconnection (OSI) model. Layer 1 interfaces with the MAC sublayer of Layer 2, and performs character encoding, transmission, reception, and character decoding.

Layer 2

The Data Link Layer of the OSI model. Examples of Layer 2 protocols are Ethernet and Frame Relay.

Layer 3

The Network Layer of the OSI model. Example of a Layer 3 protocol is Internet Protocol (IP).

Layer 4

The Transport Layer of the OSI model. An example of a Layer 4 protocol is Transfer Control Protocol (TCP).

link-state advertisement (LSA)

Packets that contain state information about directly connected links (interfaces) and adjacencies. Each Open Shortest Path First (OSPF) router generates the packets.

link-state database (LSDB)

A database built by each OSPF router to store LSA information. The router uses the LSDB to calculate the shortest path to each destination in the Autonomous System (AS), with itself at the root of each path.

Link Aggregation Control Protocol (LACP)

A protocol that exists between two endpoints to bundle links into an aggregated link group for bandwidth increase and link redundancy.

Link Aggregation Control Protocol Data Units (LACPDU)

Link aggregation control protocol data unit (LACPDU) is used for exchanging information among LACP-enabled devices.

link aggregation group (LAG)

A group that increases the link speed beyond the limits of one single cable or port, and also increases the redundancy for higher availability.

load balancing

The practice of splitting communication into two (or more) routes or servers.

management information base (MIB)

Defines system operations and parameters used for the Simple Network Management Protocol (SNMP).

multicast router discovery (MRDISC)

Provides the automatic discovery of multicast capable routers. By listening to multicast router discovery messages, Layer 2 devices can determine where to send multicast source data and Internet Group Management Protocol (IGMP) host membership reports.

multihomed AS

An AS that has multiple connections to one or more ASs and does not carry transit traffic.

multiplexing

Carriage of multiple channels over a single transmission medium; a process where a dedicated circuit is shared by multiple users. Typically, data streams intersperse on a bit or byte basis (time division), or separate by different carrier frequencies (frequency division).

next hop

The next hop to which a packet can be sent to advance the packet to the destination.

not so stubby area (NSSA)

Prevents the flooding of external link-state advertisements (LSA) into the area by providing them with a default route. An NSSA is a configuration of the Open Shortest Path First (OSPF) protocol.

out of band (OOB)

Network dedicated for management access to chassis.

packet loss

Expressed as a percentage of packets dropped over a specified interval. Keep packet loss to a minimum to deliver effective IP telephony and IP video services.

policing

Ensures that a traffic stream follows the domain service provisioning policy or service level agreement (SLA).

Protocol Data Units (PDUs)

A unit of data that is specified in a protocol of a given layer and that consists of protocol-control information of the given layer and possibly user data of that layer.

Protocol Independent Multicast, Source Specific (PIM-SSM)

Uses only shortest-path trees to provide multicast services based on subscription to a particular (source, group) channel.

Protocol Independent Multicast, Sparse Mode (PIM-SM)

Adds a Rendezvous Point router to avoid multicast-data flooding. Use PIM-SM when receivers for multicast data are sparsely distributed throughout the network.

quality of service (QoS)

Use QoS features to reserve resources in a congested network. For example, you can configure a higher priority to IP deskphones, which need a fixed bit rate, and, split the remaining bandwidth between data connections if calls in the network are important than the file transfers.

Read Write All (RWA)

An access class that lets users access all menu items and editable fields.

remote login (rlogin)

An application that provides a terminal interface between hosts (usually UNIX) that use the TCP/IP network protocol. Unlike Telnet, rlogin assumes the remote host is, or behaves like, a UNIX host.

remote monitoring (RMON)

A remote monitoring standard for Simple Network Management Protocol (SNMP)-based management information bases (MIB). The Internetwork

Engineering Task Force (IETF) proposed the RMON standard to provide guidelines for remote monitoring of individual LAN segments.

resilient packet ring (RPR)
A shared packet edge ring connection, where both paths around the ring carry traffic, that allows double bandwidth on each ring.

reverse path forwarding (RPF)
Prevents a packet from forging its source IP address. Typically, the system examines and validates the source address of each packet.

route flapping
An instability that is associated with a prefix, where the associated prefix routes can exhibit frequent changes in availability over a period of time.

routing policy
A form of routing that is influenced by factors other than the default algorithmically best route, such as the shortest or quickest path.

Secure Copy (SCP)
Securely transfers files between the switch and a remote station.

Secure Shell (SSH)
Used for secure remote logons and data transfer over the Internet. SSH uses encryption to provide security.

Secure Sockets Layer (SSL)
An Internet security encryption and authentication protocol for secure point-to-point connections over the Internet and intranets, especially between clients and servers.

Service Instance Identifier (I-SID)
The SPBM B-MAC header includes a Service Instance Identifier (I-SID) with a length of 24 bits. This I-SID can be used to identify and transmit any virtualized traffic in an encapsulated SPBM frame. These I-SIDs are used to virtualize VLANs (Layer 2 Virtual Services Network [VSN]) or VRFs (Layer 3 Virtual Services Network [VSN]) across the MAC-in-MAC backbone. With Layer 2 VSN, the I-SID is associated with a customer VLAN, which is then virtualized across the backbone. With Layer 3 VSN, the I-SID is associated with a customer VRF, which is also virtualized across the backbone.

service level agreement (SLA)
A service contract that specifies the forwarding service that traffic receives.

shortest path tree (SPT)
Creates a direct route between the receiver and the source for group members in a Protocol Independent Multicast—Spare Mode (PIM-SM) domain.

Simple Loop Prevention Protocol (SLPP)
Simple Hello Protocol that prevents loops in a Layer 2 network (VLAN).

single mode fiber (SMF)
One of the various light waves transmitted in an optical fiber. Each optical signal generates many modes, but in single-mode fiber only one mode is transmitted. Transmission occurs through a small diameter core (approximately ten micrometers), with a cladding that is 10 times the core

diameter. These fibers have a potential bandwidth of 50 to 100 GHz per kilometer.

small form factor pluggable (SFP)

A hot-swappable input and output enhancement component used with Avaya products to allow gigabit Ethernet ports to link with other gigabit Ethernet ports over various media types.

small form factor pluggable plus (SFP+)

SFP+ transceivers are similar to SFPs in physical appearance but SFP + transceivers provide Ethernet at 10 gigabit per second (Gb/s).

SMLT aggregation switch

One of two IST peer switches that form a split link aggregation group. It connects to multiple wiring closet switches, edge switches, or customer premise equipment (CPE) devices.

SMLT client

A switch located at the edge of the network, such as in a wiring closet or CPE connecting to two SMLT aggregation switches. An SMLT client switch performs link aggregation but does not require Split MultiLink Trunking (SMLT) intelligence.

spanning tree

A simple, fully-connected active topology formed from the arbitrary physical topology of connected bridged Local Area Network components by relaying frames through selected bridge ports. The protocol parameters and states that are used and exchanged to facilitate the calculation of the active topology and to control the bridge relay function.

Spanning Tree Group (STG)

A collection of ports in one spanning tree instance.

Switch Fabric (SF) cooling module

The SF cooling module is a hot swappable fan tray used to cool the Switch Fabric (SF) modules.

Switch Fabric (SF) module

The Switch Fabric module connects to all interface and Control Processor modules. You can install 6 SF modules in Virtual Services Platform 9000, using 5 SF modules plus 1 as a hot backup. The SF modules comprise a data path and a control path, and provide a back end switching solution in the midplane chassis.

SysAdmin, Audit, Network, Security (SANS) Institute

The research and education organization for network administrators and security professionals.

time-to-live (TTL)

The field in a packet used to determine the valid duration for the packet; the TTL determines the packet lifetime. The system discards a packet with a TTL of zero.

traffic engineering

A method that guarantees performance in a network.

Trivial File Transfer Protocol (TFTP)
A protocol that governs transferring files between nodes without protection against packet loss.

trunk
A logical group of ports that behaves like a single large port.

unshielded twisted pair (UTP)
A cable with one or more pairs of twisted insulated copper conductors bound in a single plastic sheath.

user-based security model (USM)
A security model that uses a defined set of user identities for authorized users on a particular Simple Network Management Protocol (SNMP) engine.

User Datagram Protocol (UDP)
In TCP/IP, a packet-level protocol built directly on the Internet Protocol layer. TCP/IP host systems use UDP for application-to-application programs.

view-based access control model (VACM)
Provides context, group access, and group security levels based on a predefined subset of management information base (MIB) objects.

Virtual Link Aggregation Control Protocol (VLACP)
Virtual Link Aggregation Control Protocol (VLACP) is a Layer 2 handshaking protocol that can detect end-to-end failure between two physical Ethernet interfaces.

Virtual Router Redundancy Protocol (VRRP)
A protocol used in static routing configurations, typically at the edge of the network. This protocol operates on multiple routers on an IP subnet and elects a primary gateway router. When the primary router fails, a backup router is quickly available to take its place.

Voice over IP (VOIP)
The technology that delivers voice information in digital form in discrete packets using the Internet Protocol (IP) rather than the traditional circuit-committed protocols of the public switched telephone network (PSTN).

wavelength division multiplexing (WDM)
Simultaneously transmits many colors (wavelengths) of laser light down the same optical fiber to increase the amount of transferred information.

wiring closet
A central termination area for telephone or network cabling or both.