Network Design Reference for Avaya Virtual Services Platform 9000

Release 4.0
NN46250-200
Issue 06.01
December 2014

© 2014 Avaya Inc.

All Rights Reserved.

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 may generally make available to users of its products and Hosted Services. 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 websites 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 Avaya hardware and software. 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 website: http://support.avaya.com or such successor site as designated by Avaya. Please note that if You acquired the product(s) from an authorized Avaya Channel Partner outside of the United States and Canada, the warranty is provided to You by said Avaya Channel Partner and not by Avaya.

Licenses

THE SOFTWARE LICENSE TERMS AVAILABLE ON THE AVAYA WEBSITE, HTTP://SUPPORT.AVAYA.COM/LICENSEINFO OR SUCH SUCCESSOR SITE AS DESIGNATED BY AVAYA, ARE APPLICABLE TO ANYONE WHO DOWNLOADS, USES AND/OR INSTALLS AVAYA SOFTWARE, PURCHASED FROM AVAYA INC., ANY AVAYA AFFILIATE, OR AN AVAYA CHANNEL PARTNER (AS APPLICABLE) UNDER A COMMERCIAL AGREEMENT WITH AVAYA OR AN AVAYA CHANNEL PARTNER. UNLESS OTHERWISE AGREED TO BY AVAYA IN WRITING, AVAYA DOES NOT EXTEND THIS LICENSE IF THE SOFTWARE WAS OBTAINED FROM ANYONE OTHER THAN AVAYA, AN AVAYA AFFILIATE OR AN AVAYA CHANNEL PARTNER; AVAYA RESERVES THE RIGHT TO TAKE LEGAL ACTION AGAINST YOU AND ANYONE ELSE USING OR SELLING THE SOFTWARE WITHOUT A LICENSE. BY INSTALLING, DOWNLOADING OR USING THE SOFTWARE, OR AUTHORIZING OTHERS TO DO SO, YOU, ON BEHALF OF YOURSELF AND THE ENTITY FOR WHOM YOU ARE INSTALLING, DOWNLOADING OR USING THE SOFTWARE (HEREINAFTER REFERRED TO INTERCHANGEABLY AS “YOU” AND “END USER”), AGREE TO THESE TERMS AND CONDITIONS AND CREATE A BINDING CONTRACT BETWEEN YOU AND AVAYA INC. OR THE APPLICABLE AVAYA AFFILIATE (“AVAYA”).

Avaya grants You a license within the scope of the license types described below, with the exception of Heritage Nortel Software, for which the scope of the license is detailed below. Where the order documentation does not expressly identify a license type, the applicable license will be a Designated System License. The applicable number of licenses and units of capacity for which the license is granted will be one (1), unless a different number of licenses or units of capacity is specified in the documentation or other materials available to You. “Software” means computer programs in object code, provided by Avaya or an Avaya Channel Partner, whether as stand-alone products, pre-installed on hardware products, and any upgrades, updates, patches, bug fixes, or modified versions thereto. “Designated Processor” means a single stand-alone computing device. “Server” means a Designated Processor that hosts a software application to be accessed by multiple users. “Instance” means a single copy of the Software executing at a particular time: (i) on one physical machine; or (ii) on one deployed software virtual machine (“VM”) or similar deployment.

License types

Designated System(s) License (DS). End User may install and use each copy or an Instance of the Software only on a number of Designated Processors up to the number indicated in the order. Avaya may require the Designated Processor(s) to be identified in the order by type, serial number, feature key, Instance, location or other specific designation, or to be provided by End User to Avaya through electronic means established by Avaya specifically for this purpose.

Heritage Nortel Software

“Heritage Nortel Software” means the software that was acquired by Avaya as part of its purchase of the Nortel Enterprise Solutions Business in December 2009. The Heritage Nortel Software currently available for license from Avaya is the software contained within the list of Heritage Nortel Products located at http://support.avaya.com/LicenseInfo under the link “Heritage Nortel Products” or such successor site as designated by Avaya. For Heritage Nortel Software, Avaya grants You a license to use Heritage Nortel Software provided hereunder solely to the extent of the authorized activation or authorized usage level, solely for the purpose specified in the Documentation, and solely as embedded in, for execution on, or for communication with Avaya equipment. Charges for Heritage Nortel Software may be based on extent of activation or use authorized as specified in an order or invoice.

Copyright

Except where expressly stated otherwise, no use should be made of materials on this site, the Documentation, Software, Hosted Service, or hardware provided by Avaya. All content on this site, the documentation, Hosted Service, and the product provided by Avaya including the selection, arrangement and design of the content is owned either by Avaya or its licensors and is protected by copyright and other intellectual property laws including the sui generis rights relating to the protection of databases. You may not modify, copy, reproduce, republish, upload, post, transmit or distribute in any way any content, in whole or in part, including any code and software unless expressly authorized by Avaya. Unauthorized reproduction, transmission, dissemination, storage, and or use without the express written consent of Avaya can be a criminal, as well as a civil offense under the applicable law.

Third Party Components

“Third Party Components” mean certain software programs or portions thereof included in the Software or Hosted Service may contain software (including open source software) distributed under third party agreements (“Third Party Components”), which contain terms regarding the rights to use certain portions of the Software (“Third Party Terms”). As required, information regarding distributed Linux OS source code (for those products that have distributed Linux OS source code) and identifying the copyright holders of the Third Party Components and the Third Party Terms that apply is available in the products, Documentation or on Avaya’s website at: http://support.avaya.com/Copyright or such successor site as designated by Avaya. You agree to the Third Party Terms for any such Third Party Components.

Preventing Toll Fraud

“Toll Fraud” is the unauthorized use of your telecommunications system by an unauthorized party (for example, a person who is not a corporate employee, agent, subcontractor, or is not working on your company's behalf). Be aware that there can be a risk of Toll Fraud associated with your system and that, if Toll Fraud occurs, it can

result in substantial additional charges for your telecommunications services.

Avaya Toll Fraud intervention

If You suspect that You are being victimized by Toll Fraud and You need technical assistance or support, call Technical Service Center Toll Fraud Intervention Hotline at +1-800-643-2353 for the United States and Canada. For additional support telephone numbers, see the Avaya Support website: http://support.avaya.com or such successor site as designated by Avaya. Suspected security vulnerabilities with Avaya products should be reported to Avaya by sending mail to: securityalerts@avaya.com.

Downloading Documentation

For the most current versions of Documentation, see the Avaya Support website: http://support.avaya.com, or such successor site as designated by Avaya.

Contact Avaya Support

See the Avaya Support website: http://support.avaya.com for product or Hosted Service notices and articles, or to report a problem with your Avaya product or Hosted Service. For a list of support telephone numbers and contact addresses, go to the Avaya Support website: http://support.avaya.com (or such successor site as designated by Avaya), scroll to the bottom of the page, and select Contact Avaya Support.

Trademarks

The trademarks, logos and service marks (“Marks”) displayed in this site, the Documentation, Hosted Service(s), 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, Hosted Service(s) 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. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries.

Contents

Chapter 1: Introduction 8
Purpose 8
Related resources 8
Documentation 8
Training 8
Viewing Avaya Mentor videos 8
Support 9
Searching a documentation collection 9
Chapter 2: New in this release 11
Features 11
Other changes 12
Chapter 3: Network design fundamentals 14
Chapter 4: Hardware fundamentals and guidelines 16
Chassis considerations 16
Chassis power 16
Chassis cooling 17
Modules 20
Switch Fabric modules 20
Control Processor module 20
Input/Output modules 21
Optical device guidelines 22
1000BASE-X and 10GBASE-R reach 22
Dispersion considerations for long reach 23
10/100BASE-X and 1000BASE-TX reach 24
10/100/1000BASE-TX Auto-Negotiation recommendations 24
Auto MDIX 25
CANA 25
Chapter 5: Platform redundancy 26
Power redundancy 26
Input/output port redundancy 27
Control plane redundancy 27
Switch Fabric redundancy 27
Configuration redundancy 28
Link redundancy 28
Switch redundancy 28
High Availability mode 29
Chapter 6: Link redundancy 32
Physical layer redundancy 32
Multilink Trunking 36

802.3ad-based link aggregation ... 37
Chapter 7: Redundant network design ... 39
Network Load Balancing ... 39
Network Load Balancing topologies ... 40
Chapter 8: Lossless networks ... 44
General configuration limitations ... 45
Interface module configuration limitations ... 45
PFC configuration limitations ... 47
MLT configuration limitations ... 47
Lossless deployments ... 48
Chapter 9: Layer 2 switch clustering and SMLT ... 53
Modular design for redundant networks ... 53
Network edge redundancy ... 57
Split MultiLink Trunk configuration ... 57
SMLT full-mesh recommendations with OSPF ... 66
Chapter 10: Layer 3 switch clustering and RSMLT ... 67
Routed SMLT ... 67
Switch clustering topologies and interoperability with other products ... 73
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 ... 81
Chapter 12: Layer 2 loop prevention ... 83
Loop prevention and detection ... 83
CPU protection and loop prevention compatibility ... 88
Chapter 13: Spanning tree ... 89
Spanning tree and protection against isolated VLANs ... 89
MSTP and RSTP considerations ... 90
Chapter 14: Layer 3 network design ... 92
VRF Lite ... 92
Virtual Router Redundancy Protocol ... 96
Subnet-based VLAN guidelines ... 103
Open Shortest Path First ... 104
Border Gateway Protocol ... 108
IPv6 BGP+ ... 112
IP routed interface scaling ... 113
IPv6 ... 113
Chapter 15: SPBM design guidelines ... 117
802.1aq standard ... 117
VLANs without member ports ... 119
Provisioning ... 120

Implementation options ... 121
Reference architectures ... 127
Best practices ... 140
Migration best practices ... 142
SPBM restrictions and limitations ... 144
IP multicast over SPBM restrictions ... 146
Chapter 16: IP multicast network design ... 148
Multicast and VRF-lite ... 148
Multicast and MultiLink Trunking considerations ... 149
Multicast scalability design rules ... 149
IP multicast address range restrictions ... 151
Multicast MAC address mapping considerations ... 151
Dynamic multicast configuration changes ... 153
IGMPv3 backward compatibility ... 153
IGMP Layer 2 Querier ... 154
TTL in IP multicast packets ... 154
Multicast MAC filtering ... 155
Guidelines for multicast access policies ... 155
Split-subnet and multicast ... 156
Protocol Independent Multicast-Sparse Mode guidelines ... 156
Protocol Independent Multicast-Source Specific Multicast guidelines ... 168
IGMP and PIM-SM interaction ... 168
Layer 3 multicast extended VLANs ... 169
Multicast for multimedia ... 171
Chapter 17: System and network stability and security ... 174
Control plane rate limit (CP-Limit) ... 174
DoS protection mechanisms ... 175
Damage prevention ... 176
Security and redundancy ... 178
Data plane security ... 178
Control plane security ... 182
Additional information ... 188
Chapter 18: QoS design guidelines ... 189
QoS mechanisms ... 189
QoS interface considerations ... 192
Network congestion and QoS design ... 193
QoS examples and recommendations ... 194
Chapter 19: Layer 1, 2, and 3 design examples ... 198
Layer 1 examples ... 198
Layer 2 examples ... 201
Layer 3 examples ... 205
RSMLT redundant network with bridged and routed VLANs in the core ... 209
Chapter 20: Optical routing design ... 213

Optical routing system components 213
Chapter 21: Software and hardware scaling capabilities 215
Hardware scaling capabilities 215
Software scaling capabilities 217
Chapter 22: Supported standards, request for comments, and Management Information Bases 222
Supported standards 222
Supported RFCs 223
Quality of service 226
Network management 226
MIBs 227
Standard MIBs 228
Proprietary MIBs 231
Glossary 232

Chapter 1: Introduction

Purpose

This document 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.

Related resources

Documentation

See *Documentation Reference for Avaya Virtual Services Platform 9000,* NN46250-100 for a list of the documentation for this product.

Training

Ongoing product training is available. For more information or to register, you can access the website at http://avaya-learning.com/.

Course code	Course title
4D00010E	Knowledge Access: ACIS - Avaya ERS 8000 and VSP 9000 Implementation
5D00040E	Knowledge Access: ACSS - Avaya VSP 9000 Support

Viewing Avaya Mentor videos

Avaya Mentor videos provide technical content on how to install, configure, and troubleshoot Avaya products.

About this task

Videos are available on the Avaya Support website, listed under the video document type, and on the Avaya-run channel on YouTube.

Procedure

- To find videos on the Avaya Support website, go to http://support.avaya.com and perform one of the following actions:
  - In **Search**, type `Avaya Mentor Videos` to see a list of the available videos.
  - In **Search**, type the product name. On the Search Results page, select **Video** in the **Content Type** column on the left.
- To find the Avaya Mentor videos on YouTube, go to www.youtube.com/AvayaMentor and perform one of the following actions:
  - Enter a key word or key words in the Search Channel to search for a specific product or topic.
  - Scroll down Playlists, and click the name of a topic to see the available list of videos posted on the website.

Note:

Videos are not available for all products.

Support

Go to the Avaya Support website at http://support.avaya.com for the most up-to-date documentation, product notices, and knowledge articles. You can also search for release notes, downloads, and resolutions to issues. Use the online service request system to create a service request. Chat with live agents to get answers to questions, or request an agent to connect you to a support team if an issue requires additional expertise.

Searching a documentation collection

On the Avaya Support website, you can download the documentation library for a specific product and software release to perform searches across an entire document collection. For example, you can perform a single, simultaneous search across the collection to quickly find all occurrences of a particular feature. Use this procedure to perform an index search of your documentation collection.

Before you begin

- Download the documentation collection zip file to your local computer.
- You must have Adobe Acrobat or Adobe Reader installed on your computer.

Procedure

1. Extract the document collection zip file into a folder.

2. Navigate to the folder that contains the extracted files and open the file named <*product_name_release*>.pdx.
3. In the Search dialog box, select the option **In the index named <*product_name_release*>.pdx**.
4. Enter a search word or phrase.
5. Select any of the following to narrow your search:
   - Whole Words Only
   - Case-Sensitive
   - Include Bookmarks
   - Include Comments
6. Click **Search**.

   The search results show the number of documents and instances found. You can sort the search results by Relevance Ranking, Date Modified, Filename, or Location. The default is Relevance Ranking.

Chapter 2: New in this release

The following sections detail what is new in *Network Design Reference for Avaya Virtual Services Platform 9000,* NN46250-200, for Release 4.0.

Features

See the following sections for information about feature changes.

9012QQ-2 I/O module

Release 4.0.1 introduces a second generation 9012QQ-2 Input/Output (I/O) module.

The 9012QQ-2 module is a 12-port 40-gigabit-per-second (Gbps) module that supports QSFP+ fiber transceivers, and also directly-attached cables (DACs).

For more information about the 9012QQ-2 I/O module, see:

- Chassis power on page 16.
- Input/Output modules on page 21.
- Hardware scaling capabilities on page 215.
- SPBM restrictions and limitations on page 144.

For more information about the 9012QQ-2 module specifications, see *Installing Modules in Avaya Virtual Services Platform 9000,* NN46250-301.

Access-level and access-strict

If you configure an access policy mode to deny, the system does not check **`accesslevel`** or **`access-strict`** information. With the access policy mode configured to deny, the system checks the mode and service, and if they match the system denies the connection. If you configure the access policy mode to allow, the system continues to check the **`accesslevel`** and **`access-strict`** information. For more information, see Data plane security on page 178.

First and second generation modes

Release 4.0 introduces a second generation mode for Virtual Services Platform 9010 or Virtual Services Platform 9012 to support second generation modules. You can use second generation I/O modules in first generation mode or second generation mode. If you use second generation modules for Virtual Services Platform 9012, you must install two 9012FCHS cooling modules. You can achieve increased scaling capabilities for IPv4 FIB routes and IPv6 FIB routes if you use second generation modules in second generation mode.

For more information about first and second generation mode, see:

- Virtual Services Platform 9012 cooling on page 18.
- Switch Fabric modules on page 20.
- Input/Output modules on page 21.
- Switch Fabric redundancy on page 27.
- Hardware scaling capabilities on page 215.
- Software scaling capabilities on page 217

For more information about how to configure the minimum module generation support, see *Configuring Ethernet Modules on Avaya Virtual Services Platform 9000,* NN46250-508.

9048XS-2 I/O module

Release 4.0 introduces a second generation 9048XS-2 I/O module.

The 9048XS-2 module is a 48-port 10-gigabit-per-second (Gbps) module that supports the 40GBASE-R transceivers.

For more information about the 9048XS-2 I/O module, see

- Chassis power on page 16.
- Input/Output modules on page 21.
- Hardware scaling capabilities on page 215.

For more information about the 9048XS-2 module specifications, see *Installing Modules in Avaya Virtual Services Platform 9000,* NN46250-301.

Virtual Services Platform 9012 cooling module

This document is updated to include information about the 9012FCHS high-speed cooling module for use with Virtual Services Platform 9012.

For more information about the 9012FCHS cooling module, see:

- Virtual Services Platform 9012 cooling on page 18.
- Virtual Services Platform 9012 chassis power considerations on page 17.

For more information about how to install the 9012FCHS cooling module, see *Installing Cooling Modules in Avaya Virtual Services Platform 9000,* NN46250-302.

VLANs without member ports

VLANs without member ports on page 119 is added to the document to explain how Ethernet Routing Switch 8800, Virtual Services Platform 9000, and Virtual Services Platform 4000 manage VLANs without member ports differently.

Other changes

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

Document title change

In Release 4.0, the title of this document changed from *Avaya Virtual Services Platform 9000 Planning and Engineering — Network Design*, NN46250-200 to *Network Design Reference for Avaya Virtual Services Platform 9000,* NN46250-200.

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.

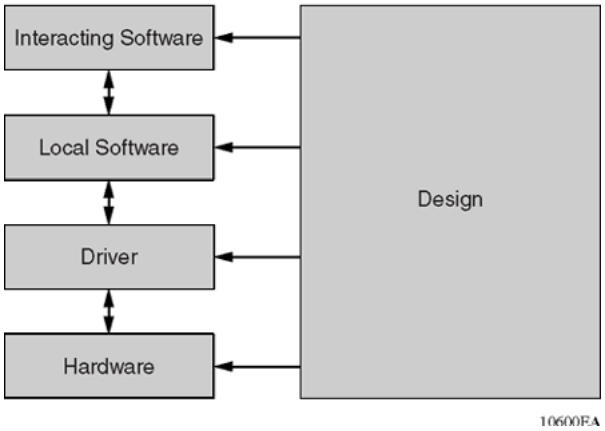


Figure 1: Hardware and software interaction

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

This section identifies the Virtual Services Platform 9010 and Virtual Services Platform 9012 power requirements for chassis components.

Virtual Services Platform 9010 chassis power considerations

The Virtual Services Platform 9010 AC chassis supports up to eight 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.

Power requirements

The following table describes the power requirements for the Virtual Services Platform 9010 chassis components.

Table 1: Component power draw for the VSP 9010 chassis

Component	Module type	Power required (Watt)
9080CP	Control Processor module	80 W
9095SF	Switch Fabric module	70 W (slots SF1 and SF4) 50 W (slots SF2, SF3, SF5, and SF6)
9010CM	Cooling module	1,800 W for two cooling modules at full speed
9048GT	48-port 10/100/1000BASE-T module	350 W

Component	Module type	Power required (Watt)
9048GB	48-port 1000BASE-X SFP module	340 W, with short range SFP
9024XL	24-port 10GBASE-R SFP+ module	575 W, with short range SFP+
9048XS-2	48-port 10GBASE-R SFP+ and 1000BASE-X SFP I/O module	600 W, with short range SFP+
9012QQ-2	12-port 40GBASE-R QSFP+ I/O module	600 W, with short range QSFP+

Virtual Services Platform 9012 chassis power considerations

Virtual Services Platform 9012 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.

Power requirements

The following table describes the power requirements for Virtual Services Platform 9012 chassis components.

Table 2: Component power draw for the VSP 9012 chassis

Component	Module type	Power required (Watt)
9080CP	Control Processor module	80 W
9090SF	Switch Fabric (SF) module	70 W (slots SF1 and SF4) 50 W (slots SF2, SF3, SF5, and SF6)
9012RC	SF cooling module	65 W
9012FC	Input/Output (I/O) cooling module	150 W
9012FCHS	I/O cooling module	150 W
9048GT	48-port 10/100/1000BASE-T module	350 W
9048GB	48-port 1000BASE-X SFP module	340 W, with short range SFP
9024XL	24-port 10GBASE-R SFP+ module	575 W, with short range SFP+
9048XS-2	48-port 10GBASE-R SFP+ and 1000BASE-X SFP I/O module	600 W, with short range SFP+
9012QQ-2	12-port 40GBASE-R QSFP+ I/O module	600 W, with short range QSFP+

Chassis cooling

This section identifies the types of cooling modules required for the Virtual Services Platform 9010 or Virtual Services Platform 9012 when the chassis is configured in first generation mode or second generation mode.

Virtual Services Platform 9010 cooling

You can configure your Virtual Services Platform 9010 to operate in first generation mode or second generation mode. Both modes support the same cooling module type for the Virtual Services

Platform 9010 chassis. You must install two 9010CM cooling modules in the chassis. Both slots are located at the front of the chassis and are installed horizontally. Each cooling module includes four fans, and provides cooling from front-to-back for the Input Output (I/O) , the Control Processor (CP), and the Switch Fabric (SF) modules.

The following messages are associated with the Virtual Services Platform 9010:

Danger:

Remove the cooling modules using the provided handles. Be careful to keep your fingers out of the fan blades. When you remove a cooling module, allow time for the fans to spin down before you fully withdraw the cooling module.

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.

Important:

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.

Caution:

Risk of electromagnetic interference

This device is a Class A product. Operation of this equipment in a residential area is likely to cause harmful interference, in which case users are required to take appropriate measures necessary to correct the interference at their own expense.

Virtual Services Platform 9012 cooling

You can configure your Virtual Services Platform 9012 to use first generation or second generation I/O modules. If you use second generation I/O modules, you must install two 9012FCHS cooling modules at the front of the Virtual Services Platform 9012.

First generation mode

In first generation mode, the software supports first and second generation I/O modules together in the same chassis. There are four cooling modules in the chassis, two at the front and two at the back. Each cooling module includes eight fans.

You require the following cooling modules:

- Two cooling modules in the front of the chassis. These cooling modules provide cooling for the I/O modules and the Control Processor (CP) modules. If you use second generation I/O modules, the two front cooling modules must be 9012FCHS cooling modules.
- Two 9012RC cooling modules in the back of the chassis. These cooling modules provide cooling for the SF modules.

Second generation mode

In second generation mode, the software supports only second generation I/O modules. There are four cooling modules in the chassis, two at the front and two at the back. Each cooling module includes eight fans.

You require the following cooling modules:

- Two 9012FCHS cooling modules in the front of the chassis. These cooling modules provide cooling for the I/O modules and the Control Processor (CP) modules.
- Two 9012RC cooling modules in the back of the chassis. These cooling modules provide cooling for the SF modules.

Note:

The 9012FCHS replaces the existing 9012FC cooling modules at the front of the chassis.

Important:

You must use the 9012FCHS cooling module if you install and configure second generation I/O modules; otherwise, the second generation I/O modules remain offline until you install the 9012FCHS cooling modules.

Example

The following messages are associated with the Virtual Services Platform 9012 chassis:

Danger:

Remove the cooling modules using the provided handles. Be careful to keep your fingers out of the fan blades. When you remove a cooling module, allow time for the fans to spin down before you fully withdraw the cooling module.

If you remove an I/O module from the chassis, you must replace it with another I/O 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 I/O modules can overheat and shutdown. An empty slot can also affect the mean time between failure (MTBF) of the I/O modules.

Important:

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.

Caution:

Risk of electromagnetic interference

This device is a Class A product. Operation of this equipment in a residential area is likely to cause harmful interference, in which case users are required to take appropriate measures necessary to correct the interference at their own expense.

Modules

This section provides design guidelines and considerations for Virtual Services Platform 9000 modules.

Use modules to interface the device to the network. There are three module types:

- Switch Fabric modules
- Control Processor modules
- Input/Output modules

Switch Fabric 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 following table shows the supported SF modules by chassis type.

Chassis type	Supported module
Virtual Services Platform 9010 chassis	9095SF module
Virtual Services Platform 9012 chassis	9090SF module

The slot location determines the module function. Slots SF1 and SF4 provide the arbitration and scheduling for traffic (and therefore, bandwidth management) from the I/O 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.

Note:

- If you install a second generation module in Virtual Services Platform 9010 or Virtual Services Platform 9012, you must have a minimum of five SF modules installed. Populate slots SF1 and SF4, and use any other slots for the remaining three SF modules.
- If you install a first generation module in Virtual Services Platform 9010 or Virtual Services Platform 9012, you must have a minimum of three SF modules installed. Populate slots SF1 and SF4, and use any other slot for the remaining SF module.

Control Processor module

The Control Procssor (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.

The Virtual Services Platform 9010 and Virtual Services Platform 9012 support the 9080CP module.

Note:

The 9080CP module requires a minimum software version of 3.4 if you install it in a Virtual Services Platform 9010AC chassis.

For more information about hardware and software compatibility, see *Release Notes for Avaya Virtual Services Platform 9000,* NN46250-401.

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

Input/Output modules

Input/Output (I/O) modules provide support for a variety of technologies, interfaces, and feature sets. They also support up to 1 and 10 Gbps port rates.

You can install a maximum number of I/O modules in the following chassis:

- Virtual Services Platform 9010 chassis: 8 modules
- Virtual Services Platform 9012 chassis: 10 modules

First and second generation I/O modules

The following tables show the first and second generation I/O modules for the Virtual Services Platform 9000.

Table 3: First generation I/O modules

I/O module	Description
9048GT	48-port 10/100/1000BASE-T
9048GB	48-port 1000BASE-X SFP
9024XL	24-port 10GBASE-R SFP+ / 1000BASE-X SFP **Note:** 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 10 Gigabit. 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.

Table 4: Second generation I/O modules

I/O module	Description
9048XS-2	48-port 10GBASE-R SFP+ / 1000BASE-X SFP
9012QQ-2	12-port 40GBASE-R QSFP+

For more information about SFP and SFP+ specifications, see *Installing Transceivers and Optical Components on Avaya Virtual Services Platform 9000,* 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 215. 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 *Installing Transceivers and Optical Components on Avaya Virtual Services Platform 9000,* NN46250-305.

1000BASE-X and 10GBASE-R reach

You can use various SFP (1 Gbps) and SFP+ (10 Gbps) 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 *Installing Transceivers and Optical Components on Avaya Virtual Services Platform 9000,* NN46250-305.

Table 5: SFP optical devices and maximum reach

SFP	Maximum reach
1000BASE-SX	Up to 550 m
1000BASE-LX	Up to 10 km
1000BASE-BX	Up to 10 km
1000BASE-XD	Up to 40 km
1000BASE-ZX	Up to 70 km
1000BASE-EX	Up to 120 km

Table 6: SFP+ optical devices and maximum reach

SFP+	Maximum reach
10GBASE-LRM	Up to 220 m
10GBASE-SR/SW	Up to 400 m
10GBASE-LR/LW	Up to 10 km

SFP+	Maximum reach
10GBASE-ER/EW	Up to 40 km
10GBASE-ZR/ZW	Up to 70 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 bit rate 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 7: Maximum cable distances

	10BASE-T	100BASE-TX	1000BASE-TX
IEEE standard	802.3 Clause 14	802.3 Clause 21	802.3 Clause 40
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 8: 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

The Virtual Services Platform supports multiple power supplies for custom power redundancy configurations. Two typical configurations are N+1 or N+N, where N is the number of required power supplies to power the chassis and modules.

- An N+ 1 redundant configuration provides power in the event of a single power supply failure, single circuit breaker trip, or knocked-out power cord.
- An N+N redundant configuration provides power in the event of a loss of a single power phase in the building. Balance the line side voltage source between building phases.

Note:

The Virtual Services Platform 9010 chassis can support N+N redundancy with the 2 kW power supply.

The input voltage range for the 9006AC power supply is 90-140 VAC for a 120 VAC nominal connection, and 185–275 VAC for a 240 VAC nominal connection.

The output power for the 9006AC is limited to 1,200 W maximum at 90-140 VAC input voltage conditions. To obtain full output power of 2,000 W, you must connect the 9006AC power supply to a 185-275 VAC input voltage source.

Important:

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

For more information about power supplies, see *Installing AC Power Supplies in Avaya Virtual Services Platform 9000,* 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 *Administering Avaya Virtual Services Platform 9000,* NN46250-600.

Switch Fabric redundancy

Avaya recommends that you install SF modules in both SF1 and SF4. Install an SF module in 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.

Important:

If you install a second generation I/O module in Virtual Services Platform 9010 or Virtual Services Platform 9012, you must have a minimum of five SF modules installed. Populate slots SF1 and SF4, and use any other slots for the remaining three SF modules.

If you install a first generation I/O module in Virtual Services Platform 9010 or Virtual Services Platform 9012, you must have a minimum of three SF modules installed. Populate slots SF1 and SF4, and use any other slot for the remaining SF module.

For more information about the SF modules, see *Installing Modules in Avaya Virtual Services Platform 9000,* 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 32.

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 53, Layer 3 switch clustering and RSMLT on page 67, 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 synchronize. In non HA-CPU mode, also called Warm Standby, the two CPUs do not synchronize.

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

Table 9: Feature support for HA-CPU

Feature	Release 3.4
Modules	Yes
Platform	Yes
Layer 2	Yes
Layer 3	Yes; partial-HA for Border Gateway Protocol and IPv6
Multicast	Partial HA
Multicast virtualization (virtualized IGMP only)	Full HA
Security	Yes
Applications	Partial HA

Table 10: Synchronization capabilities in HA-CPU mode

Synchronization of	Release 3.4
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
Shortest Path Bridging MAC (MAC)	Yes
Connectivity Fault Management (CFM)	Yes
SMLT parameters	Yes
VLAN parameters	Yes
Layer 3	
ARP entries	Yes
Border Gateway Protocol (BGP)	Partial (configuration only)

Synchronization of	Release 3.4
Dynamic Host Configuration Protocol (DHCP) Relay	Partial (configuration only)
Internet Group Management Protocol (IGMP)	Yes
IGMP virtualization Snooping SPB	Yes
IGMP Snooping routed-SPB	Yes
IP filters	Yes
IPv6 filters	Yes
IPv6	Partial (configuration only)
Layer 3 Filters: access control entries, access control lists	Yes
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
Shortest Path Bridging MAC (SPBM)	Yes
Connectivity Fault Management (CFM)	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
Terminal Access Controller Access-Control System plus (TACACS+)	Partial (configuration only)
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 *Administering Avaya Virtual Services Platform 9000,* NN46250-600.

Partial HA implementation

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
- IPv6
- Link Layer Discovery Protocol (LLDP), 802.1AB
- Protocol Independent Multicast-Sparse Mode (PIM-SM)
- Protocol Independent Multicast-Source Specific Mode (PIM-SSM)
- Terminal Access Controller Access-Control System plus (TACACS+)
- 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).

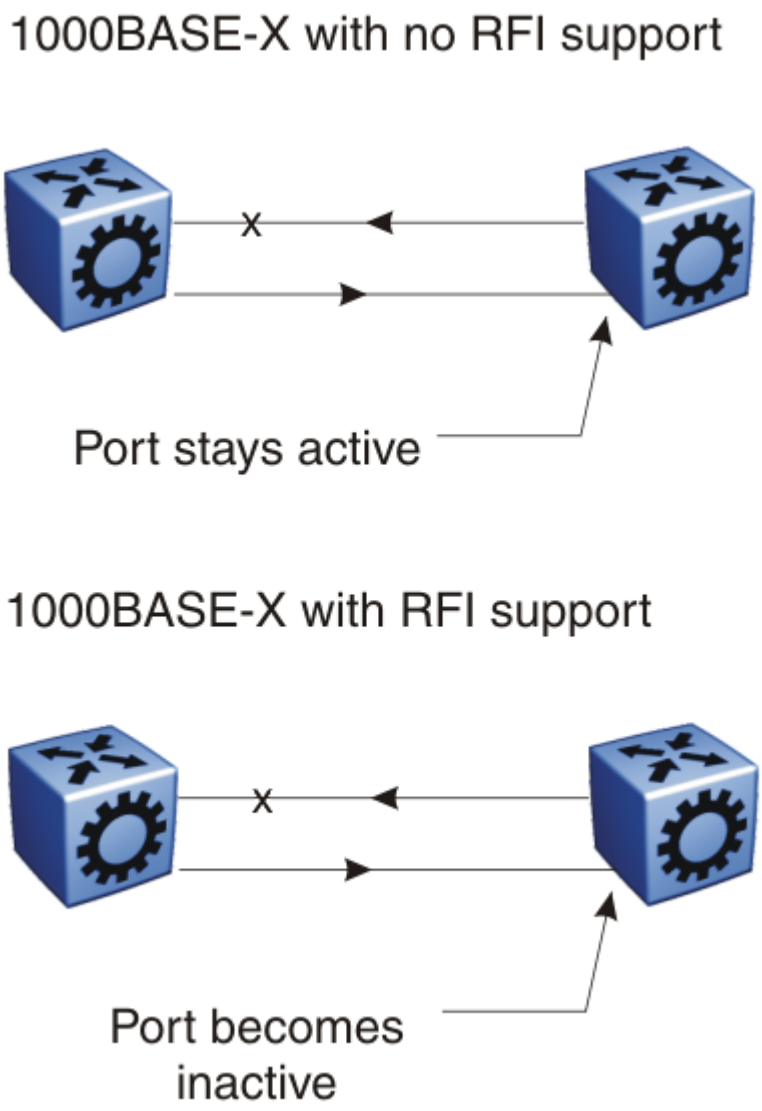


Figure 2: 1000BASE-X RFI

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.

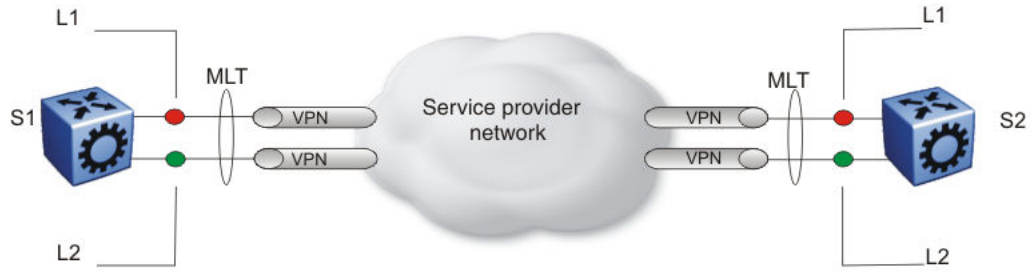


Figure 3: Problem description (1 of 2)

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.

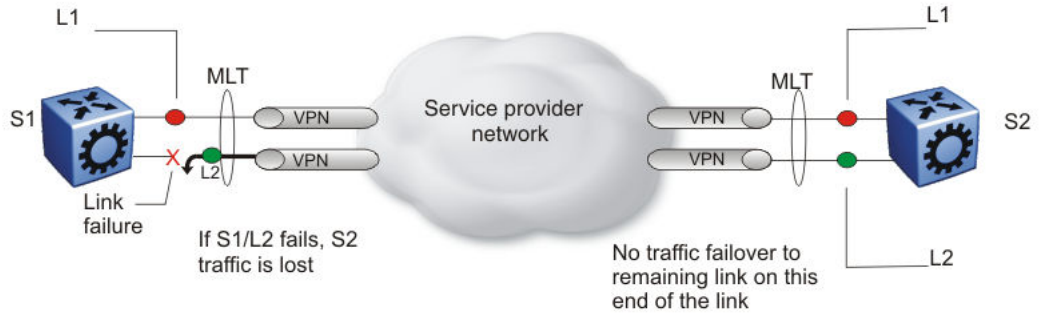


Figure 4: Problem description (2 of 2)

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 11: 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 eight active links.

For LACP fundamentals and configuration procedures, see *Configuring Link Aggregation, MLT, and SMLT on Avaya Virtual Services Platform 9000,* 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 (NLB) is a clustering technology available with Microsoft Windows 2000, Microsoft Windows 2003, Microsoft Windows 2008, and Microsoft Windows 2012 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.

You can alter the operation of Network Load Balancing clustering on the Virtual Services Platform 9000 by configuring the global IP ARP Multicast MAC flooding feature. This feature applies only when you configure NLB for Multicast or Internet Group Management Protocol (IGMP)-multicast modes. If you enable IP ARP multicast MAC flooding, all traffic sent to the virtual IP address for the cluster floods all member ports in the NLB VLAN.

Unicast mode inherently floods all ports in an NLB VLAN. Unicast mode supports connectivity to a secondary virtual IP address. For more information about software scaling capabilities, see *Avaya Virtual Services Platform 9000 — Release Notes*, NN46250-401.

Multicast and IGMP-multicast modes support multiple clusters for each NLB VLAN and connectivity to a secondary virtual IP address with Global IP ARP multicast MAC flooding enabled.

Note:

SPBM supports Network Load Balancing (NLB) unicast. SPBM does not support NLB multicast or NLB multicast with IGMP.

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

Network Load Balancing topologies

Virtual Services Platform 9000 supports the Network Load Balancing topologies in the following figures.

Topology 1

Use the topology in the following figure to deploy Network Load Balancing clusters in unicast, multicast, and IGMP-multicast modes without routing. Do not enable IP routing on the switching platform.

Note:

If you configure NLB unicast mode on a VLAN, Virtual Services Platform 9000 floods Network Load Balancing cluster traffic to every port in that VLAN. If you deploy NLB clusters using multicast or IGMP-multicast mode on the VLAN, Virtual Services Platform 9000 only forwards traffic to ports where the NLB clusters are attached. If you enable global IP ARP multicast MAC flooding, every port in the NLB configured VLAN gets flooded.

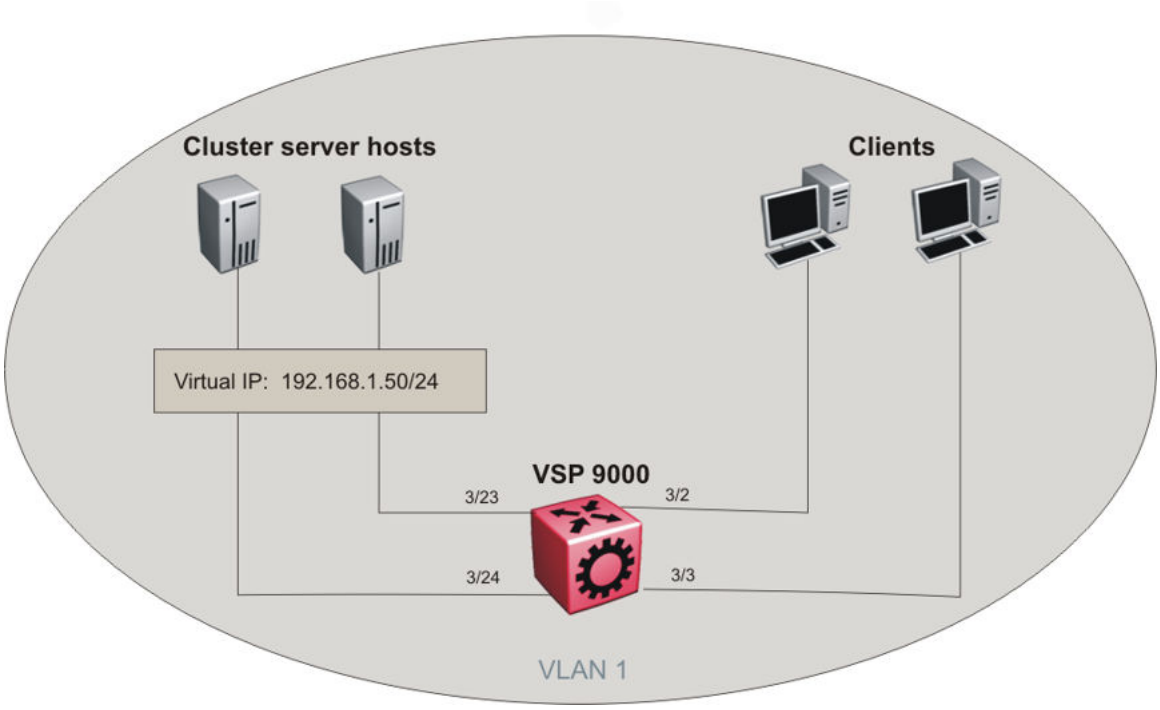


Figure 5: Network server load balancing — Topology 1

Topology 2

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.

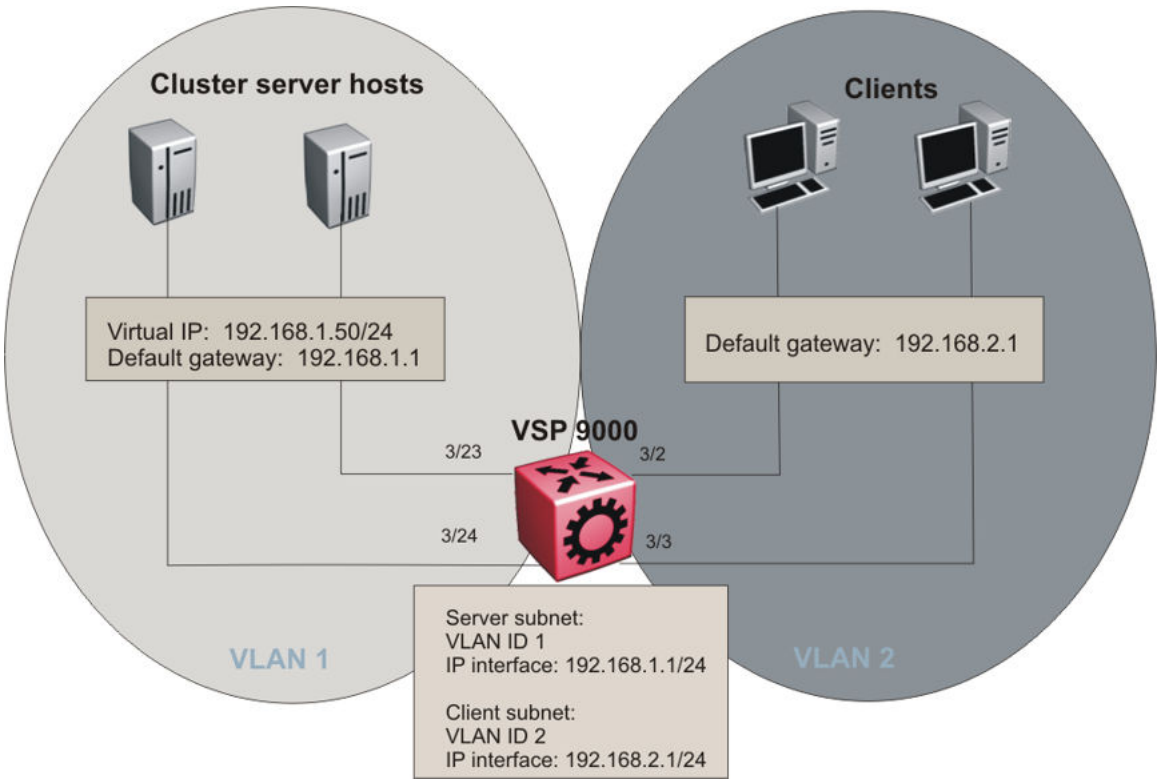


Figure 6: Network server load balancing — Topology 2

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

Topology 3

In the following topology, this figure shows more than one cluster for each VLAN. If you configure NLB unicast mode on a VLAN, Virtual Services Platform 9000 floods Network Load Balancing cluster traffic to every port in that VLAN. If you deploy NLB clusters using multicast or IGMP-multicast mode on a VLAN, Virtual Services Platform 9000 only forwards traffic to ports where the NLB clusters are attached. However, because Virtual Services Platform 9000 only supports one NLB multicast ARP entry in the IP ARP table, you must enable global IP ARP multicast MAC flooding to support multiple clusters with multicast and IGMP-multicast modes.

Important:

If you enable global IP ARP multicast MAC flooding, every port in the NLB configured VLAN gets flooded. If you do not enable global IP ARP multicast MAC flooding and you have configured multicast or IGMP-multicast mode, the topology of multiple clusters on a single VLAN cannot be supported because there is no guarantee which NLB cluster will be reachable from the Virtual Services Platform 9000.

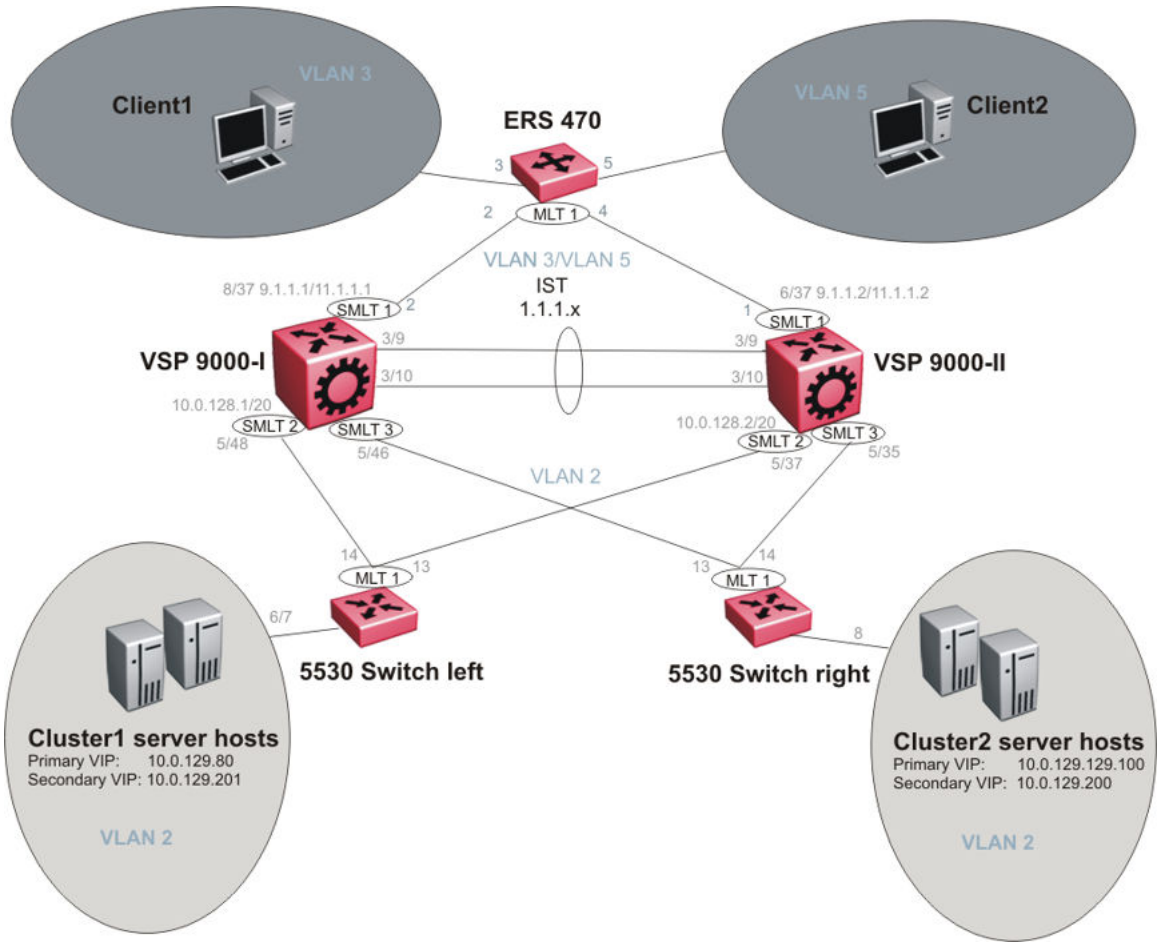


Figure 7: Network server load balancing — Topology 3

Chapter 8: Lossless networks

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

On Virtual Services Platform 9000, you can configure 10 Gb 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 10Gb 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 12: Lossless treatment based on pause type

Ingress port	Egress port	Lossless guaranteed
PFC	PFC	802.1p only
PFC	FC	802.1p only
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 Gb ports.

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

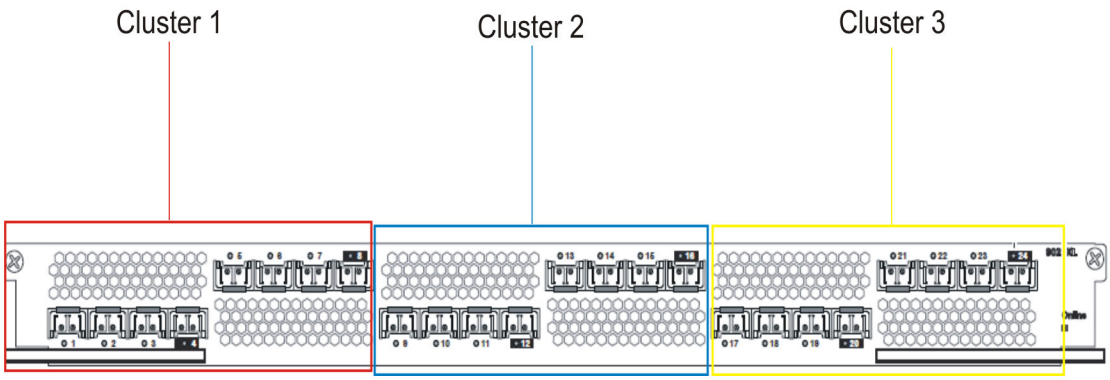


Figure 8: 9024XL Lossless Ethernet clusters

- 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

This section shows multiple lossless deployments.

Pause frame over a single hop

The following figure shows pause-frame support over a single hop. All links are 10 Gb/s. 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.

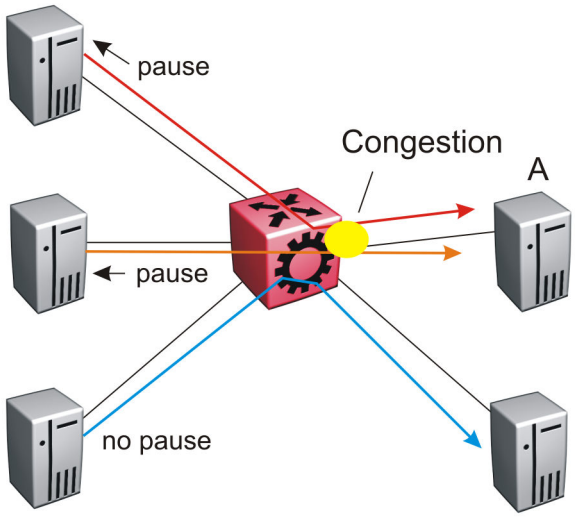


Figure 9: Pause frame over a single hop

Pause frame over two hops

The following figure shows pause-frame support over two hops. All links are 10 Gb/s. 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.

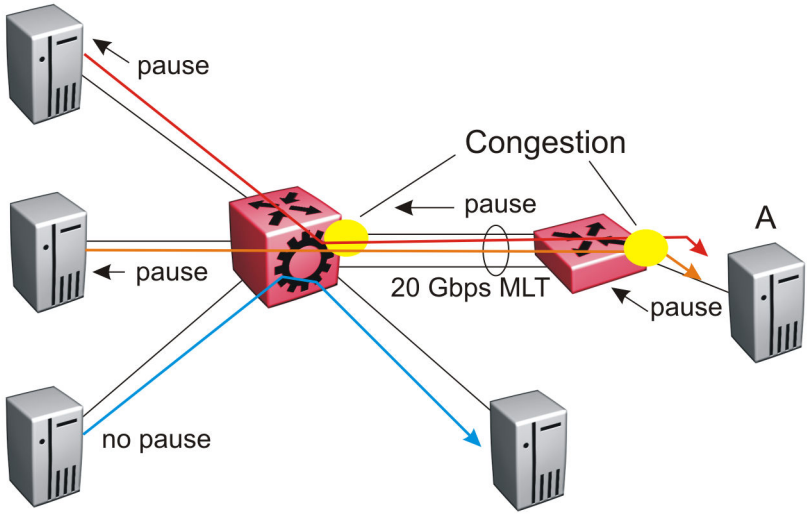


Figure 10: Pause frame over two hops

PFC frame over single hop

The following figure shows PFC-frame support over a single hop. All links are 10 Gb/s. 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.

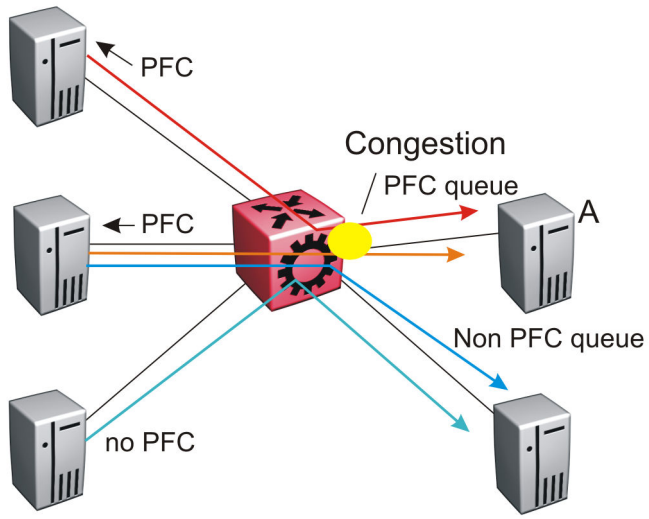


Figure 11: PFC frame over single hop

PFC frame over two hops

The following figure shows PFC-frame support over two hops. All links are 10 Gb/s. 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.

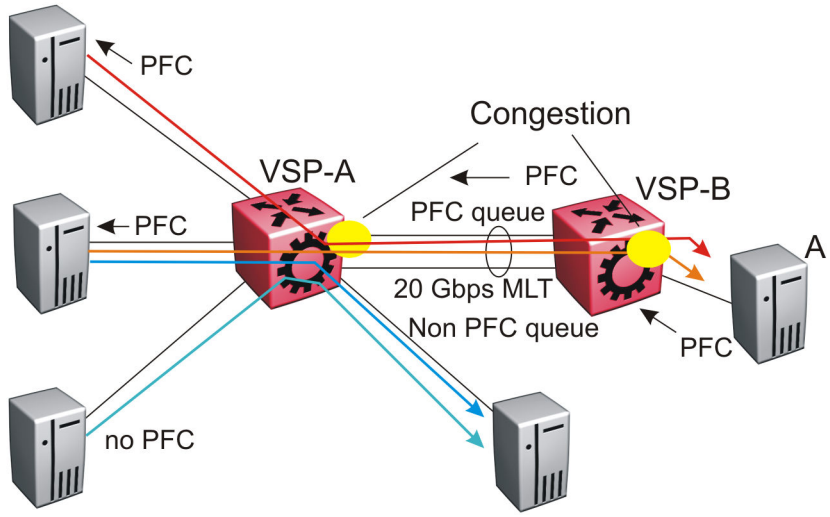


Figure 12: PFC frame over two hops

Data-center deployment

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.

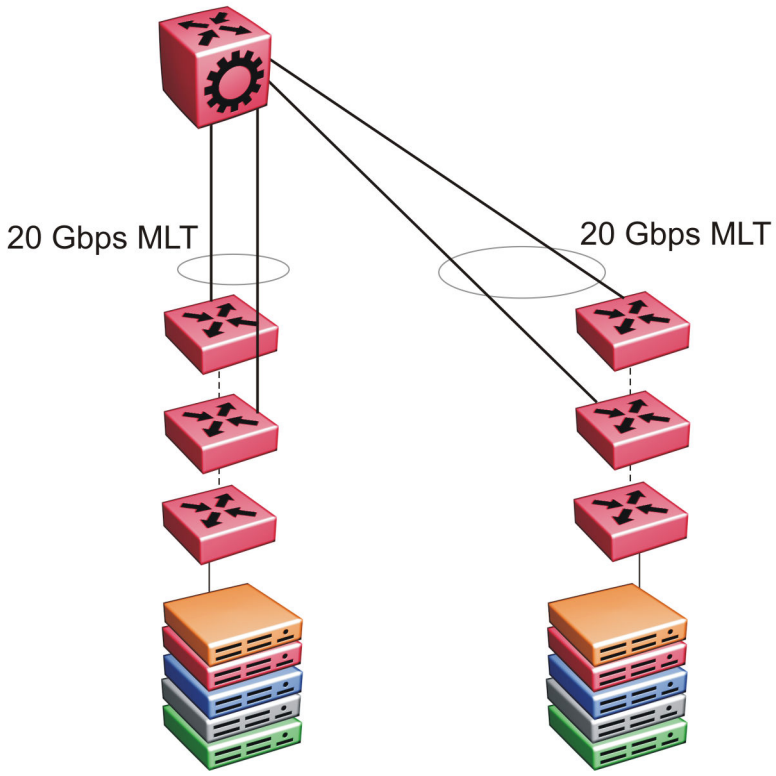


Figure 13: Data-center deployment with one-switch core

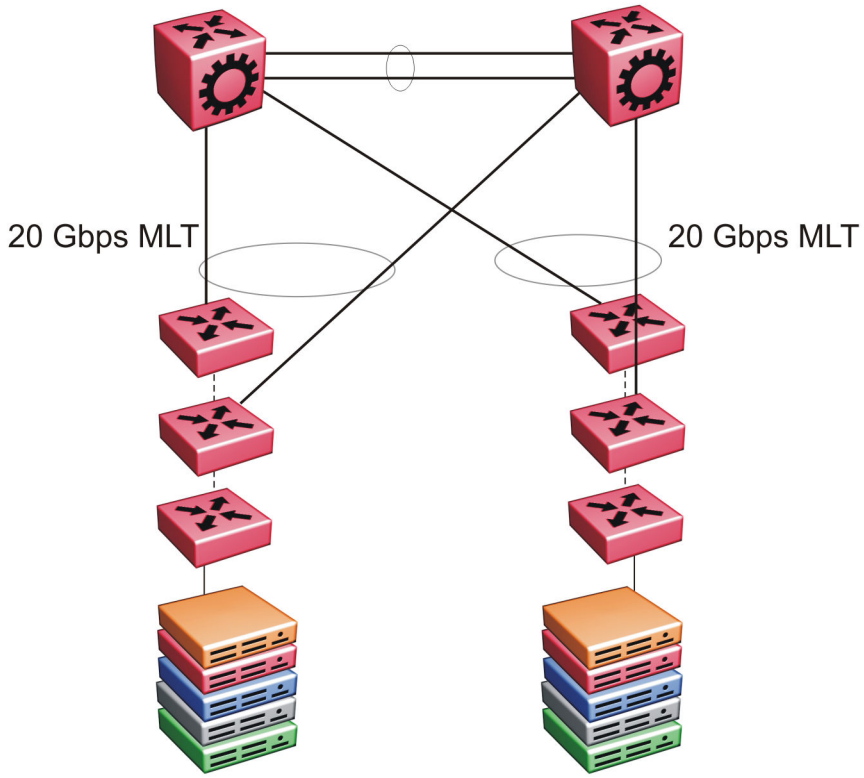


Figure 14: Data-center deployment with two-switch core

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.

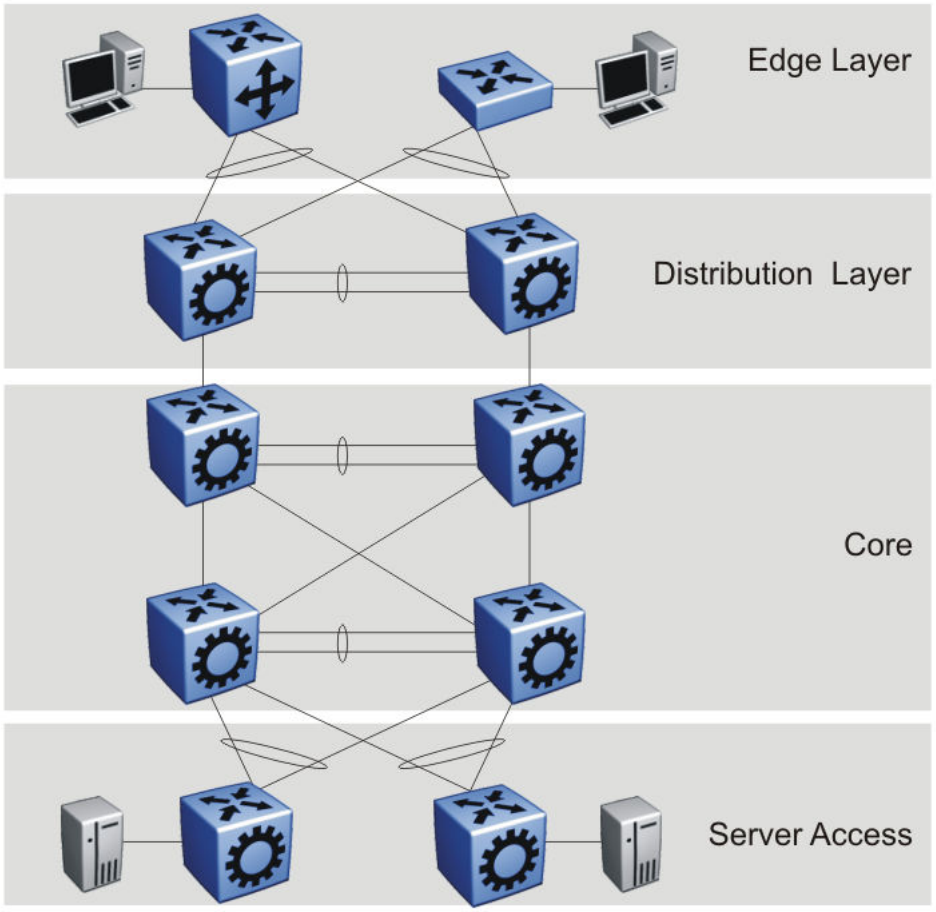


Figure 15: Three-tiered architecture plus data center

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.

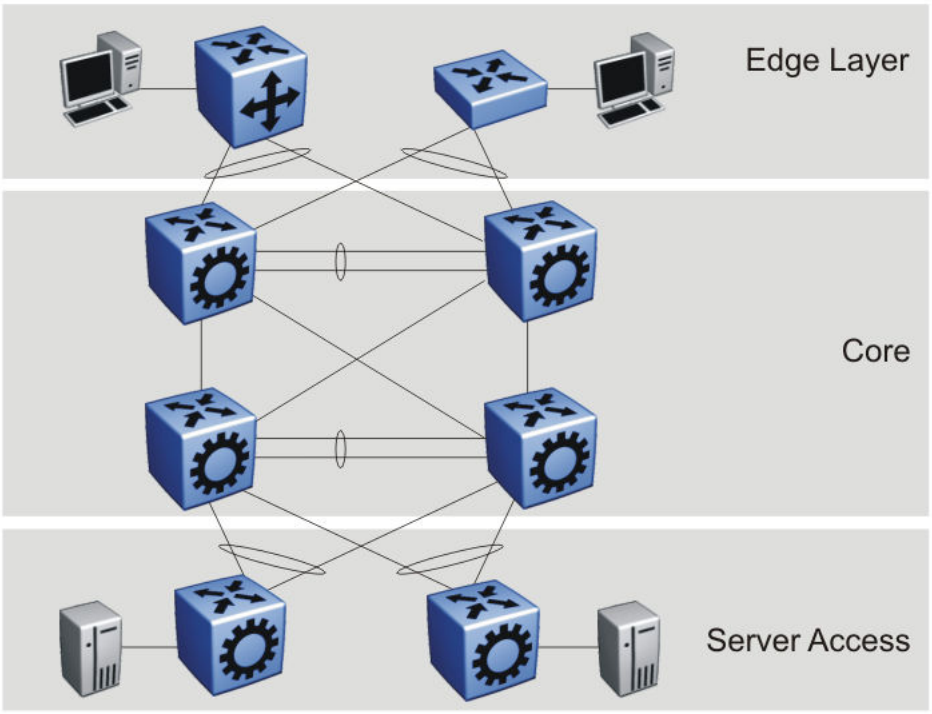


Figure 16: Two-tiered architecture with four-switch core plus data center

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

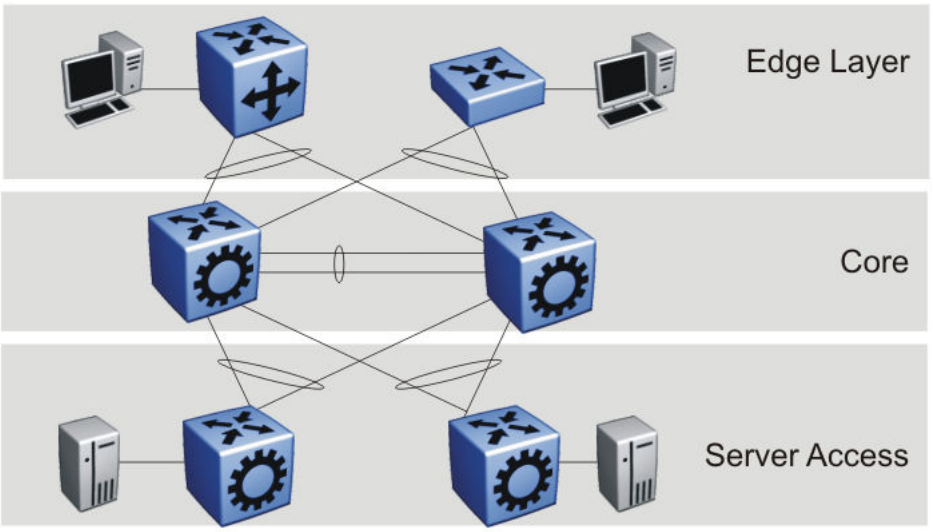


Figure 17: Two-tiered architecture with two-switch core plus data center

For more information about 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.

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.

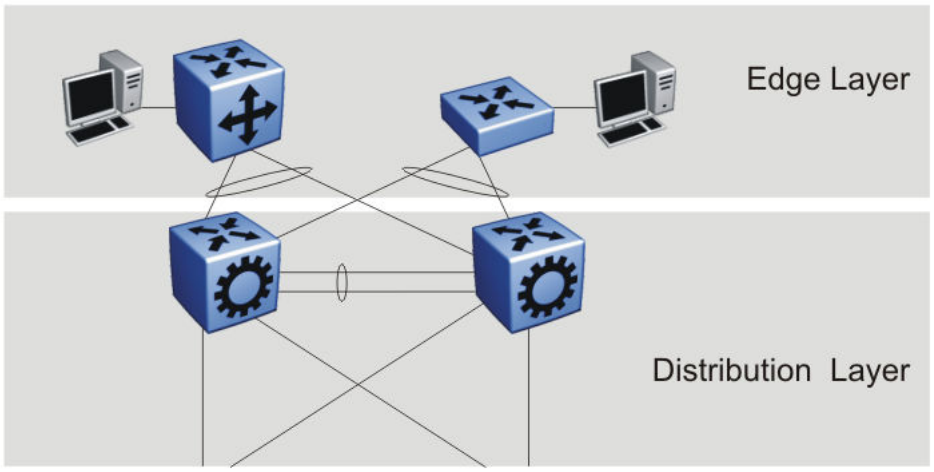


Figure 18: Redundant network edge diagram

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.

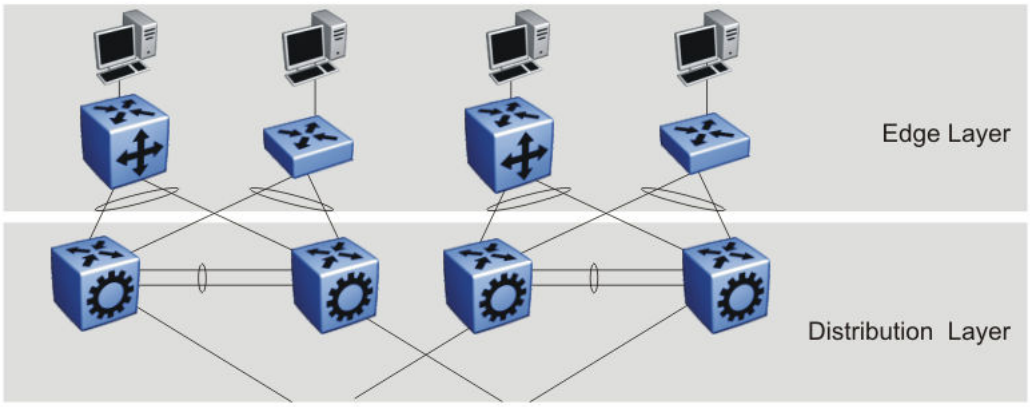


Figure 19: Recommended network edge design

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.

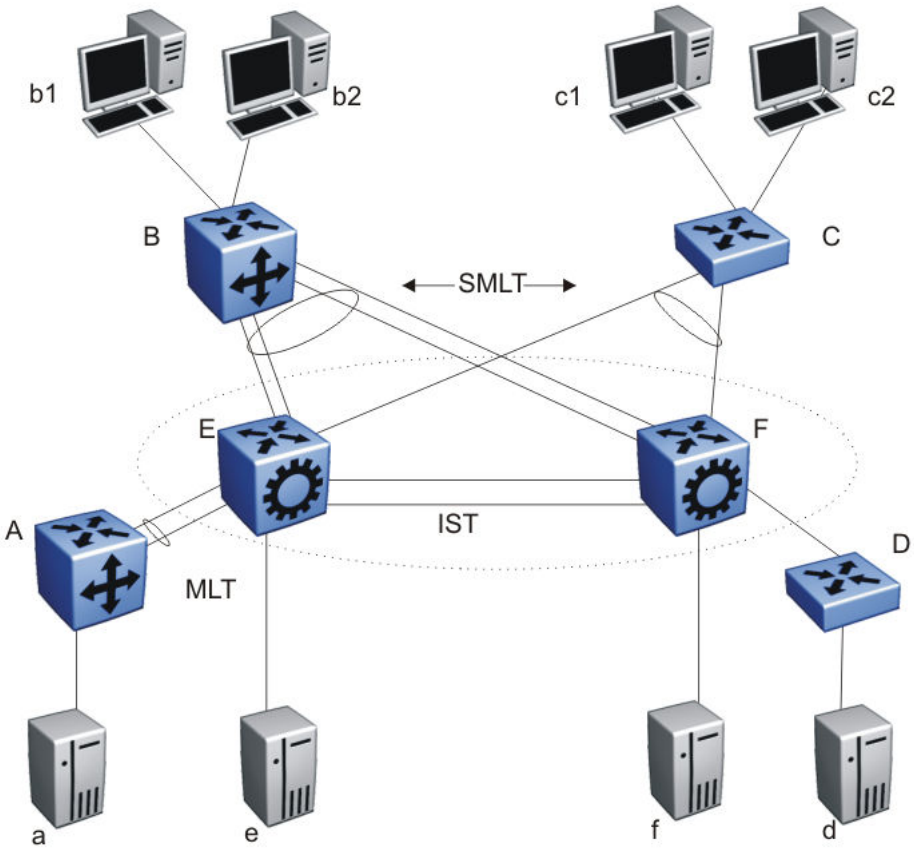


Figure 20: SMLT triangle configuration

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 13: 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 20: SMLT triangle configuration on page 58 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 *Configuring Link Aggregation, MLT, and SMLT on Avaya Virtual Services Platform 9000,* 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.

Comments? infodev@avaya.com

- 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 *Configuring Link Aggregation, MLT, and SMLT on Avaya Virtual Services Platform 9000,* 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 *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* 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 511 SMLT edge switch connections.

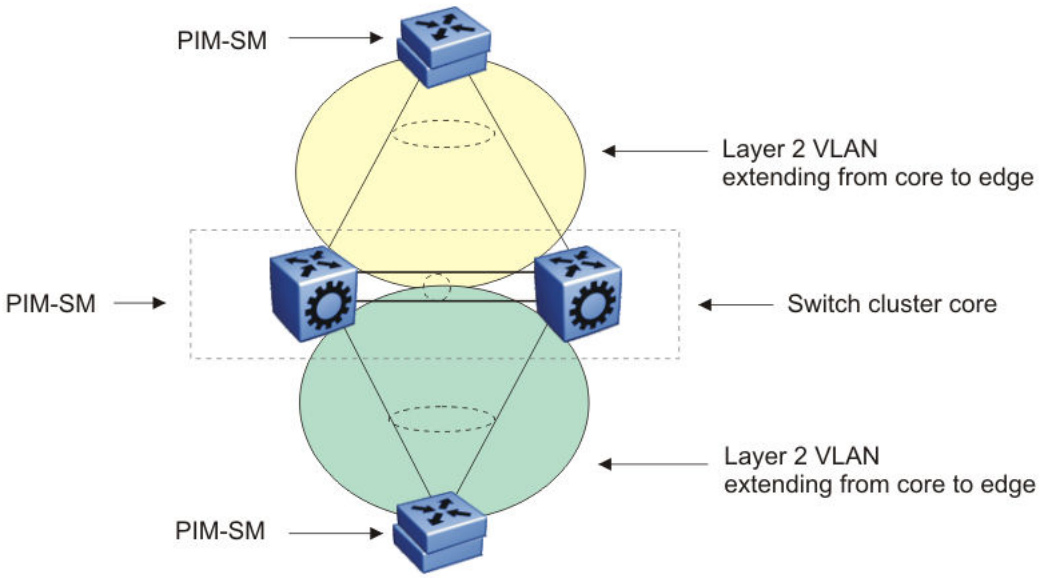


Figure 21: SMLT triangle configuration

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.

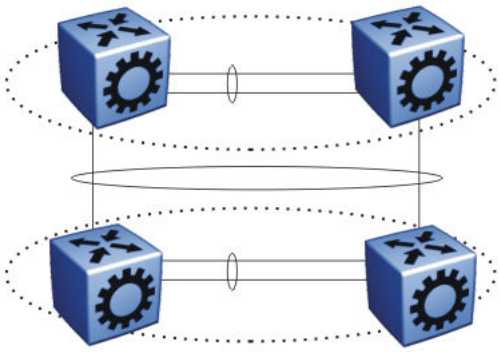


Figure 22: SMLT square configuration

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.

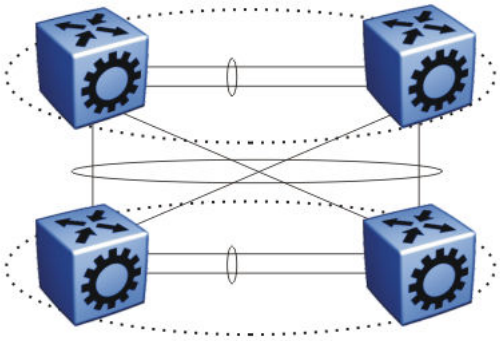


Figure 23: SMLT full-mesh configuration

Virtual Services Platform 9000 supports up to 512 MLT groups of 16 ports. Configure SMLT groups as shown in Figure 24: SMLT scaling on page 65 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.

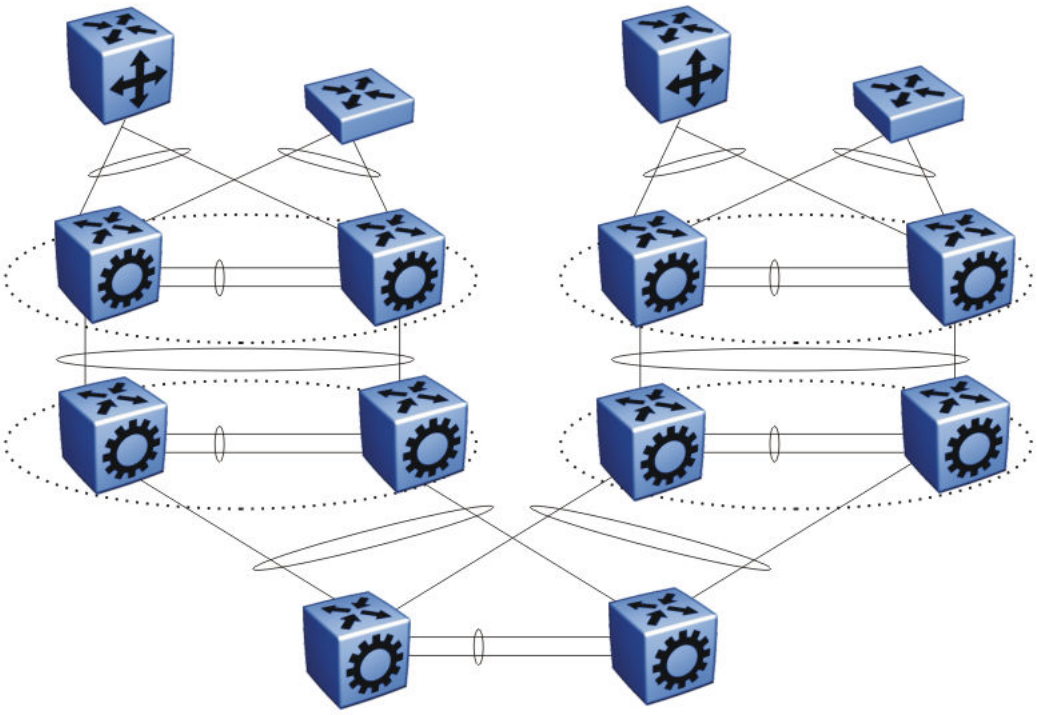


Figure 24: SMLT scaling

For more information about the SMLT triangle, square, and full-mesh designs, see *Configuring Link Aggregation, MLT, and SMLT on Avaya Virtual Services Platform 9000,* 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 25: SMLT and RSMLT in Layer 2 and 3 environments on page 68 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.

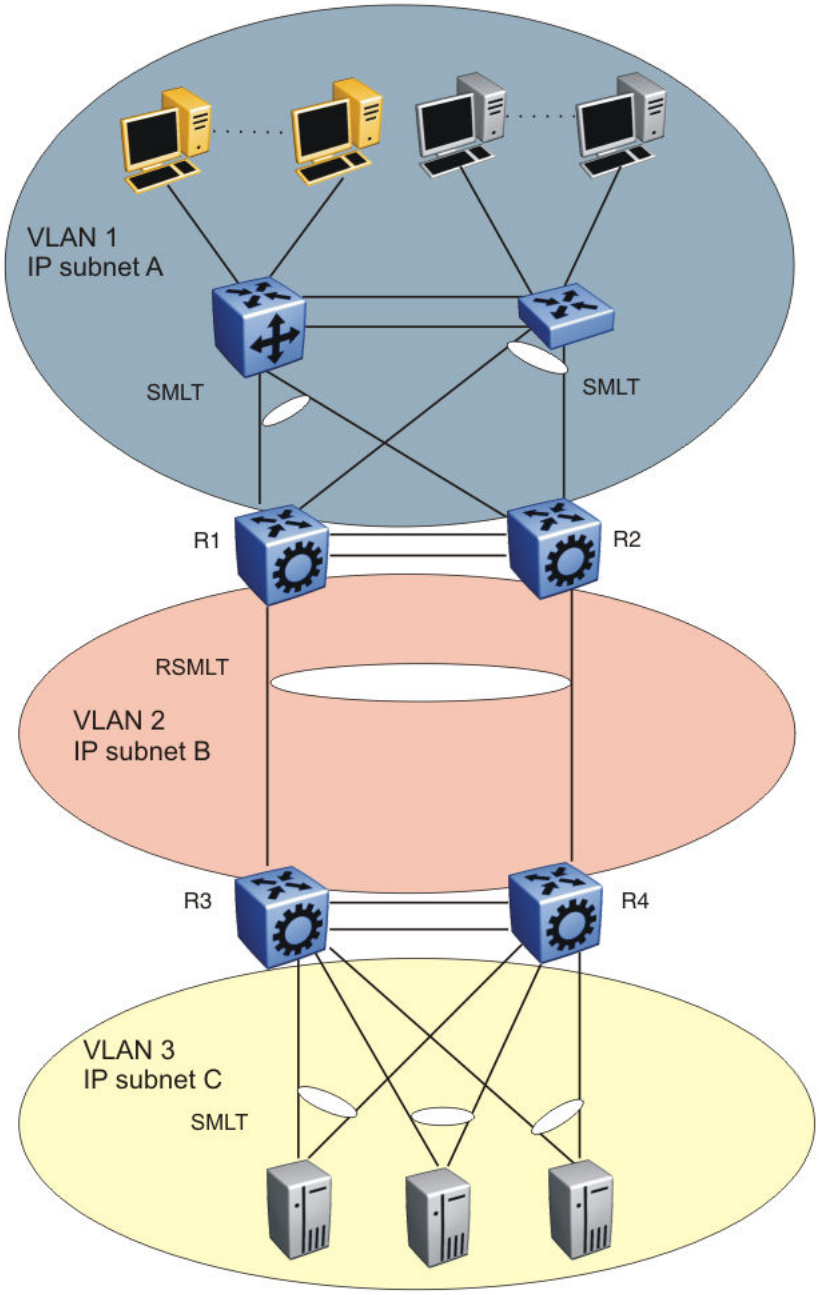


Figure 25: SMLT and RSMLT in Layer 2 and 3 environments

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 209.

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 25: SMLT and RSMLT in Layer 2 and 3 environments on page 68.

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 209.

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.
- 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.

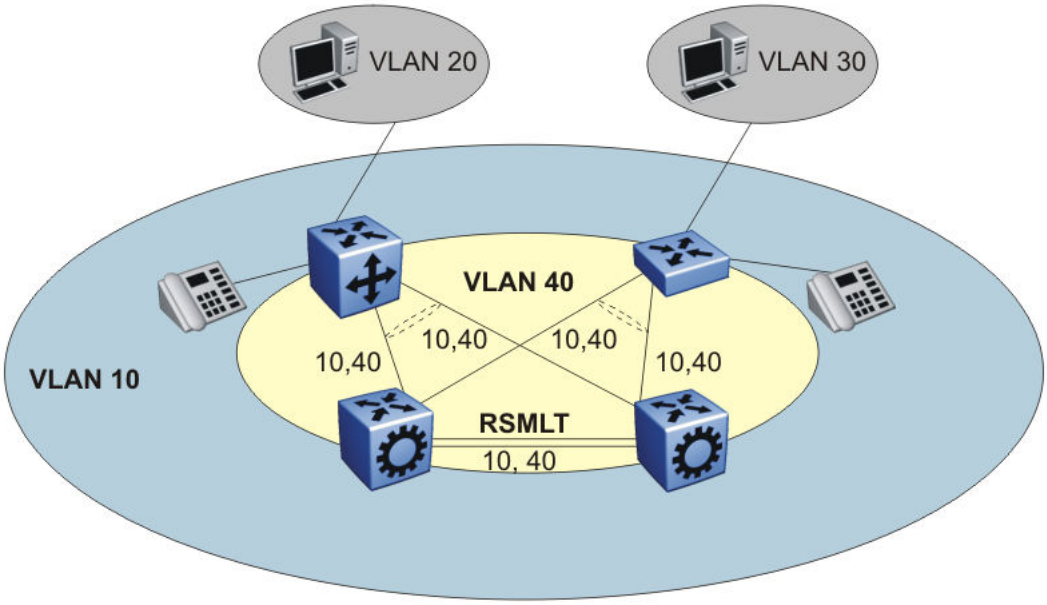


Figure 26: VLAN with all IP Deskphones as members

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.

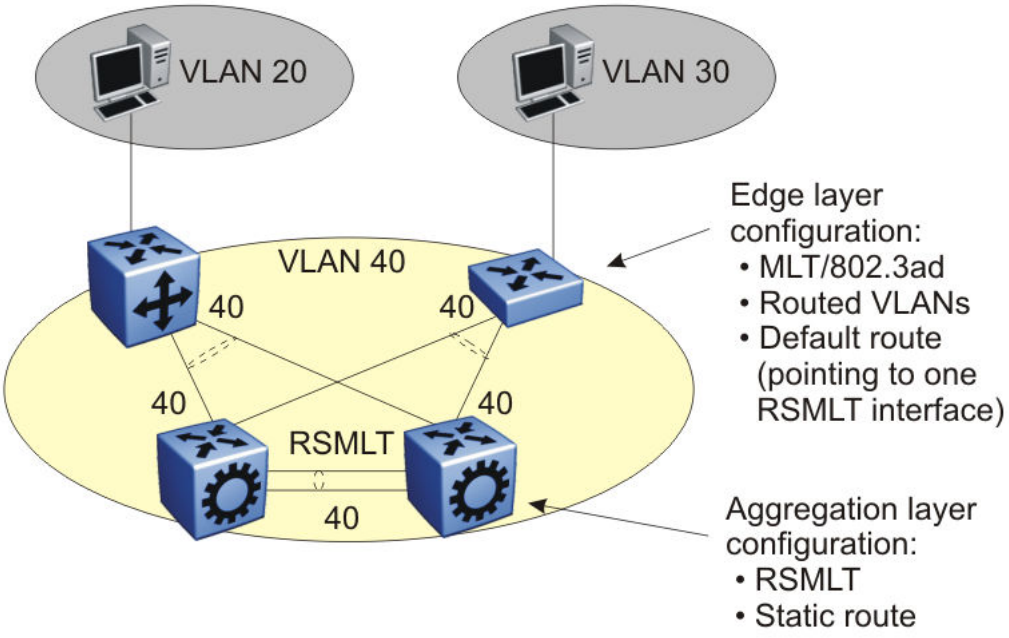


Figure 27: VLAN edge configuration

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.

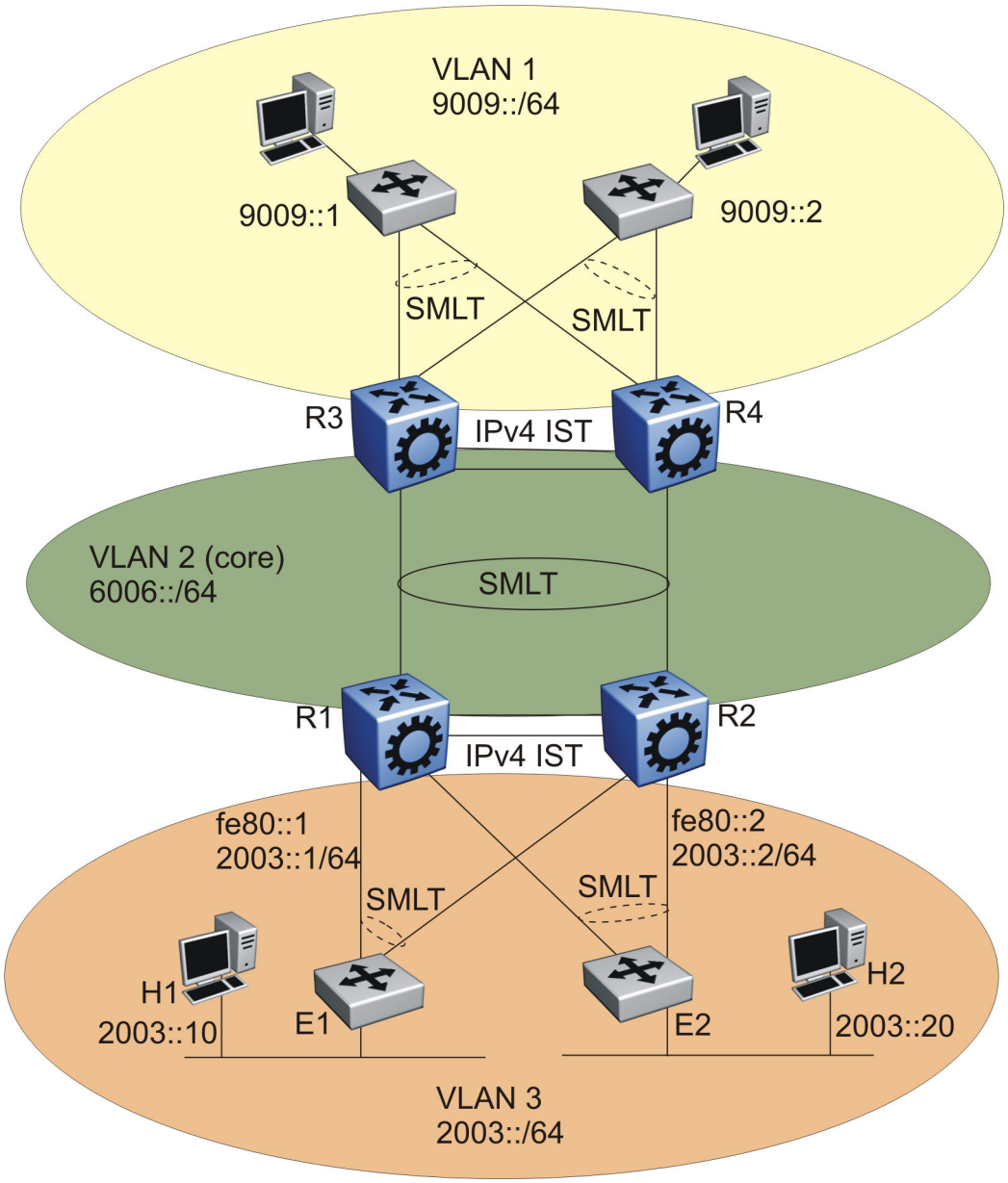


Figure 28: IPv6 RSMLT topology

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.

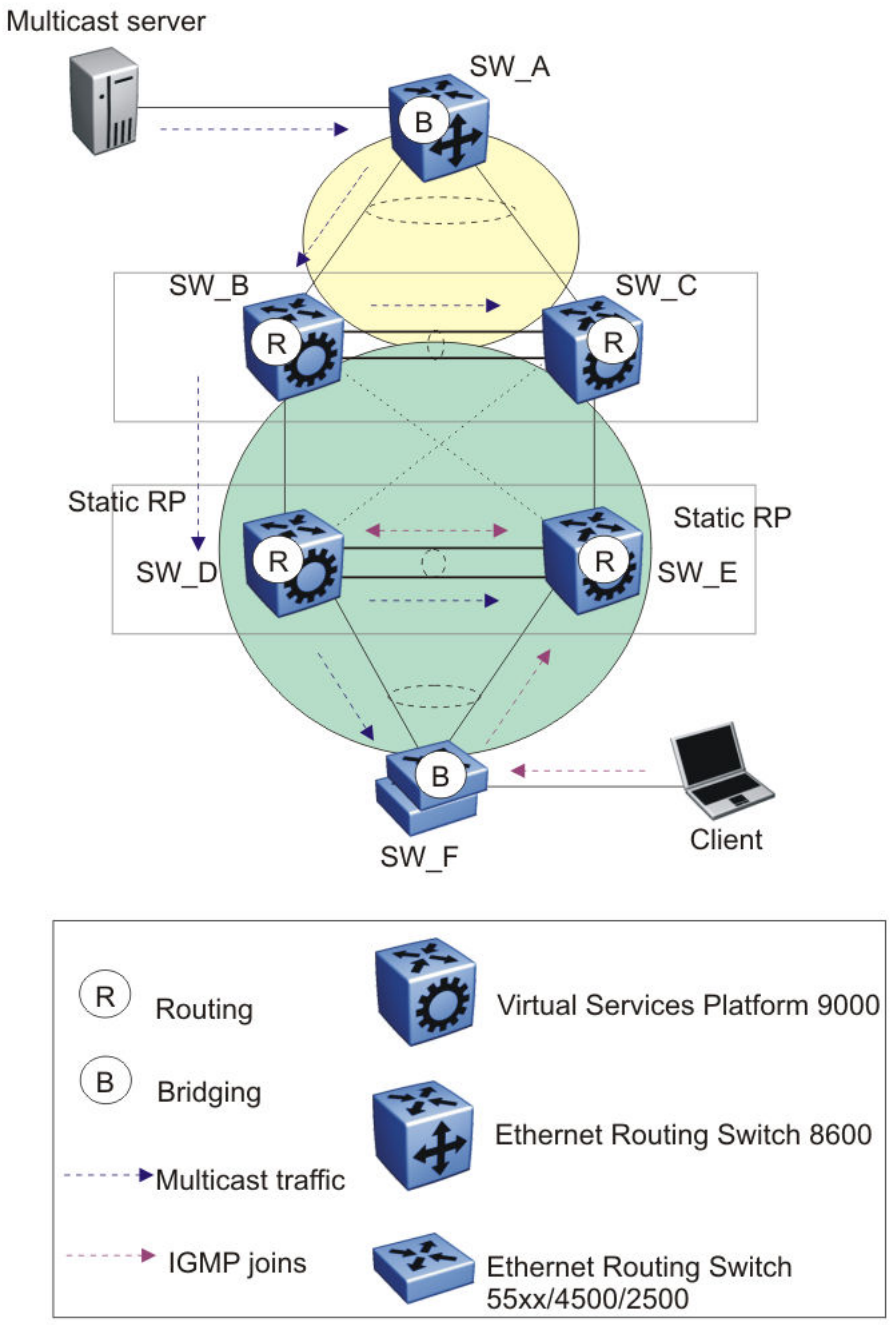


Figure 29: Multicast behavior in SMLT environment

In Figure 29: 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.

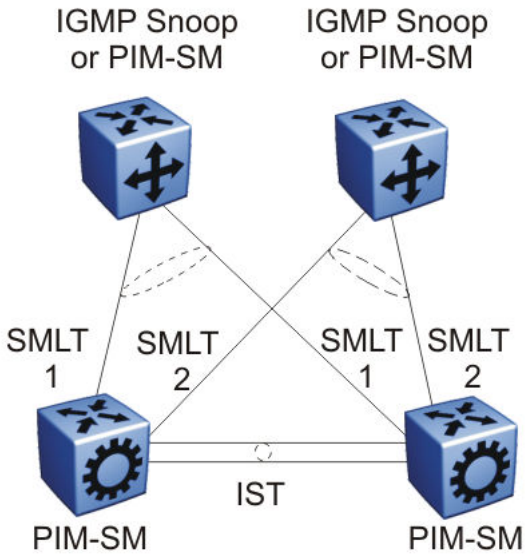


Figure 30: Multicast routing using PIM-SM

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.

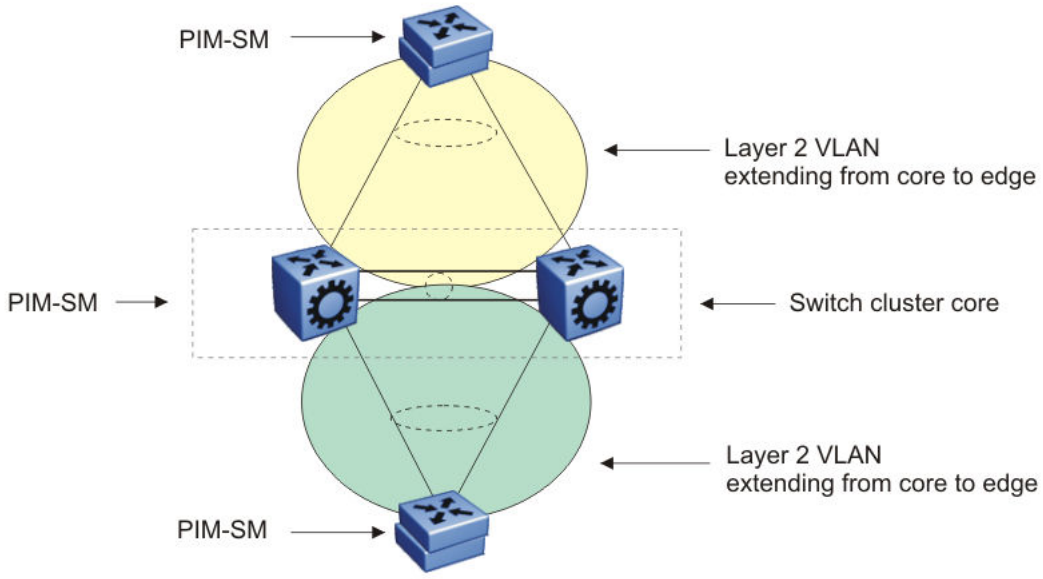


Figure 31: Multicast SMLT triangle

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.

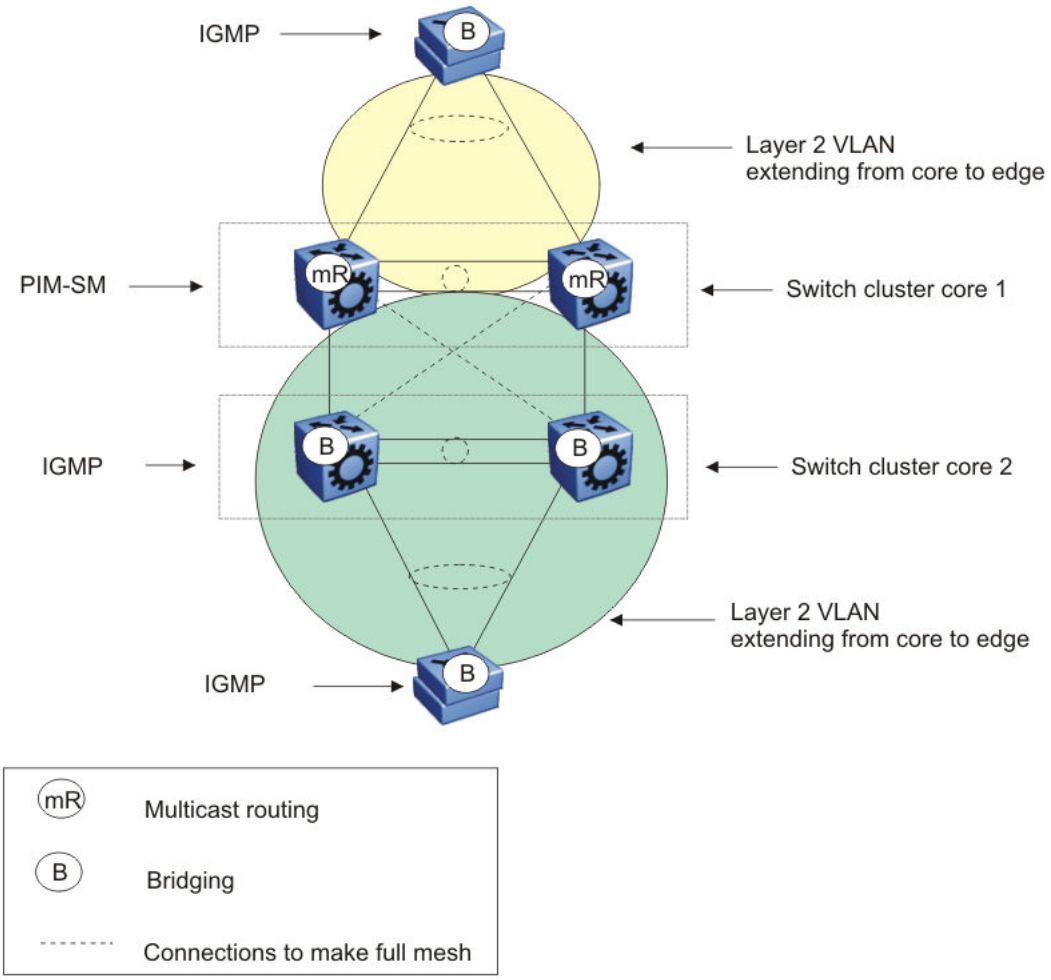


Figure 32: Multicast SMLT square 1

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.

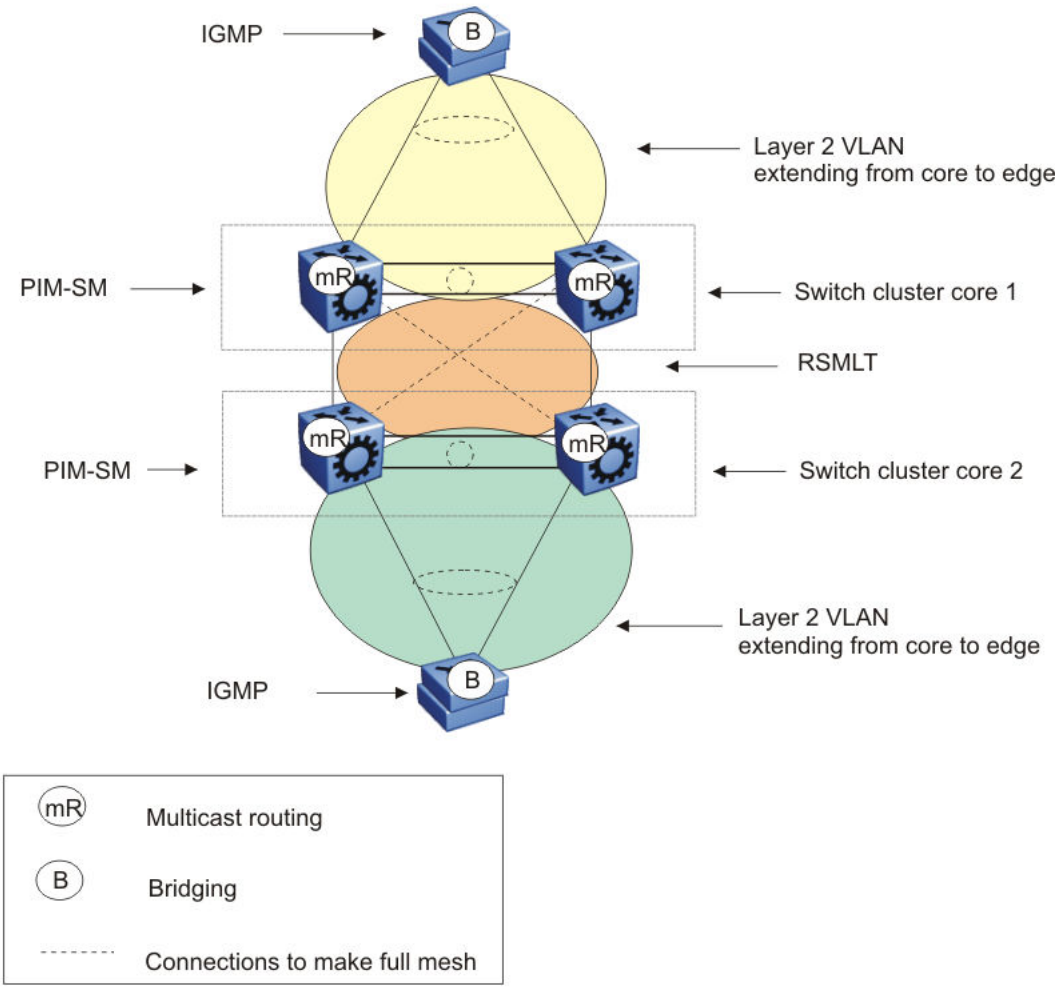


Figure 33: Multicast SMLT square 2

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.

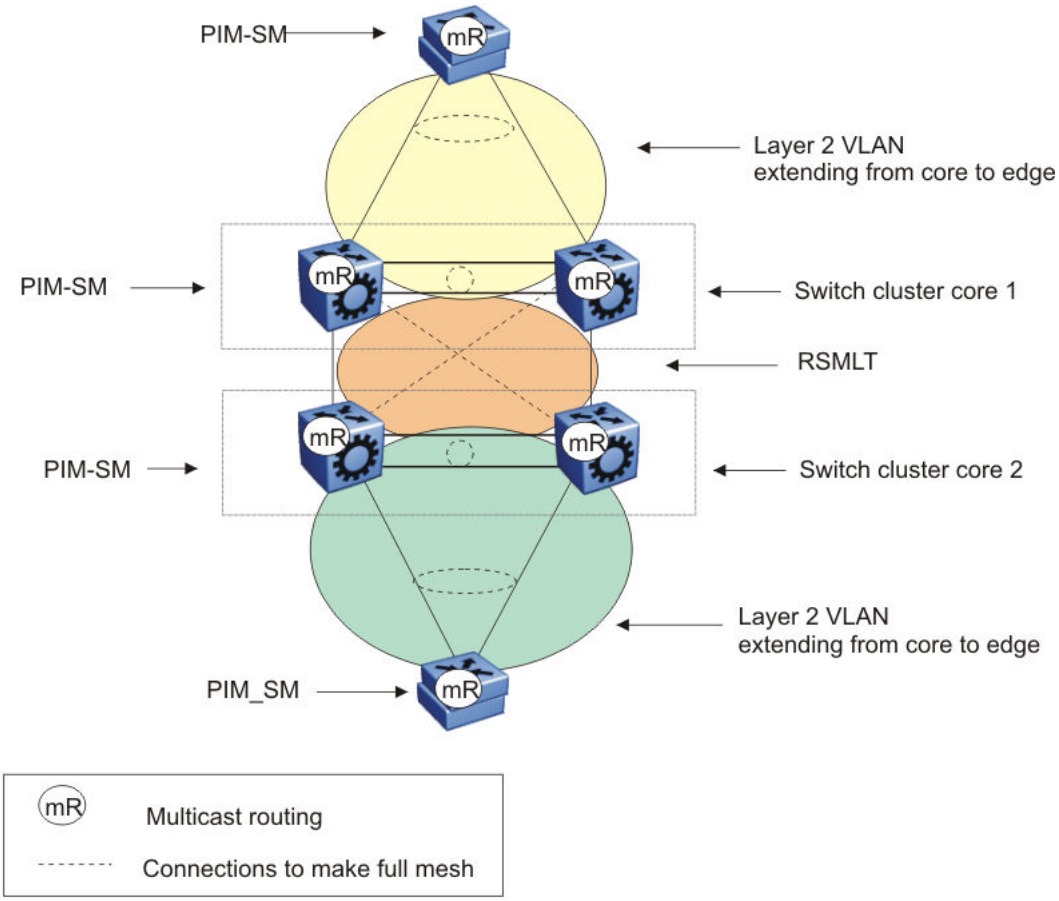


Figure 34: Multicast SMLT square 3

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.

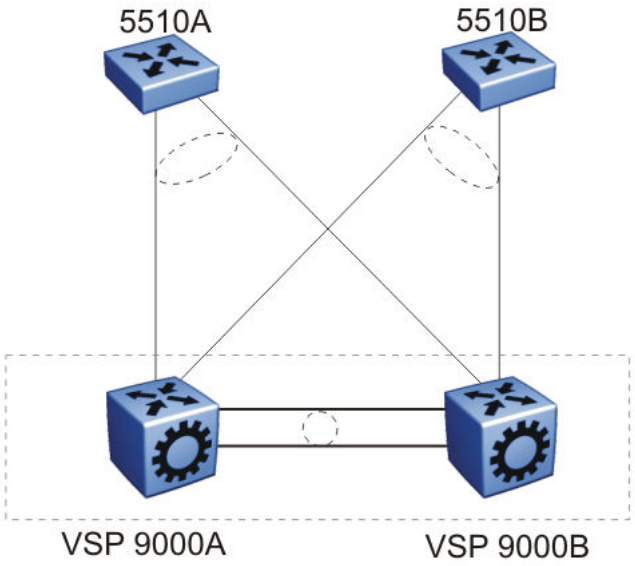


Figure 35: Unicast route example

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 *Administering Avaya Virtual Services Platform 9000,* NN46250-600. For more information about SLPP and loop detection, see *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* 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.

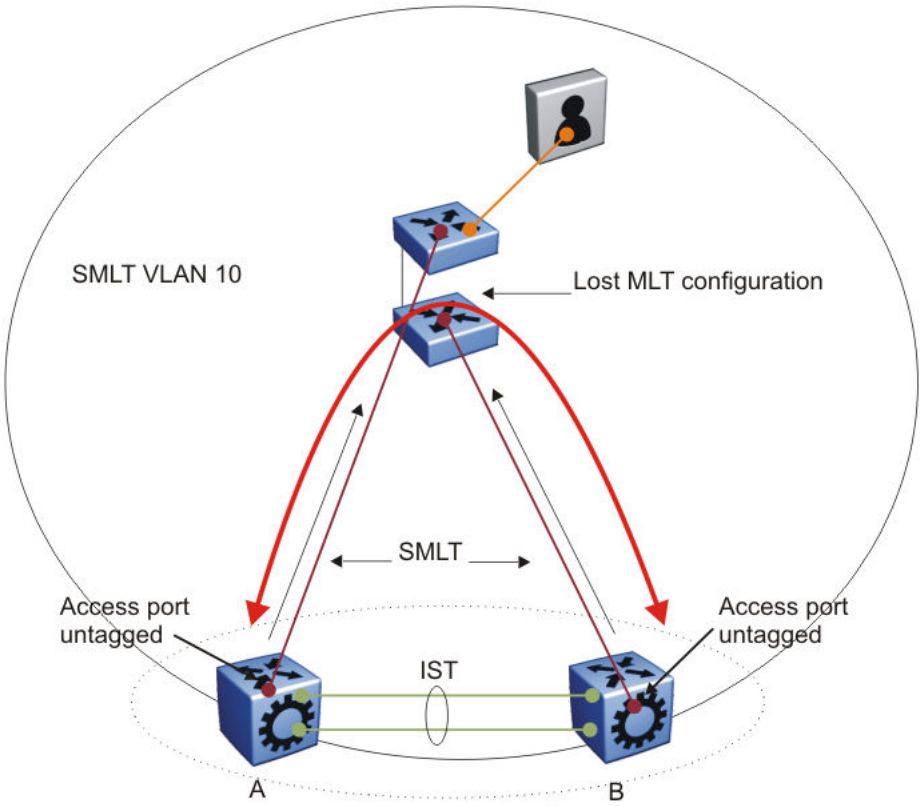


Figure 36: Untagged SMLT links

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 36: Untagged SMLT links on page 84 detects loops and avoids edge isolation. For tagged data, consider the configuration in the following figure.

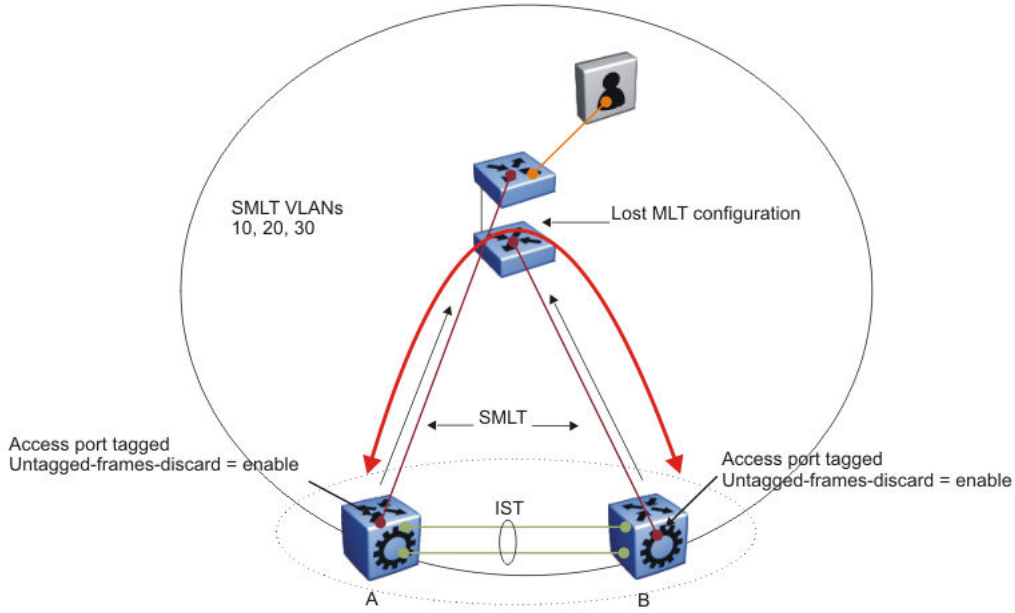


Figure 37: Tagged SMLT links

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 14: SLPP recommended values

	Configuration
Enable SLPP	
Access SMLT	Yes
Core SMLT	No
IST	No
Primary switch	

Comments? infodev@avaya.com

	Configuration
Enable SLPP	
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 59

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 *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* NN46250-500.

Spanning tree and protection against isolated VLANs

Virtual Local Area Network (VLAN) isolation disrupts packet forwarding. Figure 38: VLAN isolation on page 90 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.

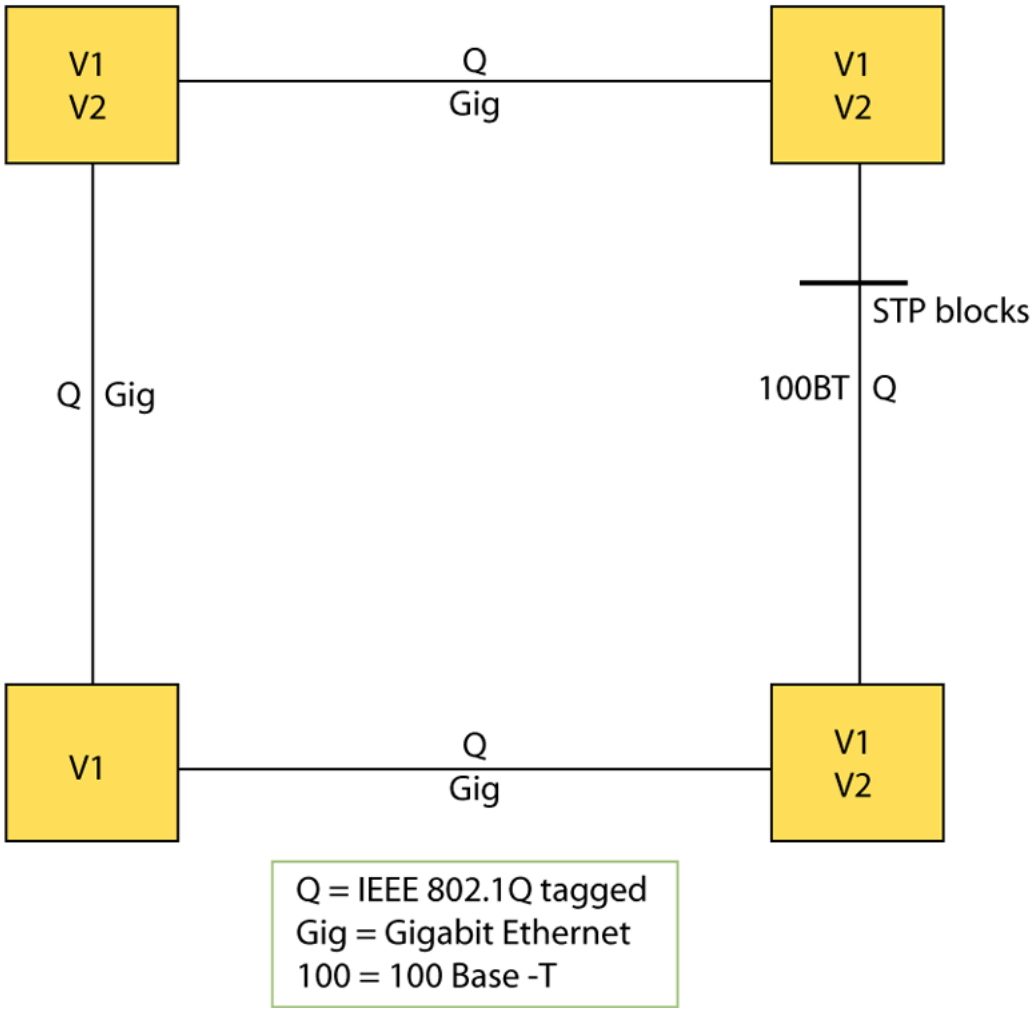


Figure 38: VLAN isolation

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 *Configuring IP Routing on Avaya Virtual Services Platform 9000,* 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.

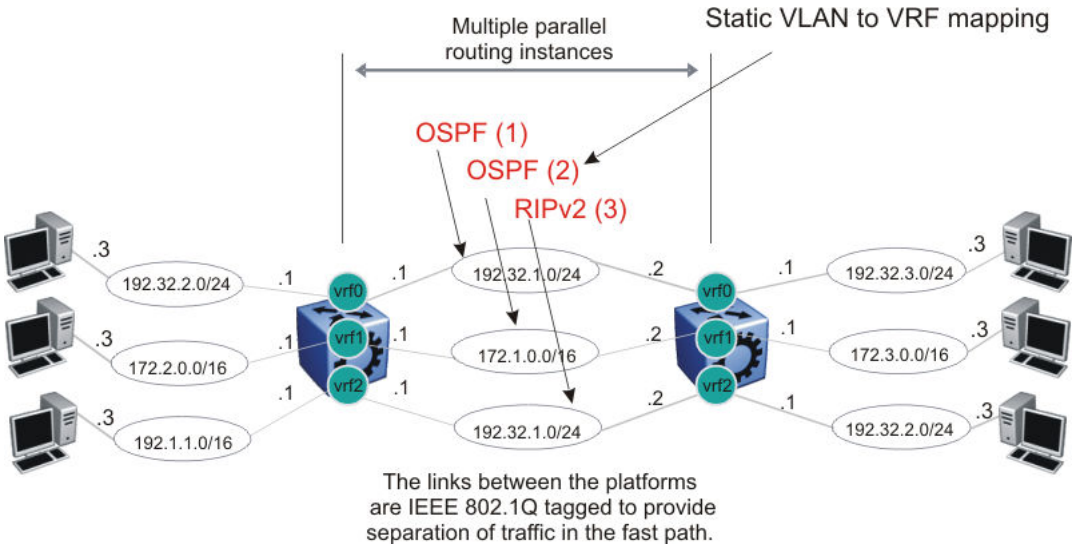


Figure 39: VRF Lite example

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

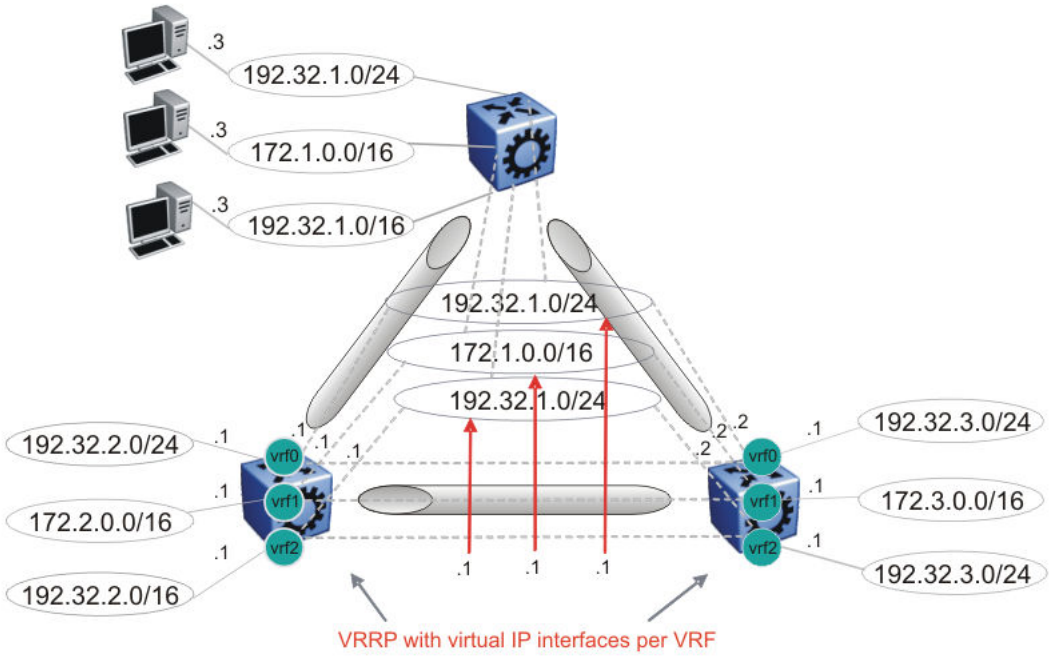


Figure 40: VRRP and VRF in SMLT topology

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

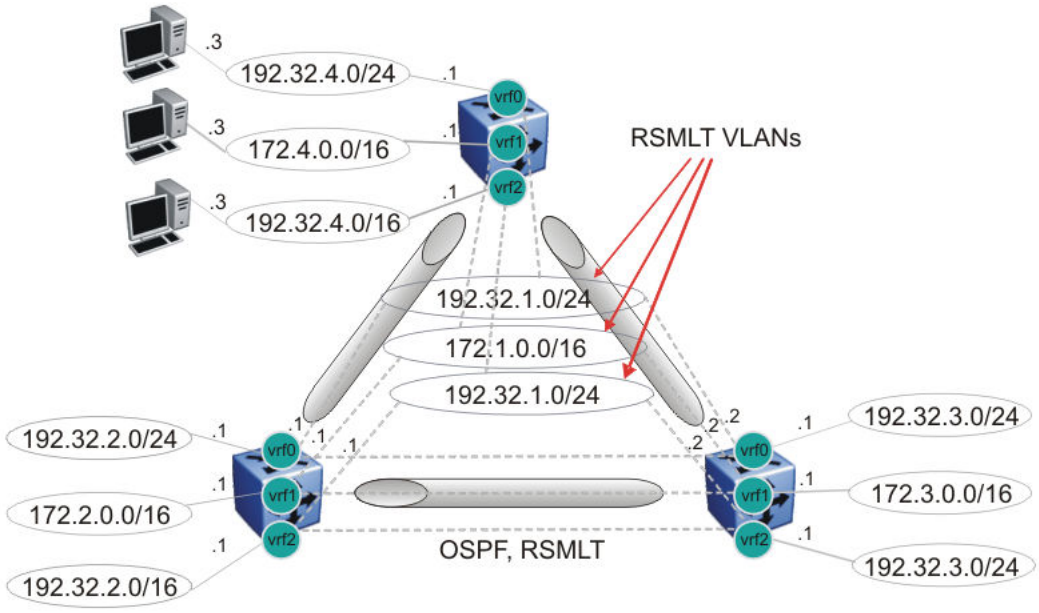


Figure 41: Router redundancy for multiple routing instances (using RSMLT)

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

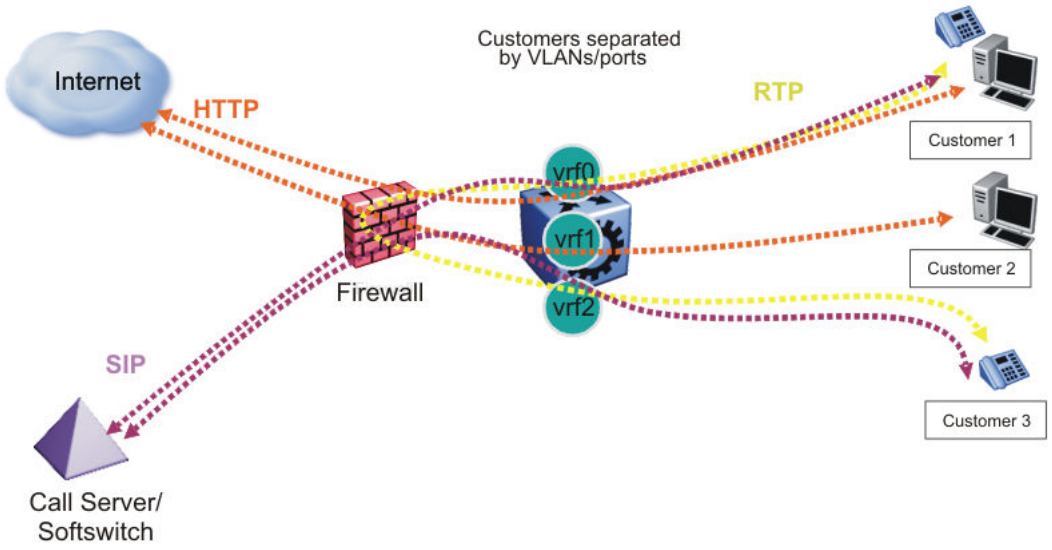


Figure 42: Inter-VRF forwarding based on external firewall

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.

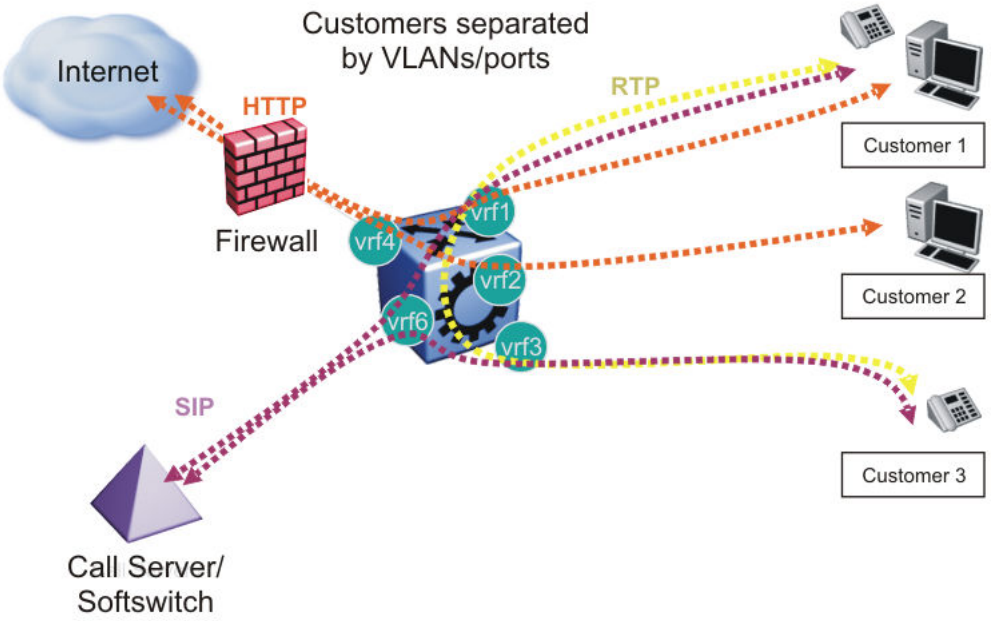


Figure 43: Inter VRF communication, internal inter-VRF forwarding

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.

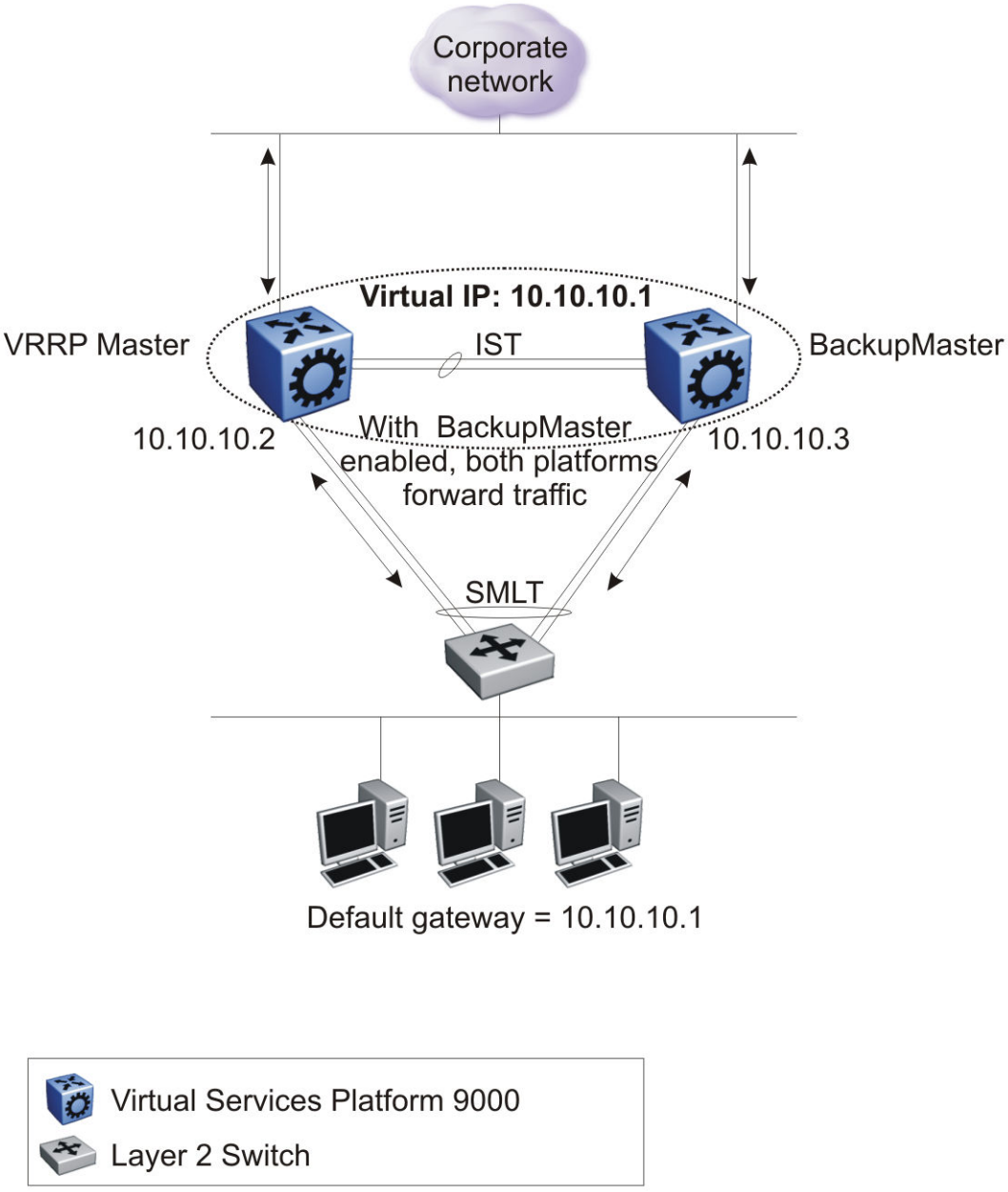


Figure 44: VRRP with BackupMaster

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 *Configuring IP Routing on Avaya Virtual Services Platform 9000,* NN46250-505.

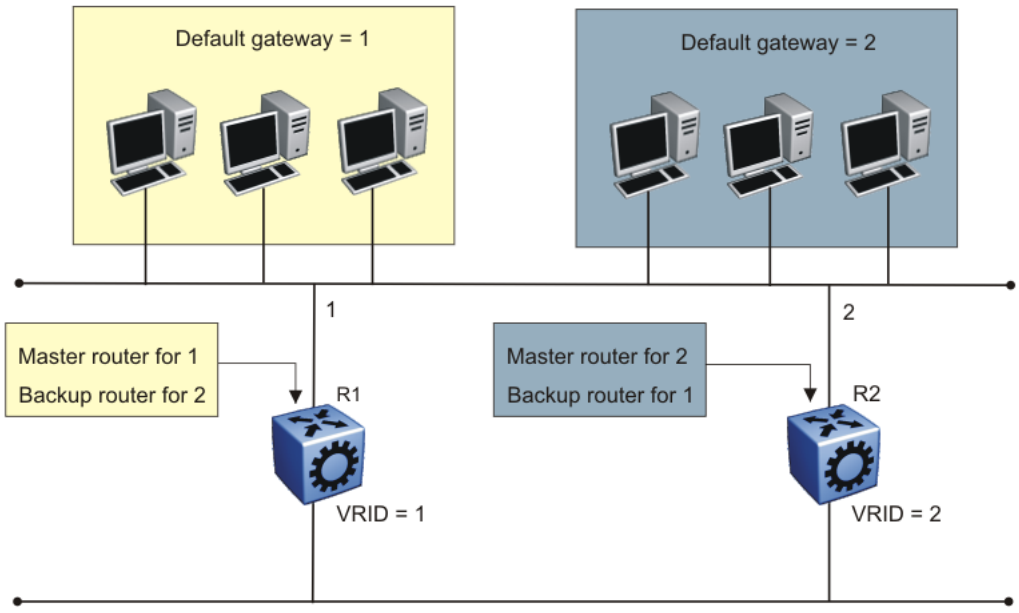


Figure 45: VRRP network configuration

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

  Note:

  Avaya does not support a VRRP virtual IP address to be the same as the local physical address of the device.
- 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.

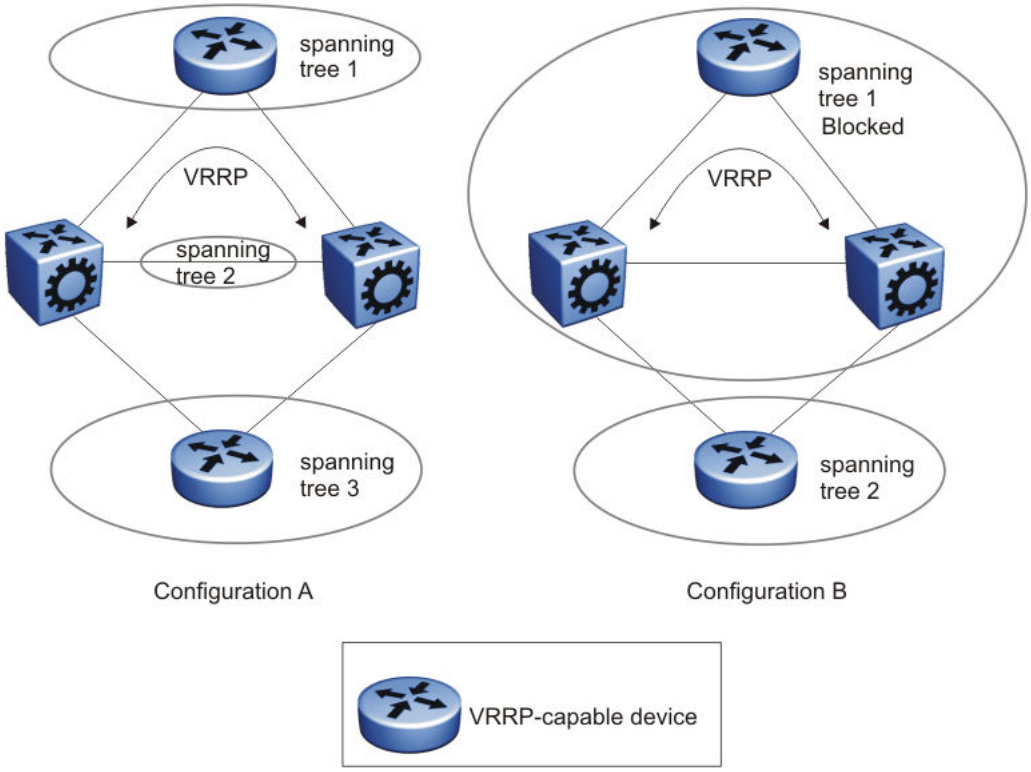


Figure 46: VRRP and STG configurations

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 61.

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.

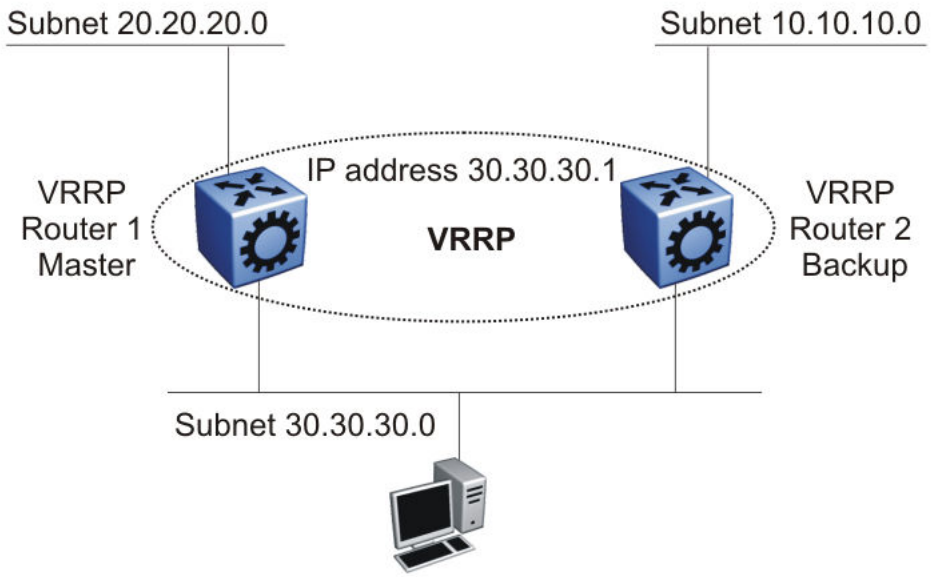


Figure 47: ICMP redirect messages

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 53.

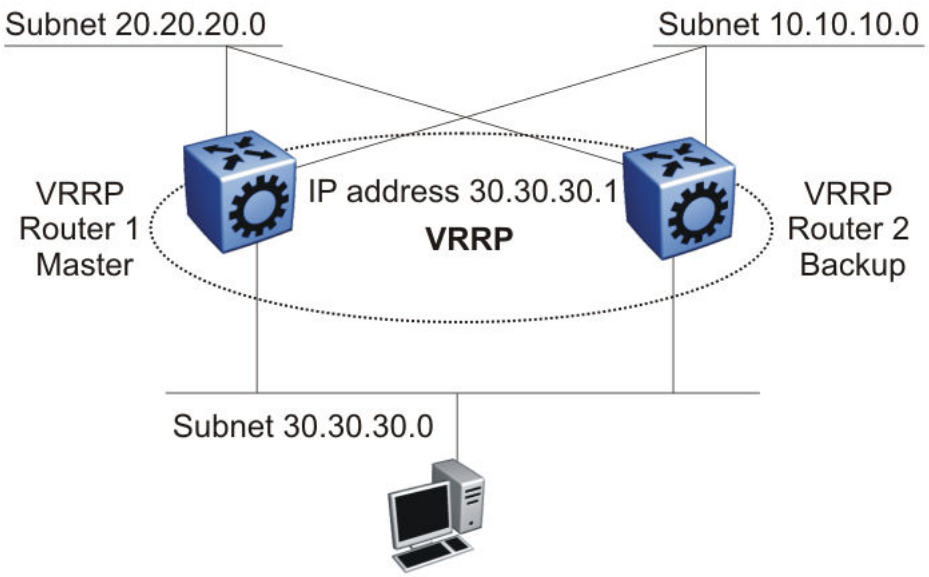


Figure 48: Avoiding excessive ICMP redirect messages without SMLT

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.

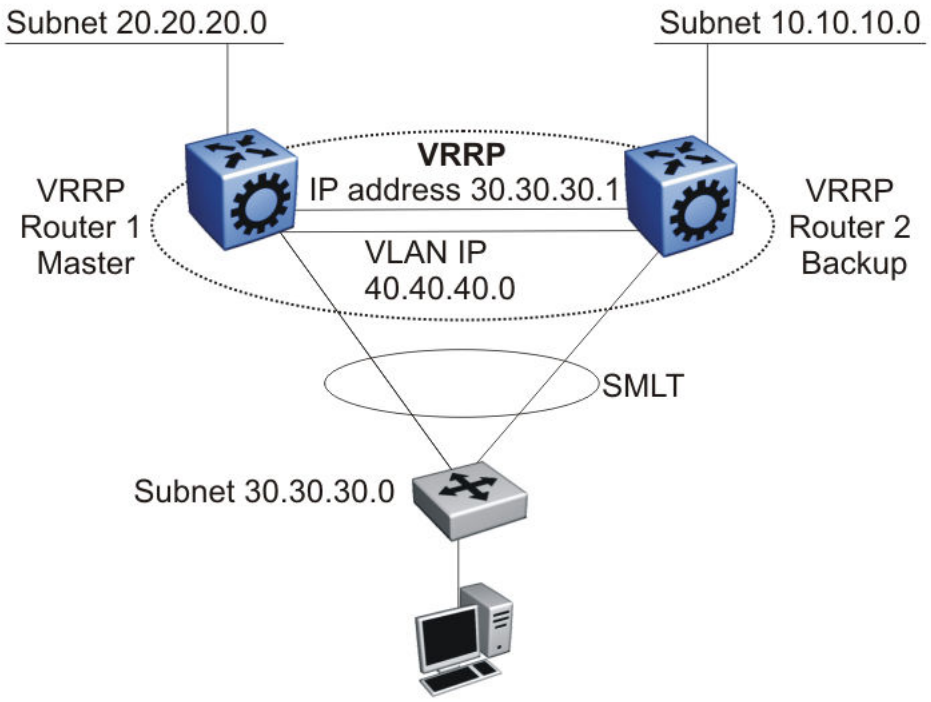


Figure 49: Avoiding excessive ICMP redirect messages with SMLT

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 67.

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 *Configuring OSPF and RIP on Avaya Virtual Services Platform 9000,* 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):

   $$\sum \mathrm{Adj}_N * \mathrm{LSA_CNT}_N < 40\mathrm{k}$$

   N = 1 to the number of areas for each switch

   Adj_N = number of adjacencies for each Area N

   $\mathrm{LSA_CNT}_N$ = 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 198.

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 50: 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.

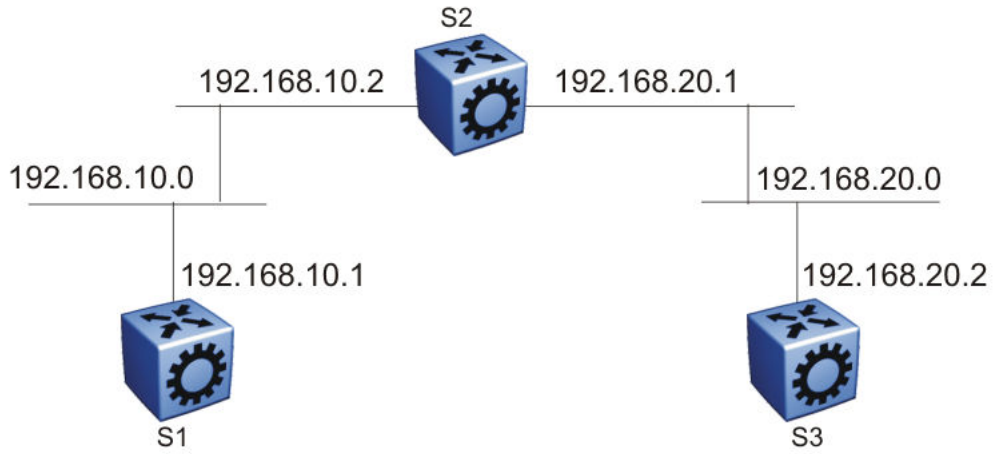


Figure 51: Example 2: OSPF on two subnets in one area

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.

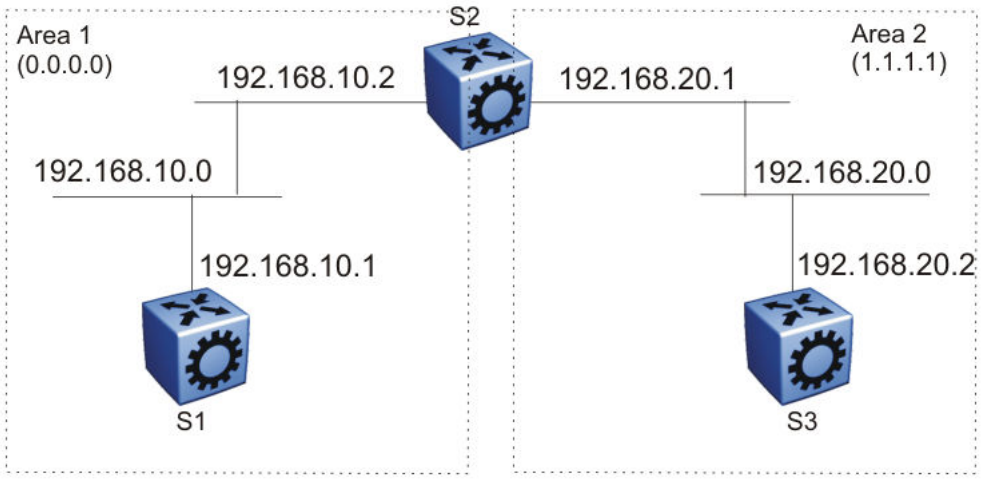


Figure 52: Example 3: OSPF on two subnets in two areas

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 *Configuring BGP Services on Avaya Virtual Services Platform 9000,* 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.

Comments? infodev@avaya.com

- 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 *Configuring BGP Services on Avaya Virtual Services Platform 9000,* 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.

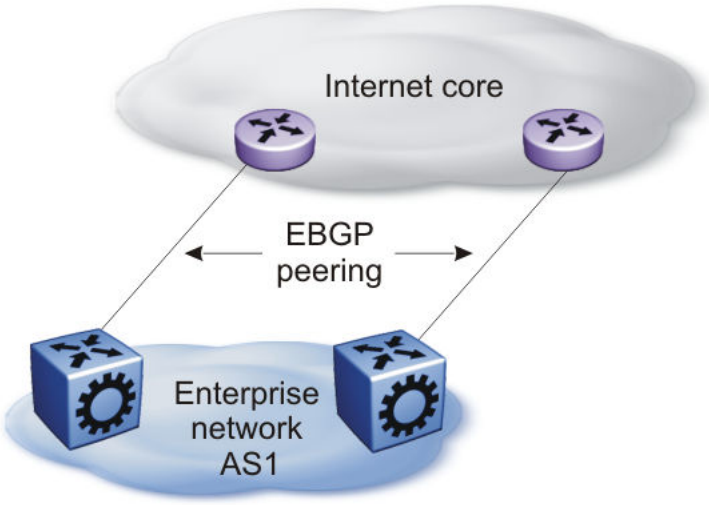


Figure 53: BGP and Internet peering

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).

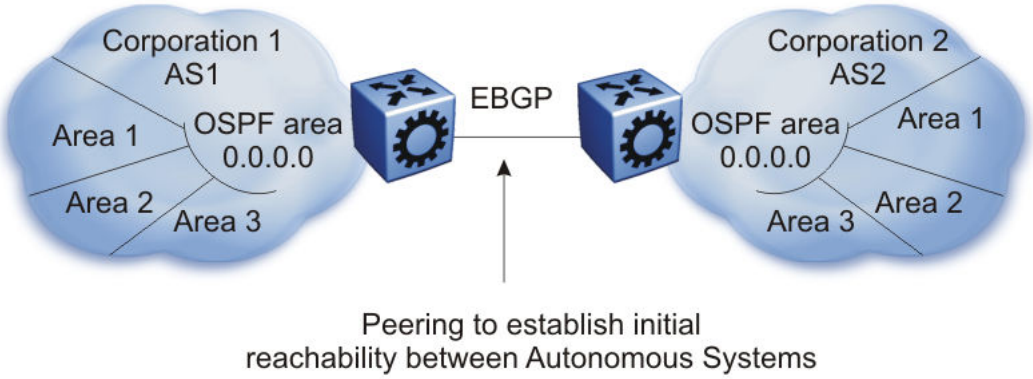


Figure 54: Routing domain interconnection with BGP

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).

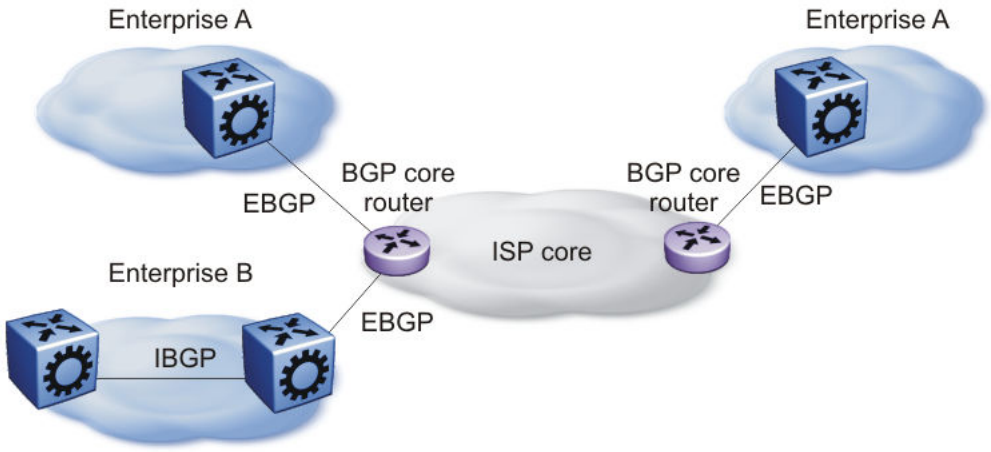


Figure 55: BGP and edge aggregation

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).

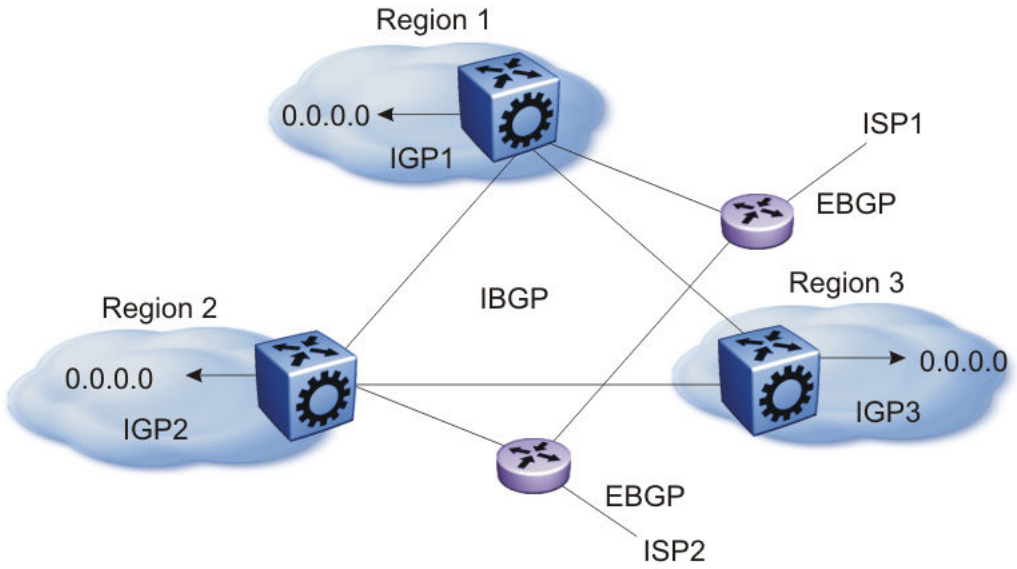


Figure 56: Multiple regions separated by IBGP

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.

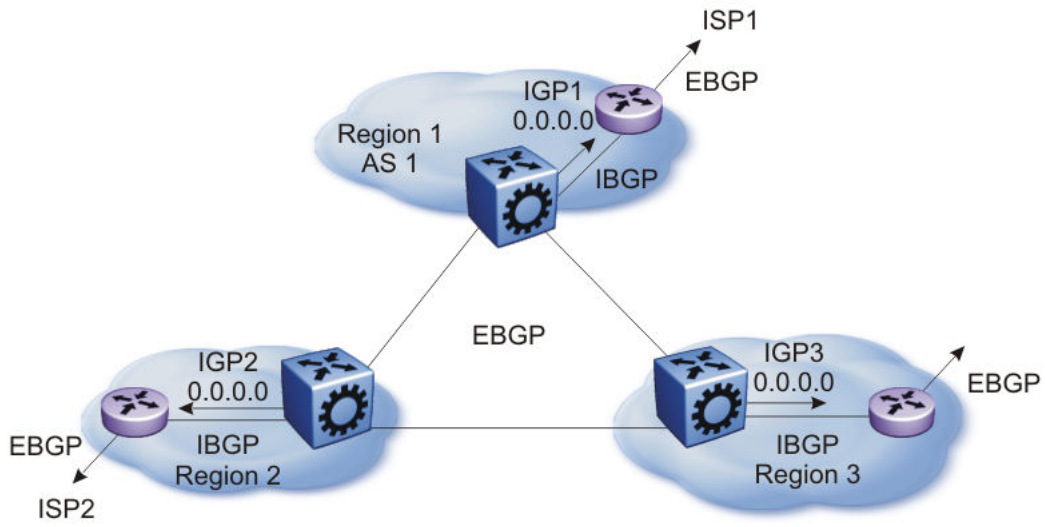


Figure 57: Multiple regions separated by EBGP

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.

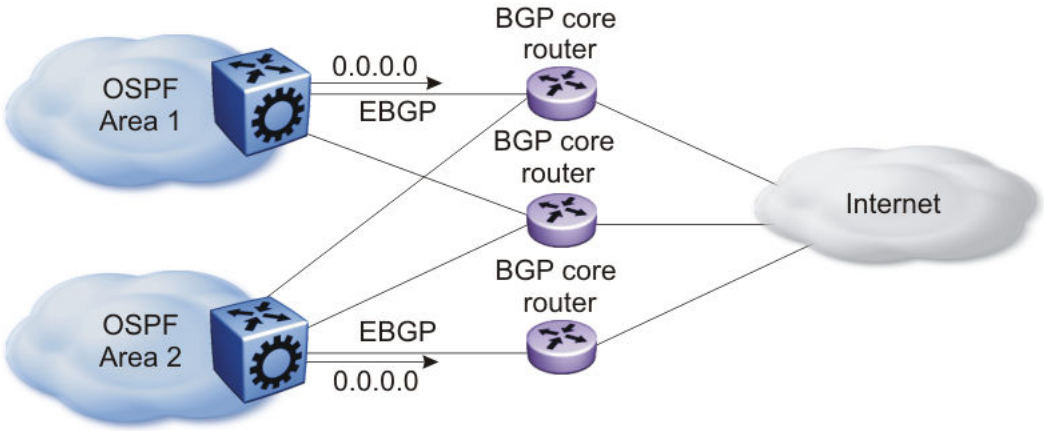


Figure 58: Multiple OSPF regions peering with the Internet

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 *Configuring IPv6 Routing on Avaya Virtual Services Platform 9000,* 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

The switch 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

The switch 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.

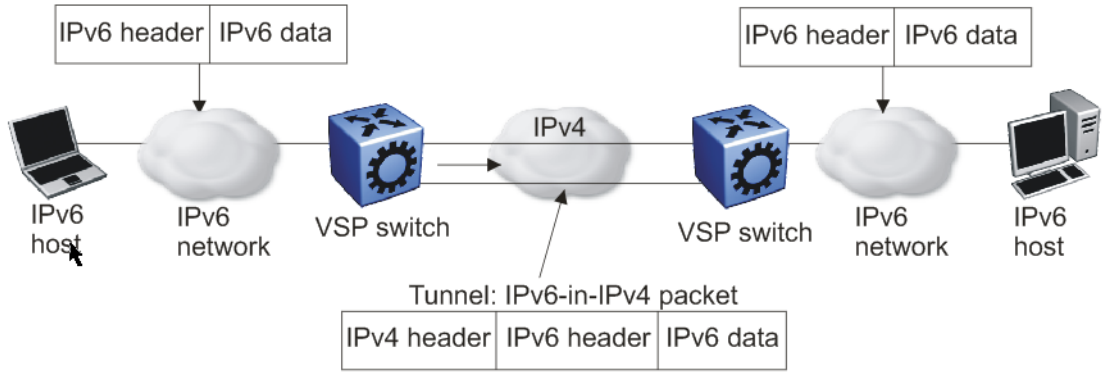


Figure 59: IPv6 over IPv4 tunneling

The tunnel in the preceding figure exists only between two VSP 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, the switch 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).

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.

Important:

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

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: SPBM design guidelines

Shortest Path Bridging MAC (SPBM) is a next generation virtualization technology that revolutionizes the design, deployment, and operations of enterprise campus core networks and data centers. The benefits of the technology are clearly evident in its ability to provide massive scalability while at the same time reducing the complexity of the network. SPBM makes network virtualization a much easier paradigm to deploy within the enterprise environment than other technologies.

This section provides design guidelines that illustrate the operational simplicity of SPBM. This section also lists best practices to configure SPBM in your network. For more information about SPBM, see the following document:

- For information about fundamental concepts, command structure, and basic configurations, see *Configuring Avaya VENA Fabric Connect on Avaya Virtual Services Platform 9000,* NN46250-510.

802.1aq standard

Virtual Services Platform 9000 supports the IEEE 802.1aq standard of SPBM. SPBM makes network virtualization easy to deploy within the enterprise environment by reducing the complexity of the network while at the same time providing greater scalability. This technology provides all the features and benefits required by carrier-grade deployments to the enterprise market without the complexity of alternative technologies traditionally used in carrier deployments, for example, Multiprotocol Label Switching (MPLS). SPBM integrates into a single control plane all the functions that MPLS requires multiple layers and protocols to support.

IS-IS

SPBM eliminates the need for multiple overlay protocols in the core of the network by reducing the core to a single Ethernet-based, link-state protocol (IS-IS). IS-IS provides virtualization services, both Layer 2 and Layer 3, using a pure Ethernet technology base. SPBM also uses IS-IS to discover and advertise the network topology, which enables it to compute the shortest path to all nodes in the SPBM network.

Spanning Tree is a topology protocol that prevents loops but does not scale very well. Because SPBM uses IS-IS, which has its own mechanisms to prevent loops, SPBM does not have to use Spanning Tree to provide a loop free Layer 2 domain.

SPBM uses the IS-IS shortest path trees to populate forwarding tables for the individual backbone MAC (B-MAC) addresses of each participating node. Depending on the topology, SPBM supports as many Equal Cost Multipath trees as there are backbone VLAN IDs (B-VIDs) provisioned (with a

maximum of 16 B-VIDs allowed by the standard and 2 allowed in this in release) per IS-IS instance. IS-IS interfaces operate in point-to-point mode only, which means that for any given port or MLT interface where IS-IS has been enabled, there can be only one IS-IS adjacency across that interface.

B-MAC

An SPBM backbone includes Backbone Edge Bridges (BEB) and Backbone Core Bridges (BCB). A BEB performs the same functionality as a BCB, but it also terminates one or more Virtual Service Networks (VSNs). A BCB does not terminate any VSNs and is unaware of the VSN traffic it transports. A BCB simply knows how to reach any other BEB in the SPBM backbone.

To forward customer traffic across the service provider backbone, the BEB for the VSN encapsulates the customer Ethernet packet received at the edge into a Backbone MAC header using the 802.1ah MAC-in-MAC encapsulation. This encapsulation hides the Customer MAC (C-MAC) address in a Backbone MAC (B-MAC) address pair. MAC-in-MAC encapsulation defines a BMAC-DA and BMAC-SA to identify the backbone source and destination addresses. The originating node creates a MAC header for delivery from end to end. Intermediate BCB nodes within the SPBM backbone perform packet forwarding using BMAC-DA alone. When the packet reaches the intended egress BEB, the B-MAC header is removed and the original customer packet is forwarded.

I-SID

SPBM introduces a service instance identifier called I-SID. SPBM uses I-SIDs to separate services from the infrastructure. After you create an SPBM infrastructure, you can add additional services (such as VLAN extensions or VRF extensions) by provisioning the endpoints only. The SPBM endpoints are BEBs, which mark the boundary between the core MAC-in-MAC SPBM domain and the edge customer 802.1Q domain. I-SIDs are provisioned on the BEBs to be associated with a particular service instance. In the SPBM core, the bridges are BCBs. BCBs forward encapsulated traffic based on the BMAC-DA.

The SPBM B-MAC header includes an I-SID. The length of the I–SID is 32 bits with a 24–bit ID. I-SIDs identify and transmit virtualized traffic in an encapsulated SPBM frame. These I-SIDs are used in a VSN for VLANs or VRFs across the MAC-in-MAC backbone:

- For a Layer 2 VSN, the I-SID is associated with a customer VLAN, which is then virtualized across the backbone. Layer 2 VSNs offer an any-any LAN service type. Layer 2 VSNs associate one VLAN per I-SID.
- For a Layer 2 VSN with multicast, the BEB associates a data I-SID with the multicast stream and a scope I-SID that defines the scope as a Layer 2 VSN. A multicast stream with a Layer 2 VSN scope can only transmit a multicast stream for the same Layer 2 VSN.
- For a Layer 3 VSN, the I-SID is associated with a customer VRF, which is also virtualized across the backbone. Layer 3 VSNs are always full-mesh topologies. Layer 3 VSNs associate one VRF per I-SID.
- For a Layer 3 VSN with multicast, the BEB associates a data I-SID with the multicast stream and a scope I-SID that defines the scope as a Layer 3 VSN. A multicast stream with a Layer 3 VSN scope can only transmit a multicast stream for the same Layer 3 VSN.
- For IP Shortcuts with multicast, the BEB associates a data I-SID with the multicast stream and defines the scope as Layer 3 GRT. A multicast stream with a scope of Layer 3 GRT can only transmit a multicast stream for the Layer 3 GRT.

Encapsulating customer MAC addresses in backbone MAC addresses greatly improves network scalability (no end-user C-MAC learning required in the core) and also significantly improves network robustness (loops have no effect on the backbone infrastructure).

The following figure shows the components of a basic SPBM architecture.

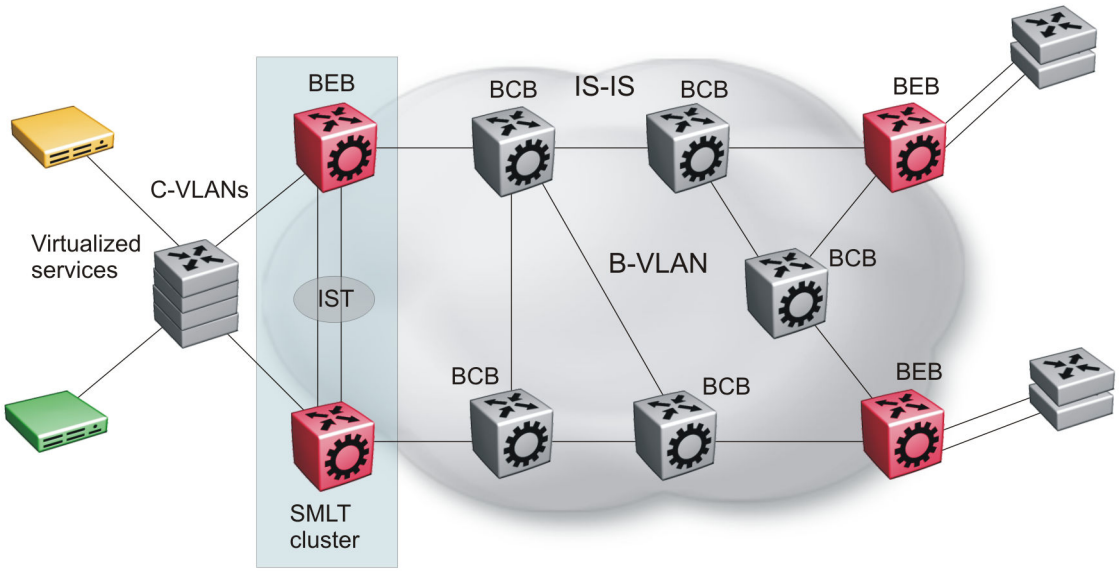


Figure 60: SPBM basic architecture

VLANs without member ports

The Avaya Ethernet Routing Switch 8800 manages VLANs without member ports differently than the Virtual Services Platform 9000 and Virtual Services Platform 4000.

- The Ethernet Routing Switch 8800 always designates the VLAN as operationally up if there is an attached I-SID.
- The Virtual Services Platform 9000 and Virtual Services Platform 4000 designate the VLAN as operationally up only if there is a matching I-SID in the SPBM network. For more information, see the following sections.

Ethernet Routing Switch 8800 implementation

If a VLAN has an IP address and is attached to an I-SID, the ERS 8800 designates that VLAN as operationally up whether it has a member port or not. When the VLAN is operationally up, the IP address of the VLAN will be in the routing table.

The ERS 8800 design behaves this way because the VLAN might be acting as an NNI in cases of Layer 2 Inter-VSN routing. If the VLAN was acting as a UNI interface, it would require a member port.

Virtual Services Platform 9000 and Virtual Services Platform 4000 implementation

If a VLAN is attached to an I-SID there must be another instance of that same I-SID in the SPBM network.

- If another instance of that I-SID exists, the device designates that VLAN as operationally up regardless of whether it has a member port or not.

  When the VLAN is operationally up, the IP address of the VLAN will be in the routing table.
- If no matching instance of the I-SID exists in the SPBM network, then that VLAN has no reachable members and does not act as an NNI interface.

  The VLAN does not act as a UNI interface because it does not have a member port.

  Therefore, the device does not designate the VLAN as operationally up because the VLAN does not act as a UNI or an NNI interface.

If the device acts as a BCB with two VLANs configured and two I-SIDs, there must be a UNI side with the corresponding I-SID existing in the network.

If the device acts as both BEB and BCB, then there must be a member port in that VLAN to push out the UNI traffic.

Provisioning

This section summarizes how to provision SPBM. For information about specific configuration commands, see *Configuring Avaya VENA Fabric Connect on Avaya Virtual Services Platform 9000,* NN46250-510.

Infrastructure provisioning

Provisioning an SPBM core is as simple as enabling SPBM and IS-IS globally, and on all the IS-IS core Ethernet links on all the BCB and BEB nodes. The IS-IS protocol operates at Layer 2 so it does not need IP addresses configured on the links to form IS-IS adjacencies with neighboring switches (like OSPF does). You do not need to configure IP addresses on any of the core links. The encapsulation of customer MAC addresses in backbone MAC addresses greatly improves network scalability.

No flooding or learning of end-user MACs occurs in the backbone. This SPBM provisioning significantly improves network robustness, as customer-introduced network loops have no effect on the backbone infrastructure.

Service provisioning

Provision I-SIDs on a BEB to associate that BEB with a particular service instance. After you map the customer VLAN or VRF into an I-SID, any BEB that has the same I-SID configured can participate in the same Layer 2 or Layer 3 VSN. This same simplicity extends to provisioning the services to run above the SPBM backbone:

- To create a Layer 2 VSN, associate an I-SID number with an edge VLAN.
- To create a Layer 3 VSN, associate an I-SID number with a VRF and configure the desired IS-IS IP route redistribution within the newly created Layer 3 VSN.

Note:

No service provisioning is needed on the core BCB SPBM switches. This provides a robust carrier grade architecture where configuration on the core switches never needs to be touched when adding new services.

IP multicast over SPBM

Provisioning IP multicast over SPBM is as simple as enabling SPBM multicast on the BEBs. You do not need to enable IP multicast over SPBM on the BCBs.

For Layer 2 VSN using IP multicast over SPBM, configure IGMP snooping on the VLAN that represents the Layer 2 VSN.

For Layer 3 VSNs using IP multicast over SPBM, configure the Layer 3 VSN as a multicast VSN, and then enable IP multicast over SPBM on each VLAN within the VRF to which IP multicast senders and receivers attach.

For IP Shortcuts using IP multicast over SPBM, enable IP multicast over SPBM on each of the VLANs within the GRT that need to support IP multicast traffic.

Note:

SPBM supports Network Load Balancing (NLB) unicast. SPBM does not support NLB multicast or NLB multicast with IGMP.

Implementation options

The SPBM architecture is architecturally simple and easy to provision, but it is not just for simple networks. SPBM supports multiple implementation options within the same network to meet the demands of the most complex network configurations. The following figure shows how SPBM supports multiple campus networks as well as multiple data centers.

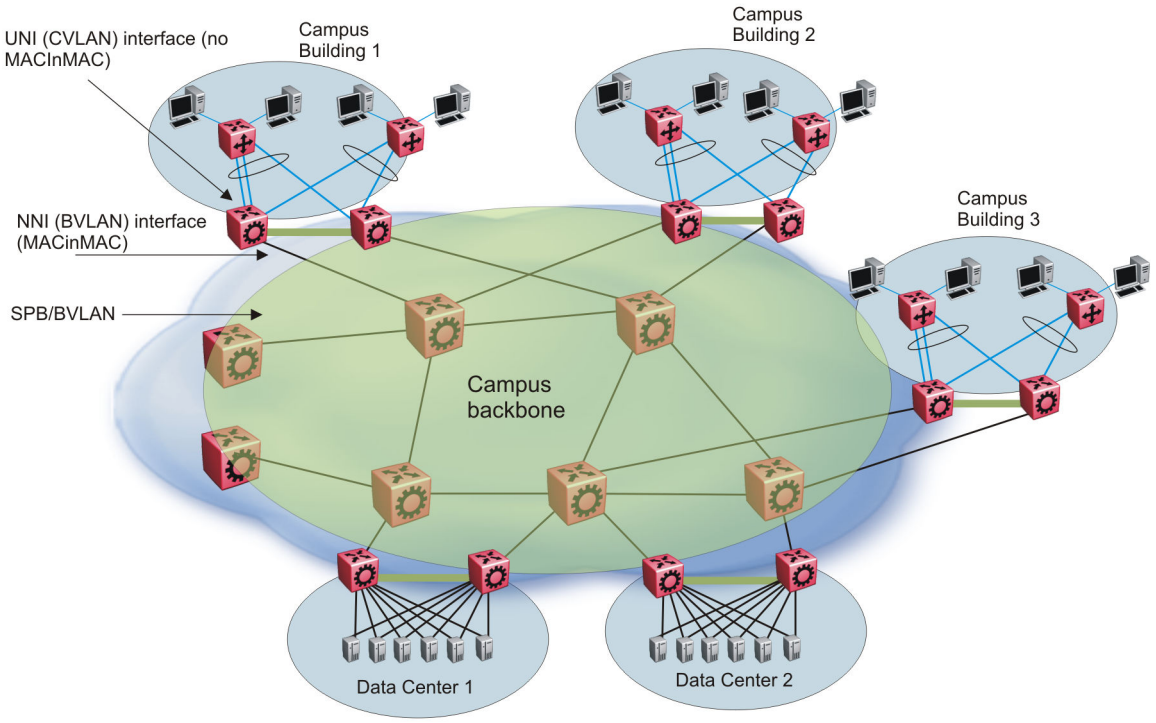


Figure 61: SPBM support for campus and data center architecture

Within the SPBM architecture, you can implement multiple options. The following figure shows all the options that SPBM supports. The current release of Virtual Services Platform 9000 does not support IPVPN-Lite over SPBM.

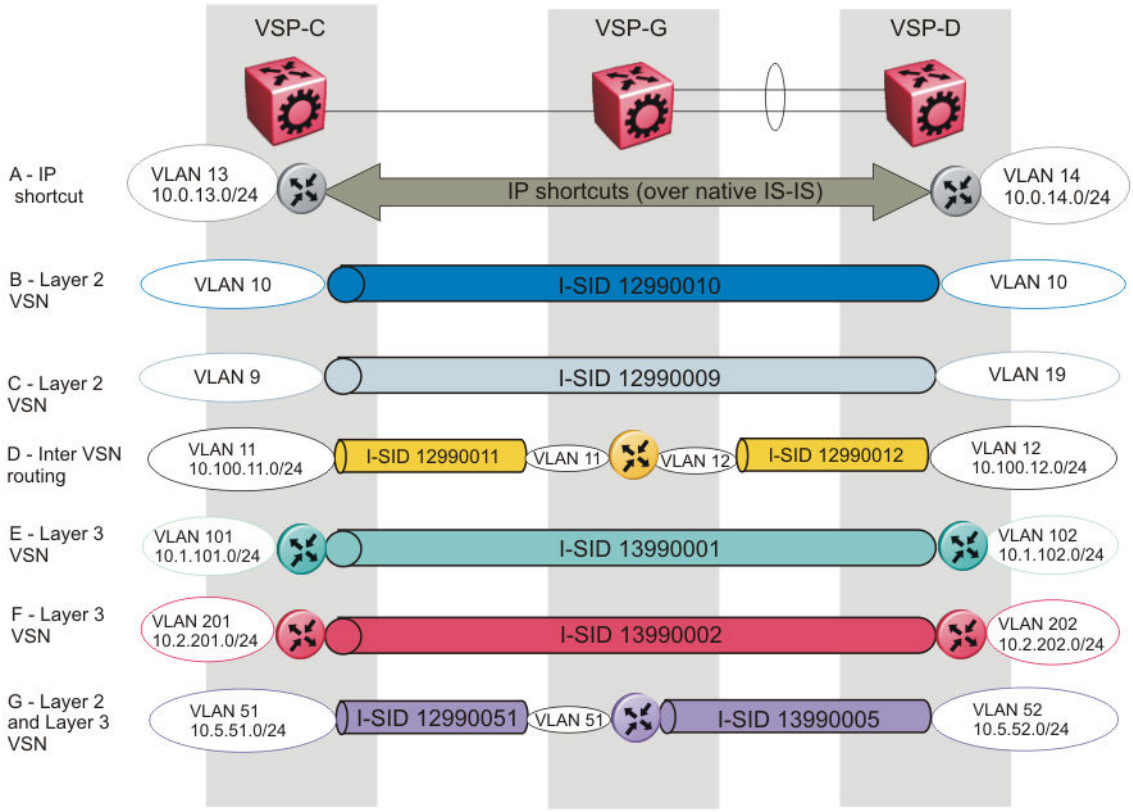


Figure 62: SPBM implementation options

A—IP shortcut

IP shortcuts forward standard IP packets over IS-IS. This option enables you to forward IP over the SPBM core, which is a simpler method than traditional IP routing or MPLS. SPBM nodes propagate Layer 3 reachability as leaf information in the IS-IS link state packets (LSP) using Extended IP reachability TLV 135, which contains routing information such as neighbors and locally configured subnets. SPBM nodes that receive the reachability information use this information to populate the routes to the announcing nodes. All TLVs announced in the IS-IS LSPs are grafted onto the shortest path tree (SPT) as leaf nodes.

An IP route lookup is only required once where the source BEB uses the routing table to identify the BEB closest to the destination subnet. All other nodes perform standard Ethernet switching based on the existing SPT. This scenario allows for end to end IP-over-Ethernet forwarding without the need for Address Resolution Protocol (ARP), flooding, or reverse learning. Because BCB SPBM nodes only forward on the MAC addresses that comprise the B-MAC header, and because unknown TLVs in IS-IS are relayed to the next hop but ignored locally, SPBM BCB nodes need not be aware of IP subnets to forward IP traffic. Only BEBs generate and receive Extended IP reachability TLV to build the routing table; BCBs just relay the TLV to the next hop based on the SPT. In fact, the Extended IP reachabilty TLV is ignored on BCBs.

With IP shortcuts there are only two IP routing hops (ingress BEB and egress BEB) as the SPBM backbone acts as a virtualized switching backplane.

IP shortcuts do not require I-SID configuration. However, you must enable IP on IS-IS, and configure the IS-IS source address to match a circuitless or loopback IP address.

In Figure 62: SPBM implementation options on page 123, node VSP-G acts as a BCB for the service, and has no IP configuration.

B—Layer 2 VSN

A Layer 2 Virtual Services Network (VSN) bridges customer VLANs (C-VLANs) over the SPBM core infrastructure. A Layer 2 VSN associates a C-VLAN with an I-SID, which is then virtualized across the backbone. All VLANs in the network that share the same I-SID can participate in the same VSN. If you use Split MultiLink Trunking (SMLT) clusters or if you want IS-IS to distribute traffic across two equal cost paths, then you need two backbone VLANs (B-VLAN) with a primary B-VLAN and a secondary B-VLAN. Otherwise, you need only a single B-VLAN.

One of the key advantages of the SPBM Layer 2 VSN is that network virtualization provisioning is achieved by configuring the edge of the network (BEBs) only. The intrusive core provisioning that other Layer 2 virtualization technologies require is not needed when new connectivity services are added to the SPBM network. For example, when new virtual server instances are created and need their own VLAN instances, they are provisioned at the network edge only and do not need to be configured throughout the rest of the network infrastructure.

Based on its I-SID scalability, this solution can scale much higher than any 802.1Q tagging based solution. Also, due to the fact that there is no need for Spanning Tree in the core, this solution does not need any core link provisioning for normal operation. Redundant connectivity between the C-VLAN domain and the SPBM infrastructure can be achieved by operating two SPBM switches in switch clustering (SMLT) mode. This allows the dual homing of any traditional link aggregation capable device into an SPBM network

In Figure 62: SPBM implementation options on page 123, nodes VSP-C and VSP-D act as BEBs for the VSN. Only these nodes have a MAC table or forwarding database for C-VLAN 10.

C—Layer 2 VSN with VLAN translation

Layer 2 VSNs with VLAN translation are basically the same as the Layer 2 VSNs, except that the BEBs on either end of the SPBM network belong to different VLANs. This option enables you to connect one VLAN to another VLAN. In Figure 62: SPBM implementation options on page 123, VLAN 9 connects to VLAN 19. The mechanism that connects them is that they use the same I-SID (12990009).

D— Inter-VSN routing

Inter-VSN routing allows routing between Layer 2 VLANs with different I-SIDs. You can use Inter-VSN routing to redistribute routes between Layer 2 VLANs. This option allows effective networking of multiple VSNs. Where Layer 2 VSN with VLAN translation enabled you to interconnect VLANs, this option takes that concept one step further and allows you to interconnect VSNs. This option also provides the ability to route IP traffic on Layer 2-VSNs that enter on NNI interfaces, which is especially useful for Layer 2 edge solutions.

As seen in Figure 62: SPBM implementation options on page 123, routing between VLANs 11 and 12 occurs on the SPBM core switch VSP-G shown in the middle of the figure. With Inter-VSN routing enabled, VSP-G transmits traffic between VLAN 11 (I-SID 12990011) and VLAN 12 (I-SID 12990012) on the VRF instance configured. Note that for these VSNs, node VSP-G acts as a BEB.

E—Layer 3 VSN

Layer 3 VSNs are very similar to Layer 2 VSNs. The difference between the two is that Layer 2 VSNs associate I-SIDs with VLANs. Layer 3 VSNs associate I-SIDs with VRFs. With the Layer 3 VSN option, all VRFs in the network that share the same I-SID can participate in the same VSN by

advertising their reachable IP routes into IS-IS and installing IP routes learned from IS-IS. Suitable IP redistribution policies need to be defined to determine what IP routes a BEB will advertise to IS-IS.

As seen in Figure 62: SPBM implementation options on page 123, the green VRF on VSP-C is configured to advertise its local or direct IP routes into IS-IS within I-SID 13990001. The VRF on node VSP-D, which is also a member of the same I-SID, installs these IP routes in its VRF IP routing table with a next-hop B-MAC address of VSP-C. Therefore, when the VRF on node VSP-D needs to IP route traffic to the IP subnet off VSP-C, it performs a lookup in its IP routing table and applies a MAC-in- MAC encapsulation with B-MAC DA of VSP-C. The SPBM core ensures delivery to the egress BEB VSP-C where the encapsulation is removed and the packet is IP routed onwards.

Note:

Like the IP shortcut service, there are only two IP routing hops (ingress BEB and egress BEB) as the SPBM backbone acts as a virtualized switching backplane.

F—Layer 3 VSN

Figure 62: SPBM implementation options on page 123 shows two VRFs (green and red) to illustrate that the BEBs can associate ISIDs with multiple VRFs. The Layer 3 VSN option provides IP connectivity over SPBM for all of your VRFs.

G—Layer 2 VSN and Layer 3 VSN

Figure 62: SPBM implementation options on page 123 shows both a Layer 2 VSN and a Layer 3 VSN to show that you can configure both options on the same BEBs. This topology is simply made up of a number of BEBs that terminate VSNs of both types. What this example shows is the flexibility to extend one or more edge VLANs (using one or more Layer 2 VSNs) to use a default gateway that is deeper into the SPBM core. From here, traffic can then be IP routed onwards as either non-virtualized with IP shortcuts or, as shown in this example, with a virtualized Layer 3 VSN. Note that in this example the central node VSP-G is now also acting as BEB for both service types as it now maintains both a MAC table for the Layer 2 VSN it terminates, and an ARP cache and IP routing table for the Layer 3 VSN it also terminates.

Multiple tenants using different SPBM services

The following figure shows multiple tenants using different services within an SPBM metro network. In this network, you can use some or all of the SPBM implementation options to meet the needs of the community while maintaining the security of information within VLAN members.

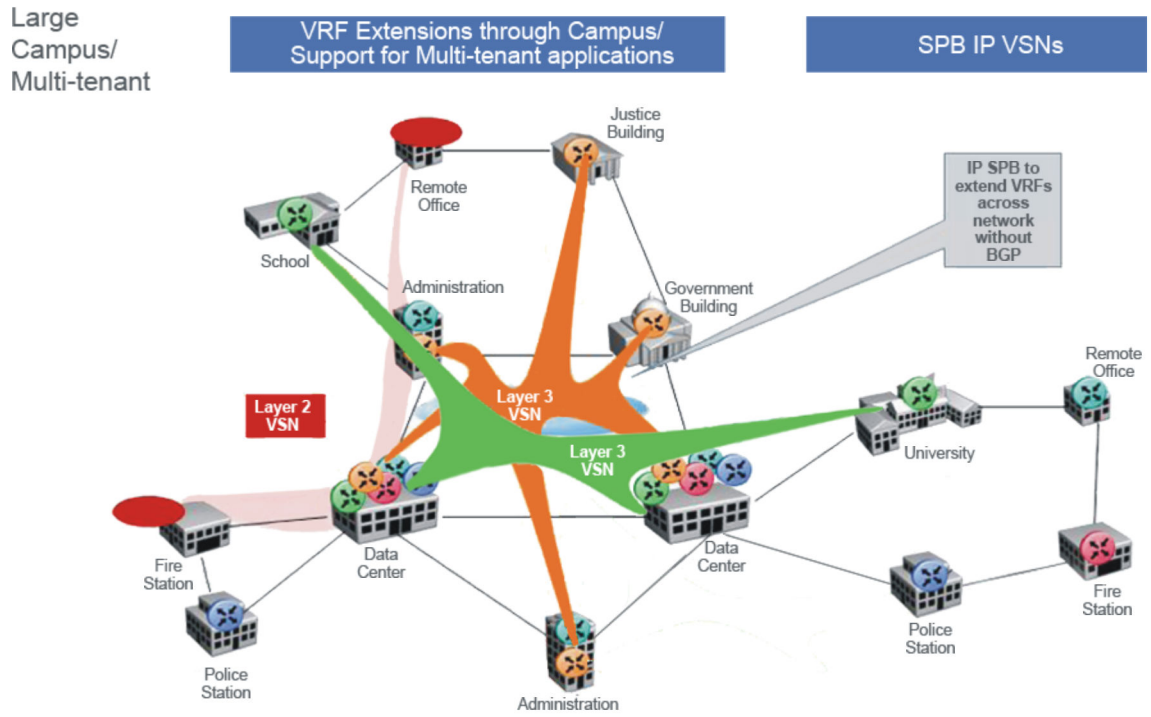


Figure 63: Multi-tenant SPBM metro network

To illustrate the versatility and robustness of SPBM even further, the following figure shows a logical view of multiple tenants in a ring topology. In this architecture, each tenant has their own domain where some users have VLAN requirements and are using Layer 2 VSNs and others have VRF requirements and are using Layer 3 VSNs. In all three domains, they can share data center resources across the SPBM network.

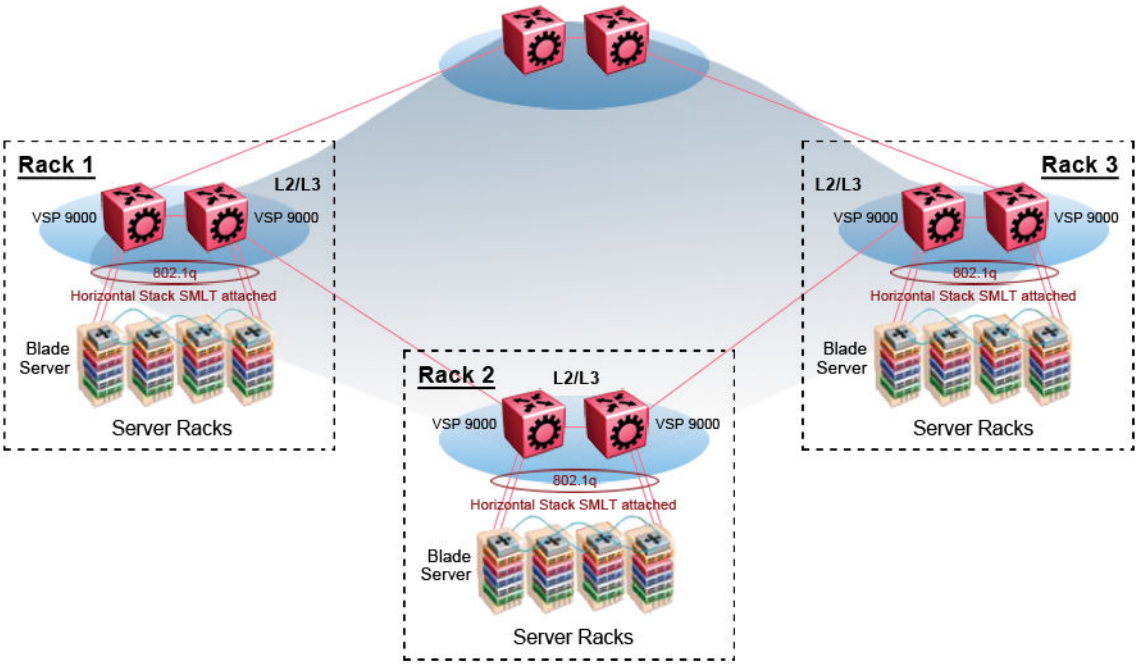


Figure 64: SPBM ring topology with shared data centers

Reference architectures

SPBM has a straightforward architecture that simply forwards encapsulated C-MACs across the backbone. Because the B-MAC header stays the same across the network, there is no need to swap a label or do a route lookup at each node. This architecture allows the frame to follow the most efficient forwarding path from end to end.

The following figure shows the MAC-in-MAC SPBM domain with BEBs on the boundary and BCBs in the core.

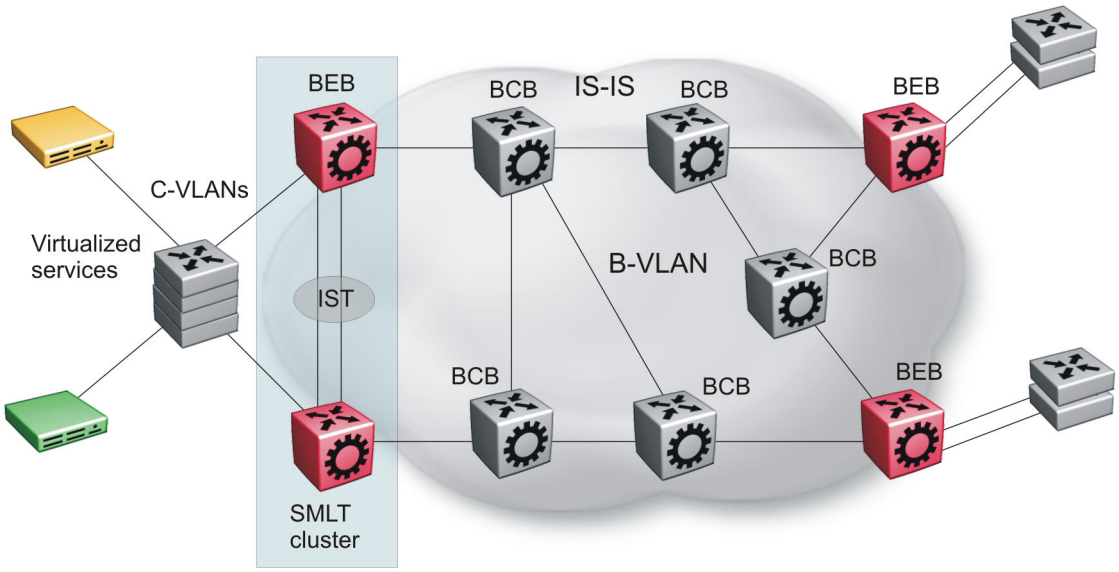


Figure 65: SPBM basic architecture

Provisioning an SPBM core is as simple as enabling SPBM and IS-IS globally on all the nodes and on the core facing links. To migrate an existing edge configuration into an SPBM network is just as simple.

The boundary between the MAC-in-MAC SPBM domain and the 802.1Q domain is handled by the BEBs. At the BEBs, VLANs or VRFs are mapped into I-SIDs based on the local service provisioning. Services (whether Layer 2 or Layer 3 VSNs) only need to be configured at the edge of the SPBM backbone (on the BEBs). There is no provisioning needed on the core SPBM nodes.

The following figure illustrates an existing edge that connects to an SPBM core.

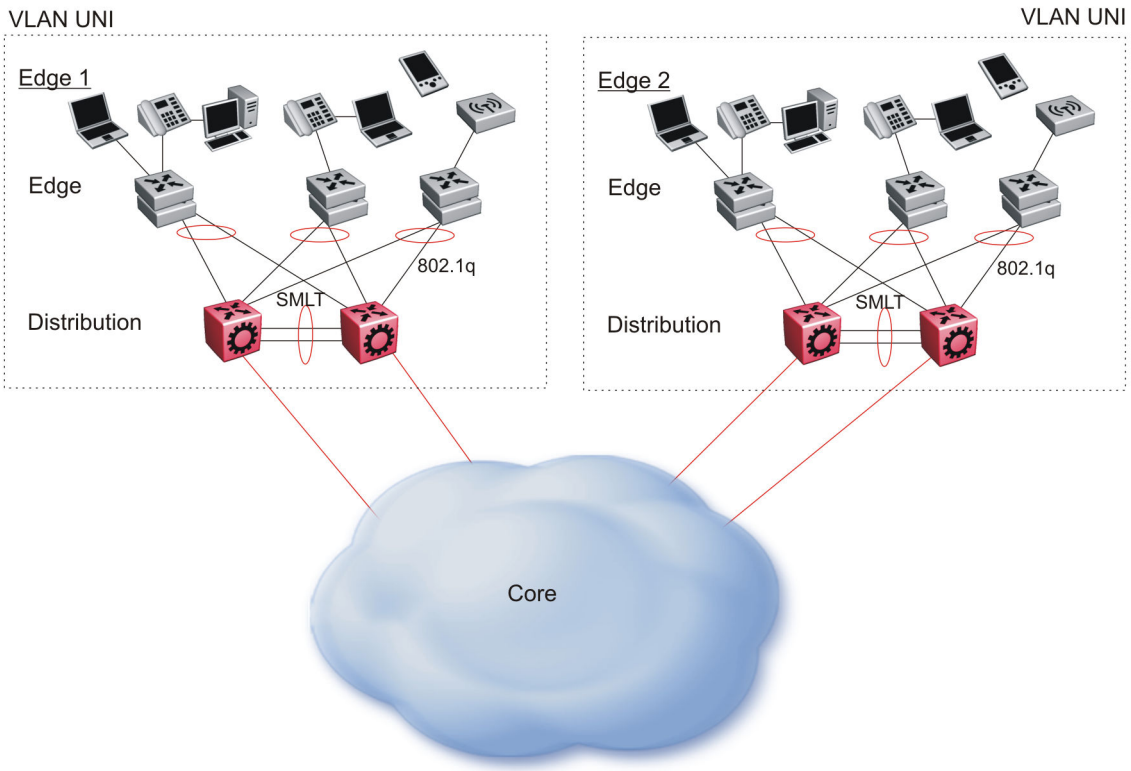


Figure 66: Access to the SPBM Core

For Layer 2 virtualized bridging (Layer 2 VSN), identify all the VLANs that you want to migrate into SPBM and assign them to an I-SID on the BEB.

For Layer 3 virtualized routing (Layer 3 VSN), map IPv4-enabled VLANs to VRFs, create an IP VPN instance on the VRF, assign an I-SID to the VRF, and then configure the desired IP redistribution of IP routes into IS-IS.

All BEBs that have the same I-SID configured can participate in the same VSN. That completes the configuration part of the migration and all the traffic flows will be back to normal.

SPBM on Virtual Services Platform 9000 supports the following traffic:

- Layer-2 bridged traffic (Layer 2 VSN)
- IPv4 unicast routed traffic on the Global Router (IP shortcuts)
- IPv4 unicast routed traffic using a VRF (Layer 3 VSN)
- IPv4 unicast routed traffic using different VSNs, which have different I-SIDs (Inter-VSN)
- Layer-2 IP multicast traffic in a bridged network (Layer 2 VSN with IP multicast over SPBM)
- IPv4 multicast routed traffic on the Global Router (IP shortcuts with IP multicast over SPBM)
- IPv4 multicast routed traffic using a VRF (Layer 3 VSN with IP multicast over SPBM)

SMLT

If your existing edge configuration uses SMLT, you can maintain that SMLT-based resiliency for services configured on the IST peer switches. SPBM requires that you upgrade both IST peer to the

current release and identify two VLANs to use as B-VLANs. SPBM then automatically creates a virtual backbone MAC for the IST pair, and advertises it with IS-IS. By operating two SPBM switches in switch clustering (SMLT) mode, you can achieve redundant connectivity between the C-VLAN domain and the SPBM infrastructure. This configuration allows the dual homing of any traditional link aggregation capable device into an SPBM network.

SMLT with IP multicast over SPBM

Layer 2 access switches use IGMP Snooping to prune multicast traffic. In IP multicast over SPBM, BEBs are the IGMP queriers, therefore access switches forward multicast data from the senders as well as IGMP control messages from receivers to the BEBs.

1. When a sender transmits multicast data to the Layer 2 access switch that has an MLT to the switch cluster, the multicast data is hashed towards one or the other BEBs in the switch cluster.
2. The receiving BEB allocates a Data I-SID and sends a TLV update on the primary B-VLAN, to announce the availability of the stream to its neighbors.
3. The BEB propagates the TLV update through the SPBM fabric in an LSP, so all BEBs are aware of this stream availability.
4. The sender information is also synchronized over the IST to the peer switch.
5. Then the peer switch allocates a Data I-SID for the multicast stream, and sends a TLV update on the secondary B-VLAN to announce the stream availability.

If a sender attaches to only one BEB in a switch cluster, the sender information does not synchronize over the IST because it is not SMLT attached. The sender information is advertised, and data is sent on either the primary or secondary B-VLAN The odd-numbered Data I-SIDs use the primary B-VLAN, and the even-numbered Data I-SIDs use the secondary B-VLAN.

For more information on switch clustering at the edge of the SPBM network, see *Avaya Virtual Services Platform 9000 Configuration — Shortest Path Bridging MAC (SPBM)*, NN46250–510.

Campus architecture

For migration purposes, you can add SPBM to an existing network that has SMLT configured. In fact, if there are other protocols already running in the network such as Open Shortest Path First (OSPF), you can leave them in place too. SPBM uses IS-IS, and operates independently from other protocols. However, Avaya recommends that you eventually eliminate SMLT in the core and other unnecessary protocols. This reduces the complexity of the network and makes it much simpler to maintain and troubleshoot.

Whether you configure SMLT in the core or not, the main point to remember is that SPBM separates services from the infrastructure. For example, in a large campus, a user may need access to other sites or data centers. SPBM enables you to grant that access by associating the user to a specific I-SID. This mechanism enables the user to work without getting access to confidential information of another department.

The following figure depicts a topology where the BEBs in the edge and data center distribution nodes are configured in SMLT clusters. Prior to implementing SPBM, the core nodes would also have been configured as SMLT clusters. When migrating SPBM onto this network design, it is important to note that you can deploy SPBM over the existing SMLT topology without network interruption. After the SPBM infrastructure is in place, you can create VSN services over SPBM or migrate them from the previous end-to-end SMLT-based design.

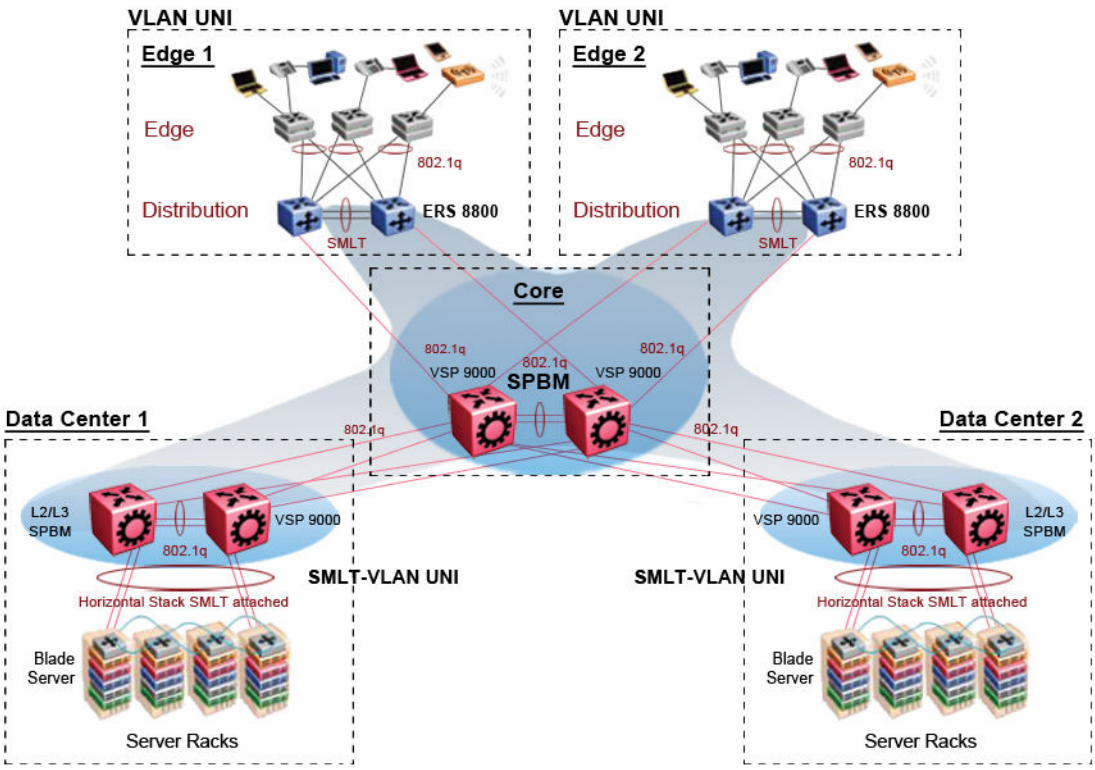


Figure 67: SPBM campus without SMLT

After migrating all services to SPBM, the customer VLANs (C-VLANs) will exist only on the BEB SMLT clusters at the edge of the SPBM network. The C-VLANs will be assigned to an I-SID instance and then associated with either a VLAN in an Layer 2 VSN or terminated into a VRF in an Layer 3 VSN. You can also terminate the C-VLAN into the default router, which uses IP shortcuts to IP route over the SPBM core.

In an SPBM network design, the only nodes where it makes sense to have an SMLT cluster configuration is on the BEB nodes where VSN services terminate. These are the SPBM nodes where C-VLANs exist and these C-VLANs need to be redundantly extended to non-SPBM devices such as Layer 2 edge stackable switches. On the BCB core nodes where no VSNs are terminated and no Layer 2 edge stackables are connected, there is no longer any use for the SMLT clustering functionality. Therefore, in the depicted SPBM design, the SMLT/IST configuration can be removed from the core nodes because they now act as pure BCBs that have no knowledge of the VSN they transport and the only control plane protocol they need to run is IS-IS.

Because SMLT BEB nodes exist in this design (the edge BEBs) and it is desirable to use equal cost paths to load balance VSN traffic across the SPBM core, all SPBM nodes in the network are configured with the same two B-VIDs.

Where Figure 67: SPBM campus without SMLT on page 131 shows the physical topology, the following two figures illustrate a logical rendition of the same topology. In both of the following figures, you can see that the core is almost identical. Because the SPBM core just serves as a transport mechanism that transmits traffic to the destination BEB, all the provisioning is done at the edge.

In the data center, VLANs are attached to Inter-VSNs that transmit the traffic across the SPBM core between the data center on the left and the data center on the right. A common application of this service is VMotion moving VMs from one data center to another.

The following figure uses IP shortcuts that route VLANs. There is no I-SID configuration and no Layer 3 virtualization between the edge distribution and the core. This is normal IP forwarding to the BEB.

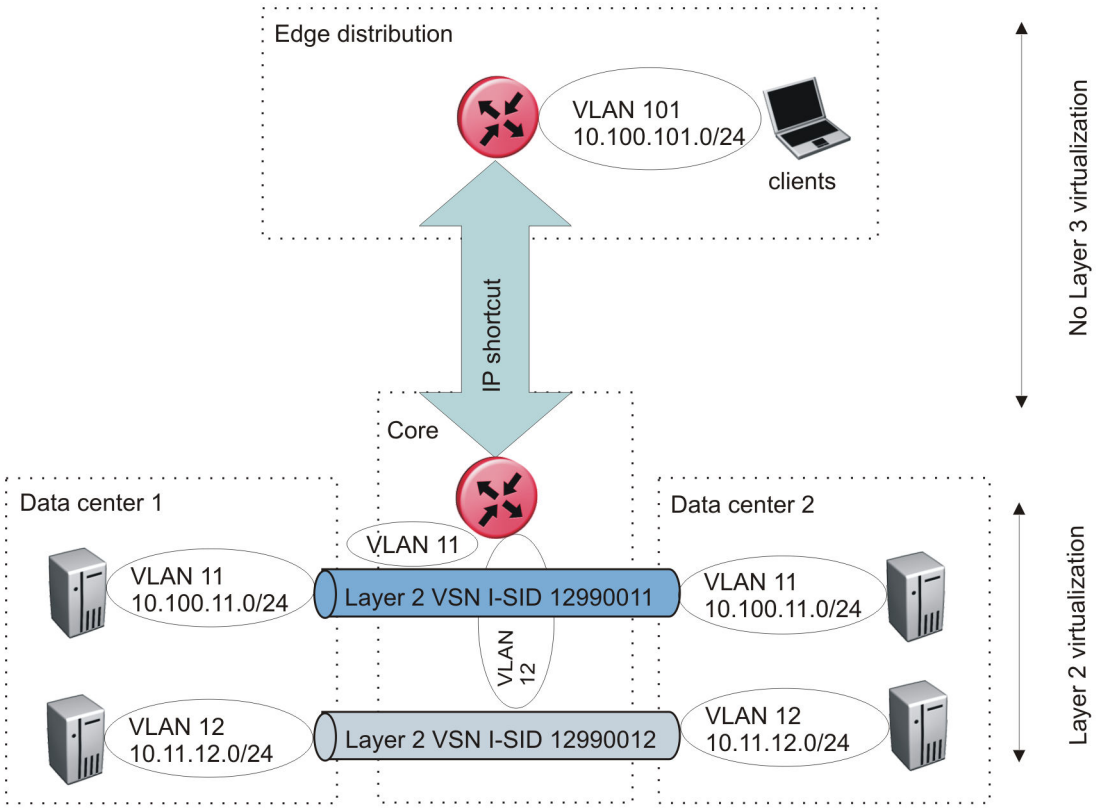


Figure 68: IP shortcut scenario to move traffic between data centers

The following figure uses Layer 3 VSNs to route VRFs between the edge distribution and the core. The VRFs are attached to I-SIDs and use Layer 3 virtualization.

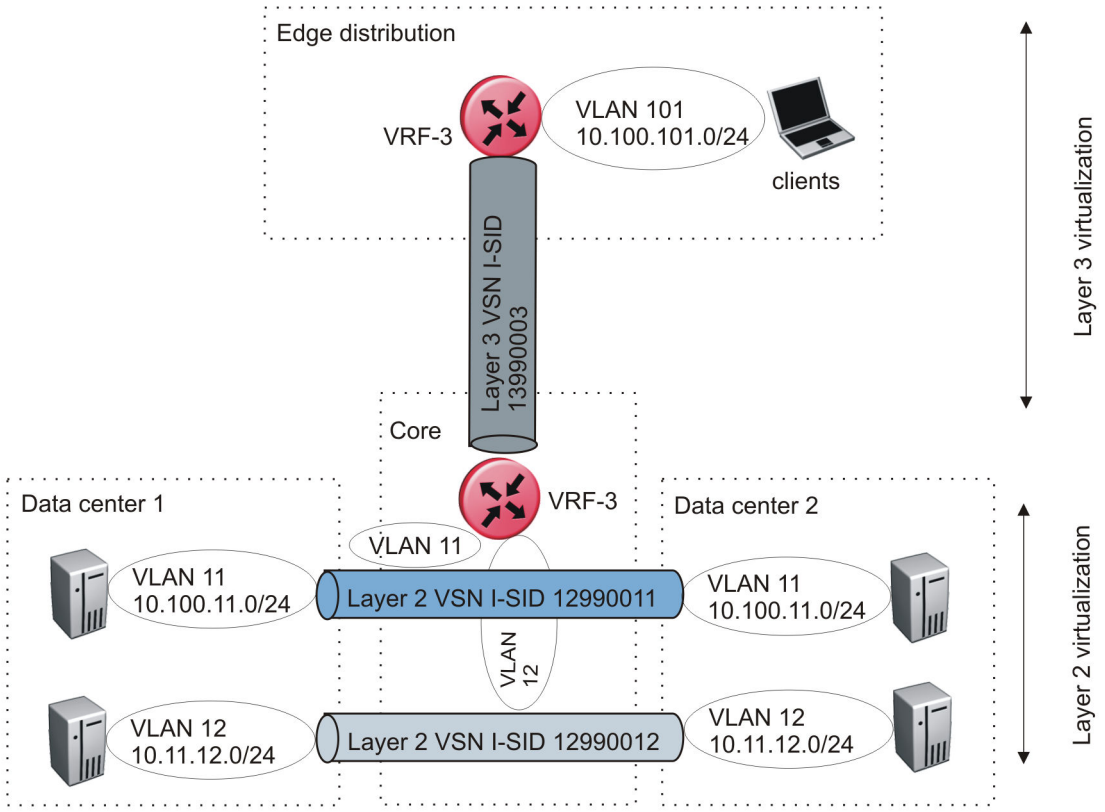


Figure 69: VRF scenario to move traffic between data centers

Multicast architecture

Networks today either have inefficient bridged IP multicast networks (IGMP) or IP multicast networks that require multiple protocols that are complex to configure and operate. IP multicast over SPBM builds on the simplicity introduced with Avaya VENA using SPBM for the control plane with support for bridged and routed IP multicast traffic, without the inefficiencies or complexities that exist in other topologies today.

This functionality extends the SPBM IS-IS control plane to additionally exchange IP multicast stream advertisement and membership information, which means that you can use SPBM for Layer 2 (unicast, broadcast, multicast) virtualization as well as Layer 3 (unicast, multicast) routing and forwarding virtualization.

IP multicast over SPBM supports three operational models:

1. Layer 2 Virtual Services Network with IGMP support on the access networks for optimized forwarding of IP multicast traffic in a bridged network (L2 VSN with multicast)
2. IP multicast routing support for Global Routing Table using SPB in the core (IP Shortcuts with multicast)
3. Layer 3 Virtual Services Network with VRF based IP multicast routing support over SPB in the core and IGMP on the access (L3 VSN with multicast)

All multicast streams are constrained within the level in which they originate, which is called the scope level. In other words, if a sender transmits a multicast stream to a BEB on a C-VLAN with IP

multicast over SPBM enabled, only receivers that are part of the same Layer 2 VSN can receive that stream. Similarly, if a sender transmits a multicast stream to a BEB on a VLAN that is part of the Layer 3 VSN with IP multicast over SPBM enabled, only receivers that are part of the same Layer 3 instance can receive that stream.

IP multicast over SPBM uses BEBs to act as senders and receivers of data. After a BEB receives IP multicast data from a sender, a BEB allocates a data I-SID in the range of 16,000,000 to 16,512,000 for the stream. The stream is identified by the S,G,V tuple, which is the source IP address, group IP Address, and the stream is identified by the local VLAN on which the stream is received. The BEB also sends a TLV update to its neighbors to inform them of the presence of an IP multicast stream, along with identifying the sender. The BEB propagates the information through the SPBM cloud through IS-IS TLV updates in LSPs that result in a multicast tree being created for that stream.

IGMP handles group membership registration to enable members to receive data. IGMP snooping listens to conversations between hosts and routers, and maintains a table of links that require IP multicast streams.

The BEBs also act as IGMP queriers and send out periodic IGMP queries. The IGMP querier enables the creation of the link table. After a BEB receives an IGMP join message from a receiver, a BEB queries the IS-IS database to check if a sender exists for the requested stream within the scope of the receiver. If the requested stream does not exist, the IGMP information is kept, but no further action is taken. If the requested stream exists, the BEB sends an IS-IS TLV update to its neighbors to inform them of the presence of a receiver and this information is propagated through the SPBM cloud.

IS-IS acts dynamically using the TLV information it receives from BEBs that connect to the sender and the receivers to create a multicast tree between them.

The following figure shows how multicast senders and receivers connect to the SPBM cloud using BEBs.

Comments? infodev@avaya.com

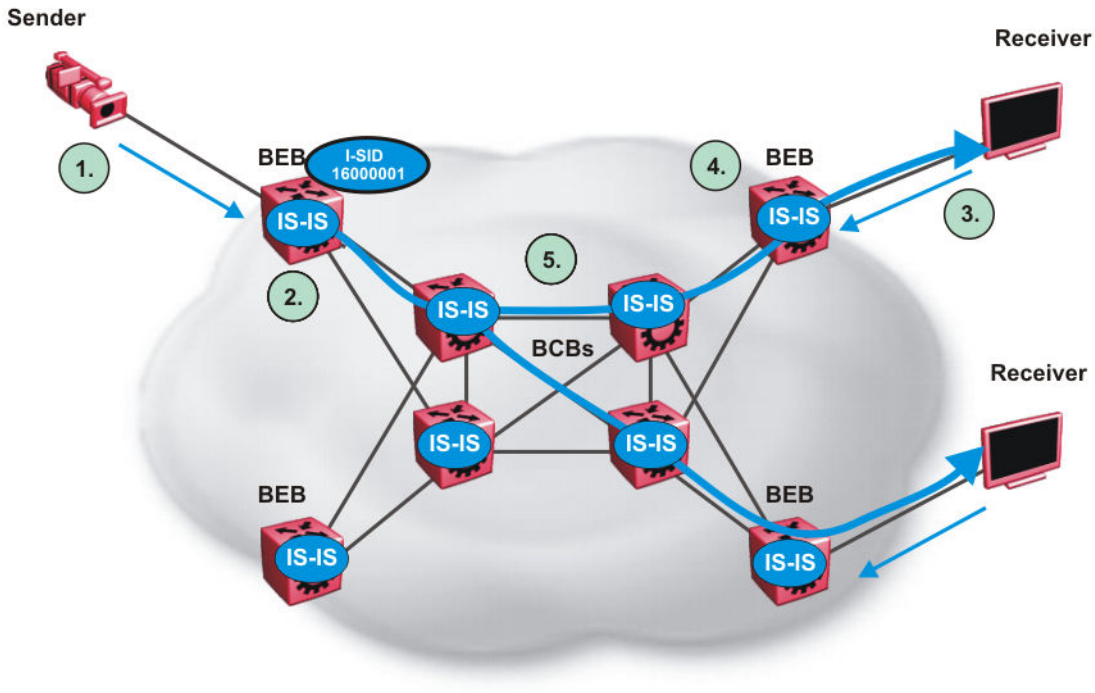


Figure 70: IP multicast over SPBM streams

The following list describes how multicast senders and receivers connect to the SPBM cloud using BEBs in the preceding diagram:

1. The sender sends multicast traffic with group IP address 233.252.0.1.
2. After the BEB receives the IP multicast stream from the sender, the BEB allocates data I-SID 16000001 for the S,G multicast stream. The BEB sends an LSP with the IPMC TLV (for Layer 2 VSN and Layer 3 VSN) or IPVPN TLV (for IP Shortcuts for multicast) with the transmit bit set. The BEB also sends an IS-IS service identifier and unicast address sub-TLV (where the unicast address has the multicast bit set and the I-SID is the data I-SID).
3. The receiver sends a join request to group 233.252.0.1.
4. The BEB (acting as the IGMP querier) queries the IS-IS database to find all senders for group 233.252.0.1. If the group exists, the BEB sends IS-IS service identifier and unicast address sub-TLV (where the unicast address has the multicast bit set and the nickname is the stream transmitter BEB and the I-SID is the data I-SID). If the requested stream does not exist, the BEB keeps the IGMP information, but no further action is taken.
5. IS-IS acts dynamically using the TLV information it receives from BEBs that connect to the sender and receivers to create a multicast tree between them and the data starts flowing from the sender.

For conceptual and configuration information about IP multicast over SPBM, see *Configuring Avaya VENA Fabric Connect on Avaya Virtual Services Platform 9000,* NN46250-510.

Large data center architecture

SPBM supports data centers with IP shortcuts, Layer 2 VSNs, or Layer 3 VSNs. If you use vMotion, you must use Layer 2 between data centers (Layer 2 VSN). With Layer 2 VSNs, you can add an IP

addresses to the VLAN on both data centers and run Virtual Router Redundancy Protocol (VRRP) between them to allow the ESX server to route to the rest of the network.

The following figure shows an SPBM topology of a large data center. This figure represents a full-mesh Virtual Enterprise Network Architecture (VENA) data center fabric using SPBM for storage over Ethernet. This topology is optimized for storage transport because traffic never travels more than two hops.

Note:

Avaya recommends that you use a two-tier, full-mesh topology for large data centers.

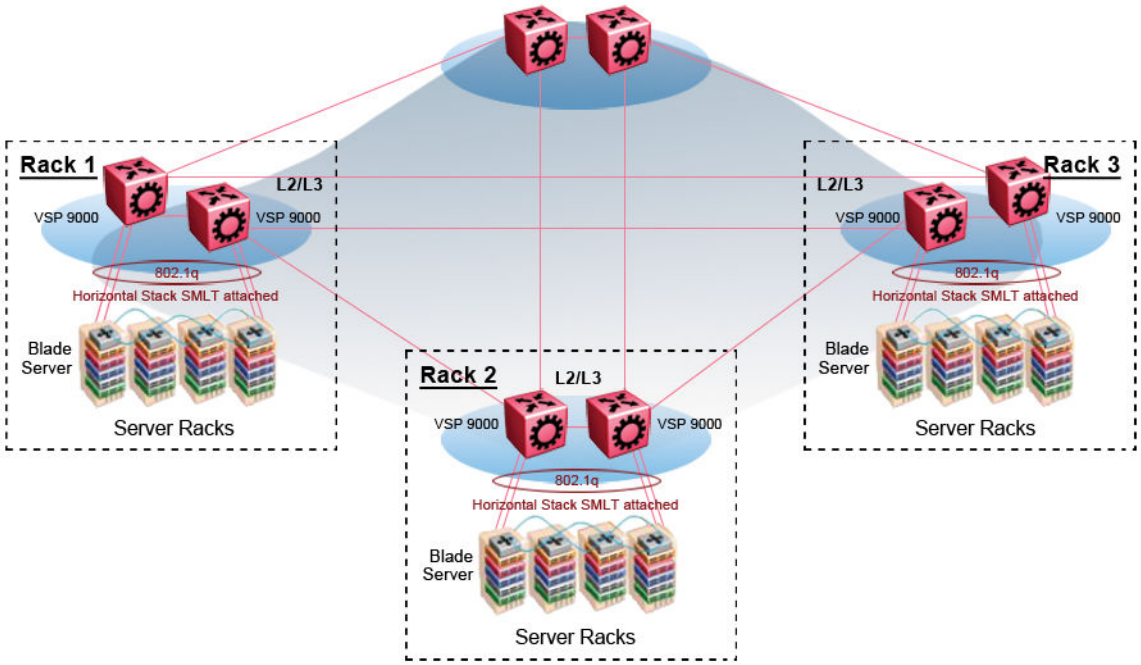


Figure 71: SPBM data center—full mesh

Traditional data center routing of VMs:

In a traditional data center configuration, the traffic flows into the network to a VM and out of the network in almost a direct path.

The following figure shows an example of a traditional data center with VRRP configured. Because end stations are often configured with a static default gateway IP address, a loss of the default gateway router causes a loss of connectivity to the remote networks. VRRP eliminates the single point of failure that can occur when the single static default gateway router for an end station is lost.

Data Center Routing — Traditional (pre VM move):

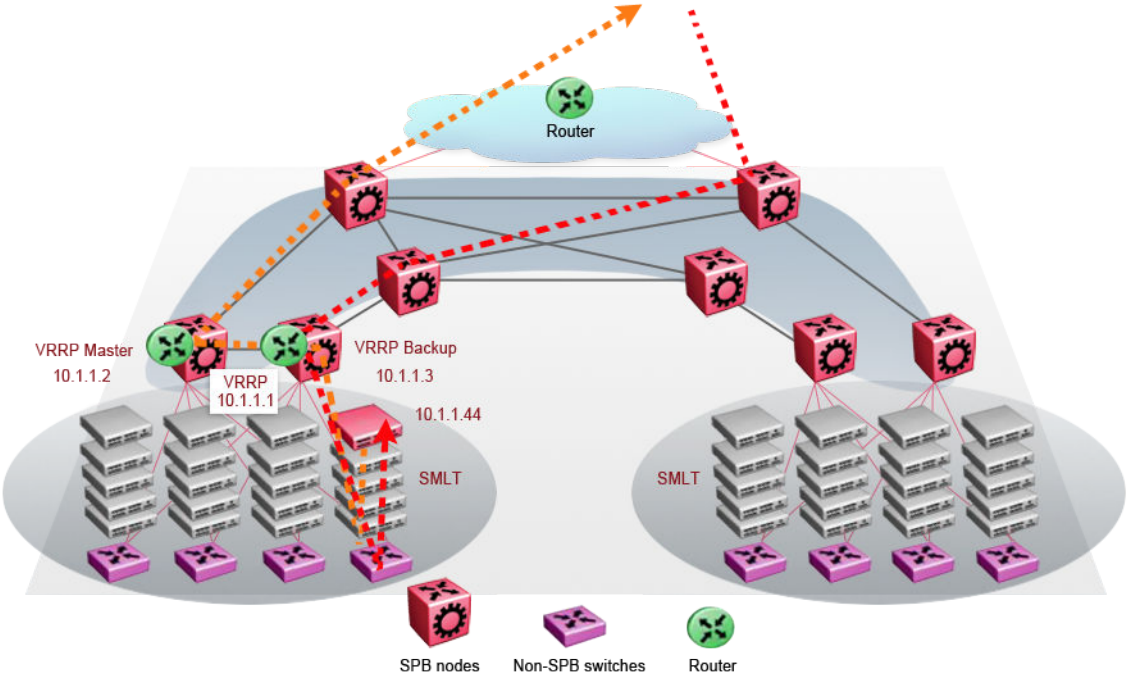


Figure 72: Traditional routing before moving VMs

A VM is a virtual server. When you move a VM, the virtual server is moved as is. This action means that the IP addresses of that server remain the same after the server is moved from one data center to the other. This in turn dictates that the same IP subnet (and hence VLAN) exist in both data centers.

In the following figure, the VM moved from the data center on the left to the data center on the right. To ensure a seamless transition that is transparent to the user, the VM retains its network connections through the default gateway. This method works, but it adds more hops to all traffic. As you can see in the figure, one VM move results in a complicated traffic path. Multiply this with many moves and soon the network look like a tangled mess that is very inefficient, difficult to maintain, and almost impossible to troubleshoot.

Data Center Routing — Traditional (post VM move):

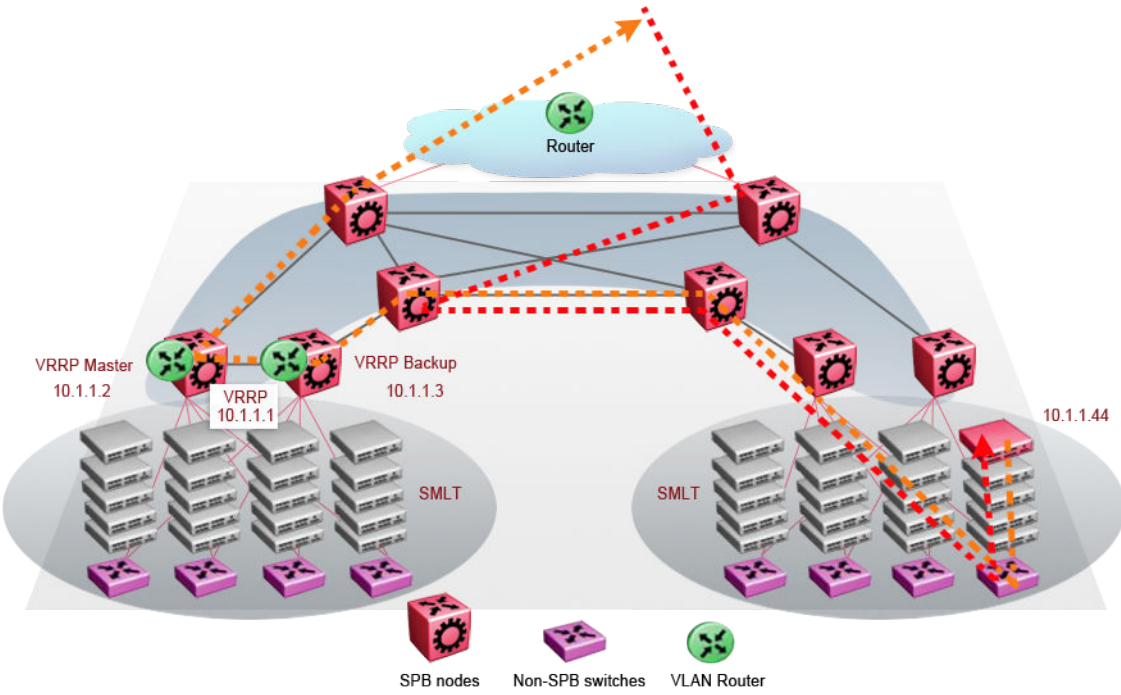


Figure 73: Traditional routing after moving VMs

Optimized data center routing of VMs:

Two features make a data center optimized:

- VLAN routers in the Layer 2 domain (green icons)
- VRRP BackupMaster

The VLAN routers use lookup tables to determine the best path to route incoming traffic (red dots) to the destination VM.

VRRP BackupMaster solves the problem of traffic congestion on the IST. Because there can be only one VRRP Master, all other interfaces are in backup mode. In this case, all traffic is forwarded over the IST link towards the primary VRRP switch. All traffic that arrives at the VRRP backup interface is forwarded, so there is not enough bandwidth on the IST link to carry all the aggregated riser traffic. VRRP BackupMaster overcomes this issue by ensuring that the IST trunk is not used in such a case for primary data forwarding. The VRRP BackupMaster acts as an IP router for packets destined for the logical VRRP IP address. All traffic is directly routed to the destined subnetwork and not through Layer 2 switches to the VRRP Master. This avoids potential limitation in the available IST bandwidth.

The following figure shows a solution that optimizes your network for bidirectional traffic flows. However, this solution turns two SPBM BCB nodes into BEBs where MAC and ARP learning will be enabled on the Inter-VSN routing interfaces. If you do not care about top-down traffic flows, you can omit the Inter-VSN routing interfaces on the SPBM BCB nodes. This makes the IP routed paths top-down less optimal, but the BCBs remain pure BCBs, thus simplifying core switch configurations.

Distributed vRouting — Optimized (pre VM move):

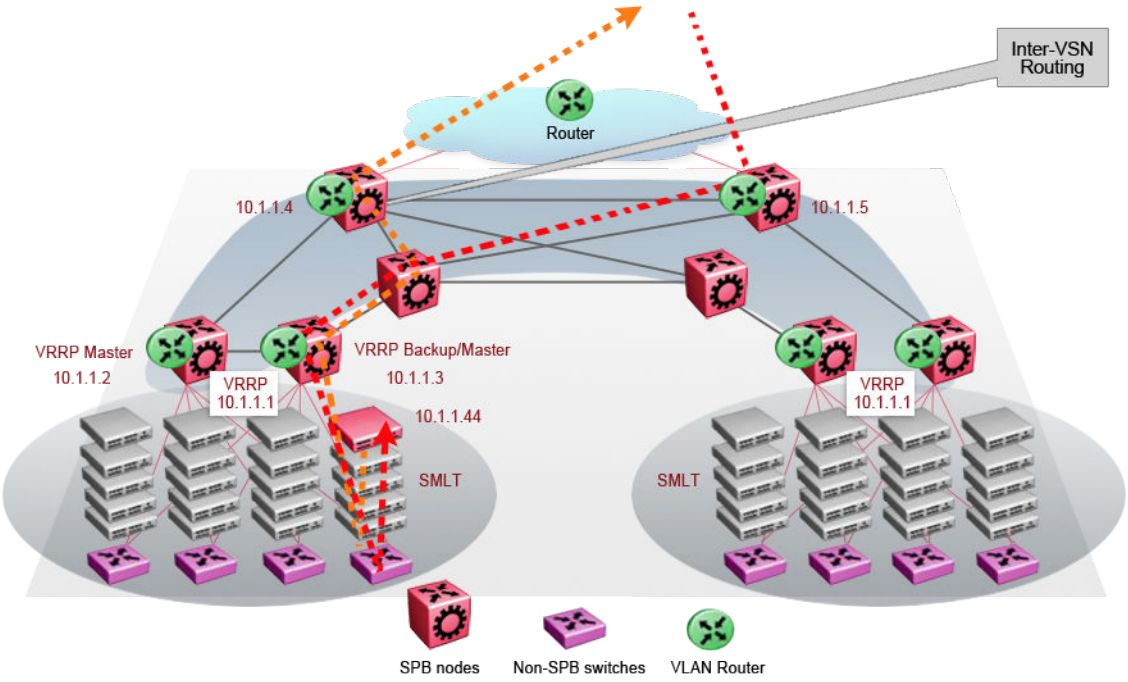


Figure 74: Optimized routing before moving VMs

In the traditional data center, chaos resulted after many VMs were moved. In an optimized data center as shown in the following figure, the incoming traffic enters the Layer 2 domain where an edge switch uses Inter-VSN routing to attach an I-SID to a VLAN. The I-SID bridges traffic directly to the destination. With VRRP BackupMaster, the traffic no longer goes through the default gateway; it takes the most direct route in and out of the network.

Distributed vRouting (post VM move):

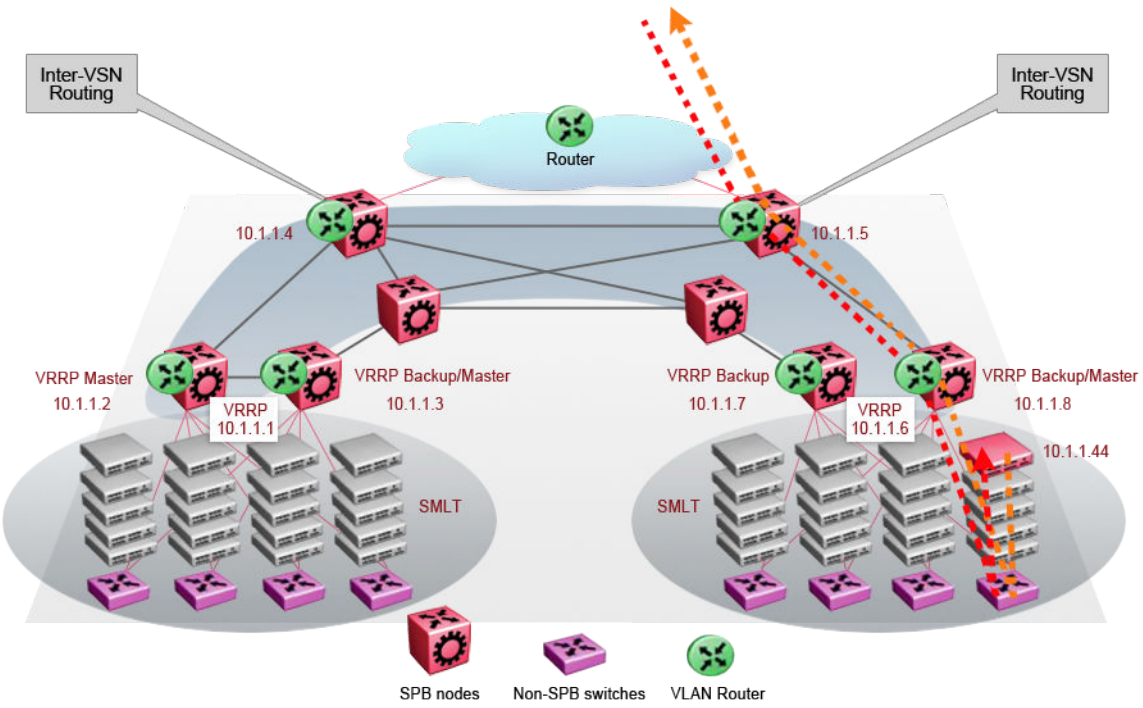


Figure 75: Optimized routing after moving VMs

Best practices

This section provides best practices to configure an SPBM network.

IS-IS

The following list identifies best practices for IS-IS:

- Avaya recommends that you change the IS-IS system ID from the default B-MAC value to a recognizable address to easily identify a switch. This change helps to recognize source and destination addresses for troubleshooting purposes.
  - If you leave the system ID as the default value (safe practice as it ensures no duplication in the network), it can be difficult to recognize the source and destination B-MAC for troubleshooting purposes.
  - If you do manually change the system ID, take the necessary steps to ensure no duplication exists in the network.
- Create two B-VLANs to allow load distribution over both B-VLANs. This configuration is required if you use SMLT. Even if you do not use SMLT in the network, this is still good practice as adding a second B-VLAN to an existing configuration allows SPBM to load balance traffic across two equal cost multipaths if the physical topology grants it.
- In a ring topology with OSPF and IS-IS configured in the core, a core link break causes slow convergence that can lead to SPBM Layer 2 traffic loss. If the last member link of an OSPF

VLAN fails, it takes down the IP interface and OSPF reconverges. While OSPF reconverges, SPBM does not have access to the CPU, which causes traffic loss.

SPBM

The following list identifies best practices for SPBM:

- Use a different, easily recognizable IS-IS nickname on each switch.
- If you enable IP shortcuts, you must configure an IS-IS IP source address on the switch.

IST

If the switch will be part of an SMLT cluster, you must create the IST prior to enabling IS-IS.

Important:

Different from Ethernet Routing Switch 8800 SPBM implementation, on Virtual Services Platform 9000 when you map a customer VLAN (C-VLAN) to a Service Instance Identifier (I-SID), you must include the IST MLT peer switches in the VLAN, which means all IST member ports must be members of this C-VLAN. In ERS 8800, you cannot configure IST MLT to be part of the C-VLAN. For more information about this and other differences between the two products, see *Platform Migration Reference for Avaya Virtual Services Platform 9000,* NN46250-107.

SMLT

The following list identifies best practices for SMLT:

- You must configure each switch in the cluster to peer with its neighbor. If the IS-IS interface between the neighbors is the IST MLT (recommended, but not strictly required), traffic paths will generally be similar to existing behavior.
- A virtual B-MAC is automatically created based on the lowest system ID in the cluster plus one.

  Important:

  The virtual B-MAC or any system ID created must not conflict with any other system ID or virtual B-MAC in the network. Verify that there is no duplication of system IDs or virtual B-MACs in the network .

  A safe practice is to leave the lowest byte in the system ID as all zeroes.
- No consistency check exists to ensure that you cannot add Layer 2 VSN VLANs to the IST.

Physical or MLT links between IS-IS switches

Virtual Services Platform supports only a single port or single MLT between a pair of IS-IS switches. For example, if two individual ports exist between a pair of IS-IS switches, you can only configure IS-IS on one of the two ports.

If you configure a single MLT between a pair of IS-IS switches, all ports in the MLT are used. You must configure the MLT before you enable IS-IS on the MLT.

Connectivity Fault Management

The following list identifies best practices for Connectivity Fault Management (CFM):

- The CFM domain name must be the same on all switches in an IS-IS area.
- The maintenance association must be the same on all switches in an IS-IS area.
- To use CFM testing over both B-VLANs, create two maintenance associations, one for each B-VLAN.

- You can configure the same maintenance domain intermediate points (MIP) on all switches in an IS-IS area or uniquely defined per switch.

  Important:

  You must configure the MIP at the same level as the maintenance association endpoints (MEP) on all switches in the SPBM network.

Example of a configuration using best practices

```
spbm-id : 1
BVID #1 & BVID #2 : 4040, 4041 (ignore warning message when configuring)
nick-name : b:b0:<node-id>
MEP-id : md.ma.<node-id>
BMAC : 00:bb:00:00:<node-id>:00
VirtBMAC : 00:bb:00:00:<node-id>:ff
MD : spbm (level 4)
MA : 4040 & 4041
mep : <node-id>
mip : (level 4)
isis manual area : 49.0001
```

Migration best practices

Before you migrate your network to SPBM, perform an audit to determine if the desired configuration and traffic is supported by SPBM in the current release. If you determine that the migration is supported, follow the high level steps in the following sections to ensure a successful migration.

Supported traffic on SPBM networks

The current release supports the following types of SPBM traffic:

- Layer-2 bridged traffic (Layer 2 VSN)
- IPv4 unicast routed traffic on the Global Router (IP shortcuts)
- IPv4 unicast routed traffic using a VRF (Layer 3 VSN)
- IPv4 unicast routed traffic using different VSNs, which have different I-SIDs (Inter-VSN)
- Layer-2 IP multicast traffic in a bridged network (Layer 2 VSN with IP multicast over SPBM)
- IPv4 multicast routed traffic on the Global Router (IP shortcuts with IP multicast over SPBM)
- IPv4 multicast routed traffic using a VRF (Layer 3 VSN with IP multicast over SPBM)

IPv6 routed traffic can only be bridged on a Layer 2 VSN and not routed in a Layer 3 VSN.

Common procedures and exclusions

Ensure that the SPBM infrastructure is enabled and functional in the network.

If a VLAN is multicast enabled or uses IPv6 routing, then you can only migrate it to an SPBM Layer 2 VSN, and route with an external router.

Identify the user-to-network interface (UNI) and network-to-network interface (NNI) ports that are currently port members of the VLAN on all the switches in the network. If you use Multiple Spanning Tree Protocol (MSTP) or Rapid Spanning Tree Protocol (RSTP) on the UNI, and the relevant

domain touches more than one SPBM-enabled switch in the network (without passing through NNI ports), then the VLAN cannot be migrated to SPBM.

Migrating a VLAN to a Layer 2 VSN (C-VLAN)

Use the following steps to provide Layer 2 connectivity for a VLAN across the SPBM core.

1. Follow the pre-migration checks in Common procedures and exclusions on page 142.
2. Identify the UNI and NNI ports that are currently port members of the VLAN on all the switches in the network.
3. On all switches in the network that currently connect by the VLAN, remove the NNI ports from the membership list of the VLAN.

   Warning:

   This step causes service interruption.
4. Make the VLAN a Layer 2 VSN using the **`vlan i-sid <1-4084> isid <0-16777215>`** command. Use the same value of I-SID on all the switches. This step restores service.

Migrating to inter-VSN routing

Inter-VSN routing provides the ability to route traffic between extended VLANs where the VLANs have different I-SIDs. All traditional IPv4 unicast routing and gateway redundancy protocols, including OSPF, Routing Information Protocol (RIP), VRRP, and Routed SMLT (RSMLT), are supported on a VLAN that maps to an I-SID.

Use the following high-level steps to migrate a configuration to use inter-VSN routing.

1. Follow the pre-migration checks in Common procedures and exclusions on page 142.
2. Perform the following for each VLAN in the SPBM core:
   - On all the switches where the VLAN is configured, remove all NNI ports.

     Warning:

     This step causes service interruption.
   - On all the switches where the VLAN is configured, map the VLAN to an I-SID. This configuration restores Layer 2 connectivity (use the **`l2 tracetree <1-4084> <1-16777215>`** command to validate Layer 2 connectivity within the VLAN). Layer 3 connectivity will be restored after the routing protocols configured on the VLAN converge.
   - If an interior gateway protocol (IGP), for example, OSPF, is used on the VLAN, you can reduce the impact on traffic during the migration by using two VLANs in each routed segment and configuring the interface cost to select the VLAN used for the routed next hop. Migrating the two VLANs one after the other reduces the duration for which there is a loss of IGP adjacency.
   - After you have assigned an I-SID to all the VLANs identified for migration, the configuration part of the migration is complete. At this point all the traffic flows should be back to normal.

IP multicast over SPBM and PIM

Before you enable IP multicast over SPBM for a VLAN, you must disable PIM on the peer switches for that VLAN.

IP multicast over SPBM and IGMP Snooping

For traditional multicast and IP multicast over SPBM to function correctly, use the following guidelines:

- If you want to use IP multicast over SPBM, enable both IP multicast over SPBM and IGMP Snoop on each BEB in the SPBM network.
- If you do not want to use IP multicast over SPBM, disable both IP multicast over SPBM and IGMP Snoop on each BEB in the SPBM network.

SPBM restrictions and limitations

This section describes the restrictions and limitations associated with SPBM.

RSTP and MSTP

The following list identifies restrictions and limitations associated with RSTP and MSTP:

- RSTP mode does not support SPBM.
- A C-VLAN-level loop across SPBM NNI ports cannot be detected and needs to be resolved at the provisional level.
- SPBM NNI ports are not part of the Layer 2 VSN C-VLAN, and BPDUs are not transmitted over the SPBM tunnel. SPBM can only guarantee loop-free topologies consisting of the NNI ports. Avaya recommends that you always use Simple Loop Prevention Protocol (SLPP) in an SMLT environment.

  Note:

  Avaya recommends that you deploy SLPP on C-VLANs to detect loops created by customers in their access networks. However, SLPP is not required on B-VLANs, and it is not supported. The B-VLAN active topology is controlled by IS-IS that has loop mitigation and prevention capabilities built into the protocol.
- SPB internally uses Multiple Spanning Tree Instance (MSTI) 62. MSTI 62 cannot be used by another VLAN or MSTI. For non-SPB customer networks, if you use MSTI 62 in the configuration, you must delete MSTI 62 before you can configure SPBM.
- You must configure SPBM B-VLANs on all devices in the same MSTP region. MSTP requires this configuration to generate the correct digest.

SPBM IS-IS

The following list identifies restrictions and limitations associated with SPBM IS-IS:

- The current release does not support IP over IS-IS as defined by RFC 1195. IS-IS protocol is only to facilitate SPBM.
- The current release uses level 1 IS-IS. The current release does not support level 2 IS-IS. The ACLI command **`show isis int-l2-contl-pkts`** is not supported in the current release because the IEEE 802.1aq standard currently only defines the use of one hierarchy, Level 1.
- The IS-IS standard defines wide (32bit) metrics and narrow (8 bits) metrics. The current release supports the wide metric.
- SPBM supports full High Availability (HA). The SPBM and IS-IS configuration and dynamic information (such as adjacencies and LSPs) are all synchronized to the standby CPU to ensure

seamless switchover. Because the HA framework cannot guarantee seamless switchover, there is a 6 to 7 seconds gap between the master CPU going down and the backup CPU coming up. To avoid IS-IS adjacencies bouncing during the switchover, the recommended hello interval is 9 seconds and the hello multiple is 3.

Pay special attention to the expected scaling of routes in the network when you select configuration values for the **`isis l1-hello-interval`** and **`isis l1-hello-multiplier`** commands on IS-IS interfaces. The default values for these commands work well for most networks, including those using moderately-scaled routes. In highly-scaled networks, you may need to configure higher values for these commands. For example, if the total number of non IS-IS routes on a given BEB exceeds 25,000 in combination with approximately 60,000 IS-IS routes that the BEB receives from other BEBs in the network, Avaya recommends that you configure a value of 12 for **`isis l1-hello-multiplier`**, rather than use the default value of 3.

SPBM NNI SMLT

For NNI-facing SMLT, the current release supports only one link between each IS-IS adjacency pair.

The following modules support NNI ports:

- 9024XL I/O module
- 9048XS-2 I/O module
- 9012QQ-2 I/O module

Internet Protocol Flow Information eXport

The switch does not support Internet Protocol Flow Information eXport (IPFIX) on Intermediate-System-to-Intermediate-System (IS-IS) interfaces.

VLACP

VLACP is generally used when a repeater or switch exists between connected Virtual Services Platform 9000 switches to detect when a connection is down even when the link LED is lit. If you configure VLACP on an SPBM link that is also an IST link, during a connection fail over (where the link LED stays lit) the IS-IS hellos time out first (after 27 seconds, using default values) and take down the IS-IS adjacency. IS-IS then calculates the new shortest path and fails over the SPBM traffic. 90 seconds after the connection failure (using default values), VLACP goes down but the IST link was already taken down by IS-IS.

In this scenario, there is no data traffic impact because IS-IS can find another path in the SPBM network before VLACP goes down.

SNMP traps

On each SPBM peer, if you configure the SPBM B-VLANs to use different VLAN IDs, for example, VLAN 10 and 20 on one switch, and VLAN 30 and 40 on the second, the system does not generate a trap message to alert of the mismatch because the two switches cannot receive control packets from one another. Configure the SPBM B-VLANs to use matching VLAN IDs.

Other

The following list identifies other restrictions and limitations:

- The current release does not support I-SID filters.
- You cannot enable C-VLAN and B-VLAN on the same port, except the IST ports.
- The current release supports NNI only on 10 GbE interface modules.

Legacy IS-IS

An SPBM node can form an adjacency with a legacy IS-IS router. This adjacency means you can introduce SPBM into existing networks and provide easy migration.

IP multicast over SPBM restrictions

Review the following restrictions for the IP multicast over SPBM feature.

IGMP

The BEB must be the only IGMP querier in the network. If the BEB receives an IGMP query from any other device, it causes unpredictable behavior, including traffic loss.

SPBM supports IGMP Snooping on a C-VLAN, but it does not support PIM on a C-VLAN. If you enable IGMP Snooping on a C-VLAN, then its operating mode is Layer 2 VSN with IP multicast over SPBM.

SPBM supports Network Load Balancing (NLB) unicast. SPBM does not support NLB multicast or NLB multicast with IGMP.

You must enable SSM snoop before you configure IGMP version 3, and you must enable both ssm-snoop and snooping for IGMPv3.

For IGMP Snooping, ensure that the IGMP version used by multicast hosts and other devices in the network is either the same as the IGMP version configured on the IGMP Snooping VLAN, or that compatibility mode is enabled.

SSM

If you delete any **`ssm-map`** in a static range group, the switch deletes the entire static range group. For example, create an ssm-map for 232.122.122.122 to 232.122.122.122.128 and after that configure this same range in a static group. If you delete any ssm-map between 232.122.122.122. to 232.122.122.128, the switch deletes the entire static range group.

PIM

There can be no interaction with PIM and multicast routers on the access.

The BEB only interacts with IGMP messages and not PIM, so all multicast traffic must be drawn towards the BEB, which acts as the IGMP querier, for the stream to be announced.

IP multicast over SPBM does not integrate PIM functionality so the following considerations apply when connecting to a PIM network:

- You must configure static IGMP receivers on the BEB access interface facing the PIM network when the sender is on the SPBM access network and the receiver is on the PIM network. Static IGMP receivers make the PIM router accept streams and avoid a Reverse Path Forwarding (RPF) check that can change the source of the stream.
- You must configure static IGMP receivers on the PIM interface facing the SPBM multicast network when the sender is on the PIM network and the receiver is on the SPBM access network.
- You must configure Access Control Lists (ACLs) on the BEB facing the PIM network for security.

Data I-SID

The BEB matches a single multicast stream to a particular data I-SID. As a result there is a one-to-one mapping between the S,G to data I-SID for each BEB.

IP address

IP multicast over SPBM only supports IPv4 multicast traffic in this release.

Supported services

The switch does not support IP multicast over SPBM routing on inter-VSN routing interfaces.

The switch supports the following modes of IP multicast over SPBM:

- Layer 2 VSN multicast service — Multicast traffic remains within the same Layer 2 VSN across the SPBM cloud.
- Layer 3 VSN multicast service — Multicast traffic remains within the same Layer 3 VSN across the SPBM cloud.
- IP Shortcuts multicast service — Multicast traffic can cross VLAN boundaries but remains confined to the subset of VLANs with the Global Routing Table that have IP multicast over SPBM enabled.

Chapter 16: 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 *Configuring IP Multicast Routing Protocols on Avaya Virtual Services Platform 9000,* NN46250-504.

For design guidelines on IP Multicast over SPBM, see SPBM design guidelines on page 117.

For more conceptual and configuration information on IP Multicast over SPBM, see *Configuring Avaya VENA Fabric Connect on Avaya Virtual Services Platform 9000,* NN46250-510.

Multicast and VRF-lite

You can configure multicast routing support with the Virtual Routing and Forwarding (VRF) Lite feature. This feature is known as multicast virtualization.

Multicast virtualization enables multiple VRF routing instances on devices and supports various unicast routing protocols so that you can provide the services of many virtual routers from one physical device.

VRF-lite configurations support IGMP.

Multicast virtualization provides support for:

- Virtualization of control and data plane
- Multicast routing table managers (MRTM)
- Virtualized IGMPv1, v2, and v3
- Support for overlapping multicast address spaces
- Support for the Global Router (VRF0) and 512 VRFs
- SMLT and RSMLT suppport for multicast VRFs

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.

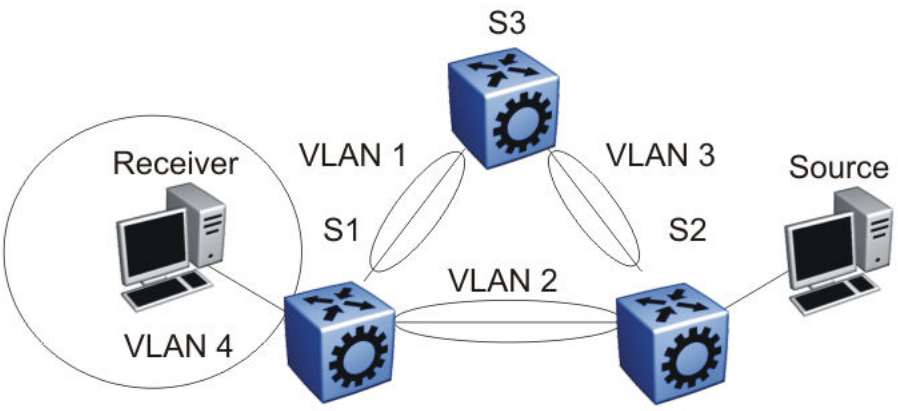


Figure 76: IP multicast sources and receivers on interconnected VLANs

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 159.
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.

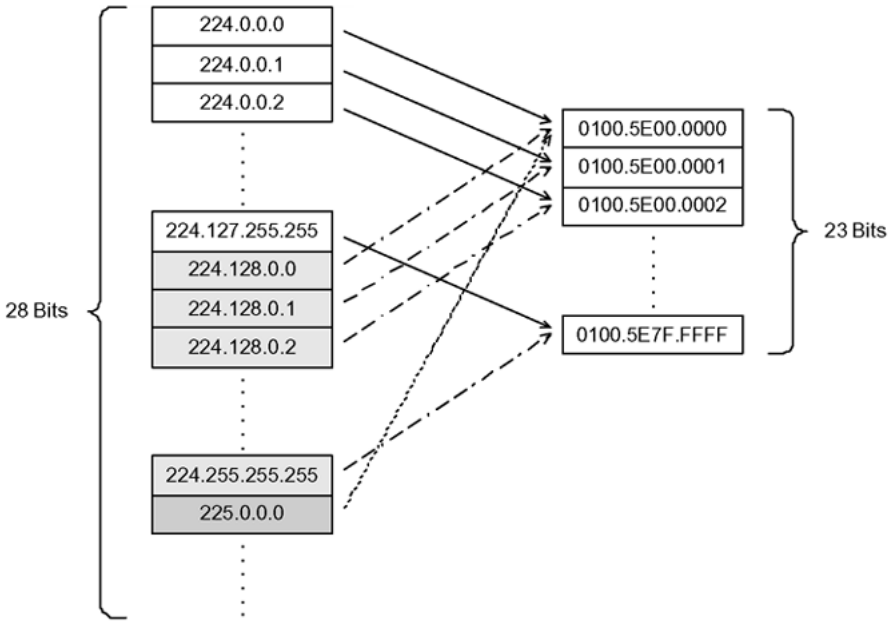


Figure 77: Multicast IP address to MAC address mapping

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.

IGMP Layer 2 Querier

In a multicast network, if you only need to use Layer 2 switching for the multicast traffic, you do not need multicast routing. However, you must have an IGMP querier on the network for multicast traffic to flow from sources to receivers. A multicast router normally provides the IGMP querier function. You can use the IGMP Layer 2 Querier to provide a querier on a Layer 2 network without a multicast router.

The Layer 2 querier function originates queries for multicast receivers, and processes the responses accordingly. On the connected Layer 2 VLANs, IGMP snoop continues to provide services as normal. IGMP snoop responds to queries and identifies receivers for the multicast traffic.

You must enable Layer 2 querier and configure an IP address for the querier before it can originate IGMP query messages. If a multicast router exists on the network, Virtual Services Platform 9000 automatically disables the Layer 2 querier.

In a Layer 2 multicast network, enable Layer 2 querier on only one of the switches in the VLAN. A Layer 2 multicast domain supports only one Layer 2 querier. No querier election exists.

For more information about how to configure IGMP Layer 2 Querier, see *Configuring IP Multicast Routing Protocols on Avaya Virtual Services Platform 9000,* NN46250-504.

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.

The current release does not support virtualized PIM. PIM is supported in the Global Routing Table only.

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 166.

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.

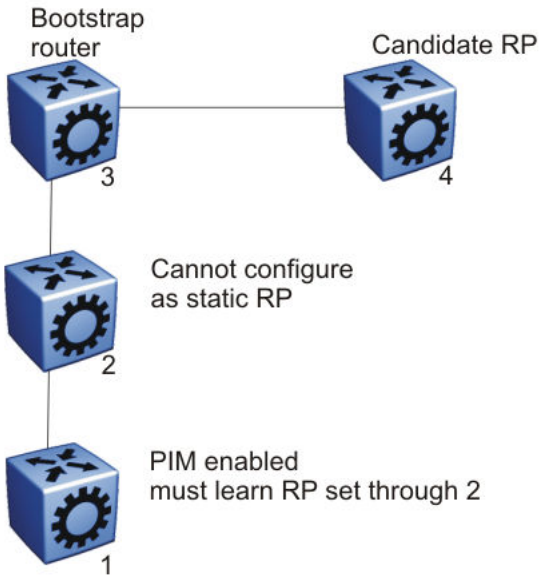


Figure 78: Example 1

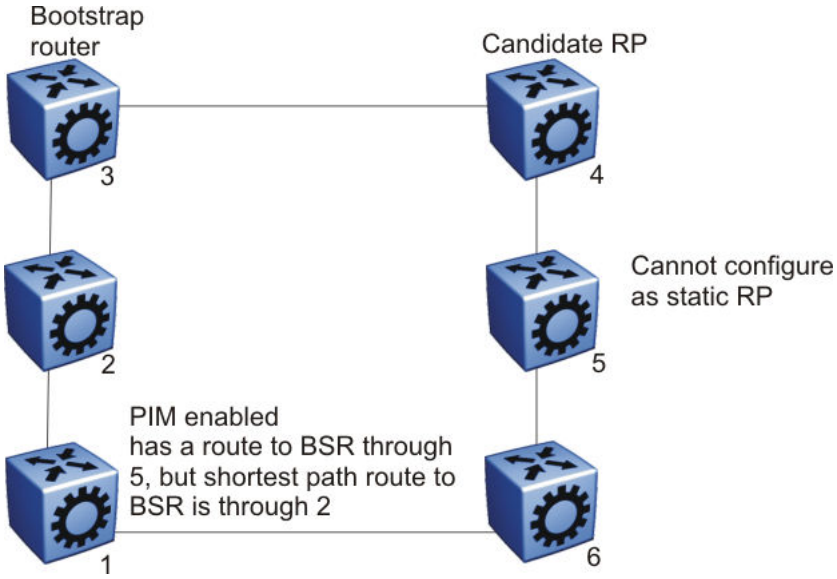


Figure 79: Example 2

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 31: Multicast SMLT triangle on page 78 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.

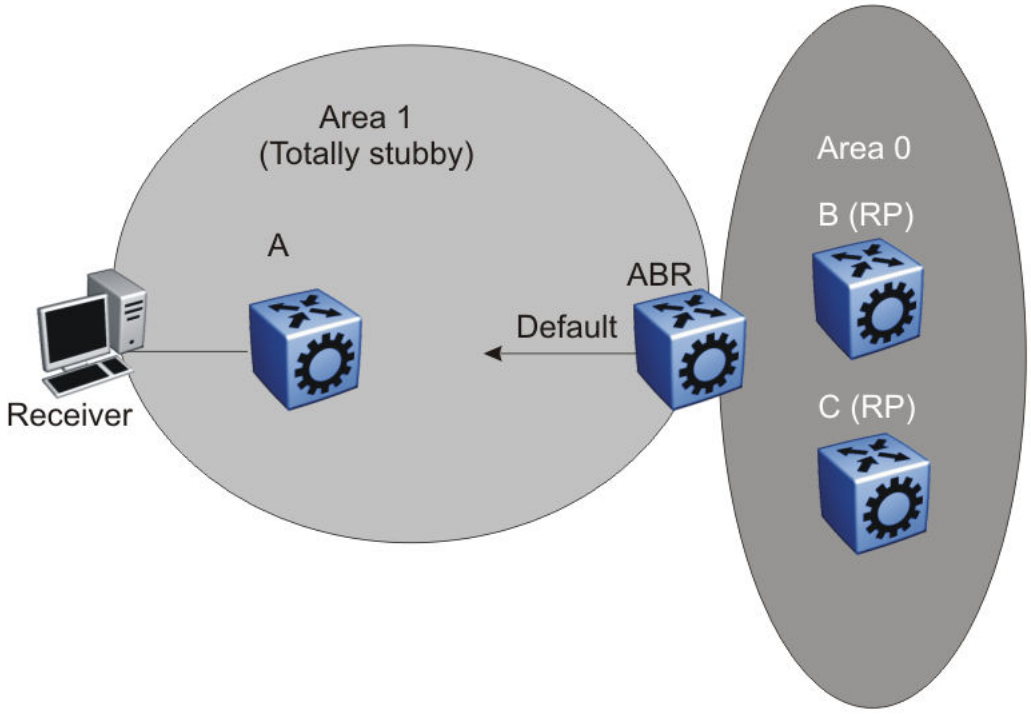


Figure 80: RP failover with default unicast routes

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.

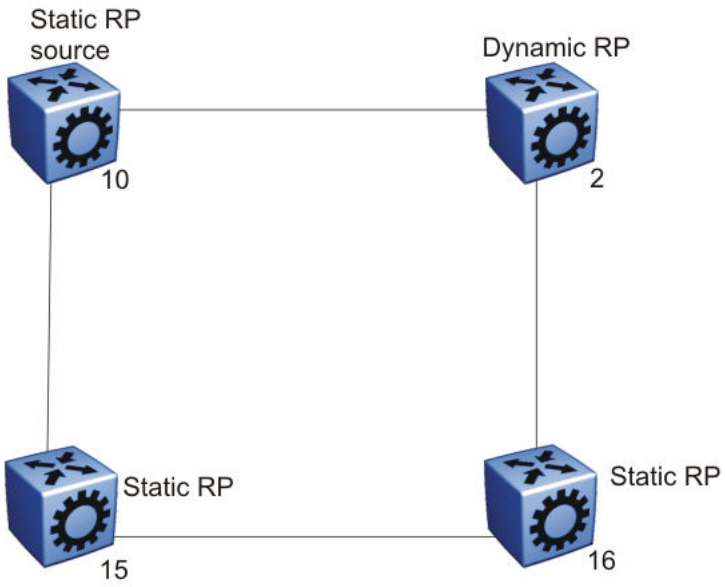


Figure 81: Unsupported static RP configuration

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.

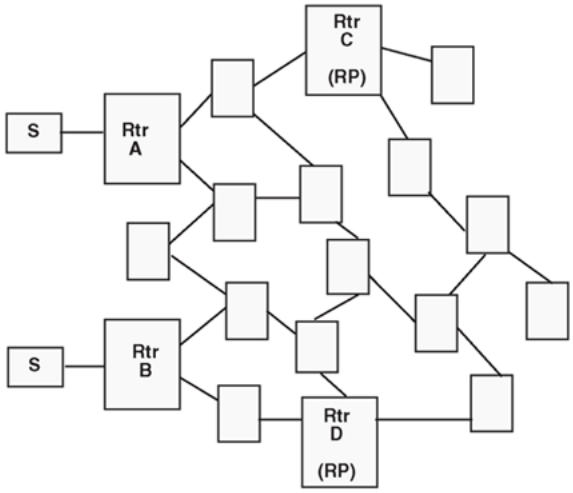


Figure 82: Example multicast network

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.

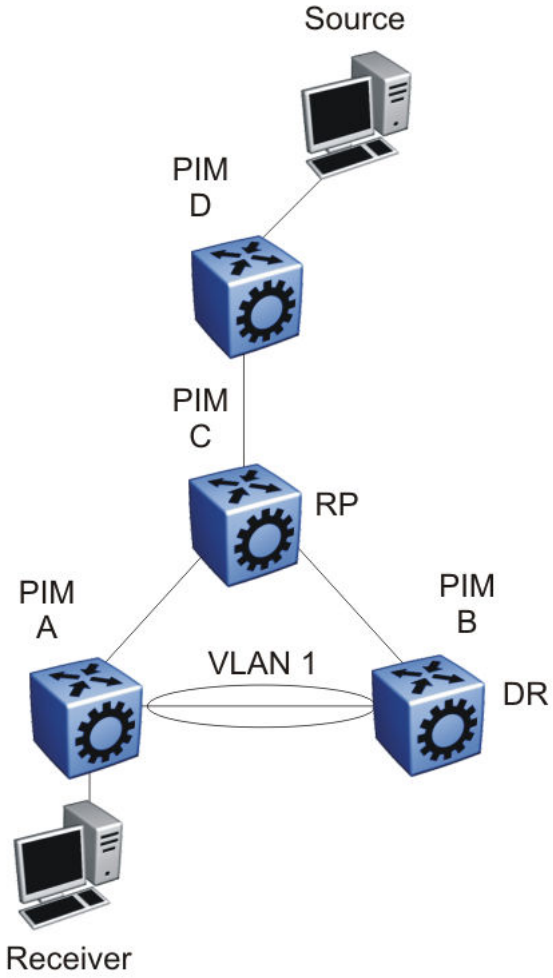


Figure 83: Receivers on interconnected VLANs

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 84: PIM network with non-PIM interfaces on page 167 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.

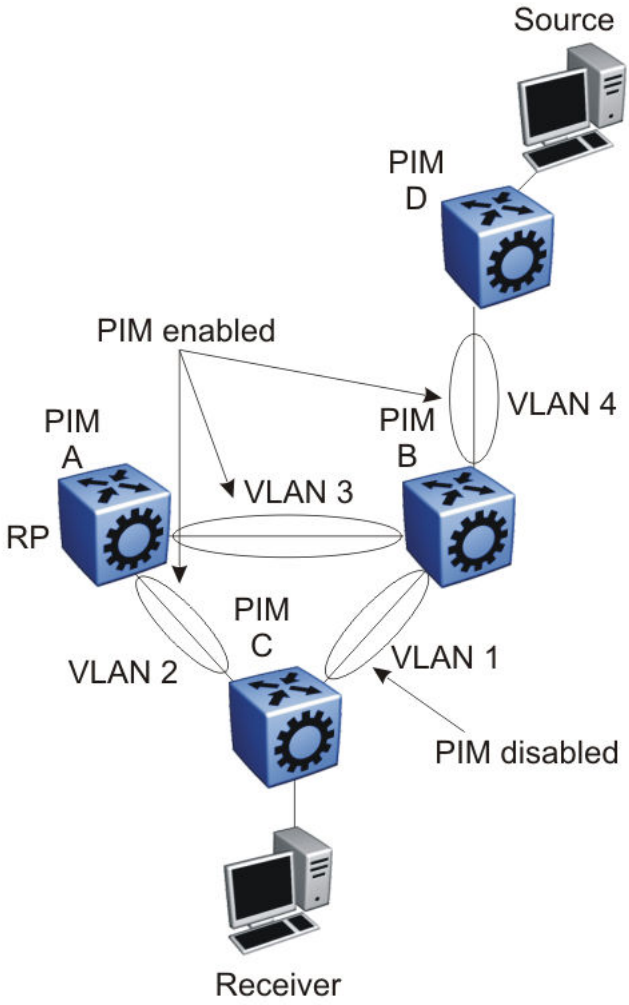


Figure 84: PIM network with non-PIM interfaces

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 *Configuring IP Multicast Routing Protocols on Avaya Virtual Services Platform 9000,* 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 specific source or sources, which is more efficient and puts less load on multicast routing devices.

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

IGMPv2 SSM extensions

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.

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 *Configuring IP Multicast Routing Protocols on Avaya Virtual Services Platform 9000,* 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.

Comments? infodev@avaya.com

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.

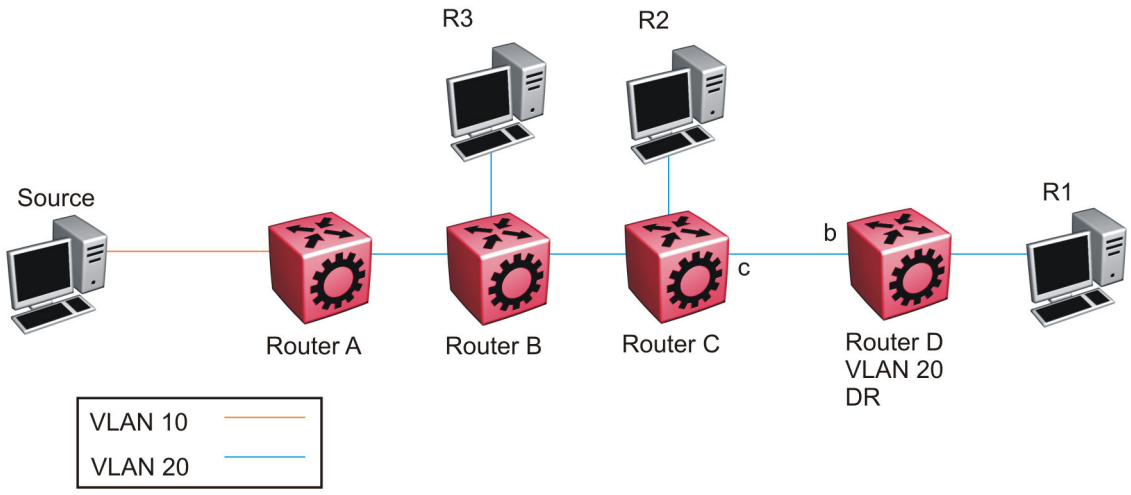


Figure 85: Extended VLAN topology

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.

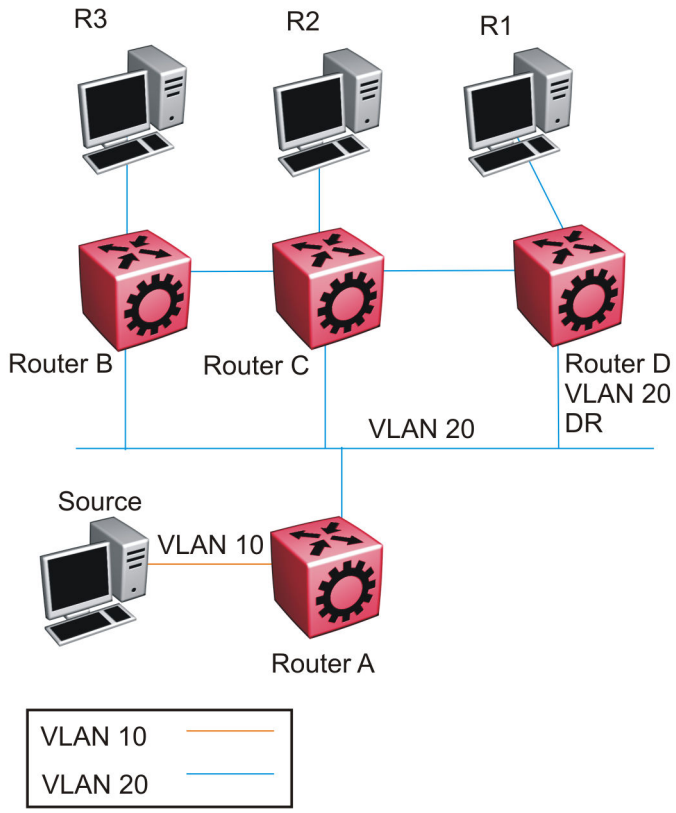


Figure 86: Shared VLAN topology

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 *Configuring IP Multicast Routing Protocols on Avaya Virtual Services Platform 9000,* 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 17: 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 39.

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 *Configuring QoS and ACL-Based Traffic Filtering on Avaya Virtual Services Platform 9000,* 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 *Configuring Security on Avaya Virtual Services Platform 9000,* 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 Resoluion Protocol (ARP) request-threshold defines the maximum number of outstanding, unresolved ARP requests. The default value for this function is 500 ARP requests. To avoid excessive amounts of subnet scanning that a virus can cause, 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 *Configuring IP Routing on Avaya Virtual Services Platform 9000,* 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 *Configuring IP Multicast Routing Protocols on Avaya Virtual Services Platform 9000,* 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 *Configuring Security on Avaya Virtual Services Platform 9000,* NN46250-601.
4. Prevent unknown devices from influencing the spanning tree topology.

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 15: 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 *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* 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 *Configuring Security on Avaya Virtual Services Platform 9000,* NN46250-601.

BPDU Filtering

Use Bridge Protocol Data Unit (BPDU) Filtering to prevent unknown devices from influencing the spanning tree topology. Virtual Services Platform 9000 supports BPDU Filtering for spanning tree groups (STG), Rapid Spanning Tree Protocol (RSTP), and Multiple Spanning Tree Protocol (MSTP).

Use BPDU Filtering to achieve the following results:

- Block the root selection process after an edge device, such as a laptop that uses Linux with STP enabled, is added to the network. Blocking the root selection process prevents unknown devices from influencing the spanning tree topology.
- Block BPDU flooding of the switch from an unknown device.

For more information about how to configure BPDU Filtering, see *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* NN46250-500.

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 they can obtain 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.

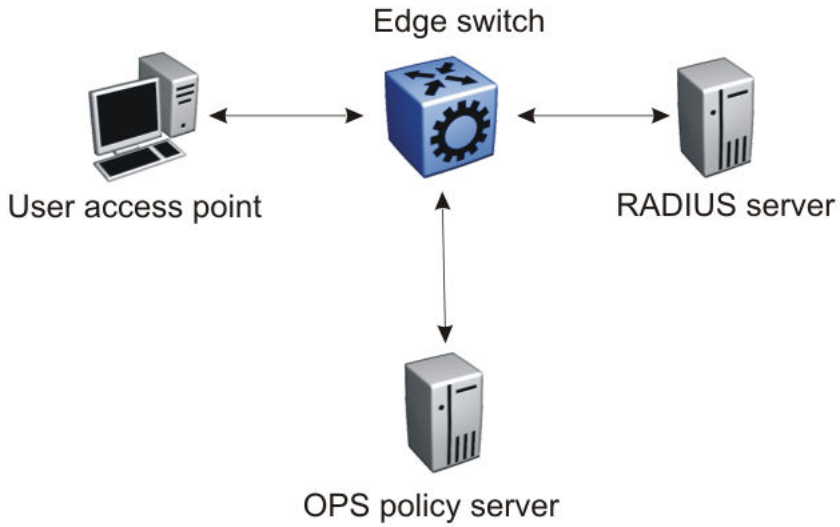


Figure 87: 802.1x and OPS interaction

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 *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* 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.

  Note:

  If you configure an access policy mode to deny, the system does not check **accesslevel** or **access-strict** information. With the access policy mode configured to deny, the system checks the mode and service, and if they match the system denies the connection. If you configure the access policy mode to allow, the system continues to check the **accesslevel** and **access-strict** information.

  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 *Configuring QoS and ACL-Based Traffic Filtering on Avaya Virtual Services Platform 9000,* 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 *Configuring VLANs and Spanning Tree on Avaya Virtual Services Platform 9000,* 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 *Configuring OSPF and RIP on Avaya Virtual Services Platform 9000,* NN46250-506 and *Configuring BGP Services on Avaya Virtual Services Platform 9000,* 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 88: Dedicated Ethernet management link on page 183) 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 89: Terminal server access on page 183).

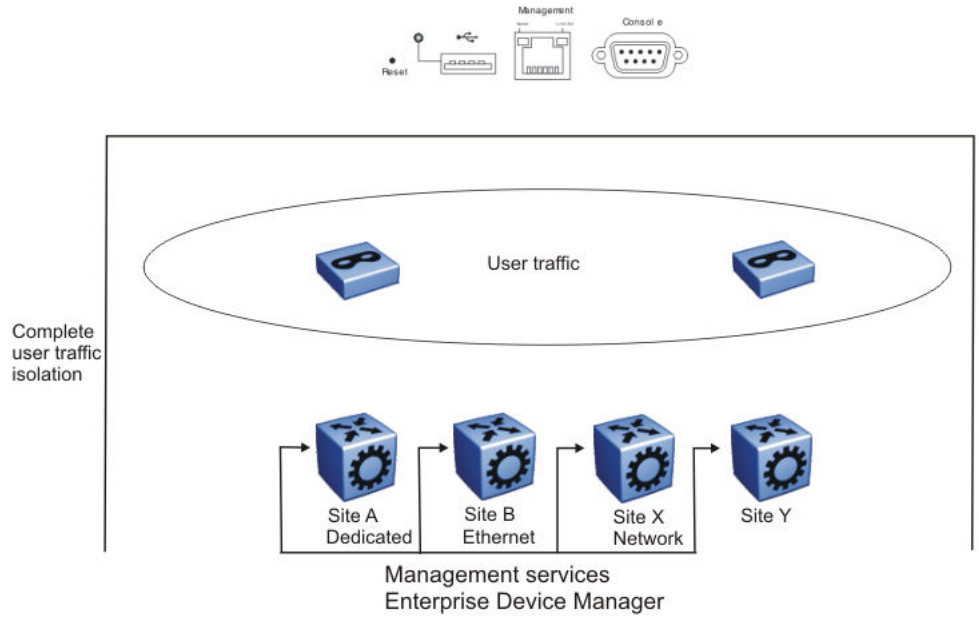


Figure 88: Dedicated Ethernet management link

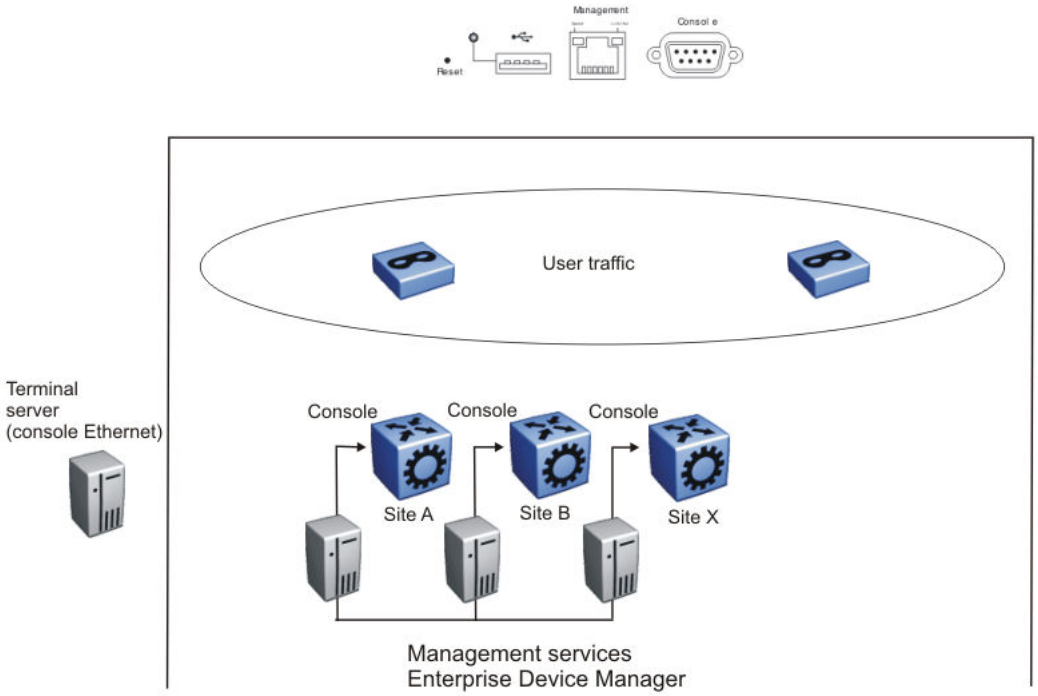


Figure 89: Terminal server access

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 16: 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 177.

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).

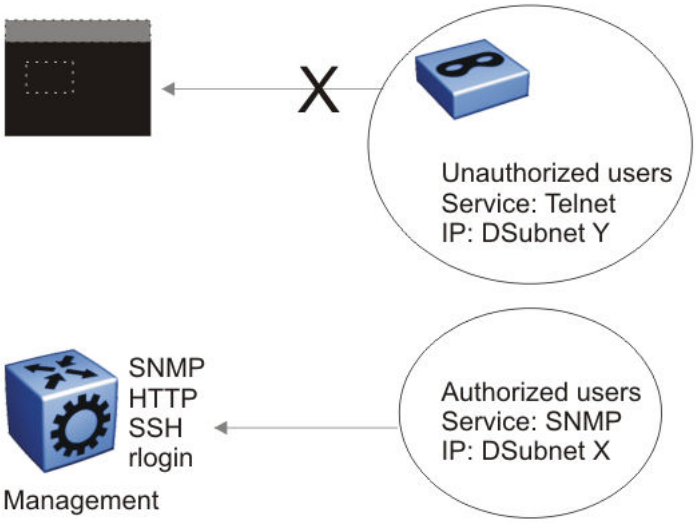


Figure 90: Access levels

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+)

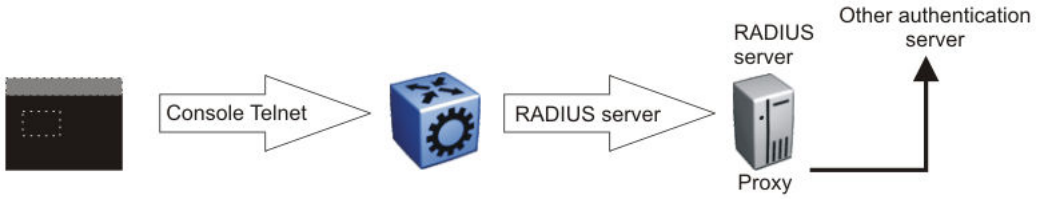


Figure 91: RADIUS server as proxy for stronger authentication

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 *Configuring Security on Avaya Virtual Services Platform 9000,* 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.

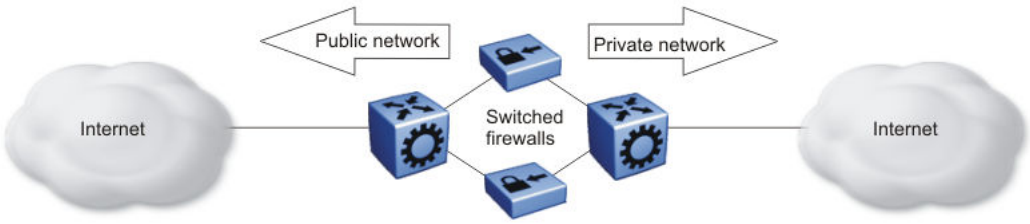


Figure 92: Firewall load balancing configuration

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 18: 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 *Configuring QoS and ACL-Based Traffic Filtering on Avaya Virtual Services Platform 9000,* 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 17: Traffic categories and ASC mappings

Traffic category	Application example	ASC
Network Control	Alarms and heartbeats	Critical

Traffic category		Application example	ASC
		Routing table updates	Network
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.

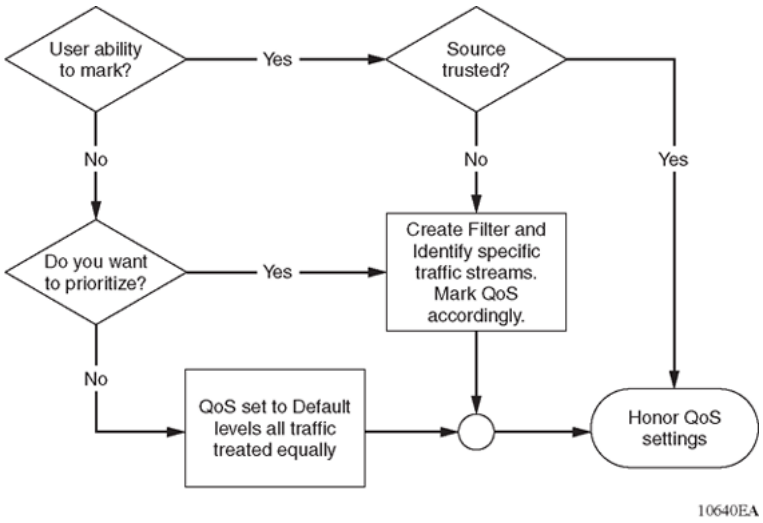


Figure 93: Filter decision-making process

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 18: 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

Comments? infodev@avaya.com

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	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.

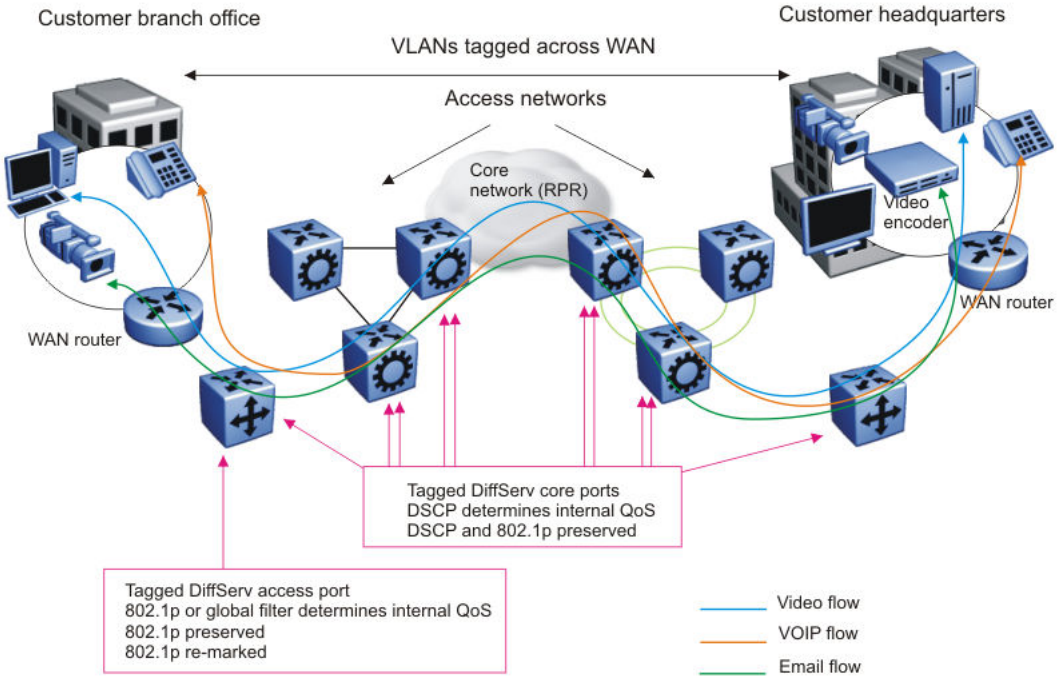


Figure 94: Trusted bridged traffic

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.

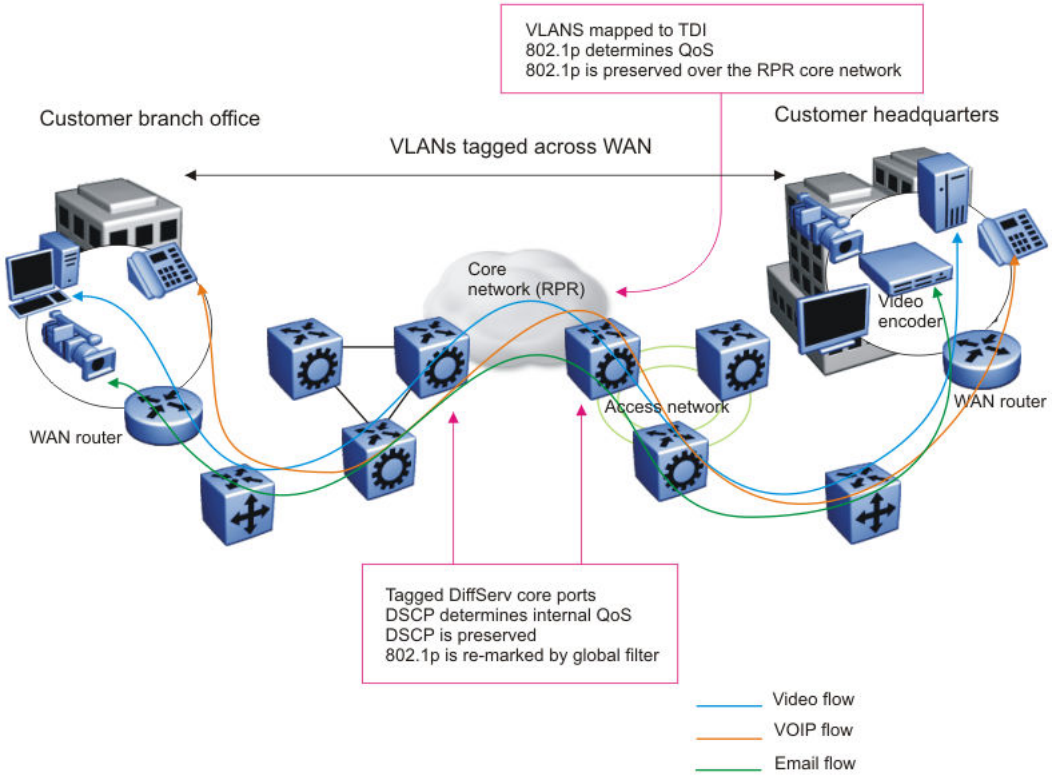


Figure 95: RPR QoS internetworking

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.

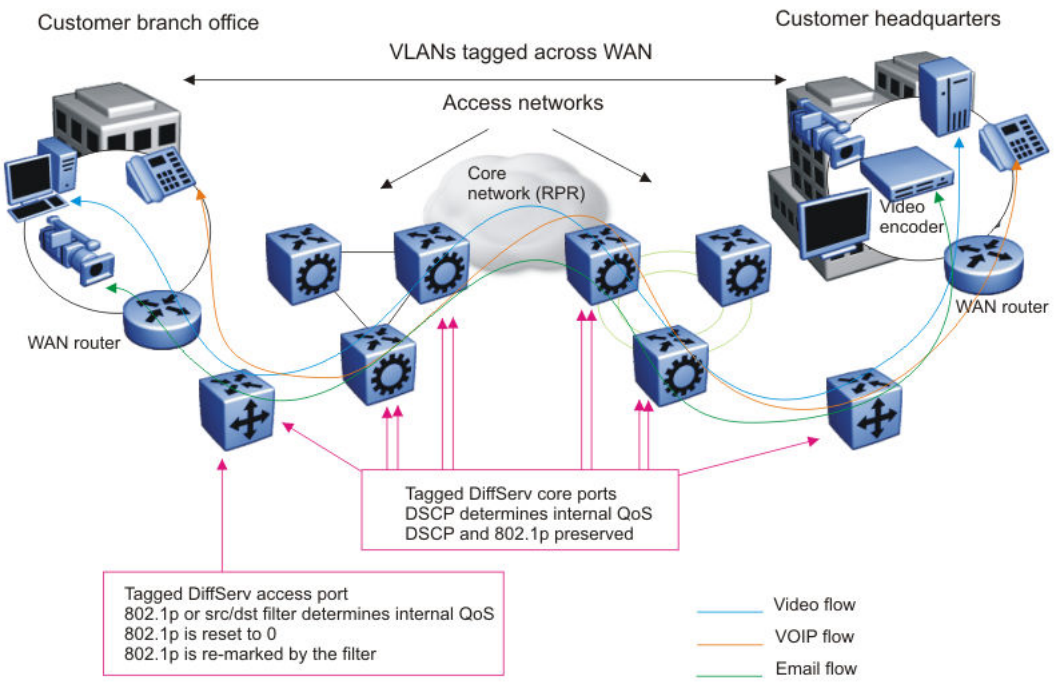


Figure 96: Trusted routed traffic

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 19: 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.

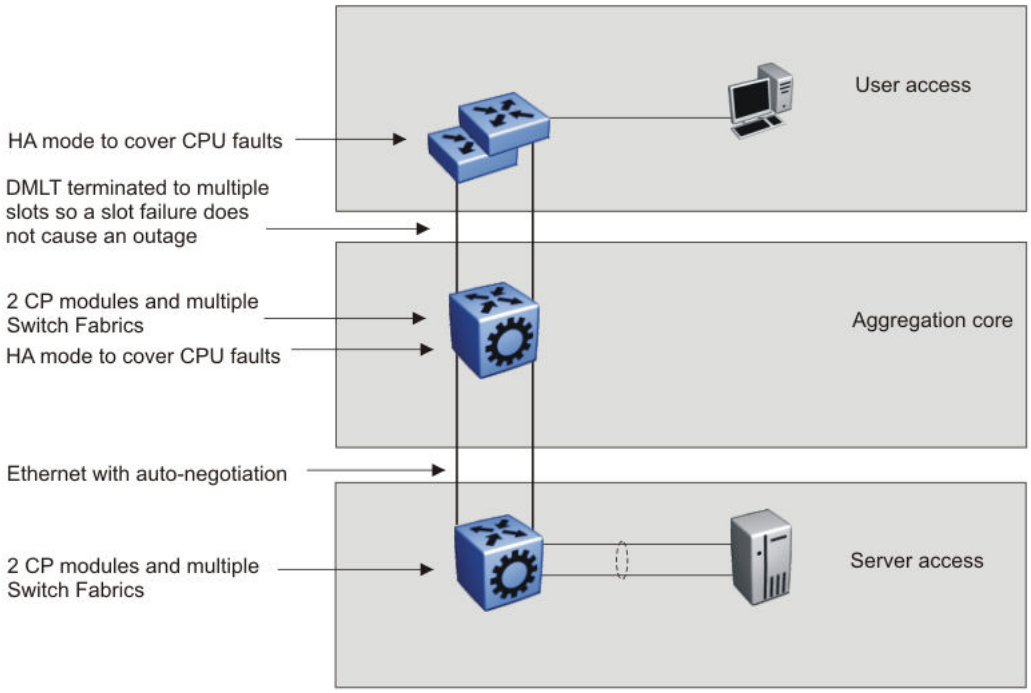


Figure 97: Layer 1 design example 1

The following figure uses Split MultiLink Trunking (SMLT) to provide switch redundancy and DMLT to provide module redundancy.

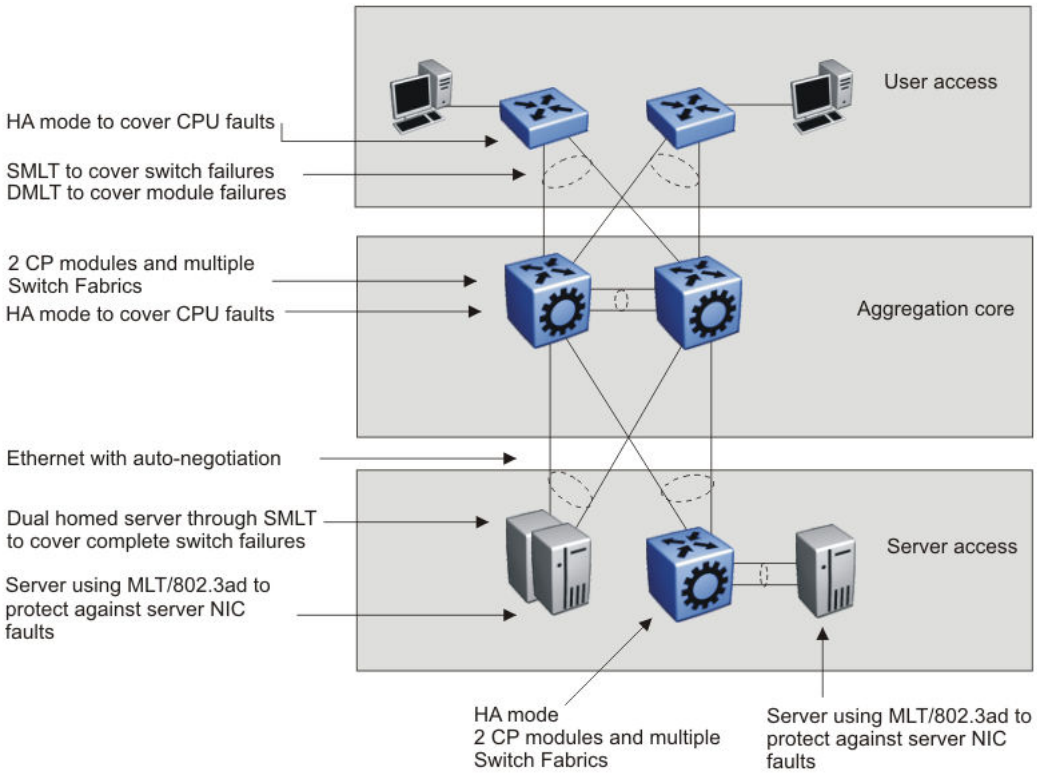


Figure 98: Layer 1 design example 2

The following figure is an example of a four-tiered topology. This example adds redundancy in the aggregation core.

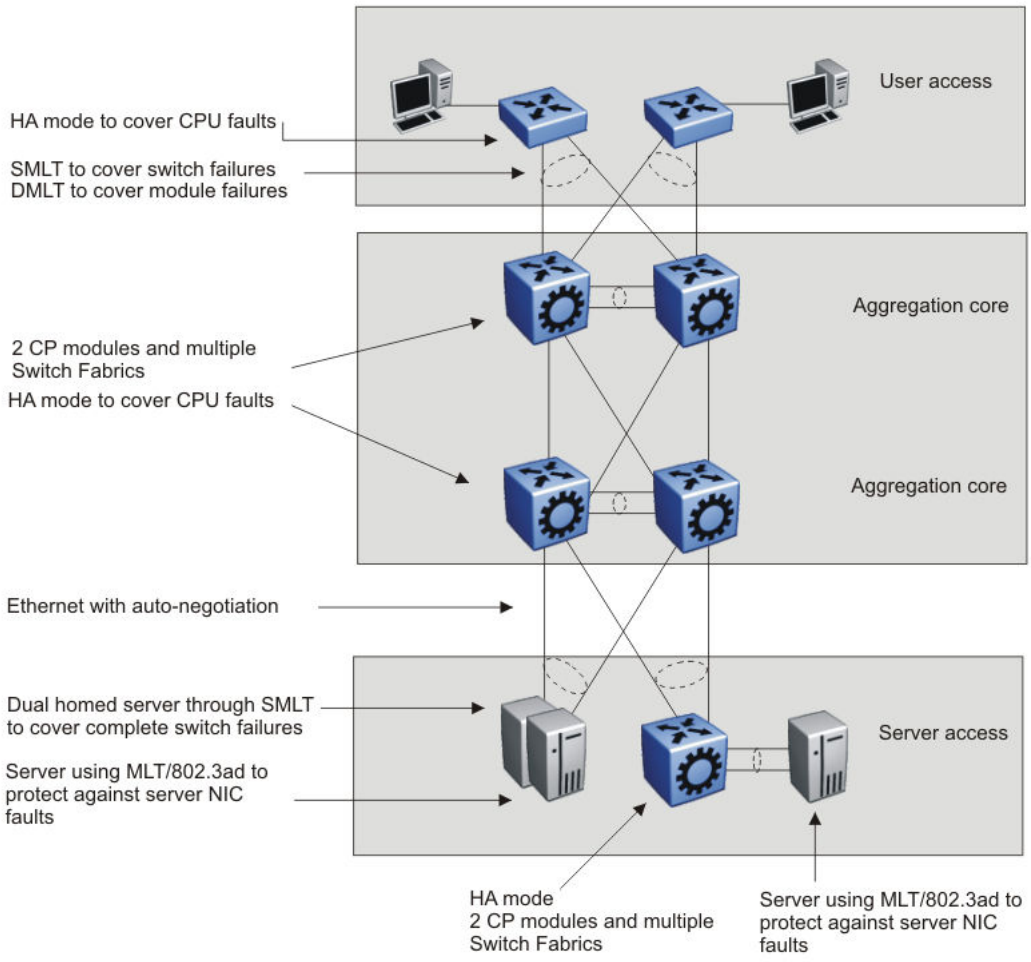


Figure 99: Layer 1 design example 3

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.

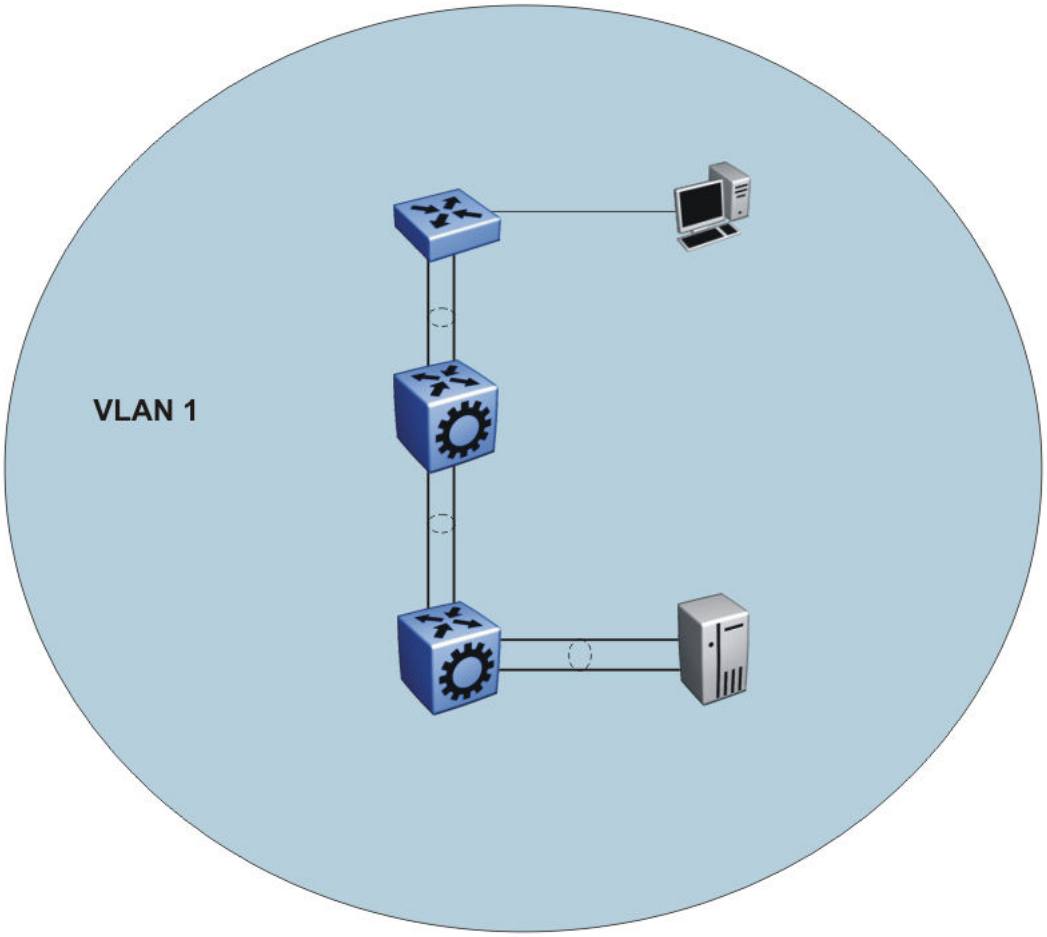


Figure 100: Layer 2 design example 1

Figure 101: Layer 2 design example 2 on page 203 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.

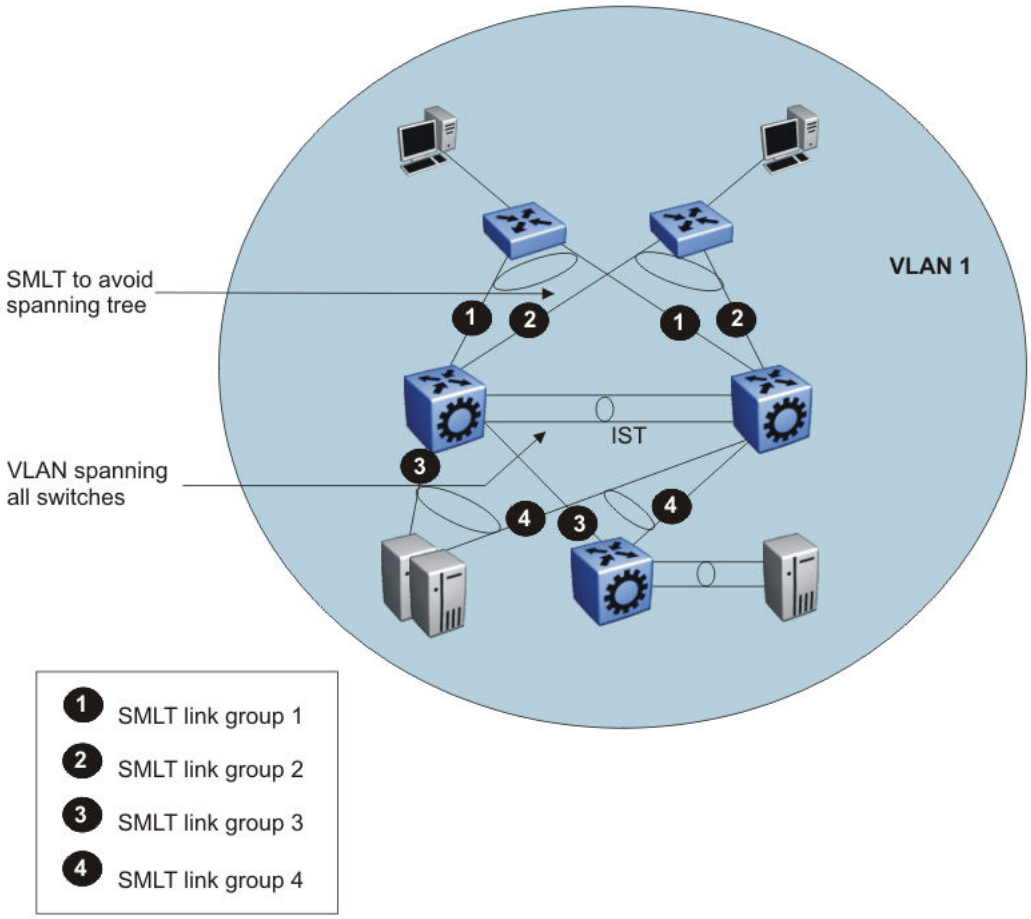


Figure 101: Layer 2 design example 2

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.

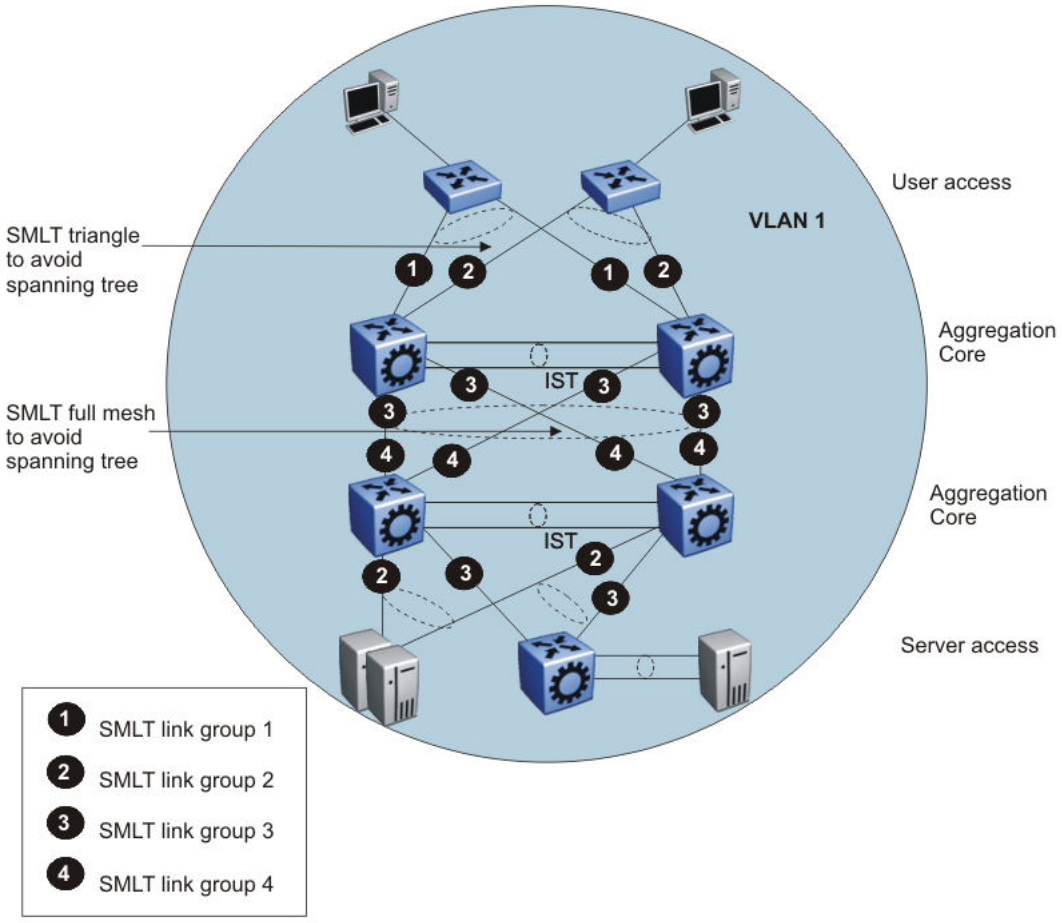


Figure 102: Layer 2 design example 3

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.

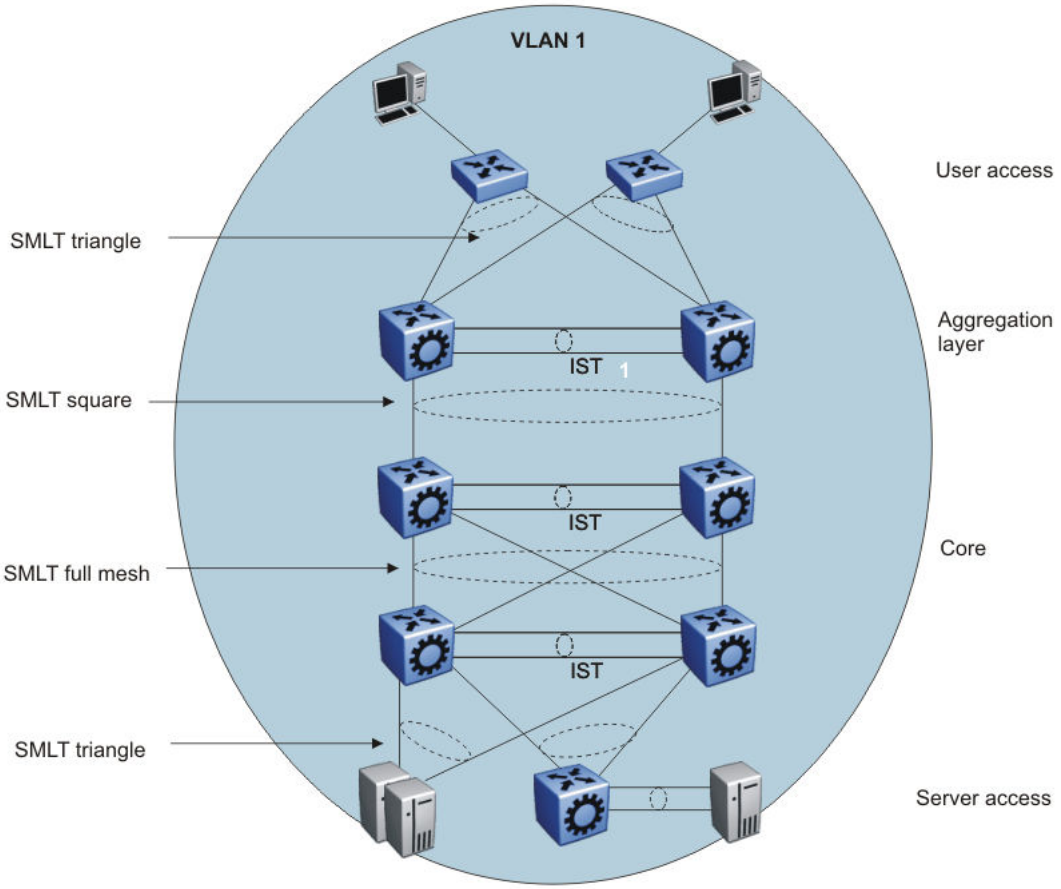


Figure 103: Layer 2 design example 4

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 104: Layer 3 design example 1 on page 206 uses redundant links.

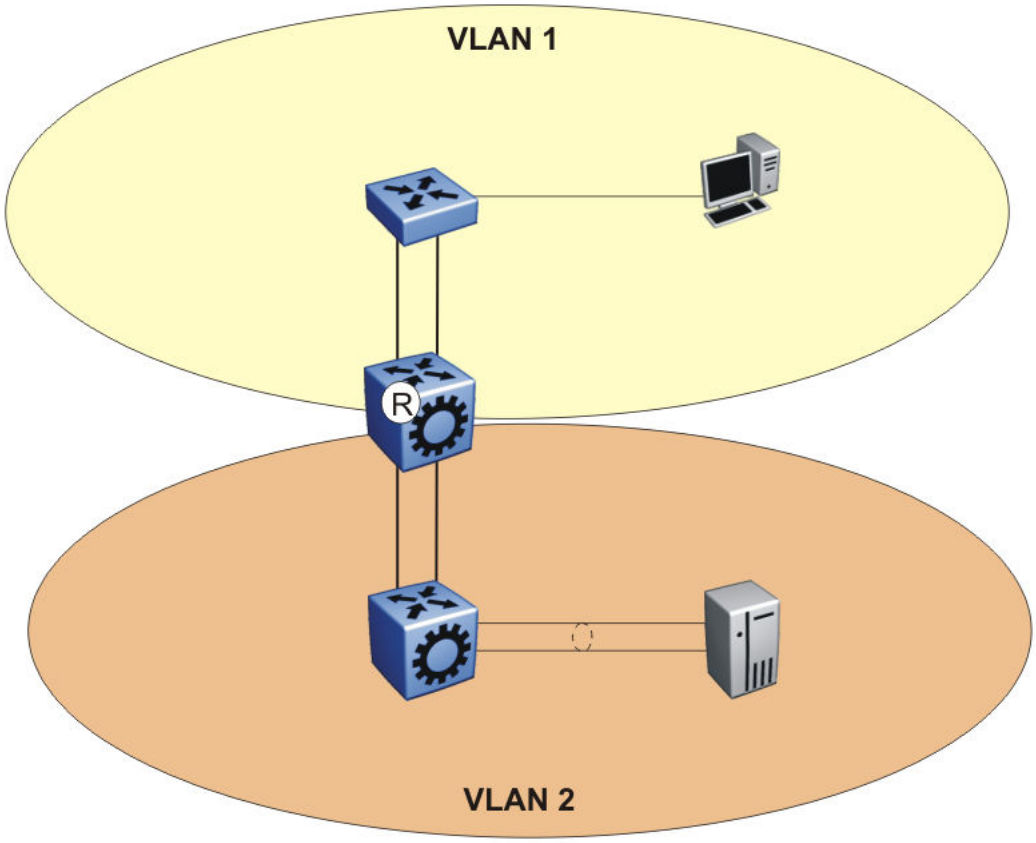


Figure 104: Layer 3 design example 1

Figure 105: Layer 3 design example 2 on page 207 uses the Virtual Router Redundancy Protocol (VRRP) to provide redundancy between the two switches.

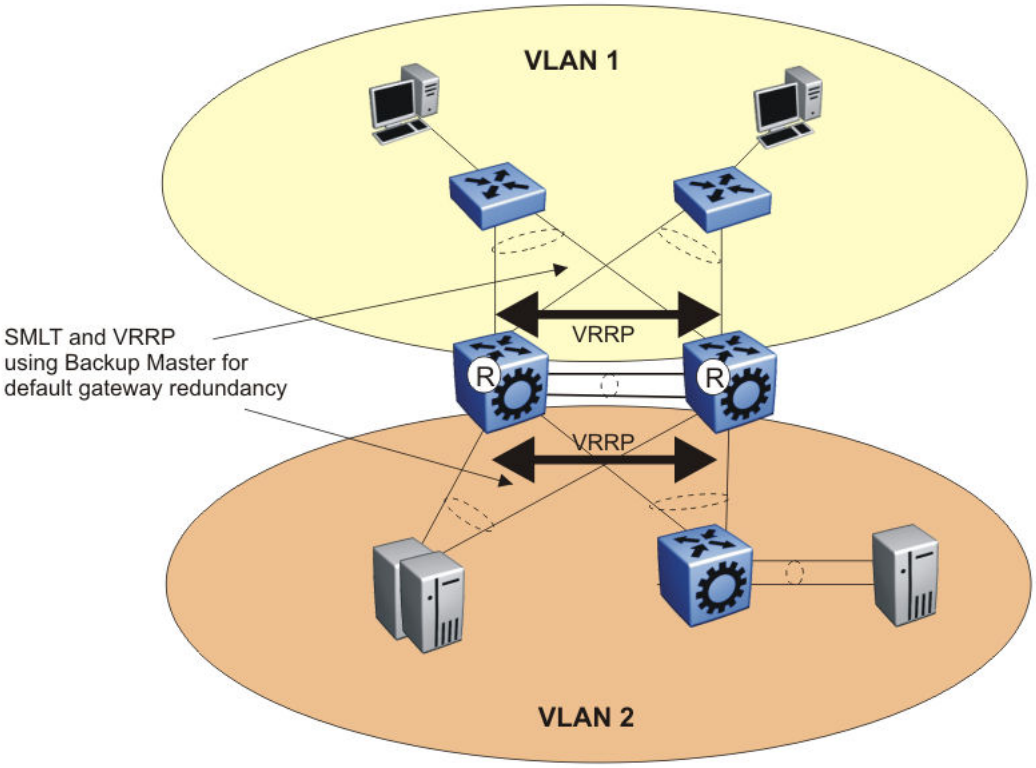


Figure 105: Layer 3 design example 2

Figure 106: Layer 3 design example 3 on page 208 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.

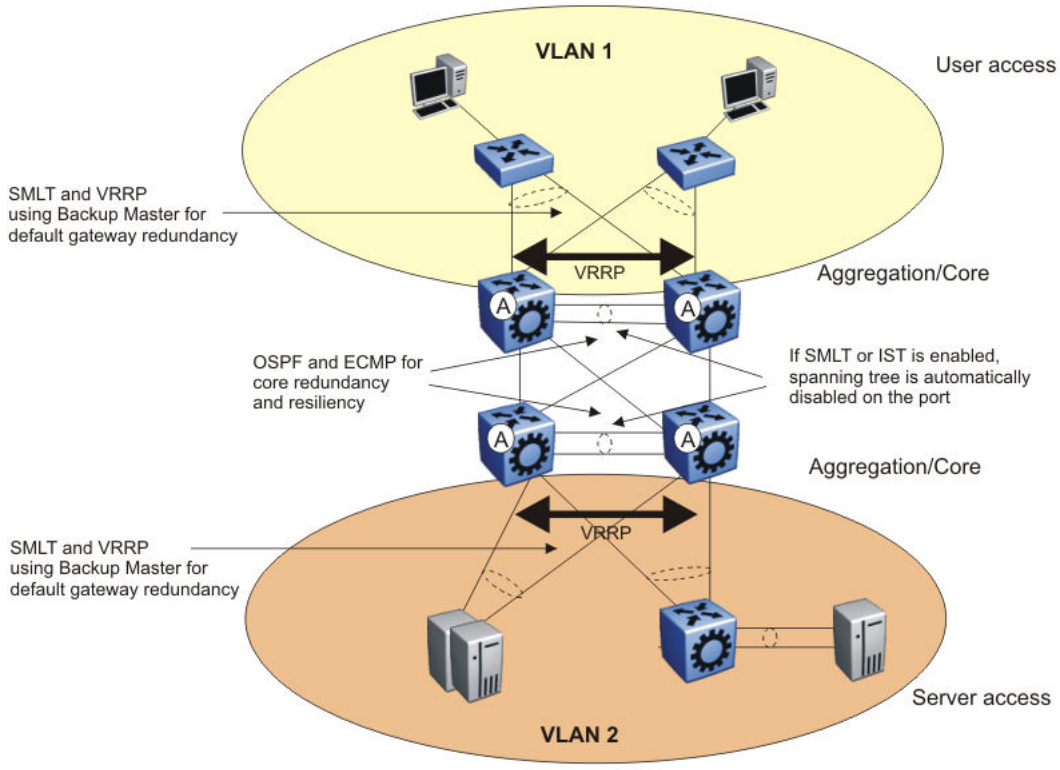


Figure 106: Layer 3 design example 3

Figure 107: Layer 3 design example 4 on page 209 uses one aggregation pair. You can connect multiple aggregation-layer and access switches to create a very large, scalable network.

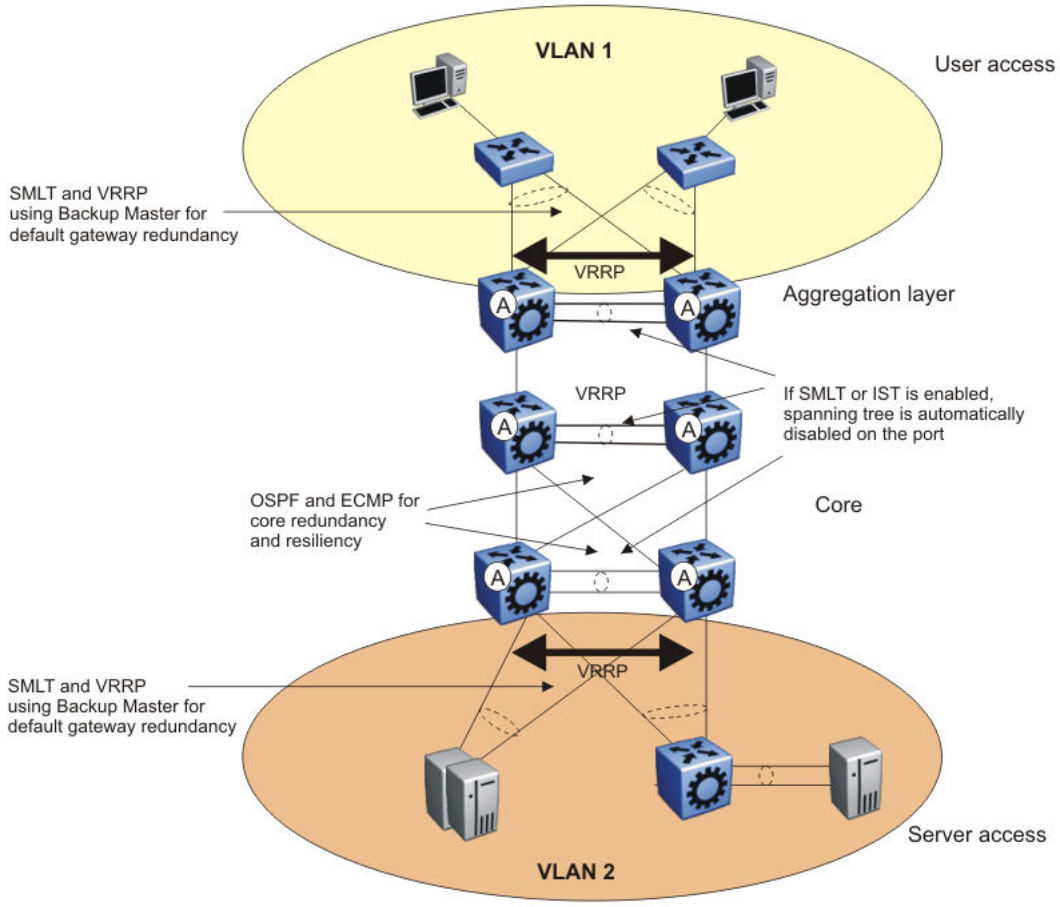


Figure 107: Layer 3 design example 4

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 108: Redundant network design on page 210 shows a redundant network design that can perform these functions.

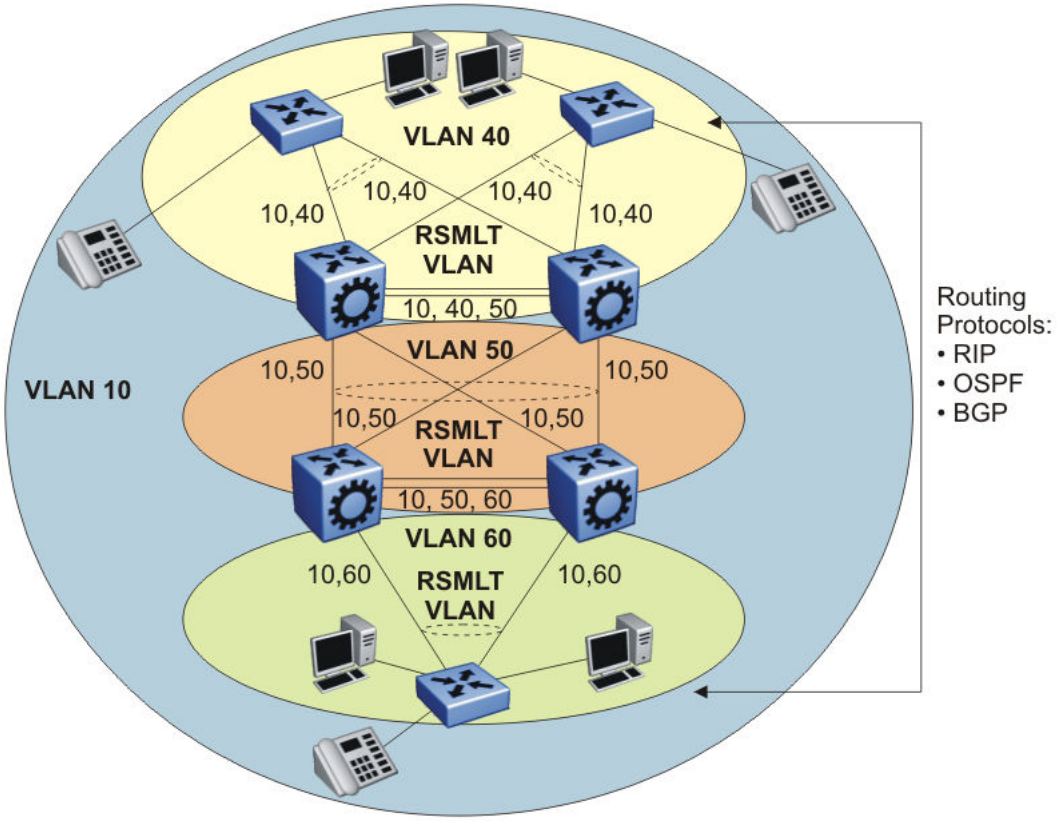


Figure 108: Redundant network design

In Figure 108: Redundant network design on page 210, 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 39 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 61.

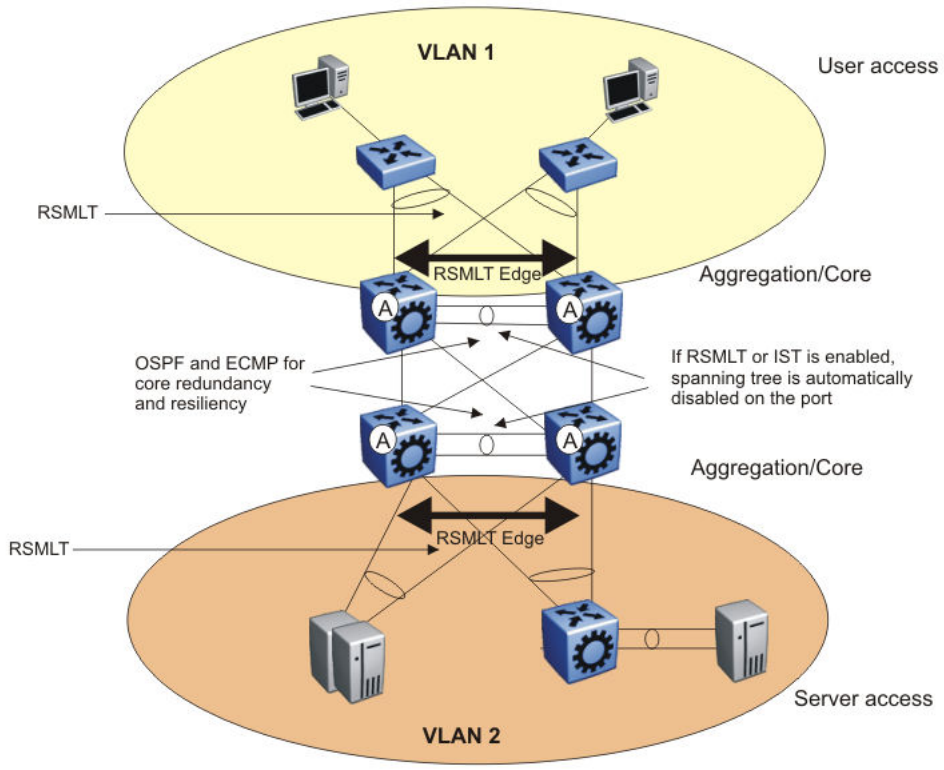


Figure 109: RSMLT Layer 2 Edge

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.

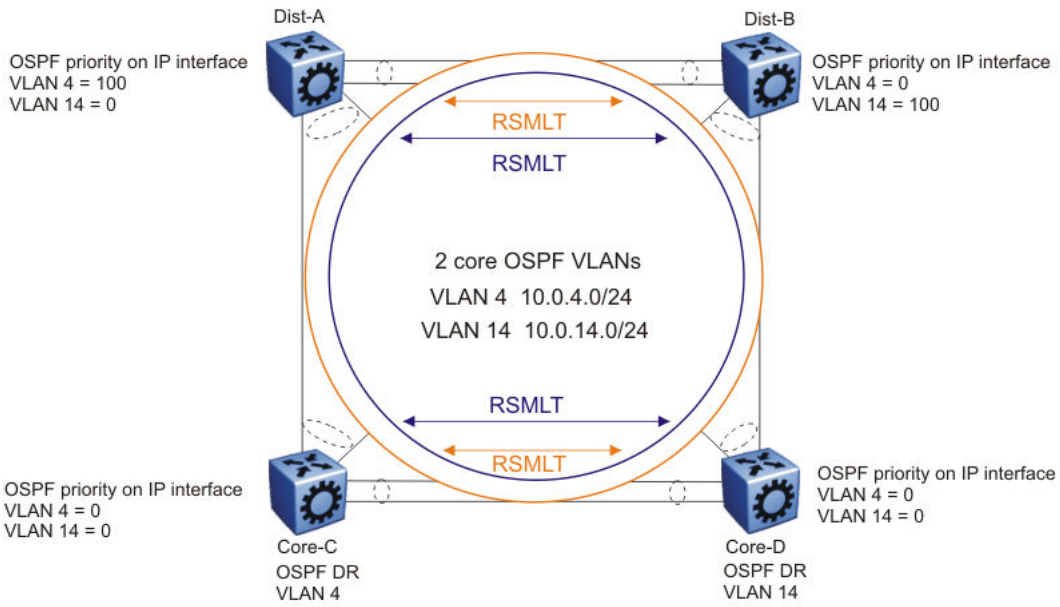


Figure 110: RSMLT with dual core VLANs

Chapter 20: 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.

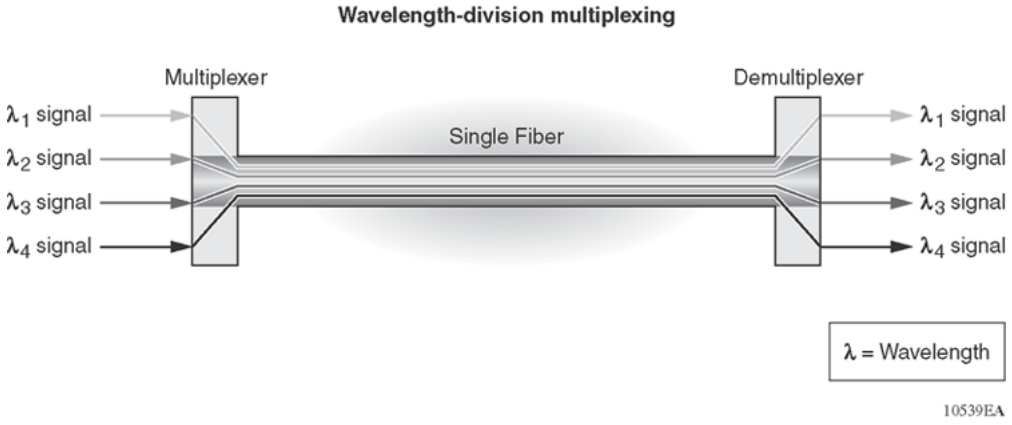


Figure 111: Wavelength division multiplexing

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 *Installing Transceivers and Optical Components on Avaya Virtual Services Platform 9000,* NN46250-305.

Chapter 21: Software and hardware scaling capabilities

This chapter details the software and hardware scaling capabilities of Avaya Virtual Services Platform 9000. The information in *Release Notes for Avaya Virtual Services Platform 9000,* 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 19: Module capabilities

Component	Maximum number supported
9012QQ-2 I/O module	
40 GbE fiber connections	9010 chassis: 96 (8 x 12) 9012 chassis: 120 (10 x 12)
Processor	800 MHz dual core
9024XL I/O module	
10 GbE fiber connections	9010 chassis: 192 (8 x 24) 9012 chassis: 240 (10 x 24)
Processor	1 GHz
9048XS-2 I/O module	
10 GbE fiber connections	9010 chassis: 384 (8 x 48) 9012 chassis: 480 (10 x 48)
Processor	800 MHz dual core
9048GB I/O module	
GbE fiber connections	9010 chassis: 384 (8 x 48) 9012 chassis: 480 (10 x 48)
Processor	1 GHz
9048GT I/O module	

Component	Maximum number supported
10/100/1000 copper connections	9010 chassis: 384 (8 x 48) 9012 chassis: 480 (10 x 48)
Processor	1 GHz
9080CP 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

Table 20: VSP 9010 AC chassis capabilities

Component	Maximum number supported
CP modules	2
Interface modules	8
SF modules	6 If you install only first generation I/O modules, you must install a minimum of three SF modules in the chassis. If you install second generation I/O modules, you must install five SF modules. Always install an SF module in both slots SF1 and SF4.
Power Supplies	8
Total power capacity	• 21.8 kW when connected to 220 VAC • 16 kW when connected to 110 VAC
Jumbo packets	9600 bytes for IPv4 9500 bytes for IPv6

Table 21: VSP 9012 chassis capabilities

Component	Maximum number supported
CP modules	2
Interface modules	10
SF modules	6 If you install only first generation I/O modules, you must install a minimum of three SF modules in the chassis. If you install second generation I/O modules, you must install five SF modules. Always install an SF module in both slots SF1 and SF4.

Component	Maximum number supported
Auxiliary slots	2
Power supplies	6
Total power capacity	• 16.3 kW when connected to 220 VAC • 12 kW when connected to 110 VAC
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 22: 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
LACP MAC Addresses	20
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
Source MAC-based VLANs	100
Split Multi-Link Trunking (SMLT)	511 groups plus 1 IST MLT
SMLT ports per group	16
VLACP Interfaces	128
VLACP MAC address	20
Layer 3	

	Maximum number supported
Address Resolution Protocol (ARP) for each port, VRF, or VLAN	64,000 entries total
BGP peers	256
BGP Internet peers (full)	3
BGP routes	1.5 million
BGP+ routes	128,000
Circuitless IP interfaces	256
ECMP routes	64,000
Number of paths per ECMP route (fastpath)	8
FIB IPv4 routes	500,000 for first or second generation Input/Output (I/O) modules in first generation mode. **Note:** Both first generation I/O modules and second generation I/O modules require a Premier license.
FIB IPv6 routes (Not supported in Release 4.0)	128,000 for first or second generation Input/Output (I/O) modules in first generation mode. 256,000 for second generation I/O modules in second generation mode. **Note:** Both first generation I/O modules and second generation I/O modules require a Premier license.
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. • For first or second generation modules in first generation mode: 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. • For second generation modules in second generation mode: The maximum number of 1,000,000 IPv4 routes is possible when no IPv6 routes are configured, and the maximum number of IPv6 routes is 256,000 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

	Maximum number supported
	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
IPv6 dynamic neighbors/interface (Not supported in Release 4.0)	64K
IPv6 interfaces (Not supported in Release 4.0)	4,087 (4,084 VLAN and 3 management [1/1, 2/1, virtual IP])
IPv6 routes (fastpath) (Not supported in Release 4.0)	128,000
IPv6 static neighbors (Not supported in Release 4.0)	1,000
IPv6 static routes (Not supported in Release 4.0)	10,000
IPv6 tunnels (Not supported in Release 4.0)	2,000
Multicast IGMP interfaces	4,084
Multicast IGMP instances	on 64 VRFs
Multicast source and group (S, G)	6,000 for each system, including VRFs
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

	Maximum number supported
PIM interfaces	512 active; 4084 passive
PIM instances	on GRT only
RIB IPv4 routes	3 * fastpath routes
RIP instances	64 (one for each VRF)
RIP interfaces	200
RIP routes	2,500 for each VRF 10,000 for each system
RSMLT interfaces (IPv4/IPv6) (IPv6 not supported in Release 4.0)	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) (Not supported in Release 4.0)	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) (IPv6 not supported in Release 4.0)	16,000
Port shapers (IPv4 and IPv6) (IPv6 not supported in Release 4.0)	480
Access control lists (ACL) for each chassis (IPv4)	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
SPBM	
ARP entries (routed)	64,000

	Maximum number supported
MAC entries	128,000 (combination of ARP entries and Layer 2 MACs)
Backbone MAC	1,000
IP routes in the Global Router	100,000 (combination of OSPF and IS-IS)
IS-IS adjacencies	128
Layer 2 VSNs	4,000 **Note:** Layer 2 VSNs for SPB services available under the base license.
VLANs in VRF	1,600
Layer 3 VSNs	512
OSPF Routers per area	620
The memory used to store rmon info is shared between alarms and events. This causes the max limits to vary on how many event-alarm pairs there are. The following scaling numbers are minimum config alarms/ events. • For 250K memory: - configuring 1 event leaves room for 111 alarms - configuring 86 events leaves room for 84 alarms • For 4M memory: - configuring 1 event leaves room for 2243 alarms - configuring 1692 events leaves room for 1689 alarms	

Chapter 22: 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 23: Supported standards

Standard	Description
802.1ag	Connectivity Fault Management
802.1ah	Provider Backbone Bridges (MacInMac encapsulation)
802.1aq	Shortest Path Bridging (SPB)
802.1AX	Link Aggregation Control Protocol (LACP)
802.1D	MAC bridges (Spanning Tree)
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 (MSTP)
802.1t	802.1D maintenance
802.1v	VLAN Classification by Protocol and Port
802.1w-2001	Rapid Spanning Tree protocol (RSTP)
802.1X	Extensible Authentication Protocol (EAP), and EAP over LAN (EAPoL)

Standard	Description
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)
802.3ae	10 Gigabit Ethernet
802.3an	10 Gigabit Copper
802.3i	10BaseT
802.3u	100BaseT
802.3x	flow control
802.3z	Gigabit Ethernet
802.3ba	40Gbs Ethernet

Supported RFCs

The following table and sections list the RFCs that Avaya Virtual Services Platform 9000 supports.

Table 24: Supported request for comments

Request for comment	Description
draft-grant-tacacs-02.txt	TACACS+ Protocol
RFC 768	UDP Protocol
RFC 783	Trivial File Transfer Protocol (TFTP)
RFC 791	Internet Protocol (IP)
RFC 792	Internet Control Message Protocol (ICMP)
RFC 793	Transmission Control Protocol (TCP)
RFC 826	Address Resolution Protocol (ARP)
RFC 854	Telnet protocol
RFC 894	A standard for the Transmission of IP Datagrams over Ethernet Networks
RFC 896	Congestion control in IP/TCP internetworks
RFC 903	Reverse ARP Protocol
RFC 906	Bootstrap loading using TFTP
RFC 950	Internet Standard Subnetting Procedure
RFC 951	BootP
RFC 959, RFC 1350, and RFC 2428	IPv6 FTP and TFTP client and server

Request for comment	Description
RFC 1027	Using ARP to implement transparent subnet gateways/Nortel Subnet based VLAN
RFC 1058	RIPv1 Protocol
RFC 1112	IGMPv1
RFC 1122	Requirements for Internet Hosts
RFC 1253	OSPF
RFC 1256	ICMP Router Discovery
RFC 1258	IPv6 Rlogin server
RFC 1305	Network Time Protocol v3 Specification, Implementation and Analysis
RFC 1340	Assigned Numbers
RFC 1519	Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy
RFC1541	Dynamic Host Configuration Protocol1
RFC1542	Clarifications and Extensions for the Bootstrap Protocol
RFC 1583	OSPFv2
RFC 1587	The OSPF NSSA Option
RFC 1591	DNS Client
RFC 1723	RIP v2 – Carrying Additional Information
RFC 1745	BGP / OSPF Interaction
RFC 1771 and RFC 1772	BGP-4
RFC 1812	Router requirements
RFC 1866	HyperText Markup Language version 2 (HTMLv2) protocol
RFC 1965	BGP-4 Confederations
RFC 1966	BGP-4 Route Reflectors
RFC 1981	Path MTU discovery
RFC 1997	BGP-4 Community Attributes
RFC 1998	An Application of the BGP Community Attribute in Multi-home Routing
RFC 2068	Hypertext Transfer Protocol
RFC 2131	Dynamic Host Control Protocol (DHCP)
RFC 2138	RADIUS Authentication
RFC 2139	RADIUS Accounting
RFC 2178	OSPF MD5 cryptographic authentication / OSPFv2
RFC 2236	IGMPv2 for snooping
RFC 2270	BGP-4 Dedicated AS for sites/single provide

Request for comment	Description
RFC 2328	OSPFv2
RFC 2338	VRRP: Virtual Redundancy Router Protocol
RFC 2362	PIM-SM
RFC 2385	BGP-4 MD5 authentication
RFC 2439	BGP-4 Route Flap Dampening
RFC 2453	RIPv2 Protocol
RFC 2460	IPv6 base stack
RFC 2464	Transmission of IPv6 packets over Ethernet networks
RFC 2545 and RFC 4710	IPv6 capable BGPv4+
RFC 2616	IPv6 HTTP server
RFC 2710 and RFC 3810	MLD (host-mode only)
RFC 2740	OSPFv3
RFC 2796	BGP Route Reflection – An Alternative to Full Mesh IBGP
RFC 2819	RMON
RFC 2874	DNS Extensions for IPv6
RFC 2918	Route Refresh Capability for BGP-4
RFC 2992	Analysis of an Equal-Cost Multi-Path Algorithm
RFC 3046	DHCP Option 82
RFC 3065	Autonomous System Confederations for BGP
RFC 3162	IPv6 RADIUS client
RFC 3315	IPv6 DHCP Relay
RFC 3376	Internet Group Management Protocol, v3
RFC 3411 and RFC 2418	SNMP over IPv6 networks
RFC 3513	Internet Protocol Version 6 (IPv6) Addressing Architecture
RFC 3569	An overview of Source-Specific Multicast (SSM)
RFC 3587	IPv6 Global Unicast Address Format
RFC 3768 and draft-ietf-vrrp-ipv6-spec-08.txt	IPv6 capable VRRP
RFC 4213	IPv6 configured tunnel
RFC 4250–RFC 4256	SSH server and client support
RFC 4443	Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
RFC 4861	IPv6 Neighbor discovery
RFC 4862	IPv6 stateless address autoconfiguration
RFC 4893	BGP support for Four-octet AS Number Space

Request for comment	Description
RFC 6329	IS-IS Extensions supporting Shortest Path Bridging

Quality of service

Table 25: Supported request for comments

Request for comment	Description
RFC 2474 and RFC 2475	DiffServ Support
RFC 2597	Assured Forwarding PHB Group
RFC 2598	An Expedited Forwarding PHB

Network management

Table 26: Supported request for comments

Request for comment	Description
RFC 959	File Transfer Protocol
RFC 1155	SMI
RFC 1157	SNMP
RFC 1215	Convention for defining traps for use with the SNMP
RFC 1258	BSD Rlogin
RFC 1269	Definitions of Managed Objects for the Border Gateway Protocol: v3
RFC 1271	Remote Network Monitoring Management Information Base
RFC 1305	Network Time Protocol v3 Specification, Implementation and Analysis3
RFC 1350	The TFTP Protocol (Revision 2)
RFC 1354	IP Forwarding Table MIB
RFC 1389	RIP v2 MIB Extensions
RFC 1757	Remote Network Monitoring Management Information Base
RFC 1907	Management Information Base for Version 2 of the Simple Network Management Protocol (SNMPv2)
RFC 1908	Coexistence between v1 & v2 of the Internet-standard Network Management Framework
RFC 1930	Guidelines for creation, selection, and registration of an Autonomous System (AS)

Request for comment	Description
RFC 2428	FTP Extensions for IPv6
RFC 2541	Secure Shell Protocol Architecture
RFC 2571	An Architecture for Describing SNMP Management Frameworks
RFC 2572	Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)
RFC 2573	SNMP Applications
RFC 2574	User-based Security Model (USM) for v3 of the Simple Network Management Protocol (SNMPv3)
RFC 2575	View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)
RFC 2576	Coexistence between v1, v2, & v3 of the Internet standard Network Management Framework
RFC 2616	IPv6 HTTP server
RFC 2819	Remote Network Monitoring Management Information Base
RFC 4250	SSH Protocol Assigned Numbers
RFC 4251	SSH Protocol Architecture
RFC 4252	SSH Authentication Protocol
RFC 4253	SSH Transport Layer Protocol
RFC 4254	SSH Connection Protocol
RFC 4255	DNS to Securely Publish SSH Key Fingerprints
RFC 4256	Generic Message Exchange Authentication for SSH

MIBs

Table 27: Supported request for comments

Request for comment	Description
RFC 1156	MIB for network management of TCP/IP
RFC 1212	Concise MIB definitions
RFC 1213	TCP/IP Management Information Base
RFC 1354	IP Forwarding Table MIB
RFC 1389	RIP v2 MIB Extensions
RFC 1398	Ethernet MIB
RFC 1442	Structure of Management Information for version 2 of the Simple Network Management Protocol (SNMPv2)

Request for comment	Description
RFC 1450	Management Information Base for v2 of the Simple Network Management Protocol (SNMPv2)
RFC 1573	Interface MIB
RFC 1650	Definitions of Managed Objects for the Ethernet-like Interface Types
RFC 1657	BGP-4 MIB using SMIv2
RFC 1724	RIPv2 MIB extensions
RFC 1850	OSPF MIB
RFC 2096	IP Forwarding Table MIB
RFC 2452	IPv6 MIB: TCP MIB
RFC 2454	IPv6 MIB: UDP MIB
RFC 2466	IPv6 MIB: ICMPv6 Group
RFC 2578	Structure of Management Information v2 (SMIv2)
RFC 2674	Bridges with Traffic MIB
RFC 2787	Definitions of Managed Objects for the Virtual Router Redundancy Protocol
RFC 2863	Interface Group MIB
RFC 2925	Remote Ping, Traceroute & Lookup Operations MIB
RFC 2932	IPv4 Multicast Routing MIB
RFC 2933	IGMP MIB
RFC 2934	PIM MIB
RFC 3416	v2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)
RFC 4022	Management Information Base for the Transmission Control Protocol (TCP)
RFC 4113	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.

Comments? infodev@avaya.com

Table 28: 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
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	RFC1907	rfc1907.mib

Standard MIB name	Institute of Electrical and Electronics Engineers/Request for Comments (IEEE/RFC)	File name
Network Management Protocol version 2 (SNMPv2)		
STDMIB21—Interfaces Group MIB using SMIv2	RFC2233	rfc2233.mib
STDMIB26a—An Architecture for Describing SNMP Management Frameworks	RFC2571	rfc2571.mib
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
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

Standard MIB name	Institute of Electrical and Electronics Engineers/Request for Comments (IEEE/RFC)	File name
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 29: 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
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

Glossary

Backbone Core Bridge (BCB)	Backbone Core Bridges (BCBs) form the core of the SPBM network. The BCBs are SPBM nodes that do not terminate the VSN services. BCBs forward encapsulated VSN traffic based on the Backbone MAC Destination Address (B-MAC-DA). A BCB can access information to send that traffic to any Backbone Edge Bridges (BEBs) in the SPBM backbone.
Backbone Edge Bridge (BEB)	Backbone Edge Bridges (BEBs) are SPBM nodes where Virtual Services Networks (VSNs) terminate. BEBs handle the boundary between the core MAC-in-MAC Shortest Bath Bridging MAC (SPBM) domain and the edge customer 802.1Q domain. A BEB node performs 802.1ah MAC-in-MAC encapsulation and decapsulation for the Virtual Services Network (VSN).
Backbone MAC (B-MAC)	Provider Backbone Bridging (PBB) MAC-in-MAC encapsulation encapsulates customer MAC addresses in Backbone MAC (B-MAC) addresses. MAC-in-MAC encapsulation defines a BMAC-DA and BMAC-SA to identify the backbone source and destination addresses. The originating node creates a MAC header that SPBM uses for delivery from end to end. As the MAC header stays the same across the network, no need exists to swap a label or perform a route lookup at each node, allowing the frame to follow the most efficient forwarding path end to end. In Shortest Path Bridging MAC (SPBM), each node has a System ID, which is used in the topology announcement. This same System ID also serves as the switch Backbone MAC address (B-MAC), which is used as the source and destination MAC address in the SPBM network.
Backbone VLAN identifier (B-VID)	The Backbone VLAN identifier (B-VID) indicates the Shortest Path Bridging MAC (SPBM) B-VLAN associated with the SPBM instance.
bit error rate (BER)	The ratio of the number of bit errors to the total number of bits transmitted in a specific time interval.
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.
Connectivity Fault Management (CFM)	Connectivity Fault Management is a mechanism to debug connectivity issues and to isolate faults within the Shortest Path Bridging MAC (SPBM) network. CFM operates at Layer 2 and provides the equivalent of ping and traceroute. IEEE 802.1ag Connectivity Fault Management (CFM) divides or

	separates a network into administrative domains called Maintenance Domains (MD).
Control Processor Unit High Availability (CPU-HA)	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 one second with the most recent topology data.
cooling module (9010CM)	The cooling module is a hot swappable fan tray used to cool the Control Processor, I/O, and Switch Fabric modules in the Virtual Services Platform 9010. Two cooling modules are installed horizontally in the front of the chassis.
Customer MAC (C-MAC)	For customer MAC (C-MAC) addresses, which is customer traffic, to forward across the service provider back, SPBM uses IEEE 802.1ah Provider Backbone Bridging MAC-in-MAC encapsulation. The system encapsulates C-MAC addresses within a backbone MAC (B-MAC) address pair made up of a BMAC destination address (BMAC-DA) and a BMAC source address (BMAC-SA).
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 separate by wavelength at the distant end.
Designated Intermediate System (DIS)	A Designated Intermediate System (DIS) is the designated router in Intermediate System to Intermediate System (IS-IS) terminology. You can modify the priority to affect the likelihood of a router being elected the designated router. The higher the priority, the more likely the router is to be elected as the DIS. If two routers have the same priority, the router with the highest MAC address (Sequence Number Packet [SNP] address) is elected as the DIS.
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).
I/O cooling module (9012FC)	The I/O cooling module is a hot swappable fan tray used to cool the I/O and CP modules in the Virtual Services Platform 9012.
Intermediate System to Intermediate System (IS-IS)	Intermediate System to Intermediate System(IS-IS) is a link-state, interior gateway protocol. ISO terminology refers to routers as Intermediate Systems (IS), hence the name Intermediate System to Intermediate System (IS-IS). IS-IS operation is similar to Open Shortest Path First (OSPF). In Shortest Path Bridging MAC (SPBM) networks, IS-IS discovers network topology and builds shortest path trees between network nodes that IS-IS uses for forwarding unicast traffic and determining the forwarding table for

	multicast traffic. SPBM employs IS-IS as the interior gateway protocol and implements additional Type-Length-Values (TLVs) to support additional functionality.
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 Internet Group Management Protocol (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	Layer 1 is the Physical Layer of the Open System Interconnection (OSI) model. Layer 1 interacts with the MAC sublayer of Layer 2, and performs character encoding, transmission, reception, and character decoding.
Layer 2	Layer 2 is the Data Link Layer of the OSI model. Examples of Layer 2 protocols are Ethernet and Frame Relay.
Layer 2 Virtual Services Network	The Layer 2 Virtual Services Network (L2 VSN) feature provides IP connectivity over SPBM for VLANs. Backbone Edge Bridges (BEBs) handle Layer 2 virtualization. At the BEBs you map the end-user VLAN to a Service Instance Identifier (I-SID). BEBs that have the same I-SID configured can participate in the same Layer 2 Virtual Services Network (VSN).
Layer 3	Layer 3 is the Network Layer of the OSI model. An example of a Layer 3 protocol is Internet Protocol (IP).
Layer 3 Virtual Services Network	The Layer 3 Virtual Services Network (L3 VSN) feature provides IP connectivity over SPBM for VRFs. Backbone Edge Bridges (BEBs) handle Layer 3 virtualized. At the BEBs through local provisioning, you map the end-user IP enabled VLAN or VLANs to a Virtualized Routing and Forwarding (VRF) instance. Then you map the VRF to a Service Instance Identifier (I-SID). VRFs that have the same I-SID configured can participate in the same Layer 3 Virtual Service Network (VSN).

Layer 4	The Transport Layer of the OSI model. An example of a Layer 4 protocol is Transmission Control Protocol (TCP).
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 increases the redundancy for higher availability.
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.
load balancing	The practice of splitting communication into two (or more) routes or servers.
MAC-in-MAC encapsulation	MAC-in-MAC encapsulation defines a BMAC-DA and BMAC-SA to identify the backbone source and destination addresses. The originating node creates a MAC header that the device uses for delivery from end to end. As the MAC header stays the same across the network, there is no need to swap a label or do a route lookup at each node, allowing the frame to follow the most efficient forwarding path end to end.
management information base (MIB)	The 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 autonomous system that has multiple connections to one or more autonomous systems 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 specific layer and that consists of protocol-control information of the specific layer and possibly user data of that layer.
Protocol Independent Multicast, Source Specific (PIM-SSM)	PIM-SSM is a multicast routing protocol for IP networks. PIM-SSM uses only shortest-path trees to provide multicast services based on subscription to a particular (source, group) channel. PIM-SSM eliminates the need for starting with a shared tree by immediately joining a source through the shortest path tree. This method enables PIM-SSM to avoid using a rendezvous point (RP) and RP-based shared tree, which can be a potential bottleneck.
Protocol Independent Multicast, Sparse Mode (PIM-SM)	PIM-SM is a multicast routing protocol for IP networks. PIM-SM provides multicast routing for multicast groups that can span wide-area and inter-domain networks, where receivers are not densely populated. PIM-SM sends multicast traffic only to those routers that belong to a specific multicast group and that choose to receive the traffic. 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.
Provider Backbone Bridge (PBB)	To forward customer traffic across the service-provider backbone, SPBM uses IEEE 802.1ah Provider Backbone Bridging (PBB) MAC-in-MAC encapsulation, which hides the customer MAC (C-MAC) addresses in a backbone MAC (B-MAC) address pair. MAC-in-MAC encapsulation defines a BMAC-DA and BMAC-SA to identify the backbone source and destination addresses.
quality of service (QoS)	QoS features reserve resources in a congested network, allowing you to configure a higher priority for certain devices. For example, you can configure a higher priority for IP deskphones, which need a fixed bit rate

	and split the remaining bandwidth between data connections if calls in the network are more 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, which 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)	Secure Copy securely transfers files between the switch and a remote station.
Secure Shell (SSH)	SSH uses encryption to provide security for remote logons and data transfer over the Internet.
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. SPBM uses this I-SID to identify and transmit any virtualized traffic in an encapsulated SPBM frame. SPBM uses I-SIDs 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 VSNs, you associate the I-SID with a customer VLAN, which is then virtualized across the backbone. With Layer 3 VSNs, you associate the I-SID 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 Bridging (SPB)	Shortest Path Bridging is a control Link State Protocol that provides a loop-free Ethernet topology. There are two versions of Shortest Path Bridge: Shortest Path Bridging VLAN and Shortest Path Bridging MAC. Shortest Path Bridging VLAN uses the Q-in-Q frame format and encapsulates the source bridge ID into the VLAN header. Shortest Path Bridging MAC uses the 802.1 ah MAC-in-MAC frame format and encapsulates the source bridge identifier into the B-MAC header.
Shortest Path Bridging MAC (SPBM)	Shortest Path Bridging MAC (SPBM) uses the Intermediate-System-to-Intermediate-System (IS-IS) link-state routing protocol to provide a loop-free Ethernet topology that creates a shortest-path topology from every node to every other node in the network based on node MAC addresses. SPBM uses the 802.1ah MAC-in-MAC frame format and encapsulates the source bridge identifier into the B-MAC header. SPBM eliminates the need for multiple overlay protocols in the core of the network by reducing the core to a single Ethernet-based link-state protocol, which can provide virtualization services, both layer 2 and layer 3, using a pure Ethernet technology base.
shortest path tree (SPT)	Creates a direct route between the receiver and the source for group members in a Protocol Independent Multicast-Sparse 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 10 micrometers), with a cladding that is 10 times the core diameter. These fibers have a potential bandwidth of 50 to 100 gigahertz (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 gigabits per second (Gbps).
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 (9012RC)	The SF cooling module is a hot swappable fan tray used to cool the Switch Fabric (SF) modules in the Virtual Services Platform 9012.
Switch Fabric (SF) module	The Switch Fabric module connects to all I/O 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 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.
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.

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.