Extreme SLX-OS
IP Multicast Configuration Guide, 20.7.3

Supporting ExtremeRouting and ExtremeSwitching
SLX 9740, SLX 9640, SLX 9540, SLX 9250, SLX 9150,
Extreme 8820, Extreme 8720, and Extreme 8520

9039521-00 Rev AA
December 2025



Copyright © 2025 Extreme Networks, Inc. All rights reserved.

Legal Notice

Extreme Networks, Inc. reserves the right to make changes in specifications and other information contained in this document and its website without prior notice. The reader should in all cases consult representatives of Extreme Networks to determine whether any such changes have been made.

The hardware, firmware, software or any specifications described or referred to in this document are subject to change without notice.

Trademarks

Extreme Networks and the Extreme Networks logo are trademarks or registered trademarks of Extreme Networks, Inc. in the United States and/or other countries.

All other names (including any product names) mentioned in this document are the property of their respective owners and may be trademarks or registered trademarks of their respective companies/owners.

For additional information on Extreme Networks trademarks, see: https://www.extremenetworks.com/about-extreme-networks/company/legal/trademarks

Open Source Declarations

Some software files have been licensed under certain open source or third-party licenses. End-user license agreements and open source declarations can be found at: https://www.extremenetworks.com/support/policies/open-source-declaration/

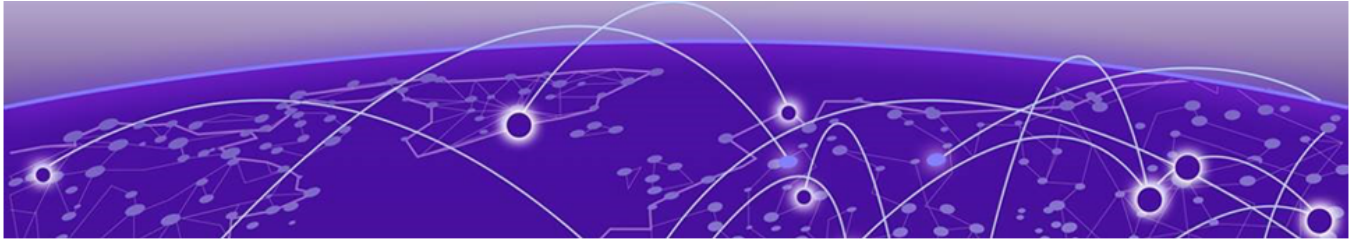


Table of Contents

Preface **6**
- Text Conventions 6
- Documentation and Training 7
  - Open Source Declarations 8
  - Training 8
- Help and Support 8
  - Subscribe to Product Announcements 9
- Send Feedback 9

About This Document **10**
- What's New in this Document 10
- Supported Hardware 10

IP Multicast **12**
- IP Multicast Overview 12
- IP Multicast Snooping 12
- IP Multicast Queries and Responses 13

IPv4 Multicast Routing **14**
- Internet Group Management Protocol 14
  - IGMP Versions 15
  - Change the IGMP Version 16
  - IGMP EVPN Routes 16
  - Disable the IGMP Router Alert Option 20
  - Configure IGMP SSM Mapping 20
  - Configure IGMPv2 SSM Mapping 21
- IPv4 Protocol Independent Multicast 22
  - IPv4 PIM Over Secondary Addresses 23
  - IPv4 PIM-SM 23
  - Enable IPv4 PIM Globally 30
  - Configure Options on an IPv4 PIM Router 30
  - Configure Options on an IPv4 PIM Interface 31
  - Configure Options for IPv4 PIM Multi-VRF 31
  - Display IPv4 PIM Information 33
  - Explicitly Select the IPv4 PIM Rendezvous Point 34
  - IPv4 PIM Anycast RP 35
  - Configure the IPV4 PIM Anycast RP 36
- IPv4 Multicast ECMP Dynamic Rebalance 37
  - Hash-based load distribution 38
  - Path failure behavior 38
  - New path behavior 39
  - Dynamic rebalancing 39
  - Considerations 39

Enable ECMP Dynamic Rebalance....39
Multicast Traceroute Diagnostics....39
Primary components of an mtrace implementation....40
Configure Mtrace....41
Layer 2 Multicast Over MCT....41
IGMP Query Packet Processing....42
IGMP Membership Reports....42
Duplicate IGMP Query Packets on CCEP....42
IGMP Leave....43
Mrouter Synchronization....43
Device Support....43
Layer 2 Multicast Traffic Forwarding....43
Optimal Traffic Forwarding....43
Layer 2 Multicast Data Encapsulation....44
Layer 3 Multicast over MCT....44
Device Support....44

IPv4 Multicast Traffic Reduction....45

IGMP Traffic Snooping....45
Multicast Routing and IGMP Snooping....45
Enable IGMP Snooping on a VLAN....46
Configure IGMP Snooping on a VLAN....46
Configure IGMP Snooping on a Bridge Domain....47
Monitor IGMP Snooping....48
IGMP Snooping and Unknown Multicast Traffic....50
Disable the IGMP Router Alert Option....51
IPv4 PIM-SM Traffic Snooping....51
Overview....51
Assumptions and Dependencies....51
PIM-SM Snooping on a Bridge Domain....52
PIM Snooping in an SSM Range....52
IPv4 PIM-SM Snooping Example....52
Enable IPv4 PIM Snooping on a VLAN....53
Enable IPv4 PIM Snooping on a Bridge Domain....53
PIM Multicast Router Presence Detection....54

IP Multicast Fabric....55

IP Multicast Fabric Overview....55
Supported Platforms....55
Reference Network....56
Ingress Replication....56
Overview of Multicast Solution....57
Multicast Distribution for BUM Traffic....58
Ideal Multicast Distribution....58
Unicast VXLAN Tunnel Establishment....58
VTEP Auto-discovery....58
Unicast VXLAN Tunnel Establishment....59
Multicast Distribution Tree....60
Per VNI MDT....61
Default MDT....62

Configure Optimization Replication....64
MDT Scale....65
Configure Multicast IP Fabric with L3 VNI....65
Logical VTEPs and MCT (Multi-homing)....67
MDT Signaling for LVTEP/MCT Cluster....67
Primary MDT Node Selection....68
LVTEP/MCT Cluster Link Failure....69
Route Sync Between LVTEP/MCT Cluster....69
Datapath....70
Bud Node Topology....70

IPv6 Multicast Routing....72
IPv6 Multicast over MCT....72
MLD query packet processing....72
MLD membership reports....72
Duplicate MLD query packets on CCEP....73
MLD listener done....73
Mrouter synchronization....73
Scalability....73
Device support....74
IPv6 Protocol Independent Multicast....74
IPv6 PIM-SM....74
IPv6 PIM-SSM....75
Enable IPv6 PIM Globally....75
Configure Options on an IPv6 PIM Router....75
Configure Options on an IPv6 PIM Interface....77
Display IPv6 PIM Information....77
Clear and Troubleshoot IPv6 PIM Information....80

IPv6 Multicast VLAN Traffic Reduction....82
IPv6 MLD Traffic Snooping....82
Overview....82
Scalability....82
Device support....83
IPv6 MLD Queries and Messages....83
IPv6 MLD Traffic and Data Forwarding....83
Enable IPv6 MLD Snooping....84
Configure IPv6 MLD Snooping....84
Display IPv6 MLD Snooping Information....85
Clear and Troubleshoot IPv6 MLD Information....86
IPv6 PIM Traffic Snooping....87
Enable IPv6 PIM Snooping....87
Display IPv6 PIM Snooping Information....87
Clear IPv6 PIM Snooping Information....88

Preface

Read the following topics to learn about:

- The meanings of text formats used in this document.
- Where you can find additional information and help.
- How to reach us with questions and comments.

Text Conventions

Unless otherwise noted, information in this document applies to all supported environments for the products in question. Exceptions, like command keywords associated with a specific software version, are identified in the text.

When a feature, function, or operation pertains to a specific hardware product, the product name is used. When features, functions, and operations are the same across an entire product family, such as Extreme Networks switches, the product is referred to as *the switch*.

Table 1: Notes and warnings

Icon	Notice type	Alerts you to...
	Tip	Helpful tips and notices for using the product
	Note	Useful information or instructions
	Important	Important features or instructions
	Caution	Risk of personal injury, system damage, or loss of data
	Warning	Risk of severe personal injury

Table 2: Text

Convention	Description
`screen displays`	This typeface indicates command syntax, or represents information as it is displayed on the screen.
The words *enter* and *type*	When you see the word *enter* in this guide, you must type something, and then press the Return or Enter key. Do not press the Return or Enter key when an instruction simply says *type*.
Key names	Key names are written in boldface, for example **Ctrl** or **Esc**. If you must press two or more keys simultaneously, the key names are linked with a plus sign (+). Example: Press **Ctrl**+**Alt**+**Del**
Words in italicized type	Italics emphasize a point or denote new terms at the place where they are defined in the text. Italics are also used when referring to publication titles.
NEW!	New information. In a PDF, this is searchable text.

Table 3: Command syntax

Convention	Description
bold text	Bold text indicates command names, keywords, and command options.
italic text	Italic text indicates variable content.
`[ ]`	Syntax components displayed within square brackets are optional. Default responses to system prompts are enclosed in square brackets.
`{ x | y | z }`	A choice of required parameters is enclosed in curly brackets separated by vertical bars. You must select one of the options.
`x | y`	A vertical bar separates mutually exclusive elements.
`< >`	Nonprinting characters, such as passwords, are enclosed in angle brackets.
...	Repeat the previous element, for example, *member*`[`*member*`...]`.
\	In command examples, the backslash indicates a “soft” line break. When a backslash separates two lines of a command input, enter the entire command at the prompt without the backslash.

Documentation and Training

Find Extreme Networks product information at the following locations:

Current Product Documentation
Release Notes
Hardware and Software Compatibility for Extreme Networks products
Extreme Optics Compatibility
Other Resources such as articles, white papers, and case studies

Open Source Declarations

Some software files have been licensed under certain open source licenses. Information is available on the Open Source Declaration page.

Training

Extreme Networks offers product training courses, both online and in person, as well as specialized certifications. For details, visit the Extreme Networks Training page.

Help and Support

If you require assistance, contact Extreme Networks using one of the following methods:

Extreme Portal

Search the GTAC (Global Technical Assistance Center) knowledge base; manage support cases and service contracts; download software; and obtain product licensing, training, and certifications.

The Hub

A forum for Extreme Networks customers to connect with one another, answer questions, and share ideas and feedback. This community is monitored by Extreme Networks employees, but is not intended to replace specific guidance from GTAC.

Call GTAC

For immediate support: (800) 998 2408 (toll-free in U.S. and Canada) or 1 (408) 579 2800. For the support phone number in your country, visit www.extremenetworks.com/support/contact.

Before contacting Extreme Networks for technical support, have the following information ready:

- Your Extreme Networks service contract number, or serial numbers for all involved Extreme Networks products
- A description of the failure
- A description of any actions already taken to resolve the problem
- A description of your network environment (such as layout, cable type, other relevant environmental information)
- Network load at the time of trouble (if known)
- The device history (for example, if you have returned the device before, or if this is a recurring problem)
- Any related RMA (Return Material Authorization) numbers

Subscribe to Product Announcements

You can subscribe to email notifications for product and software release announcements, Field Notices, and Vulnerability Notices.

1. Go to The Hub.
2. In the list of categories, expand the **Product Announcements** list.
3. Select a product for which you would like to receive notifications.
4. Select **Subscribe**.
5. To select additional products, return to the **Product Announcements** list and repeat steps 3 and 4.

You can modify your product selections or unsubscribe at any time.

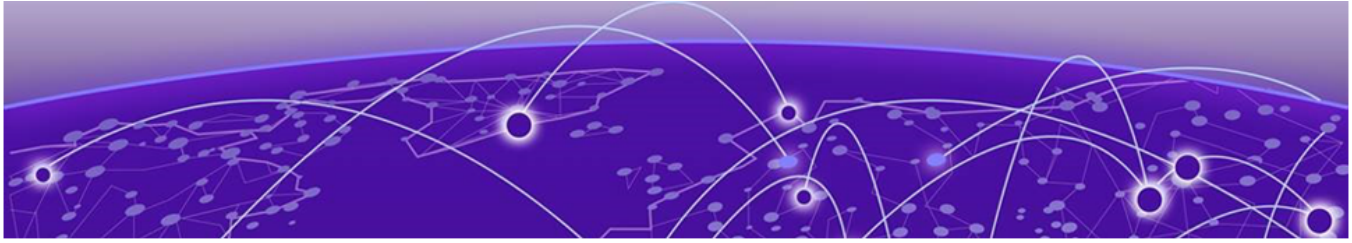
Send Feedback

The User Enablement team at Extreme Networks has made every effort to ensure that this document is accurate, complete, and easy to use. We strive to improve our documentation to help you in your work, so we want to hear from you. We welcome all feedback, but we especially want to know about:

- Content errors, or confusing or conflicting information.
- Improvements that would help you find relevant information.
- Broken links or usability issues.

To send feedback, email us at Product-Documentation@extremenetworks.com.

Provide as much detail as possible including the publication title, topic heading, and page number (if applicable), along with your comments and suggestions for improvement.



About This Document

What's New in this Document on page 10
Supported Hardware on page 10

What's New in this Document

This document is released with the SLX-OS 20.7.3 software release. No changes were made to this document for this version.

For additional information, refer to the *Extreme SLX-OS Release Notes* for this version.

Supported Hardware

SLX-OS 20.7.3 supports the following hardware platforms.

- Extreme 8820
- Extreme 8720
- Extreme 8520
- ExtremeSwitching SLX 9540
- ExtremeSwitching SLX 9250
- ExtremeSwitching SLX 9150
- ExtremeRouting SLX 9740
- ExtremeRouting SLX 9640

Note

All configurations and software features that are applicable to SLX 9150 and SLX 9250 devices are also applicable for the Extreme 8520 and Extreme 8720 devices respectively.

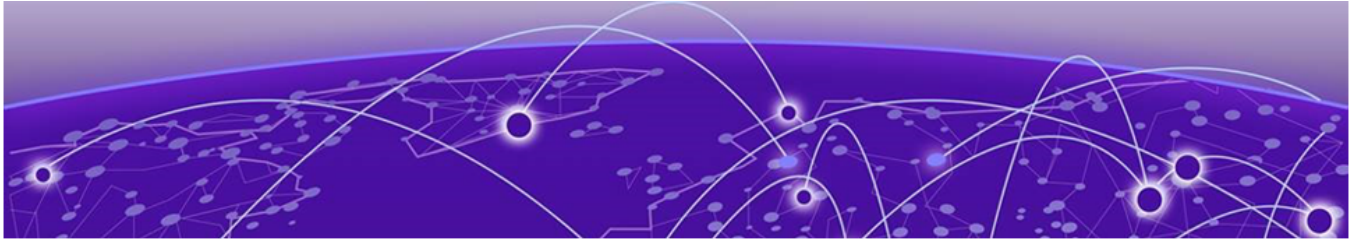
All configurations and software features that are applicable to SLX 9740 devices are also applicable for the Extreme 8820 devices.

The "Measured Boot with Remote Attestation" feature is only applicable to the Extreme 8520, Extreme 8720, and Extreme 8820 devices. It is not supported on the SLX 9150 and SLX 9250 devices.

Note

Although many software and hardware configurations are tested and supported for this release, documenting all possible configurations and scenarios is beyond this document's scope.

For information about other releases, see the documentation for those releases.



IP Multicast

IP Multicast Overview on page 12
IP Multicast Snooping on page 12
IP Multicast Queries and Responses on page 13

IP Multicast Overview

To receive and transmit multicast data, stations (clients) use multicast protocols to access a group or a channel over different networks. Distribution of stock quotes, video transmissions such as news services and remote classrooms, and video conferencing are all examples of applications that use multicast routing.

Extreme Networks devices support the following multicast protocols: Protocol-Independent Multicast (PIM) and Internet Group Management Protocol (IGMP).

With the reverse path lookup and pruning features of PIM, source-specific multicast delivery trees reach all group members. Each source and destination host group has its own multicast tree.

IPv4 hosts use IGMP to report their multicast group memberships to immediate neighboring multicast routers.

IPv6 hosts and routers use MLD to exchange group information.

For more information, see IPv4 Multicast Routing on page 14.

IP Multicast Snooping

A Layer 2 switch forwards all multicast control packets and data received on all the member ports of a VLAN interface. This simple approach is not bandwidth efficient, because only a subset of member ports may be connected to devices that want to receive these multicast packets. In a worst-case scenario, the data is forwarded to all port members of a VLAN even if only one VLAN member wants to receive the data. Such a scenario can lead to loss of throughput for switches that receive a high rate of multicast data traffic.

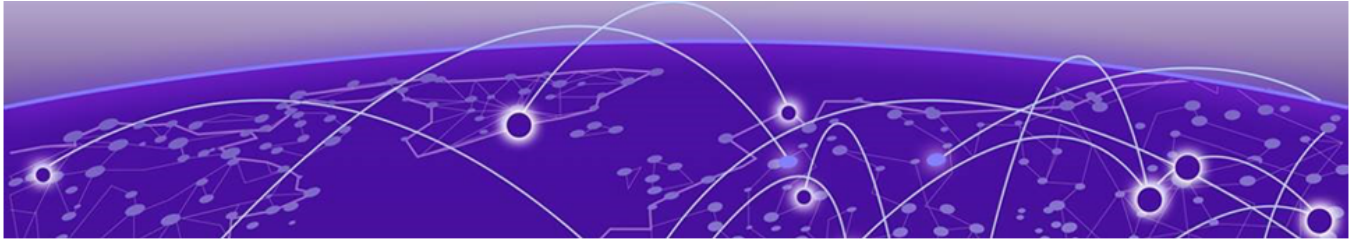
IGMP helps save bandwidth and throughput by forwarding traffic only to interested receivers. IGMP snooping provides the specification for forwarding IPv4 data traffic. For more information, see IGMP Traffic Snooping on page 45.

With PIM snooping, switches learn which multicast router ports need to receive multicast traffic. For more information, see IPv4 PIM-SM Traffic Snooping on page 51.

IP Multicast Queries and Responses

Multicast routers use IGMP to learn which groups have interested listeners on their attached physical networks. In a subnet, one multicast router is elected as an IGMP querier.

The querier sends queries to hosts. Hosts that are multicast listeners respond to the queries with requests to join or leave a multicast group. For more information, see Internet Group Management Protocol on page 14.



IPv4 Multicast Routing

Internet Group Management Protocol on page 14
IPv4 Protocol Independent Multicast on page 22
IPv4 Multicast ECMP Dynamic Rebalance on page 37
Multicast Traceroute Diagnostics on page 39
Layer 2 Multicast Over MCT on page 41
Layer 2 Multicast Traffic Forwarding on page 43
Layer 2 Multicast Data Encapsulation on page 44
Layer 3 Multicast over MCT on page 44

Internet Group Management Protocol

The Internet Group Management Protocol (IGMP) allows an IPv4 system to communicate IP multicast group membership information to its neighboring routers. The routers, in turn, limit the multicast of IP packets with multicast destination addresses to only those interfaces on the router that are identified as IP multicast group members.

In IGMPv2, when a router sends a query to the interfaces, the clients on the interfaces respond with a membership report of multicast groups to the router. The router can then send traffic to these groups, regardless of the traffic source. When an interface no longer needs to receive traffic from a group, it sends a leave message to the router, which in turn sends a group-specific query to that interface to see if any other clients on the same interface are still active.

In contrast, IGMPv3 provides selective filtering of traffic based on the traffic source. A router running IGMPv3 sends queries to every multicast-enabled interface at the specified interval. These general queries determine if any interface wants to receive traffic from the router.

There are different types of query messages.

- **General Query**: Sent by a multicast router to learn the complete multicast reception state of the neighboring interfaces. In this query, both the Group Address field and the Number of Sources (N) field are zero.
- **Group-Specific Query**: Sent by a multicast router to learn the reception state, with respect to one multicast address, of the neighboring interfaces. In this query, the Group Address field contains the multicast address of interest, and the Number of Sources (N) field contains zero.

- Group-and-Source-Specific Query: Sent by a multicast router to learn whether neighboring interfaces want to receive packets sent to a specified multicast address, from any of a specified list of sources. In this query, the Group Address field contains the multicast address of interest, and the Source Address fields contain the source addresses of interest.

Interfaces respond to these queries by sending a membership report that contains one or more of the following records that are associated with a specific group.

- The current-state record indicates from which sources the interface wants to receive and not receive traffic. The record contains the source address of the interfaces and whether traffic will be received or included (IS_IN) or not received or excluded (IS_EX) from that source.
- The filter-mode-change record indicates that if the interface changes its current state from IS_IN to IS_EX, a TO_EX record is included in the membership report. Likewise, if the interface changes its current status from IS_EX to IS_IN, a TO_IN record appears in the membership report.
- The IGMPv2 Leave report is equivalent to a TO_IN (empty) record in IGMPv3. This record indicates that no traffic from this group will be received regardless of the source.
- The IGMPv2 group report is equivalent to an IS_EX (empty) record in IGMPv3. This record indicates that all traffic from this group will be received regardless of the source.
- The source-list-change record indicates that if the interface wants to add or remove traffic sources from its membership report, the membership report can have an ALLOW record, which contains a list of new sources from which the interface wishes to receive traffic. It can also contain a BLOCK record, which lists current traffic sources from which the interface wants to stop receiving traffic.

In response to membership reports from the interfaces, the router sends a Group-Specific Query or a Group-and-Source Specific Query to the multicast interfaces. For example, a router receives a membership report with a source-list-change record to block old sources from an interface. The router sends Group-and-Source Specific Queries to the source-group pair (S,G) identified in the record. If none of the interfaces is interested in the (S,G), it is removed from the (S,G) list for that interface on the router.

Each IGMPv3-enabled router maintains a record of the state of each group and each physical port in a virtual routing interface. This record contains the group, group-timer, filter mode, and source records information for the group or interface. Source records contain information about the source address of the packet and source timer. If the source timer expires when the state of the group or interface is in include mode, the record is removed.

IGMP Versions

Default IGMP Version

IGMPv2 is enabled by default when snooping or multicast routing are enabled on the system.

You can specify which version of IGMP you want to run on a device on a per-VLAN basis. You can change the IGMP version for router ports, but not for VE interfaces. If you do not specify an IGMP version, IGMPv2 is used.

Compatibility Between IGMPv1 and IGMPv2

Different multicast groups, interfaces, and routers can run their own versions of IGMP. The version of IGMP is reflected in the membership reports that the hosts send to the router. Routers and interfaces must be configured to recognize the version of IGMP you want them to process.

Interfaces can recognize a query or report that has a different version. For example, an interface running IGMPv2 can recognize IGMPv3 packets, but cannot process them. When the router sends out IGMP queries over an IGMPv2 interface, the equal or lower version of reports is supported. A higher version of reports is not supported.

Reports sent by interfaces to routers that contain different versions of IGMP do not trigger warning messages. The version of IGMP can be specified per interface (physical port or virtual routing interface) and per physical port in a virtual routing interface.

The IGMP version on a Layer 3 physical interface or under a VLAN of the virtual routing interface supersedes the version on a physical or virtual routing interface.

Change the IGMP Version

You can change the IGMP version for a interface or VLAN.

1. Enter global configuration mode.
```
device# configure terminal
```
2. Enter interface configuration mode.
```
device(config)# interface ethernet 0/5
```
3. Specify the IGMP version. This example changes the version to 3.
```
device(config-if-0/5)# ip igmp version 3
```

IGMP EVPN Routes

In data center (DC) applications, Ethernet VPN (EVPN) is used to standardize inter-PoD communication for both intra-DC and inter-DC applications.

A subnet can span multiple points of delivery (PoD) and DCs. EVPN provides a robust multitenant solution with extensive multihoming capabilities to stretch a subnet (VLAN) across multiple PoDs and DCs. Many hosts or VMs (up to several hundred) can be attached to a subnet that is stretched across several PoDs and DCs.

Hosts or VMs express their interest in multicast groups on a subnet or VLAN by sending IGMP membership reports (Joins) for their interested multicast groups. Also, an IGMP router (for example, IGMPv1) periodically sends membership queries to determine whether any hosts on that subnet still want to receive multicast traffic for that group.

The ARP (Address Resolution Protocol) and NDP (Network Discovery Protocol) suppression mechanism in EVPN reduces the flooding of ARP messages over EVPN.

Reducing the flood of IGMP messages (both Queries and Reports) over EVPN is achieved through IGMP Join Sync and IGMP Leave Sync routes as specified in the IETF draft IGMP and MLD Proxy for EVPN.

Table 4: EVPN routes

EVPN route type	Route name	Purpose
7	IGMP Join Sync Route	To exchange (S,G)/(*,G) learned on BGP EVPN peers.
8	IGMP Leave Sync Route	To exchange group leave between BGP EVPN peers.

IGMP Join Sync Route Packets and Flags

A multicast group is assigned a multicast address (that is, an IP address in the 224.0.0.0/4 class D address space). Hosts registered to receive the stream join the group by sending an IGMP report to the upstream multicast router. The router then adds that group to the list of multicast groups that should be forwarded onto the local subnet.

The router does not maintain state about which hosts on the subnet are to receive traffic for the group. Instead, the router continues to send traffic to the subnet until either a timeout expires or there are no more hosts in that group on the subnet.

Table 5: IGMP Join Sync route packets

Packets	Description
RD (8 Octets)	Route Distinguisher
Ethernet Segment Identifier (10 Octets)	0 if (S,G) / (*,G) is learned on Cluster Edge Port (CEP)
Ethernet Tag ID (4 octets)	0 or optionally set to the VLAN ID over which a multicast route is learned
Multicast Source Length (1 octet)	32 for IPv4 address and 0 for wildcard address (*)
Multicast Source Address (variable)	IP address of multicast source
Multicast Group Length (1 octet)	32 for IPv4 address
Multicast Group Address (variable)	IP address of multicast group
Originator Router Length (1 octet)	32 for IPv4 address
Originator Router Address (variable)	The IP address of the originating router
Flags (1 octets) (optional)	The flag fields are defined in the *Flags and values* table.

Table 6: Parameters

Parameter	Value
Querier Config Sync	Multicast Group is set to 224.0.0.2.
Mrouter Sync	Multicast Group is set to 224.0.0.1.

Table 6: Parameters (continued)

Parameter	Value
Originator router length	IPv4 addresses: 32
Originator router address	IP address of the originating router

Table 7: Flags and values

0	1	2	3	4	5	6	7
Reserved	Querier config sync	Mrouter sync	PIM snooping	IE	V3	V2	V1
	Bit 1: QuerierConfigSync	Bit 2: MrouterSync	Bit 3: PIM snooping	Bit 4: (S, G) information carries the route-type of Include Group (bit value 0) or Exclude Group (bit value 1). The Exclude Group is ignored if bit 5 is set for IGMP v2, IGMP v1.	Bit 5: support for IGMP v3	Bit 6: support for IGMP v2	Bit 7: support IGMP v1
Bits 1, 2, and 3 are proprietary fields.							

IGMP Leave Sync Route Packets and Flags

A host sends an IGMP Leave message to stop receiving multicast traffic. After receiving the Leave message, the router sends a query to the local subnet to determine whether any group members remain, sending the message to all hosts on the subnet at the multicast All-Hosts address (224.0.0.1). If a host responds, the router continues to send to the group. If no host responds, the router removes the multicast group from its forwarding list and stops sending to the group.

Note
Leave Group synchronization and maximum response times are used during Leave Group procedures.

Table 8: IGMP Leave Sync route packets

Packets	Description
RD (8 Octets)	
Ethernet Segment Identifier (10 octets)	0 if (S,G) / (*,G) is learned on CEP
Ethernet Tag ID (4 octets)	0 or optionally, set to the VLAN ID over which the multicast route is learned
Multicast Source Length (1 octet)	32 for IPv4 address and 0 for wildcard address (*)

Table 8: IGMP Leave Sync route packets (continued)

Packets	Description
Multicast Source Address (variable)	IP address of multicast source
Multicast Group Length (1 octet)	32 for IPv4 address
Multicast Group Address (variable)	IP address of multicast group
Originator Router Length (1 octet)	32 for IPv4 address
Originator Router Address (variable)	The IP address of the originating router
Leave Group Synchronization # (4 octets)	Leave group synchronization
Maximum Response Time (1 octet)	The maximum response time
Flags (1 octets) (optional)	The flag fields are defined in the following table

Table 9: Flags

0	1	2	3	4	5	6	7
Reserved	Reserved	Reserved	Reserved	IE	V3	V2	V1
				Bit 4: (S, G) information carries the route-type of Include Group (bit value 0) or Exclude Group (bit value 1). The Exclude Group is ignored if bit 5 is set for IGMP v2, IGMP v1.	Bit 5: support for IGMP v3	Bit 6: support for IGMP v2	Bit 7: support for IGMP v1
Bits 0, 1, 2, and 3 are reserved and used.							

IGMP Join Group and Leave Group Process

IGMP routes are originated by IGMP module and are sent to L2RIB (MCT module) through RIB-LIB and transported to MCT Peers. IGMP Routes received from the remote peers are added to L2 RIB table in MCT and forwarded to IGMP through RIB-LIB.

Currently IGMP EVPN routes sent to MCT L2RIB are not exchanged with BGP. In future, this solution will be extended to support Multicast over BGP-EVPN, where MCT will exchange IGMP EVPN Routes with BGP and MCT Peers.

EVI Route Target Extended Community

The EVI RT is an EVPN extended community of Type 6 and a subtype yet to be defined by IANA. The EVI RT extended community is NOT supported for IGMP routes.

Instead, the RT extended community with Type 0x00 and Subtype 0x02 is supported. This extended community carries the RT associated with the EVI (VLAN or bridge domain), so that the receiving PE can identify the EVI properly.

ES-Import Route Target Extended Community

ES-Import (ES-I) Route Target (RT) is an EVPN extended community of Type 0x06 and Subtype 0x02. The 6-byte value calculated is based on the ES-Import.

The IGMP Join and Leave Sync routes carry the ES-Import RT for the ES on which the IGMP membership report was received. The report goes only to the PEs attached to that ES.

Encapsulation Support

Only the VXLAN tunnel encapsulation type is supported. MPLS is not supported for ICLs in MCT scenarios.

Disable the IGMP Router Alert Option

By default, IGMP snooping checks for the presence of the router alert option in the IP packet header of an IGMP message. Packets that do not include this option are dropped. When you disable the router alert, you also disable the snooping check for the presence of the router alert option.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Disable the router alert option.

```
device(config)# ip igmp router-alert-check-disable
```

Configure IGMP SSM Mapping

With Source-Specific Multicast (SSM) mapping, a host receives multicast traffic directly from identified sources.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter VLAN configuration mode.

```
device (config)# vlan 101
```

3. Configure the IGMP query interval for the VLAN.

```
device (config-vlan-101)# ip igmp snooping query-interval 101
```

4. Configure the prefix list, which contains the IP addresses to map.

```
device (config)# ip igmp ssm-map ?
Possible completions:
               <Word:1-32>   IP prefix-list name
               enable        Enables IGMPv2 SSM Mapping
```

5. Configure SSM mapping, which associates an IGMPv1 or IGMPv2 report packet with the configured source address, which is 203.0.0.1 in this example.

```
device (config-vlan-101)# ip igmp ssm-map enab
device (config-vlan-101)# ip igmp ssm-map prefix-list1 203.0.0.1
```

```
device (config-vlan-101)# show ip igmp ssm-map
Fri Jul  21 11:30:06.878 UTC-07:00
+-----------------------------------+-------------------+
|         PrefixList Name           |   Source Address  |
+-----------------------------------+-------------------+
  prefix-list1                        203.0.0.1
```

6. Specify the SSM group range, which is 238.0.0.0/8 in this example.

```
device (config-router-pim-vrf-default-vrf)# router pim
device (config-router-pim-vrf-default-vrf)# ssm-enable
device (config-router-pim-vrf-default-vrf)# ssm-enable range prefix-list1
device# show ip pim settings
Fri Jul  21 11:31:45.333 UTC-07:00
vrf : default-vrf
  Maximum mcache              : 32768     Current Count               : 0
  Hello interval              : 30        Neighbor timeout            : 105
  Join/Prune interval         : 60        Inactivity interval         : 180
  Hardware drop enabled       : 1         Prune wait interval         : 3
  Register Suppress Time      : 60        Register Probe Time         : 10
  Register Stop Delay         : 0         Register Suppress interval  : 0
  SSM Enabled                 : Yes       SPT Threshold               : 1
  SSM Group Range             : 232.0.0.0/8
  SSM Range Prefix_name       : prefix-list1
  Route Precedence            : uc-non-default uc-default none
```

This example sends an IGMPv2 report for group 238.0.0.1.

```
device # show ip igmp group
Total Number of Groups: 1
IGMP Connected Group Membership
Group Address    Interface            Uptime      Expires    Last Reporter     Version
238.0.0.1        vlan101              00:04:40    00:03:31   101.0.0.10        3
  Member Ports:  eth0/2

device # show ip igmp group detail
Group : 238.0.0.1
        Interface        vlan101
        Uptime           00:04:53
        Expires:         00:03:18
        Last Reporter:   101.0.0.10
        Member Ports:    eth0/2
        Last Reporter Mode: 3
        Interface : eth0/2
                INCL_SRC_LIST: 203.0.0.1
                EXCL_SRC_LIST: Nil

device # show ip pim mc
Fri Jul  21 11:58:17.278 UTC-07:00
Total entries in mcache: 1
1    (203.0.0.1, 238.0.0.1) in Eth 0/16, Uptime 00:05:23
     SSM=1, RPT=0 SPT=1 Reg=0 RegSupp=0 RegProbe=0 JDUp=1 LSrc=0 LRcv=1
     upstream neighbor=91.0.0.2
     AgeSltMsk: 0 KAT timer: Expired
     num_oifs = 1
           Ve101(00:05:23/0) Flags: MI
     Flags (0x080684d4)
         ssm=1 needRte=0
```

Configure IGMPv2 SSM Mapping

Source-Specific Multicast (SSM) requires all IGMP hosts to send IGMPv3 reports, which creates a compatibility issue with IGMPv2 hosts. In particular, reports from an IGMPv2

host contain a Group Multicast Address but do not contain source addresses. The IGMPv3 reports contain both the Group Multicast Address and one or more source addresses. Use the **ip igmp ssm-map** command and a properly configured prefix list to convert IGMPv2 reports into IGMPv3 reports.

The prefix list filters for the group multicast address. The prefix list is then associated with one or more source addresses. When the **ip igmp ssm-map enable** command is configured, IGMPv3 reports are sent for IGMPv2 hosts.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable SSM mapping.

```
device(config)# ip igmp ssm-map enable
```

This example configures the SSM map at the global configuration level.

```
device(config)# ip igmp ssm-map enable
device(config)# ip igmp ssm-map ssm-map-230-to-232 203.0.0.10
device(config)# ip igmp ssm-map ssm-map-233-to-234 204.0.0.10
```

This example configures the prefix list for an SSM range.

```
device(config)# ip prefix-list ssm-map-230-to-232 seq 5 permit 230.0.0.0/8
device(config)# ip prefix-list ssm-map-230-to-232 seq 10 permit 231.0.0.0/8
device(config)# ip prefix-list ssm-map-230-to-232 seq 15 permit 232.0.0.0/8

device(config)# ip prefix-list ssm-map-233-to-234 seq 5 permit 233.0.0.0/8
device(config)# ip prefix-list ssm-map-233-to-234 seq 10 permit 234.0.0.0/8
device(config)# ip prefix-list ssm-map-230-to-232 seq 15 permit 232.0.0.0/8
```

IPv4 Protocol Independent Multicast

IP multicast transmits a data stream to multiple hosts simultaneously. Protocol-Independent Multicast (PIM) is one of several protocols designed for IP multicast. PIM does not rely on a specific routing protocol to create its network topology state. Instead, PIM uses routing information supplied by other traditional routing protocols, such as Open Shortest Path First and Border Gateway Protocol.

PIM messages are sent encapsulated in an IP packet where the IP protocol is 103. Depending on the type of message, packets are sent to the PIM All-Router-Multicast address (224.0.0.13) or sent as unicast to a specific host.

With PIM, a source sends the same information to multiple receivers by using one stream of traffic. With its processing load minimized, the source needs to maintain only one session irrespective of the number of actual receivers. The load on the IP network is also minimized, because packets are sent only on links that lead to an interested receiver.

Extreme Networks supports PIM-SM (Sparse Mode) and PIM-SSM (Source Specific Multicast). PIM-SM explicitly builds unidirectional shared trees that are rooted at a rendezvous point (RP) for a group, and, optionally, creates shortest-path trees per source.

IPv4 PIM Over Secondary Addresses

Overview

Extreme Networks devices support PIM over secondary addresses in an IPv4 environment by using an IPv4 address with a secondary keyword.

When a secondary address is configured on a interface, all secondary addresses on the interface are sent out on the PIM Hello using the secondary address option.

When a receiver uses a secondary address as its source and sends a IGMP group report, the PIM Join and Prune messages are propagated up the network.

When a secondary address is configured as a Rendezvous Point (RP), the packets are processed appropriately.

Displaying the Secondary Address

In this example, the PIM neighbor on VE10 has multiple IP addresses configured on the interface.

```
device# show ip pim neighbor
Total Number of Neighbors : 1
Port         Phy_Port     Neighbor        Holdtime Age           UpTime  Priority
                                          sec      sec      Dd HH:MM:SS
Ve10         Ve10          10.10.10.17    105      10          00:26:10         1
                          +20.20.20.21
```

IPv4 PIM-SM

PIM-SM (Sparse Mode) is most effective in large networks in which few hosts receive multicast traffic. IPv4 PIM-SM devices are organized into domains. A PIM-SM domain is a contiguous set of devices that all implement PIM and are configured to operate in a common boundary.

In PIM-SM, unidirectional, shared distribution trees are rooted at a common node in the network called the rendezvous point (RP). The RP acts as the messenger between the source and the destination hosts or routers. An RP is configured statically per PIM router or by means of a bootstrap router (BSR). The RP must always be upstream from the destination hosts or routers.

Hosts and routers send Join messages to the RP for the group. To reduce the number of Join messages incoming to an RP, the local network selects one of its upstream routers as the designated router (DR). All hosts below a DR send Join messages to the DR. The DR sends only one Join message to the RP on behalf of its interested hosts.

IPv4 PIM-SM, you can create a source-based distribution tree, which is rooted at the router that is adjacent to the source. With this option, destination hosts can switch from the shared tree to the source-based tree if the latter has a shorter path between the source and the destination.

PIM-SM Device Types

Devices configured with PIM-SM interfaces can fill one or more roles.

Bootstrap router (BSR)

A router that distributes rendezvous point (RP) information to the other PIM-SM devices in the domain. Each PIM-SM domain has one active BSR. For redundancy, you can configure ports on multiple devices as candidate BSRs (C-BSRs). PIM-SM uses an election process to select one C-BSR as the BSR for the domain. The BSR with the highest BSR priority (a user-configurable parameter) is elected. If the priorities result in a tie, then the C-BSR interface with the highest IP address is elected.

The BSR must be configured as part of the Layer 3 core network.

For more information, see Bootstrap Router Protocol on page 24.

Rendezvous point (RP)

The meeting point for PIM-SM sources and receivers. A PIM-SM domain can have multiple RPs, but each PIM-SM multicast group address can have only one active RP. PIM-SM devices learn the addresses of RPs and the groups for which they are responsible from messages that the BSR sends to each PIM-SM device.

The RP must be configured as part of the Layer 3 core network.

A best practice is to configure the same ports as C-BSRs and RPs.

Designated router (DR)

Hosts and routers send Join messages to the RP for the group in which they are interested. The local network selects one of its upstream routers as the DR. All hosts below a DR send IGMP Join messages to the DR. The DR sends only one Join message to the RP on behalf of all its interested hosts. The RP receives the first few packets of the multicast stream, encapsulated in the PIM register message, from the source hosts. These messages are sent as a unicast to the RP. The RP decapsulates these packets and forwards them to the respective DRs.

DR election is based first on the router with the highest configured DR priority for an interface and next on the router with the highest IP address. You can use the **`ip pim dr-priority`** command to configure DR priority.

Bootstrap Router Protocol

In PIM-SM, every PIM router must know the rendexvous point (RP) in the network, so that it can map multicast groups to the available RP addresses. The Bootstrap Router (BSR) Protocol is a mechanism by which a PIM router learns the RP information.

An RP address is the root of a multicast group-specific distribution tree, the branches of which extend to all the nodes interested in receiving the traffic for that multicast group. All PIM routes use the same group-to-RP address mapping, allowing multicast sources to reach all receivers. Each node learns the same RP information using the following methods.

- Statically configuring the RP information on each PIM router
- Using the BSR protocol, which distributes the RP information to each PIM router

Some PIM routers act as candidate RPs (C-RPs), of which one C-RP is elected as RP for a particular group range. Some PIM routers are configured as candidate BSRs (C-BSRs), of which one is elected as the BSR. All PIM routers learn the elected BSR through bootstrap messages (BSMs). All C-RPs then report to the elected BSR, which forms the

RP-set available in the network and distributes it to all the PIM routers. All PIM routers eventually have the same RP-set information.

The BSR protocol mechanism converges in the following phases.

- BSR election
- Candidate RP advertisement and RP-set formation
- RP-set distribution

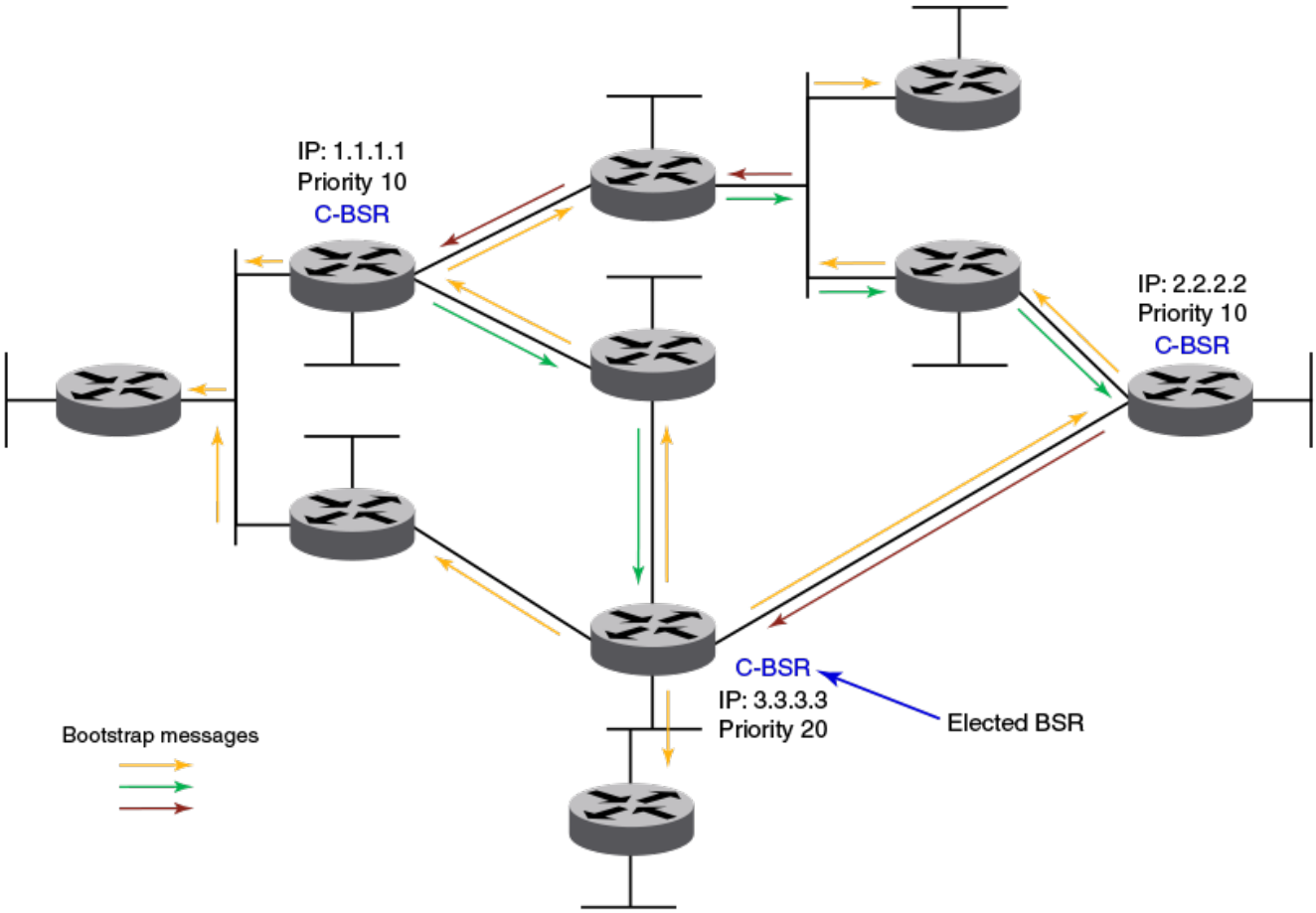


Figure 1: BSR election

Each candidate BSR periodically generates a BSM, which carries the configured BSR priority. Every PIM router in the domain floods these BSMs. Other C-BSRs that receive a BSM with higher priority suppress their own BSMs. Eventually, there is only one C-BSR with BSMs that flood periodically into the network. This lone C-BSR becomes the elected BSR and its BSM informs all routers that it is the elected BSR.

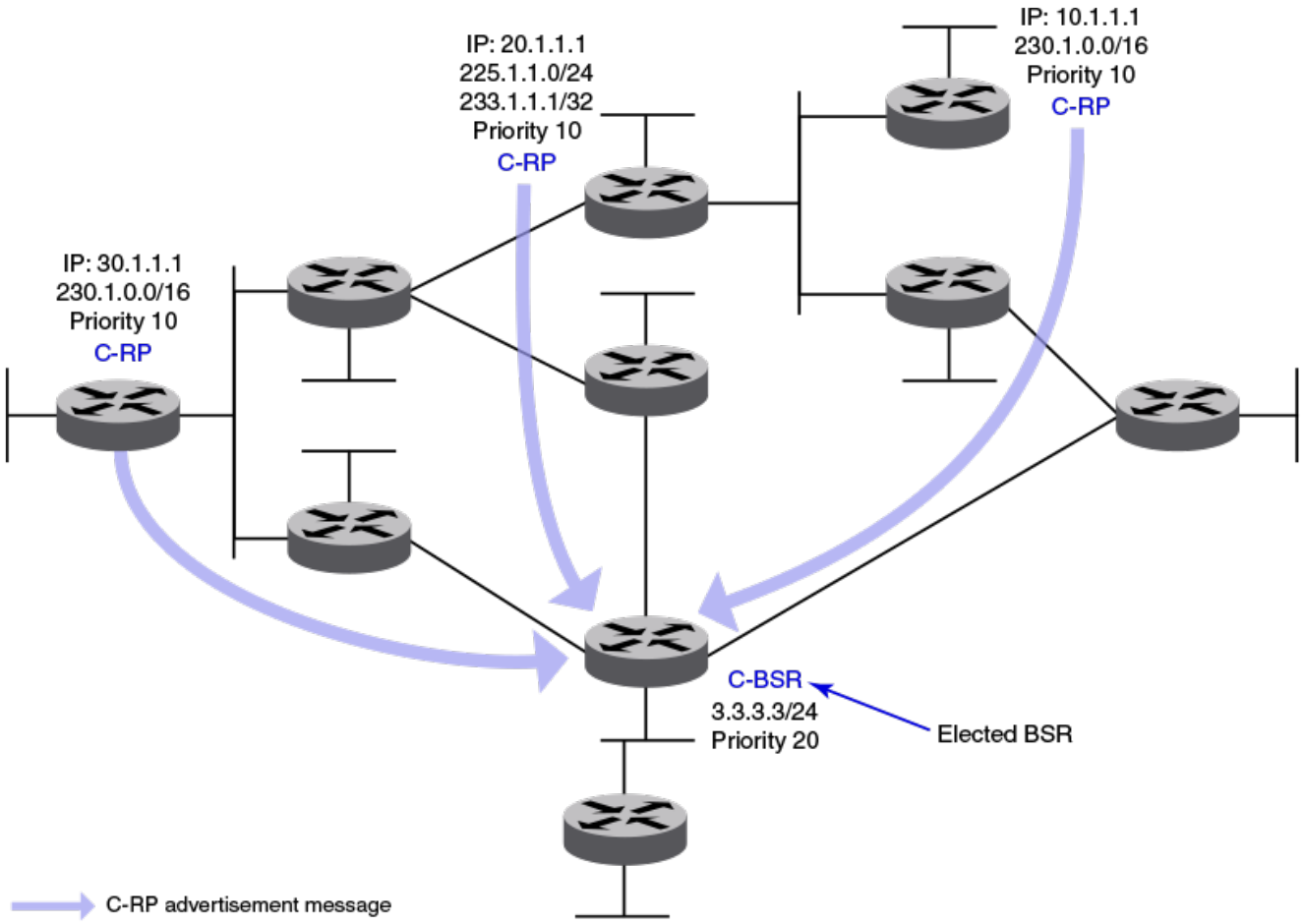


Figure 2: Candidate RP advertisement and RP-set formation

Each C-RP sends periodic candidate RP advertisement (C-RP-Adv) messages to the elected BSR. These messages contain the candidate's priority and a list of multicast group ranges for which this C-RP wants to act as RP. The messages also carry a hold time, after which the BSR discards this C-RP. In this way, the elected BSR learns about all C-RPs that are up and reachable. When the BSR starts receiving C-RP advertisements, it builds the RP-set information. This RP-set contains the list for multicast group ranges and C-RP addresses available for each of these group ranges, along with their respective priorities and hold times.

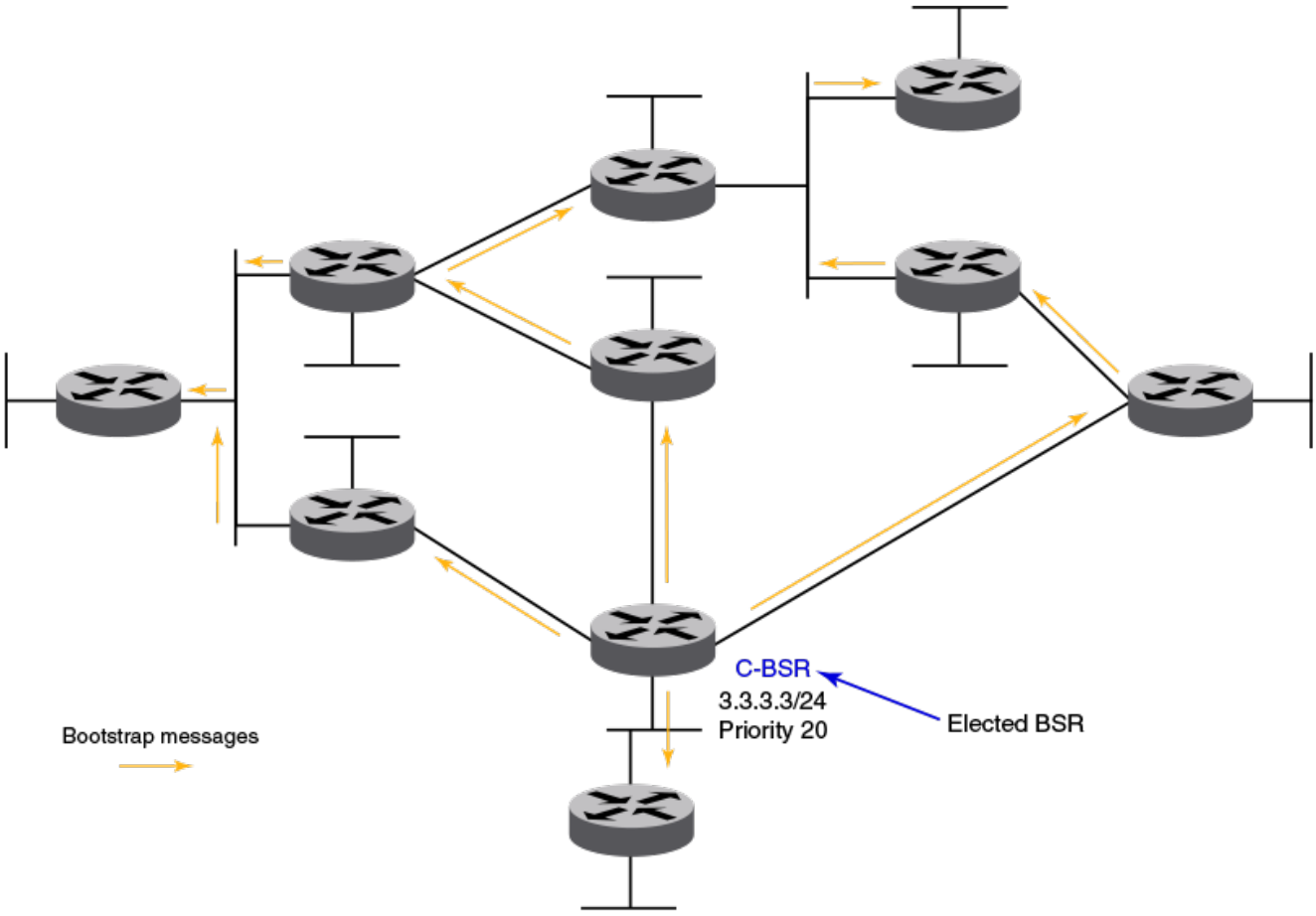


Figure 3: RP-set distribution

The RP-set built by the BSR is sent through the same BSM message. Because these BSMs are flooded, the RP-set information rapidly reaches each PIM router. When a PIM router receives the RP-set, it adds all group-to-RP mappings to its pool of mappings, also created from static RP configurations. Every PIM router runs the same RP hash algorithm to ensure the same C-RP is elected for a particular multicast group throughout the domain. In this way, all PIM routers can build the multicast group-specific distribution tree rooted to the same RP.

BSR Timers and Values

The BSR mechanism uses timers to ensure that the protocol provides reliability and faster convergence. These timers can be configured.

Timer	Default value	Description
Bootstrap message interval	60 seconds	The interval after which a BSM is generated by a BSR.
Bootstrap timeout	130 seconds	The interval after which a BSR times out if no BSM is received from it.
Bootstrap minimum interval	10 seconds	The minimum interval after which a BSR should send a BSM.

Timer	Default value	Description
C-RP mapping expiry timer	From message	Hold time from a C-RP advertisement message. The hold time for C-RP is 2.5 times the RP advertisement interval.
RP mapping expiry timer	From message	Hold time from BSM.
Candidate RP advertisement interval	60 seconds	The interval after which a C-RP generates an advertisement message to the BSR.

RP Election Algorithm (Group-to-RP Hashing)

The RP-set information received from the BSR is stored locally and updated by each PIM router periodically upon receiving BSMs. This RP-set contains the list for group prefixes and the corresponding list for C-RP for each group prefix.

The following steps explain the RP election process for a multicast group address.

1. A longest match look-up is performed on all the group prefixes in the RP-set.
2. If the look-up finds more than one C-RP, the C-RP with the lowest priority is elected.
3. If more than one C-RP has the same lowest priority, the BSR hash function is used to elect the RP.
4. If the hash functions return the same hash value for more than one C-RP, the C-RP with the highest IP address is elected.

Loopback Interfaces as RPs

Because loopback interfaces are operationally always up, it is preferable to use them as RPs. All existing PIM-SM protocol features are supported on loopback interfaces. Layer 3-enabled loopback interfaces can act as static RPs or candidate-RPs. They can also be configured as candidate-BSRs.

PIM-SM and PIM-SSM for Layer 3 Multicast Over MCT

Overview

PIM Sparse Mode (PIM-SM) and PIM Source Specific Multicast (PIM-SSM) are supported without synchronizing PIM-SM or PIM-SSM states across the multichassis tunnel (MCT) cluster.

The PIM snooping mechanism snoops the Join and Prune messages that are exchanged in the MCT VLANs and learns the interested VLAN member interfaces. Optimal traffic forwarding is achieved with PIM snooping states synchronized between the MCT peers, using the BGP EVPN Join Sync method.

The Reverse Path Forwarding (RPF) check and L3 multicast route lookup are based on the unicast routing protocol running in the MCT cluster.



Note

*The L3 PIM SM Multicast **doesn't work** for 232.0.0.1 - 232.255.255.255 after upgrade due to support of PIM-SSM.*

PIM-SM and PIM-SSM Behavior in MCT

PIM-SM and PIM-SSM can be enabled on VLAN router interfaces that are extended over MCT. These VE interfaces act as normal PIM routers, and converge just like any PIM router in an L3 domain. PIM-enabled VE interfaces on these MCT VLANs see each other as peers. They can exchange Join and Prune messages natively to form the L3 multicast forwarding tree over the interchassis link (ICL).

PIM-SM and PIM-SSM control traffic travels over the ICL link of the cluster, encapsulated with the tunnel encapsulation method used by MCT. For example MPLS/VxLAN.

For PIM-SM, any MCT node can be the rendezvous point (RP) and root of the shared multicast tree. This RP can be a PIM router on a VE interface for any extended VLAN, or it can be any other PIM-enabled L3 interface on the MCT node. Generally a PIM-enabled loopback interface performs the RP functionalities. All other functions of PIM are supported natively, such as FHR source registration, LHR thresholding and SPT switch-over, BSR protocol, and Anycast RP functionality.

SSM reserved(default) range 232.0. 0.0 through 232.255. 255.255

Non-SSM(non-default range) range is 224.0.0.0 to 231.255.255.255 -233.0.0.0 -239.255.255.255.

Before 20.2.3g code, with 18r200bd code, SSM(Source-Specific Multicast) support was not there and both default(Reserved) SSM and Non-SSM default(non-reserved) groups where all groups come in one range from 224.0. 0.0 to 239.255. 255.255 that works with PIM sparse mode, which is programmed on the box appropriately.

The PIM Sparse mode configured with IGMP V2 or IGMP V3 are joined and processed with Tailend router LL-OC2-SLX2, and (*, G) is created. The PIM joins for the group are sent upstream

towards RP along the RPT=T. So PIM Sparse mode works for all SSM groups(232.1.4.1-250 to 232.1.0.1-250) and for other groups.

The new SSM features supported SSM (source-specific Multicast)

SSM default groups include 232.0.0.1-232.255.255.255, where the “ssm-enable” option is required for default-groups(reserved) groups to work.

The SSM is programmed with SSM reserved range by default. And If IGMP v3 joins from the non-default SSM groups on the Tailend router i.e., groups 224.0.0.0 to 231.255.255.255 and 233.0.0.0 -239.255.255.255(other than 232.0.0.1 to 232.255.255) then we need to use “ssm-enable range” option So SSM will work for non-default groups and default-groups.

Command usage:

“`ssm-enable range <Prefix-list>` “ or

“`ssm-enable`”

Since you have IGMP V3 reports and no IGMP v2 reports, it is better to configure “`range`” for non-default(reserved) groups as below option1. Therefore, the PIM SSM is configured and programmed for all the IGMP V3 receivers.

Mrouter Detection and FHR Source Registration

The PIM Hello packet over ICL is snooped for multicast router (mrouter) detection. With this information, traffic from a directly connected multicast source in the MCT VLAN, which has a non-designated router (DR) PIM router interface, can forward on the ICL port that was learned as mrouter. The PIM-DR on the other MCT node can see the source traffic and perform the FHR functionality, which registers the multicast source with the RP.

Enable IPv4 PIM Globally

The **`router pim`** command enables IPv4 PIM routing and enters PIM router configuration mode.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable IPv4 PIM and enter router configuration mode.

```
device(config)# router pim
device(config-pim-router)#
```

Configure Options on an IPv4 PIM Router

When IPv4 PIM is enabled globally, you can configure several PIM options for a router. The following steps configure the default VRF. To configure a specific VRF, see Configure Options for IPv4 PIM Multi-VRF on page 31.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter router configuration mode.

```
device(config)# router pim
device(config-router-pim-vrf-default-vrf)#
```

3. Run the following commands as needed.
   a. Configure the frequency with which a device sends PIM hello messages to its neighbors.

```
hello-interval 40
```

   b. Configure the length of time a PIM device waits for hello messages before considering a neighbor to be absent.

```
nbr-timeout 160
```

c. Configure a bootstrap router as a candidate RP.

```
bsr-candidate interface loopback 11 mask 32
```

d. Configure the length of time a PIM device waits to stop traffic after receiving a Leave message.

```
prune-wait 5
```

e. Configure the frequency with which PIM Join and Prune messages are sent.

```
message-interval 180
```

f. Configure sources to use the shared RP tree instead of the Shortest Path Tree.

```
spt-threshold infinity
```

g. Configure sources to use the Shortest Path Tree instead of the shared RP tree.

```
no spt-threshold
```

h. Enable Source-Specific Multicast mode for a specific address range

```
ssm-enable range PL_ssm_range-230-to-234
```

i. Enable Source-Specific Multicast mode for the default address range of 232.0.0.0/8. This default range is displayed in the **`show ip pim settings`** output.

```
ssm-enable
```

Configure Options on an IPv4 PIM Interface

When IPv4 PIM is enabled globally, you can configure PIM options for an interface (Ethernet, loopback, or VE).

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter device configuration mode.

```
device(config)# interface ethernet 0/1
```

3. Run the following commands as needed.

   For more information, including examples, see the *Extreme SLX-OS Command Reference*.

   a. Enable IPv4 PIM on an interface.

```
device(conf-if-eth-0/1)# ip pim-sparse
```

   b. Specify the designated router (DR) priority of an IPv4 PIM interface.

```
device(conf-if-eth-0/1# ip pim dr-priority 200
```

   c. Configure the Time to Live (TTL) threshold for an IPv4 PIM interface.

```
device(conf-if-eth-0/1# ip pim ttl-threshold 50
```

Configure Options for IPv4 PIM Multi-VRF

With multi-VRF (Virtual Routing and Forwarding) support, all Layer 3 multicast protocols operate as separate instances, per VRF, depending on the VRF-specific multicast configuration. All the required configuration and mcast routing tables have multiple instances, per VRF, and function simultaneously, allowing network paths to be segmented without using multiple routers.

Multi-VRF supports multiple instances of Layer 3 multicast protocols on the same router at the same time. When IPv4 PIM is enabled globally, you can configure various options for IPv4 multicast over multi-VRF, such as the rendezvous point and the PIM hello interval.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter router configuration mode for the VRF.

```
device(config)# router pim vrf red
device(config-router-pim-vrf-red)#
```

3. Run the following commands as needed.

   For more information, including more examples, see the *Extreme SLX-OS Command Reference*.

   a. Configure a device interface as a rendezvous point (RP).

```
rp-address 100.1.1.1
```

   b. Configure a static RP from a group list.

```
rp-address 4.4.4.4 static-rp-list
```

   c. Configure Anycast RPs in multicast domains.

```
anycast-rp 100.1.1.1 anycast-rp-set
```

   d. Configure a bootstrap router as a candidate RP to distribute RP information.

```
bsr-candidate interface loopback 11 mask 32
```

   e. Configure the frequency with which a device sends PIM hello messages to its neighbors.

```
hello-interval 40
```

   f. Configure the frequency with which PIM Join and Prune messages are sent.

```
message-interval 180
```

   g. Configure the length of time a PIM device waits for hello messages before considering a neighbor to be absent.

```
nbr-timeout 160
```

   h. Configure the length of time a PIM device waits to stop traffic after receiving a Leave message.

```
prune-wait 5
```

   i. Configure a device as a candidate RP.

```
rp-candidate interface loopback 11
```

   j. Enable multicast ECMP load sharing with dynamic rebalancing.

```
rpf ecmp rebalance
```

   k. Enable Source-Specific Multicast mode for a specific address range.

```
ssm-enable range PL_ssm_range-230-to-234
```

Display IPv4 PIM Information

You can use show commands to display information about the internal state of IPv4 PIM. You can run the commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

1. Display various PIM settings, such as Hello interval and route precedence.

```
device# show ip pim settings
  Maximum mcache              : 24576     Current Count               : 0
  Hello interval              : 30        Neighbor timeout            : 105
  Join/Prune interval         : 60        Inactivity interval         : 180
  Hardware drop enabled       : 1         Prune wait interval         : 3
  Register Suppress Time      : 60        Register Probe Time         : 10
  Register Stop Delay         : 0         Register Suppress interval  : 0
  SSM Enabled                 : No        SPT Threshold               : 1
  Route Precedence            : uc-non-default uc-default none
```

2. Display route entries in the multicast mcache table.

```
device# show ip pim mcache 50.1.1.101 230.1.1.1
IP Multicast Mcache Table
Entry Flags    : sm  - Sparse Mode, ssm - Source Specific Multicast
                 RPT - RPT Bit, SPT - SPT Bit, LSrc - Local Source
                 LRcv - Local Receiver, RegProbe - Register In Progress
                 RegSupp - Register Suppression Timer, Reg - Register Complete
                 needRte - Route Required for Src/RP
Interface Flags: IM - Immediate, IH - Inherited, WA - Won Assert
                 MJ - Membership Join, BR - Blocked RPT, BA - Blocked Assert
                 BF - Blocked Filter
Total entries in mcache: 8
1    (50.1.1.101, 230.1.1.1) in Ve 40, Uptime 00:03:29
     Sparse Mode, RPT=0 SPT=1 Reg=0 RegSupp=0 RegProbe=0 LSrc=0 LRcv=1
     upstream neighbor=40.1.1.3
     num_oifs = 2
           Ve 2(00:03:29/181) Flags: IM
           Ve 10(00:03:29/0) Flags: MJ
     Flags (0x400784d1)
         sm=1 ssm=0 needRte=0
```

3. Display traffic statistics for PIM-enabled interfaces.

```
device# show ip pim traffic
Port      |HELLO  |JOIN  |PRUNE  |ASSERT |GRAFT/REGISTER |REGISTER-STOP |BSR-MSGS  |
RPC-MSGS
          |Rx     |Rx    |Rx     |Rx     |Rx             |Rx            |Rx        |
Rx
-----------+---------+---------+---------+---------+---------+---------+---------
+---------+
Ve10         54       0       0       0        0           0
0          0
Lo 1         0        0       0       0        0           0
0          0

device# show ip pim traffic
Port        |HELLO |JOIN  |PRUNE |ASSERT |GRAFT/REGISTER |REGISTER-STOP
|BSR-MSGS |RPC-MSGS
            | Tx   |Tx    | Tx   |Tx     |Tx             |Tx
-----------+---------+---------+---------+---------+---------+---------+---------
+---------+---
Ve10         29        0     0        0         0            0
0         0
Lo 1         28        0     0        0         0            0
0         0
```

4. Display information about active PIM neighbors.

```
device(config)# show ip pim neighbor
-----+--------+------------+--------+---+---------+---------+-----+----------
+-----------+----
Port |PhyPort |Neighbor    |Holdtime|T  |PropDelay|Override |Age
|UpTime    |VRF        |Prio
     |        |            |sec     |Bit|msec     |msec     |
sec  |          |           |
-----+--------+------------+--------+---+---------+---------+-----+----------
+-----------+----
v2     e0/1     2.1.1.2     105      1   500       3000
0     00:44:10   default-vrf  1
v4     e0/2     4.1.1.2     105      1   500       3000
10    00:42:50   default-vrf  1
v5     e0/1     5.1.1.2     105      1   500       3000
0     00:44:00   default-vrf  1
v22    e0/1     22.1.1.1    105      1   500       3000
0     00:44:10   default-vrf  1
Total Number of Neighbors : 4
```

5. Display information about the bootstrap router and the candidate RP.

```
device# show ip pim bsr
PIMv2 Bootstrap information for Vrf Instance : default-vrf
------------------------------------------------------------------------------
  This system is the Elected BSR
  BSR address: 1.51.51.1. Hash Mask Length 32. Priority 255.
  Next bootstrap message in 00:01:00
  Configuration:
    Candidate loopback 2 (Address 1.51.51.1). Hash Mask Length 32. Priority 255.
Next Candidate-RP-advertisment in 00:01:00
  RP: 1.51.51.1
    group prefixes:
    224.0.0.0 / 4
  Candidate-RP-advertisement period: 60
```

6. Display information about the candidate RP and its group mappings.

```
device# show ip pim rp-candidate
Next Candidate-RP-advertisement in 00:00:10
  RP: 207.95.7.1
    group prefixes:
    224.0.0.0 / 4
  Candidate-RP-advertisement period: 60
```

Explicitly Select the IPv4 PIM Rendezvous Point

With the PIM-SM election process, a backup RP automatically takes over if the active RP router becomes unavailable. However, you can explicitly select the RP, so that the device uses the selected RP for all group-to-RP mappings and overrides the set of candidate RPs supplied by the BSR.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter router PIM configuration mode.

```
device(config)# router pim
```

3. Specify the IP address of the RP.

```
device(config-pim-router)# rp-address 4.4.4.4
```

   The command in this example identifies the device interface at IP address 4.4.4.4 as the RP for the PIM-SM domain. The device uses the specified RP and ignores group-to-RP mappings received from the BSR.

4. To configure static RP with specific group ranges, run the following commands.

```
device(config-pim-router)# rp-address 4.4.4.4 static-rp-list
device(config)# ip prefix-list static-rp-list permit 225.1.1.0/24
```

   The following commands configure the RP candidate.

```
device(config-pim-router)# rp-candidate interface loopback 11
device(config-pim-router)# rp-candidate prefix my-rp-cand-list
device(config)# ip prefix-list my-rp-cand-list permit 226.1.1.0/24
device(config)# ip prefix-list my-rp-cand-list permit 228.1.1.0/24
```

IPv4 PIM Anycast RP

PIM Anycast RP (rendezvous point) provides load balancing and fast convergence to PIM RPs in an IPv4 multicast domain. The RP address of the Anycast RP is a shared address used among multiple PIM routers, known as PIM RP. The PIM RP routers create an Anycast RP set. Each router in the Anycast RP set is configured using two IP addresses: a shared RP address in their loopback address and a separate, unique IP address. The loopback address must be reachable by all PIM routers in the multicast domain. The separate, unique IP address is configured to establish static peering with other PIM routers and communication with the peers.

When the source is activated in a PIM Anycast RP domain, the PIM First Hop (FH) registers the source to the closest PIM RP. When there are external peers in the Anycast RP set, the router re-encapsulates the packet with the local peering address as the source address of the encapsulation. The router unicasts the packet to all Anycast RP peers. The re-encapsulation of the data register packet to Anycast RP peers ensures source state distribution to all RPs in a multicast domain.

The following example is a PIM Anycast-enabled network with 3 RPs and 1 PIM-FH router connecting to its active source and local receiver. Loopback 1 in RP1, RP2, and RP3 have the same IP address: 100.1.1.1. Each Loopback 2 in RP1, RP2, and RP3 has a separate IP address configured to communicate with their peers in the Anycast RP set.

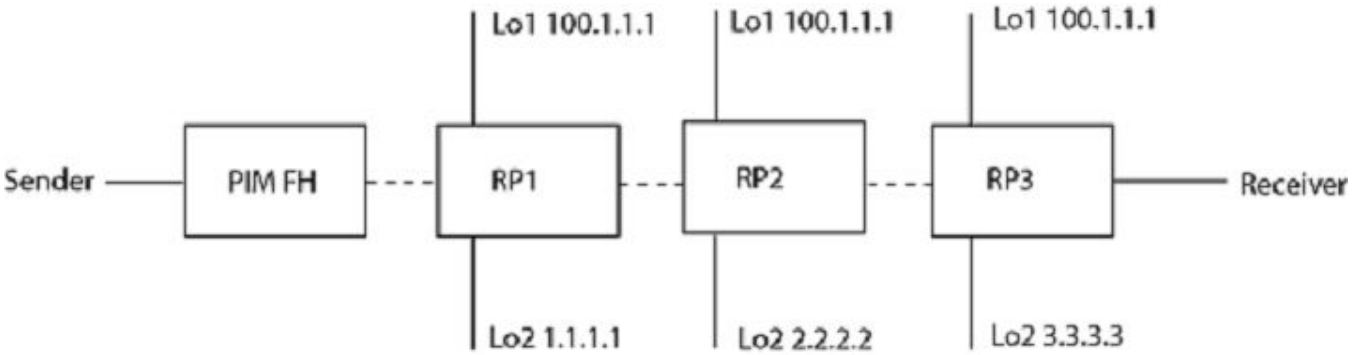


Figure 4: Example of a PIM Anycast RP network

Configure the IPV4 PIM Anycast RP

The **anycast-rp** command maps the RP and the Anycast RP peers.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter router PIM configuration mode.

```
device(config)# router pim
```

3. Identify the RP address for the PIM-SM domain.

```
device(config-pim-router)# rp-address 100.1.1.1
```

4. Specify the Anycast RP, including the prefix list that identifies the Anycast peers that are configured with the same Anycast RP address, also called the Anycast RP set.

```
device(config-pim-router)# anycast-rp 100.1.1.1 anycast-rp-set
```

The following is a configuration of PIM Anycast RP 100.1.1.1. The example avoids using the Loopback 1 interface when configuring PIM Anycast RP because the Loopback 1 address could be used as a router-id. A PIM First Hop router registers the source with the closest RP. The first RP that receives the register re-encapsulates the register to all other Anycast RP peers.

The RP shared address 100.1.1.1 is used in the PIM domain. IP addresses 1.1.1.1, 2.2.2.2, and 3.3.3.3 are listed in the ACL that forms the self-inclusive Anycast RP set. Multiple Anycast RP instances can be configured on a system, with each peer having the same or a different Anycast RP set.

```
device(config)# interface loopback 2
device(config-lbif-2)# ip address 100.1.1.1/24
device(config-lbif-2)# ip pim-sparse
device(config-lbif-2)# interface loopback 3
device(config-lbif-3)# ip address 1.1.1.1/24
device(config-lbif-3)# ip pim-sparse
device(config-lbif-3)# router pim
device(config-pim-router)# rp-address 100.1.1.1
device(config-pim-router)# anycast-rp 100.1.1.1 anycast-rp-set
device(config)# ip prefix-list anycast-rp-set permit 1.1.1.1/32
device(config)# ip prefix-list anycast-rp-set permit 2.2.2.2/32
device(config)# ip prefix-list anycast-rp-set permit 3.3.3.3/32
```

IPv4 Multicast ECMP Dynamic Rebalance

If there are multiple equal-cost paths between PIM routers to reach the source or the RP, the multicast RPF algorithms distribute the load across the available paths.

The following diagram shows a topology in which equal-cost multi-path (ECMP) support is not enabled. R1 through R11 have IP addresses in ascending order (R1 having the lowest IP address and R11 having the highest). All the routers are PIM-enabled routers. The links emanating from each router are ECMP links. The existing path usage is indicated in red. With the highest IP address neighbor chosen for the available ECMP paths, the multicast cache entries use only the R1-R4-R11-SRC path.

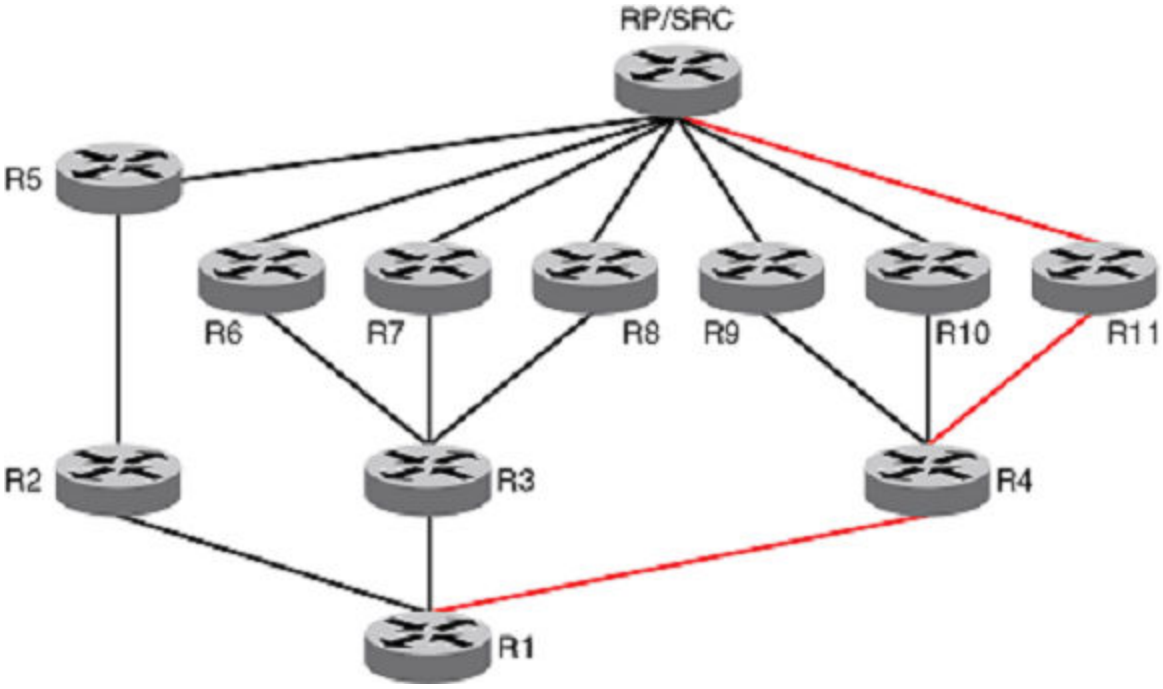


Figure 5: Path usage without multicast ECMP support

In the following figure, with the ECMP support turned on, the multicast entries are distributed among the equal-cost next hops as indicated in green for better usage of the available paths.

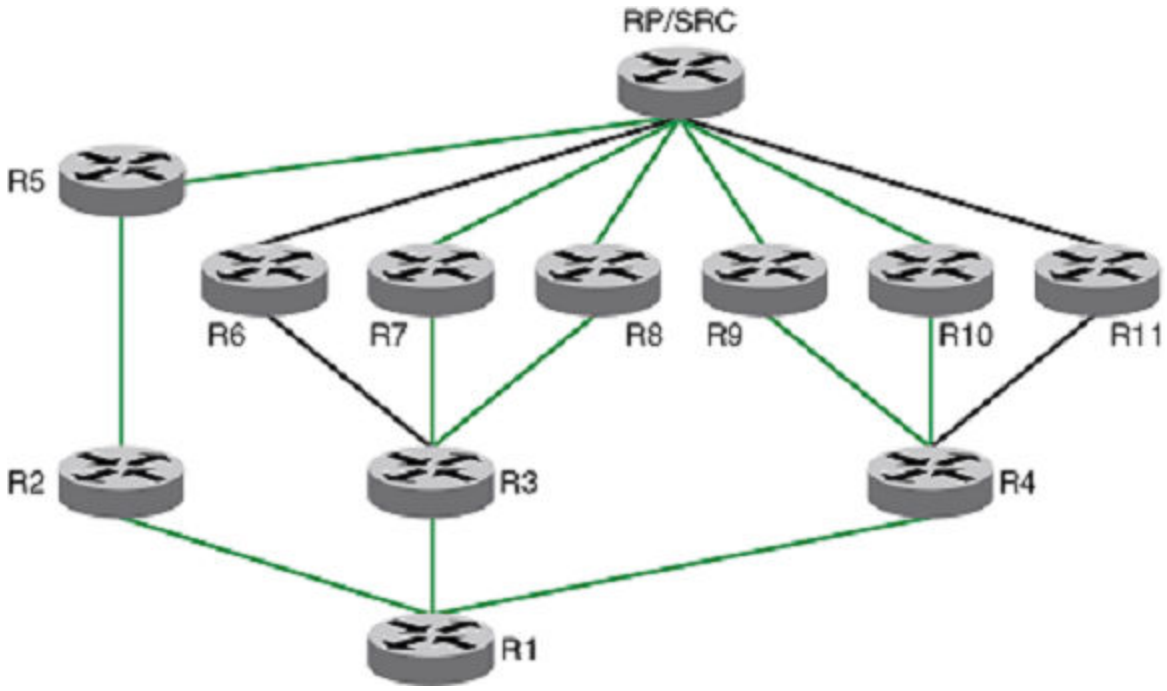


Figure 6: Path use with multicast ECMP support

The load distribution is achieved by distributing the multicast cache entries (*,G or S,G) to the available paths. Two methods are widely used to achieve this distribution.

- **Hash based**: Load splitting
- **Least-used path based**: Load balancing

Extreme Networks devices support the hash-based method of load distribution for multicast ECMP.

Hash-based load distribution

Hash-based load distribution depends on a hash function to distribute the multicast cache entries. The S, G, next-hop addresses are hash function based. This method splits the cache entries by choosing a different RPF neighbor and splits the traffic. Load balancing is based on the distribution of the keys S, G, next-hop. This method of distribution is the least disruptive because the hashing redistributes only those cache entries that are affected during link flaps. Some paths may not be used for the distribution of the multicast entries. For example, for the ECMP paths from R3 to R6, R7 and R8, only paths R3 to R7 and R3 to R8 are used.

Path failure behavior

When an ECMP path is not operating, all the multicast entries using that path are redistributed among the other available paths.

New path behavior

When a new path is added to the ECMP set, there is no redistribution (default behavior without the rebalance option) of the cache entries. Here, optimal use of the paths is chosen in favor of not disturbing the existing flow. This method sometimes requires a full branch setup toward the source or RP of the multicast distribution tree. When a path flaps (stops and starts), the multicast entries that had been using this path do not use this path anymore. The situation worsens if a subset of paths stops operating and then starts up one path at a time. In this scenario, the only paths that carry all entries are the paths that did not flap.

Dynamic rebalancing

This option, which uses the hash method, rebalances the traffic immediately upon the addition of a new next-hop or path, and helps in new next-hop and path addition and path flap cases. There is less disruption in existing flows by using the hash method.

Considerations

The following considerations apply to the configuration of ECMP load balancing.

- Because the hash method is a load-splitting method, traffic load balancing is not supported.
- S-based and S,G-based hashing is not supported.
- The load-balancing effect due to load splitting the multicast entries is only a best effort. Splitting is actually based on the number of S, G flows, the number of next-hops, and the actual distribution of the S,G and the next-hop addresses.
- If the rebalancing is not configured, then link flap results in sub-optimal use of the ECMP links.
- Multicast ECMP and unicast ECMP both support 32 paths.

Enable ECMP Dynamic Rebalance

ECMP dynamic rebalance enables hash-based distribution among the ECMP paths when a new next-hop is added.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter router PIM configuration mode.

```
device(config)# router pim
```

3. Enable ECMP load sharing with dynamic rebalance.

```
device(config-pim-router)# rpf ecmp rebalance
```

Multicast Traceroute Diagnostics

Multicast traceroute (mtrace) is a diagnostic tool that traces the multicast path from a specified destination to a source for a multicast group. It runs over the IGMP protocol.

The unicast traceroute program allows the tracing of a path from one device to another. The key mechanism for unicast traceroute is the ICMP TTL exceeded message, which is specifically excluded as a response to multicast packets. The mtrace facility allows the tracing of an IP multicast routing path. Mtrace also requires special implementations on the part of routers.

Mtrace uses any information available in the router to determine a previous hop to forward the trace toward the source. Multicast routing protocols vary in the type and amount of state they keep. Mtrace endeavors to work with all of them by using whatever is available. For example, if a PIM-SM router is on the (*,G) tree, mtrace chooses the parent toward the RP as the previous hop. In such a case, no source or group-specific state is available, but the path can still be traced.

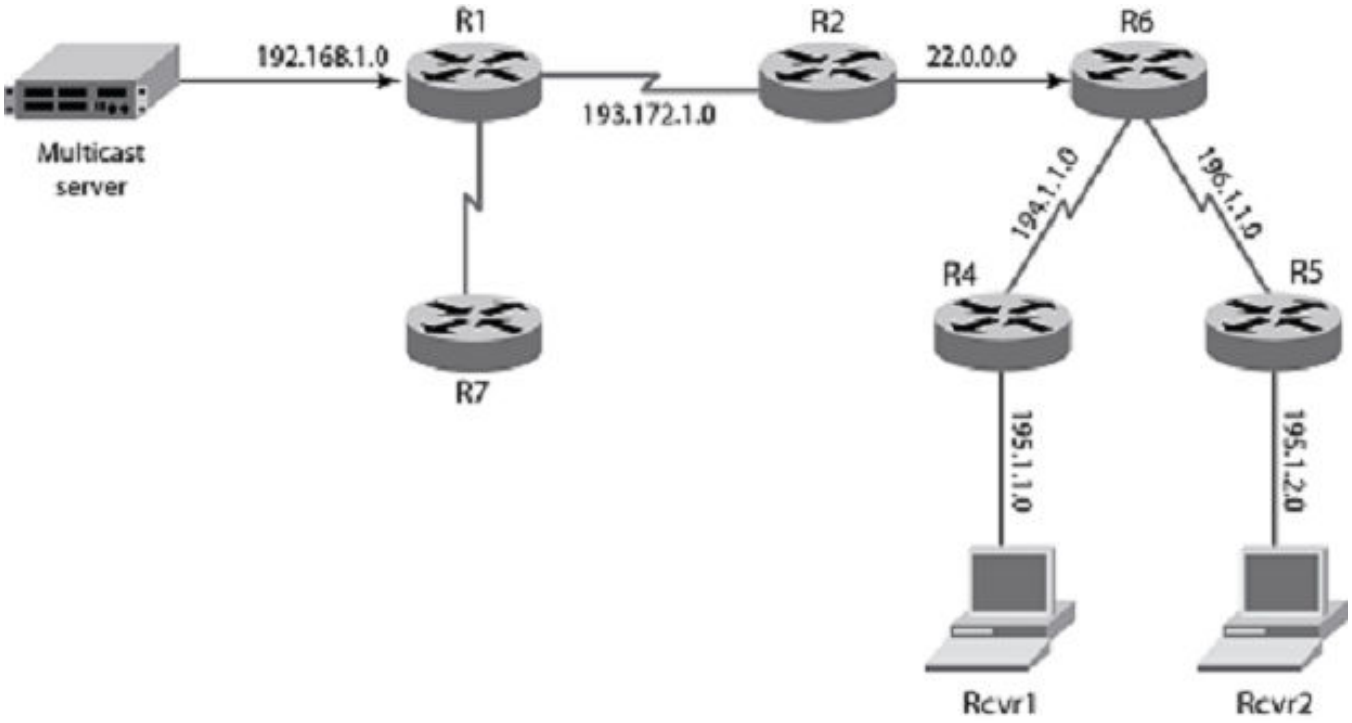


Figure 7: Network topology

Primary components of an mtrace implementation

Mtrace query

The party requesting the traceroute sends a traceroute query packet to the last-hop (LH) multicast router for the given destination. The query and request have the same opcode. The receiving router can distinguish between a query and a request by checking the size of the packet. A query is a request packet with none of the response fields filled up.

Mtrace request

The LH router changes the query packet into a request packet by adding a response data block containing its interface addresses and packet statistics. The LH router then forwards the request packet by unicast to the router that it believes is the proper previous hop for the given source and group. Each hop adds its response

data to the end of the request packet, then forwards it by unicast to the previous hop.

Mtrace response

The first hop (FH) router is the router that believes that packets from the source originate on one of its directly connected networks. The FH router changes the packet type to a response packet and sends the completed response to the response destination address. The response may be returned before reaching the FH router if a fatal error condition, such as "no route," is encountered along the path.

Configure Mtrace

Mtrace can be started on any router on the network.

Enable mtrace and specify the source, destination, and group IP addresses.

```
SLX# mtrace source 10.1.1.2 destination 40.1.1.1 group 225.0.0.1

2020/01/17-06:48:37.451759 MTRACE.EVT: fwd code changed from MTRACE_NO_ERR to
2020/01/17-06:48:37.451828 MTRACE_NO_ERRSending mtrace query from src 10.1.1.2 to dest
40.1.1.1 through group 225.0.0.1

Collecting Statistics, waiting for 5 seconds.....

Type Control-c to abort
0          30.1.1.1         NONE       THRES      1  MTRACE_NO_ERR
1          15.1.1.2         NONE       THRES      1  MTRACE_NO_ERR
2          10.1.1.1         NONE       THRES      1  MTRACE_REACHED_RP
```

Topology:

SRC-(10.1.1.2)----------(10.1.1.1)--R1(FHR)--(15.1.1.1)----------(15.1.1.2)--
R2--(30.1.1.2)---------(30.1.1.1)--R3(LHR)--(40.1.1.1)---------(40.1.1.2)--HOST

In this example the destination IP is 40.1.1.1, the source IP is 10.1.1.2, and the group IP is 225.0.0.1. The mtrace query is sent from R3 to R2. The initial header is not modified by any of the routers. R2 adds a response block based on the (S, G) or the (*, G) entry, adds its incoming interface, outgoing interface, and other information, and sends it to its upstream neighbor. The process continues until the query reaches the First Hop Router (FHR), which is R1. R1 then determines that it is the FHR, completes the response block, and sends the response back to R3. R3 reads the information from the packet and prints it out.

Layer 2 Multicast Over MCT

SLX-OS devices support Layer 2 multicast control packets over multichassis trunks (MCT). Multicast state information is synchronized between MCT peers using MP-BGP EVPN transport. Multicast protocol packets are sent on the peer link only when required.

Internet Group Management Protocol (IGMP) protocol packets are of three types.

- **IGMP Query**: General query and Group-specific query

- **IGMP Report**: Version 1, Version 2, and Version 3 Membership reports. In a report, the multicast address field contains the specific multicast address to which the sender is listening.
- **IGMP Leave**: Version 2 group leave

Protocol Independent Multicast (PIM) protocol packets are of 4 types.

- PIM hello
- PIM join/prune
- BSR and candidate-RP advertisements
- RP registration & null-registration

IGMP Query Packet Processing

Each EVI is associated with a multicast group ID (MGID) that is BUM-suppressed (Broadcast, unknown-unicast and multicast). However, query packets need to be transmitted on an interchassis link (ICL) to address the following scenarios.

- The querier connected to only one of the MCT peer switches is the elected querier.
- Only one of the peer switches is configured as a querier.
- The switch ages out IGMP routes if memberships are not confirmed during the timeout interval. Although query packets are received on the MCT peer link, the mrouter port is not learned or considered on that peer link.

IGMP Membership Reports

- Traditionally, each peer switch learns about Layer 2 multicast memberships by snooping the IGMP membership reports. The membership reports are then flooded on multicast router (mrouter) ports.
- For MCT, because an mrouter port is not learned on the peer link, membership reports are not flooded between the peer switches. Peer switches exchange learned routes by using EVPN NLRI messages between Border Gateway Protocol (BGP) peers running on the MCT cluster control VLAN.
- MCT module (L2RIB) which handles the exchange of information across MCT cluster, communicates Multicast routes to Multicast module.
- If a general query or group-specific query is received from any port other than a peer link, each peer switch generates a proxy report for the IGMP routes learned across MCT.

Duplicate IGMP Query Packets on CCEP

If a query is configured on both MCT peers of a member VLAN, duplicate query packets reach clients that are connected to the MCT domain by means of a Cluster Client Edge Port (CCEP) or Cluster Edge Port (CEP).

Same BGP EVPN IGMP Join Sync Route is used to exchange IGMP Querier configuration on the member VLANs. Only one MCT Peer per VLAN is be elected as IGMP Querier based on higher MCT Peer IP address.

IGMP Leave

When fast-leave is not configured and an MCT peer receives a leave membership report from one of its clients for group G, the switch or router informs other MCT peers about the group-specific query and latency by using Leave Sync Route. The peer switch, which runs the querier, sends group-specific queries and group queries to the local VLAN ports.

Mrouter Synchronization

Mrouter synchronization helps in achieving optimal path selection for unknown multicast traffic and optimal MP-BGP message exchange between MCT peers. Mrouter port information is synchronized to MCT Peer using the same BGP EVPN IGMP Join Sync route. For Mrouter detection on CCEP client port, the traffic is forwarded based on Local Bias forwarding behavior. For all CEP Mrouter ports learnt, only one Sync is used, first add and last delete, unlike CCEP Mrouter ports which is per Client ESI.

Device Support

Extreme Networks supports Layer 2 multicast over MCT on the following devices:

- SLX 9640
- SLX 9540
- SLX 9250
- SLX 9150
- Extreme 8720
- Extreme 8520

Layer 2 Multicast Traffic Forwarding

When a receiver connected on a Cluster Client Edge Port (CCEP) sends a membership report to join group G, the election of the designated forwarder (DF) for the CCEP and MGID prevents duplication of multicast data packets destined to group G on CCEP and on the peer link.

DF election is always honored for programming receivers on CCEP in the multicast group ID (MGID). However, the path given by DF election may not be optimal, because it might direct the multicast data traffic originating at one peer switch to a receiver on CCEP over the peer link, even though (*,G) membership does not include the Cluster Edge Port (CEP) on the peer switch that does not host the source.

Optimal Traffic Forwarding

When a group is learned over member VLANs, the DF for the IGMP route is elected by hashing on the IVID parameter of the member VLAN, Source-IP, and Group-IP.

The OIF (outgoing interface) list of the IGMP route on the DF includes the following.

- Receivers connected through CCEP

- Interchassis link (ICL), if its multichassis trunk (MCT) peer has receivers connected through CEP
- Receivers connected through local CEP

The OIF list of the IGMP routes on the non-DF includes the following.

- ICL to redirect the multicast stream to the DF of the stream
- Receivers connected through CEP

Layer 2 Multicast Data Encapsulation

Data encapsulation of Layer 2 multicast from CEP and CCEP received on a member VLAN is similar to that for Layer 2 flooding traffic.

Local forwarding occurs in the following scenarios.

- The Cluster Edge Port (CEP) is not operating.
- The remote Cluster Client Edge Port (CCEP) is not operating.
- The ingress is the CEP.
- The ingress is a different CCEP.

Local forwarding does not occur in the following scenarios.

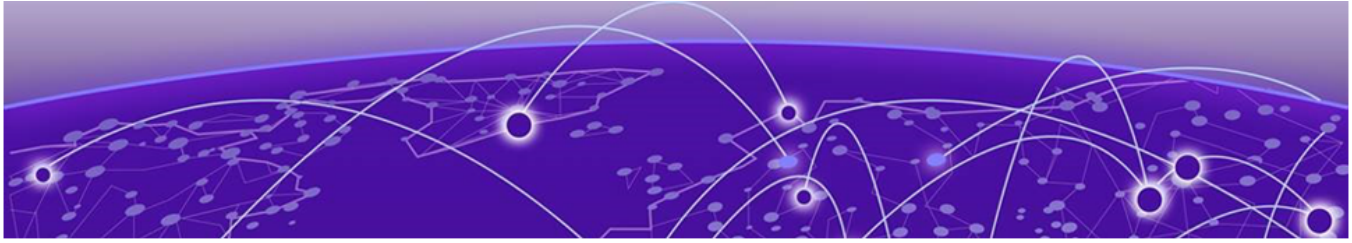
- The local CCEP is not operating.
- The ingress is the interchassis link (ICL).

Layer 3 Multicast over MCT

Device Support

Extreme Networks supports Layer 3 multicast over MCT on the following devices:

- SLX 9150
- SLX 9250
- SLX 9540
- SLX 9640
- Extreme 8520
- Extreme 8720



IPv4 Multicast Traffic Reduction

IGMP Traffic Snooping on page 45
IPv4 PIM-SM Traffic Snooping on page 51

IGMP Traffic Snooping

A Layer 2 switch forwards all multicast control packets and data received on all the member ports of a VLAN interface. This simple approach is not bandwidth efficient, because only a subset of member ports may be connected to devices that want to receive these multicast packets.

In a worst-case scenario, the data is forwarded to all port members of a VLAN, even if only one VLAN member is interested in receiving the data. Such scenarios can lead to loss of throughput for switches that receive a high rate of multicast data traffic.

IGMP snooping is a mechanism by which a Layer 2 device can effectively address the issue of inefficient multicast forwarding to VLAN port members. Snooping involves "learning" forwarding states for multicast data traffic on VLAN port members from the IGMP control (join and leave) packets received on them. The Layer 2 device also provides for a way to configure forwarding states statically through the CLI.

Multicast Routing and IGMP Snooping

Multicast routers use IGMP snooping to learn which groups have members on their attached physical networks. A multicast router keeps a list of multicast group memberships for each attached network and a timer for each membership. In a multicast group membership, at least one member of a multicast group on an attached network is available.

Hosts join multicast groups in one of the following ways.

- By sending an unsolicited IGMP join request
- By sending an IGMP join request in response to a general query from a multicast router

In response to the request, the host creates an entry in its Layer 2 forwarding table for that VLAN. When other hosts send join requests for the same multicast, the device adds them to the existing table entry. Only one entry is created per VLAN in the Layer 2 forwarding table for each multicast group.

VLANs can be configured as snooping only or routing with snooping. When Layer 3 multicast routing is enabled on a VE, snooping for the underlying VLAN is enabled implicitly. Both explicit and implicit snooping can be enabled on a VLAN. Implicit snooping is the default. When explicit snooping is disabled on a VE that has routing enabled, snooping reverts back to implicit snooping. This does not change the functionality, but only removes the configuration.

When routing is disabled on a VE where explicit snooping is configured, the routing side of the programming stops and the snooping side of the programming takes over. When routing is enabled, the Layer 3 IGMP querier takes precedence on that VLAN. When routing is disabled and the snooping querier is configured, then the snooping querier takes effect.

Enable IGMP Snooping on a VLAN

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter VLAN configuration mode.

```
device(config)# vlan 1
device(config-vlan-1)#
```

3. Enable IGMP snooping.

```
device(config-vlan-1)# ip igmp snooping enable
```

Configure IGMP Snooping on a VLAN

Use IGMP snooping in a VLAN when PIM is not configured. The IGMP snooping querier sends IGMP queries to trigger IGMP responses from devices that are to receive IP multicast traffic. The IGMP snooping querier listens for these responses to map the appropriate forwarding addresses.

Note
The IGMP snooping querier is suspended if Layer 3 IGMP is enabled on any of the cluster nodes.

1. Access global configuration mode.

```
device# configure terminal
```

2. Enter VLAN configuration mode.

```
device(config)# vlan 25
```

3. Specify the IGMP query interval for the VLAN.

```
device(config-vlan-25)# ip igmp snooping query-interval 125
```

The valid range is from 1 through 18000 seconds. The default is 125 seconds.

4. Specify the last member query interval.

```
device(config-vlan-25)# ip igmp snooping last-member-query-interval 1000
```

The valid range is from 1000 through 25500 milliseconds. The default is 1000 milliseconds.

5. Configure the static mrouter port.

```
device(config-vlan-25)# ip igmp snooping mrouter interface ethernet 0/2
```

6. Configure a static IGMP group.

```
device(config-vlan-25)# ip igmp snooping static-group 225.0.0.1 interface ethernet 0/15
```

7. Configure the IGMP version.

```
device(config-vlan-25)# ip igmp snooping version v3
```



Note
Version 2 is enabled by default. When you change the version, existing static or dynamic groups are deleted. These groups are relearned when the next query is sent.

8. Activate the IGMP snooping querier functionality for the VLAN.

```
device(config-Vlan-25)# ip igmp snooping querier enable
```



Note
The IGMP snooping querier and the static mrouter can be configured together on a VLAN interface.

Configure IGMP Snooping on a Bridge Domain

A bridge domain is a set of different types of service endpoints, such as pseudowire and VxLAN tunnel, grouped into one broadcast domain that allows any-to-any bridging. IGMP snooping on a bridge domain learns the multicast group on specific ports that are associated with the bridge domain.

IGMP snooping traps the IGMP control packets and programs the hardware entries with learned multicast groups and a list of interested ports that are part of the bridge domain.

When multicast traffic comes from a source, the traffic is sent to the interested receivers instead of flooding the multicast traffic on all ports of the bridge domain. IGMP on the bridge domain internally works the same as it works on a VLAN. A bridge domain contains logical interfaces (LIFs), so the corresponding multicast groups contain the LIFs as the outgoing interfaces.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter bridge domain configuration mode.

```
device(config)# bridge-domain 10
device(config-bridge-domain-10)#
```

3. Run the following commands as needed.

   For more information, including examples, see the *Extreme SLX-OS Command Reference*.

   a. Enable IGMP snooping on a bridge domain.

   ```
   ip igmp snooping enable
   ```

   b. Enable the IGMP querier on a bridge domain.

   ```
   ip igmp snooping querier enable
   ```

   c. Specify the IGMP version on a bridge domain.

   ```
   ip igmp version 2
   ```

   d. Enable fast-leave processing on a bridge domain, which allows the removal of an interface from the forwarding table without sending group-specific queries to the interface.

   ```
   ip igmp snooping fast-leave
   ```

   e. Specify the interval between snooping queries.

   ```
   ip igmp snooping query interval 30
   ```

   f. Specify the maximum amount of time to wait for a response from a snooping query.

   ```
   ip igmp snooping query-max-response-time 20
   ```

   g. Specify the time limit for sending last member queries.

   ```
   ip igmp snooping last-member-query-interval  150
   ```

The following example shows all the possible completions for the **ip igmp snooping** command.

```
device(config-bridge-domain-10)# ip igmp snooping ?
Possible completions:
  enable                        IGMP Enable
  fast-leave                    Fast Leave Processing
  last-member-query-interval    Last Member Query Interval
  mrouter                       Multicast Router
  querier                       Querier
  query-interval                Query Interval
  query-max-response-time       IGMP Max Query Response Time
  static-group                  Static Group to be Joined
  version                       IGMP Snooping Version
```

Monitor IGMP Snooping

You can run the following commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

1. Display all information about IGMP multicast groups for the device, including configured entries for all groups on all interfaces, all groups on specific interfaces, or specific groups on specific interfaces.

```
device# show ip igmp groups
Total Number of Groups: 2
IGMP Connected Group Membership
Group Address  Interface Uptime      Expires      Last Reporter   Version
225.1.1.1      vlan25    00:05:27     00:02:32     25.1.1.1202
Member Ports: eth 0/24
```

2. View snooping configuration information.

```
device# show ip igmp snooping
 Vlan ID: 10
 Multicast Router ports:  eth0/1
 Querier - Disabled
 IGMP Operation mode: IGMPv3
 Is Fast-Leave Enabled : Enabled
 Max Response time = 10
 Last Member Query Interval = 1
 Query interval = 125
 Number of Multicast Groups: 0
```

3. View information about the multicast cache.

```
device# show ip multicast snooping mcache
Flags : V2|V3 : IGMP Receiver, P_G : PIM (*,G) Join, P_SG: PIM (S,G) Join
                 BR : PIM Blocked RPT
 Vlan ID : 10
 -------------
1     (20.20.20.20, 232.0.0.10 )  22:37:48    NumOIF: 1
        Outgoing Ports:
                eth0/34        Flags: 0x24 ( V3)  00:00:08/252s
```

4. Display the IGMP statistics for a VLAN or interface.

```
device# show ip igmp statistics vlan 1

IGMP packet statistics for all interfaces in vlan 1:
IGMP Message type     Edge-Received   Edge-Sent   Edge-Rx-Errors   ISL Received
Membership Query                  0           0                0              0
V1 Membership Report              0           0                0              0
V2 Membership Report              0           0                0              0
Group Leave                       0           0                0              0
V3 Membership Report              0           0                0              0
PIM hello                         0           0                0              0

IGMP Error Statistics:
Unknown types            0
Bad Length               0
Bad Checksum             0
```

5. Display the Layer 3 IGMP interface configuration information.

```
device# show ip igmp interface
Interface Ve100
IGMP enabled
IGMP query interval 30 seconds
IGMP other-querierinterval 65 seconds
IGMP query response time 10 seconds
IGMP last-member query interval 1 seconds
IGMP immediate-leave disabled
IGMP querier100.0.0.1(this system)
IGMP version 2
```

6. Display multicast router port information.

```
device# show ip igmp snooping mrouter vlan 10
Vlan     Interface    Expires (Sec)
10       eth0/4       250
10       eth0/1       238
```

7. Display the SSM mapping with the prefix list name and source address details.

```
device# show ip igmp ssm-map

+-----------------------------------+-------------------+
|         PrefixList Name           |   Source Address  |
```

```
+-----------------------------------+-------------------+
   ssm-map-230-to-232                 203.0.0.10
  ssm-map-233-to-234                 204.0.0.11
```

IGMP Snooping and Unknown Multicast Traffic

IGMP snooping floods unknown multicast data packets on the member ports of VLAN. This unknown data traffic is the traffic sent to multicast groups that are not learned by means of IGMP membership reports or static IGMP group configuration.

Restricting the unknown multicast is achieved by redirecting the unknown multicast traffic (traffic destined to multicast MAC 01:00:5E:XX:XX:XX) to a multicast group identifier (MGID) that has no replication ports. Unknown multicast traffic is dropped because there are no ports to send the traffic.

Traffic Forwarding Scenarios

The following scenarios describe how multicast data traffic is forwarded when the restrict unknown multicast feature is enabled.

Multicast traffic learned from IGMP reports and static groups

IGMP v2 groups are programmed in LEM table as (*,G,V) entries.

IGMP v3 groups are programmed in KAPS as (S,G,V) entries.

The multicast data traffic is forwarded based on a match from LEM or KAPS table, which provides an MGID. The multicast traffic is replicated to an MGID that has egress ports.

Unknown multicast traffic

All the unknown multicast data traffic is restricted or dropped instead of flooded on the VLAN members.

Mrouter port behavior

Mrouter ports that are learned by means of receiving IGMP queries, PIM hello messages, or static configuration are added as part of the restrict unknown multicast MGID. Unknown multicast data traffic is then flooded only to the mrouter ports of the VLAN.

TCAM Profile

The restrict unknown multicast functionality is supported in default TCAM profile.

Limitations

- The MGIDs used for redirecting the unknown multicast traffic are allocated from the IGMP MGID space. The maximum number for IGMP groups without restrict unknown multicast is 16384.
- When you enable restrict unknown multicast on a VLAN or bridge domain, the IGMP groups maximum scaling number is reduced by 1.

Note
The maximum number of 15872 IGMP groups can be learned, if restrict unknown multicast is enabled on 512 VLANs or bridge domains.

Disable the IGMP Router Alert Option

By default, IGMP snooping checks for the presence of the router alert option in the IP packet header of an IGMP message. Packets that do not include this option are dropped. When you disable the router alert, you also disable the snooping check for the presence of the router alert option.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Disable the router alert option.

```
device(config)# ip igmp router-alert-check-disable
```

IPv4 PIM-SM Traffic Snooping

Overview

By default, a device does not examine the IP multicast information in a packet. Instead, the device forwards the packet out all ports except the port that received the packet. In some networks, this method causes unnecessary traffic overhead. For example, if the device is attached to only one group source and two group receivers, but has devices attached to every port, the device forwards group traffic out all ports in the same broadcast domain except the port attached to the source, even though there are only two receivers for the group.

PIM Sparse Mode (PIM-SM) traffic snooping eliminates the superfluous traffic by configuring the device to forward IP multicast group traffic only on the ports that are attached to receivers for the group.

PIM-SM traffic snooping requires IP multicast traffic reduction to be enabled on the device. IP multicast traffic reduction configures the device to listen for IGMP messages. PIM-SM traffic snooping provides a finer level of multicast traffic control by configuring the device to listen specifically for PIM-SM Join and Prune messages sent from one PIM-SM router to another through the device.



Note
This snooping feature applies only to PIM-SM version 2 (PIM v2).

Assumptions and Dependencies

- Standalone PIM snooping is not supported.
- IGMP snooping must be enabled in order to enable PIM snooping on the VLAN.
- Enabling Layer 3 multicast on a VE interface implicitly enables IGMP snooping and PIM snooping by default.
- PIM snooping is disabled when IGMP snooping is disabled.
- A global PIM snooping CLI is not available.
- Disabling Layer 3 multicast disables IGMP and PIM snooping unless configured by the user.

- When Layer 3 multicast is configured, you cannot disable IGMP or PIM snooping
- Clearing the IP IGMP groups clears the whole snooping database.

PIM-SM Snooping on a Bridge Domain

A bridge domain is a set of different types of service endpoints, such as pseudowire and VxLAN tunnel, grouped into one broadcast domain that allows any-to-any bridging. A Layer 3 VE interface can be bound to a bridge domain. Because PIM-SM can be configured on the VE, PIM-SM snooping is supported on a bridge domain.

PIM Snooping in an SSM Range

Source-Specific Multicast (SSM) is required only at the last-hop router (LHR) and is not required for the intermediate switch or router. At the LHR, if the SSM map is enabled in the IGMP, the v2 report in the SSM group range is converted to (S,G) join and sent to PIM. If a host sends a v3 report in the same SSM range, and the router is v3 enabled with an SSM map configured on this router, then this report is dropped. From the PIM snooping perspective, the (S,G) join is only created in the software.

IPv4 PIM-SM Snooping Example

In this example, a device is connected through an IP router to a PIM-SM group source that is sending traffic for a multicast group. The device also is connected to a receiver for the group.

When PIM-SM traffic snooping is enabled, the device listens for PIM-SM Join and Prune messages and IGMP group membership reports. Until the device receives a Join message or a membership report, it forwards IP multicast traffic out on all ports. When the device receives a Join message or group membership report, the device forwards subsequent traffic for that group only on the ports from which the Join messages or reports were received.

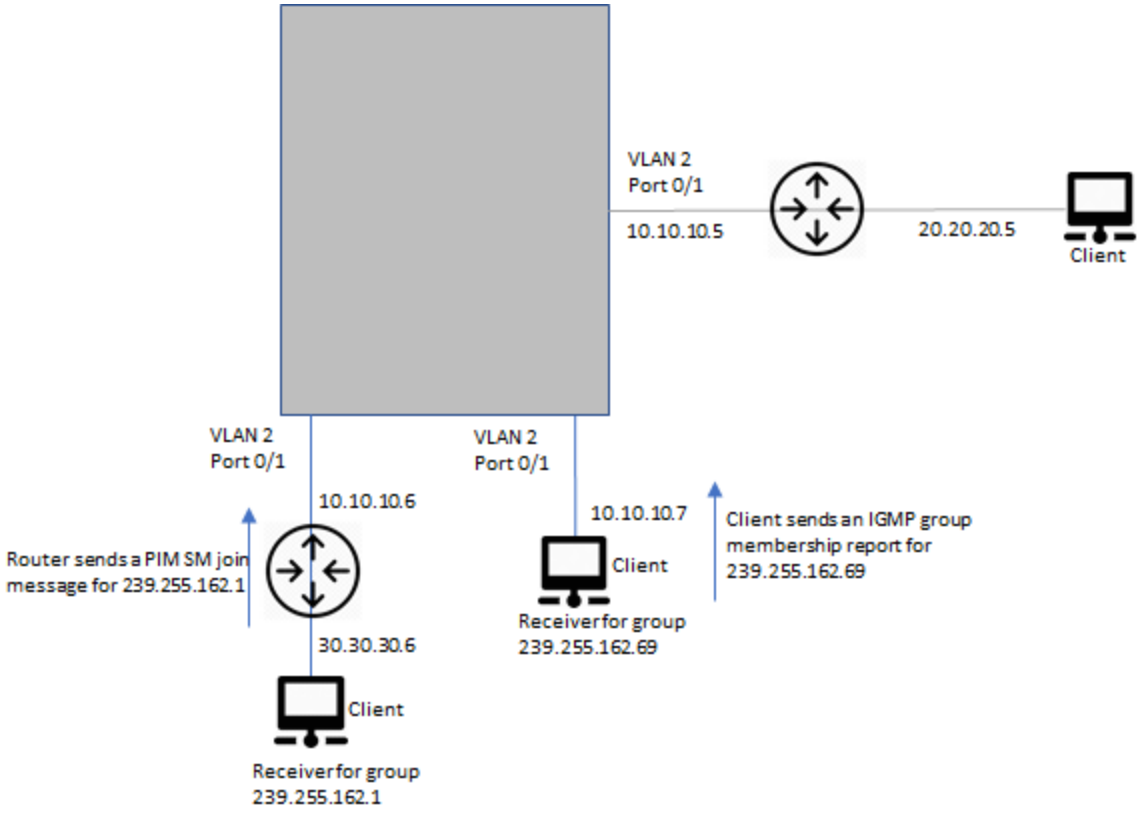


Figure 8: PIM SM traffic reduction in an enterprise network

In this example, the switch snoops for PIM SM Join and Prune messages. The router connected to the receiver for group 239.255.162.1 sends a Join message toward the group's source. Because PIM-SM traffic snooping is enabled on the device, the device examines the Join message to learn the group ID, then makes a forwarding entry for the group ID and the port connected to the receiver's router. The next time the device receives traffic for 239.255.162.1 from the group's source, the device forwards the traffic only on port 0/1, because that is the only port connected to a receiver for the group.

Enable IPv4 PIM Snooping on a VLAN

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter VLAN configuration mode.

```
device(config)# vlan 1
```

3. Enable PIM snooping.

```
device(config-vlan-1)# ip pim snooping enable
```

Enable IPv4 PIM Snooping on a Bridge Domain

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter bridge domain configuration mode.

```
device(config)# bridge-domain 10
```

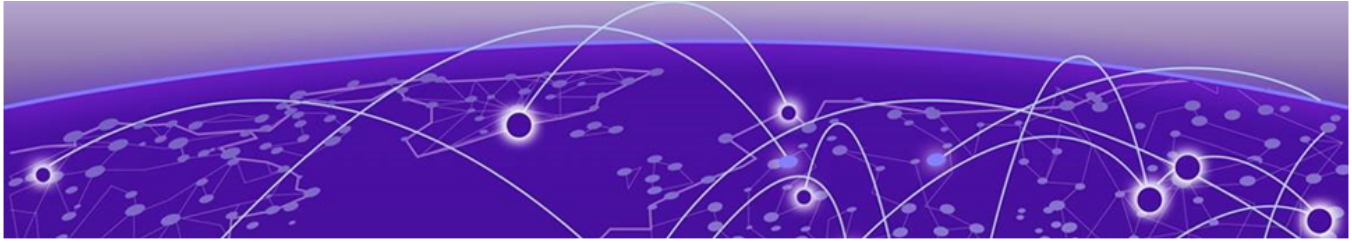
3. Enable PIM snooping.

```
device(config-bridge-domain-10)# ip pim snooping enable
```

PIM Multicast Router Presence Detection

The PIM multicast router presence detection feature scans the network traffic for incoming PIM hellos. This feature is enabled when multicast routing or snooping is enabled.

When a PIM hello is detected, that port is marked for the presence of a multicast router and the information is saved. This action prevents unnecessary flooding if the PIM designated router (DR) goes offline, because IGMP reports are forwarded to the multicast routers and not only the snooping-enabled router.



IP Multicast Fabric

IP Multicast Fabric Overview on page 55
Reference Network on page 56
Ingress Replication on page 56
Overview of Multicast Solution on page 57
Ideal Multicast Distribution on page 58
Unicast VXLAN Tunnel Establishment on page 58
Multicast Distribution Tree on page 60
Configure Optimization Replication on page 64
MDT Scale on page 65
Configure Multicast IP Fabric with L3 VNI on page 65
Logical VTEPs and MCT (Multi-homing) on page 67
Bud Node Topology on page 70

IP Multicast Fabric Overview

IP Multicast Fabric replaces ingress replication, which sends packets once over each unicast vxLAN tunnel. IP Multicast Fabric uses Multicast Distribution Trees to deliver traffic effectively while minimizing packet replication in the fabric.

Supported Platforms

IP multicast fabric is supported on the following platforms:

- SLX 9150-48Y (leaf/spine)
- SLX 9150-48XT (leaf/spine)
- SLX 9250-32C (leaf/spine)
- SLX 9540 (leaf/spine)
- SLX 9640 (leaf/spine)
- SLX 9740 (spine)
- Extreme 8520-48Y (leaf/spine)
- Extreme 8520-48XT (leaf/spine)
- Extreme 8720-32C (leaf/spine)
- Extreme 8820 (spine)

Reference Network

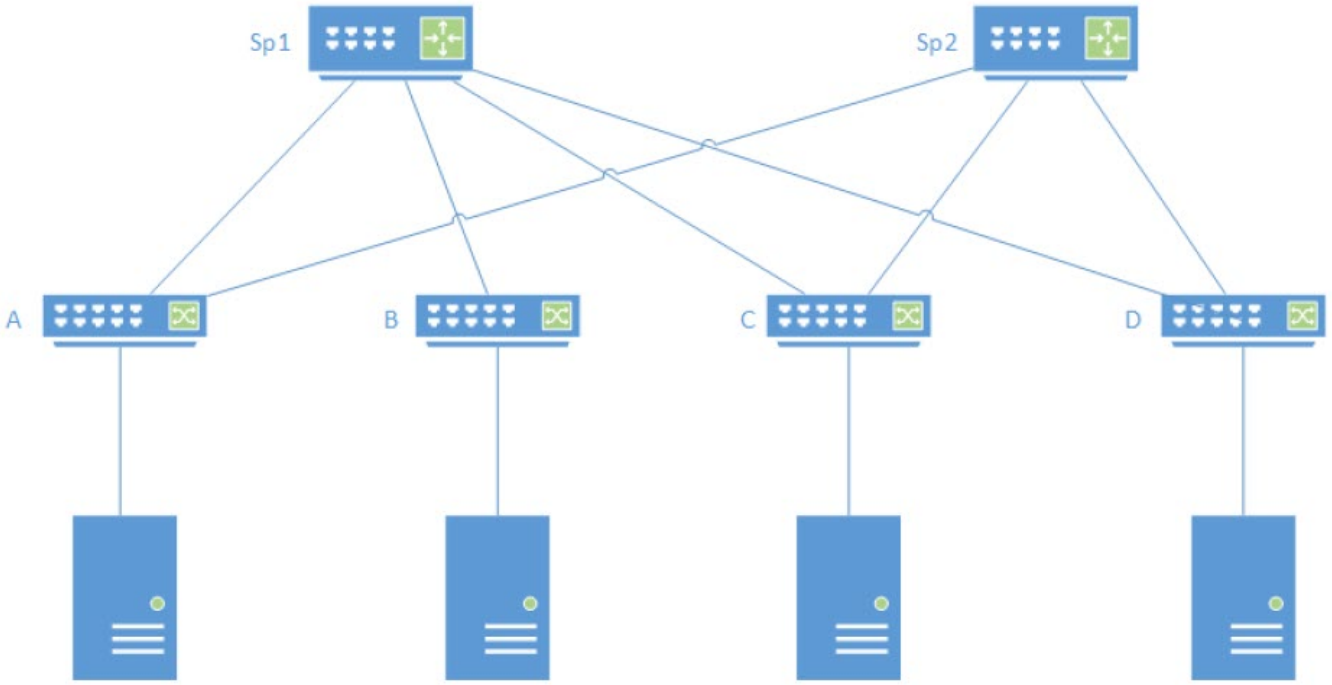


Figure 9: Overview of BGP EVPN Network

A BGP EVPN network is a full mesh of unicast vxLan tunnels between all leaf switches that are part of the same vLAN.

In a BGP EVPN network:

- IP network in the fabric is used as underlay to deliver L2 or L3 traffic between the hosts. VxLAN tunnels are used as overlay for delivering traffic from leaf to leaf with each leaf serving as a VTEP.
- A BGP session is established between each pair of directly connected routers.
- At each node the user configures the VLANs to exchange traffic with the Hosts. Each VLAN is associated with a VNI, the identifier for the vxLAN tunnel in the fabric between leaf switches
- BGP distributes an IMR (Inclusive Multicast Route) which associates the VNI and participating VTEPs. This route information eventually reaches all leaf switches, where the vxLan tunnel to the remote peer is created.

Ingress Replication

In ingress replication, all leaf switches participate in the same flooding domain or VLAN. A flood or multicast packet entering the network at one node is replicated on the tunnels to other leaf switches. A high volume of multicast traffic can congest the network when the same packet is transmitted multiple times on the same link.

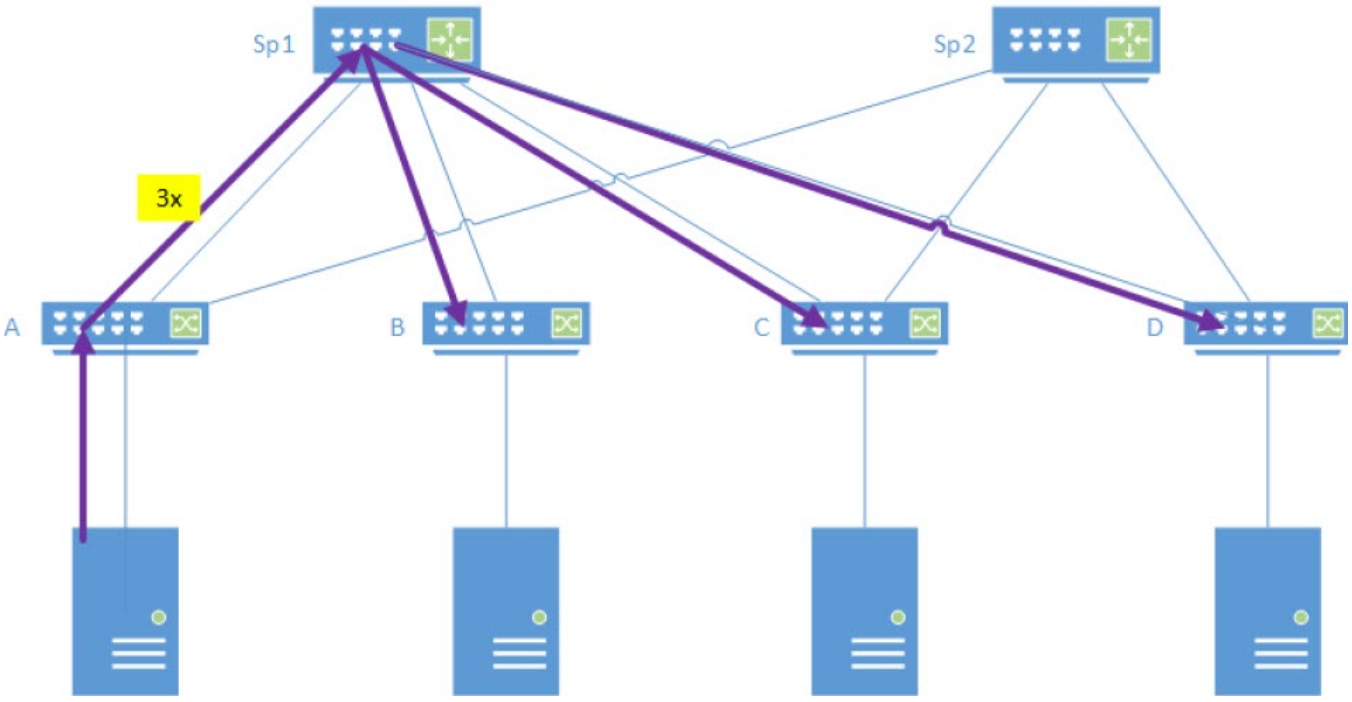


Figure 10: Overview of Ingress Replication

Overview of Multicast Solution

In an ideal multicast solution, the multicast packet is sent only once from a leaf into the spine layer of the network and reduce the replication of flood or multicast packets.

Since the number of spine nodes is less than the number of leaf nodes in a network, the packet replication can be avoided by not sending the same packet multiple times over the same link.

An MDT can be established using BGP and PIM. BGP discovers the VTEPs and PIM establishes the maintains the tree.

There are different types of MDT traffic:

- General or default MDT: The tree is formed from all nodes that participate in VTEP discovery within a domain. Any BUM traffic entering the network is transmitted to all participating leaf nodes and is dropped at egress if it does not apply to the leaf. This may happen if the VNI is not configured at that leaf for user multicast traffic or VLAN does not participate in the c-multicast group.
- VLAN or VNI specific: BUM traffic on the tree is delivered to all members of the VLAN. The user multicast traffic is dropped at egress if the group in the packet is not part of this VLAN at the egress node.
- Group specific: Multicast traffic for a specific group and VLAN can form its own MDT for optimal distribution.

The IP multicast fabric allows user-configured mapping between VNIs and MDTs which supports a default MDT for VLANs or VRFs and VNI specific MDTs, as required. The

multicast vxLan tunnels are used to deliver packets. The format of the IP packet is same as a vxLAN packet:

- Destination IP: Group address of the MDT (instead of the remote VTEP)
- Source IP: Sender (leaf node) IP
- VNI: VNI which represents the VLAN in the network

Multicast Distribution for BUM Traffic

The replication of BUM packet must be controlled and reduced. If there are N VTEPs (neighbors) participating in a VLAN, ingress replication sends N copies of the BUM packet.

Ideal Multicast Distribution

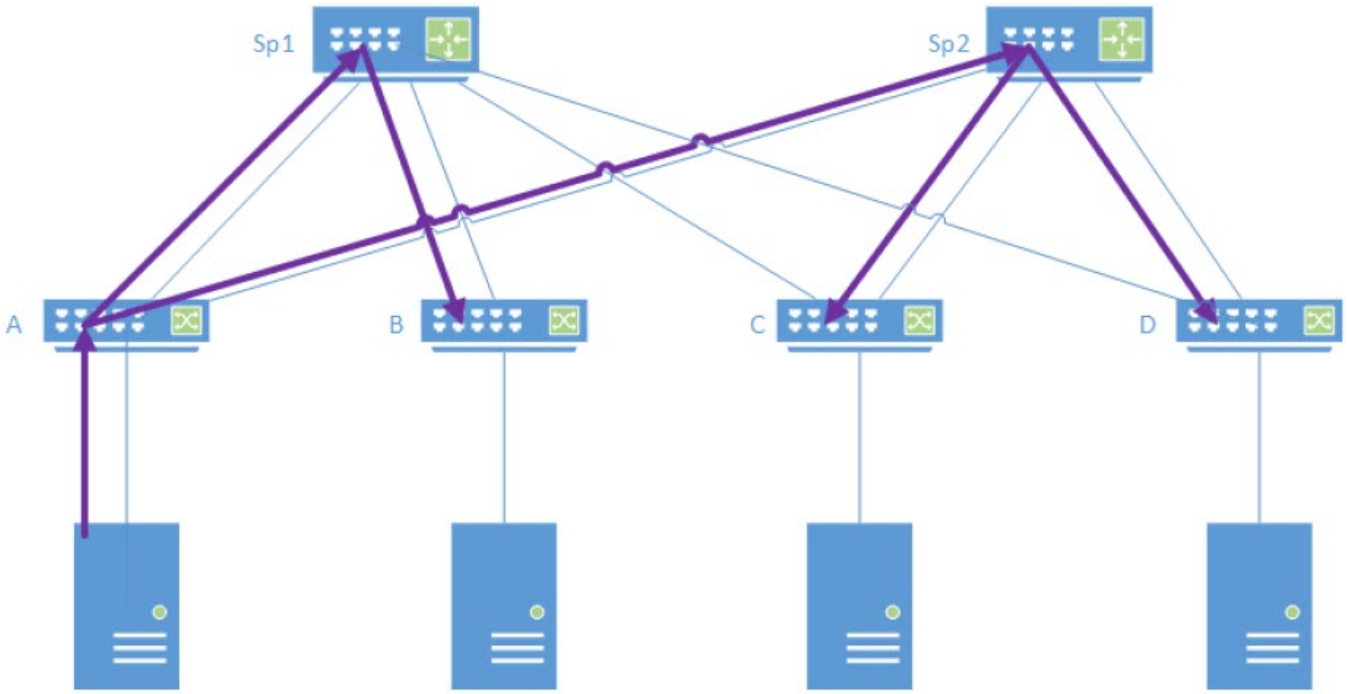


Figure 11: Overview of Ideal Multicast Distribution

BUM traffic at entry node is sent over the vxLAN tunnel for the default MDT, reaching all nodes. The traffic is dropped at egress as required. The topology in which a single spine is utilized is also possible. This depends on where the protocol messages or PIM Join to establish the tree are transmitted.

Unicast VXLAN Tunnel Establishment

VTEP Auto-discovery

MDT in the fabric builds upon the EVPN VTEP auto discovery. After the user configures VLAN to VNI mappings for each leaf, BGP Inclusive Multicast Ethernet Tagged (IMET) routes are initiated and distributed within the fabric. Each node that participates in a

flood domain initiates an IMET or EVPN Type 3 route. Any node that receives an IMET route, establishes a VXLAN tunnel to the remote destination if it belongs to the same domain.

The following figure shows an overview of VTEP Auto-discovery. Node A initiates an IMET route signal in the network. Since nodes B and D are part of the same flooding domain, establish VXLAN tunnels with node A. Similarly, the IMET route signals initiated by nodes B and D form VXLAN tunnels with other leaf nodes.

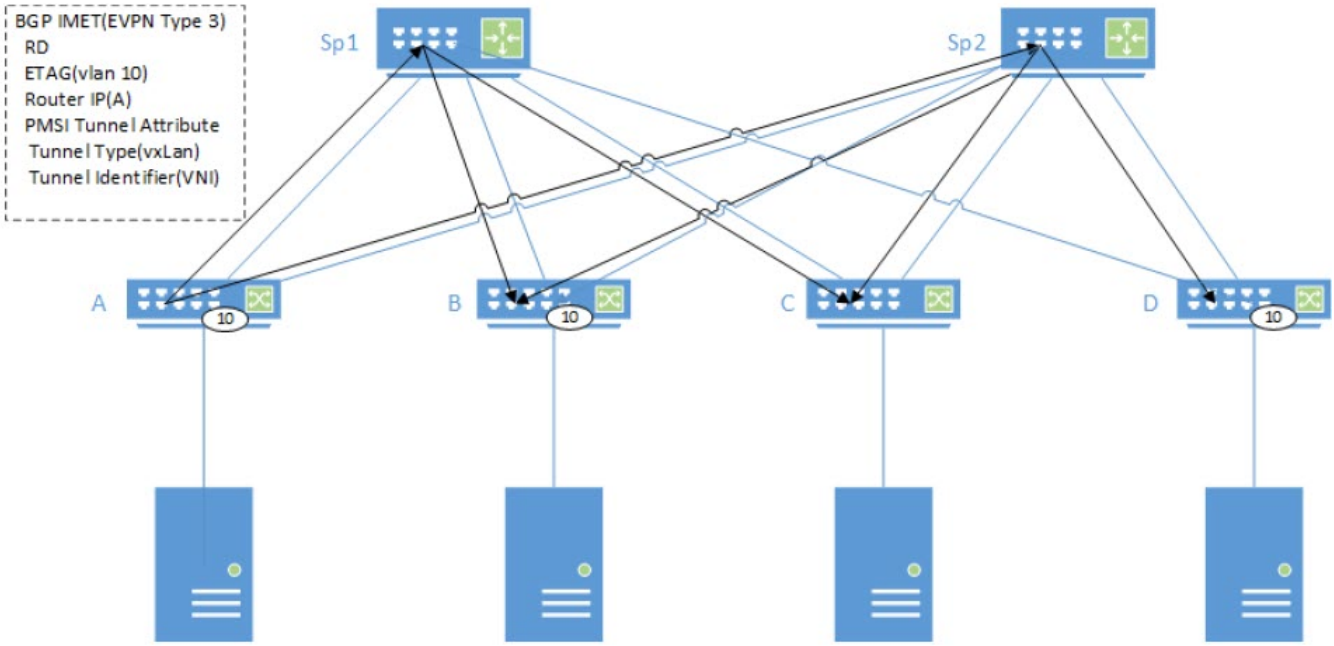


Figure 12: Overview of VTEP auto-discovery

Unicast VXLAN Tunnel Establishment

The following figure shows an overview of unicast VXLAN tunnel establishment. Each leaf creates a tunnel with node A after receiving the IMET route. Eventually, a full mesh of tunnels from each of the nodes participating in VLAN 10, using the same VNI are established.

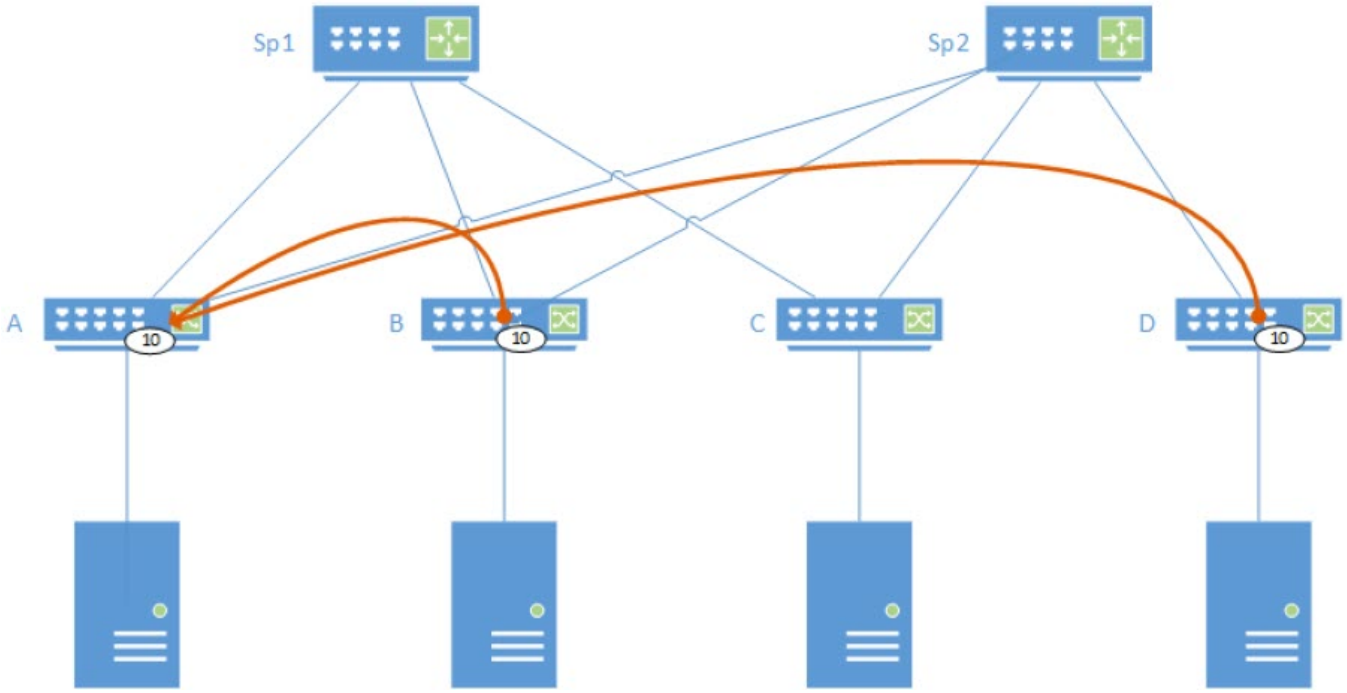


Figure 13: Overview of Unicast VXLAN tunnel establishment

Multicast Distribution Tree

PIM-SSM can create the MDTs using PIM Protocol. After the IMET route is received which identifies the core multicast group for the PIM joins to establish the MDT, PIM signals a join on the RPF path towards the route originator. This results in at least one MDT rooted at each VTEP node.

The number of MDTs is determined by the scope of the traffic inserted into the tree. It is not possible to create an MDT for each C-Multicast group as the number of such groups is virtually unbound. There are several options for creating MDTs:

- One MDT for each VNI
- Single MDT for the entire fabric
- Combination of the above options with additional MDTs for high volume flows

In the PIM-SSM model, a leaf creates a P2MP tree rooted locally to send traffic to all participants. The number of MDTs in the fabric is bound by the number of leaf nodes/ VTEPs.

Separate MDTs such as a Default MDT for flood or BUM traffic and a Data MDT for user multicast traffic can be created. However, a separate MDT just for flood traffic is not required as network congestion is negligible. IP Multicast fabric solution uses the same MDT for both types of traffic.

Per VNI MDT

The following figure shows an example of per VNI MDT signaling. When nodes B, C, and D receive IMET routes identifying A as part of flood domain based on VNI, each node establishes a unicast VXLAN tunnel towards A with the appropriate VNI.

VNIs can be configured to core multicast group address relationship for multicast traffic. In this example, two different group addresses are configured for the two VNIs. PIM creates two source specific trees routed at A, one for each group address.

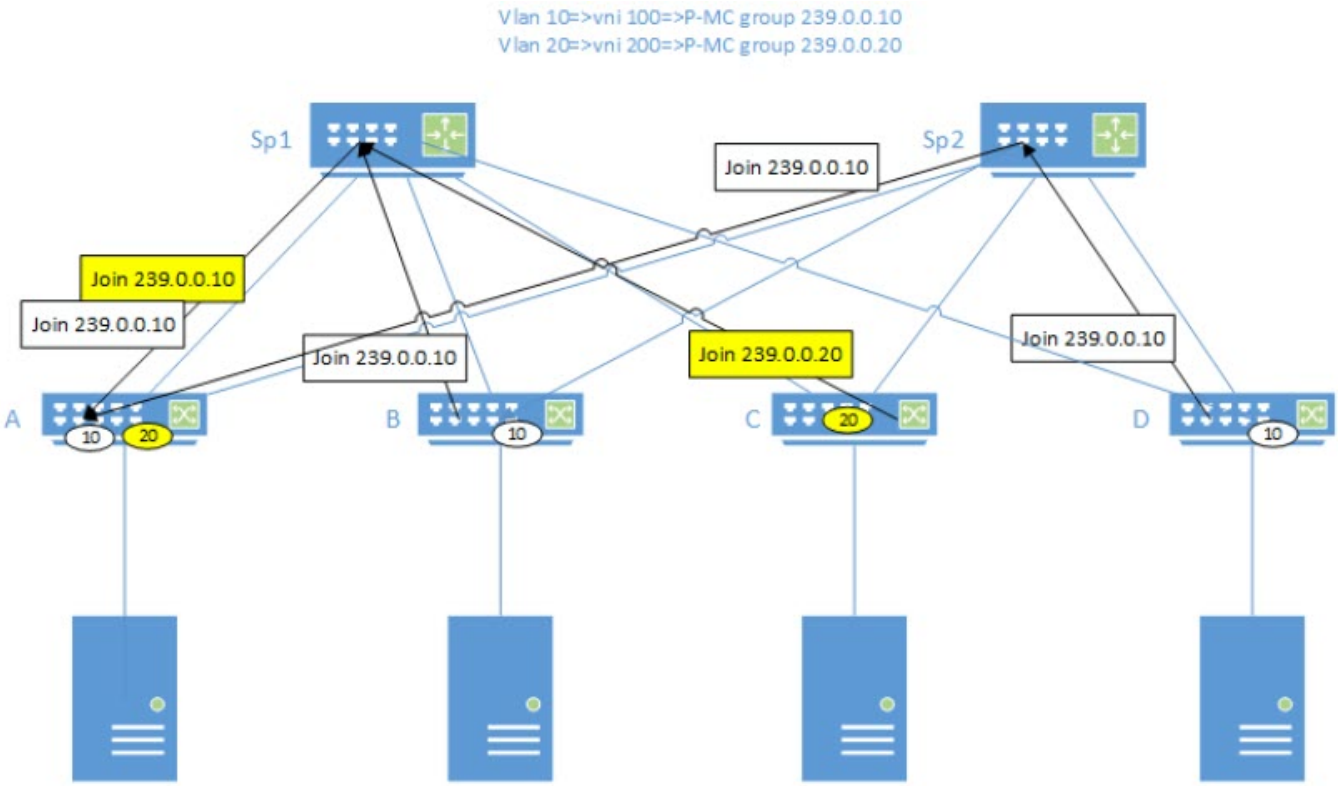


Figure 14: Per VNI MDT signaling

The following figure shows an example of per VNI MDT datapath. The multicast traffic is forwarded within the core based on the MDT group that corresponds to the VLAN and VNI which were signaled with PIM. Traffic for vlan 20 follows a different path than traffic for vlan 10.

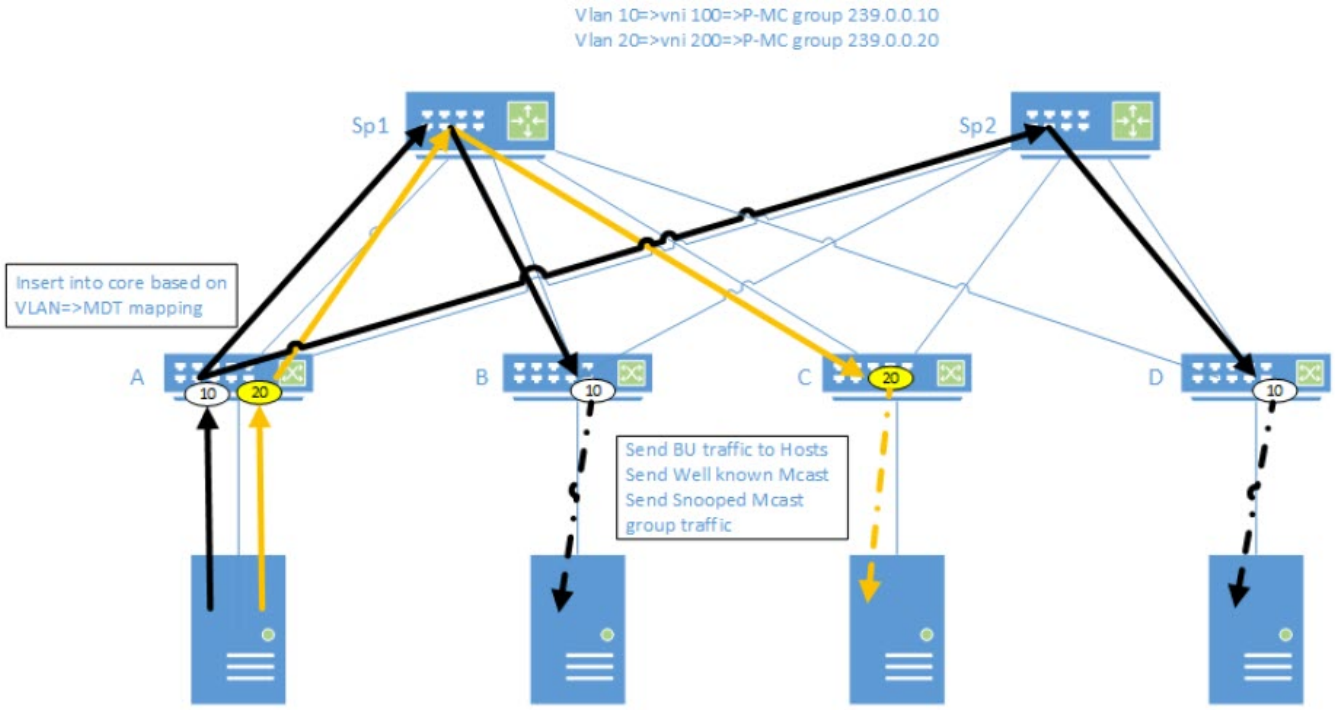


Figure 15: Per VNI MDT Datapath

Default MDT

The following figure shows an example where each VNI in the network is configured with the same multicast group address for the core. A single MDT is formed and rooted at A for all VNIs.

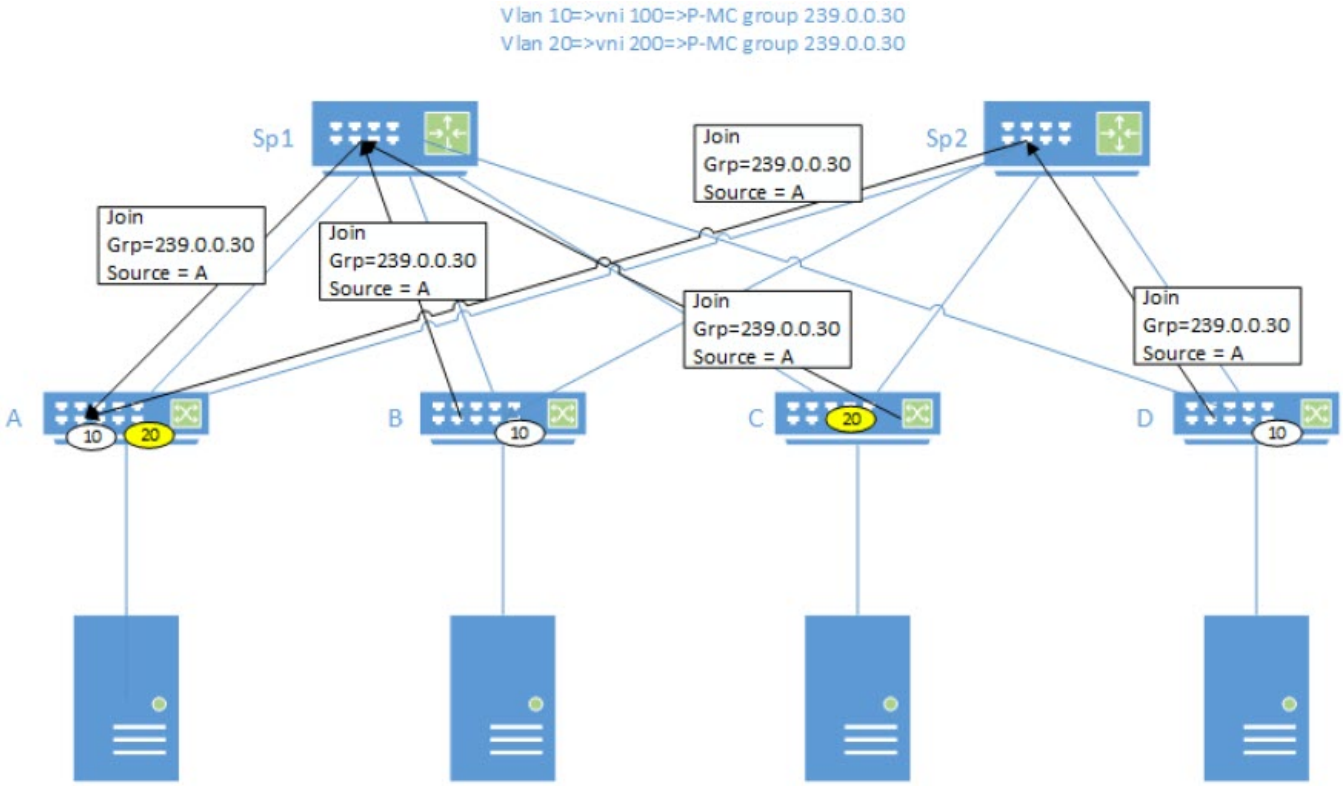


Figure 16: Default MDT signaling

The following figure shows an example where the traffic for both VLANs is injected into the same tree, and filtered at egress as required.

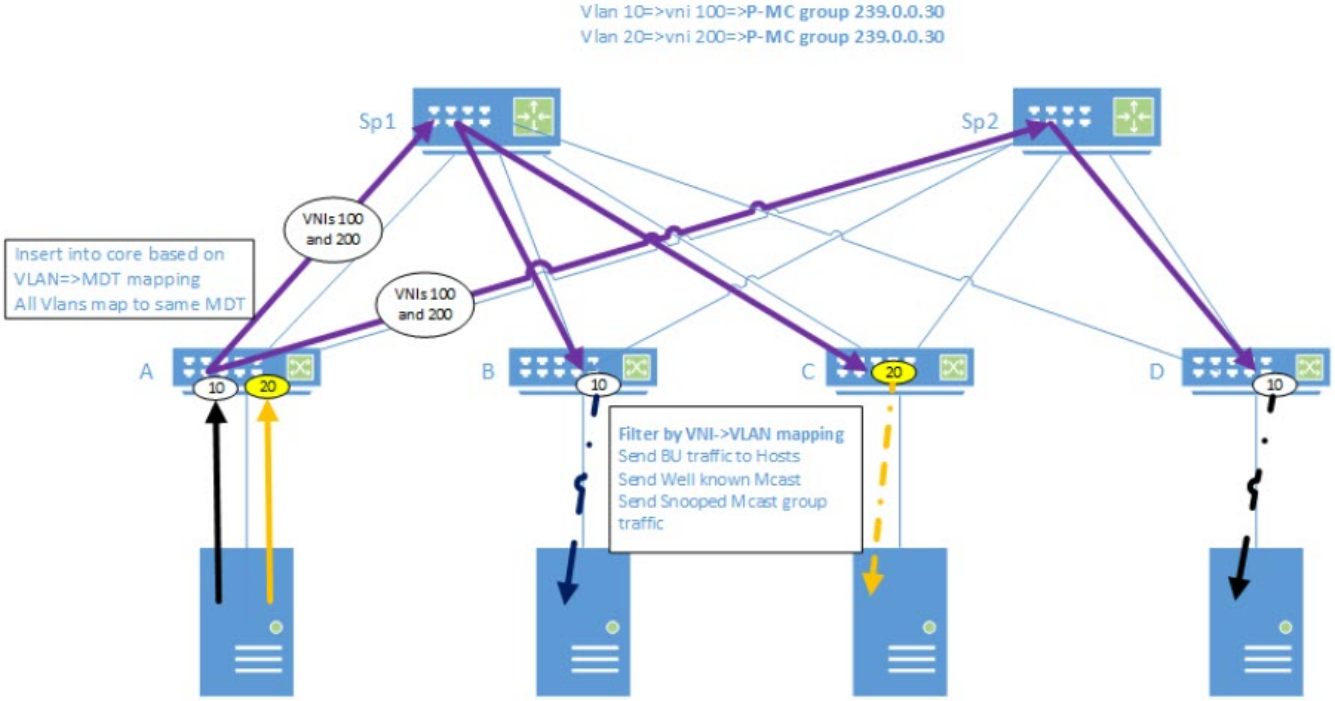


Figure 17: Default MDT datapath

Configure Optimization Replication

1. Configure the IP fabric.
2. On all fabric nodes, configure IP prefix-list for the Multicast Group address range you want to reserve and use with PIM-SSM to form MDT.
3. Enable the following:
   - Router PIM under default VRF
   - PIM-SSM with prefix-list

```
ip prefix-list mdt-range seq 5 permit 239.0.0.0/8
!
router pim vrf default-vrf
ssm-enable range mdt-range
!
```

4. Enable PIM-SM on all Fabric links.

```
interface Ethernet 0/6
ip address 11.1.1.200/24
ip pim-sparse
no shutdown
!
```

5. On all leaf nodes, enable Optimized-Replication under Overlay-Gateway with the default MDT group.

   If a separate MDT is required for an EVPN domain, the MDT group address can be assigned to the associated VLAN/BD.

```
SLX# show running-config overlay-gateway
overlay-gateway tunnel
ip interface Loopback 1
map vni auto
optimized-replication
underlay-mdt-default-group 239.1.1.1
underlay-mdt-group 239.1.1.103 vlan add 300
underlay-mdt-group 239.1.1.110 vlan add 110
underlay-mdt-group 239.1.1.200 bridge-domain add 1
!
activate
!
```

6. (Optional) Remove Optimization Replication.

```
# no optimized-replication
```

```
# vlan 5
# vlan 6
# vlan 10
# vlan 11
# vlan 12
# vlan 20
# vlan 30
# vlan 40
# overlay-gateway 10
# type layer2-extension
# optimized-replication
# underlay-mdt-default-group 239.0.0.100 ! default MDT for all VNIs
# underlay-mdt-group 239.0.0.1 vlan add 10-12,20 ! shared MDT; but not the default
# underlay-mdt-group 239.0.0.3 vlan add 30 ! MDT dedicated to vlan 30
# underlay-mdt-group 239.0.0.4 vlan add 40
# underlay-mdt-group 239.0.0.5 bridge-domain add 50,60-70
```

MDT Scale

MDT scale for IP Fabric depends on total VxLAN scale. Each VxLAN tunnel consumes two types of hardware resource: Tunnel Termination and Encapsulation.

IP Fabric with Optimized Replication also contains unicast VxLAN tunnels. The hardware resources are shared between unicast and Multicast tunnels, with FCFS policy.

The following table shows the total hardware resources.

Table 10: Total hardware resources

		Max limit	
SLX 9150/Extreme 8520 SLX 9250/ Extreme 8720	Tunnel Termination	256	FCFS (Unicast + Multicast Tunnels)
	Encapsulation	16k	15.5k LIF + 512 FCFS Unicast/Multicast
SLX 9540/SLX 9640	Tunnel Termination	512	FCFS (Unicast + Multicast VxLAN)
	Encapsulation	96k	FCFS (LIF + L3 + VxLAN)

Total unicast VxLAN tunnels in a Fabric depends on the number leaf nodes. Each leaf node has unicast VxLAN tunnels. Each unicast VxLAN tunnel consumes 1 Termination and 2 Encap entries. Depending on size of the fabric, MDT can scale based on the remaining hardware resources.

The following example shows how an MDT scales when there are no unicast tunnels and all resources are available for multicast tunnels.

1 MDT = 1 Multicast VxLAN Tunnel = (1 Termination) + (1 Encap per Spine)

SLX 9250:

- 256 MDT with 1 Spine = 256 Tunnels x 1 Encap (Max 256 Termination limit reached)
- 128 MDT with 4 Spines = 128 Tunnels x 4 Encap (Max 512 Encap limit reached)
- 1 MDT with 512 Spines = 1 Tunnel x 512 Encap (Max 512 Encap limit reached)

Configure Multicast IP Fabric with L3 VNI

- Each L3 VNI is treated as a separate VNI that needs to be mapped to MDT.
- For deployments where user VNIs are not configured on all leaf nodes, L3 VNI serves as an Intermediate Hop to forward the traffic.
- There is no change in the Unicast traffic forwarding behavior as Unicast traffic does not use MDT.
- Each L3 VNI must be configured and mapped to MDT Group. It can have a unique MDT or a shared VNI MDT.

To support IP Multicast routing using L3 VNI, each L3 VNI must be configured as PIM interface. This enables the PIM protocol to form Multicast forwarding tree over L3 VNIs.

1. Configure the following on the device:
   - VRF
   - Route target for EVPN
   - Interface for L3 VNI

   Note
   L3 VNI interface must be configured with IP address.
   On LVTEP, the VLAN associated with L3VNI must also be added as MCT VLAN member.

```
vrf mc
rd 102:3
evpn irb ve 300
address-family ipv4 unicast
route-target export 102:102 evpn
route-target import 102:102 evpn
!
!
```

2. For L3 multicast routing, enable Router PIM for the selected VRF domain on each leaf node, as required.

```
router pim vrf mc
!
```

3. Select a leaf node to act as RP per VRF Domain and create a loopback interface on the node.
4. Enable PIM-SM for the loopback interface.

```
interface Loopback 2
vrf forwarding mc
ip address 30.30.30.30/32
ip pim-sparse
no shutdown
!
```

5. Configure IP address of the selected RP interface as RP-address on each leaf node where PIM is running for the selected VRF.

```
router pim vrf mc
rp-address 30.30.30.30
!
```

6. Enable the following on each leaf node as required.
   - PIM on L3 VNI interface.
   - For L3 Multicast Routing, PIM on all selected VE interfaces associated with EVPN VLANs/BDs.

```
interface Ve 300
vrf forwarding mc
ip address 3.4.5.3/24
ip pim-sparse
no shutdown
!
```

```
vrf mc
rd 102:3
```

```
evpn irb ve 300
address-family ipv4 unicast
route-target export 102:102 evpn
route-target import 102:102 evpn
!
address-family ipv6 unicast
route-target export 102:102 evpn
route-target import 102:102 evpn
!
!
vlan 300
router-interface Ve 300
!
interface Ve 300
vrf forwarding mc
ip address 3.4.5.3/24
ip pim-sparse
no shutdown
!
overlay-gateway 10
type layer2-extension
optimized-replication
underlay-mdt-group 239.0.0.100 default ! default MDT for all VNIs
underlay-mdt-group 239.0.0.1 vlan add 300
```

Logical VTEPs and MCT (Multi-homing)

The following sections provide information on logical VTEPs and MCT (multi-homing).

MDT Signaling for LVTEP/MCT Cluster

The following figure shows an example of multi-homed configuration. The server attached to A supports a LAG connection to A/A' which forms an MCT cluster. The join for the core MDT coming from the spines are directed at the logical VTEP address for the A/A' cluster, so that any particular join from a spine can land at either switch. Also, when LVTEP/MCT cluster joins the IP Fabric MDT, the PIM-SSM Join is initiated only by the MCT Peer node which is elected as Primary MDT node.

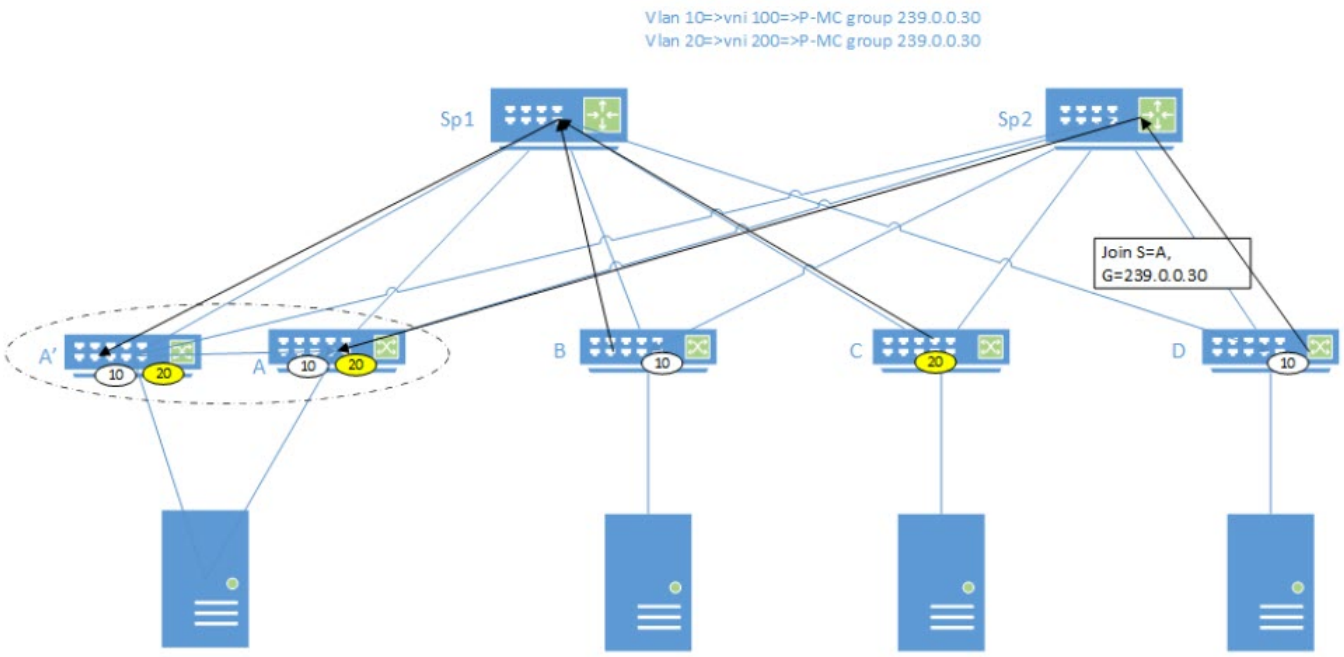


Figure 18: MDT signaling for LVTEP/MCT cluster

Primary MDT Node Selection

The Primary MDT node is selected in MCT cluster to determine which MCT peer must initiate the PIM-SSM Join to all discovered remote VTEPs and join the IP Fabric MDT to receive the traffic. The selection is based on MCT keep-alive (KA) client isolation role configuration and local and peer node status.

	MCT KA Role Configured/Selected	
	Primary	Secondary
Local node status	Up	Shut
Peer node status	Up	Down

Note

The MCT KA role is either configured or selected. The MCT node with highest IP address is the Primary node.

The local node status is based on shutdown status in cluster configuration.

The peer node status is based on Keep-Alive status.

If both MCT peer nodes have the MCT KA role configured as Primary, MDT join role is also selected as Primary in both the nodes and both MCT peer nodes receive MDT traffic.

LVTEP/MCT Cluster Link Failure

On MCT ICL link fail, Loose-Loose mode is activated. MCT peers continue with their existing MCT KA and MDT join Primary/Secondary states and only the Primary MDT node receives the MDT traffic. This results in the MDT traffic to be forwarded to CCEP clients based on Local-Biased forwarding. The CEP ports on Secondary MDT node do not receive the traffic.

MCT peer node with KA role as Secondary, shuts down its CCEP interfaces and the LVTEP loopback interface. The spine nodes that join the Secondary MCT node initially, re-direct the PIM-SSM joins to the Primary MCT node. All CCEP MDT traffic egress out into the EVPN Fabric from the Primary MCT node. The traffic from the CEP port on Secondary MCT node cannot egress out into the EVPN Fabric and is dropped.

Route Sync Between LVTEP/MCT Cluster

Any IGMP Group discovered over the IP Fabric Multicast Tunnel is termed as IGMP EVPN route. For all remotely discovered IGMP EVPN routes, Multicast tunnel is added as OIF to the L2/L3 Mcast routes.

The following figure shows an example of LVTEP/MCT cluster. When IGMP EVPN route is discovered by node A, Join Sync Route (EVPN Type 7) is sent between nodes A and A'. The purpose of the communication is to ensure that group traffic is directed to all branches of the tree.

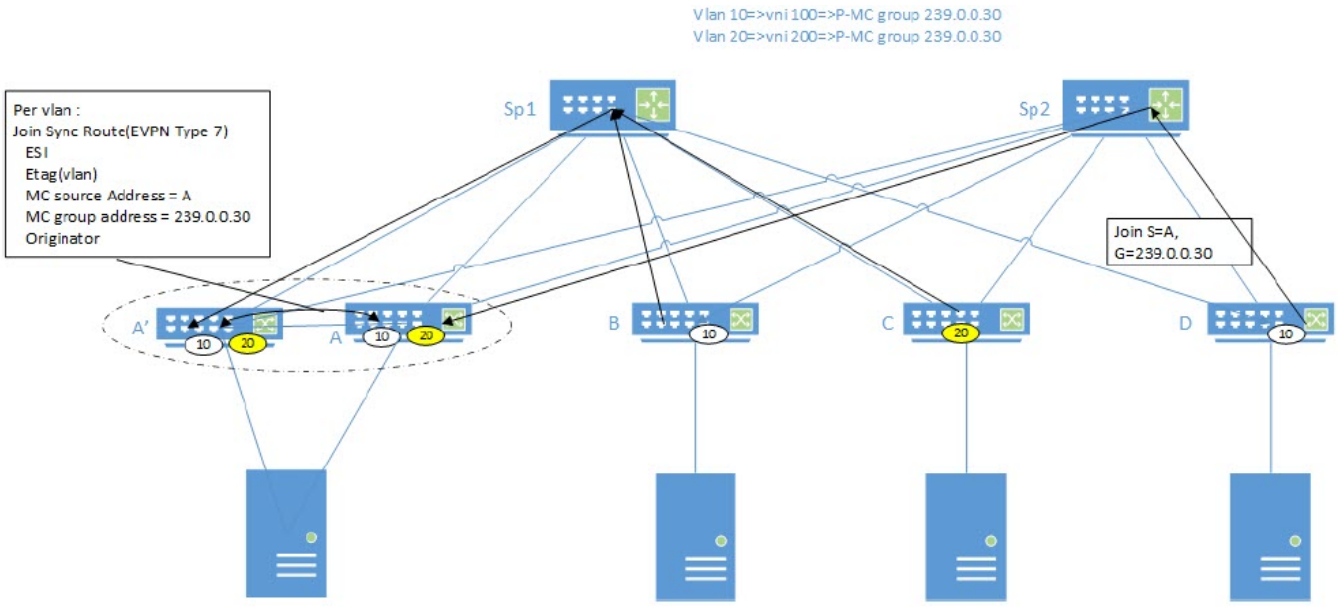


Figure 19: Overview of Route sync between LVTP/MCT cluster

Datapath

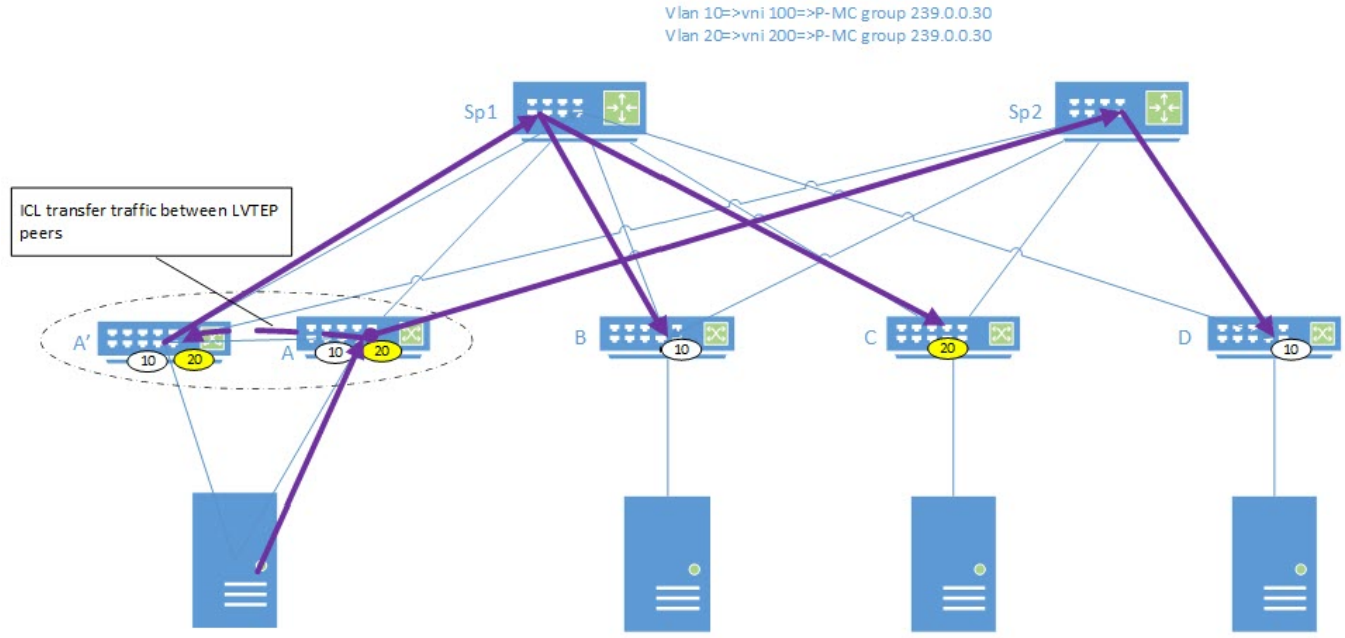


Figure 20: Datapath

The following figure shows an example of the datapath. Node A has a sync route from node A' and has added the ICL link to the OIF list for the traffic. When multicast traffic reaches node A, the traffic is sent not just to the multicast VxLAN tunnel which sends the traffic to Sp2, but also to the ICL connecting the nodes A and A'. Node A' picks up the traffic and directs it into its multicast VxLAN tunnel to deliver it to nodes B and C through Sp1. Traffic delivery to the servers connected to nodes B and D depends on the presence of VNI and C-multicast group on those nodes or routers.

Bud Node Topology

In some cases, traffic on the MDT must be terminated / decapsulated and then forwarded as part of the MDT. This can happen easily in a non-CLOS architecture, especially in a ring topology.

The following figure shows an example of a bud node in CLOS. Initially, the links between Sp2 and A/A' and Sp1 to D are down. The tree that is formed includes router C (leaf node) in the path from A to D. The CLOS devolves to a straight-line network between A --> Sp1 --> C --> Sp2 --> D.

In this case, it is not enough for C to decapsulate the traffic from the VxLAN and forward to the servers and its multicast VxLAN tunnel which includes Sp2 as an immediate destination. Hence, the packets are recycled by processing in two passes. The first pass includes forwarding the packet along any downstream links that are part of the MDT. The second pass includes decapsulating the packet and sending it to the local ports. This requires internal or external ports to be reserved for packet recycling depending on the specific platform.

Router C detects this scenario once it has received a Join for a source that is not itself. For example, receiving a (S, G) join where `S ! = C` and `G = the group address for the vxLan MDT` is enough for C to detect that bud node behavior is required.

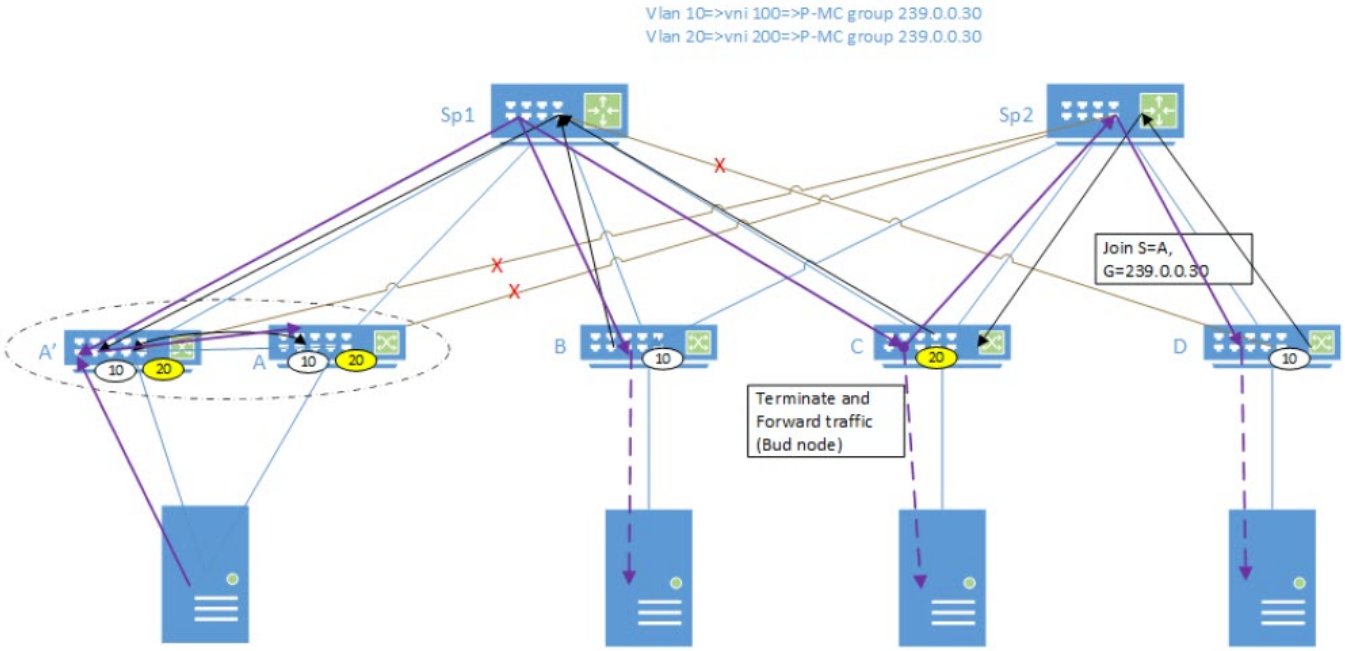
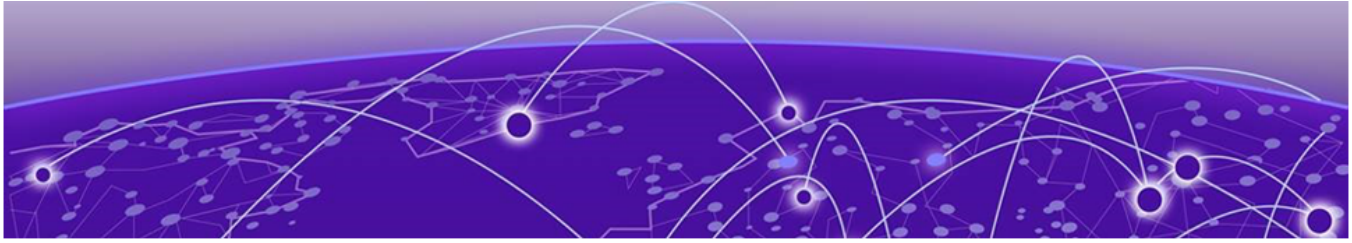


Figure 21: Bud node topology

Any problems in recycling due to recycle port configuration and recycle port limits can be addressed with the configuration model and allowing multiple recycle ports or higher bandwidth recycle ports.



IPv6 Multicast Routing

IPv6 Multicast over MCT on page 72
IPv6 Protocol Independent Multicast on page 74

IPv6 Multicast over MCT

Multicast state information is synchronized between the MCT peers using MP-BGP EVPN transport, specifically the Join Sync Route and the Leave Sync Route. Multicast protocol packets are sent on the peer link only when required.

MLD (Multicast Listener Discovery) protocol packets are of three types.

- **Multicast listener query**: General query and Group-specific query
- **Multicast listener reports**: Version 1 and Version 2
- **Multicast listener done**: Version 1 and Version 2

MLD query packet processing

Each EVI is associated with a multicast group ID (MGID) that is BUM-suppressed (Broadcast, unknown unicast and multicast). However, query packets need to be transmitted on an interchassis link (ICL) to address the following scenarios.

- The querier connected to only one of the MCT peer switches is the elected querier.
- Only one of the peer switches is configured as a querier.
- The switch ages out MLD routes if memberships are not confirmed in the timeout interval. Although query packets are received on the MCT peer link, the mrouter port is not learned or considered on that peer link.

MLD membership reports

- Traditionally, each peer switch learns about Layer 2 multicast memberships by snooping the MLD membership reports. The membership reports are then flooded on multicast router (mrouter) ports.
- For MCT, because an mrouter port is not learned on the peer link, membership reports are not flooded between the peer switches. Peer switches exchange learned routes by using EVPN NLRI messages between Border Gateway Protocol (BGP) peers running on the MCT cluster control VLAN.

- The BGP on the peer switch communicates with the MLD routes to the multicast module, which adds and deletes group memberships. The module also allocates or rejects the MGIDs and installs or uninstalls the routes in hardware.
- If a general query or group-specific query is received from any port other than a peer link, each peer switch generates a proxy report for the MLD routes learned from EVPN NLRI.

Duplicate MLD query packets on CCEP

If a query is configured on both MCT peers of a member VLAN, duplicate query packets reach clients that are connected to the MCT domain by means of a Cluster Client Edge Port (CCEP).

Multicast modules use inclusive multicast routes of EVPN NLRI to exchange the multicast querier configuration on the member VLANs. The MCT peer that is the DF of the member VLAN is elected as the querier.

Multicast modules handle querier election for the member VLAN when peer links are active or inactive.

MLD listener done

When fast-leave is not configured and an MCT peer receives MLD Listener Done from one of its clients for group G, the switch or router informs other MCT peers about the group-specific query and latency by using the Leave Sync Route. The peer switch, which runs the querier, sends group-specific queries and group queries.

Mrouter synchronization

Mrouter synchronization helps in achieving optimal path selection for unknown multicast traffic and optimal MP-BGP message exchange between MCT peers. The node that is elected multicast DF (designated forwarder) for the VLAN forwards data traffic on CCEP as an mrouter port.

Scalability

Scalability feature	Maximum
MLD snooping-enabled VLANs or bridge domains over MCT	256
IPv6 PIM-enabled VEs over MCT	32
MLD groups over MCT	4000
IPv6 PIM cache routes over MCT	2000

Device support

Extreme Networks supports IPv6 multicast over MCT on the following devices.

- ExtremeSwitching SLX 9540
- ExtremeRouting SLX 9640

IPv6 Protocol Independent Multicast

IP multicast transmits a data stream to multiple hosts simultaneously. Protocol-Independent Multicast (PIM) is one of several protocols designed for IP multicast. PIM does not rely on a specific routing protocol to create its network topology state, but uses routing information from other protocols.

PIM messages are sent encapsulated in an IP packet where the IP protocol is 103 and the version is 6 for IPv6. Depending on the type of message, packets are sent to the PIM All-Router-Multicast address (`FF02::D`) or sent as unicast to a specific host.

Extreme Networks supports IPv6 PIM in the IPv6 multicast TCAM hardware profile.

Extreme Networks supports PIM-SM (Sparse Mode) and PIM-SSM (Source Specific Multicast) on the following devices.

- ExtremeSwitching SLX 9540
- ExtremeRouting SLX 9640

IPv6 PIM-SM

PIM-SM (Sparse Mode) is most effective in large networks in which few hosts receive multicast traffic. PIM-SM devices are organized into domains. A PIM-SM domain is a contiguous set of devices that all implement PIM and are configured to operate in a common boundary.

In PIM-SM, unidirectional, shared distribution trees are rooted at a common node in the network called the rendezvous point (RP). The RP acts as the messenger between the source and the destination hosts or routers. An RP is configured statically per PIM router or by means of a bootstrap router (BSR). The RP must always be upstream from the destination hosts or routers.

Hosts and routers send Join messages to the RP for a group. To reduce the number of Join messages incoming to an RP, the local network selects one of its upstream routers as the designated router (DR). All hosts below a DR send Join messages to the DR. The DR sends only one Join message to the RP on behalf of its hosts.

In IPv6 PIM-SM, you can create a source-based distribution tree, which is rooted at a router that is adjacent to the source. With this option, destination hosts can switch from the shared tree to the source-based tree if the latter has a shorter path between the source and the destination.

IPv6 PIM-SSM

In PIM Source-Specific Multicast (PIM-SSM), which uses a subset of PIM-SM and Multicast Listener Discovery (MLD) v2, a hosts receives multicast traffic directly from the source.

PIM-SSM uses PIM-SM functionality to create a Shortest Path Tree (SPT) between the receiver and the source without the help of a rendezvous point (RP).

In a PIM-SSM network, a host subscribes to an SSM channel (by means of MLDv2) and sends a request to join group G and source S. The directly connected PIM-SM router, the receiver's designated router (DR), sends an (S,G) Join message to its reverse path forwarding (RPF) neighbor for the source. The receiver doesn't contact the RP in this process, unlike standard PIM-SM operations. The (S,G) Join message initiates the source tree and then builds it, hop by hop, toward the source. Multicast traffic is delivered to the subscribing host by way of the source tree.

The default IPv6 SSM group range is `FF3x::/32`, where *x* is one of the following valid values: `4, 5, 8, E`.

Enable IPv6 PIM Globally

Use the **ipv6 pim router pim** command to enable IPv6 PIM routing and enter IPv6 PIM router configuration mode. Extreme Networks supports IPv6 PIM in the IPv6 multicast TCAM hardware profile.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable IPv6 PIM and enter router configuration mode for default virtual routing and forwarding (VRF).

```
device(config)# ipv6 router pim
device(config-ipv6-router-pim-vrf-default-vrf)#
```



Tip
To enable IPv6 PIM for a specific VRF, use the following form of the command: **ipv6 router pim** vrf *VRF-name*.

Configure Options on an IPv6 PIM Router

When IPv6 PIM is enabled globally, you can configure several PIM options for the router.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter router configuration mode for the default VRF.

```
device(config)# ipv6 router pim
device(config-ipv6-router-pim-vrf-default-vrf)#
```

3. Run the following commands as needed.

   For more information, including more examples, see the *Extreme SLX-OS Command Reference*.

   a. Configure a device interface as a rendezvous point (RP).

   ```
   rp-address 1234::22
   ```

   b. Configure PIM Anycast RPs in multicast domains.

   ```
   anycast-rp 1234::100 anycast-rp-peer-prefix-list
   ```

   c. Configure a bootstrap router as a candidate RP to distribute RP information to other PIM-SM devices.

   ```
   bsr-candidate interface ve 30 mask 126 priority 10
   ```

   d. Configure a device as a candidate RP.

   ```
   rp-candidate interface loopback 1
   ```

   e. Enable Source-Specific Multicast (PIM-SSM) mode.

   ```
   ssm-enable
   ```

   f. Configure the maximum size of the multicast cache.

   ```
   max-mcache 500
   ```

   g. Configure the frequency with which a device sends PIM hello messages to its neighbors.

   ```
   hello-interval 50
   ```

   h. Configure the length of time a PIM device waits for hello messages before considering a neighbor to be absent.

   ```
   nbr-timeout 600
   ```

   i. Configure the length of time a forwarding entry can remain unused before a device deletes it.

   ```
   inactivity-timer 90
   ```

   j. Configure the frequency with which PIM Join and Prune messages are sent.

   ```
   message-interval 3600
   ```

   k. Configure the length of time a PIM device waits to stop traffic after receiving a Prune message.

   ```
   prune-wait 10
   ```

   l. Configure sources to use the shared RP tree instead of the Shortest Path Tree.

   ```
   spt-threshold infinity
   ```

   m. Define the order in which unicast routes are selected from the routing tables during Reverse Path Forwarding (RPF) resolution.

   ```
   route-precedence none priority-3 uc-non-default priority-2 uc-default priority-1
   ```

   n. Enable multicast ECMP load sharing with dynamic rebalancing.

   ```
   rpf ecmp rebalance
   ```

Configure Options on an IPv6 PIM Interface

When IPv6 PIM is enabled globally, you can configure several PIM options for the interface.

Note
Perform all Layer 3 configuration for VLANs, bridge domains, and link aggregation groups under their respective Layer 3-level interfaces.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter device configuration mode.

```
device(config)# interface ve 30
```

3. Run the following commands as needed.

   For more information, including examples, see the *Extreme SLX-OS Command Reference*.

   a. Enable IPv6 PIM on an interface.

```
device(config-if-ve-30)# ipv6 pim-sparse
```

   b. Specify the designated router (DR) priority of an IPv6 PIM interface.

```
device(config-if-ve-30)# ipv6 pim dr-priority 100
```

   c. Configure the Time to Live (TTL) threshold for an IPv6 PIM interface.

```
device(config-if-ve-30)# ipv6 pim ttl-threshold 64
```

Display IPv6 PIM Information

You can use show commands to display information about the internal state of IPv6 PIM. You can run the commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

1. Display the mapping of group addresses to Rendezvous Point (RP) addresses.

```
device# show ipv6 pim rp-map
Number of group-to-RP mappings: 3

     Group address       RP address
----------------------------------------
1    ff04:3000:4000::2   1234::1234
2    ff08:1001:2001::1   1234::5
3    ff0e:1000:2000::1   1234::32
```

2. Display the RPs for a PIM-SM (Sparse Mode) group.

```
device# show ipv6 pim rp-hash ff0e:1000:2000::1
RP: 1234::32, v2
    Info source: 1234::5, via bootstrap
```

3. Display the list of RPs that the switch has learned.

```
device# show ipv6 pim rp-set
Vrf Instance : default-vrf
--------------------------
Static RP
---------
 Static RP count: 1
   RP: 1234::1234
```

```
Number of group prefixes Learnt from BSR: 3
Group prefix = ff05::      # RPs: 1
      RP 1: 1234::5    priority=0    age=30    holdtime=150
Group prefix = ff08::      # RPs: 1
      RP 1: 1234::5    priority=0    age=30    holdtime=150
Group prefix = ff0e::      # RPs: 1
      RP 1: 1234::32   priority=0    age=40    holdtime=150
```

4. Display the Reverse Path Forwarding (RPF) for an address.

```
device# show ipv6 pim rpf 1234:32
upstream nbr fe80::2e0:cff:fe03:a on Eth 0/8
```

5. Display a list of PIM multicast groups.

```
device# show ipv6 pim group
Total number of groups: 3
1    Group ff04:3000:4000::2
     Group member at  Ve100: Ve100
2    Group ff08:1001:2001::1
     Group member at  Ve200: Ve200
3    Group ff0e:1000:2000::1
     Group member at  eth 0/1: eth 0/1
```

6. Display information about the bootstrap router (BSR) and the candidate RP.

```
device# show ipv6 pim bsr
PIMv2 Bootstrap information :
-------------------------------------------------------------------
  This system is the Elected BSR
  BSR address: 1234::5. Hash Mask Length 32. Priority 0.
  Next bootstrap message in 00:00:50
  Configuration:
    Candidate Lo 1 (Address 1234::5). Hash Mask Length 32. Priority 0.
Next Candidate-RP-advertisment in 00:00:50
  RP: 1234::5
    group prefixes:
    ff08:: / 16
    ff05:: / 16
  Candidate-RP-advertisement period: 60
```

7. Display information about the candidate RP and its group mappings.

```
device# show ipv6 pim rp-candidate
Next Candidate-RP-advertisment in 00:00:10
  RP: 1234::5
    group prefixes:
    ff08:: / 16
    ff05:: / 16
  Candidate-RP-advertisement period: 60
```

8. Display information about the Anycast RP.

```
device# show ipv6 pim anycast-rp
Number of Anycast RP: 2
Anycast RP :2003:1::1
Prefix Name :AnyCastRPList-1
Peer Addresses :
2002:1:2::1
2002:1:5::1
2002:1:8::1
Anycast RP :2003:2::1
Prefix Name :AnyCastRPList-2
Peer Addresses :
2001:2:5:25::5
2001:2:8:28::8
2001:2:8:28::2
2001:5:8:58::5
```

```
2001:2:5:25::2
2001:5:8:58::8
```

9. Display information about active PIM neighbors.

```
device# show ipv6 pim neighbor
Total Number of Neighbors : 4
Port         Phy_Port     Neighbor                       Holdtime Age
UpTime         Priority
                                                         sec      sec      Dd HH:MM:SS
Eth 0/6      Eth 0/6      fe80::da84:66ff:feea:3b1f      105      16
01:32:17         1
                         +2001:6:8:68::6
Ve20         Ve20         fe80::7aa6:e1ff:fe45:9815      105      19
01:42:21         1
                         +2001:5:8:58::5
Ve40         Ve40         fe80::748e:f8ff:fe05:7801      105      10       7d
07:10:24         1
                         +2001:2:8:28::2
Ve41         Ve41         fe80::748e:f8ff:fe05:7801      105      10       7d
07:10:23         1
                         +2001:8:2:82::2
```

10. Display the IPv6 PIM status and configuration information for an interface.

```
device# show ipv6 pim interface
-------------+--------------------------------+----+-----
+-------------------------------------------+---+--------|--------------------
Interface    |Local                           |Ver
|Mode |Designated                    Router       |TTL| DR     |        Vrf
             |Address                         |
|     |Address                       Port         |Thr| Prio   |        Name
-------------+--------------------------------+----+-----
+-------------------------------------------+---+--------|--------------------
Eth 0/6       fe80::629c:9fff:fe5a:ee1c
v2    SM    fe80::da84:66ff:feea:3b1f     None          1   1       default-vrf
Eth 0/47      fe80::629c:9fff:fe5a:ee45
v2    SM    Itself                        None          1   1       default-vrf
Ve20          fe80::629c:9fff:fe5a:ee15      v2
SM    fe80::7aa6:e1ff:fe45:9815     None          1   1       default-vrf
Ve40          fe80::629c:9fff:fe5a:ee15      v2
SM    fe80::748e:f8ff:fe05:7801     None          1   1       default-vrf
Ve41          fe80::629c:9fff:fe5a:ee15      v2
SM    fe80::748e:f8ff:fe05:7801     None          1   1       default-vrf
Ve700         fe80::629c:9fff:fe5a:ee15      v2
SM    Itself                        None          1   1       default-vrf
```

11. Display route entries in the multicast cache.

```
device# show ipv6 pim mcache
IP Multicast Mcache Table
Entry Flags    : sm  - Sparse Mode, ssm - Source Specific Multicast
                 RPT - RPT Bit, SPT - SPT Bit, LSrc - Local Source
                 LRcv - Local Receiver, RegProbe - Register In Progress
                 RegSupp - Register Suppression Timer, Reg - Register Complete
                 needRte - Route Required for Src/RP, JDUp - Join Desire Upstream
Interface Flags: IM - Immediate, IH - Inherited, WA - Won Assert
                 MJ - Membership Join, MI - Membership Include, ME - Membership Exclude
                 BR - Blocked RPT, BA - Blocked Assert, BF - Blocked Filter
                 BI - Blocked IIF, UJ - Unintended Join oif
Total entries in mcache: 2
1    (*, ff0e:1000:2000::1) RP 1234::32 in Eth 0/8, Uptime 00:16:18
     Sparse Mode, RPT=1 SPT=0 Reg=0 RegSupp=0 RegProbe=0 JDUp=1 LSrc=0 LRcv=1
     upstream neighbor=fe80::2e0:cff:fe03:a (2008::32)
     AgeSltMsk: 0     num_oifs = 1
           Ve100(00:16:18/0) Flags: MJ
     Flags (0x01260480)
```

```
        sm=1 needRte=0
2    (1010::100, ff0e:1000:2000::1) in Eth 0/8, Uptime 00:09:27
     Sparse Mode, RPT=0 SPT=1 Reg=0 RegSupp=0 RegProbe=0 JDUp=1 LSrc=0 LRcv=1
     upstream neighbor=fe80::2e0:cff:fe03:a (2008::32)
     AgeSltMsk: 1 KAT timer value: 240
     num_oifs = 1
           Ve100(00:09:27/0) Flags: MJ
     Flags (0x510680f4)
         sm=1 needRte=0
```

12. Display traffic statistics for IPv6 PIM-enabled interfaces.

```
device# show ipv6 pim traffic
Port        |HELLO              |JOIN               |PRUNE              |
ASSERT             |GRAFT/REGISTER     |REGISTER-STOP      |BSR-MSGS           |RPC-
MSGS
            |Rx        Tx       |Rx        Tx       |Rx        Tx       |Rx
Tx       |Rx        Tx       |Rx        Tx       |Rx        Tx       |Rx        Tx
-----------+---------+---------+---------+---------+---------+---------+---------+---------
+---------+---------+---------+---------+---------+---------+---------+---------
+---------
Eth 0/6      197       201       490       0         0         0         0
0         0         0         0         0         0         0         0         0
Eth 0/47     0         21087     0         0         0         0         0
0         0         0         0         0         0         0         0         0
Ve20         219       220       1068      108       0         214       0
0         0         643       428       0         0         0         0         0
Ve40         21034     21035     18264     64000     57        31535     0
0         73972     41170     61336     36982     0         0         0         0
Ve41         21032     21037     4         12        16        450       0
14913     4         0         0         0         0         0         0         0
Ve700        0         21086     0         0         0         0         0
0         0         0         0         0         0         0         0         0
```

13. Display settings such as neighbor timeout and hello interval.

```
device# show ipv6 pim settings
vrf : default-vrf
Maximum mcache           : 32768     Current Count             : 0
Hello interval           : 30        Neighbor timeout          : 105
Join/Prune interval      : 60        Inactivity interval       : 180
Hardware drop enabled    : 1         Prune wait interval       : 3
Register Suppress Time   : 60        Register Probe Time       : 10
Register Stop Delay      : 0         Register Suppress interval: 0
SSM Enabled              : No        SPT Threshold             : 1
Route Precedence         : uc-non-default uc-default none
```

Clear and Troubleshoot IPv6 PIM Information

Use clear commands to remove information about the internal state of IPv6 PIM. Use the debug command to enable debug flags for troubleshooting. You can run the commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

Important
Diagnostic commands are developed and intended for specialized troubleshooting. Work closely with Extreme Networks technical support in running debug commands and interpreting their results.

1. Remove the RP-to-group mappings from the static IPv6 PIM forwarding table.

```
device# clear ipv6 pim rp-map
```

2. Clear the IPv6 PIM forwarding cache.

```
device# clear ipv6 pim mcache
```

3. Remove the IPv6 PIM traffic counters.

```
device# clear ipv6 pim traffic
```

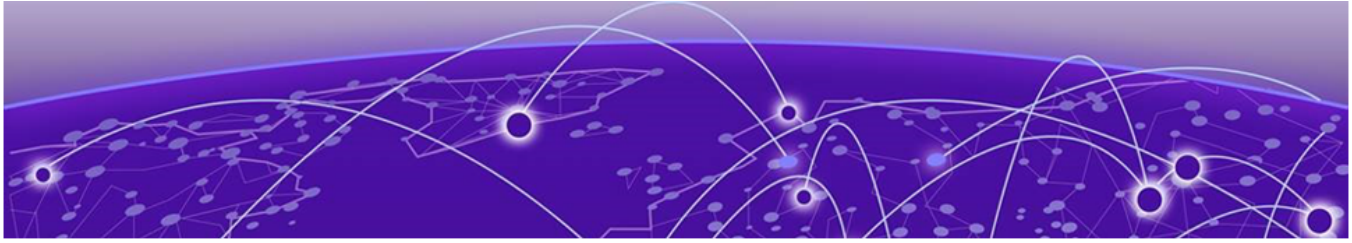
4. Enable debug flags.

```
device# debug ipv6 pim parameter
```

Parameter	Flag to enable
add-del-oif	OIF change
bootstrap	Bootstrap processing
group	Group processing
join-prune	Join and Prune processing
nbr-change	Neighbor changes
packets	Packet processing
parent	Parent change processing
route-change	Route changes
rp	Rendezvous Point processing
source	Source processing
state	State processing
all	State of all flags

5. Display the status of all IPv6 PIM debug flags.

```
device# show debug ipv6 pim
PIM debugging status:
--------------------------
add-del-oif     : off
bootstrap       : off
group           : off
join-prune      : on
nbr-change      : off
packets         : off
parent          : off
regproc         : on
route-change    : on
rp              : off
source          : off
--------------------------
```



IPv6 Multicast VLAN Traffic Reduction

IPv6 MLD Traffic Snooping on page 82
IPv6 PIM Traffic Snooping on page 87

IPv6 MLD Traffic Snooping

Overview

Multicast Listener Discovery (MLD) v1 and v2 snooping is a multicast-constraining mechanism that runs on Layer 2 or Layer 3 devices to manage and control IPv6 multicast groups.

A Layer 2 switch forwards all multicast control packets and data received on all member ports of a VLAN interface. This approach, though simple, is not bandwidth efficient, because only a subset of member ports may be connected to devices that are interested in receiving those multicast packets.

In the worst-case scenario, data is forwarded to all port members of a VLAN, even if only one VLAN member is interested in receiving the data. Such scenarios can lead to loss of throughput for a switch when it receives high-rate multicast data traffic.

MLD snooping is a multicast constraining mechanism that runs on Layer 2 or Layer 3 devices to manage and control IPv6 multicast groups. MLD snooping sends IPv6 multicast traffic only to interested listeners. By listening to and analyzing MLD messages, a Layer 2 device running MLD snooping establishes mappings between ports and multicast MAC addresses or multicast IP addresses, and forwards multicast data accordingly. Multicast routers in a network are found by means of static configuration, dynamic learning, or PIM hello-based mrouter detection.

Scalability

Scalability feature	Maximum
MLD group	4000
Number of snooping-enabled VLANS or bridge domains	256
Report learning rate	512 frames per second

Device support

Extreme Networks supports MLD snooping on the following devices.

- ExtremeSwitching SLX 9540
- ExtremeRouting SLX 9640

IPv6 MLD Queries and Messages

In a subnet, one multicast router is elected as an MLD querier. The MLD querier sends the following types of queries to hosts.

- **Multicast Listener General query**: (MLD v1, v2) Asks whether any host is listening to any group.
- **Multicast Listener Group-specific query**: (MLD v1, v2) Asks whether any host is listening to a specific multicast group. This query is sent in response to a host leaving the multicast group. The router can quickly determine whether any remaining hosts are interested in the group.
- **Multicast Listener Group-and-source-specific query**: (MLD v2) Asks whether any host is listening to group multicast traffic from a specific multicast source. This query is sent in response to a host indicating that it is not interested in receiving group-specific multicast traffic from the multicast source. The router can quickly determine whether any remaining hosts are interested in receiving group multicast traffic from that source.

Hosts that are multicast listeners send the following kinds of messages.

- **Multicast Listener Report**: Indicates that the host wants to join a particular multicast group.
- **Multicast Listener Done**: Indicates that the host wants to leave a particular multicast group.

IPv6 MLD Traffic and Data Forwarding

Traffic forwarding

MLD traffic is forwarded as follows.

- MLD general queries received on a multicast-router interface are forwarded to all other interfaces in the VLAN or bridge domain.
- MLD group-specific queries received on a multicast-router interface are forwarded to only those interfaces in the VLAN or bridge domain that are members of the group.
- MLD reports received on a host interface are forwarded to multicast-router interfaces in the same VLAN or bridge domain, but not to the other host interfaces in the VLAN or bridge domain.
- All unrecognized multicast packets must be flooded on to all (STP) unblocked member ports of the VLAN or bridge domain to ensure that no data traffic is dropped.
- MLD reports with an EXCLUDE source list are not supported. Traffic are flooded for the source lists in EXCLUDE mode

Data forwarding

Data forwarding rules ensure that the multicast traffic received at the switch is forwarded to all interested downstream port members. Forwarding rules are based on the Layer 3 multicast destination IP address.

- If a switch is already in a learned multicast group, multicast packets are forwarded only to those host interfaces in the VLAN or bridge domain that are members of the multicast group and to all multicast-router interfaces in the VLAN or bridge domain.
- If a switch is not in a learned multicast group (a group that has no current members), multicast packets are flooded to all member ports of the VLAN or bridge domain and to all multicast-router interfaces in the VLAN or bridge domain.

Enable IPv6 MLD Snooping

The **ipv6 mld snooping enable** command enables snooping on a VLAN or bridge domain.

1. Enter global configuration mode

```
device# configure terminal
```

2. Enter VLAN or bridge domain configuration mode.

```
device(config)# vlan 10
device(config)# bridge-domain 10
```

3. Enable MLD snooping.

```
device(config-vlan-10)# ipv6 mld snooping enable
device(config-bridge-domain-10)# ipv6 mld snooping enable
```

Configure IPv6 MLD Snooping

When IPv6 MLD snooping is enabled, you can configure various options for a VLAN or a bridge domain.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter VLAN or bridge domain configuration mode.

```
device(config)# interface vlan 10
```

3. Run the following commands as needed.

   For more information, including examples, see the *Extreme SLX-OS Command Reference*.

   a. Configure a VLAN port member to be a multicast router (mrouter) port.

```
device(config-vlan-10)# ipv6 mld snooping mrouter interface Ethernet 0/2
```

   b. Configure a logical member port of a bridge domain to be an mrouter port.

```
device(config-bridge-domain-10)# ipv6 mld snooping mrouter logical-interface
Ethernet 0/2.200
```

   c. Activate the MLD snooping querier function.

```
device(config-vlan-10)# ipv6 mld snooping querier enable
```

d. Configure a static group on an Ethernet or a port channel.

```
device(config-vlan-10)# ipv6 mld snooping static-group ff28::1 interface Ethernet
0/3
```

e. Specify the time limit for sending last member queries.

```
device(config-vlan-10)# ipv6 mld snooping last-member-query-interval 3000
```

f. Specify the interval between MLD snooping queries.

```
device(config-bridge-domain-10)# ipv6 mld snooping query-interval 600
```

g. Specify the amount of time to wait for a response from an MLD snooping query.

```
device(config-bridge-domain-10)# ipv6 mld snooping query-max-response-time 15
```

h. Specify the MLD version that the querier uses for snooping.

```
device(config-vlan-10)# ipv6 mld snooping version 2
```

i. Enable MLD snooping fast-leave processing.

```
device(config-bridge-domain-10)# ipv6 mld snooping fast-leave
```

j. Enable Source-Specific Multicast mapping.

```
device(config-vlan-10)# ssm-enable prefixlist 2::100
```

Display IPv6 MLD Snooping Information

You can use show commands to display information about IPv6 MLD snooping for a VLAN or a bridge domain. You can run the commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

1. Display information about learned groups.

```
device# show ipv6 mld groups vlan 100
Total number of groups: 1
MLD Connected Group Membership
Group Address         Interface        Uptime        Expires     Last Reporter
ff28::1               Vlan 100         00:02:03       Never      Static
Member Port: Eth 0/1
```

2. Display snooping information.

```
device# show ipv6 mld snooping bridge-domain 10
Bridge-domain ID: 10
 Multicast Router ports: Eth 0/44.200
 Querier - disabled
 MLD Operation mode: MLDv2
 Is Fast-Leave Enabled : Disabled Max Response time = 10
 Last Member Query Interval = 10 Query interval = 125
 Number of Multicast Groups: 0
 Group: ff28::1
 Member Ports: Eth 0/44.200
```

3. Display multicast router (mrouter) port-related information.

```
device# show ipv6 mld snooping mrouter vlan 100
Vlan         Interface       Expires (Sec)
100          Eth 0/1         Never
```

4. Display MLD packet statistics.

```
device# show ipv6 mld statistics bridge-domain 100
MLD packet statistics for Bridge-Domain 10:
MLD Message Type         Edge-Received     Edge-Sent    Edge-Rx-Errors
General Query            1                      0                 0
Group Specific Query     0                      0                 0
```

```
V1 Membership Report     0                    1                    0
V2 Membership Report     0                    0                    0
Group Leave              0                    0                    0
```

Clear and Troubleshoot IPv6 MLD Information

Use clear commands to remove information about the internal state of IPv6 MLD. Use the debug command to enable debug flags for troubleshooting. You can run the commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.



Important
Diagnostic commands are developed and intended for specialized troubleshooting. Work closely with Extreme Networks technical support in running debug commands and interpreting their results.

1. Remove the accumulated information about learned groups for a specified bridge domain or VLAN.

```
device# clear ipv6 mld groups ?
Possible completions:
    <IPv6 address>
    bridge-domain   Bridge-Domain ID
    vlan            Vlan interface
    |               Output modifiers
    <cr>
device# clear ipv6 mld groups
device# clear ipv6 mld groups ff28::1
device# clear ipv6 mld groups ff28::1 bridge-domain 10
device# clear ipv6 mld groups ff38::2 vlan 100
device# clear ipv6 mld groups bridge-domain 10
device# clear ipv6 mld groups vlan 10
```

2. Remove the accumulated MLD statistics for a specified VLAN or bridge domain.

```
device# clear ipv6 mld statistics ?
Possible completions:
    bridge-domain   Bridge-Domain ID
    vlan            Vlan interface
    |               Output modifiers
    <cr>
device# clear ipv6 mld statistics bridge-domain 100
device# clear ipv6 mld statistics vlan 100
```

3. Enable debug flags.

```
device# debug ipv6 mld ?
Possible completions:
    all         All
    errors      Errors
    group       Multicast group
    interface   Interface information
    l2-port     l2-port
    packet      packet
    rx          packet recieved
    tx          packet transmitted
    vlan        Vlan interface

device# debug ipv6 mld errors
```

IPv6 PIM Traffic Snooping

When IPv6 PIM snooping is enabled, switches listen for Hello, Join, and Prune messages to determine which multicast router ports need to receive multicast traffic in a VLAN.

The snooping functionality of IPv4 PIM is supported for IPv6 PIM. For more information, see IPv4 PIM-SM Traffic Snooping on page 51.

Enable IPv6 PIM Snooping

The **ipv6 pim snooping enable** command enables snooping on a VLAN or a bridge domain. By default, IPv6 PIM snooping is enabled when IPv6 PIM is enabled on a Layer 3 VLAN interface. By default, IPv6 PIM snooping is disabled on a Layer 2 interface.

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter VLAN or bridge domain configuration mode.

```
device(config)# vlan 30
```

3. Enable IPv6 PIM snooping.

```
device(config-vlan-30)# ipv6 pim snooping enable
```

Display IPv6 PIM Snooping Information

You can run these show commands from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

1. Display the snooping status on a VLAN or a bridge domain.

```
device# show ipv6 multicast snooping
Vlan id            : 100
MLD  snooping      : Enabled
PIM  snooping      : Enabled
MLD  Version       : MLDv1
```

2. Display the IPv6 multicast forwarding cache for a VLAN or a bridge domain.

```
show ipv6 multicast snooping mcache vlan 100
Flags : V1|V2 : MLD Receiver, P_G : PIM (*,G) Join, P_SG: PIM (S,G) Join
                BR : PIM Blocked RPT
 Vlan ID : 100
 -------------
1     (*, ff04:3000:4000::2 )  00:21:48    NumOIF: 1
        Outgoing Ports:
                eth0/14        Flags: 0x84 ( V1)  00:21:48/202s
2     (*, ff08:1001:2001::1 )  00:27:09    NumOIF: 1
        Outgoing Ports:
                eth0/14        Flags: 0x84 ( V1)  00:27:09/131s
3     (*, ff0e:1000:2000::1 )  00:27:09    NumOIF: 1
        Outgoing Ports:
                eth0/14        Flags: 0x84 ( V1)  00:27:09/131s
```

Clear IPv6 PIM Snooping Information

You can run this command from any level of the CLI. For more information, see the *Extreme SLX-OS Command Reference*.

Clear the IPv6 PIM snooping database for a VLAN or a bridge domain.

```
device# clear ipv6 mld snooping vlan 10
```