



Extreme SLX-OS

Layer 2 Switching Configuration Guide, 18x.1.00a

9035673-01 Rev 02
November 2020



Copyright © 2020 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: www.extremenetworks.com/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.....	9
Text Conventions.....	9
Documentation and Training.....	10
Getting Help.....	11
Subscribe to Service Notifications.....	11
Providing Feedback.....	11
About This Document.....	13
Supported hardware and software.....	13
What's new in this document.....	13
Regarding Ethernet interfaces and chassis devices.....	13
Link Aggregation.....	14
Link aggregation overview.....	14
Link Aggregation Control Protocol.....	15
LAG distribution process and conditions.....	15
Configuring and managing Link Aggregation.....	15
Configuring a new port channel interface.....	16
Deleting a port channel interface.....	16
Adding a member port to a port channel.....	17
Deleting a member port from a port channel.....	18
Configuring the minimum number of LAG member links.....	18
Configuring the LACP system priority	19
Configuring the LACP port priority.....	19
Configuring the LACP timeout period.....	20
LACP PDU forwarding.....	20
Configuring LACP default Up.....	21
Troubleshooting LACP.....	22
Hash-based load balancing.....	23
Configuring LAG hashing.....	23
Configuring header protocols for load-balancing.....	23
Load balancing mechanism on different traffic types.....	24
Show and clear LAG commands.....	25
Displaying port-channel information.....	25
Displaying LAG hashing.....	26
Displaying LACP system-id information.....	26
Displaying LACP statistics.....	26
Clearing LACP counter statistics on a LAG.....	27
Clearing LACP counter statistics on all LAG groups.....	27
VLANs.....	28
802.1Q VLAN overview.....	28
Configuring VLANs.....	28

Configuring a VLAN.....	28
Configuring a switchport interface.....	29
Configuring the switchport interface mode.....	29
Configuring the switchport access VLAN type.....	30
Configuring a VLAN in trunk mode.....	30
Configuring a native VLAN on a trunk port.....	31
Enabling VLAN tagging for native traffic.....	32
Displaying the status of a switchport interface.....	33
Displaying the switchport interface type.....	33
Verifying a switchport interface running configuration.....	34
Displaying VLAN information.....	34
Enabling Layer 3 routing for VLANs.....	35
VLAN statistics.....	35
Enabling statistics on a VLAN.....	36
Displaying statistics for VLANs.....	36
Clearing statistics on VLANs.....	37
Endpoint Tracking - MAC Authentication using RADIUS Protocol.....	37
Overview.....	37
Additional details and considerations.....	38
RADIUS VSA support.....	38
Enabling and Disabling Endpoint Tracking on a Port.....	39
VXLAN Layer 2 Gateway.....	44
VXLAN Layer 2 gateway overview.....	44
VXLAN Layer 2 gateway considerations and limitations.....	45
Configuring VXLAN Layer 2 gateway.....	45
VXLAN Layer 2 gateway support for bridge domains.....	47
Configuring VXLAN Layer 2 Gateway support for bridge domains.....	47
VXLAN Layer 2 gateway payload tag processing.....	48
VXLAN Layer 2 support for LVTEP.....	50
LVTEP control plane.....	50
LVTEP data plane.....	50
Configuring VXLAN LVTEP support.....	52
LVTEP support for other features.....	56
Nondefault TPID	58
Configuring TCAM profiles to support LVTEP.....	61
LVTEP show commands.....	61
QoS for VXLAN Layer 2 gateways.....	66
Configuring QoS for VXLAN Layer 2 gateways.....	67
Multi-Chassis Trunking (MCT).....	68
MCT Overview.....	68
MCT terminology.....	69
SLX-OS MCT control plane.....	69
MCT data plane.....	72
MAC management.....	77
MCT configuration considerations.....	80
General considerations.....	80
Peer considerations.....	80
VLAN considerations.....	81

LACP considerations.....	81
LSP considerations.....	82
Configuring the BGP EVPN peer.....	82
Configuring MCT.....	84
Taking the MCT node offline for maintenance.....	85
Configuring additional MCT cluster parameters.....	86
Changing the client-isolation mode	86
Changing the designated-forwarder hold timer value.....	86
Moving the traffic from an MCT node to the remote node.....	87
Configuring an auto-generated ESI for a cluster client.....	87
Displaying MCT information.....	87
Displaying the cluster information	87
Displaying the cluster client information.....	88
Displaying member VLAN information.....	88
Displaying and clearing the MAC address table cluster information.....	88
Layer 3 routing over MCT.....	89
Configuration considerations.....	89
Layer 3 MCT VLAN configuration example.....	90
Layer 3 MCT bridge-domain configuration example.....	91
Using MCT with VRRP and VRRP-E.....	92
MCT short path forwarding configuration using VRRP-E example.....	93
PE1 configuration.....	95
PE2 configuration.....	96
MCT use cases.....	97
L2 MCT in the data center core.....	97
L2 MCT in a data center with a collapsed core and aggregation.....	99
Logical Interfaces.....	101
Logical interfaces overview.....	101
LIFs and bridge domains.....	101
Configuration considerations.....	101
Configuring a logical interface on a physical port or port-channel (LAG).....	102
LIF considerations for the SLX 9030.....	103
No local switching.....	104
VC mode raw-passthrough.....	104
Dual-tag support on the SLX 9030.....	104
TPID support on the SLX 9030.....	107
PW profile support on the SLX 9030.....	107
Bridge Domains.....	109
Bridge domain overview.....	109
Bridge domain statistics.....	109
Configuring a bridge domain.....	110
Displaying bridge-domain configuration information.....	111
Enabling statistics on a bridge domain.....	114
Displaying statistics for logical interfaces in bridge domains.....	115
Clearing statistics on bridge domains.....	115
802.1ag Connectivity Fault Management.....	116
Maintenance Domain (MD).....	117
Maintenance Association (MA).....	118

Maintenance End Point (MEP).....	118
Maintenance Intermediate Point (MIP).....	118
CFM Hierarchy.....	119
Mechanisms of Ethernet IEEE 802.1ag OAM.....	119
Fault detection (continuity check message).....	119
Fault verification (Loopback messages).....	119
Fault isolation (Linktrace messages).....	119
Enabling or disabling CFM.....	119
Creating a Maintenance Domain.....	120
Creating and configuring a Maintenance Association.....	120
Displaying CFM configurations.....	121
show cfm.....	121
show cfm connectivity.....	122
show cfm brief.....	122
802.1d Spanning Tree Protocol.....	123
Spanning Tree Protocol overview.....	123
Spanning Tree Protocol configuration notes.....	123
Optional features.....	123
STP states.....	124
BPDUs.....	124
TCN BPDUs	125
STP configuration guidelines and restrictions.....	125
Understanding the default STP configuration.....	126
STP features.....	126
Root guard.....	126
BPDU guard.....	127
Error disable recovery.....	128
PortFast.....	128
STP parameters.....	128
Bridge parameters.....	128
Error disable timeout parameter.....	130
Port-channel path cost parameter.....	130
Configuring STP.....	130
Enabling and configuring STP globally.....	130
Enabling and configuring STP on an interface	132
Configuring basic STP parameters	134
Re-enabling an error-disabled port automatically	137
Clearing spanning tree counters.....	137
Clearing spanning tree-detected protocols	138
Shutting down STP	138
802.1w Rapid Spanning Tree Protocol.....	140
Rapid Spanning Tree Protocol overview	140
RSTP parameters.....	141
Edge port and automatic edge detection.....	141
Configuring RSTP.....	142
Enabling and configuring RSTP globally	142
Enabling and configuring RSTP on an interface	143
Configuring basic RSTP parameters.....	146

Clearing spanning tree counters.....	148
Clearing spanning tree-detected protocols	148
Shutting down RSTP	148
Per-VLAN Spanning Tree+ and Rapid Per-VLAN Spanning Tree+.....	150
PVST+ and R-PVST+ overview.....	150
PVST+ and R-PVST+ guidelines and restrictions.....	150
PVST+ and R-PVST+ parameters.....	151
PortFast.....	151
Edge port and automatic edge detection.....	151
Configuring PVST+ and R-PVST+.....	152
Enabling and configuring PVST+ globally	152
Enabling and configuring PVST+ on an interface	153
Enabling and configuring PVST+ on a system.....	155
Enabling and configuring R-PVST+ globally.....	161
Enabling and configuring R-PVST+ on an interface	163
Enabling and configuring R-PVST+ on a system.....	165
Clearing spanning tree counters.....	171
Clearing spanning tree-detected protocols	171
Shutting down PVST+ or R-PVST+	172
802.1s Multiple Spanning Tree Protocol.....	174
MSTP overview.....	174
Common Spanning Tree (CST)	174
Internal Spanning Tree (IST).....	174
Common Internal Spanning Tree (CIST).....	175
Multiple Spanning Tree Instance (MSTI)	175
MST regions.....	175
MSTP regions.....	175
MSTP global level parameters.....	176
MSTP interface level parameters.....	176
Edge port and automatic edge detection.....	176
BPDU guard.....	177
Configuring MSTP.....	177
Enabling and configuring MSTP globally.....	178
Enabling and configuring MSTP on an interface	181
Enabling MSTP on a VLAN.....	183
Configuring basic MSTP parameters.....	184
Clearing spanning tree counters.....	186
Clearing spanning tree-detected protocols	186
Shutting down MSTP	187
Topology Groups.....	188
Topology groups.....	188
Master VLAN, member VLANs, and bridge-domains.....	188
Control ports and free ports.....	189
Configuration considerations.....	189
Configuring a topology group.....	190
Configuring a master VLAN.....	190
Adding member VLANs.....	190
Adding member bridge-domains.....	191

Replacing a master VLAN.....	192
Displaying topology group information.....	192
Loop Detection.....	194
LD protocol overview.....	194
Strict mode.....	194
Loose mode.....	195
LD PDU format.....	196
LD PDU transmission.....	197
LD PDU reception.....	197
LD parameters.....	198
LD PDU processing.....	199
Support for EPVN VLAN tunnels.....	200
Configuration considerations.....	200
LD use cases.....	200
MCT strict mode.....	200
MCT loose mode.....	201
Configuring LD protocol.....	202
Loop detection for VLAN.....	205
Configuring loop detection for VLAN.....	205



Preface

This section describes the text conventions used in this document, where you can find additional information, and how you can provide feedback to us.

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 ExtremeSwitching switches or SLX routers, the product is referred to as *the switch* or *the router*.

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 <i>enter</i> and <i>type</i>	When you see the word <i>enter</i> 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 <i>type</i> .
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.
<i>italic</i> 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, <i>member</i> [<i>member</i> ...].
\	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 white papers, data sheets, and case studies

Extreme Networks offers product training courses, both online and in person, as well as specialized certifications. For details, visit www.extremenetworks.com/education/.

Getting Help

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 2826. 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 Service Notifications

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

1. Go to www.extremenetworks.com/support/service-notification-form.
2. Complete the form (all fields are required).
3. Select the products for which you would like to receive notifications.



Note

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

4. Select **Submit**.

Providing Feedback

The Information Development team at Extreme Networks has made every effort to ensure the accuracy and completeness of this document. We are always striving to improve our documentation and help

you work better, 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 in the document.
- Broken links or usability issues.

If you would like to provide feedback, you can do so in three ways:

- In a web browser, select the feedback icon and complete the online feedback form.
- Access the feedback form at <https://www.extremenetworks.com/documentation-feedback/>.
- Email us at documentation@extremenetworks.com.

Provide the publication title, part number, and as much detail as possible, including the topic heading and page number if applicable, as well as your suggestions for improvement.



About This Document

[Supported hardware and software](#) on page 13

[What's new in this document](#) on page 13

[Regarding Ethernet interfaces and chassis devices](#) on page 13

Supported hardware and software

The following platforms are supported by this release:

- ExtremeSwitching SLX 9030 Series

What's new in this document

The following table includes descriptions of new information added to this guide for SLX-OS Release 18x.1.00a.

Table 4: Summary of enhancements in SLX-OS Release 18x.1.00a

Feature	Description	Described in
Endpoint tracking updates	The endpoint tracking feature minimizes the configuration and management of VLANs on switches in the data center.	Endpoint tracking

Regarding Ethernet interfaces and chassis devices

The Ethernet interface configuration and output examples in this document may appear as either O/X or N/X assignments, where N is an integer greater than 0.

Be aware of the interface configuration options of your particular device.

In addition, some legacy show outputs may reflect output from a variety of devices, including chassis devices.



Link Aggregation

[Link aggregation overview on page 14](#)

[Configuring and managing Link Aggregation on page 15](#)

[Hash-based load balancing on page 23](#)

[Show and clear LAG commands on page 25](#)

Link aggregation overview

The aggregated trunk is referred to as a Link Aggregation Group (LAG) or *port-channel*. The LAG is viewed as a single link by connected devices, the Spanning Tree Protocol, IEEE 802.1Q VLANs, and so on. When one physical link in the LAG fails, the other links stay up. A small drop in traffic is experienced when the link carrying the traffic fails.

To configure links to form a LAG, the physical links must be of the same speed. Link aggregation can be done by statically configuring the LAG, or by dynamically configuring the LAG using the IEEE 802.1AX Link Aggregation Control Protocol (LACP).

When queuing traffic from multiple input sources to the same output port, all input sources are given the same weight, regardless of whether the input source is a single physical link or a trunk with multiple member links.

The benefits of link aggregation are as follows:

- Increased bandwidth (The logical bandwidth can be dynamically changed as the demand changes.)
- Increased availability
- Load sharing
- Rapid configuration and reconfiguration

Each LAG consists of the following components:

- A MAC address that is different from the MAC addresses of the LAG's individual member links.
- An interface index for each link to identify the link to the neighboring devices.
- An administrative key for each link. Only the links with the same administrative key value can be aggregated into a LAG. On each link configured to use LACP, LACP automatically configures an administrative key value equal to the port-channel identification number.

Two LAG types are supported:

- Static LAG— In static link aggregation, links are added into a LAG without exchanging any control packets between the partner systems. The distribution and collection of frames on static links is determined by the operational status and administrative state of the link.

- Dynamic, standards-based LAG using LACP—Dynamic link aggregation uses LACP to negotiate with links that can be added and removed from a LAG. Typically, two partner systems sharing multiple physical Ethernet links can aggregate a number of those physical links using LACP. LACP creates a LAG on both partner systems and identifies the LAG by the LAG ID. All links with the same administrative key, and all links that are connected to the same partner switch become members of the LAG. LACP continuously exchanges LACPDU s to monitor the health of each member link.

Link Aggregation Control Protocol

Link Aggregation Control Protocol (LACP) is an IEEE 802.1AX standards-based protocol that allows two partner systems to dynamically negotiate attributes of physical links between them to form logical trunks. LACP determines whether a link can be aggregated into a LAG. If a link can be aggregated into a LAG, LACP puts the link into the LAG. All links in a LAG inherit the same administrative characteristics.

LACP operates in two modes:

- *Active mode*— LACP initiates the LACPDU exchange regardless of whether the partner system sends LACPDU s.
- *Passive mode* — LACP responds to Link Aggregation Control Protocol Data Units (LACPDU s) initiated by its partner system but does not initiate the LACPDU exchange.

LAG distribution process and conditions

The LAG aggregator is associated with the collection and distribution of Ethernet frames. The collection and distribution process implements:

- Inserting and capturing control PDU s.
- Restricting the traffic of a given conversation to a specific link.
- Load balancing between individual links.
- Handling dynamic changes in LAG membership.

On each port, link aggregation control does the following:

- Maintains configuration information to control port aggregation.
- Exchanges configuration information with other devices to form LAGs.
- Attaches ports to and detaches ports from the aggregator when they join or leave a LAG.
- Enables or disables an aggregator's frame collection and distribution functions.

LAG configuration guidelines:

- Interfaces configured as switchport interfaces cannot be aggregated into a LAG. However, a LAG can be configured as a switchport.

Configuring and managing Link Aggregation

The following sections discuss working with Link Aggregation on Extreme devices.

Configuring a new port channel interface

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface port-channel** command to create a new port channel interface at the global configuration level.

```
device(config)# interface port-channel 30
```



Note

The port-channel interface ranges from 1 to 64.

Example

The following example creates a new port channel interface of 30.

```
device# configure terminal
device(config)# interface port-channel 30
```

What to Do Next

After creating a new port channel, you can do "no shutdown" or "shutdown" to bring up or down the port-channel as follows.

```
device# configuration terminal
device(config)# interface Port-channel 30
2016/10/17-20:31:21, [NSM-1004], 302, M2 | Active | DCE, INFO, SLX, Port-channel 30 is
created.
device(config-Port-channel-30)#
device(config-Port-channel-30)# no shutdown
2016/10/17-20:31:26, [NSM-1019], 303, M2 | Active | DCE, INFO, SLX, Interface Port-
channel 30 is administratively up.
device(config-Port-channel-30)#
```

Deleting a port channel interface

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. To delete a port-channel interface, enter the **no interface port-channel** command.

```
device(config)# no interface port-channel 30
```

Example

The following example deletes port-channel interface 30.

```
device# configure terminal
device(config)# no interface port-channel 30
```


Adding a member port to a port channel

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **interface port-channel** command to add a port channel interface at the global configuration level.

```
device(config)# interface port-channel 30
device(conf-Port-channel-30)#
```

3. Configure the **interface ethernet** command to enable the interface.

```
device(conf-Port-channel-30)# interface ethernet 0/5
device(conf-if-eth-0/5)#
```

4. Add a port to the port channel interface as static.

```
device(conf-if-eth-0/5)# channel-group 30 mode on
```

5. Add a port to the port channel interface as a dynamic (using LACP), active or passive mode.

```
device(conf-if-eth-0/5)# channel-group 30 mode active
```

Example

The following example is for a static LAG configuration with the mode ON.

```
device# configure terminal
device(config)# interface port-channel 30
device(conf-Port-channel-30)# interface ethernet 0/5
device(conf-if-eth-0/5)# channel-group 30 mode on
```

The following example adds a port 0/5 to the existing dynamic port channel interface 30 with the mode active.

```
device# configure terminal
device(config)# interface port-channel 30
device(conf-Port-channel-30)# interface ethernet 0/5
device(conf-if-eth-0/5)# channel-group 30 mode active
```



Note

Run the **no shutdown** command to bring the above interface online.

```
device(conf-if-eth-0/5)# no shutdown
2016/10/18-03:47:15, [NSM-1019], 528, M2 | Active | DCE, INFO, SLX, Interface
Ethernet 0/5 is administratively up.2016/10/18-03:47:15, [NSM-1001], 529, M2 |
Active | DCE, INFO, SLX, Interface Ethernet 0/5 is online.
```

The following example adds a port 0/5 to the existing dynamic port channel interface 30 with the mode passive.

```
device# configure terminal
device(config)# interface port-channel 30
device(conf-Port-channel-30)# interface ethernet 0/5
device(conf-if-eth-0/5)# channel-group 30 mode passive
```

Deleting a member port from a port channel

Procedure

Delete a port from the port channel interface.

```
device(conf-if-eth-0/5)# no channel-group
```

Example

The following example deletes a port 0/5 from the existing port channel interface 30.

```
device# configure terminal
device(config)# interface ethernet 0/5
device(conf-if-eth-0/5)# no channel-group
```

Configuring the minimum number of LAG member links

About This Task

This configuration allows a port-channel to operate at a certain minimum bandwidth at all times. If the bandwidth of the port-channel drops below the minimum number, then the port-channel is declared operationally DOWN even though it has operationally UP members.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **interface port-channel** command at the global configuration level.

```
device(config)# interface port-channel 30
device(conf-Port-channel-30)#
```

3. Configure the minimum number of LAG member links at the port-channel interface configuration mode.

```
device(conf-Port-channel-30)# minimum-links 5
```



Note

The number of links ranges from 1 to 32. The default minimum links is 1.

Example

The following example sets min-link 5 to the existing port channel interface 30.

```
device# configure terminal
device(config)# interface port-channel 30
device(conf-Port-channel-30)# minimum-links 5
```

Configuring the LACP system priority

About This Task

You configure the LACP system priority on each switch running LACP. LACP uses the system priority with the switch MAC address to form the system ID and also during negotiation with other switches.

The system priority value must be a number in the range of 1 through 65535. The higher the number, the lower the priority. The default priority is 32768.

To configure the global LACP system priority, perform the following steps:

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Specify the LACP system priority.

```
device(config)# lacp system-priority 25000
```

3. To reset the system priority to the default value.

```
device(config)# no lacp system-priority
```

Configuring the LACP port priority

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface port-channel** command to add a port channel interface at the global configuration level.

```
device(config)# interface port-channel 30  
device(conf-Port-channel-30)#
```

3. Configure the **interface ethernet** command and add the port to the port-channel interface.

```
device(conf-Port-channel-30)# interface ethernet 0/5  
device(conf-if-eth-0/5)# channel-group 30 mode active
```

4. Configure the LACP port priority 12 for the member port.

```
device(conf-if-eth-0/5)# lacp port-priority 12
```



Note

The LACP port priority value ranges from 1 to 65535. The default value is 32768.

5. To reset the configured port priority to the default value.

```
device(conf-if-eth-0/5)# no lacp port-priority
```

Example

The example sets the port priority as 12.

```
device# configure terminal
device(config)# interface port-channel 30
device(conf-Port-channel-30)# interface ethernet 0/5
device(conf-if-eth-0/5)# channel-group 30 mode active
device(conf-if-eth-0/5)# lacp port-priority 12
```

Configuring the LACP timeout period

About This Task

The **short** timeout period is 3 seconds and the **long** timeout period is 90 seconds. The default is **long**. The **short** timeout period specifies that the PDU is sent every second and the port waits three times this long (three seconds) before invalidating the information received earlier on this PDU. The **long** timeout period specifies that the PDU is sent once in 30 seconds and the port waits three times this long (90 seconds) before invalidating the information received earlier on this PDU.

To configure the LACP timeout period on an interface, perform the following steps:

Procedure

1. Enter the **configure terminal** command to access global configuration mode.
2. Enter the **interface** command, specifying the interface type and the slot/port.

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

3. Enter the **no shutdown** command to enable the interface.
4. Specify the LACP timeout short period for the interface.

```
device(conf-if-eth 0/1)# lacp timeout short
```

5. Specify the LACP timeout long period for the interface.

```
device(conf-if-eth 0/1)# lacp timeout long
```

LACP PDU forwarding

Since the destination address of the PDU is a multicast MAC, the frame will be flooded on the VLAN. If the VLAN on which the LACP PDU is received is a regular VLAN, the PDU will be flooded on the VLAN. If the VLAN on which the PDU is received is a service delimiter for a bridge domain, the LACP PDU is flooded on the bridge domain accordingly.

LACP PDU forwarding is supported only on physical interfaces and static port channel interfaces. LACP PDUs cannot be forwarded if they are received on a LACP based dynamic port channel. LACP PDU forwarding enabled on a static port channel applies to all the member ports. If LACP is enabled on a port, it overrides the LACP PDU forwarding configuration and the PDUs are trapped in the CPU.

*Configuring LACP PDU forwarding on a physical interface***Procedure**

1. Enter the global configuration mode.

```
device# configure terminal
device(config)#
```

2. Specify the physical interface on which LACP PDU forwarding needs to be enabled.

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

3. Configure LACP PDU forwarding on the physical interface.

```
device(conf-if-eth-0/1)# lacp-pdu-forward enable
```

Example

The following example enables LACP forwarding on a port-channel interface.

```
device# configure terminal
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# lacp-pdu-forward enable
```

*Configuring LACP PDU forwarding on a port-channel interface***Procedure**

1. Enter the global configuration mode.

```
device# configure terminal
```

2. Enter the **interface port-channel** command to add a port channel interface at the global configuration level.

```
device(config)# interface port-channel 10
```

3. Configure LACP PDU forwarding on the port-channel interface.

```
device(conf-Port-channel-10)# lacp-pdu-forward enable
```

LACP PDU forwarding is supported only on static port channel interfaces.

Example

The following example enables LACP forwarding on a port-channel interface.

```
device# configure terminal
device(config)# interface port-channel 10
device(conf-Port-channel-10)# lacp-pdu-forward enable
```

Configuring LACP default Up

About This Task

Consider the following when using the **lacp default-up** command:

- The command is available only if the configured interface is a dynamic member of a port-channel interface.
- The command is not supported on static LAGs.
- The command is not supported on port-channel interfaces.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.
2. Enter the **interface** command, specifying the interface type and the slot/port.

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

3. Specify LACP default-up for the interface.

```
device(conf-if-eth-0/1)# lacp default-up
```

4. Enter the no form of the command to disable the configuration.

```
device(conf-if-eth-0/1)# no lacp default-up
```

Troubleshooting LACP

To troubleshoot problems with your LACP configuration, use the following troubleshooting tips.

If a standard IEEE 802.1AX-based dynamic trunk is configured on a link and the link is not able to join the LAG, do the following:

- Make sure that both ends of the link are configured as **standard** for the trunk type.
- Make sure that both ends of the link are *not* configured for **passive** mode. They must be configured as **active /active**, **active /passive**, or **passive /active**.
- Make sure that the port-channel interface is in the administrative "up" state by ensuring that the **no shutdown** command was entered on the interface on both ends of the link.
- Make sure that the links that are part of the LAG are connected to the same neighboring switch.
- Make sure that the system ID of the switches connected by the link is unique. You can verify this by entering the **show lacp sys-id** command on both switches.
- Make sure that LACPDUs are being received and transmitted on both ends of the link and that there are no error PDUs. You can verify this by entering the **show lacp counters number** command and looking at the receive mode (rx) and transmit mode (tx) statistics. The statistics should be incrementing and should not be at zero or a fixed value. If the PDU rx count is not incrementing, check the interface for possible CRC errors by entering the **show interface link-name** command on the neighboring switch. If the PDU tx count is not incrementing, check the operational status of the link by entering the **show interface link-name** command and verifying that the interface status is "up."

When a link has problem, the **show port-channel** command displays the following message:

```
Mux machine state: Deskew not OK.
```

If a static trunk is configured on a link and the link is not able to join the LAG, do the following:

- Make sure that both ends of the link are configured as **standard** for trunk type and verify that the mode is "on."
- Make sure that the port-channel interface is in the administrative "up" state by ensuring that the **no shutdown** command was entered on the interface on both ends of the link.

Hash-based load balancing

Configuring LAG hashing

To configure symmetric LAG hashing on supported devices, complete the following tasks.

1. Define where to start picking headers for the key generation, using the **lag hash hdr-start** command.
 - **fwd**—Start from the header that is used for the forwarding of the packet (inner header). This is the default option.
 - **term**—Start from the last terminated header (outer header)—the header after the forwarding header. For switching traffic, as there is no header below the forwarding header, hashing is not visible.
2. Configure the number of headers to be considered for LAG hashing, using the **lag hash hdr-count** command. The default value is 1. There can be a maximum of 3 headers—based on the first header selected using the command in the previous step.

The following options provide other LAG configurations to achieve specific tasks:

- Configure hash rotate using the **lag hash rotate** command to provide different options for randomness of hashing. The number can be between 0 and 15. The default value is 3.
- If there is a need to use the same hash in both directions, configure hash normalize, using the **lag hash normalize** command. The normalize option is disabled by default.
- Allow the source port to be included in the hashing configuration using the **lag hash srcport** command. The source port is not used for hashing by default.
- Enter the **lag hash pwctrlword** command to skip the psuedowire control word in the hashing configuration.

Configuring header protocols for load-balancing

Select the protocol header type using one of the following commands. By default, all the header parameters are enabled as shown here. If you disable a header, you can then re-enable its parameters one-by-one.

- Ethernet headers:
 - **load-balance hash ethernet da-mac**
 - **load-balance hash ethernet etype**
 - **load-balance hash ethernet sa-mac**
 - **load-balance hash ethernet vlan**
- IPv4 and L4 headers
 - **load-balance hash ip dst-ip**
 - **load-balance hash ip dst-l4-port**
 - **load-balance hash ip protocol**
 - **load-balance hash ip src-ip**
 - **load-balance hash ip src-l4-port**

- IPv6 and L4 headers
 - `load-balance hash ipv6 ipv6-dst-ip`
 - `load-balance hash ipv6 ipv6-dst-l4-port`
 - `load-balance hash ipv6 ipv6-next-hdr`
 - `load-balance hash ipv6 ipv6-src-ip`
 - `load-balance hash ipv6 ipv6-src-l4-port`

Load balancing mechanism on different traffic types

The following table provides information about load balancing on different traffic types.

Table 5: Load balancing on different traffic types

Traffic type	Header field	Description
Layer 2/ Layer 3 packet load balancing	<ul style="list-style-type: none"> • Ethernet DA, SA, Etype, Vlan-id • IPv4/v6 dst IP, src IP • L4 Src-Port, Dst-Port 	<ul style="list-style-type: none"> • Ethernet destination address, source address, ethernet type, VLAN ID load balancing • IPv4/v6 destination address, source address load balancing • Layer 4 source and destination port-based load balancing
VPLS/ VLL packet load balancing	<p>CE to PE router traffic can use the following fields for load-balancing similar to the Layer 2/ Layer 3 traffic)</p> <ul style="list-style-type: none"> • Ethernet DA, SA, Etype, Vlan-id • IPv4/v6 dst IP, src IP • L4 Src-Port, Dst-Port <p>PE to CE router traffic can use the following fields for load-balancing</p> <ul style="list-style-type: none"> • Customer (inner) ethernet DA, SA, Etype, Vlan-id • Customer (inner) IPv4/v6 dst IP, lpv4/ lpv6 src IP, protocol • Customer (inner) L4 Src-Port, Dst-Port 	<p>CE to PE router traffic</p> <ul style="list-style-type: none"> • Ethernet destination address, source address, ethernet type, VLAN ID load balancing • IPv4/v6 destination address, source address load balancing • Layer 4 source and destination port-based load balancing <p>PE to CE router traffic</p> <ul style="list-style-type: none"> • Customer ethernet destination and source address, ethernet type, VLAN ID load balancing • Customer IPv4/v6 destination address, source address load balancing • Customer Layer 4 source and destination port-based load balancing

Show and clear LAG commands

This section contains tasks for showing port-channel information and statistics and for clearing the relevant counters.

Displaying port-channel information

Procedure

1. Use the **show port-channel summary** command to display brief information of all port-channels.

```
device# show port-channel summary
Flags: D - Down                P - Up in port-channel (members)
      U - Up (port-channel)    * - Primary link in port-channel
      S - Switched
      M - Not in use. Min-links not met

=====
Group Port-channel  Protocol  Member ports
=====
1     Po 1    (D)    None      Eth 0/25 (D)
                        Eth 0/26 (D)
2     Po 2    (D)    None      Eth 0/27 (D)
                        Eth 0/28 (D)
10    Po 10   (U)    LACP      Eth 0/4  (P)
                        Eth 0/18 (P)
100   Po 100  (U)    None      Eth 0/10 (P)
                        Eth 0/11 (P)
```

2. Use the **show port-channel detail** command to display detailed information of all the port-channels.

```
device# show port-channel detail
Static Aggregator: Po 1
Aggregator type: Standard
Number of Ports: 2
Member ports:
  Eth 0/25
  Eth 0/26

Static Aggregator: Po 2
Aggregator type: Standard
Number of Ports: 2
Member ports:
  Eth 0/27
  Eth 0/28

Static Aggregator: Po 100
Aggregator type: Standard
Number of Ports: 2
Member ports:
  Eth 0/10
  Eth 0/11

LACP Aggregator: Po 10
Aggregator type: Standard
Actor System ID - 0x8000,76-8e-f8-0a-98-00
```

```

Admin Key: 0010 - Oper Key 0010
Receive link count: 2 - Transmit link count: 2
Individual: 0 - Ready: 1
Partner System ID - 0x8000,76-8e-f8-0a-68-00
Partner Oper Key 0010
Number of Ports: 2
Member ports:
  Link: Eth 0/4 (0x18820016) sync: 1
  Link: Eth 0/18 (0x18890084) sync: 1

```

- Use the **show port-channel** *number* command to display detailed information of a specific port-channel interface

```

device# show port-channel 10
LACP Aggregator: Po 10
Aggregator type: Standard
Admin Key: 0010 - Oper Key 0010
Partner System ID - 0x8000,76-8e-f8-0a-68-00
Partner Oper Key 0010
Number of Ports: 2
Member ports:
  Link: Eth 0/4 (0x18820016) sync: 1
  Link: Eth 0/18 (0x18890084) sync: 1

```

Displaying LAG hashing

```

device# show port-channel load-balance
Header parameters
  Ethernet Mask: sa-mac da-mac etype vlan
  ip: src-ip dst-ip protocol src-l4-port dst-l4-port
  ipv6: ipv6-src-ip ipv6-dst-ip ipv6-next-hdripv6-src-l4-port ipv6-dst-l4-port

Hash Settings
  hdr-start:FWD, hdr-count:1, bos-start:0, bos-skip:0, skip-cw:0
  normalize:0, rotate:3, include_src_port:0, Disable: L2 0, ipv4 0, ipv6 0

load-balance-type hash-based

```

Displaying LACP system-id information

Procedure

Enter the **show lacp sys-id** command to display LACP information for the system ID and priority.

```

device# show lacp sys-id
System ID: 0x8000,76-8e-f8-0a-98-00

```

Displaying LACP statistics

Procedure

Enter the **show lacp counters** command to display LACP statistics for a port-channel.

Clearing LACP counter statistics on a LAG

Procedure

Enter the **clear lacp** *LAG_group_number* **counters** command to clear the LACP counter statistics for the specified LAG group number.

```
device# clear lacp 42 counters
```

Clearing LACP counter statistics on all LAG groups

Procedure

Enter the **clear lacp counter** command to clear the LACP counter statistics for all LAG groups.

```
device# clear lacp counter
```



VLANs

[802.1Q VLAN overview on page 28](#)

[Configuring VLANs on page 28](#)

[Enabling Layer 3 routing for VLANs on page 35](#)

[VLAN statistics on page 35](#)

[Endpoint Tracking - MAC Authentication using RADIUS Protocol on page 37](#)

802.1Q VLAN overview

IEEE 802.1Q VLANs provide the capability to overlay the physical network with multiple virtual networks. VLANs allow you to isolate network traffic between virtual networks and reduce the size of administrative and broadcast domains.

A VLAN contains end stations that have a common set of requirements that are independent of physical location. You can group end stations in a VLAN even if they are not physically located in the same LAN segment. VLANs are typically associated with IP subnetworks and all the end stations in a particular IP subnet belong to the same VLAN. Traffic between VLANs must be routed. VLAN membership is configurable on a per-interface basis.

Configuring VLANs

About This Task

The following sections discuss working with VLANs on Extreme devices.

Configuring a VLAN

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **vlan** command to create a topology group at the global configuration level.

```
device(config)# vlan 5  
device(config-vlan-5)#
```



Note

The **no vlan** command removes the existing VLAN instance from the device.

Configuring a switchport interface

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface ethernet** command to configure the interface mode.

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

3. Enter the **switchport** command to configure a switchport interface.

```
device(conf-if-eth-0/1)# switchport
```

Configuring the switchport interface mode

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface ethernet** command to configure the interface mode.

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

3. Enter the **switchport** command to set the interface as switchport.

```
device(conf-if-eth-0/1)# switchport
```

4. Enter the **switchport mode** command to configure the switchport interface in trunk mode.

```
device(conf-if-eth-0/1)# switchport mode trunk
```



Note

The default mode is access. Enter the **switchport mode access** command to set the mode as *access*.

What to Do Next



Note

Before you change the switch port mode from **switchport mode access** with an explicit **switchport access vlan** to **switchport mode trunk-no-default-native**, you must enter the **no switchport** command on the interface level, and then enter the **switchport** command to set the interface as a switchport. Now you can configure the **switchport mode trunk-no-default-native** command.

Configuring the switchport access VLAN type

Before You Begin

Ensure that reserved VLANs are not used. Use the **no switchport access vlan** command to set the default VLAN as the access VLAN.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface ethernet** command to specify an Ethernet interface.

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

3. Enter the **switchport** command to set the interface as switchport.

```
device(conf-if-eth-0/1)# switchport
```

4. Enter the **switchport access vlan** command to set the mode of the interface to *access* and specify a VLAN.

```
device(conf-if-eth-0/1)# switchport access vlan 10
```

Example

This example sets the mode of a specific port-channel interface to *trunk*.

```
device# configure terminal
device(config)# interface port-channel 35
device(config-port-channel-35)# switchport mode trunk
```

Configuring a VLAN in trunk mode

Before You Begin

Ensure that reserved VLANs are not used.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface ethernet** command to specify an Ethernet interface.

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

3. Enter the **switchport** command to set the interface as switchport.

```
device(conf-if-eth-0/1)# switchport
```

4. Enter the **switchport trunk allowed vlan** command to set the mode of the interface to *trunk* and add a VLAN.

```
device(conf-if-eth-0/1)# switchport trunk allowed vlan add 5
```

Example

The example sets the mode of the Ethernet interface to *trunk*.

```
device# configure terminal
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# switchport mode trunk
```

Example

The example sets the mode of a port-channel interface to *trunk* and allows all VLANs.

```
device# configure terminal
device(config)# interface port-channel 35
device(config-Port-channel-35)# switchport trunk allowed vlan all
```

Configuring a native VLAN on a trunk port

Before You Begin

Ensure that reserved VLANs are not used.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **interface ethernet** command to configure the interface mode.

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

3. Enter the **switchport** command to set the interface as switchport.

```
device(conf-if-eth-0/1)# switchport
```

4. Enter the **switchport trunk native-vlan** command to set native VLAN characteristics to access and specify a VLAN.

```
device(conf-if-eth-0/1)# switchport trunk native-vlan 300
```

Example

This example removes the configured native VLAN on the Ethernet interface.

```
device# configure terminal
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# no switchport trunk native-vlan 300
```

Enabling VLAN tagging for native traffic

Before You Begin

Ensure that reserved VLANs are not used.

The following table describes the acceptable frame types, as well as system behavior, for tagged native VLAN, untagged native VLAN, and no native VLAN.

Table 6: Acceptable frame types and system behavior for native VLANs

	Tagged native VLAN	Untagged native VLAN	No native VLAN
Configuration	switchport trunk tag native-vlan (Default) and Globally: <code>vlan dot1q tag native</code>	no switchport trunk tag native-vlan or Global config: <code>no vlan dot1q tag native</code>	switchport mode trunk-no-default-native
Acceptable frame type	VLAN-tagged only	All (tagged and untagged)	VLAN-tagged only
Receive untagged	Drop	Forward/flood in native VLAN	Drop
Receive tagged on native VLAN	Forward/flood in native VLAN	Forward/flood in native VLAN	Drop
Transmit on native VLAN	Tagged with native VLAN	Untagged packet	Will not forward on native VLAN

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
```

2. Enter the **interface ethernet** command to configure the interface mode.

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

3. Enter the **switchport** command to set the interface as switchport.

```
device(conf-if-eth-0/1)# switchport
```

4. Enter the **switchport trunk tag native-vlan** command to enable tagging for native traffic data VLAN characteristics on a specific interface.

```
device(conf-if-eth-0/1)# switchport trunk tag native-vlan
```

Example

This example enables tagging for native traffic data on a specific Ethernet interface.

```
device# configure terminal
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# switchport trunk tag native-vlan
```


Example

This example disables the native VLAN tagging on a port-channel.

```
device# configure terminal
device(config)# interface port-channel 35
device(config-Port-channel-35)# no switchport trunk tag native
```

Displaying the status of a switchport interface

Procedure

Enter the **show interface switchport** to display the detailed Layer 2 information for all interfaces.

```
device# show interface switchport
Interface name      : Eth 0/1
Switchport mode    : access
Ingress filter     : enable
Acceptable frame types : all
Default Vlan       : 1
Active Vlans       : 1
Inactive Vlans     : -
Interface name     : Port-channel 5
Switchport mode    : access
Ingress filter     : enable
Acceptable frame types : all
Default Vlan       : 1
Active Vlans       : 1
```

Displaying the switchport interface type

Procedure

Enter the **show interface switchport** to display the detailed Layer 2 information for a specific interface.

```
device# show interface ethernet 0/1 switchport
Interface name      : ethernet 0/1
Switchport mode    : trunk
Fcoepoint enabled  : no
Ingress filter     : enable
Acceptable frame types : vlan-tagged only
Native Vlan        : 1
Active Vlans       : 1,5-10
Inactive Vlans     : -
```

Example

The example displays the detailed Layer 2 information for a port-channel interface.

```
device# show interface port-channel 5 switchport
Interface name      : Port-channel 5
Switchport mode    : access
Fcoepoint enabled  : no
Ingress filter     : enable
Acceptable frame types : vlan-untagged only
```

```

Default Vlan      : 1
Active Vlans     : 1
Inactive Vlans   : -

```

Verifying a switchport interface running configuration

Procedure

Enter the **show running-config interface** to display the running configuration information for a specific interface.

```

device# show running-config interface ethernet 0/1 switchport
interface interface Eth 0/1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan add 5-10
  switchport trunk tag native-vlan

```

Example

This example displays the running configuration information for a port-channel interface.

```

device# show running-config interface port-channel 5 switchport
interface Port-channel 5
  switchport
  switchport mode access
  switchport access vlan 1

```

Displaying VLAN information

Procedure

1. Enter the **show vlan** to display information about VLAN 1.

```

device# show vlan 1
VLAN Name State Ports
(u)-Untagged, (t)-Tagged
(c)-Converged
=====
1 default ACTIVE Eth 0/1(t) Eth 0/4(t) Eth 0/5(t) Eth 0/8(t)

```

2. Enter the **show vlan detail** command to display detailed information.

```

device# show vlan det
VLAN: 1, Name: default
Admin state: ACTIVE, Config status: Static
Number of interfaces: 7
  Eth 0/4, tagged, Static
  Eth 0/3, tagged, Static
  Eth 0/2, tagged, Static
  Eth 0/8, tagged, Static
  Eth 0/6, tagged, Static
  Eth 0/9, untagged, Static
  Po 20, tagged, Static
VLAN: 10, Name: VLAN0010
Admin state: ACTIVE, Config status: Static
Number of interfaces: 3
  Eth 0/3, tagged, Static
  Eth 0/2, tagged, Static

```

```

    Eth 0/4, tagged, Static
    Po 20, tagged, Static
VLAN: 11, Name: VLAN0011
Admin state: ACTIVE, Config status: Static
Number of interfaces: 3
    Eth 0/3, tagged, Static
    Eth 0/2, tagged, Static
    Eth 0/4, tagged, Dynamic (MVRP)
VLAN: 12, Name: VLAN0012
Admin state: ACTIVE, Config status: Dynamic (MVRP)
Number of interfaces: 1
    Eth 0/4, tagged, Dynamic (MVRP)
VLAN: 13, Name: VLAN0013
Admin state: ACTIVE, Config status: Dynamic (EP tracking)
Number of interfaces: 1
    Eth 0/6, tagged, Dynamic (EP tracking)
VLAN: 14, Name: VLAN0014
Admin state: INACTIVE(member port down), Config status: Static
Number of interfaces: 1
    Eth 0/8, tagged, Static

```

Enabling Layer 3 routing for VLANs

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Create a VLAN.

```
device(config)# vlan 200
```

3. Create a virtual Ethernet (VE), assign an IP address and mask, and enable the interface.

```
device(config)# interface ve 200
device(config-Ve-200)# ip address 10.2.2.1/24
device(config-Ve-200)# no shutdown
```

A VE interface can exist without a VLAN configuration, but it must be provisioned in the VLAN in order to be used.

4. Enter the **router-interface** command and specify the VLAN.

```
device(config-vlan-200)# router-interface ve 200
```

VLAN statistics

Use the **statistics** command in the VLAN configuration mode to enable statistics on a VLAN.



Note

Statistics has to be manually enabled for a specific VLAN, since it is not enabled by default for VLANs.

Please note that:

- The statistics reported are not real-time statistics since they depend upon the load on the system.
- Statistics has to be manually enabled for a specific VLAN. This ensures better utilization of the statistics resources in the hardware.

- Statistics for VLANs with VE interfaces consider only the switched frames. Packets which are routed into or out of the VE interface are not counted.
- Enabling statistics on a VLAN has a heavy impact on the data traffic.

Enabling statistics on a VLAN

1. Enter the global configuration mode.

```
device# configure terminal
```

2. Enter the **vlan** command to specify a VLAN for statistics collection.

```
device(config)# vlan 5
device(config-vlan-5)#
```

3. Enter the **statistics** command to enable statistics for all ports and port channels on configured VLANs.

```
device(config-vlan-5)# statistics
```



Note

Use the **no statistics** command to disable statistics on VLANs.

```
device(config-vlan-5)# no statistics
```

Displaying statistics for VLANs

Enter the **show statistics vlan** command to view the statistics for all ports and port channels on all configured VLANs.

```
device# show statistics vlan

Vlan 10 Statistics
Interface      RxPkts      RxBytes      TxPkts      TxBytes
eth 0/1        821729      821729      95940360    95940360
eth 0/2        884484      885855      95969584    95484555
po 1           8884        8855        9684        9955

Vlan 20 Statistics
Interface      RxPkts      RxBytes      TxPkts      TxBytes
eth 0/6        821729      821729      95940360    95940360
eth 0/21       8884        8855        9684        9955
po 2           884484      885855      95969584    95484555
```

Table 7: Output descriptions of the show statistics vlan command

Field	Description
Interface	The interface whose counter statistics are displayed.
RxPkts	The number of packets received at the specified port.
RxBytes	The number of bytes received at the specified port.
TxPkts	The number of packets transmitted from the specified port.
TxBytes	The number of bytes transmitted from the specified port.

Displaying VLAN statistics for a specific VLAN

Enter the **show statistics vlan** *vlan ID* command to view the statistics for a specific VLAN. Here *vlan ID* is the specific VLAN ID.

```
device# show statistics vlan 10

Vlan 10 Statistics
Interface      RxPkts      RxBytes      TxPkts      TxBytes
eth 0/1        821729      821729      95940360    95940360
eth 0/2        884484      885855      95969584    95484555
po 1           8884        8855        9684        9955
```

Clearing statistics on VLANs

Enter the **clear statistics vlan** command to clear the statistics for all ports and port channels on all configured VLANs.

```
device# clear statistics vlan
```

Clearing statistics for a specific VLAN

Enter the **clear statistics vlan** *vlan ID* command to clear the statistics for a specific VLAN. Here *vlan ID* is the specific VLAN ID.

```
device# clear statistics vlan 10
```

Endpoint Tracking - MAC Authentication using RADIUS Protocol

Overview

Statically provisioning VLANs in a data center network has the following drawbacks:

- Managing VLANs on top-of-rack (TOR) switches is tedious for the administrator.
- Having VLANs provisioned ahead of time increases the size of the active topology for control protocols such as STP and RSTP, increasing convergence times.
- Flood (unknown unicast, broadcast) traffic can unnecessarily eat up bandwidth on the TOR-to-EOR (end of row) links.
- For the case when the virtual machine (VM) sends or receives tagged traffic, flood traffic can consume CPU cycles on every server that is connected to the network.
- The dynamic VLAN feature allows SLX-OS to create, prune, and open VLANs on the switch dynamically as they are needed by the VMs. This enables the VLAN to follow the VM as it migrates between servers in the data center.

This feature provides the following to remedy the above drawbacks:

- Association of MAC addresses to specific VLANs.
- Once a MAC address is authorized and the VLAN is not already provisioned on the switch, the dynamic (1) creation of the VLAN to which this MAC is associated, (2) configuration of the tag, and (3) addition of the feature-enabled port on which the MAC was detected.

- Once the last MAC address using the VLAN is deleted or aged out, deletion of the VLAN and associated resources on the switch.

The endpoint tracking feature also authorizes the VM. When a VM (and MAC address) is authorized, SLX-OS dynamically creates the VLAN that is required for the VM to send traffic. If a VM shuts down or is moved, its VLAN is pruned to preserve bandwidth. In this way the network responds to changes in the VM network.

Additional details and considerations

Note the following:

- The user enables or disables this feature on a Layer 2 port (switchport), by means of the **endpoint-tracking enable** command. This feature is also supported on LAG and MCT ports.
- SLX-OS communicates with a RADIUS server or XMC-NAC (Extreme Management Center Network Access Control) for MAC authentication information, which can map a MAC address to a VLAN. RADIUS VLAN assignment triggers the creation of the VLAN and the port VLAN membership. SLX-OS expects MAC-to-VLAN binding information from RADIUS. VNI (VXLAN Network Identifier) assignment is not supported.
- The maximum number of authenticated or nonauthenticated MAC addresses supported for each endpoint-tracking-enabled port is 2000 per system.
- Port VLAN membership is local to SLX-OS. The removal of port VLAN membership or the deletion of the VLAN is not propagated to the RADIUS server.
- On an access port, if RADIUS does not provide MAC-to-VLAN mapping, SLX-OS assigns a default VLAN. Only the first RADIUS VLAN assignment is honored, and subsequent MAC addresses cannot override the existing assignment.
- For VMs sending tagged traffic, if the RADIUS authentication response does not have a VLAN assignment for the MAC, SLX-OS creates the VLAN (assuming the VLAN does not already exist) with the received packet's tag and adds the port to the VLAN as tagged. Dynamically created VLANs are not saved across reboots, and therefore are not part of the running configuration.
- When a port becomes part of a VLAN dynamically, this is not reflected in the running configuration of the interface.
- In reauthentication, RADIUS sends two attributes as part of a Change of Authorization (CoA) request (a CoA is an unsolicited message from RADIUS to the switch to trigger an authentication-related action): (1) Calling-station-id, a value field that holds the MAC being authenticated, and VSA, a vendor-specific attribute value field that holds commands and parameters for commands. RADIUS matches the outstanding requests response by means of a CoA identifier.

RADIUS VSA support

This feature supports the following VSA format.

Table 8: VSA format

Type (26)	Length	Vendor ID (Extreme Vendor ID)	
Vendor ID (cont'd)		Sub-attr (Sub-type)	Sub-length
Value			

The VSA fields are as follows.

Table 9: VSA fields

Field	Description
Type	8-bit field. Always 26 as it represents a VSA attribute.
Length	8-bit field. Length of entire attribute, including type and length fields.
Vendor ID	4 octets, encoding the Extreme Vendor ID.
Sub-attr (Sub-type)	8-bit field, indicating class of command.
Sub-length	8-bit field. Length of the Sub-attr (attribute), including sub-type and sub-length fields.
Value	A string, integer, or IP address based on Sub-type

As a RADIUS Access response, only one VSA attribute is provided, for Egress-vlan. The Egress-vlan VSA format has sub-type 216 and a value field of type integer. The format is shown below.

Table 10: Egress-vlan VSA format

Type (26)	Length	Vendor ID (Extreme Vendor ID)	
Vendor ID (cont'd)		Sub-type (216)	Sub-length
egress VLAN_ID			



Note

Both Cisco and Extreme Networks Vendor IDs are supported for the above.

For MAC reauthentication, RADIUS sends a Change of Authorization (CoA) of Code 43 with the VSA. This has a sub-type of 1 and the Value field is the string "subscriber:command=reauthenticate". The format is shown below.

Table 11: MAC reauthentication format

Type (26)	Length	Vendor ID (Extreme Vendor ID)	
Vendor ID (cont'd)		Sub-type (1)	Sub-length
subscriber:command=reauthenticate			



Note

Both Cisco and Extreme Networks Vendor IDs are supported for the above.

Enabling and Disabling Endpoint Tracking on a Port



Note

For configuration details, refer to the "VLANs" chapter in this guide.

Endpoint tracking on an access port

When endpoint tracking is enabled on an access port, initial MAC learning occurs on the default VLAN and after authentication if there is RADIUS VLAN assignment. Note the following:

- If RADIUS assigns the VLAN, the VLAN is created if it does not exist. The Layer 2 forwarding entry is added with the RADIUS-assigned VLAN.
- If RADIUS does not assign the VLAN, the Layer 2 forwarding entry is added with the default VLAN.

The following illustrates the use of the **endpoint-tracking enable** command in switchport access mode (the default).

```
device# configure terminal
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# switchport
device(conf-if-eth-0/1)# endpoint-tracking enable
```

Endpoint tracking on a trunk port

On a trunk port, packets are always tagged. To allow any tagged traffic on the port, ingress VLAN filtering is disabled in the forwarding plane when endpoint tracking is enabled. As a result, all the tagged packets are trapped to the CPU as a Layer 2 learning event, which is later authenticated by RADIUS. A MAC learning event that is generated has wire tag information as part of the VLAN information. After authentication, a RADIUS-assigned dynamic VLAN is added for egress forwarding on the port.

The following illustrates the use of the **endpoint-tracking enable** command in switchport trunk mode.

```
device# configure terminal
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# switchport
device(conf-if-eth-0/1)# switchport mode trunk
device(conf-if-eth-0/1)# endpoint-tracking enable
```

By default, reauthentication for each session on the port is disabled. However, you can optionally set a timer value for reauthentication, by means of the **endpoint-tracking timeout reauth-period** command, as in the following example.

```
device(conf-if-eth-0/1)# endpoint-tracking timeout reauth-period 86400
```

Once endpoint tracking is enabled on a trunk port, the following occurs:

1. The source MAC lookup fails to find a matching entry in hardware and the packet is sent to the CPU for Layer 2 learning with a tag in place of the IVID.
2. When received by the CPU, the wire tag information is stored as a VLAN, and the MAC address is sent for authentication.
3. If authentication is successful and the RADIUS VLAN assignment matches the wire tag, or if there is no VLAN assignment from RADIUS, VLAN creation is triggered and the egress VIF is set for the port to allow forwarding and flooding.

Verifying configurations

A variety of **show** commands are available to verify the configuration of endpoint tracking, as described in the following table.

Table 12: Endpoint tracking show commands

Command	Description
show mac-address-table endpoint-tracking authenticated	Displays authenticated MAC addresses that are learned on all ports that are enabled for endpoint tracking.
show mac-address-table endpoint-tracking authentication-failed	Displays nonauthenticated MAC addresses that are learned on all ports that are enabled for endpoint tracking.
show mac-address-table endpoint-tracking authenticated interface	Displays authenticated MAC addresses that are learned on a specific port that is enabled for endpoint tracking.
show mac-address-table endpoint-tracking authentication-failed interface	Displays nonauthenticated MAC addresses that are learned on a specific port that is enabled for endpoint tracking.
show vlan detail	Displays detailed VLAN information.

Examples are shown below respectively for these commands.

show mac-address-table endpoint-tracking authenticated

```
device# show mac-address-table endpoint-tracking authenticated
VlanId/BDId  Mac-address      Type      State(R-Radius,F-FailOpen)  Ports
100 (V)      0001.0100.0001  Dynamic   Authenticated(R)           Eth 0/5
```

show mac-address-table endpoint-tracking authenticated interface

```
device# show mac-address-table endpoint-tracking authenticated interface ethernet 0/5
VlanId/BDId  Mac-address      Type      State      Ports
100 (V)      0001.0100.0001  Dynamic   Authenticated  Eth 0/5
```

show mac-address-table endpoint-tracking authentication-failed

```
device# show mac-address-table endpoint-tracking authentication-failed interface ethernet 0/5
VlanId/BDId  Mac-address      Type      State(R-Radius,F-FailOpen)  Ports
100 (V)      0001.0100.0001  Dynamic   Authenticated(R)           Eth 0/5
```

show mac-address-table endpoint-tracking authentication-failed interface

```
device# show mac-address-table endpoint-tracking authentication-failed interface ethernet 0/5
VlanId/BDId  Mac-address      Type      State      Ports
100 (V)      0001.0100.0001  Dynamic   Pending    Eth 0/5
100 (V)      0001.0100.0002  Dynamic   Failed     Eth 0/5
```

show vlan detail

```
device# show vlan detail
VLAN: 10, Name: VLAN0010
Admin State: ACTIVE, Config status: Dynamic (Endpoint Tracking)
Number of interfaces: 6
  Eth 0/6: tagged, trunk, dynamic (EP tracking)
  Eth 0/10: tagged, trunk, dynamic (EP tracking)
  Tu 61441: tagged, trunk, dynamic(BGP-EVPN), vni : 10010
  Tu 61442: tagged, trunk, static, vni : 10010
```

VM MAC aging and flush

When a VMs MAC address is deleted because of aging or a flush operation, if it is the last MAC on the port, SLX-OS removes the port from the VLAN and checks to see if the VLAN is associated with other ports. If there are no other associations, the VLAN is also deleted.

MAC reauthentication

MAC reauthentication lets the RADIUS server send unsolicited messages to SLX-OS, to relearn the MAC address of the VM. Note the following:

- SLX-OS stores the CoA request identifier and uses the same identifier in the response (ACK/NAK).
- A RADIUS reauthentication request without calling-station-id is returned with a NAK.
- A RADIUS reauthentication request with a calling-station-id that is not present in the switch is returned with a NAK.
- A RADIUS request with a different vendor-id is silently ignored by the switch.
- Duplicate CoA messages are rejected at the switch.
- Re-authentication can be achieved by means of either a CoA or a Disconnect message. In the case of a CoA message, the VM is not removed and traffic is not disrupted during re-authentication. In the case of a Disconnect message, the VM is removed from the switch. When new traffic from the VM is relearned, authentication is triggered, and traffic is disrupted.

Additional Considerations for Endpoint Tracking

Please note the following:

- Logical Interfaces are not supported on endpoint tracking ports.
- Spanning-tree Protocol (STP) should be disabled on a port before the port is configured for endpoint tracking.

MCT support

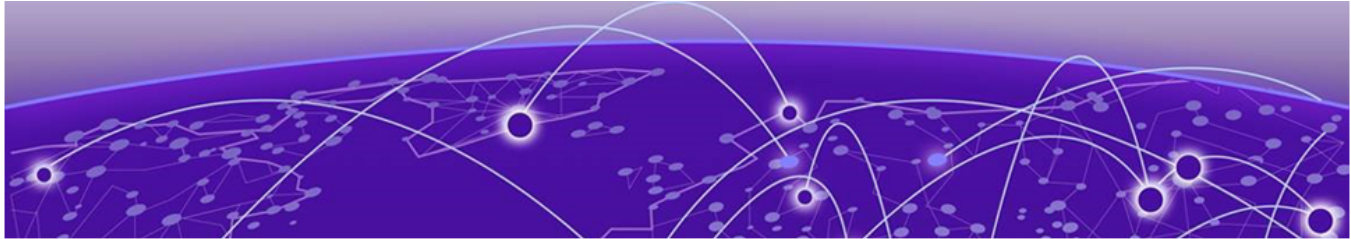
A VLAN that is created dynamically on one MCT peer node is communicated to other peer node. Similarly, if a cluster client edge port (CCEP) port becomes part of a VLAN dynamically on one MCT peer node, this is communicated to the other peer node. The receiving MCT peer node, depending on the message, creates a dynamic VLAN or dynamic port VLAN membership.

The dynamic deletion of a VLAN or port VLAN membership is triggered by the last local MAC deletion. In addition, similar to the addition case, this information is also communicated to the peer.

Static VLAN dependency

A statically configured VLAN has a higher precedence over a dynamically created VLAN. Note the following:

- If a dynamically created VLAN is configured as static as well, removing the static VLAN also removes the dynamic VLAN information from the system.
- If a dynamically created port/VLAN membership is configured as static as well, removing the static port/VLAN membership also removes the dynamic port/VLAN membership from the system.



VXLAN Layer 2 Gateway

- [VXLAN Layer 2 gateway overview on page 44](#)
- [VXLAN Layer 2 gateway considerations and limitations on page 45](#)
- [Configuring VXLAN Layer 2 gateway on page 45](#)
- [VXLAN Layer 2 gateway support for bridge domains on page 47](#)
- [VXLAN Layer 2 support for LVTEP on page 50](#)

VXLAN Layer 2 gateway overview

The following figure illustrates an example Layer 2 gateway topology.

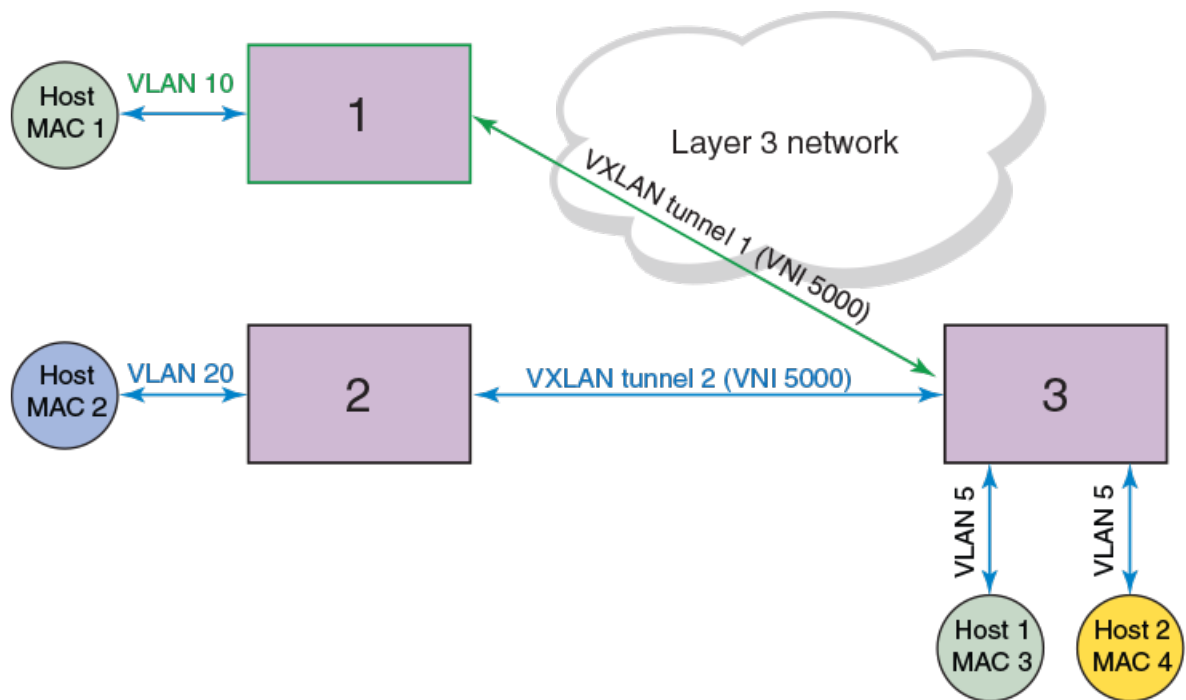


Figure 1: Layer 2 gateway topology

Device	IP address	Mapping
1: SLX 9850 series	1.1.1.1	VNI 5000 < > VLAN 10
2: VDX 6740 series	2.2.2.2	VNI 5000 < > VLAN 20
3: SLX 9540 series	3.3.3.3	VNI 5000 < > VLAN 5

VLANs on each node are extended through common Virtual Network Instance (VNI) 5000. The MAC addresses of local hosts are learned on access points, and MAC addresses are learned on VXLAN tunnels. A split-horizon topology is supported. All nodes participating in the VLAN must be connected through VXLAN tunnels, because there is no tunnel-to-tunnel flooding of broadcast, unknown unicast, and multicast (BUM) traffic.

In this topology, Devices 1, 2, and 3 are VXLAN Layer 2 gateway devices. On Device 3, tunnel 1 and tunnel 2 are mapped to VLAN 5. VLAN 5 has two hosts, MAC 3 and MAC 4. Device 3 is connected to two other hosts, Device 1 and Device 2, which connect to hosts MAC 1 and MAC 2, respectively, through VXLAN tunnels 1 and 2, respectively. If MAC 3 needs to establish traffic to MAC 1, initially there will be BUM flooding and upon a response from MAC 1, MAC 1 is learned through tunnel 1. Subsequent traffic goes directly from MAC 3 to Device 1 on tunnel 1. Traffic in the reverse direction comes from Device 1, is decapsulated, and goes to MAC 3.

VXLAN Layer 2 gateway considerations and limitations

- A maximum of 256 tunnels are supported.
- A maximum of 8 ECMP paths are supported.
- Layer 2 snooping is not supported.
- VRRPe source IP addresses are not supported.
- QoS, TTL, and MTU values are not configurable. The MTU is based on the IP interface MTU. If a packet is bigger than the IP interface MTU minus the VXLAN header, the packet is dropped. The default TTL value is 255. The default QoS value is 0, which is applied to the DSCP field of the IP header.
- VXLAN tunnels have the standard UDP header encapsulated with the standard defined value of 4789. This value is not configurable. SLX-OS expects VXLAN tunnel packets to be received with this value.
- Only BGP EVPN-based VXLAN tunnels are supported. Static VXLAN tunnels are not supported.
- The tunnel TX bytes statistics do not account for the outer VLAN header size.
- When BUM packets received on a tunnel are flooded, split horizon drops the packets to the same tunnel. However, the TX Statistics counter on that tunnel increments.
- BD VE does not support a VXLAN underlay next hop.
- Only one next hop can be selected from any physical port as a VXLAN next hop.

Configuring VXLAN Layer 2 gateway

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter the **overlay-gateway** command, specify the name of a gateway, and enter VXLAN overlay gateway configuration mode.

```
device(config)# overlay-gateway GW1
```

3. Enter the **type** command and specify **layer2-extension**.

```
device(config-overlay-gw-GW1)# type layer2-extension
```

4. Enter the **map vlan vni** command and specify **l2-extension**.

```
device(config-overlay-gw-GW1)# map vlan 5 vni 5000
```

- Enter the **map bridge-domain** command and specify a bridge domain and VNI.

```
device(config-overlay-gw-GW1)# map bridge-domain 1 vni 2000
```

- Enter the **ip interface** command and specify a loopback ID.

```
device(config-overlay-gw-GW1)# ip interface loopback 1
```

- Enter the **site** command, specify a site name, and enter VXLAN overlay gateway site configuration mode.

```
device(config-overlay-gw-GW1)# site mysitel
```

This mode configures a VXLAN tunnel to the site node. Only regular VTEPs are supported; logical VTEPs are not supported.

- Enter the **ip address** command and specify an IP address.

```
device(config-overlay-gw-GW1-site-mysitel)# ip address 1.1.1.1
```

- Enter the **extend vlan add** command and specify a VLAN.

```
device(config-overlay-gw-GW1-site-mysitel)# extend vlan add 5
```

- Enter the **extend bridge-domain add** command and specify a bridge domain.

```
device(config-overlay-gw-GW1-site-mysitel)# extend bridge-domain add 1
```

- Enter the **activate** command to activate the site.

```
device(config-overlay-gw-GW1-site-mysitel)# activate
```

- In privileged EXEC mode, enter the **show overlay-gateway** command to confirm the gateway configuration.

```
device# show overlay-gateway
Overlay Gateway "GW1", ID 1,
Admin state up
IP address 3.3.3.3 (loopback 1), Vrfdefault-vrf
Number of tunnels 1
Packet count: RX 17909 TX 1247
Byte count : RX (500125) TX 356626
```

- In privileged EXEC mode, enter the **show tunnel** command to confirm the tunnel configuration.

```
device# show tunnel 61441
Tunnel 61441, mode VXLAN
Ifindex 0x7c00f001, Admin state up, Operstate up
Source IP 3.3.3.3, Vrf: default-vrf
Destination IP 1.1.1.1
Active next hops on node 1:
IP: 4.4.4.5, Vrf: default-vrf
Egress L3 port: Ve45, Outer SMAC: 609c.9f5a.4415
Outer DMAC: 609c.9f5a.0015, ctag: 0
BUM forwarder: yes
```

- In privileged EXEC mode, enter the **show vlan** command to confirm the VLAN configuration.

```
device# show vlan 5
VLAN          Name          State          Ports
Classification
(R)-RSPAN
=====
5              VLAN05        ACTIVE         Eth 2/1(t)
               tu61441 vni5000
```

15. In privileged EXEC mode, enter the **show mac-address-table** command to confirm the MAC configuration.

```
device# show mac-address-table
VlanId/BDId  Mac-address      Type      State      Ports/LIF/PW
35 (V)       609c.9f5a.5b15   Dynamic   Active     Po 35
45 (V)       609c.9f5a.4415   Dynamic   Active     Po 45
5 (V)        0000.0400.0011   Dynamic   Active     tu61441
5 (V)        0000.0500.0011   Dynamic   Active     Eth 0/5
5 (V)        0000.0400.0011   Dynamic   Active     tu61441
5 (V)        0000.0500.0011   Dynamic   Active     Eth 0/5
Total MAC addresses : 6
device#
```

VXLAN Layer 2 gateway support for bridge domains

Since a bridge domain supports different port and VLAN endpoints, all of its traffic can be extended to a remote node through one VNI.

Also, VXLAN gateway support to bridge domains enables VLAN translation of traffic on both sides of the network. The local VLANs can use different VLAN tags on either side of the network and map to the same VNI.



Note

Only point-to-multipoint bridge domains are supported to extend over VXLAN tunnels. Point-to-point bridge domains are not supported.

You can extend the bridge domain under a site configuration. You can configure the bridge domain to VNI mapping automatically with auto mode where the bridge domain to the VNI is mapped implicitly. For example, VLANs 1 through 4096 are mapped to VNI 1 through 4090, respectively, and the bridge domain 1 is mapped to 4097. You can also configure the bridge domain to a VNI map manually, similarly to that of a VLAN.



Note

The default tagging mode for a bridge domain is raw mode.

Configuring VXLAN Layer 2 Gateway support for bridge domains

Before You Begin

Before performing this configuration, configure a point-to-multipoint bridge domain.

About This Task

Follow these steps configure a VXLAN Layer 2 gateway to support bridge domains.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Create a VXLAN overlay gateway and access the overlay gateway configuration mode.

```
device(config)# overlay-gateway gateway1
device(config-overlay-gw-gateway1)#
```

- Configure the Layer 2 type extension.

```
device(config-overlay-gw-gateway1)# type layer2-extension
```

- Specify a loopback interface.

```
device(config-overlay-gw-gateway1)# ip interface loopback 1
```

- Enable the mapping of a bridge domain to a VNI.

```
device(config-overlay-gw-gateway1)# map bridge-domain 1 vni 999
```

- Create a remote Layer 2 extension site in a VXLAN overlay gateway and access site configuration mode.

```
device(config-overlay-gw-gateway1)# site bd1
```

- Specify the destination IPv4 address of a tunnel.

```
device(config-site-bd1)# ip address 10.67.67.1
```

- Configure a bridge domain for the tunnel to the site.

```
device(config-site-bd1)# extend bridge-domain add 1
```

- Exit site configuration mode.

```
device(config-site-bd1)# exit
```

- Activate the gateway.

```
device(config-overlay-gateway-gateway1)# activate
```

Example

The following summarizes the configuration example.

```
device# configure terminal
device(config)# overlay-gateway gateway1
device(config-overlay-gw-gateway1)# type layer2-extension
device(config-overlay-gw-gateway1)# ip interface loopback 1
device(config-overlay-gw-gateway1)# map bridge-domain 1 vni 999
device(config-overlay-gw-gateway1)# site bd1
device(config-site-bd1)# ip address 10.67.67.1
device(config-site-bd1)# extend bridge-domain add 1
device(config-site-bd1)# exit
device(config-overlay-gateway-gateway1)# activate
```

VXLAN Layer 2 gateway payload tag processing

An SLX-OS device provides the following modes for the processing of the payload tag that is received on the attachment-circuit packets:

- VXLAN RFC-compliant mode
- Enhanced payload tag transport mode

VXLAN RFC-compliant mode

In RFC-compliant mode, the VLAN tag in a packet is not carried in the packet and must be stripped at the ingress device before the VXLAN encapsulated packet is sent into the network.

To configure RFC-compliant mode, configure the bridge domain in raw mode as shown in the following example.

```
pw-profile test
vc-mode raw
```



```
bridge-domain 10 p2mp
  pw-profile test
```

Then extend the bridge domain in the overlay gateway, as shown in the following example.

```
overlay-gateway gateway1
  type layer2-extension
  ip interface loopback 1
  map bridge-domain 10 vni 999
  site vcs1
  ip address 10.67.67.1
  extend bridge-domain add 10
```



Note

An SLX-OS device supports RFC-compliant mode with the bridge domain-based VXLAN service only.

Enhanced payload tag transport mode

In enhanced payload tag transport mode, one VLAN tag from the traffic is carried as part of the VXLAN encapsulated packet as an inner payload tag. This tag can carry the PCP value to include the priority information and also can interoperate with other devices. This tag is removed in the remote device capable of this behavior.



Note

This mode does not interoperate with RFC-compliant mode.

This mode is supported for VLAN-based VXLAN service and bridge domain-based VXLAN service with tag mode.

To configure enhanced payload tag transport mode, configure the bridge domain in tagging mode as shown in the following example.

```
pw-profile test
  vc-mode tag

bridge-domain 10 p2mp
  pw-profile test
```

Then extend the bridge domain in the overlay gateway, as shown in the following example.

```
overlay-gateway gateway1
  type layer2-extension
  ip interface Loopback 1
  map bridge-domain 10 vni 999
  site vcs1
  ip address 10.67.67.1
  extend bridge-domain add 10
```

VXLAN Layer 2 support for LVTEP

This section details support for a logical VXLAN tunnel end point (LVTEP) at Layer 2.

LVTEP control plane

The BGP session between the cluster peers must be configured with an encapsulation of MCT. The BGP session with the remote peers must be configured with an encapsulation of VXLAN.

The LVTEP control plane uses MCT Control Plane Designated Forwarder Election among the cluster peers. BGP VXLAN tunnels that are discovered automatically are treated as cluster client end points (CCEPs).

The following table shows Ethernet Segment Identifier (ESI) values for the VXLAN tunnels.

Table 13: ESI values for VXLAN tunnels

Parameter	Value
ESI type	0
Destination IP address (DIP)	4 bytes
Source IP address (SIP)	4 bytes

The ESI label is allocated globally for the LVTEP and is the same for all LVTEP tunnels. The tunnel operational status that is used for LVTEP tunnels is the same as that used for cluster clients.

For BGP EVPN deployment, MAC addresses that are learned on the LVTEP VXLAN tunnels are not synchronized between the cluster peers. However, for static LVTEP, the data-plane MAC addresses that are learned on the CCEPs of the LVTEP tunnel are synchronized between the cluster peers.



Note

LVTEP is supported only with the TCAM profile `VxlanExtended`, as configured by the `profile tcam vxlan-ext` command in hardware configuration mode. It is not supported in other TCAM profiles.

All the `show` commands that apply to MCT clients also apply to tunnel cluster clients.

LVTEP data plane

Example topology

The following figure illustrates a basic LVTEP topology for the data plane, with cluster nodes supporting remote peers and client end points (CEPs) and cluster client end points (CCEPs).

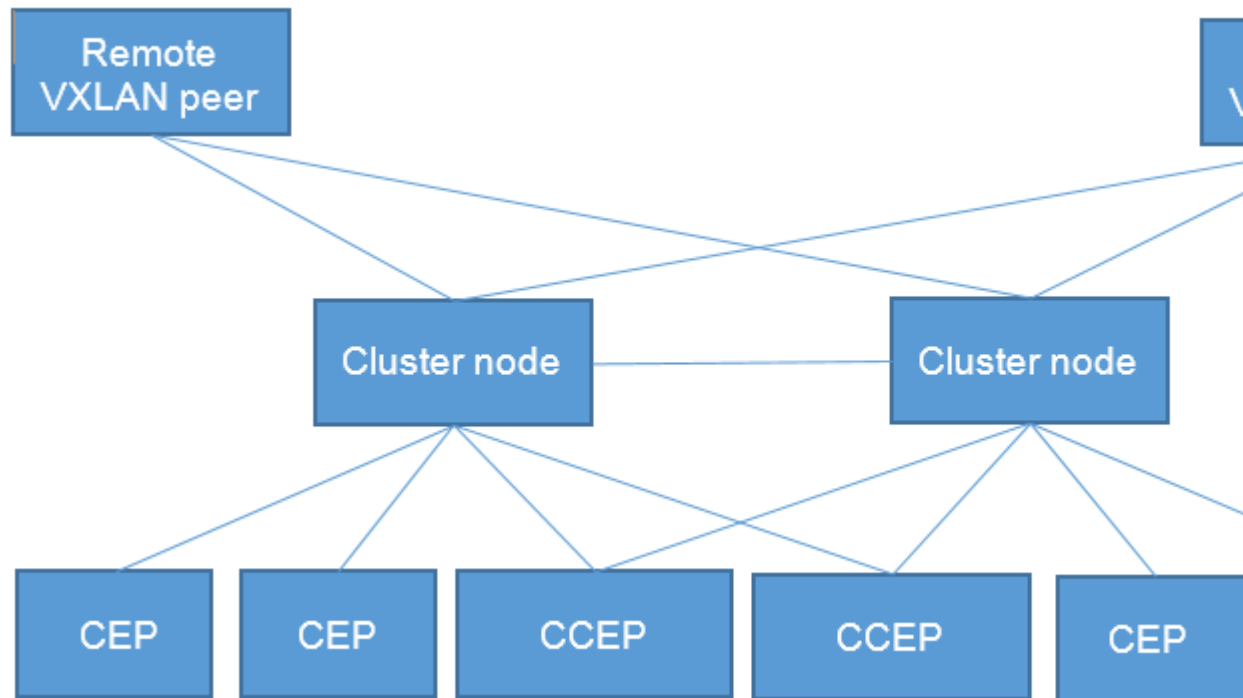


Figure 2: LVTEP topology

Dynamic cluster configuration support

When the cluster is dynamically changed from "Deploy" to "No Deploy" and conversely, the LVTEP tunnel CCEP is accordingly converted to a single VTEP and conversely.

When one or both cluster nodes are configured with "No Deploy" of the cluster configuration, the traffic behavior becomes undeterministic. Recovery scenarios are presented here.

Deploy to No Deploy

Do the following to recover from this error scenario:

1. Deactivate the overlay gateway on both cluster nodes.
2. Undeploy (or remove) the cluster configuration on both the cluster nodes.
3. Change the loopback address on both the cluster nodes to be different from the LVTEP loopback address.
4. Activate the overlay gateway on both the cluster nodes.

No Deploy to Deploy

Do the following to recover from this error scenario:

1. Deactivate the overlay gateway on both the cluster nodes.
2. Change the loopback address on both the cluster nodes to a common LVTEP loopback address that is different from the previous loopback address.

3. Deploy (or configure) a cluster on both the cluster nodes.
4. Activate the overlay gateway on both the cluster nodes.

Deployment scenario 1: BGP session and tunnel down, remote peers

For a BGP EVPN deployment, the MAC addresses learned on the LVTEP tunnel CCEP are flushed. Traffic destined to a remote VXLAN tunnel, although the tunnel may be up on the cluster node, is still flooded on the corresponding VLAN/BD. This behavior differs from that for the Layer 2 CCEP, where traffic is forwarded to the cluster node with the PWE unicast label. The primary reason for this behavior is that the MAC addresses are not synced between the cluster nodes (they are learned on the LVTEP tunnel CCEP).

For the static LVTEP, the behavior is the same as that for the Layer 2 CCEP, where the MAC addresses are synced between the cluster peers. When a BGP session or tunnel down event occurs, the MAC addresses point to the corresponding MCT EVPN PWE and are forwarded to the cluster peer; they are not flooded on the corresponding VLAN/BD.

Configuring VXLAN LVTEP support

Before You Begin

Configure the appropriate TCAM profile for this feature. Refer to [Configuring TCAM profiles to support LVTEP](#).

Procedure

1. Configure multiple loopback interfaces to support BGP neighbor address-family and the LVTEP IP address.

- a. Enter global configuration mode.

```
device# configure terminal
```

- b. In global configuration mode, specify a loopback port number.

```
device(config)# interface loopback 1
```

- c. Configure a loopback interface with OSPF area 0 and an IP address, and enable the interface to support BGP neighbor address-family.

```
device(config-Loopback-1)# ip ospf area 0
device(config-Loopback-1)# ip address 6.6.100.6/32
device(config-Loopback-1)# no shutdown
```

- d. Configure a second loopback interface to support the LVTEP IP address.

```
interface Loopback 2
 ip ospf area 0
 ip address 6.7.100.67/32
 no shutdown
```

The same address is used for both nodes in the cluster.

2. In global configuration mode, create two VLANs to support a pair of logical interfaces (LIFs) and BDs.

```
device(config)# vlan 11-12
```

3. Configure the LIFs and BDs.

- a. Specify an Ethernet interface.

```
device(config)# interface ethernet 0/5
```

- b. Configure the parent interface as switchport.

```
device(config-if-eth-0/5)# switchport
```

- c. Specify trunk mode.

```
device(config-if-eth-0/5)# switchport mode trunk
```

- d. Enable the interface.

```
device(config-if-eth-0/5)# no shutdown
```

- e. Specify a service instance and enter LIF configuration mode.

```
device(config-if-eth-0/5)# logical-interface ethernet 0/5.1
```

- f. Specify an interface and create a dual-tagged (inner VLAN) VLAN.

```
device(config-if-eth-lif-0/5.1)# vlan 10 inner-vlan 1
```

The VLAN in the LIF configuration is for VLAN tag classification. This example shows a dual-tagged LIF being configured. The expected packet that enters through this port must be dual-tagged, without VLAN 10 and the inner VLAN 1, in order to be classified as a packet received for this LIF.

- g. (Optional) By default, the administrative state of the LIF is "no shutdown." To remove the port from participating in any data traffic without having to shut down the physical interface, enter the
- no**
- form of the
- shutdown (LIF)**
- command.

```
device(config-if-eth-lif-0/5.1)# no shutdown
```

- h. Repeat Step 3e through Step 3g for the second logical interface, and specify a second inner VLAN.

```
logical-interface ethernet 0/5.2
  vlan 10 inner-vlan 2
```

4. Create and configure a BD.

- a. Create BD 1.

```
device(config)# bridge-domain 1 p2mp
```

By default, the bridge-domain service type is point-to-multipoint (**p2mp**).

- b. Bind the logical interfaces for attachment circuit (AC) endpoints to the BD.

```
device(config-bridge-domain-1)# logical-interface ethernet 0/5.1
```

Logical interfaces representing BD endpoints must be created before they can be bound to a BD. For further information, refer to *Logical Interfaces*.

- c. Ensure that local switching is enabled for BD 1.

```
device(config-bridge-domain-1)# local-switching
```

Local switching is enabled by default.

- d. Enable the dropping of Layer 2 bridge protocol data units (BPDUs) for BD 1.

```
device(config-bridge-domain-1)# bpdu-drop-enable
```

A default pseudowire (PW) profile is automatically configured, with the following defaults:

```
Vc_mode = RAW Mode
mtu = 1500
mtu_enforce = NO
pw_profile_control_word = 0
pw_profile_flow_label = 0
```

- e. Repeat the above BD configuration for the second BD, as in the following example.

```
bridge-domain 2 p2mp
 logical-interface ethernet 0/5.2
 pw-profile default
 bpdu-drop-enable
 local-switching
```

5. Configure an overlay gateway.

- a. In global configuration mode, specify a gateway.

```
device(config)# overlay-gateway gw1
```

- b. Specify the type as Layer 2 extension.

```
device(config-overlay-gw-gw1)# type layer-2-extension
```

- c. Specify the LVTEP loopback interface.

```
device(config-overlay-gw-gw1)# ip interface loopback 2
```

- d. Configure the automatic mapping of VLANs/BDs to Virtual Network Identifiers (VNIs).

```
device(config-overlay-gw-gw1)# map vni auto
```

- e. Activate the gateway.

```
device(config-overlay-gw-gw1)# activate
```

6. In global configuration mode, enable EVPN configuration mode and configure the EVPN instance.

- a. Enter default EVPN configuration mode.

```
device(config)# evpn
```

Default mode is the only available mode.

- b. Enable the auto-generation of the import and export route-target community attributes for the default EVPN instance.

```
device(config-evpn-default)# route-target both auto
```

- c. Enable the auto-generation of a route distinguisher (RD) for the default EVPN instance.

```
device(config-evpn-default)# rd auto
```

- d. Add the BDs to the default EVPN instance.

```
device(config-evpn-default)# bridge-domain add 1-2
```

- e. Add the VLANs to the default EVPN instance.

```
device(config-evpn-default)# vlan add 11-12
```

7. Configure the cluster.

- a. In global configuration mode, specify an MCT cluster name and cluster ID (in this example, "c1" and "1", respectively) to enable cluster configuration mode.

```
device(config)# cluster c1 1
```

- b. Ensure that client-isolation loose mode is enabled.

```
device(config-cluster-1)# client-isolation-loose
```

By default, the node with the lower peer IP address is set to the client-isolation mode of loose. Note that loose mode is recommended for LVTEP. This mode ensures that both cluster nodes forward traffic received from the remote peer when the ICL link is down. Without this default configuration, one node becomes loose and one becomes strict (as a result of the peer IP address configuration). The node that becomes strict is not able to forward traffic to the CCEP, because the CCEP links are shut down.

- c. Specify a virtual Ethernet (VE) interface through which to reach the MCT cluster peer.

```
device(config-cluster-1)# peer-interface ve 2
```

- d. Specify the IP address of the MCT cluster peer.

```
device(config-cluster-1)# peer 7.7.100.7
```

- e. Deploy the cluster.

```
device(config-cluster-1)# deploy
```

- f. Exit to Privileged EXEC mode.

8. In Privileged EXEC mode, enter the **cluster management** command and specify a cluster.

```
device# cluster management 2
```

The cluster ID for each node in the cluster must be unique.

9. Configure BGP routing with neighbor and address-family attributes.

- a. In global configuration mode, enable BGP routing and enter BGP router configuration mode.

```
device(config)# router bgp
```

- b. Specify the autonomous system number (ASN) for the AS in which the remote neighbor resides.

```
device(config-bgp-router)# neighbor 7.7.100.7 remote-as 100
```

- c. Configure the BGP device to communicate with a neighbor through a specified interface, in this case loopback 1.

```
device(config-bgp-router)# neighbor 7.7.100.7 update-source loopback 1
```

- d. Repeat the above two substeps for the other peer address, as in the following example.

```
neighbor 8.8.100.8 remote-as 100
neighbor 8.8.100.8 update-source loopback 1
```

- e. Enable IPv4 and IPv6 unicast address-family.

```
device(config-bgp-router)# address-family ipv4 unicast
device(config-bgp-router)# address-family ipv6 unicast
```

10. Enable the L2VPN address-family configuration mode to configure a variety of BGP EVPN options.

- a. Enable L2VPN address-family configuration mode and enter BGP EVPN configuration mode.

```
device(config-bgp-router)# address-family l2vpn evpn
```

- b. Specify VXLAN encapsulation for the first peer.

```
device(config-bgp-evpn)# neighbor 8.8.100.8 encapsulation vxlan
```

- c. Enable the exchange of information with BGP neighbors and peer groups.

```
device(config-bgp-evpn)# neighbor 8.8.100.8 activate
```

- d. Repeat the above two substeps for the other peer, but specify MCT encapsulation, as in the following example.

```
neighbor 7.7.100.7 encapsulation mct
neighbor 7.7.100.7 activate
```

11. Repeat the above steps for the other node in the cluster, with modifications as appropriate.

LVTEP support for other features

AC LIF

AC LIFs of type untagged, single tagged, and double tagged are supported.

Layer 2 ACLs

Layer 2 end points (leaf) support Layer 2 ACLs.

Rate limiting

Layer 2 end points (leaf) do not support rate limiting.

QoS

QoS behavior is similar to that for a single VTEP gateway. Details are listed in the following tables for DiffServe tunneling uniform and pipe modes in an overlay network.

Table 14: QoS behavior for DiffServ tunneling uniform mode

Traffic type	VXLAN origination	VXLAN tunnel termination
VLAN cases (untagged)	DSCP 0 and TTL 255 are sent on the tunnel header.	NA
VLAN cases (tagged)	Priority Code Point (PCP) is mapped to IP DCSP of VXLAN tunnel and TTL is set to 255.	DSCP is remapped to PCP of L2 traffic.
BD (raw, single tagged)	PCP of VLAN is mapped to DSCP and TTL is set to 255.	DSCP is remapped to PCP of VLAN.
BD (raw, double tagged)	PCP of outer VLAN is mapped to DSCP and TTL is set to 255.	DSCP is remapped to PCP of both inner and outer VLAN. Original PCP inner VLAN is not retained.
BD (raw, untagged)	DSCP 0 and TTL 255 are sent on the tunnel header	NA
BD (tagged, single tagged)	PCP is mapped to DSCP and TTL is set to 255	DSCP is remapped to PCP of L2 traffic.

Table 14: QoS behavior for DiffServ tunneling uniform mode (continued)

Traffic type	VXLAN origination	VXLAN tunnel termination
BD (tagged, double tagged)	Outer VLAN PCP is mapped to DSCP and TTL is 255 Inner VLAN header is carried with original PCP.	DSCP is remapped to PCP of outer VLAN and Inner VLAN of L2 traffic.
BD (tagged, untagged)	DSCP 0 and TTL 255 are sent on the tunnel header. Dummy VLAN is added: VLAN ID is BD ID and PCP is 0.	NA

Table 15: QoS behavior for DiffServ tunneling pipe mode

Traffic type	VXLAN origination	VXLAN tunnel termination
VLAN cases (untagged)	DSCP 0 and TTL 255 are sent on the tunnel header.	NA
VLAN cases (tagged)	DSCP 0 and TTL 255 are sent on the tunnel header.	PCP value is set to 0 as inner VLAN is not carried.
BD (raw, single tagged)	DSCP 0 and TTL 255 are sent on the tunnel header.	PCP value is set to 0 as inner VLAN is not carried.
BD (raw, double tagged)	DSCP 0 and TTL 255 are sent on the tunnel header.	PCP value is set to 0 as inner VLAN is not carried. PCP is cleared for both tags.
BD (raw, untagged)	DSCP 0 and TTL 255 are sent on the tunnel header.	NA
BD (tagged, single tagged)	DSCP 0 and TTL 255 are sent on the tunnel header. VLAN is carried and PCP is carried.	PCP value is set to zero.
BD (tagged, double tagged)	DSCP 0 and TTL 255 are sent on the tunnel header. Inner VLAN is carried and PCP is carried.	PCP value is set to zero.
BD (tagged, untagged)	DSCP 0 and TTL 255 are sent on the tunnel header. Dummy VLAN is added: VLAN ID is BD ID and PCP is 0.	NA

MTU

MTU behavior is similar to that for a single VTEP gateway.

Inner packet tag behavior

Inner packet tag behavior is similar to that for a single VTEP gateway.

The following table summarizes this behavior for VLANs and BDs.

Table 16: Inner packet tag behavior for VLANs and BDs

VLAN/BD Configuration	Inner packet tag behavior	Remarks
VLAN	Inner packet is always untagged.	
BD raw mode	Inner packet is always untagged.	This is RFC-compliant behavior.
BD tagged mode	Inner packet is always tagged.	This mode can be used if a tag must always be sent in the inner packet.

Statistics

The LVTEP tunnel CCEP supports statistics. For BUM traffic, although the traffic is suppressed (through split horizon), it is still accounted for as part of the statistics. This behavior is the same as existing single VTEP behavior.

Nondefault TPID

The TPID field is located at the same position as the EtherType/length field in untagged frames, and is thus used to distinguish the frame from untagged frames. If you require support for dual tagging or provider backbone bridge (PBB), the outer TPID of the packet must be configured to a value different from the default.

The raw pass-through support for an untagged LIF requires you to configure the interface TPID to a value that allows it to treat all traffic received on that port as untagged. You can configure the TPID on port and LAG interfaces.



Important

When the tag type is changed on interface, the interface is brought down first, causing all learned MAC addresses to be flushed.

High availability (HA) MM failover supports the TPID feature. The TPID configuration is automatically synchronized to the standby module if a standby MM is installed on an SLX-OS device.

Hardware limitations

The TPID feature has the following limitations:

- **Tagging:** The TPID configuration is supported for an outer tag. If dual-tagging needs to be supported on the interface, the inner tag must be 0x8100.
- **TPID:** Up to four TPIDs are supported system-wide. One is used by default (TPID = 0x8100) and cannot be changed.
- **LSP FRR:** Hardware support for LSP FRR is available only for TPID 0x8100. If you require a label switched path with fast reroute (LSP FRR) configuration, note that none of the routable interfaces (whether a router port or a LIF of a VE) can have any nondefault TPID configuration, because FRR always assumes that the link layer has the default TPID of 0x8100.
- **OpenFlow:** If OpenFlow is enabled while the default TCAM profile is used, then no TPID can be configured. This restriction is also enforced in the opposite direction; if any port has TPID configured, then OpenFlow cannot be enabled while the default profile is used.

*Configuring a nondefault TPID***About This Task**

Perform the following steps to configure a nondefault TPID. The interface can be a port or a port-channel (LAG).

Procedure

1. Do the following to configure a nondefault TPID on an Ethernet interface.
 - a. Enter global configuration mode.

```
device# configuration terminal
```

- b. Enter interface configuration mode and specify an Ethernet interface.

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

- c. Use the **tag-type** command to configure the TPID.

```
device(config-if-eth-1/1)# tag-type 0x9100
```

By default, all interfaces in the system have default TPID value of 0x8100.

- d. Enter the **show interface ethernet** command to confirm the configuration.

```
device# show interface ethernet 1/1
Ethernet 1/1 is up, line protocol is up (connected)
Hardware is Ethernet, address is 609c.9f5f.5005
  Current address is 609c.9f5f.5005
Pluggable media present
Interface index (ifindex) is 203431936
MTU 1548 bytes
10G Interface
LineSpeed Actual      : 1000 Mbit
LineSpeed Configured : Auto, Duplex: Full
Tag-type: 0x9100
Priority Tag disable
Forward LACP PDU: Disable
Route Only: Disabled
Last clearing of show interface counters: 23:12:47
Queueing strategy: fifo
Receive Statistics:
  0 packets, 0 bytes
  Unicasts: 0, Multicasts: 0, Broadcasts: 0
  64-byte pkts: 0, Over 64-byte pkts: 0, Over 127-byte pkts: 0
  Over 255-byte pkts: 0, Over 511-byte pkts: 0, Over 1023-byte pkts: 0
  Over 1518-byte pkts(Jumbo): 0
  Runt: 0, Jabbers: 0, CRC: 0, Overruns: 0
  Errors: 0, Discards: 0
Transmit Statistics:
  0 packets, 0 bytes
  Unicasts: 0, Multicasts: 0, Broadcasts: 0
  Underruns: 0
  Errors: 0, Discards: 0
Rate info:
  Input 0.000000 Mbits/sec, 0 packets/sec, 0.00% of line-rate
  Output 0.000000 Mbits/sec, 0 packets/sec, 0.00% of line-rate
Route-Only Packets Dropped: 0
```

```
Time since last interface status change: 23:12:45
```

- e. To revert to the default TPID value, enter the **no tag-type** command.

```
device(conf-if-eth-1/1)# no tag-type
```

You can achieve the same result by configuring a tag-type of 0x8100.

2. Do the following to configure a nondefault TPID on a port-channel interface.
 - a. Enter global configuration mode.

```
device# configuration terminal
```

- b. Enter interface configuration mode and specify a port-channel interface.

```
device(config)# interface port-channel 20
```

- c. Use the **tag-type** command to configure the TPID.

```
device(config-Port-channel-20)# tag-type 0x88a8
```

By default, all interfaces in the system have default TPID value of 0x8100.

- d. Enter the **show interface port-channel** command to confirm the configuration.

```
device# show interface port-channel 20
Port-channel 20 is up, line protocol is up
Hardware is AGGREGATE, address is 609c.9f5c.ac07
  Current address is 609c.9f5c.ac07
Interface index (ifindex) is 671088660 (0x28000014)
Minimum number of links to bring Port-channel up is 1
MTU 1548 bytes
LineSpeed Actual      : 300000 Mbit
Allowed Member Speed : 100000 Mbit
Priority Tag disable
Forward LACP PDU: Disable
Route Only: Disabled
Tag-type: 0x88a8
Last clearing of show interface counters: 2d01h50m
Queueing strategy: fifo
Receive Statistics:
  34579 packets, 4201368 bytes
  Unicasts: 0, Multicasts: 34579, Broadcasts: 0
  64-byte pkts: 0, Over 64-byte pkts: 17288, Over 127-byte pkts: 17291
  Over 255-byte pkts: 0, Over 511-byte pkts: 0, Over 1023-byte pkts: 0
  Over 1518-byte pkts(Jumbo): 0
  Runts: 0, Jabbers: 0, CRC: 0, Overruns: 0
  Errors: 0, Discards: 0
Transmit Statistics:
  34578 packets, 4201240 bytes
  Unicasts: 0, Multicasts: 34578, Broadcasts: 0
  Underruns: 0
  Errors: 0, Discards: 0
Rate info:
  Input 0.000000 Mbits/sec, 0 packets/sec, 0.00% of line-rate
  Output 0.000000 Mbits/sec, 0 packets/sec, 0.00% of line-rate
Route-Only Packets Dropped: 0
Time since last interface status change: 1d23h59m
```

Configuring TCAM profiles to support LVTEP

About This Task



Important

This profile must be applied to all LVTEP configurations in this chapter.

Procedure

1. Enter global configuration mode and enter the **hardware** command.

```
device# configure terminal
device(config)# hardware
```

2. In hardware configuration mode, enter the **profile tcam** command and specify **vxlan-ext**.

```
device(config-hardware)# profile tcam vxlan-ext
```

3. Save the running configuration to the startup configuration.

```
device# copy running-config startup config
```

4. Reboot the device.

LVTEP show commands

Cluster show commands

show cluster

```
device# show cluster 1
Cluster cl 1
=====
Cluster State: Deployed
Client Isolation Mode: Loose
DF Hold Time: 3
Configured Member Vlan Range: 100-101
Active Member Vlan Range: 100-101
Configured Member BD Range: 1000-1001
Active Member BD Range: 1000-1001
No. of Peers: 1
No. of Clients: 2

Peer Info:
=====
Peer IP: 10.38.38.38, State: Up
Peer Interface: Not Configured
ICL Tunnel Type: VXLAN, State: Up

Client Info:
=====
Name          Id          ESI          Interface          Local/Remote
State
----          -
-----
c3            3           0:a:b:1:2:3:0:0:0  Ethernet 0/11     Up / Up
```

show cluster client

```

device# show cluster 1 client 513
Client Info:
=====
Client: tu61441, client-id: 513, Deployed, State: Up
Interface: Tunnel 61441
Vlans: 11-12
Bridge Domains: 1-2
Number of DF Vlans      : 0
Elected DF for vlans  :
Number of DF Bridge Domains : 0
Elected DF for Bridge Domains :

```

*MCT show commands**show cluster management*

```

device# show cluster management

Total Number of Nodes in Cluster   : 2

Node-Id      Switch MAC                IP Address      Status
-----
1            60:9C:9F:5A:44:14                4.4.100.4      Connected
2            >60:9C:9F:5A:00:14*              5.5.100.5      Co-ordinator

Please note that 5.5.100.5 (Node-id 2) is this node, 4.4.100.4 (Node-id 1) is remote node

```

*BGP show commands**show bgp evpn routes type ethernet-segment detail*

```

device# show bgp evpn routes type ethernet-segment detail
Total number of BGP EVPN Ethernet Segment Routes : 2
Status A:AGGREGATE B:BEST b:NOT-INSTALLED-BEST C:CONFED_EBGP D:DAMPED
E:EBGP H:HISTORY I:IBGP L:LOCAL M:MULTIPATH m:NOT-INSTALLED-MULTIPATH
S:SUPPRESSED F:FILTERED s:STALE
Route Distinguisher: 6.6.100.6:1
1 Prefix: ESR:[00.000808.640806.076443][IPv4:6.6.100.6], Status: BL, Age: 1h51m49s
NEXT_HOP: 0.0.0.0, Learned from Peer: Local Router
LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
AS_PATH:
Extended Community: ExtCom:06:02:06:07:08:08:64:08 ExtCom:
06:20:01:04:00:02:00:04
RT Import 1543:134767624
esi df_type : 1 expectedNoOfADroutes : 2 flag : 00000004 seqno : 4
Adj_RIB_out count: 1, Admin distance 0
RD: 6.6.100.6:1
Route Distinguisher: 7.7.100.7:1
2 Prefix: ESR:[00.000808.640806.076443][IPv4:7.7.100.7], Status: BI, Age: 1h51m48s
NEXT_HOP: 7.7.100.7, Metric: 1, Learned from Peer: 7.7.100.7 (100)
LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
AS_PATH:
Extended Community: ExtCom:06:02:06:07:08:08:64:08 ExtCom:06:20:01:0c:
00:02:00:04 ExtCom:03:0c:00:00:00:00:0a:00
RT Import 1543:134767624
esi df_type : 1 expectedNoOfADroutes : 2 flag : 0000000c seqno : 4
Extended Community: ExtCom: Tunnel Encapsulation (Type MPLS)
RD: 7.7.100.7:1

```

show bgp evpn routes type autodiscovery detail

```

device# show bgp evpn routes type autodiscovery detail
Total number of BGP EVPN AD Routes : 4
Status A:AGGREGATE B:BEST b:NOT-INSTALLED-BEST C:CONFED_EBGP D:DAMPED
      E:EBGP H:HISTORY I:IBGP L:LOCAL M:MULTIPATH m:NOT-INSTALLED-MULTIPATH
      S:SUPPRESSED F:FILTERED s:STALE
Route Distinguisher: 6.6.100.6:1
1      Prefix: AD:[00.000808640806076443][4294967295], Status: BL, Age: 1h52m34s
      NEXT_HOP: 0.0.0.0, Learned from Peer: Local Router
      LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
      AS_PATH:
      Extended Community: ExtCom:06:01:00:00:01:32:0c:00 RT 100:1073741835 RT
100:1073741836
      ESI Label Ext Community: 20057088      All-Active (0x00000000)
      Adj_RIB_out count: 1, Admin distance 0
      RD: 6.6.100.6:1
Route Distinguisher: 6.6.100.6:33
2      Prefix: AD:[00.000808640806076443][4294967295], Status: BL, Age: 1h52m34s
      NEXT_HOP: 0.0.0.0, Learned from Peer: Local Router
      LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
      AS_PATH:
      Extended Community: ExtCom:06:01:00:00:01:32:0c:00 RT 100:1073745921 RT
100:1073745922
      ESI Label Ext Community: 20057088      All-Active (0x00000000)
      Adj_RIB_out count: 1, Admin distance 0
      RD: 6.6.100.6:33
Route Distinguisher: 7.7.100.7:1
3      Prefix: AD:[00.000808640806076443][4294967295], Status: BI, Age: 1h52m36s
      NEXT_HOP: 7.7.100.7, Metric: 1, Learned from Peer: 7.7.100.7 (100)
      LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
      AS_PATH:
      Extended Community: ExtCom:06:01:00:00:01:32:0c:00 RT 100:1073741835 RT
100:1073741836 ExtCom:03:0c:00:00:00:00:0a:00
      ESI Label Ext Community: 20057088      All-Active (0x00000000)
      Extended Community: ExtCom: Tunnel Encapsulation (Type MPLS)
      RD: 7.7.100.7:1
Route Distinguisher: 7.7.100.7:33
4      Prefix: AD:[00.000808640806076443][4294967295], Status: BI, Age: 1h52m36s
      NEXT_HOP: 7.7.100.7, Metric: 1, Learned from Peer: 7.7.100.7 (100)
      LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
      AS_PATH:
      Extended Community: ExtCom:06:01:00:00:01:32:0c:00 RT 100:1073745921 RT
100:1073745922 ExtCom:03:0c:00:00:00:00:0a:00
      ESI Label Ext Community: 20057088      All-Active (0x00000000)
      Extended Community: ExtCom: Tunnel Encapsulation (Type MPLS)
      RD: 7.7.100.7:33

```

show bgp evpn routes type inclusive-multicast detail

```

device# show bgp evpn routes type inclusive-multicast detail
Total number of BGP EVPN IMR Routes : 12
Status A:AGGREGATE B:BEST b:NOT-INSTALLED-BEST C:CONFED_EBGP D:DAMPED
      E:EBGP H:HISTORY I:IBGP L:LOCAL M:MULTIPATH m:NOT-INSTALLED-MULTIPATH
      S:SUPPRESSED F:FILTERED s:STALE
Route Distinguisher: 6.6.100.6:32779
1      Prefix: IMR:[0][IPv4:6.6.100.6], Status: BL, Age: 1h54m39s
      NEXT_HOP: 0.0.0.0, Learned from Peer: Local Router
      LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
      AS_PATH:
      Extended Community: RT 100:1073741835
      Adj_RIB_out count: 2, Admin distance 0
      L2 Label: 11
      RD: 6.6.100.6:32779

```

```

Route Distinguisher: 7.7.100.7:32779
5   Prefix: IMR:[0][IPv4:7.7.100.7], Status: BI, Age: 1h53m37s
    NEXT_HOP: 7.7.100.7, Metric: 1, Learned from Peer: 7.7.100.7 (100)
    LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
    AS_PATH:
      Extended Community: RT 100:1073741835 RT 100:268435456 RT 100:0 ExtCom:03:0c:
00:00:00:00:0a:00
    PMSI Attribute Flags: 0x00000000 Label-Stack: 817163 Tunnel-Type: 6 Tunnel-
IP: 0.0.0.0
    Extended Community: ExtCom: Tunnel Encapsulation (Type MPLS)
    L2 Label: 817163
    RD: 7.7.100.7:32779

Route Distinguisher: 8.8.100.8:32779
9   Prefix: IMR:[0][IPv4:8.8.100.8], Status: BI, Age: 1h53m29s
    NEXT_HOP: 8.8.100.8, Learned from Peer: 8.8.100.8 (100)
    LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
    AS_PATH:
      Extended Community: RT 100:1073741835 RT 100:268435467 RT 100:11 ExtCom:03:0c:
00:00:00:00:08:00
    PMSI Attribute Flags: 0x00000000 Label-Stack: 11 Tunnel-Type: 6 Tunnel-IP:
8.8.100.8
    Extended Community: ExtCom: Tunnel Encapsulation (Type Vxlan)
    L2 Label: 11
    RD: 8.8.100.8:32779

```

show bgp evpn routes type mac detail

```

device# show bgp evpn routes type mac detail
Total number of BGP EVPN MAC Routes : 2
Status A:AGGREGATE B:BEST b:NOT-INSTALLED-BEST C:CONFED_EBGP D:DAMPED
      E:EBGP H:HISTORY I:IBGP L:LOCAL M:MULTIPATH m:NOT-INSTALLED-MULTIPATH
      S:SUPPRESSED F:FILTERED s:STALE
Route Distinguisher: 6.6.100.6:36865
1   Prefix: MAC:[0][0000.0607.0000], Status: BL, Age: 0h0m53s
    NEXT_HOP: 0.0.0.0, Learned from Peer: Local Router
    LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
    AS_PATH:
      Extended Community: RT 100:1073745921
      Adj_RIB_out count: 2, Admin distance 0
      L2 Label: 4097
      ESI : 00.00000000000000000000
      RD: 6.6.100.6:36865
Route Distinguisher: 8.8.100.8:32779
2   Prefix: MAC:[0][0000.0807.0000], Status: BI, Age: 0h5m16s
    NEXT_HOP: 8.8.100.8, Learned from Peer: 8.8.100.8 (100)
    LOCAL_PREF: 100, MED: 0, ORIGIN: incomplete, Weight: 0
    AS_PATH:
      Extended Community: RT 100:1073741835 RT 100:268435467 RT 100:11 ExtCom:03:0c:
00:00:00:00:08:00
    Extended Community: ExtCom: Tunnel Encapsulation (Type Vxlan)
    L2 Label: 11
    ESI : 00.00000000000000000000
    RD: 8.8.100.8:32779

```

Tunnel show commands

show tunnel brief

```

device# show tunnel brief
Tunnel 61441, mode VXLAN, node-ids 1
Admin state up, Oper state up

```



```
Source IP 6.7.100.67 ( Loopback 2 ), Vrf default-vrf
Destination IP 8.8.100.8
```

show tunnel

```
device# show tunnel 61441
Tunnel 61441, mode VXLAN, node-ids 1
Ifindex 0x7c00f001, Admin state up, Oper state up
Overlay gateway "gw1", ID 1
Source IP 6.7.100.67 ( Loopback 2 ), Vrf default-vrf
Destination IP 8.8.100.8
Configuration source BGP-EVPN
MAC learning BGP-EVPN
Tunnel QOS mode UNIFORM
Active next hops on node 1:
  IP: 10.6.8.8, Vrf: default-vrf
  Egress L3 port: Ve 3, Outer SMAC: 609c.9f5a.3d15
  Outer DMAC: 609c.9f5a.4515, ctag: 0
  BUM forwarder: yes

Packet count: RX 167993610      TX 995395
Byte count   : RX 29902862580  TX 226950060
```



QoS for VXLAN Layer 2 gateways

A VXLAN L2 gateway can interconnect tenants in the same subnet through the VXLAN configuration. The following figure illustrates a simple use case where a packet is sent from VM1. If it is a BUM packet, Leaf-A floods through the packet to all the VTEPs in the same VXLAN segment, until the MAC table at Leaf-A is populated with corresponding entries through mechanisms such as EVPN. Meanwhile, the known unicast packet is forwarded in unicast mode to the corresponding VTEP only. At Leaf-A, the Traffic Class/802.1p marking of the tenant frame determines the DSCP of the added IP header, and the IP header is followed by the VXLAN header. At terminating leaf nodes, the DSCP of the IP header is ignored and the Traffic Class/802.1p marking of the tenant frame is thereby exposed, and the original tenant frame is used to determine the QoS policy for switching the frame to the destination VM.

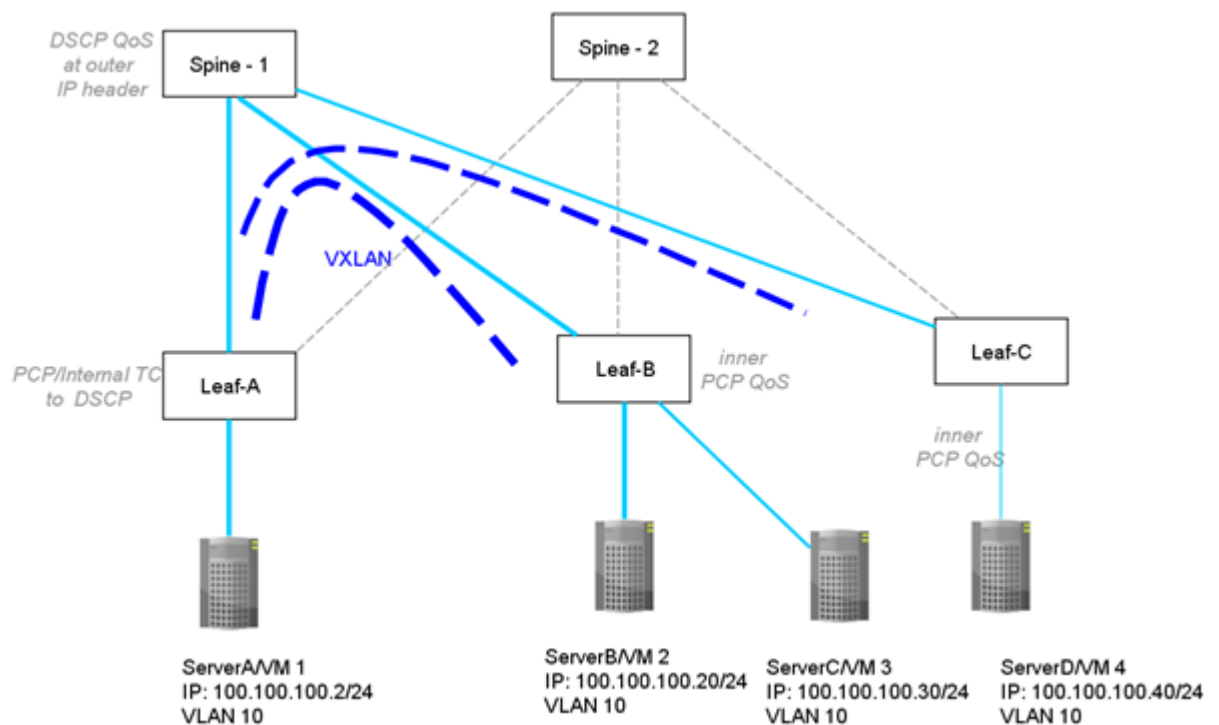


Figure 3: VXLAN L2 gateway interconnection of tenants in the same subnet

Configuring QoS for VXLAN Layer 2 gateways

Enter the **qos-ttl-mode uniform** command to set the QoS type mode to uniform, which is the default, when configuring the VXLAN L2 gateway. For example, enter the following commands when configuring the gateway for the ingress (tunneling) side of packet forwarding:

```
device# configure terminal
device(config)# overlay-gateway gateway_L2
device(config-overlay-gw-gateway_L2)# type layer2-extension
device(config-overlay-gw-gateway_L2)# ip interface loopback 1
device(config-overlay-gw-gateway_L2)# qos-ttl-mode uniform
device(config-overlay-gw-gateway_L2)# map vni auto
device(config-overlay-gw-gateway_L2)# activate
```



Note

For information about QoS configuration on VXLAN L3 gateways and on VLAN L2 and L3 gateway interconnections, refer to "QoS for VXLAN Layer 3 gateways" and "QoS for VXLAN Layer 2 and Layer 3 gateway interconnections".



Multi-Chassis Trunking (MCT)

[MCT Overview](#) on page 68

[MCT configuration considerations](#) on page 80

[Configuring the BGP EVPN peer](#) on page 82

[Configuring MCT](#) on page 84

[Configuring additional MCT cluster parameters](#) on page 86

[Configuring an auto-generated ESI for a cluster client](#) on page 87

[Displaying MCT information](#) on page 87

[Layer 3 routing over MCT](#) on page 89

[Using MCT with VRRP and VRRP-E](#) on page 92

[MCT use cases](#) on page 97

MCT Overview



Note

The SLX-OS device does not support Layer 2 protocols over MCT. You must provide a loop-free topology.



Note

MCT does not support any variant of Spanning Tree Protocol (STP). STP is disabled by default and must not be enabled with MCT.

In a datacenter network environment, LAG trunks provide link level redundancy and increased capacity. However, they do not provide switch-level redundancy. If the switch connected to the LAG trunk fails, the entire trunk loses network connectivity.

With MCT, member links of the LAG trunk are connected to two MCT-aware devices. A configuration between the devices enable data flow and control messages between them to establish a logical Inter-Chassis Link (ICL). In this model, if one MCT device fails, a data path remains through the other device.

SLX-OS Layer 2 MCT is based on RFC 7432 (BGP Ethernet VPN). The MP-BGP EVPN extension is the control plane on the SLX-OS device to perform both MAC synchronization and cluster management. MAC synchronization between the MCT peers synchronizes the MAC tables between the MCT nodes for node resiliency and faster convergence.

For the data plane, the SLX-OS device uses VXLAN encapsulation between the MCT Provider Edge devices (PEs).

SLX-OS MCT provides Layer 3 protocol support for IPv4 and IPv6 BGP, and OSPF through a VLAN or bridge domain VE interface. The VE over MCT interface MAC address is synchronized to the MCT peer through BGP by using an EVPN MAC route.

SLX-OS supports MCT over IPv4 or IPv6 Virtual Routing Redundancy Protocol (VRRP) and VRRP Extended (VRRP-E).

MCT terminology

<i>Inter-Chassis Link (ICL)</i>	VxLAN tunnel created between the MCT peer that forms the data path.
<i>Ethernet Segment (ES)</i>	Set of links that connect a client to the MCT nodes.
<i>Ethernet Segment Identifier (ESI)</i>	Globally unique 10-byte identifier for each ES.
<i>Designated Forwarder (DF)</i>	MCT PE that is elected to send BUM traffic to the client for a particular VLAN on an ES.
<i>Cluster Client Edge Port (CCEP)</i>	Client-facing ports or ports connecting dual-homed hosts on an MCT node.
<i>Cluster Edge Port (CEP)</i>	Orphan or edge ports connected to one of the MCT nodes.

SLX-OS MCT control plane

The control plane consist of the following components:

- EVPN instance—Mapped to a Layer 2 VLAN that the RFC refers to the VLAN-Based service interface.
- Ethernet segment ID (ESI)—10-byte integer that uniquely identifies the set of links connecting MCT PEs to the client CE. SLX-OS MCT supports both dynamic and static LAG between MCT PE and CE and uses ESI type 0 that is encoded as follows:
 - 1-byte ESI type = 0
 - 9-byte ESI value = user input through the SLX-OS **esi** command

When the client interface is a port channel and LACP is running on the port channel, MCT supports an automatically generated ESI value, as defined in RFC 7432. This ESI is encoded as type 1, as follows:

- 1-byte ESI type = 1
- 9-byte ESI value = 6-byte LACP system MAC address of the client followed by the 2-byte LACP port key, and then a 1-byte 0x00
- MP-BGP route distinguisher (RD)—Encoded using RD type 1 as defined in RFC 4364 that consists of the following subfields:
 - 4-byte administrator subfield that is set with the 4-byte router ID
 - 2-byte assigned number subfield that is encoded with the all zeros for the Ethernet segment (ES) route, client ID for the Ethernet Auto-Discovery route, and EVPN ID (VLAN ID) for MAC and multicast routes.
- MP-BGP EVPN capability—When a BGP session is brought up to a MCT peer, BGP indicates to the peer that it is EVPN capable using BGP capability advertisement with the following information:
 - Capability code = 1 (MP-BGP)
 - AFI = 25 (L2VPN)

- SAFI = 70 (EVPN)

If a BGP session already existed to the same peer, the existing BGP session is flapped to allow the advertisement of the EVPN capability.

- MP-BGP EVPN route types—Includes Ethernet Auto-Discovery (A-D), MAC/IP Advertisement, Inclusive Multicast Ethernet Tag, and Ethernet Segment routes.
- Designated forwarder—A PE in a set of multi-homing PEs connected to the same Ethernet segment that is elected for sending BUM traffic to a client for a VLAN ID on an Ethernet segment.

SLX-OS MCT operates in dual-homing mode.

Inter-chassis link

The underlay interface carrying the traffic can be any Layer 3 interface between the MCT peers. All EVPN VLANs or bridge domains (BDs) are extended to the MCT peer.

By default, VLAN-VNI mapping is automatically configured for the ICL VxLAN tunnel. Since a single VLAN-VNI mapping domain is supported, any change to this mapping under the overlay gateway changes the mapping for the ICL and temporarily affects its traffic.

MP-BGP EVPN Routes

RFC 7432 defines EVPN Network Layer Reachability Information (NLRI) with the format as shown in the following figure:

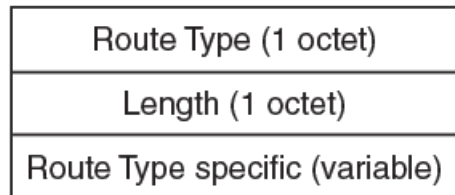


Figure 4: EVPN NLRI format

SLX-OS MCT supports the following route types.

Table 17: SLX-OS MCT route types

Route type	Route name	SLX-OS usage
1	Ethernet Auto-Discovery (for each Ethernet Segment only)	<p>Mass MAC address withdraw and Designated forwarder election.</p> <p>The PE can advertise multiple Ethernet A-D routes for each deployed client interface. Each A-D route has a maximum of 128 route targets (RTs).</p> <p>When the client interface goes down, the PE withdraws the Ethernet A-D route which is served as a trigger for the remote PE to remove all MAC addresses learned over the affected client interface instead of withdrawing an individual MAC address.</p>
2	MAC/IP Advertisement	<p>MAC synchronization of the MAC addresses between two MCT peers.</p> <p>Type 2 MAC/IP routes are also referred as ARP and Neighbor Discovery (ARP/ND) routes. ARP/ND addresses that are learned on the VLAN or bridge domain are automatically exported into BGP. Then BGP advertises the ARP routes to all of its EVPN peers. ARP/ND resolution is dependent on the MAC address. ARP/ND entries for Virtual IP and IP address configured on VE interfaces are automatically advertised.</p> <p>For more information on a Type 2 route, refer to <i>Extreme SLX-OS Layer 3 Routing Configuration Guide</i>.</p>
3	Inclusive Multicast Ethernet Tag	Advertisement of ingress replication usage and multicast label expected for each EVPN instance when receiving BUM traffic.
4	Ethernet Segment (ES)	<p>Designated forwarder election.</p> <p>The ES route is used to update the remote peer when the MCT client is deployed and undeployed.</p> <p>The ES route is used as an acknowledgment (ACK) mechanism to handle the client interface up scenario. The ACK ensures that the DF election does not occur on the node where the interface comes up until the A-D routes are processed by the remote peer.</p>
5	IP Prefix route	<p>This route is an extension to RFC 7432 described in the IETF draft, "draftietf-bess-evpn-prefix-advertisement." It is used to support advertisement of IP prefix routes. This route exchanges IPv4 and IPv6 routes in the overlay network.</p> <p>Prefix route exchange is supported over BGP-EVPN encapsulation and peering on the data plane.</p> <p>For more information on Type 5 route, refer to <i>Extreme SLX-OS Layer 3 Routing Configuration Guide</i>.</p>

Designated forwarder election

To avoid traffic duplication to clients, BUM traffic received by MCT nodes is forwarded by only one node that is elected as the designated forwarder (DF) for the VLAN or BD. The DF decides the forwarding

node for BUM traffic only by updating the MGID for the corresponding VLAN or BD. Unicast traffic is always locally forwarded if the local client port is up.

DF election is done based on the router IDs based on RFC 7432. For dual-homing, one node as the DF is for all odd VLANs or BDs and the other node as the DF is for even VLANs or BDs.

DF election is triggered only after the client is deployed on both MCT devices and in the following scenarios:

Table 18: DF election triggers

Local client state	Remote client state	DF election
Up	Undeployed	No DF election
Up (VLANs configured)	Up (no VLANs configured)	Local node DF for all VLANs
Up (VLANs configured)	Up (VLANs configured)	DF election done—Local DF for odd VLANs, remote DF for even VLANs
Up	Down	Local DF for all VLANs
Up	CCP Down (BGP session down)	Default isolation—Loose/loose Both nodes DF for all VLANs

To elect a designated forwarder (DF) for a VLAN ID on an Ethernet segment, each PE exchanges its router IP address with its multi-homing PEs through the Ethernet Segment route. The following algorithm uses the IP address to select the DF.

1. Upon the discovery of a new ES, a PE advertises the ES route and waits a default of three seconds for its peer to advertise the ES route.
2. When the timer expires, the PE builds a sorted list of PE IP addresses including its own address connected to the same ES.
3. The PE with the ordinal number that equals $(V \bmod N)$ is elected the DF. V is the VLAN ID and N is the number of PEs.
4. When the ES link fails, the PE withdraws its ES route which triggers the selection process to select a new DF. When a PE node failure occurs, DF election is also triggered when the PE is up and down.

MCT data plane

For the discussion of the MCT data plane, refer to the following topology diagram.

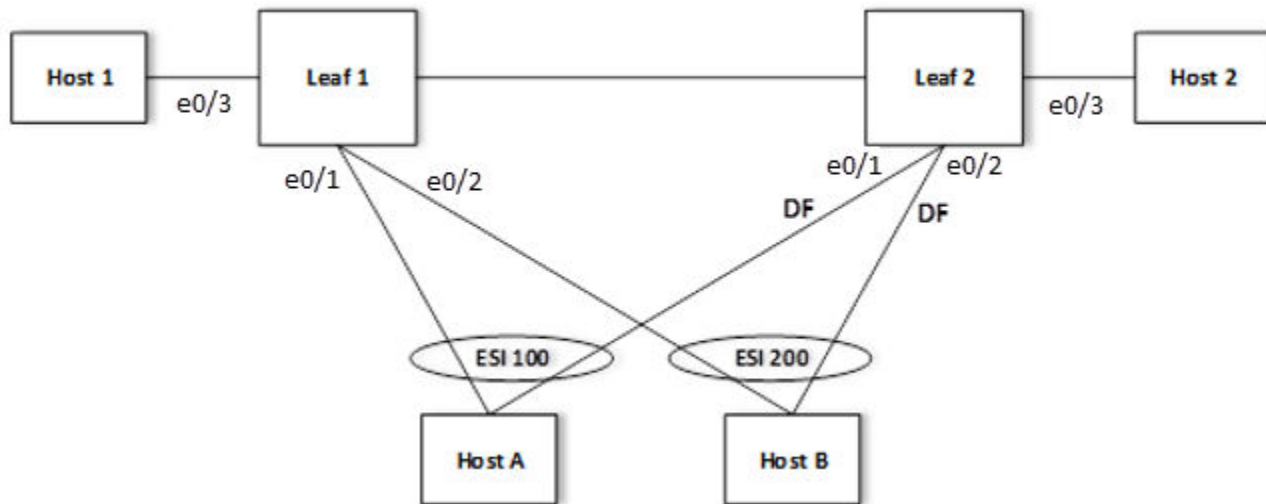


Figure 5: MCT data plane topology example

- Leaf 1 and Leaf 2 are MCT peers.
- Host A and Host B are client nodes configured with ESI 100 and 200, respectively. They are connected to Leaf1 and Leaf2 through cluster client edge ports (CCEPs) Ethernet 1/1 and 1/2.
- Host 1 and Host 2 are connected to edge ports and are not dual homed. They are connected to Leaf1 and Leaf2 through cluster edge port (CEP) Ethernet 1/3.
- Traffic for a single VLAN or BD is depicted where Leaf 2 is elected as the designated forwarder (DF).

Forwarding unicast traffic

For the unicast traffic from the CCEP to the CEP, refer to the following figure.

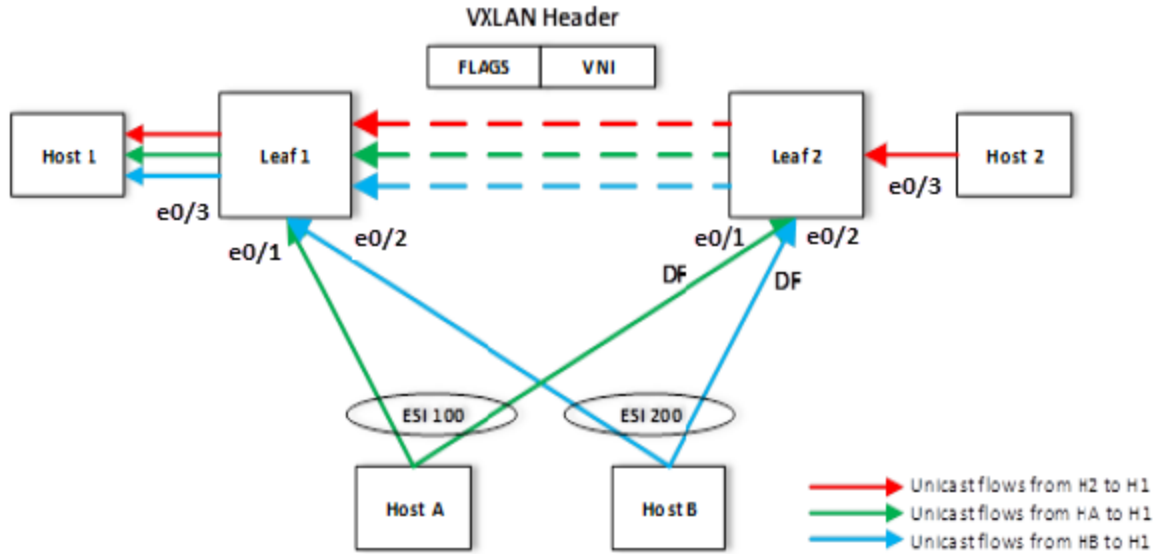


Figure 6: Unicast forwarding traffic from the CCEP to the CEP

- Unicast traffic from Host A and Host B to Host 1 can be hashed to Leaf 1 and Leaf 2.
- Leaf 1 learns the MAC address of Host 1. This traffic is locally switched.
- Leaf 2 has the remote MAC address of Host 1 as EVPN against the tunnel. This traffic is sent over the ICL.
- The MAC addresses of Host A and Host B are learned as CCL and CCR pointing to the local client port (Ethernet 1/1 and 1/2).

For the unicast traffic from the CEP to the CCEP, refer to the following figure.

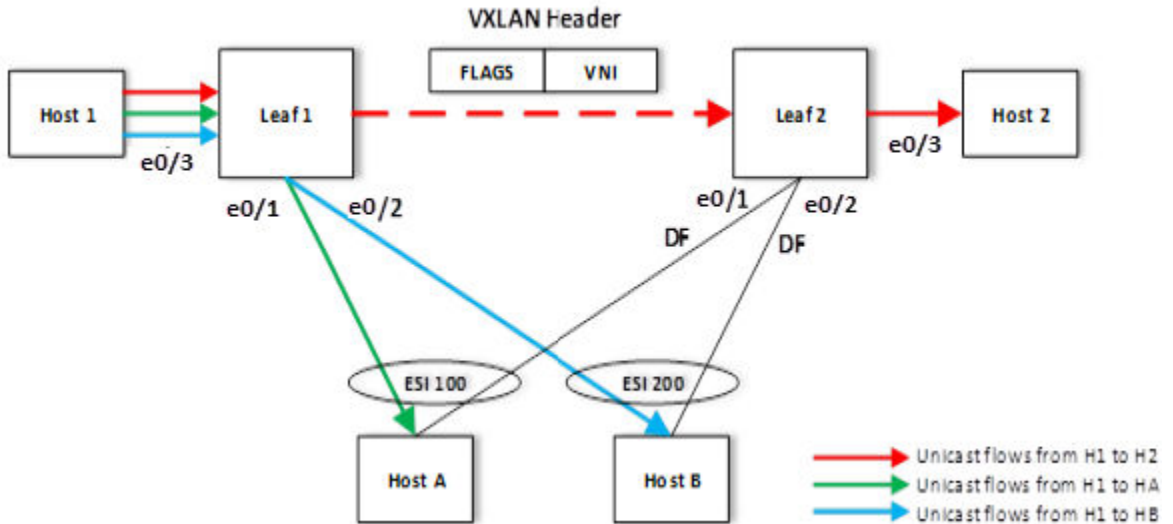


Figure 7: Unicast forwarding traffic from the CEP to the CCEP

- Leaf2 learns the MAC address of Host 2 that is synchronized to Leaf 1 as EVPN. The Unicast traffic from Host 1 to Host 2 is sent over the ICL tunnel to Leaf2 and then is locally switched.
- Based on the learned CCL and CCR MAC addresses, Leaf1 locally switches the unicast traffic from Host 1 to Host A and Host B. The designated forwarder does not affect the unicast traffic.

If the local client port goes down, the client MAC address is reprogrammed in the hardware to point to the ICL tunnel. Then the traffic is sent over the ICL to Leaf2 where the learned CCL and CCR MAC addresses ensure the traffic is switched to the client.

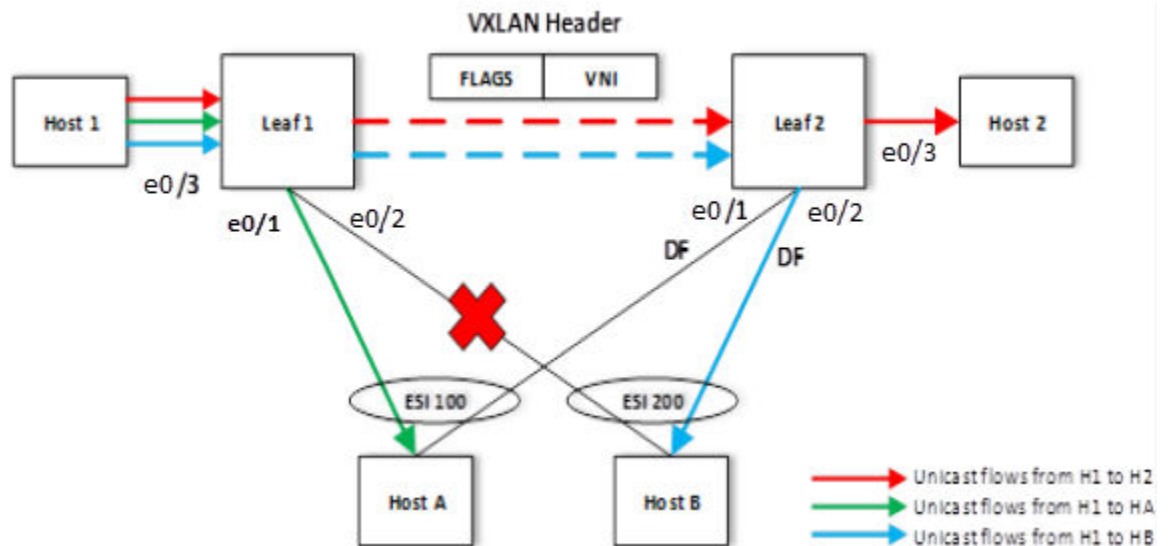


Figure 8: Unicast forwarding traffic from the CEP to the CCEP (local client down)

Flooding traffic from a CEP

For BUM traffic received from Host 1 (CEP) on Leaf 1:

- Leaf1 sends a copy to Leaf2 over the VXLAN tunnel. This copy is encapsulated with a VXLAN header with a VNI and default ESI of 0.
- Copies to Host A and Host B are suppressed on Ethernet 1/1 and 1/2 because Leaf1 is not the DF; the MGID is not programmed with Ethernet 1/2 and 1/2.
- Leaf 2 decapsulates the packet and floods the packet to Host 2, Host A and Host B because it is the DF for both clients based on the MGID membership (Ethernet 1/1, 1/2, and 1/3).

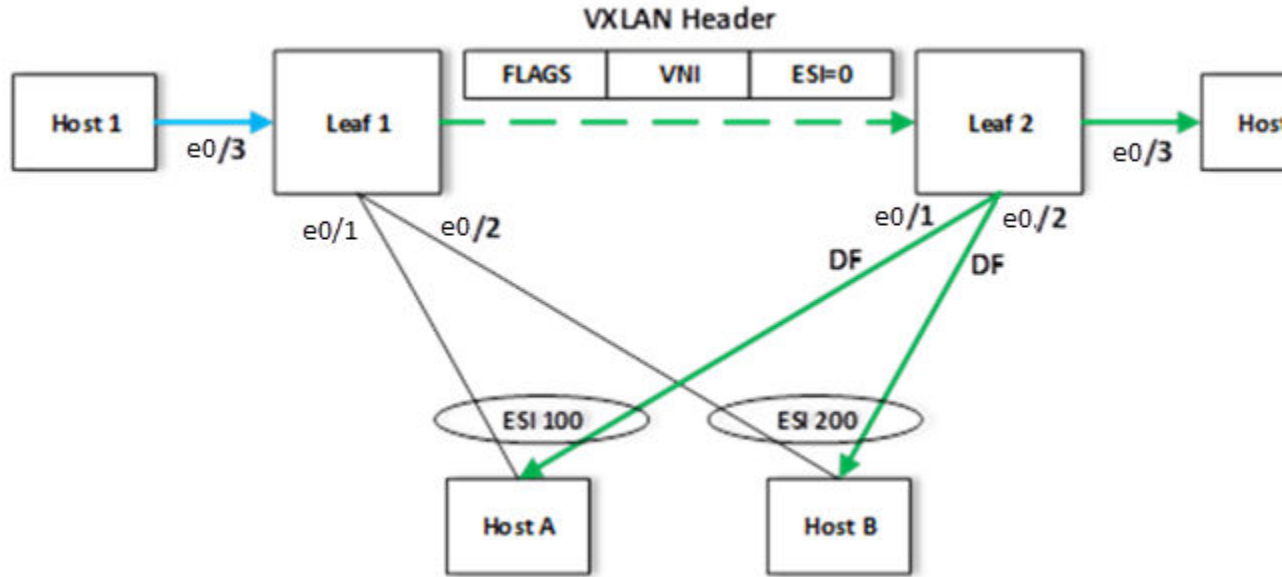


Figure 9: BUM forwarding between cluster edge ports

Flooding traffic received from the CCEP

Traffic received from CCEP Host A on Leaf 1 is flooded to Host 1 over Ethernet 1/3 (Layer 2 flooding), and to Leaf 2 over the VXLAN tunnel with ESI 100. Flooding is suppressed to Host A and Host B over Ethernet 1/1 and 1/2 because Leaf 1 is not the DF.

Then Leaf 2 floods the packet to Host 2 over Ethernet 1/3, and Host B over Ethernet 1/2 because Leaf 2 is the DF for the VLAN based on the MGID membership. A copy to Host A is suppressed due to the matching ESI stamping based on the ACL rules.

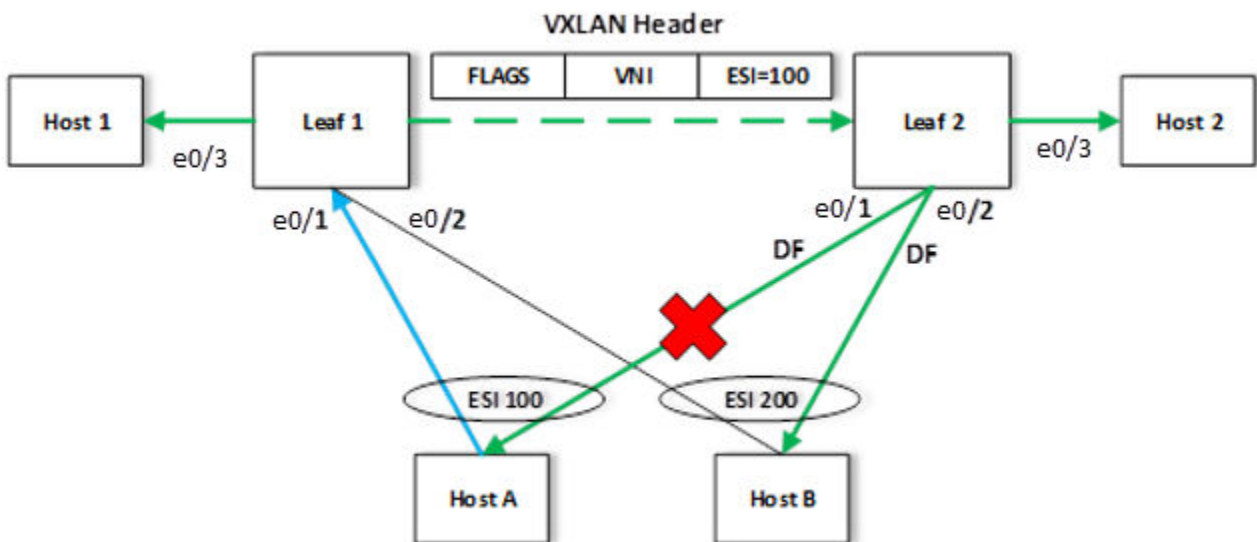


Figure 10: Flooding traffic from the CCEP with DF selection

MAC management

Each MAC entry maintains information about all the owners of the MAC entry who learned and advertised it. The information about each owner is maintained in the form of the MAC database (MDB).

A MDB entry contains peer and client information with the cost. A local MAC entry has cost 0 and MAC addresses learned from an MCT peer has a cost of 1. The MDB entry with the lowest cost is chosen as the filtering database (FDB).

In MCT topologies, MAC addresses can be learned either on local CCEP or CEP interfaces, or from the remote MCT node. If the same MAC is learned from both MCT nodes, then MAC entries learned locally have higher priority than the one learned from the remote peer.

For remote MAC addresses, aging is disabled, and they can only be deleted when delete notifications are received from the remote node that advertised it before for learning.

The following terminologies are associated with MCT MAC entries.

- Dynamic—MAC addresses learned locally as EVPN on CEP ports
- EVPN—MAC addresses learned on remote CEP ports
- Cluster Client Local (CCL)—MAC addresses learned locally on a client interface
- Cluster Client Remote (CCR)—MAC addresses learned on a remote client interface

Static MAC handling

Configuration of static MAC entries is allowed over MCT enabled VLANs and CEP and CCEP interfaces.

The MCT static MAC addresses configured on a local node are advertised to remote MCT node for learning. While advertising the MAC using the BGP MAC advertisement route, it uses the MAC mobility extended-community route to identify the MAC as static using the sticky MAC field. On the remote node, when MAC advertisement is received for a static MAC address, the sticky MAC information is saved along with the MAC entry.

When an MCT static MAC address is deleted, a BGP MAC withdrawal route is sent to the remote peer to delete the MAC entry from its database.

When a CEP interface is down and if any static MAC entries are present, MAC Delete messages are sent to the remote node to flush the entries.

When a CCEP interface is down and if any static MAC entries are associated with the client, the MAC addresses are moved to point to the remote MCT peer. The MAC addresses are moved back to the CCEP when the interface comes back up.

On a local MCT node, when a cluster is UP and you configure a static MAC on a CEP or CCEP interface, the node synchronizes the MAC address to the remote MCT node. The remote node processes the MAC address and adds it to the FDB. On the remote MCT node, you can configure the same MAC as the static MAC address for the client 1 CCEP interface since it is configured on the same client CCEP interface. No additional static MAC configurations on the remote node are required since the same MAC are already part of the local MCT node.

When the cluster is down on the local and remote MCT nodes, both nodes are independent as clusters that can be independently configured with the static MAC addresses for the CEP or CCEP interface. However, when the cluster is brought up, the static MAC addresses are synchronized from both nodes

and the addresses on the remote node are rejected since the local configuration takes precedence. The misconfiguration remains until you correct it.

MAC learning

MAC learning over CEP interfaces is like basic Layer 2 learning and the EVPN MAC advertisement route is sent to the cluster peer to synchronize the learned MAC address. MAC learning over CCEP interfaces is two-step process in which the MAC entry is added into the MDB first. The best MAC entry is chosen and installed into the FDB and the EVPN MAC advertisement route is sent to the cluster peer for synchronization of the learned MAC address.

The following rules are used for MAC learning:

- If a static MAC address is configured on the CEP port, it is learned as the Static and the EVPN MAC advertisement route message is sent to the peer. In this case, the ESI is set to NULL. On the peer MCT node, the EVPN MAC address programmed on the cluster peer is static towards the MCT peer.
- If a static MAC address is configured on the CCEP port, an EVPN MAC advertisement route is sent to the peer. In this case, the MAC entry is associated with the ESI of the MCT client. The peer MCT node programs the MAC address as static over the local CCEP interface.
- Dynamic MAC learning from the CEP is similar to basic Layer 2 MAC learning. An EVPN MAC advertisement route is sent to the peer. In this case, the ESI is set to invalid or NULL. On the peer MCT node, the EVPN MAC address programmed on the cluster peer is static towards the MCT peer.
- Dynamic MAC learning from the CCEP occurs as a CCL MAC. An EVPN MAC advertisement route is sent to the peer. In this case, the MAC entry is associated with the ESI of the cluster client. The peer MCT node programs the MAC address as static over the local CCEP interface.

MAC aging rules

The following rules are defined for MAC aging:

- The local MAC age over CEP interface is similar to the Layer 2 MAC age. After the local MAC delete, an EVPN MAC withdrawal route is sent to the MCT peer.
- The local MAC age over CCEP interface is considered aged only if all MCT nodes age out the entry. When the MAC that ages on one of the MCT node local MDB is deleted, if the remote MDB present MAC is reprogrammed as the CCR, else the MAC is removed from the local FDB, an EVPN MAC withdrawal route is sent to MCT peer.
- The remote MAC addresses of the EVPN and CCR that are learned through the EVPN MAC advertisement route does not age out. They can only be removed by the EVPN MAC withdraw messages from the peer.

MAC movement

A MAC address is considered to be moved when the same MAC address is received on a different interface with same VLAN. In MCT, a MAC movement is allowed on both local and remote nodes.

The following table describes the allowed MAC movements in MCT.

Table 19: MCT MAC movement

MAC movement scenario	Behavior
Local dynamic MAC move from CEP1 to the CEP2 edge interface on MCT1.	On local node MCT1, the MAC address is updated to point to the new CEP2 interface. There is no MAC route update required to the remote MCT node. As on the remote node, the MAC always point towards the MCT peer for all EVPN MAC addresses.
Local dynamic MAC move from CEP1 edge interface to the CCEP1 client interface on MCT1.	On local node MCT1, the MAC address is updated to point to the new client interface CCEP1. A MAC update route is sent with the new ESI of client 1. The remote node updates the MAC address to point to the CCEP of client 1.
CCEP1 interface (client 1) to CCEP2 interface (client 2) on MCT1.	On local node MCT1, the MAC address is updated to point to the new client interface CCEP2. A MAC update is sent with the new ESI of client 2 to the remote node. The remote node updates the MAC address to point to the CCEP of client 2.
Local dynamic MAC move from CCEP1 interface (client 1) to CEP1 edge interface on MCT1.	On local node MCT1, the MAC address is updated to point to the new edge interface CEP1. A MAC update is sent with the new ESI 0 to the remote node. The remote node MCT2 updates the MAC address pointing to the MCT1 node.
For a MAC learned on a CEP port locally (MCT1). Dynamic MAC move to a CEP port on the remote node (MCT2)	On the MCT2 node for the EVPN MAC learned from MCT1, it is considered as a MAC move when it is learned on a CEP port. The MAC is updated as local on MCT2 and now points to the Dynamic on the CEP port on MCT2 instead of pointing to MCT1 node MCT2 sends an updated MAC to MCT1. MCT1 updates the MAC as remote and points to the MCT2 .
For a MAC learned on a CEP port locally (MCT1). Dynamic MAC move to CCEP1 on MCT2.	On the MCT2 node for the EVPN MAC learned from MCT1, it is considered as a MAC move when the same MAC is learned on a CCEP1 port. The MAC is updated as CCL on MCT2 and now points to the local CCEP1 port on MCT2 instead of pointing to the MCT1 node. MCT2 sends a CCL MAC updated to MCT1. MCT1 updates the MAC as CCR and point to the CCEP1 port.
For a MAC CCL learned on a CCEP1 port locally (MCT1). Dynamic MAC move to CCEP2 on remote MCT2 node.	On the MCT2 node for the CCR MAC learned from MCT1 for client 1, it is considered as a MAC move when the same MAC is learned on client 2 over the CCEP2 port. The MAC is updated as CCL on MCT2 and now points to the local CCEP2 port on MCT2 instead of pointing to CCEP1. From MCT2, it sends a CCL MAC updated to MCT1. MCT1 updates the MAC as CCR and points to the CCEP2 port.
For a MAC CCL learned on a CCEP1 port locally (MCT1). Dynamic MAC move to CEP port on remote MCT2 node.	On the MCT2 node for the CCR MAC learned from MCT1 for client1, it is considered as a MAC move when the same MAC is learned on the CEP port. The MAC is updated as Dynamic on MCT2 and points to the local CEP port on MCT2 instead of pointing to CCEP1. From MCT2, it sends the EVPN MAC updated to MCT1. MCT1 updates the MAC as EVPN and points to the MCT2.

MAC address deletion

The following rules are defined for MAC address deletion. Note that every deletion triggers the MAC resolution algorithm and reprograms the MAC entry if required.

- If the CCEP interface is down, MAC addresses are deleted locally and individual MAC deletion messages are sent to the peer.
- If the CCEP local port is down and the remote CCEP is down, MAC addresses are deleted locally and the ESI withdraw message is sent to the MCT peer instead of sending individual MAC delete messages.
- If the CCEP local port is down and the remote CCEP is up, all local MAC addresses are moved to point to the remote MCT peer including the static MAC addresses associated with the CCEP.
- When the client entry is undeployed, all MAC addresses are deleted locally, and the ESI withdraw message is sent to the MCT peer to delete all associated client MAC addresses.

MCT configuration considerations

General considerations

- MCT Peers need to run the same version of SLX-OS.
- When the primary physical interface is also an MCT cluster client interface, logical interfaces that belong to that primary interface cannot be in the same MCT bridge domain (BD).
- MCT does not support any variant of Spanning Tree Protocol (STP) on ICL links, cluster client ports, or any VLAN that is part of the cluster. STP is disabled by default.



Note

If you enable STP on MCT nodes, each node acts as an independent (not interconnected) STP switch. This situation results in STP state flap at the node connected to the CCEP port, because it receives two different Bridge Protocol Data Units (BPDU) on that CCEP port

- SLX-OS supports dynamic and static LAG between the MCT PE and CE.
- SLX-OS uses Ethernet segment identifier (ESI) type 0 (static LAG) and type 1 (dynamic LAG). Consider the following for ESI type 0:
 - Configure the 9-byte ESI value that is used to form a globally unique 10-byte integer ESI.
 - Configure the same 9-byte ESI value for each client on both MCT devices.
- SLX MCT configurations (Layer 2 or Layer 3) do not require the Advance Feature license.

Peer considerations

- For both MCT peers, the MCT peer address must match the peer's BGP router ID (loopback address).
- You must configure the same client ID on both MCT peers for the link or CCEP that is connected to the same client.
- Layer 3 connectivity should allow a BGP session to be established between MCT peers.
- SLX MCT peers must be physically connected to each other.
- You must configure and activate a BGP EVPN neighbor as the peer interface.
- (Required) Prevent native VLAN traffic from going across the ICL by disabling the default native VLAN on the ICL trunk port between the MCT peers. To prevent native VLAN traffic, use the

switchport mode trunk-no-default-native command from interface configuration mode. For example:

```
device(conf-if-eth-0/17)# switchport mode trunk-no-default-native
```

- MCT MPLS interfaces are automatically created when you configure MCT without the use of router MPLS. By design, automatically created MCT MPLS interfaces do not offer FRR (fast reroute) protection, because it is difficult to remove the FRR configuration from outside of MPLS.



Note

You need an Advance Feature license if you want to enable FRR for non-MCT paths

VLAN considerations

- To avoid traffic looping, consider these guidelines:
 - Overlay EVPN VLANs or BDs must not be natively configured on the peer interface (underlay interface).
 - If you use a VE as a peer interface, use the corresponding VLAN for underlay alone and do not extend it into the EVPN domain.
- The peer interface is in a separate control VLAN that is not used for MCT members
- If you configure an MCT port channel for multiple VLAN or VE interfaces, use static routes instead of OSPF ECMP.
 - For example, you have two VE interfaces, VE 10 (an MCT VLAN) and VE 20 (a non-MCT VLAN). When OSPF is configured on the interfaces, the route to the BGP router-id (or MCT) peer address is through the interfaces. If you then configure BGP BFD and BGP EVPN sessions between the MCT peers for faster recovery, BFD flaps continuously. Flapping occurs because OSPF does per-packet load balancing of BGP BFD packets, which are not reassembled properly.
 - As a best practice, configure a static route for MCT peers through the MCT VLAN (VE 10 in this example). A static route has a lower administrative distance than OSPF and places only one route in the routing table.

LACP considerations

- When the client interface is a port channel on which LACP is running, MCT supports an automatically generated ESI value, as defined in RFC 7432.
- To ensure that two MCT peers send the same system ID and key but a different port ID to each client, you must configure the same cluster ID and client ID on both nodes. The system ID is derived by appending a cluster ID to the MCT base system ID, which is a device-defined value.

```
System ID = mct_base_system_id (0180.c200) + cluster_id
```

- The LACP fields have the following settings:

```
Key = MCT_LACP_KEY_BASE (3000) + client_ID
```

```
Port ID (16-bit unique value) =  
5-bit (slot value) + 8-bit (port value + 3-bit) (MCT position offset)
```

LSP considerations

LSP is automatically created when a cluster is deployed. No user configuration is required.

- The SLX-OS device uses the peer interface to set up the MPLS data path automatically. The LSP is the outgoing peer interface and has an MCT prefix when it is displayed by the **show mpls lsp** command.
- The configured peer interface must be the best next hop to reach the MCT peer. If not, the automatically created LSP fails to come up.
- The automatically created LSP can co-exist with other explicit MPLS configurations. You can configure more LDPs and RSVP LSPs, but not with the same name as the automatically created LSP.
- When co-existence occurs on the device:
 - As part of the MCT cluster configuration, the MPLS daemon automatically creates and signals the RSVP LSP with the default template and a unique auto-generated name.
 - If you create an MCT LSP with the same name as an existing LSP, then the MCT LSP is created with a similar name but with a different number suffix to make it unique. For example:
`MCT_20.21.22.23_9876543_123456`.
 - You cannot create an LSP with same name as an existing MCT LSP. The attempted configuration is rejected with an error message.
 - When the router port or the VE interface is MPLS-enabled from the MCT configuration, you cannot disable the router port or the VE interface. The configuration to disable the router port or the VE interface from the CLI is rejected with an error message.
 - You can disable MPLS with the **no router mpls** command. Upon command execution, the automatically created LSP and the MPLS-enabled peer-interface configuration are disabled and re-enabled.
- An automatically created LSP is restricted to using the configured MCT peer interface as its next hop connection to the MCT peer. The IGP shortest path between the two peers must match the configured MCT peer interface. Otherwise, the LSP remains operationally down.
 - For example, suppose there are two paths between the MCT peers, eth 1/1 and eth 1/2, and the node configuration for the two peers is symmetrical. Furthermore, the user has configured eth 1/1 as the MCT peer interface.
 - If the IGP SPF path between the two peers uses eth 1/1 as the next hop link, then the LSP becomes operational and uses eth 1/1 as the next hop interface. If the IGP cost of eth 1/1 increases such that eth 1/2 is selected by the IGP SPF calculation, then the LSP remains operationally down.
 - You can use eth 1/2 as a normal MPLS interface. Do not use eth 1/1 (which is the MCT interface) for normal MPLS purposes.

Configuring the BGP EVPN peer

Before You Begin

Before configuring a BGP EVPN peer, ensure you configure a loopback interface.

About This Task

You also create an EVPN instance to add the bridge domains and VLANs associated with the MCT cluster.

Procedure

1. In privileged EXEC mode, enter global configuration mode.

```
device# configure terminal
```

2. Enable BGP routing.

```
device(config-terminal)# router bgp
```

3. Configure the EVPN peer with the autonomous system number (ASN).

```
device(config-bgp-router)# neighbor 10.1.1.1 remote-as 100
```

4. Configure the EVPN peer through the loopback interface.

```
device(config-bgp-router)# neighbor 10.1.1.1 update-source loopback 1
```

5. Enter EVPN address family configuration mode

```
device(config-bgp-router)# address-family l2vpn evpn
```

6. Configure MCT encapsulation on the peer.

```
device(config-bgp-evpn)# neighbor 10.1.1.1 encapsulation mct
```

7. Activate the EVPN peer.

```
device(config-bgp-evpn)# neighbor 10.1.1.1 activate
```

8. Access global configuration mode.

```
device(config-bgp-router)# exit
```

9. Access EVPN configuration mode.

```
device(config)# evpn myinstance
```

10. Configure the bridge domains.

```
device(config-evpn-myinstance)# bridge-domain add 1-2000,4000-4096
```

For more information on the bridge domain configuration, see the Bridge Domain chapter.

11. Configure the VLANs.

```
device(config-evpn-myinstance)# vlan add 1-2000
```

12. Enable auto-generation of a route distinguisher (RD) for this EVPN instance.

```
device(config-evpn-myinstance)# rd auto
```

13. Enable the auto-generation of the import and export route-target community attributes and ignore the autonomous system (AS) number.

```
device(config-evpn-myinstance)# route-target both auto ignore-as
```

Example

The following example shows the steps in the previous configuration.

```
device# configure terminal
device(config-terminal)# router bgp
device(config-bgp-router)# neighbor 10.1.1.1 remote-as 100
device(config-bgp-router)# neighbor 10.1.1.1 update-source loopback 1
device(config-bgp-router)# address-family l2vpn evpn
device(config-bgp-evpn)# neighbor 10.1.1.1 encapsulation mct
device(config-bgp-evpn)# neighbor 10.1.1.1 activate
device(config-bgp-router)# exit
device(config)# evpn myinstance
device(config-evpn-myinstance)# bridge-domain add 1-2000,4000-4096
device(config-evpn-myinstance)# vlan add 1-2000
device(config-evpn-myinstance)# rd auto
device(config-evpn-myinstance)# route-target both auto ignore-as
```

Configuring MCT

Before You Begin

Before configuring MCT, ensure that the following configurations exist:

- Layer 3 interface if you configure an optional cluster peer interface
- VLANs and bridge domains under an EVPN configuration to function as the MCT members
- Port channel for Link Aggregation or an Ethernet interface as a client interface
- Disable the default native VLAN on the trunk port between the MCT peers by using the interface configuration mode **switchport mode trunk-no-default-native** command. For example:

```
device(conf-if-eth-0/17)# switchport mode trunk-no-default-native
```



Note

A peer interface is in a separate VLAN that is not an MCT member. Do not configure VLANs and bridge domains configured under EVPN on the trunk port between the MCT peers.

About This Task

Perform the following steps.

Procedure

1. In privileged EXEC mode, enter global configuration mode.

```
device# configure terminal
```

2. Create a cluster on the device.

```
device (config)# cluster MCT1 1
```

3. Configure the peer IP address.

```
device(config-cluster-1)# peer 10.1.1.1
```

The peer IP address must be the remote peer's router ID. This address corresponds with the neighbor in BGP EVPN address family configuration for the peer.

4. Optionally, configure the peer interface.

```
device(config-cluster-1)# peer-interface Ve 10
```

The peer interface should be a valid Layer 3 interface. You should configure the peer interface before deploying the configuration.

5. Deploy the cluster.

```
device(config-cluster-1)# deploy
```

6. Create the client for the cluster and access cluster client configuration mode.

```
device(config-cluster-1)# client MCT1-client 200
```

On both MCT nodes, you must configure the same client ID.

7. Configure the interface to the cluster client instance.

```
device(config-cluster-client-200)# client-interface port-channel 3
```

The port channel specifies the LAG ID.

The client interface can also be a physical interface, for example:

```
device(config-cluster-client-200)# client-interface Ethernet 2/5
```

The client interface cannot be added under multiple client entries.

- Set the 9-octet Ethernet Segment ID (ESI) value which is used to uniquely identify the cluster client.

```
device(config-cluster-client-200)# esi 00.a1.b2.c3.d4.e5.f6.89.00
```

You must configure the same value on both MCT nodes to create the MCT client LAG.

The ESI cannot be added under multiple client entries.

- Deploy the cluster client.

```
device(config-cluster-client-200)# deploy
```

- After configuring the local MCT cluster and client, configure the remote MCT cluster and client.

Example

The following example is the steps in the previous configuration.

```
device# configure terminal
device (config)# cluster MCT1 1
device(config-cluster-1)# peer 10.1.1.1
device(config-cluster-1)# peer-interface Ve 10
device(config-cluster-1)# deploy
device(config-cluster-1)# client MCT1-client 200
device(config-cluster-client-200)# client-interface port-channel 3
device(config-cluster-client-200)# esi 00.a1.b2.c3.d4.e5.f6.89.00
device(config-cluster-client-200)# deploy
```

Taking the MCT node offline for maintenance

About This Task

If you need to take an MCT device offline for maintenance or an upgrade, perform the following steps to minimize traffic loss.

Procedure

- Verify that the MCT node that is peer to the node being taken offline is in loose client-isolation mode.

```
device# show cluster 1
Cluster MCT1 1
=====
Cluster State: Deploy
Client Interfaces Shutdown: FALSE
Client Isolation Mode: Strict
DF Hold Time: 3
Configured Member Vlan Range: 2, 4-7
Active Member Vlan Range: 2, 4-7
...
```

- If the peer node is in strict client-isolation mode, configure it to loose mode.

```
device(config)# cluster MCT1 1
device(config-cluster-1)# client-isolation loose
```

- Disable the MCT clients from the MCT node that you will take offline, as shown in the following example.

```
device(config)# cluster MCT1 1
device(config-cluster-1)# client-interfaces-shutdown
```

- Isolate the MCT node that you will take offline from the core of the network by shutting down all uplink interfaces.

What to Do Next



Note

Do not write the configuration changes made in the previous steps to the startup-configuration file.

To bring the MCT node back online, perform one of the following actions.

- If you upgraded or downgraded the device, select the **coldboot** option under the firmware download menu.
- For any other reason, execute the **reload system** command. Since the changed configuration was not saved, the reload reverts the configuration changes that had taken the MCT node offline.

Configuring additional MCT cluster parameters

Changing the client-isolation mode

Isolation mode defines the action to be taken when the BGP control session goes down between the MCT nodes while the cluster is in deployed state. When the client-isolation mode is strict, the client interface will be shutdown.

By default, client-isolation mode is loose. In loose mode, both peers act as DF masters. When the EVPN control session goes down, the peer device performs the master/slave negotiation. After negotiation, the slave shuts down its peer ports, and the master peer ports continue to forward the traffic.

You can configure strict mode. In this mode, when the EVPN control session goes down, the interfaces on both the cluster devices are administratively shut down. In strict mode, the client is completely isolated from the network if the control session is not operational.

Use the **client-isolation strict** command to configure the strict mode on both nodes, as shown in the following example.

```
device(config-cluster-1)# client-isolation strict
```

Use the **client-isolation loose** command to configure the loose mode on both nodes, as shown in the following example.

```
device(config-cluster-1)# client-isolation loose
```

Changing the designated-forwarder hold timer value

Upon expiration of the designated-forwarder hold timer, the reelection of the designated forwarder is considered.

By default, the hold time is three seconds. Use the **designated-forwarder-hold-time** command to change the time in seconds from 1 to 60 seconds, as shown in the following example.

```
device(config-cluster-1)# designated-forwarder-hold-time 35
```

Moving the traffic from an MCT node to the remote node

Use the **client-interfaces-shutdown** command to move all the traffic on the node to the remote MCT node by disabling the local client interfaces administratively, as shown in the following example.

```
device(config-cluster-1)# client-interfaces-shutdown
```

Configuring an auto-generated ESI for a cluster client

The following example shows how to configure an auto-generated ESI for the cluster client.

```
device(config)# cluster MCT1 1
device(config-cluster-1)# client MCT1-client 200
device(config-cluster-client-200)# esi auto lacp
```

Displaying MCT information

To display the EVPN neighbor information, use **show ip bgp neighbors** command. This information includes the peer configured for the EVPN address family, the undeployed MCT cluster, and the negotiation of the EVPN address family.

Displaying the cluster information

The following example shows the information of the cluster on the SLX-OS device.

```
device# show cluster 1

Cluster MCT1 1
=====
Cluster State: Deploy
Client Isolation Mode: Loose
DF Hold Time: 3
Configured Member Vlan Range: 2, 4-7
Active Member Vlan Range: 2, 4-7
Cluster Control Vlan: 4090
Configured Member BD Range:
Active Member BD Range:
No. of Peers: 1
No. of Clients: 2

Peer Info:
-----
Peer IP: 10.10.10.20, State: Up
Peer Interface: Not Configured
ICL Tunnel Type: VXLAN, State: Up

Client Info:
-----
Name      Id  ESI                               Interface      Local/Remote State
c1        1  0:11:22:33:80:0:0:0:0:0          Ethernet 0/3   Up/Up
c1        2  0:11:22:33:81:0:0:0:0:0          Port-channel 20 Up/Up
```

Displaying the cluster client information

The following example displays all client information for cluster 1.

```
device# show cluster 1 client
Client Info:
-----
Name      Id    Label (Lo/Re)  Interface      Local/Remote State
access1   100   NA/ 798721     Ethernet 1/8   UnDep/Dep
access2   200   798722/798722 Port-Channel 3   Up/UP
access3   300   798723/798723 Ethernet 2/3    Down/Up
```

The following example displays client 100 information for cluster 10.

```
device# show cluster 10 client 100
Client Info:
-----
Client: access1, client-id: 100, Deployed, State: Up
Interface: Ethernet 1/8
Vlans : 1-10, 100
Elected DF for vlans: 2, 4, 8, 10, 100
```

Displaying member VLAN information

The following example displays the member VLAN information for the cluster.

```
device# show cluster member vlan
VLAN-ID    Mcast-label (Lo/Re)    Unicast-label (Lo/Re)    Forwarding State
101        817253 / 817253        800869 / 800869         Up
102        817254 / 817254        800870 / 800870         Up
103        817255 / 817255        800871 / 800871         Up
104        817256 / 817256        800872 / 800872         Up
105        817257 / 817257        800873 / 800873         Up
106        817258 / 817258        800874 / 800874         Up
```

Displaying and clearing the MAC address table cluster information

The following example shows how to display the MCT cluster information in the MAC address table.

```
device# show mac-address-table cluster 1
Vlan/Bd'Id Mac-address      Type      State   Ports
100 (V)    0010.a111.aaaa   CCL       Active  ETH 3/1
100 (V)    0010.a111.aa22   Static-CCL Active  ETH 3/1
100 (V)    0010.a111.bbbb   CCR       Active  ETH 3/1
200 (V)    003d.a111.1111   Dynamic   Active  Eth 1/1
200 (V)    003d.a111.1122   Static    Active  Eth 1/1
200 (V)    003d.a111.3333   EVPN     Active  10.2.2.2
200 (V)    003d.a111.3322   EVPN-Static Active  10.2.2.2
```

The MAC Type for an MCT cluster displays the following information:

- For the client MAC behavior, MAC addresses are learned as CCL on the local MCT node and CCR on the remote MCT node pointing to the CCEP interface.
- Static MAC addresses configured on CEP AC endpoints are learned as Static. The corresponding remote MAC addresses are learned as EVPN-Sticky in the remote node.
- For static MAC addresses over client interfaces, Static-CCL and CCR are displayed.

You can also view the MAC entries for a specific client.

Clearing the MCT cluster MAC table entries

You can clear all cluster entries from the MAC address table or the entries for a specified client. The following example clears the MAC entries for client 3 of cluster 1.

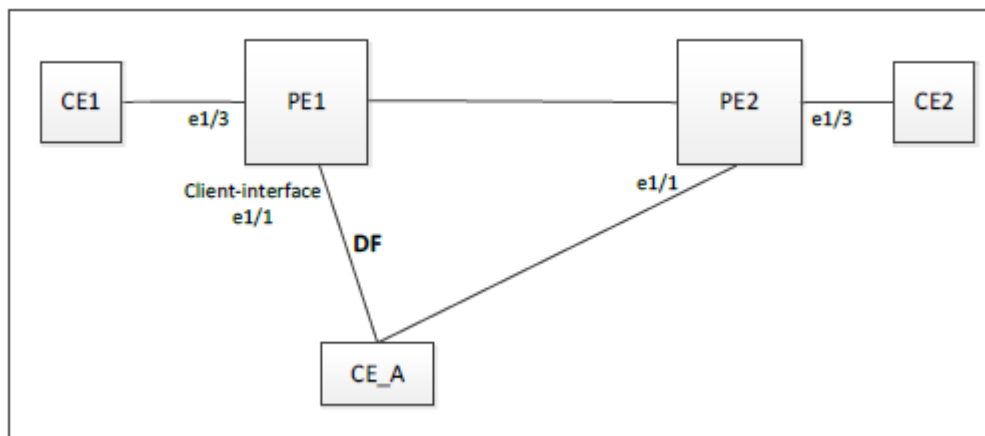
```
device# clear mac-address-table cluster 1 client 3
```

Only the local MAC entries are deleted from the current node. Individual MAC withdrawal flush messages are sent through the EVPN. However, BGP still batches multiple routes to the remote node.

When the remote MCT peer receives the MAC withdrawal message, it only deletes the remote MAC entry. To clear MAC addresses on both nodes, you must issue **clear mac-address-table** commands on both MCT nodes.

Layer 3 routing over MCT

The following diagram is the Layer 3 MCT data plane.



All devices are on the same VE and receive protocol Hello packets. Over the ICL link, the following packets are flooded:

- Hello and multicast packets from PE1 and PE2
- ARP and ND6 packets

Any two devices have direct Layer 3 communications.

- IP traffic is sent to the DA MAC of the target IP address or next-hop IP address.
- Traffic from CE_A to PE1 or CE1 might be sent through the PE2 link first, and switched by PE2 over the ICL link.

Configuration considerations

You must first create the VE interface for the MCT VLAN or bridge domain on the MCT pair.

Enabling L3 protocols is the same as enabling them on a VE interface.

For routes learned over Layer 3 protocols, the next-hop IP address is usually the peer IP address and not necessarily the MCT router address.

To bind the VE interface to an MCT VLAN or bridge domain, use the **router-interface ve** command under VLAN or bridge-domain configuration mode, respectively. The following example shows how to bind VE 200 to bridge domain 2.

```
device# configure terminal
device(config)# bridge-domain 2
device(config-bridge-domain-2)# router-interface ve 200
```

Layer 3 MCT VLAN configuration example

The following configuration example shows how to enable OSPFv2 and OSPFv3 protocols on PE1, PE2, and CE_A over VE 200 for the MCT member VLAN 2.

PE1:

```
router ospf
  area 0

ipv6 router ospf
  area 0

vlan 2
  router-interface Ve 200

interface Ve 200
  ipv6 address 2001::1/64
  ip address 10.2.2.1/24

  ip ospf area 0
  ipv6 ospf area 0
  !
  no shutdown
  !
```

PE2:

```
router ospf
  area 0

ipv6 router ospf
  area 0

vlan 2
  router-interface Ve 200

interface Ve 200
  ipv6 address 2001::2/64
  ip address 10.2.2.2/24

  ip ospf area 0
  ipv6 ospf area 0
  !
  no shutdown
  !
```

CE_A:

```
router ospf
  area 0

ipv6 router ospf
  area 0
```

```

vlan 2
  router-interface Ve 200

interface Ve 200
  ipv6 address 2001::10/64
  ip address 10.2.2.10/24

  ip ospf area 0
  ipv6 ospf area 0
  !
  no shutdown
  !

```

Layer 3 MCT bridge-domain configuration example

The following configuration example shows how to enable OSPFv2 and OSPFv3 protocols on PE1, PE2, and CE_A over VE 200 for the MCT member bridge-domain 2.

PE1:

```

router ospf
  area 0

ipv6 router ospf
  area 0

interface Ve 200
  ipv6 address 2001::1/64
  ip address 10.2.2.1/24

bridge-domain 2 p2mp
  router-interface Ve200
  logical-interface ethernet 0/1.2
  pw-profile default
  bpdu-drop-enable
  local-switching
  !
  ip ospf area 0
  ipv6 ospf area 0
  !
  no shutdown
  !

```

PE2:

```

router ospf
  area 0

ipv6 router ospf
  area 0

interface Ve 200
  ipv6 address 2001::2/64
  ip address 10.2.2.2/24

bridge-domain 2 p2mp
  router-interface Ve200
  logical-interface ethernet 0/1.2
  pw-profile default
  bpdu-drop-enable
  local-switching

```

```

!
ip ospf area 0
ipv6 ospf area 0
!
no shutdown
!

```

CE_A:

```

router ospf
  area 0

ipv6 router ospf
  area 0

interface Ve 200
  ipv6 address 2001::10/64
  ip address 10.2.2.10/24

bridge-domain 2 p2mp
  router-interface Ve200
  logical-interface ethernet 0/1.2
  pw-profile default
  bpdu-drop-enable
  local-switching
!
ip ospf area 0
ipv6 ospf area 0
!
no shutdown
!

```

Using MCT with VRRP and VRRP-E

The MCT device that acts as the Virtual Routing Redundancy Protocol (VRRP) and VRRP Extended (VRRP-E) backup router performs as a Layer 2 switch to pass the packets to the VRRP or VRRP-E master router for forwarding. Through MAC synchronization, the VRRP or VRRP-E backup router learns the virtual MAC (VMAC) on the Inter-Chassis Link (ICL). The data traffic and control traffic both pass through the ICL from the backup router. If VRRP-E short path forwarding is enabled, the backup router can forward the packets directly, instead of sending them to the master.



Note

Short path forwarding is only supported on VRRP-E.

In the diagram below, when an ARP request from the S1 switch device is sent through the direct link to the VRRP or VRRP-E backup router (PE2), a broadcast packet is sent to the VRRP/E master router (PE1) for processing through the ICL. When the ARP request is received by the PE1 device, PE1 sends a reply through the direct link to S1. If the ARP reply was received before the MAC address for the MCT on S1 is learned, the reply packet may be flooded to both the Customer Client Edge Port (CCEP) ports and ICL ports.

Using VRRP or VRRP-E, data traffic received from a client device on a backup router is Layer 2 switched to the master device. If VRRP-E short path forwarding is enabled, traffic received on the backup device may be forwarded by the backup if the route to the destination device is shorted than through the master device.

MCT short path forwarding configuration using VRRP-E example

In this example configuration, we are assuming that MCT is using the VRRP-E short path forwarding. When short path forwarding is enabled, packets from either the E1 or E2 devices with a destination of the E4 device can be routed through the PE 2 device which is a VRRP-E backup device. Short path forwarding is designed for load-balancing and allows packets to use the shortest path, and in this case, PE2 is directly connected to E4 so the packets will travel through PE2.

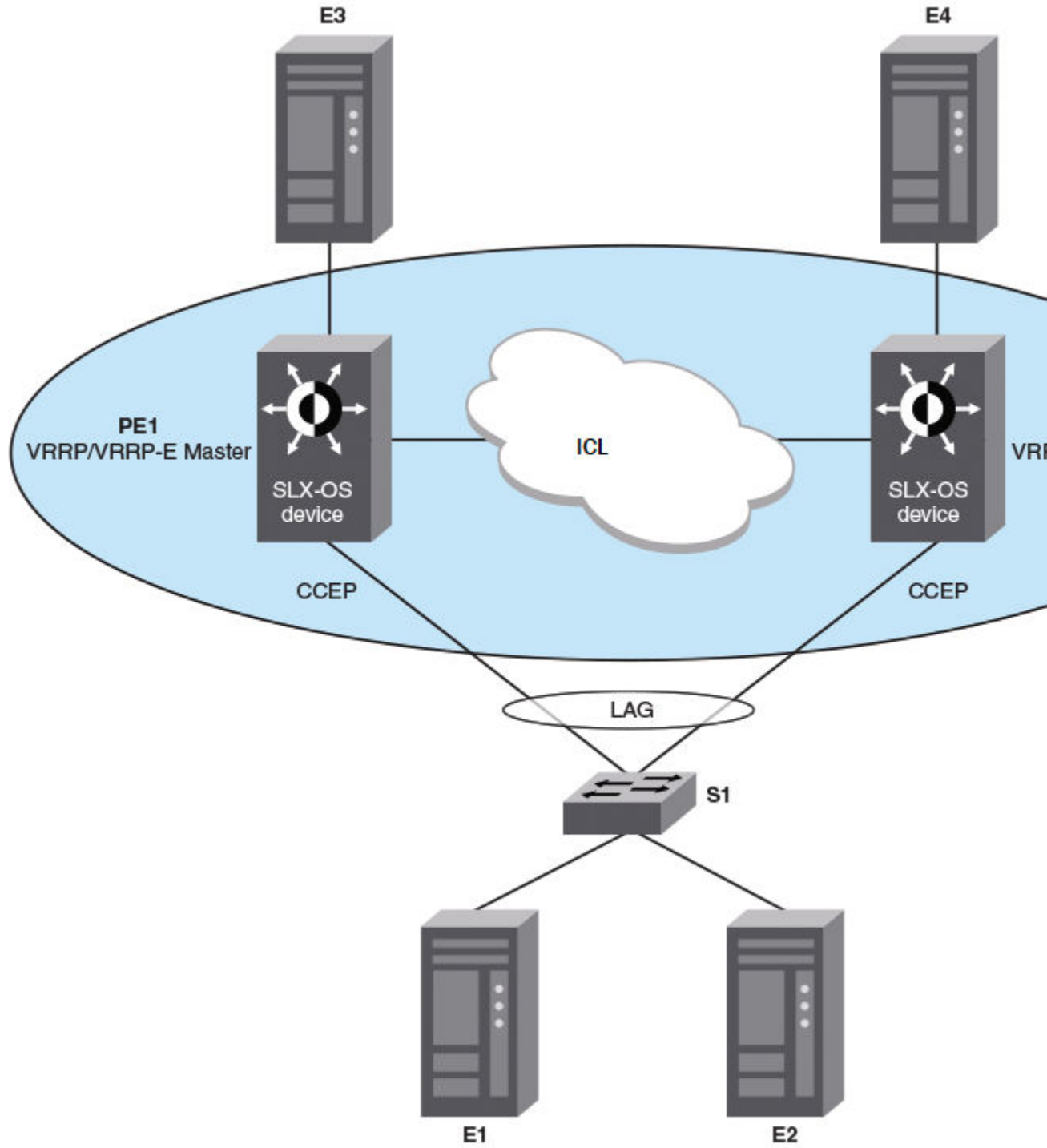


Figure 11: MCT short path forwarding

PE1 configuration

The following example configures the OSPF and BGP protocols with cluster configuration for MCT for the PE1 router in the diagram. A VRRP-E priority value of 110 (higher than the device at PE2) allows the PE1 device to assume the role of VRRP-E master.

```
ip router-id 10.19.19.19
router ospf
  area 0

interface Loopback 200
  no shutdown
  ip ospf area 0
  ip address 10.19.19.19/32

router bgp
  local-as 100
  neighbor 10.32.32.32 remote-as 100
  neighbor 10.32.32.32 update-source loopback 200
  address-family ipv4 unicast
  !
  address-family ipv6 unicast
  address-family l2vpn evpn
  neighbor 10.32.32.32 encapsulation mct
  neighbor 10.32.32.32 activate
  exit
  !
interface Ethernet 2/3
  ip address 10.1.8.19/24
  ip ospf area 0
  no shutdown
  !
vlan 100
  !
interface Ethernet 2/5
  switchport
  switchport mode trunk-no-default-native
  switchport trunk allow vlan add 100
  no shutdown
  !
  evpn myinstance
  vlan add 100
  rd auto
  route-target both auto ignore-as
  !
cluster c1 1
  peer 10.32.32.32
  deploy
  client c1 1
  client-interface Ethernet 2/5
  esi 01:02:03:04:05:06:07:08:09
  deploy
  !
vlan 100
  router-interface Ve 100
  !
protocol vrrp-extended
interface Ve 100
  ip proxy-arp
  ip address 10.2.3.6/24
  vrrp-extended-group 1
  priority 110
  short-path-forwarding
```

```

    virtual-ip 10.2.3.4
    no shutdown
    !
interface Ve 100
    ipv6 address fe80::1:2 link-local
    ipv6 address 3313::2/64
    ipv6 vrrp-extended-group 1
    virtual-ip 3313::1

```

PE2 configuration

The following example configures the OSPF and BGP protocols with cluster configuration for MCT for the PE2 router in the diagram. A VRRP-E priority value of 80 (lower than the device at PE1) allows the PE2 device to assume the role of a VRRP-E backup device.

```

ip router-id 10.32.32.32
router ospf
    area 0

interface Loopback 100
    no shutdown
    ip ospf area 0
    ip address 10.32.32.32/32

router bgp
    local-as 100
    neighbor 10.19.19.19 remote-as 100
    neighbor 10.19.19.19 update-source loopback 100
    address-family ipv4 unicast
    !
    address-family ipv6 unicast
    !
    address-family l2vpn evpn
    neighbor 10.19.19.19 encapsulation mct
    neighbor 10.19.19.19 activate
    !
    !
interface Ethernet 2/3
    ip address 10.1.8.32/24
    no shutdown
    ip ospf area 0
    !
vlan 100
    !
interface Ethernet 2/7
    switchport
    switchport mode trunk-no-default-native
    switchport trunk allow vlan add 100
    no shutdown
    !
    evpn myinstance
    vlan add 100
    rd auto
    route-target both auto ignore-as
    !
cluster c1 1
peer 10.19.19.19
deploy
    client c1 1
        esi 01:02:03:04:05:06:07:08:09
        client-interface Ethernet 2/7
    deploy

```



```
!  
vlan 100  
  router-interface Ve 100  
!  
protocol vrrp-extended  
interface Ve 100  
  ip proxy-arp  
  ip address 10.2.3.5/24  
  vrrp-extended-group 1  
    priority 80  
    short-path-forwarding  
    virtual-ip 10.2.3.4  
  no shutdown  
!  
interface Ve 100  
  ipv6 address fe80::1:1 link-local  
  ipv6 address 3313::3/64  
  ipv6 vrrp-extended-group 1  
  virtual-ip 3313::1
```

MCT use cases

An L2 MCT solution can be deployed at the access, aggregation, and the core of the data center. However, SLX-OS device is targeted for the data center core.

L2 MCT in the data center core

The following diagram shows a typical 3-tier data center where access and aggregation layers are running Layer 2 and the core is running Layer 2 and Layer 3. The access and aggregation can be standalone Extreme switches or any other third party switches.

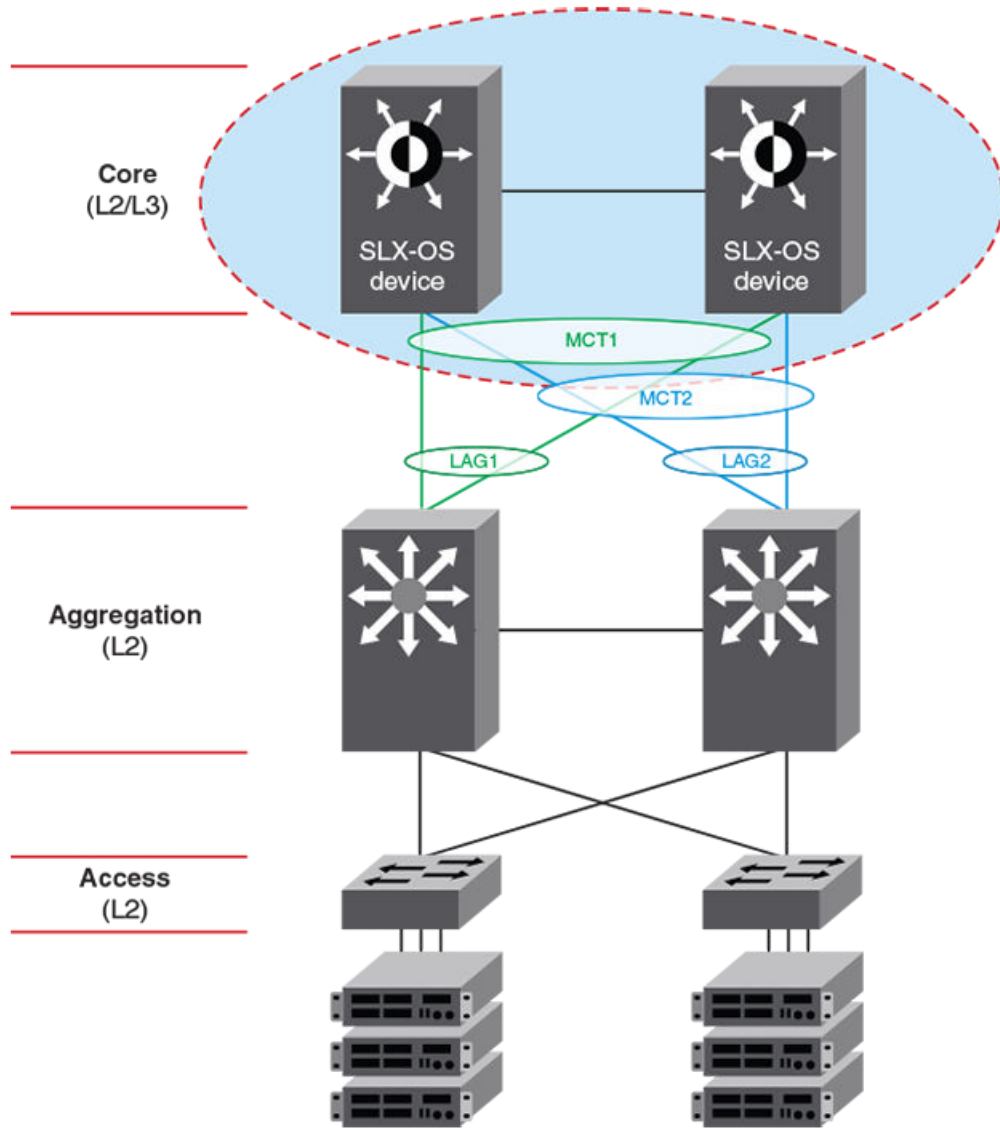


Figure 12: Typical 3-tier data center

Another variation of this use case is when the aggregation layer is a virtual cluster of switches which is transparent to SLX-OS devices in the core layer.

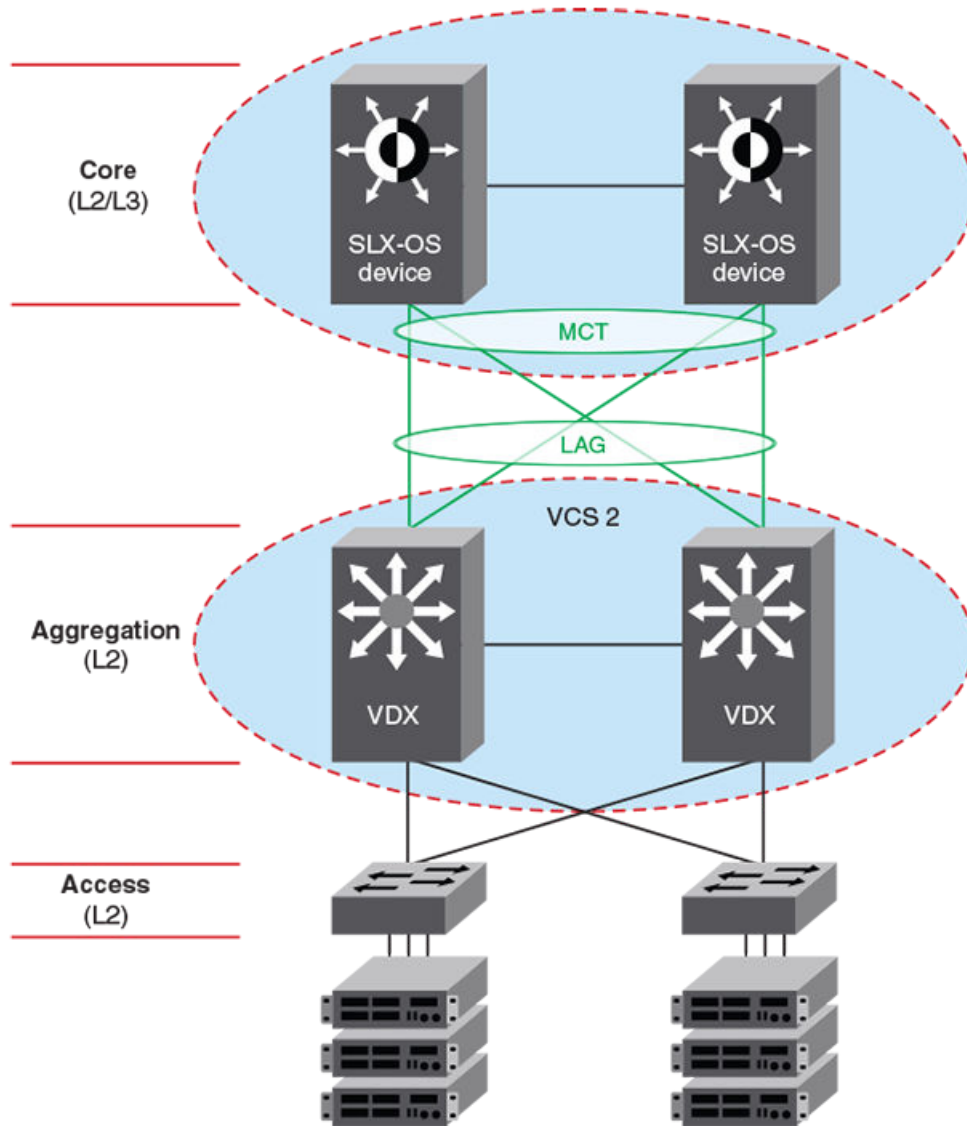


Figure 13: L2 MCT in the data center core connecting to a VCS

L2 MCT in a data center with a collapsed core and aggregation

The following diagram describes a scenario where VCS fabric of VDX 8870 switches is deployed at the access layer. With the availability of 10G and 40G interface, access switches can connect directly to the core without the need to have a separate aggregation layer.

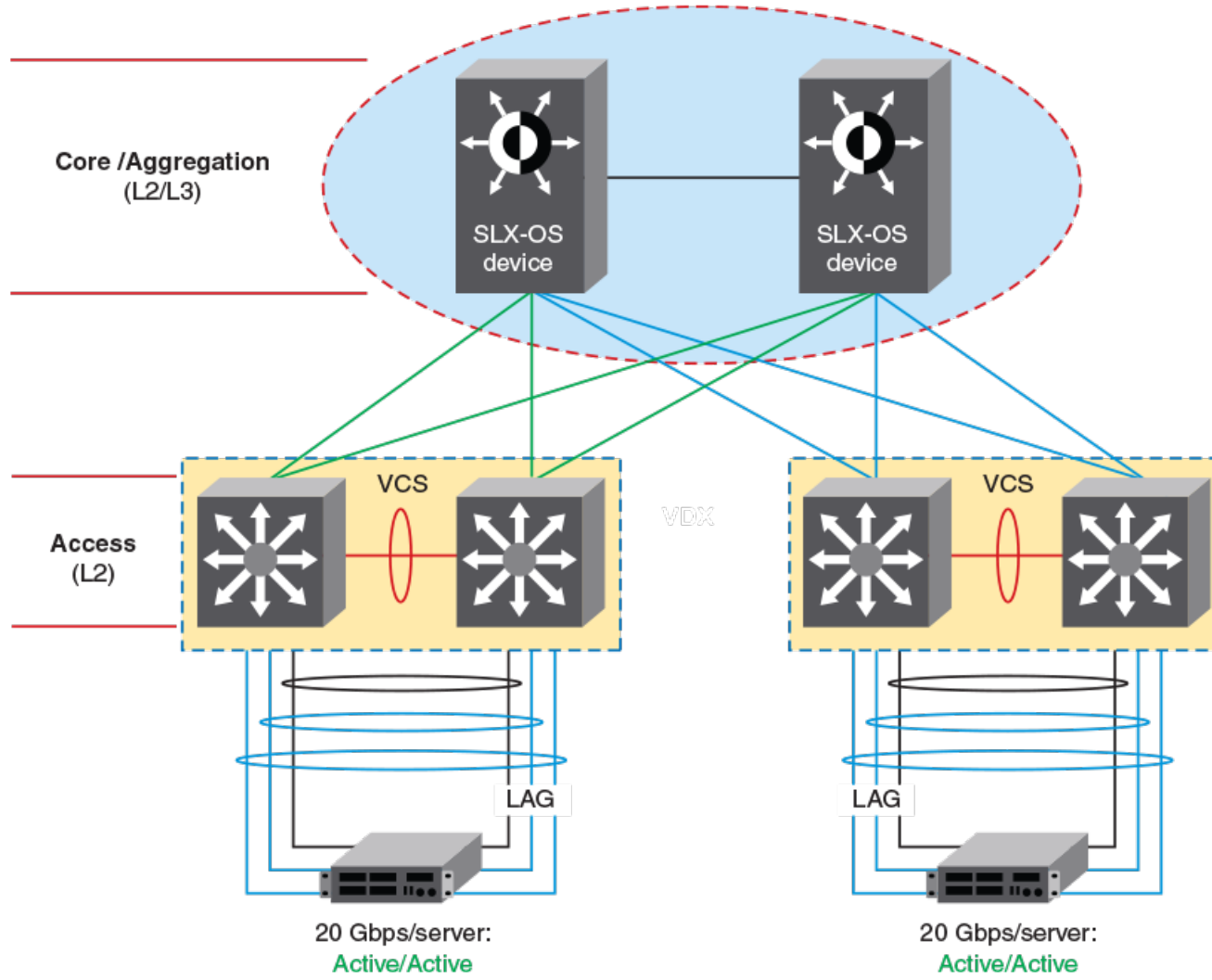


Figure 14: L2 MCT with collapsed core and aggregation



Logical Interfaces

[Logical interfaces overview](#) on page 101

[Configuring a logical interface on a physical port or port-channel \(LAG\)](#) on page 102

[LIF considerations for the SLX 9030](#) on page 103

Logical interfaces overview

This feature facilitates the support of future forwarding technologies without the need to modify code design in various software components.

A forwarding interface is also known as "main interface." It can be a physical port, a port-channel (Link Aggregation Group, or LAG), a pseudowire (PW), a tunnel, and so on. A logical interface can also be thought of as a subinterface configuration on top of a main interface.



Note

Currently the only LIFs that require explicit user configuration are attachment circuit (AC) LIFs.

LIFs and bridge domains

A Layer 2 application for LIFs is for bridge domains (BDs). A BD is an infrastructure that supports the implementation of different switching technologies; it is essentially a generic broadcast/forwarding domain that is not tied to a specific transport technology. Bridge domains support a wide range of service endpoints, including regular Layer 2 endpoints and Layer 2 endpoints over Layer 3 technologies. Logical interfaces representing BD endpoints must be created before they can be bound to a BD. For more information and configuration details, refer to the [Bridge Domains](#) chapter in this guide.

Configuration considerations

The following are some common rules to consider in configuring logical interfaces:

- By default, when the LIF is created it is configured as "no shutdown."
- By default, when the LIF is created, it is "tagged" unless it is explicitly configured with the "untagged" option.
- Allowed LIF service instance ID ranges are from 1 through 12288.
- An LIF service instance ID has no correlation to the VLAN ID of the LIF.

- Each physical/LAG-based LIF must have an associated VLAN configured or else it will not be usable when the user attempts to add it to a service (such as VPLS, Layer 2). Such a configuration request to add the LIF to a service will be rejected.
- Once the LIF is associated with a Layer 2 service, its VLAN value cannot be changed or deleted unless it is first removed from the associated service. In case the LIF is not yet associated to a service, the user is free to remove the VLAN configuration or change the VLAN assignment.
- The **no** option to the **logical-interface** command can be applied at any time.
- The "untagged" configuration is allowed for only one LIF under the same physical port or LAG. If one LIF is already configured as untagged, all subsequent attempts on the same physical port or LAG are rejected.
- Once the "untagged" option is selected, it will only have one VLAN as the next classification option. There is no dual-tag support for the untagged case.
- In order to configure an untagged LIF, the main interface must be configured as "switchport mode trunk-no-default-native". If it is configured set to regular trunk mode, the native VLAN is already associated with a regular Layer 2 VLAN LIF and no explicit untagged LIF can be configured on that interface.
- Once the LIF is associated with a service (Layer 2) such as a bridge domain, its "untagged/tagged" configuration cannot be changed. The service instance or its current VLAN classification must be deleted by the user first and then added back with the proper "untagged/tagged" option.
- VLANs 4091 through 4095 are reserved VLANs and these should not be used as the VLAN ID for either the inner or outer VLAN of the LIF.
- The VLAN specified under the LIF ensures that such a VLAN is not already configured under the **switchport** command for a regular Layer 2 allowed VLAN.

If the interface is already configured as "switchport access," then it is not allowed to be configured with LIF. The reverse condition is also not allowed: the interface cannot be changed to mode access if a LIF is still configured under the main interface.

Configuring a logical interface on a physical port or port-channel (LAG)

About This Task

Refer to the Usage Guidelines for the **logical-interface** command for complete details.

Procedure

1. Do the following to configure a logical interface on an Ethernet port.

- a. Enter global configuration mode.

```
device# configure terminal
```

- b. Specify an Ethernet interface.

```
device(config)# interface ethernet 2/6
```

- c. Enter the **switchport** command to configure the parent interface as switchport.

```
device(conf-if-eth-2/6)# switchport
```

- d. Enter the **switchport mode trunk-no-default-native** command to enable an explicit untagged LIF to be configured.

```
device(conf-if-eth-2/6)# switchport mode trunk-no-default-native
```

- e. Enable the interface.

```
device(conf-if-eth-2/6)# no shutdown
```

- f. Enter the **logical-interface** command, specify a service instance, and enter LIF configuration mode.

```
device(conf-if-eth-2/6)# logical-interface ethernet 2/6.120
```

- g. (Optional) Enter the **name** command to facilitate the management of the LIF.

```
device(conf-if-eth-lif-2/6.120)# name myLIF120
```

- h. Enter the **vlan** command with the **inner-vlan** option to specify an interface and create dual-tag VLANs.

```
device(conf-if-eth-lif-2/6.120)# vlan 120 inner-vlan 200
```

- i. Alternatively, enter the **untagged vlan** command to specify that the LIF is to receive untagged packets.

```
device(conf-if-eth-lif-2/6.120)# untagged vlan 120
```

See the Usage Guidelines for the **vlan (LIF)** command.

- j. (Optional) By default, the administrative state of the LIF is "no shutdown." To remove the port from participating in any data traffic without having to shut down the physical interface, enter the **no** form of the **shutdown (LIF)** command.

```
device(conf-if-eth-lif-2/6.120)# no shutdown
```

- k. (Optional) For convenience, you can also enter up to two options in a single command line, as in the following examples.

```
device(conf-if-eth-2/6)# logical-interface ethernet 2/6.120 name myLIF120
```

```
device(conf-if-eth-2/6)# logical-interface ethernet 2/6.120 vlan 120
```

2. To configure a port-channel, configure the basic LIF parameters and options as in Step 1.
 - a. Specify a port-channel, set its mode to "trunk-no-default-native," and specify a logical interface service instance.

```
device(config)# interface port-channel 10
device(config-port-channel-10)# switchport mode trunk-no-default-native
device(config-port-channel-10)# logical-interface port-channel 10.3
device(config-if-po-lif-10.3)#
```

- b. Repeat additional substeps in Step 1 as appropriate.

LIF considerations for the SLX 9030

The SLX 9030 supports all the logical interface (LIF) features supported by other platforms in this release, with exceptions as noted in the following table:

Table 21: LIF feature support on the SLX 9030

Supported	Not supported
Attachment circuit (AC) LIF configuration (untagged, single tag, dual tag) on port/LAG interfaces	Under bridge domain configuration, the no local-switching command is not allowed on the SLX 9030.
AC LIF admin shut/no shut	Raw-passthrough mode

Table 21: LIF feature support on the SLX 9030 (continued)

Supported	Not supported
Tag protocol identifier (TPID) configuration on port/LAG	
LIF statistics	
LIF QoS mapping	
Tag/raw mode	
Up to 15744 VLAN ports	

The unsupported features are detailed below.

No local switching

Under bridge domain configuration, the **no local-switching** command is not allowed on the SLX 9030. The command is rejected as shown in the following example.

```
device(config)# bridge 3
dfevice(config-bridge-domain-3)# no local-switching
% Error: no local switching is not supported on this HW.
device(config-bridge-domain-3)#
```

VC mode raw-passthrough

Because the SLX 9030 does not support MPLS, the pseudo-wire (PW) profile configuration for virtual connection (VC) mode does not allow the "raw-passthrough" option. Raw-passthrough mode is valid only for VPLS/VLL features. A bridge domain that has a VXLAN extension does use the PW profile VC mode configuration to decide whether an internal VLAN TAG should be passed towards the VXLAN tunnel, so a VC mode with a **raw** or **tag** option is allowed.



Note

For routing over a BD-based VXLAN, the **tag** keyword is needed if the VXLAN peer is a previous SLX platform that requires VC mode to be tagged.

Below is an example configuration that demonstrate how VC mode is rejected if it is configured on an SLX 9030.

```
device(config)# pw-profile RPT
device(config-pw-profile-RPT)# vc-mode
Possible completions:
  raw          Raw Mode
  raw-passthrough Raw-passthrough Mode
  tag          Tag Mode
device(config-pw-profile-RPT)# vc-mode raw-passthrough
% Error: vc-mode raw-passthrough not supported on this HW.
device(config-pw-profile-RPT)#
```

Dual-tag support on the SLX 9030

Support for dual tagging is introduced on the SLX 9030 with this release. However, because that platform is unable to parse more than two recognized VLAN tags, resulting in a specific scenario where

dual-tag VLAN editing is not supported. As long as the user is aware of this particular scenario and takes care in designing the network, full dual-tag service delimiting can be achieved as with other SLX platforms.

Dual-tag service delimiting definition

When a single-tagged LIF is configured on an interface, any packet that is received on that interface with a matching outer VLAN ID (VID) is accepted as valid traffic sourced from that LIF. This includes dual-tagged packets where only the outer VLAN tag matches the LIF. If a LIF is configured with dual-tag service delimiting, only packets with matching both outer and inner VLAN tags are classified as sourced from that dual-tagged LIF. Once the packet's LIF is classified, its bridge domain is identified and the packet is switched within that BD. The service-delimiting VLAN tags are either stripped or replaced, depending on the destination LIF's service delimiting VLAN(s).

The following show outputs illustrate this scenario.

```
device# show running-config interface ethernet 0/1
interface Ethernet 0/1
  switchport
  switchport mode trunk-no-default-native
  no shutdown
  logical-interface ethernet 0/1.100
    vlan 100
  !
  logical-interface ethernet 0/1.120
    vlan 100 inner-vlan 200
  !
  logical-interface ethernet 0/1.300
    untagged vlan 300
  !
  logical-interface ethernet 0/1.400
    vlan 400
  !
  logical-interface ethernet 0/1.560
    vlan 500 inner-vlan 600
  !
device# show run inter eth 0/2
interface Ethernet 0/2
  switchport
  switchport mode trunk-no-default-native
  no shutdown
  logical-interface ethernet 0/2.200
    untagged vlan 200
  !
  logical-interface ethernet 0/2.230
    vlan 200 inner-vlan 300
  !
  logical-interface ethernet 0/2.300
    vlan 300
  !
device# show running-config bridge-domain
bridge-domain 1 p2mp
  pw-profile default
  logical-interface ethernet 0/1.100
  logical-interface ethernet 0/2.200
  logical-interface ethernet 0/2.230
  logical-interface ethernet 0/2.300
  bpdu-drop-enable
  local-switching
  !
```

```
bridge-domain 2 p2mp
pw-profile default
logical-interface ethernet 0/1.120
logical-interface ethernet 0/1.300
logical-interface ethernet 0/1.400
logical-interface ethernet 0/1.560
bpdu-drop-enable
local-switching
!
```

If a single tagged packet with VLAN tag 100 enters through interface Ethernet 0/1, it is classified as a packet sourced from LIF Ethernet 0/1.100 in bridge-domain 1.

If a dual-tagged packet with outer VLAN tag 100 and inner VLAN tag 200 enters through interface Ethernet 0/1, it is classified as a packet sourced from LIF Ethernet 0/1.120 in bridge-domain 2. If a dual-tagged packet with outer VLAN tag 100 and inner VLAN tag 300 enters through interface Ethernet 0/1, because there are no LIFs configured to match both VLAN tags and only the outer tag has a match, the packet is classified as sourced from LIF Ethernet 0/1.100 in bridge-domain 1.

Dual-tag VLAN editing

Both tags are service delimiting

Referring to previous show examples, if a dual-tagged packet with outer VLAN 100 and inner VLAN 200 is received on interface Ethernet 0/1, and it is a BUM packet, this packet is recognized as sourced from LIF 0/1.120 and flooded towards all other LIFs on bridge-domain 2. Because both VLAN tags are service delimiting, based on the destination, the following shows how VLAN editing is performed:

- **Destined to "untagged" LIF Ethernet 0/1.300:** The original packet's outer and inner tags are both stripped and the packet goes out as untagged.
- **Destined to "single-tagged" LIF Ethernet 0/1.400:** The original packet's outer VLAN ID is replaced from 100 to 400 and the original packet's inner VLAN tag is stripped. The packet goes out as single tagged.
- **Destined to "dual-tagged" LIF Ethernet 0/1.560:** The original packet's outer VLAN ID is replaced from 100 to 500 and the original packet's inner VLAN ID is replaced from 200 to 600. The packet goes out as dual tagged.

Only the outer tag is service delimiting

Here is another example of how a dual-tagged packet is handled when only the outer VLAN tag is service delimiting.

Consider the example of a dual-tagged BUM packet with outer VLAN 100 and inner VLAN 800 that is received on interface Ethernet 0/1. It is classified as packet sourced from LIF Ethernet 0/1.100 in bridge-domain 1. This packet is flooded to all other LIFs on bridge-domain 1. Because only the outer VLAN tag is service delimiting, the inner VLAN tag 800 is treated as a customer payload and is not modified. Any destination LIF VLAN editing must happen above this original packet's inner VLAN tag. The following shows how the VLAN editing is performed as the packet goes towards each destination LIF.

- **Destined to "untagged" LIF Ethernet 0/2.200:** The original packet's outer VLAN tag is stripped and the packet goes out as "single tagged" with VID 800. Remember that the inner VLAN tag is not service delimiting and thus should not be removed.

- **Destined to "single-tagged" LIF Ethernet 0/2.300:** The original packet's outer VLAN ID is replaced from 100 to 300 and the original packet's inner VLAN tag is untouched. The packet goes out as "dual tagged" with outer VID 300 and inner VID 800.
- **Destined to "dual tagged" LIF Ethernet 0/2.230:** Ideally the platform would replace the original packet's outer VLAN ID from 100 to 200 and insert a new inner tag with VID 300 while keeping the original packet's inner VLAN TAG 800 as the third VLAN TAG. However, the platform can handle only two VLAN tags for manipulation by VLAN editing, and the original inner VLAN tag is treated as a customer payload that cannot be altered. Consequently, destination VLAN editing does not work. The end result is that only two VLAN tags can be kept and the original customer VLAN tag with VID 800 is replaced by the destination LIF's inner VLAN tag. This causes the packet to go out as "dual-tagged" with outer VLAN 200 and inner VLAN 300, instead of as triple tagged.

**Important**

Because of the above limitation, the user must understand that to have dual-tagged packets while only the outer VLAN is service delimiting, a dual-tagged LIF must not be configured within the same bridge domain. If there is a need to configure a dual-tagged LIF with both VLANs being service delimiting, then the bridge domain traffic should never have dual-tagged packets where the inner VLAN is NOT service delimiting.

TPID support on the SLX 9030

One TPID is the default, 0x8100. This limitation applies where the TPID configuration can be applied only to the outer TPID. The inner TPID, in case of a dual-tagged packet, must be 0x8100 in order to be recognized as a valid inner VLAN TAG.

Below is a typical TPID configuration example that uses a nondefault TPID.

```
device(config)# interface ethernet 0/4
device(conf-if-eth-0/4)# tag-type 0x9100
device(conf-if-eth-0/4)#
```

**Important**

If the TPID of the port is configured to be different from the received packet's TPID value, the received packet's VLAN tag is not recognized by the SLX 9030 as a valid VLAN tag and the entire packet is treated as "untagged". Therefore, if an untagged LIF is configured on the receiving interface, then the packet is classified as being sourced from that untagged LIF. If no untagged LIF is configured on that interface, then the packet is expected to be dropped.

Consequently, the user may purposely want to configure the interface's TPID to be different from the packet's TPID value to achieve this "passthrough" behavior to meet network requirements.

PW profile support on the SLX 9030

Normally a VXLAN operates under "raw" VC mode, where all the service-delimiting VLAN tag(s) are stripped before the packet is encapsulated with the VXLAN header to be sent towards the VXLAN tunnel. However, in the case of routing it may be necessary to configure the VC mode as "tagged" in order to interoperate with other platforms where routing requires the use of tagged VC mode. Although the SLX 9030 has no such hardware limitation, it is recommended that tagged VC mode be used when routing is needed, in order to interoperate with those other platforms. Tagged VC mode means that the innermost service-delimiting VLAN tag of the AC LIF where the packet is received is kept and

transmitted towards its VXLAN peer. The purpose of this service-delimiting VLAN tag is mainly to preserve the original packet's Priority Code Point (PCP) value for QoS purposes at the receiving VXLAN's peer. The actual VLAN ID is never used for another purpose.



Note

On the SLX 9030, because of the hardware requirement to support VLAN translation, the original VLAN ID may be translated to a "normalized" value instead of the original VLAN value when the source AC LIF is a tagged LIF. Therefore, the user should not expect what this VLAN ID value should be. It is there only to carry the original PCP value.

Raw/tagged mode packet format with a VXLAN header

By default, a bridge domain uses the default PW profile, which is set with VC mode "raw". Similarly, if the VXLAN is extended on a VLAN (not a bridge domain), it is expected to be in raw mode. The packet format of a VXLAN packet does not carry the extra VLAN tag in the case of raw mode.

The following table shows the packet format of the VXLAN header and the inner Ethernet frame.

Table 22: Packet format of VXLAN header and inner Ethernet frame

VXLAN header				Inner Ethernet frame		
Outer Ethernet header (14 bytes)	Outer IP header (20 bytes)	Outer UDP header (8 bytes)	VXLAN header (8 bytes)	Inner destination and source MAC (12 bytes)	Optional VLAN tag (4 bytes) if tagged mode	Ether type, inner iP header, and payload



Bridge Domains

[Bridge domain overview on page 109](#)

[Configuring a bridge domain on page 110](#)

[Displaying bridge-domain configuration information on page 111](#)

[Enabling statistics on a bridge domain on page 114](#)

[Displaying statistics for logical interfaces in bridge domains on page 115](#)

[Clearing statistics on bridge domains on page 115](#)

Bridge domain overview

A bridge domain is a generic broadcast domain that is not tied to a specific transport technology. Bridge domains support a wide range of service endpoints including regular L2 endpoints and L2 endpoints over L3 technologies.



Note

Bridge domains are not supported for MPLS/VPLS purposes. They can be used for VLAN translation purposes only. Up to 1K bridge domains are supported at the system level.

Bridge domain statistics

Statistics must be manually enabled for a specific bridge domain, since statistics for bridge domains are not enabled by default.

Use the **statistics** command in bridge domain configuration mode to enable statistics on a bridge domain.



Note

- The statistics reported are not real-time statistics since they depend upon the load on the system.
- Enabling statistics on a bridge domain has a heavy impact on data traffic.
- For optimum utilization of the statistics resources in the hardware, statistics on a bridge domain are not enabled by default.

Configuring a bridge domain

Before You Begin

Before configuring a bridge domain, configure any logical interface that is to be bound to the bridge domain. Logical interfaces that represent bridge-domain endpoints must be created before they are bound to a bridge domain. For further information on configuration of logical interfaces, refer to *Logical Interfaces*.

There is an example at the end of this task that shows all the configuration steps in order.

About This Task

Perform the following task to configure a bridge domain.

Procedure

1. From privileged EXEC mode, enter global configuration mode.

```
device# configure terminal
```

2. Create a bridge domain.

```
device(config)# bridge-domain 5 p2p
```

By default, the bridge-domain service type is multipoint (**p2mp**). In this example, bridge domain 5 is configured as a point-to-point service (**p2p**).

- 3.



Note

Logical interfaces representing bridge-domain endpoints must be created before they can be bound to a bridge domain. For further information, refer to *Logical Interfaces*.

Bind the logical interfaces for attachment circuit endpoints to the bridge domain.

```
device(config-bridge-domain-5)# logical-interface ethernet 0/6.400
```

In this example, Ethernet logical interface 0/6.400 is bound to bridge domain 5.

4. Repeat Step 4 to bind other logical interfaces for attachment circuit endpoints to the bridge domain.

```
device(config-bridge-domain-5)# logical-interface port-channel 2.200
```

In this example, port channel logical interface 2.200 is bound to bridge domain 5.

5. (Optional) Enable local switching for bridge domain 5.

```
device(config-bridge-domain-5)# local-switching
```

By default, local switching is enabled.

6. (Optional) Enable dropping L2 bridge protocol data units (BPDUs) for bridge domain 5.

```
device(config-bridge-domain-5)# bpdu-drop-enable
```

Example

The following example creates bridge domain 5. It binds ethernet and port-channel logical interfaces to the bridge domain. It configures local switching, and enables dropping of L2 BPDUs.

```
device# configure terminal
```

```

device(config)# bridge-domain 5
device(config-bridge-domain-5)# logical-interface ethernet 0/6.400
device(config-bridge-domain-5)# logical-interface port-channel 2.200
device(config-bridge-domain-5)# local-switching
device(config-bridge-domain-5)# bpdu-drop-enable

```

Displaying bridge-domain configuration information

- Enter the **show bridge-domain** command to display information about all configured bridge domains.

```

device# show bridge-domain

Total Number of bridge-domains: 3
Number of bridge-domains: 3

Bridge-domain 1
-----
Bridge-domain Type: mp , VC-ID: 5
Number of configured end-points: 5 , Number of Active end-points: 4
VE if-indx: 1207959555, Local switching: TRUE, bpdu-drop-enable:TRUE
PW-profile: 1, mac-limit: 128000
Number of Mac's learned:90000,      Static-mac count: 10,
VLAN: 100, Tagged ports: 2(2 up), Un-tagged ports: 0 (0 up)
Tagged ports: Eth 2/6, eth 2/8
Un-tagged ports:

Total PW peers: 2 (2 Operational)
Peer address: 12.12.12.12, State: Operational, Uptime: 2 hr 55 min
  Load-balance: True , Cos enabled:False,
  Assigned LSP;s:
  Tnnl in use: tnl2[RSVP]
  Local VC lbl: 983040, Remote VC lbl: 983040
  Local VC MTU: 1500, Remote VC MTU: 1500,
  Local VC-Type: Ethernet(0x05), Remote VC-Type: Ethernet(0x05)
Peer address: 15.15.15.15, State: Operational, Uptime: 2 hr 55 min
  Load-balance: False , Cos enabled:False,
  Assigned LSP's: lsp1, lsp2
  Tnnl in use: tnl1[MPLS]
  Local VC lbl: 983041, Remote VC lbl: 983043
  Local VC MTU: 1500, Remote VC MTU: 1500 ,
  Local VC-Type: Ethernet(0x05), Remote VC-Type: Ethernet(0x05)

Bridge-domain 2
-----
Bridge-domain Type: mp , VC-ID: 100
Number of configured end-points: 5 , Number of Active end-points: 4
VE if-indx: NA, Local switching: FALSE, bpdu-drop-enable:FALSE
PW-profile: profile_1, mac-limit: 262144
Number of Mac's learned:90000,      Static-mac count: 10,
VLAN: 100, Tagged ports: 2(1 up), Un-tagged ports: 0 (0 up)
  Tagged ports: eth 2/10, eth 1/10
  Un-tagged ports:
VLAN: 150, Tagged ports: 1(1 up), Un-tagged ports: 0 (0 up)
  Tagged ports: eth 1/5
  Un-tagged ports:

Bridge-domain 3
-----
Bridge-domain Type: mp , VC-ID: 200
Number of configured end-points: 5 , Number of Active end-points: 4
VE if-indx: 120793855, Local switching: FALSE, bpdu-drop-enable:FALSE

```

```

PW-profile: 2, mac-limit: 262144
Number of Mac's learned:90000,      Static-mac count: 10,
Local switching: TRUE,
VLAN: 500, Tagged ports: 2(2 up), Un-tagged ports: 2 (1 up)
Tagged ports:      eth 11/6, eth 4/3
Un-tagged ports:

Total VPLS peers: 3 (2 Operational)
Peer address: 5.5.5.5, State: Operational, Uptime: 2 hr 35 min
      Load-balance: False , Cos enabled:False,
      Assigned LSP;s:
      Tnnl in use: tn12[RSVP]
      Local VC lbl: 983050, Remote VC lbl: 983050
      Local VC MTU: 1500,Remote VC MTU: 1500,
      Local VC-Type: Ethernet(0x05), Remote VC-Type: Ethernet(0x05)
Peer address: 20.20.20.20, State: Operational, Uptime: 0 hr 18 min
      Load-balance: False , Cos enabled:True,
      Assigned LSP's:
      Tnnl in use: NA,
      Local VC lbl: NA, Remote VC lbl: NA
      Local VC MTU: 1500,Remote VC MTU: 1500,
      Local VC-Type: Ethernet(0x05), Remote VC-Type: Ethernet(0x05)
Peer address: 10.10.10.10, State: Not-Operational (Tunnel Not Available),
      Load-balance: True , Cos enabled:False,
      Assigned LSP's: lsp10, lsp15
      Tnnl in use: NA,
      Peer Index:2
      Local VC lbl: NA, Remote VC lbl: NA
      Local VC MTU: 1500,Remote VC MTU: NA ,
      Local VC-Type: Ethernet(0x05), Remote VC-Type: NA

```

- Enter the **show bridge-domain** command specifying the bridge-domain ID to display information about a specific bridge domain. The following example displays information about bridge domain 501.

```

device# show bridge-domain 501

Bridge-domain 501
-----
Bridge-domain Type: MP , VC-ID: 501
Number of configured end-points: 2 , Number of Active end-points: 2
VE if-indx: 0, Local switching: TRUE, bpdu-drop-enable: TRUE
PW-profile: default, mac-limit: 0
VLAN: 501, Tagged ports: 1(1 up), Un-tagged ports: 0 (0 up)
Tagged Ports: eth 1/6.501
Un-tagged Ports:
Total VPLS peers: 1 (1 Operational):

VC id: 501, Peer address: 10.9.9.9, State: Operational, uptime: 2 sec
      Load-balance: False, Cos Enabled: False,
      Tunnel cnt: 1
      rsvp p101(cos_enable:False cos_value:0)
      Assigned LSPs count:0 Assigned LSPs:
      Local VC lbl: 989042, Remote VC lbl: 983040,
      Local VC MTU: 1500, Remote VC MTU: 1500,
      Local VC-Type: 5, Remote VC-Type: 5

```

The following example shows information about a bridge domain (501) in which the **load-balance** option is configured for the peer device 10.9.9.9.

```

show bridge-domain 501

Bridge-domain 501

```



```

-----
Bridge-domain Type: MP , VC-ID: 501
Number of configured end-points: 2 , Number of Active end-points: 2
VE if-indx: 0, Local switching: TRUE, bpdu-drop-enable: TRUE
PW-profile: default, mac-limit: 0
VLAN: 501, Tagged ports: 1(1 up), Un-tagged ports: 0 (0 up)
Tagged Ports: eth 1/6.501
Un-tagged Ports:
Total VPLS peers: 1 (1 Operational):

VC id: 501, Peer address: 10.9.9.9, State: Operational, uptime: 48 sec
  Load-balance: True , Cos Enabled: False,
  Tunnel cnt: 16
  rsvp p101(cos_enable:False cos_value:0)
  rsvp p102(cos_enable:False cos_value:0)
  rsvp p103(cos_enable:False cos_value:0)
  rsvp p104(cos_enable:False cos_value:0)
  rsvp p105(cos_enable:False cos_value:0)
  rsvp p106(cos_enable:False cos_value:0)
  rsvp p107(cos_enable:False cos_value:0)
  rsvp p108(cos_enable:False cos_value:0)
  rsvp p109(cos_enable:False cos_value:0)
  rsvp p110(cos_enable:False cos_value:0)
  rsvp p111(cos_enable:False cos_value:0)
  rsvp p112(cos_enable:False cos_value:0)
  rsvp p113(cos_enable:False cos_value:0)
  rsvp p114(cos_enable:False cos_value:0)
  rsvp p115(cos_enable:False cos_value:0)
  rsvp p116(cos_enable:False cos_value:0)
  Assigned LSPs count:0 Assigned LSPs:
  Local VC lbl: 989040, Remote VC lbl: 983040,
  Local VC MTU: 1500, Remote VC MTU: 1500,
  Local VC-Type: 5, Remote VC-Type: 5

```

The following example shows information about bridge domain 501 in which the **load-balance** option and four assigned label-switched paths (p101, p102, p103, and p104) are configured for the peer device 10.9.9.9.

```

device# show bridge-domain 501

Bridge-domain 501
-----
Bridge-domain Type: MP , VC-ID: 501
Number of configured end-points: 2 , Number of Active end-points: 2
VE if-indx: 0, Local switching: TRUE, bpdu-drop-enable: TRUE
PW-profile: default, mac-limit: 0
VLAN: 501, Tagged ports: 1(1 up), Un-tagged ports: 0 (0 up)
Tagged Ports: eth 1/6.501
Un-tagged Ports:
Total VPLS peers: 1 (1 Operational):

VC id: 501, Peer address: 10.9.9.9, State: Operational, uptime: 4 sec
  Load-balance: True , Cos Enabled: False,
  Tunnel cnt: 4
  rsvp p101(cos_enable:False cos_value:0)
  rsvp p102(cos_enable:False cos_value:0)
  rsvp p103(cos_enable:False cos_value:0)
  rsvp p104(cos_enable:False cos_value:0)
  Assigned LSPs count:4 Assigned LSPs:p101 p102 p103 p104
  Local VC lbl: 989041, Remote VC lbl: 983040,
  Local VC MTU: 1500, Remote VC MTU: 1500,
  Local VC-Type: 5, Remote VC-Type: 5

```

- Enter the **show bridge-domain brief** command to display summary information about all configured bridge domains.

```
device# show bridge-domain brief

Total Number of bridge-domains configured: 10
Number of VPLS bridge-domains: 5
Macs Dynamically learned: 50360, Macs statically configured: 0

BDID (VC-ID)   TYPE      Intf (up)    PWs (up)    macs
501 (501)     P2MP     5 (3)        2 (2)       50000
502 (502)     P2MP     1 (1)        1 (1)       10
503 (503)     P2MP    10 (6)        3 (1)       0
504 (504)     P2MP     1 (1)        1 (1)       350
505 (505)     P2MP     1 (1)        1 (1)       0
506 (506)     P2P      1 (1)        1 (1)       0
507 (507)     P2P      1 (1)        1 (1)       0
508 (508)     P2P      1 (1)        1 (1)       0
509 (509)     P2P      1 (1)        1 (1)       0
510 (510)     P2P      1 (1)        1 (1)       0
```

Enabling statistics on a bridge domain

About This Task



Note

By default statistics are disabled on bridge domains. After enablement, statistics should be disabled when no longer needed because the collection of statistical information has a heavy impact on data traffic.

Procedure

1. Enter the global configuration mode.

```
device# configure terminal
```

2. Enter the **bridge-domain** command to create a bridge domain at the global configuration level.

```
device(config)# bridge-domain 3
```

3. Enter the **statistics** command to enable statistics for all the logical interfaces and peers in the bridge domain.

```
device(config-bridge-domain-3)# statistics
```



Note

When statistics are no longer needed, use the **no statistics** command to disable statistics on the bridge domain.

Example

Example

The following example shows how to enable statistics on bridge domain 3.

```
device# configure terminal
device(config)# bridge-domain 3
device(config-bridge-domain-3)# statistics
```

Example

The following example shows how to disable statistics on bridge domain 3.

```
device# configure terminal
device(config)# bridge-domain 3
device(config-bridge-domain-3)# no statistics
```

Displaying statistics for logical interfaces in bridge domains

- Enter the **show statistics bridge-domain** command to display statistics for all logical interfaces and peers on all configured bridge domains.

```
device# show statistics bridge-domain

Bridge Domain 1 Statistics
Interface          RxPkts          RxBytes          TxPkts
TxBytes
eth 1/1.100        821729          821729           95940360        95940360
eth 1/21.200       884484          885855           95969584        95484555
po 1.300           8884            8855             9684             9955

Bridge Domain 20 Statistics
Interface          RxPkts          RxBytes          TxPkts
TxBytes
eth 1/6.400        821729          821729           95940360        95940360
eth 1/21.100       8884            8855             9684             9955
po 2.40            884484          885855           95969584        95484555
```

- Enter the **show statistics bridge-domain** command specifying a bridge-domain ID to view the statistics for a specific bridge domain. The following example displays statistics for bridge-domain ID 1.

```
device# show statistics bridge-domain 1

Bridge Domain 1 Statistics
Interface          RxPkts          RxBytes          TxPkts
TxBytes
eth 1/1.100        821729          821729           95940360        95940360
eth 1/21.200       884484          885855           95969584        95484555
po 1.300           8884            8855             9684             9955
```

Clearing statistics on bridge domains

- Enter the **clear statistics bridge-domain** command to clear statistics for all logical interfaces and peers on all configured bridge domains.

```
device# clear statistics bridge-domain
```

- Enter the **clear statistics bridge-domain** command specifying the bridge-domain ID to clear the statistics for a specific bridge domain. The following example shows how to clear statistics for bridge domain ID 1.

```
device# clear statistics bridge-domain 1
```



802.1ag Connectivity Fault Management

[Enabling or disabling CFM on page 119](#)

[Creating a Maintenance Domain on page 120](#)

[Creating and configuring a Maintenance Association on page 120](#)

[Displaying CFM configurations on page 121](#)

Bridges are increasingly used in networks operated by multiple independent organizations, each with restricted management access to each other's equipment. CFM provides capabilities for detecting, verifying and isolating connectivity failures in such networks.

There are multiple organizations involved in a Metro Ethernet Service: Customers, Service Providers and Operators.

Customers purchase Ethernet Service from Service Providers. Service Providers may utilize their own networks, or the networks of other Operators to provide connectivity for the requested service. Customers themselves may be Service Providers, for example a Customer may be an Internet Service Provider which sells Internet connectivity.

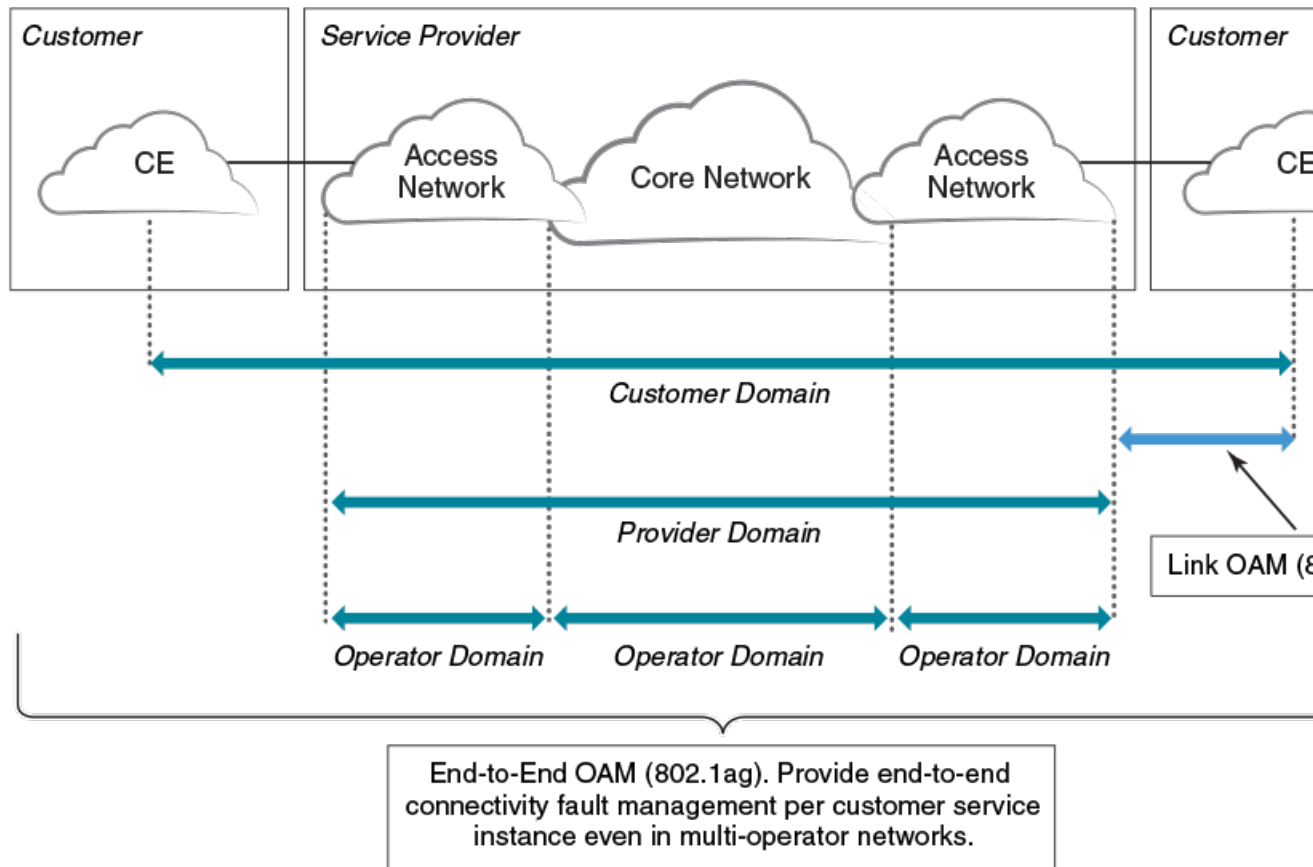


Figure 15: OAM Ethernet tools

Maintenance Domain (MD)

A Maintenance Domain is part of a network controlled by a single operator. In the following figure, a customer domain, provider domain and operator domain are described.

The Maintenance Domain (MD) levels are carried on all CFM frames to identify different domains. For example, in the following figure, some bridges belong to multiple domains. Each domain associates to an MD level.

- Customer Level: 5-7
- Provider Level: 3-4
- Operator Level: 0-2

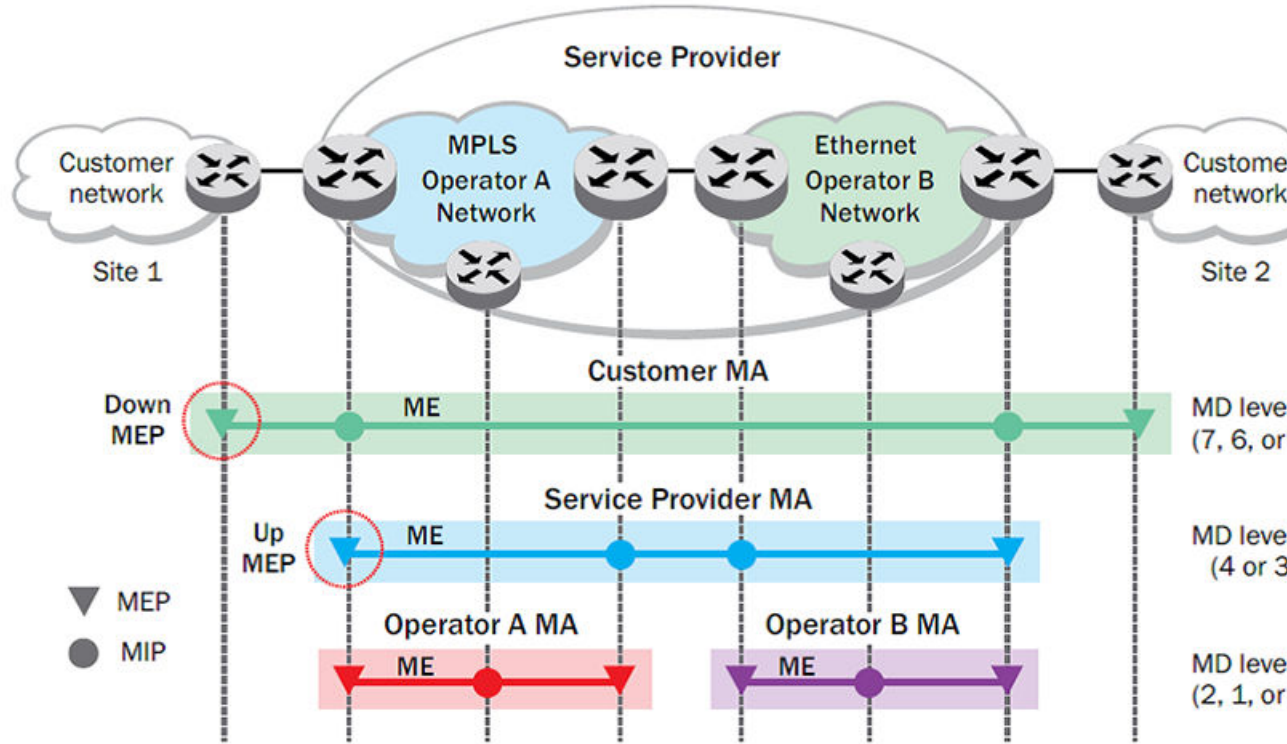


Figure 16: CFM deployment

Maintenance Association (MA)

Every MD can be further divided into smaller networks having multiple Maintenance End Points (MEP). Usually an MA is associated with a service instance (for example, a VLAN or a VPLS).

Maintenance End Point (MEP)

An MEP is located on the edge of an MA and defines the endpoint of the MA. Each MEP has unique ID (MEPID) within the MA. The connectivity in a MA is defined as connectivity between MEPs. MEPs generate a Continuity Check Messages that are multicast to all other MEPs in same MA to verify the connectivity.

Each MEP has a direction, down or up. Down MEPs receive CFM PDUs from the LAN and sends CFM PDUs towards the LAN. Up MEPs receive CFM PDUs from a bridge relay entity and sends CFM PDUs towards the bridge relay entity on a bridge. End stations support down MEPs only, as they have no bridge relay entities.

Maintenance Intermediate Point (MIP)

An MIP is located within a MA. It responds to Loopback and Linktrace messages for Fault isolation.

CFM Hierarchy

MD levels create a hierarchy in which 802.1ag messages sent by customer, service provider, and operators are processed by MIPs and MEPs at the respective level of the message. A common practice is for the service provider to set up a MIP at the customer MD level at the edge of the network, as shown in the figure above, to allow the customer to check continuity of the Ethernet service to the edge of the network. Similarly, operators set up MIPs at the service provider level at the edge of their respective networks, as shown in the figure above, to allow service providers to check the continuity of the Ethernet service to the edge of the operators' networks. Inside an operator network, all MIPs are at the respective operator level, also shown in the figure above.

Mechanisms of Ethernet IEEE 802.1ag OAM

Mechanisms supported by IEEE 802.1ag include Connectivity Check (CC), Loopback, and Link trace. Connectivity Fault Management allows for end-to-end fault management that is generally reactive (through Loopback and Link trace messages) and connectivity verification that is proactive (through Connectivity Check messages).

Fault detection (continuity check message)

Each MEP transmits periodic multicast CCMs towards other MEPs. For each MEP, there is 1 transmission and n-1 receptions per time period. Each MEP has a remote MEP database. It records the MAC address of remote MEPs.

Fault verification (Loopback messages)

A unicast Loopback Message is used for fault verification. A Loopback message helps a MEP identify the precise fault location along a given MA. A Loopback message is issued by a MEP to a given MIP along an MA. The appropriate MIP in front of the fault responds with a Loopback reply. The MIP behind the fault do not respond. For Loopback to work, the MEP must know the MAC address of the MIP to ping.

Fault isolation (Linktrace messages)

Linktrace mechanism is used to isolate faults at Ethernet MAC layer. Linktrace can be used to isolate a fault associated with a given Virtual Bridge LAN Service. Note that fault isolation in a connectionless (multi-point) environment is more challenging than a connection oriented (point-to-point) environment. In case of Ethernet, fault isolation can be even more challenging since a MAC address can age out when a fault isolates the MAC address. Consequently a network-isolating fault results in erasure of information needed for locating the fault.

Enabling or disabling CFM

To enable or disable the Connectivity Fault Management (CFM) protocol globally on the devices and enter into the CFM protocol configuration mode, enter the following command.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)#
```

The **no** form of the command disables the CFM protocol.

Creating a Maintenance Domain

A Maintenance Domain (MD) is the network or the part of the network for which faults in connectivity are to be managed. A Maintenance Domain consists of a set of Domain Service Access Points.

An MD is fully connected internally. A Domain Service Access Point associated with an MD has connectivity to every other Domain Service Access Point in the MD, in the absence of faults.

Each MD can be separately administered.

The **domain-name** command in Connectivity Fault Management (CFM) protocol configuration mode creates a maintenance domain with a specified level, name, and ID and enters the specific MD mode specified in the command argument.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)# domain-name md1 id 1 level 4
device(config-cfm-md-md1)#
```

The **no** form of the command removes the specified domain from the CFM protocol configuration mode.

Creating and configuring a Maintenance Association

Procedure

1. Create a MA within a specific domain, use the **ma-name** command.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)# domain name md1 id 1 level 4
device(config-cfm-md-md1)# ma-name ma1 id 1 vlan-id 30 priority 4
device(config-cfm-md-ma-ma1)#
```

This command changes the Maintenance Domain (MD) mode to the specific MA mode.

2. Set the time interval between two successive Continuity Check Messages (CCMs) that are sent by Maintenance End Points (MEP) in the specified MA, use the **ccm-interval** command.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)# domain name md1 id 1 level 4
device(config-cfm-md-md1)# ma-name ma1 id 1 vlan-id 30 priority 3
device(config-cfm-md-ma-ma1)# ccm-interval 10-second
device(config-cfm-md-ma-ma1)#
```

The **id** field specifies the short MAID format that is carried in the CCM frame. The default time interval is 10 seconds.

3. Add local ports as MEP to a specific maintenance association using the **mep** command in MA mode.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)# domain name md1 id 1 level 4
device(config-cfm-md-md1)# ma-name ma1 id 1 vlan-id 30 priority 3
device(config-cfm-md-ma-ma1)# mep 1 down ethernet 1/2
device(config-cfm-md-ma-mep-1)#
```

To configure a CFM packet to a **Down MEP**, you must send it out on the port on which it was configured. To configure a Connectivity Fault Management (CFM) packet to an **Up MEP**, you must

sent it to the entire VLAN for multicast traffic and the unicast traffic must be sent to a particular port as per the MAC table.

4. Configure the remote MEPs using the **remote-mep** command.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)# domain name mdl id 1 level 4
device(config-cfm-md-md1)# ma-name mal id 1 vlan-id 30 priority 3
device(config-cfm-md-ma-mal)# mep 1 down ethernet 1/2
device(config-cfm-md-ma-mep-1)# remote-mep 2
device(config-cfm-md-ma-mep-1)#
```

If a remote MEP is not specified, the remote MEP database is built based on the CCM. If one remote MEP never sends CCM, the failure cannot be detected.

5. Configure the conditions to automatically create MIPs on ports using the **mip-policy** command, in Maintenance Association mode.

```
device# configure terminal
device(config)# protocol cfm
device(config-cfm)# domain name mdl id 1 level 4
device(config-cfm-md-md1)#ma-name mal id 1 vlan-id 30 pri 7
device(config-cfm-md-ma-mal)#mip-policy explicit
device(config-cfm-md-ma-mal)#
```

A MIP can be created on a port and VLAN, only when explicit or default policy is defined for them. For a specific port and VLAN, a MIP is created at the lowest level. Additionally, the level created should be the immediate higher level than the MEP level defined for this port and VLAN.

Displaying CFM configurations

The following commands are used to display the CFM configurations and connectivity status.

show cfm

Use the **show cfm** command to display the Connectivity Fault Management (CFM) configuration.

```
device# show cfm
Domain: mdl
Index: 1
Level: 7
Maintenance association: ma5
MA Index: 5
CCM interval: 100 ms
Bridge-Domain ID: 50
Priority: 7
MAID Format: Short
```

MEP	Direction	MAC	PORT	VLAN	INNER-VLAN	PORT-STATUS-TLV
====	=====	=====	====	====	=====	=====
1	UP	609c.9f5f.700d	Eth 1/9	50	--	N



Note

For the **show cfm** command to generate output, you must first enable CFM in protocol configuration mode.

show cfm connectivity

Use the **show cfm connectivity** command to display the Connectivity Fault Management (CFM) configuration.

The following commands display the received port status tlv state at RMEP.

```
device# show cfm connectivity
Domain: mdl
Index: 1
Level: 7
Maintenance association: ma5
MA Index: 5
CCM interval: 100 ms
Bridge-Domain ID: 50
Priority: 7
MAID Format: Short
MEP Id: 1
MEP Port: Eth 1/9
RMEP  MAC                VLAN/PEER                INNER-VLAN                PORT                STATE
====  ===                =====                =====                =====                =====
2      609c.9f5e.4809          19.1.1.1                  --                    --                    OK
```



Note

For the **show cfm** command to generate output, you must first enable CFM in protocol configuration mode.

show cfm brief

Use the **show cfm brief** command to display the Connectivity Fault Management (CFM) brief output.

```
device# show cfm brief
Domain: mdl
Index: 1
Level: 7  Num of MA: 1
Maintenance association: ma5
MA Index: 5
CCM interval: 100 ms
Bridge-Domain ID: 50
Priority: 7
MAID Format: Short
Num of MEP: 1  Num of RMEP: 1
rmepfail: 0  rmepok: 1
```



802.1d Spanning Tree Protocol

[Spanning Tree Protocol overview](#) on page 123

[Spanning Tree Protocol configuration notes](#) on page 123

[STP features](#) on page 126

[STP parameters](#) on page 128

[Configuring STP](#) on page 130

Spanning Tree Protocol overview

The IEEE 802.1d Spanning Tree Protocol (STP) runs on bridges and switches that are 802.1d-compliant.

These variants are Rapid STP (RSTP), Multiple STP (MSTP), Per-VLAN Spanning Tree Plus (PVST+), and Rapid-PVST+ (R-PVST+)

When the spanning tree algorithm is run, the network switches transform the real network topology into a spanning tree topology. In an STP topology any LAN in the network can be reached from any other LAN through a unique path. The network switches recalculate a new spanning tree topology whenever there is a change to the network topology.

For each LAN, the switches that attach to the LAN select a designated switch that is the closest to the root switch. The designated switch forwards all traffic to and from the LAN. The port on the designated switch that connects to the LAN is called the designated port. The switches decide which of their ports is part of the spanning tree. A port is included in the spanning tree if it is a root port or a designated port.

STP runs one spanning tree instance (unaware of VLANs) and relies on long duration forward-delay timers for port state transition between disabled, blocking, listening, learning and forwarding states.

Spanning Tree Protocol configuration notes

The Extreme device supports STP as described in the IEEE 802.1d-1998 specification.

The STP is disabled by default on the Extreme device. Thus, any new VLANs you configure on the Extreme device have STP disabled by default.

Optional features

The following STP configuration features are optional:

- Root guard

- BPDU guard
- PortFast

STP states

A network topology of bridges typically contains redundant connections to provide alternate paths in case of link failures. The redundant connections create a potential for loops in the system. As there is no concept of time to live (TTL) in Ethernet frames, a situation may arise where there is a permanent circulation of frames when the network contains loops. To prevent this, a spanning tree connecting all the bridges is formed in real time.

Every Layer 2 interface running the STP is in one of these states:

State	Action or inaction
Blocking	The interface does not forward frames. Redundant ports are put in a blocking state and enabled when required. This is a transitional state after initialization.
Listening	The interface is identified by the spanning tree as one that should participate in frame forwarding. This is a transitional state after the blocking state for a legacy STP.
Learning	The interface prepares to participate in frame forwarding. This is a transitional state after the blocking state for a legacy STP.
Forwarding	The interface forwards frames. This is a transitional state after the learning state.
Disabled	The interface is not participating in a spanning tree because of shutdown of a port or the port is not operationally up. Any of the other states may transition into this state.

BPDUs

To construct a spanning tree requires knowledge of all the participants. The bridges must determine the root bridge and compute the port roles (root, designated, or blocked) with only the information that they have. To ensure that each bridge has enough information, the bridges use BPDUs to exchange information about bridge IDs and root path costs.

A bridge sends a BPDU frame using the unique MAC address of the port itself as a source address, and a destination address of the STP multicast address 01:80:C2:00:00:00.

BPDUs are exchanged regularly (every 2 seconds by default) and enable switches to keep track of network changes and to start and stop forwarding through ports as required.

When a device is first attached to a switch port, it does not immediately forward data. It instead goes through a number of states while it processes inbound BPDUs and determines the topology of the network. When a host is attached, after a listening and learning delay of about 30 seconds, the port always goes into the forwarding state. The time spent in the listening and learning states is determined by the forward delay. However, if instead another switch is connected, the port may remain in blocking mode if it would cause a loop in the network.

There are four types of BPDUs in the original STP specification:

- Configuration BPDU (CBPDU) is used for spanning tree computation.
- Topology Change Notification (TCN) BPDU is used to announce changes in the network topology.

- RSTP BPDU is used for RSTP
- MSTP BPDU is used for MSTP

TCN BPDUs

TCNs are injected into the network by a non-root switch and propagated to the root. Upon receipt of the TCN, the root switch will set a Topology Change flag in its normal BPDUs. This flag is propagated to all other switches to instruct them to rapidly age out their forwarding table entries.

Consider these configuration rules:

- TCN BPDUs are sent per VLAN.
- TCN BPDUs are sent only in those VLANs in which a topology change is detected.
- TCN BPDUs are sent only in those VLANs for which the bridge is not the root bridge.
- If a topology change is detected on a VLAN for which the bridge is the root bridge, the topology change flag is set in the configuration BPDU that is sent out.

For a given link, in conjunction with the configuration rules, a TCN BPDU is sent out as follows:

- On an access port, only a standard IEEE TCN BPDU is sent out. This TCN BPDU corresponds to a topology change in the access VLAN.
- On a trunk port, if VLAN 1 is allowed (either untagged or untagged), a standard IEEE TCN BPDU is sent for VLAN 1.
- On a trunk port, if the native VLAN is not 1, an untagged TCN BPDU is sent to Cisco or Extreme proprietary MAC address for that VLAN.
- On a trunk port, a tagged TCN BPDU is sent to Cisco or Extreme proprietary MAC address for a tagged VLAN.

As part of the response to TCN BPDUs, the Topology Change and Topology Change Acknowledgment flags are set in all configuration BPDUs corresponding to the VLAN for which the TCN was received.

When a topology change is detected on a trunk port, it is similar to detecting topology changes in each VLAN that is allowed on that trunk port. TCN BPDUs are sent for each VLAN as per the rules.

STP configuration guidelines and restrictions

- Only one form of a spanning tree protocol, such as STP or RSTP, can be enabled at a time. You must disable one form of xSTP before enabling another.
- When any form of STP is enabled globally, that form of STP is enabled by default on all switch ports.
- LAGs are treated as normal links for any form of STP.
- The STP is disabled by default on the SLX device. Thus, any new VLANs you configure on the SLX device have STP disabled by default.
- PVST/RPVST BPDUs are flooded only if PVST/RPVST is not enabled. STP/RSTP (IEEE) BPDUs are never flooded if STP/RSTP is not enabled.

Understanding the default STP configuration

Table 23: Default STP configuration

Parameter	Default setting
Spanning-tree mode	By default, STP, RSTP, and MSTP are disabled
Bridge priority	32768
Bridge forward delay	15 seconds
Bridge maximum aging time	20 seconds
Error disable timeout timer	Disabled
Error disable timeout interval	300 seconds
Port-channel path cost	Standard
Bridge hello time	2 seconds

The following table lists the switch defaults for the interface-specific configuration.

Table 24: Default interface specific configuration

Parameter	Default setting
Spanning tree	Enabled on the interface
Automatic edge detection	Disabled
Path cost	2000
Edge port	Disabled
Guard root	Disabled
Hello time	2 seconds
Link type	Point-to-point
Portfast	Disabled
Port priority	128
BPDU restriction	Restriction is disabled.

STP features

The following sections discuss root guard, BPDU guard, and PortFast.

Root guard

At times it is necessary to protect the root bridge from malicious attack or even unintentional misconfigurations where a bridge device that is not intended to be the root bridge becomes the root bridge, causing severe bottlenecks in the data path. These types of mistakes or attacks can be avoided by configuring root guard on ports of the root bridge.

The root guard feature provides a way to enforce the root bridge placement in the network and allows STP and its variants to interoperate with user network bridges while still maintaining the bridged network topology that the administrator requires. Errors are triggered if any change from the root bridge placement is detected.

When root guard is enabled on a port, it keeps the port in designated FORWARDING state. If the port receives a superior BPDU, which is a root guard violation, it sets the port into a DISCARDING state and triggers a Syslog message and an SNMP trap. No further traffic will be forwarded on this port. This allows the bridge to prevent traffic from being forwarded on ports connected to rogue or wrongly configured STP or RSTP bridges.

Root guard should be configured on all ports where the root bridge should not appear. In this way, the core bridged network can be cut off from the user network by establishing a protective perimeter around it.

Once the port stops receiving superior BPDUs, root guard automatically sets the port back to a FORWARDING state after the timeout period has expired.

BPDU guard

In a valid configuration, edge port-configured interfaces do not receive BPDUs. If an edge port-configured interface receives a BPDU, an invalid configuration exists, such as the connection of an unauthorized device. The BPDU Guard provides a secure response to invalid configurations because the administrator must manually put the interface back in service.

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

In some instances, it is unnecessary for a connected device, such as an end station, to initiate or participate in an STP topology change. In this case, you can enable the STP BPDU guard feature on the Extreme device port to which the end station is connected. The STP BPDU guard shuts down the port and puts it into an "error disabled" state. This disables the connected device's ability to initiate or participate in an STP topology. A log message is then generated for a BPDU guard violation, and a message is displayed to warn the network administrator of an invalid configuration.

The BPDU Guard provides a secure response to invalid configurations because the administrator must manually put the interface back in service with the **no shutdown** command if error disable recovery is not enabled by enabling the **errdisable-timeout** command. The interface can also be automatically configured to be enabled after a timeout. However, if the offending BPDUs are still being received, the port is disabled again.

Expected behavior in an interface context

When BPDU Guard is enabled on an interface, the device is expected to put the interface in Error Disabled state when BPDU is received on the port when edge-port and BPDU guard is enabled on the switch interface. When the port ceases to receive the BPDUs, it does not automatically switch to edge port mode, you must configure **error disable timeout** or **no shutdown** on the port to move the port back into edge port mode.

Error disable recovery

A port is placed into an error-disabled state when:

- A BPDU guard violation or loop detection violation occurs
- The number of inError packets exceeds the configured threshold
- An EFM-OAM enabled interface receives a critical event from the remote device (functionally equivalent to a disable state)

Once in an error disable state, the port remains in that state until it is re-enabled automatically or manually.

In STP, RSTP, MSTP, PVST+, or R-PVST+ mode, you can specify the time in seconds it takes for an interface to time out. The range is from 10 through 1000000 seconds. The default is 300 seconds. By default, the timeout feature is disabled.

PortFast

Consider the following when configuring PortFast:

- Do not enable PortFast on ports that connect to other devices.
- PortFast only needs to be enabled on ports that connect to workstations or PCs. Repeat this configuration for every port connected to workstations or PCs.
- Enabling PortFast on ports can cause temporary bridging loops, in both trunking and nontrunking mode.
- If BPDUs are received on a PortFast-enabled interface, the interface loses the edge port status unless it receives a **shutdown/no shutdown** command.
- PortFast immediately puts the interface into the forwarding state without having to wait for the standard forward time.

STP parameters

The following section discusses bridge parameters.

Bridge parameters

These parameters are set in STP, RSTP, MSTP, PVST+, and R-PVST+.

Bridge priority

Use this parameter to specify the priority of a device and to determine the root bridge.

Each device has a unique bridge identifier called the bridge ID. The bridge ID is an 8 byte value that is composed of two fields: a 2 B bridge priority field and the 6 B MAC address field. The value for the bridge priority ranges from 0 to 61440 in increments of 4096. The default value for the bridge priority is 32768. You use the **bridge-priority** command to set the appropriate values to designate a device as the root bridge or root device. A default bridge ID may appear as 32768.768e.£805.5800. If the bridge priorities are equal, the device with the lowest MAC address is elected the root.

After you decide what device to designate as the root, you set the appropriate device bridge priorities. The device with the lowest bridge priority becomes the root device. When a device has a bridge priority that is lower than that of all the other devices, it is automatically selected as the root.

The root device should be centrally located and not in a "disruptive" location. Backbone devices typically serve as the root because they usually do not connect to end stations. All other decisions in the network, such as which port to block and which port to put in forwarding mode, are made from the perspective of the root device.

You may also specify the bridge priority for a specific VLAN. If the VLAN parameter is not provided, the priority value is applied globally for all per-VLAN instances. However, for the VLANs that have been configured explicitly, the per-VLAN configuration takes precedence over the global configuration.

Bridge Protocol data units (BPDUs) carry information between devices. All the devices in the Layer 2 network, participating in any variety of STP, gather information on other devices in the network through an exchange of BPDUs. As the result of exchange of the BPDUs, the device with the lowest bridge ID is elected as the root bridge.

When setting the bridge forward delay, bridge maximum aging time, and the hello time parameters keep in mind that the following relationship should be kept:

$$(2 \times (\text{forward-delay} - 1)) \geq \text{max-age} \geq (2 \times (\text{hello-time} + 1))$$

Bridge forward delay

The bridge forward delay parameter specifies how long an interface remains in the listening and learning states before the interface begins forwarding all spanning tree instances. The valid range is from 4 through 30 seconds. The default is 15 seconds.

Additionally, you may specify the forward delay for a specific VLAN. If the VLAN parameter is not provided, the bridge forward delay value is applied globally for all per-VLAN instances. However, for the VLANs that have been configured explicitly, the per-VLAN configuration takes precedence over the global configuration.

Bridge maximum aging time

You can use this setting to configure the maximum length of time that passes before an interface saves its BPDU configuration information.

Keeping with the inequality shown above, when configuring the maximum aging time, you must set the value greater than the hello time. The range of values is 6 through 40 seconds while the default is 20 seconds.

You may specify the maximum aging for a specific VLAN. If the VLAN parameter is not provided, the priority value is applied globally for all per-VLAN instances. However, for the VLANs that have been configured explicitly, the per-VLAN configuration takes precedence over the global configuration.

Bridge hello time

You can use this parameter to set how often the device interface broadcasts hello BPDUs to other devices.

Use the **hello-time** command to configure the bridge hello time. The range is from 1 through 10 seconds. The default is 2 seconds.

You may also specify the hello time for a specific VLAN. If the VLAN parameter is not provided, the priority value is applied globally for all per-VLAN instances. However, for the VLANs that have been configured explicitly, the per-VLAN configuration takes precedence over the global configuration.

Error disable timeout parameter

These parameters are be set in STP, RSTP, MSTP, PVST+, and R-PVST+.

When the STP BPDU guard disables a port, the port remains in the disabled state unless the port is enabled manually. The parameter specifies the time in seconds it takes for an interface to time out. The range is from 10 through 1000000 seconds. The default is 300 seconds.

By default, the timeout feature is disabled.

Port-channel path cost parameter

This parameter can be set in STP, RSTP, MSTP, PVST+, and R-PVST+ mode.

There are two path cost options:

- Custom - Specifies that the path cost changes according to the port channel bandwidth.
- Standard - Specifies that the path cost does not change according to the port channel bandwidth.

The default port cost is standard.

Configuring STP

The following section discusses configuring STP.

Enabling and configuring STP globally

About This Task

You can enable STP or STP with one or more parameters enabled.

The parameters can be configured individually by:

1. Entering the commands in steps 1 and 2
2. Running the relevant parameter command
3. Verifying the result
4. Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable STP globally.

```
device(config)# protocol spanning-tree stp
```

A spanning tree can be disabled by entering the **no protocol spanning-tree stp** command.

3. Describe or name the STP.

```
device(config-stp)# description stp1
```

A description is not required.

4. Specify the bridge priority.

```
device(config-stp)# bridge-priority 4096
```

The bridge with the lowest priority number (highest priority) is designated the root bridge. The range of values is 0 through 61440; values can be set only in increments of 4096. The default priority is 32678.

5. Specify the bridge forward delay.

```
device(config-stp)# forward-delay 20
```

The forward delay specifies how long an interface remains in the listening and learning states before it begins forwarding all spanning tree instances. The valid range is from 4 through 30 seconds. The default is 15 seconds.

6. Configure the maximum aging time.

```
device(config-stp)# max-age 25
```

This parameter controls the maximum length of time that passes before an interface saves its BPDU configuration information. You must set the maximum age to be greater than the hello time. The range is 6 through 40 seconds. The default is 20 seconds.

7. Configure the maximum hello time.

```
device(config-stp)# hello-time 8
```

The hello time determines how often the switch interface broadcasts hello BPDUs to other devices. The default is 2 seconds while the range is from 1 through 10 seconds.

8. Enable the error disable timeout timer.

```
device(config-stp)# error-disable-timeout enable
```

This parameter enables a timer that brings a port out of the disabled state. By default, the timeout feature is disabled.

9. Set the error disable timeout timer.

```
device(config-stp)# error-disable-timeout interval 60
```

When enabled the default is 300 seconds and the range is from 10 through 1000000 seconds.

10. Configure the port channel path cost.

```
device(config-stp)# port-channel path-cost custom
```

Specifying **custom** means the path cost changes according to the port channel's bandwidth.

11. Return to privileged EXEC mode.

```
device(config-stp)# end
```

12. Save the configuration.

```
device# copy running-config startup-config
```

STP configuration example

```
device# configure terminal
device(config)# protocol spanning-tree stp
device(config-stp)# description stpForInterface
device(config-stp)# bridge-priority 4096
device(config-stp)# forward-delay 20
device(config-stp)# max-age 25
device(config-stp)# hello-time 8
device(config-stp)# error-disable-timeout enable
device(config-stp)# error-disable-timeout interval 60
device(config-stp)# port-channel path-cost custom
device(config-stp)# end
device# show spanning-tree brief
device# copy running-config startup-config
```

Enabling and configuring STP on an interface

Before You Begin

Globally enable STP and STP parameters.

About This Task

The parameters can be configured individually by:

1. Entering the commands in steps 1-3
2. Running the relevant parameter command
3. Verifying the result
4. Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter interface configuration mode.

```
device(config)# interface ethernet 0/20
```

3. Enable the interface.

```
device(conf-if-eth-0/20)# no shutdown
```

4. Configure the path cost for spanning tree calculations on the interface.

```
device(conf-if-eth-0/20)# spanning-tree cost 10000
```

The lower the path cost means a greater chance that the interface becomes the root port. The range is 1 through 200000000. The default path cost is assigned as per the port speed.

5. Enable BPDU guard on the interface.

```
device(conf-if-eth-0/20)# spanning-tree port-fast bpdu-guard
```

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

6. Configure Root Guard on the interface.

```
device(conf-if-eth-0/20)# spanning-tree guard root
```

Root Guard protects the root bridge from malicious attacks and unintentional misconfigurations where a bridge device that is not intended to be the root bridge becomes the root bridge.

7. Specify an interface link-type.

```
device(conf-if-eth-0/20)# spanning-tree link-type point-to-point
```

Specifying a point-to-point link enables rapid spanning tree transitions to the forwarding state. Specifying a shared link disables spanning tree rapid transitions. The default setting is point-to-point.

8. Specify port priority to influence the selection of root or designated ports.

```
device(conf-if-eth-0/20)# spanning-tree priority 64
```

The range is from 0 through 240 in increments of 16. The default value is 128.

9. Verify the configuration.

```
device# show spanning-tree brief
```

```
Spanning-tree Mode: Spanning Tree Protocol
```

```
Root ID      Priority 4096
             Address 768e.f805.5800
             Hello Time 8, Max Age 25, Forward Delay 20
```

```
Bridge ID   Priority 4096
             Address 768e.f805.5800
             Hello Time 8, Max Age 25, Forward Delay 20
```

Interface	Role	Sts	Cost	Prio	Link-type	Edge
Eth 0/2	DES	FWD	2000	128	P2P	No
Eth 0/20	DES	FWD	1000	64	P2P	No
Eth 0/25	DIS	DIS	20000000	128	P2P	No
Eth 0/30	DIS	DIS	20000000	128	P2P	No

```
Eth 0/31    DIS    DIS    2000000    128    P2P        No
```

**Note**

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $38 \geq 25 \geq 18$

```
device# show running-config interface ethernet 0/20
interface ethernet 0/20
switchport
switchport mode access
switchport access val 1
spanning-tree cost 1000
spanning-tree guard root
spanning-tree link-type point-to-point
spanning-tree portfast bpdu-guard
spanning-tree priority 64
```

10. Save the settings by copying the running configuration to the startup configuration.

```
device# copy running-config startup-config
```

STP on an interface configuration example

```
device# configure terminal
device(config)# interface ethernet 0/20
device(conf-if-eth-0/20)# no shutdown
device(conf-if-eth-0/20)# spanning-tree cost 10000
device(conf-if-eth-0/20)# spanning-tree port-fast bpdu-guard
device(conf-if-eth-0/20)# spanning-tree guard root
device(conf-if-eth-0/20)# spanning-tree link-type point-to-point
device(conf-if-eth-0/20)# spanning-tree priority 64
device(conf-if-eth-0/20)# end
device# show spanning-tree brief
device# copy running-config startup-config
```

Configuring basic STP parameters

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable STP globally

```
device(config)# protocol spanning-tree stp
```

3. Name the STP.

```
device(config-stp)# description stp1
```

4. Designate the root switch.

```
device(config-stp)# bridge-priority 28672
```

The priority values can be set only in increments of 4096. The range is 0 through 61440.

5. Specify the bridge forward delay.

```
device(config-stp)# forward-delay 20
```

6. Configure the maximum aging time.

```
device(config-stp)# max-age 25
```

7. Configure the maximum hello time.

```
device(config-stp)# hello-time 8
```

8. Enable the error disable timeout timer.

```
device(config-stp)# error-disable-timeout enable
```

9. Set the error disable timeout timer interval.

```
device(config-stp)# error-disable-timeout interval 60
```

10. Enable port fast on switch ports.

- a. Configure port fast on Ethernet port 0/1.

```
device(config)# interface ethernet 0/1
device(config-if-eth-0/1)# spanning-tree portfast
device(config-if-eth-0/1)# exit
```

Spanning trees are automatically enabled on switch ports.

- b. Configure port fast on Ethernet port 0/2.

```
device(config)# interface ethernet 0/2
device(config-if-eth-0/2)# spanning-tree portfast
device(config-if-eth-0/2)# exit
```

- c. Repeat these commands for every port connected to workstations or PCs.

```
device(config)# interface ethernet ...
```

11. Specify port priorities to influence the selection of the root and designated ports.

```
device(config)# interface ethernet 0/1
device(config-if-eth-0/1)# spanning-tree priority 1
device(config-if-eth-0/1)# exit
```

12. Enable the guard root feature.

```
device(config)# interface ethernet 0/12
device(config-if-eth-0/12)# no shutdown
device(config-if-eth-0/12)# spanning-tree guard root
```

Root guard lets the device top participate in the STP but only when the device does not attempt to become the root.

13. Return to privileged exec mode.

```
device(conf-if-eth-0/12)# end
```

14. Verify the configuration.

```
device# show spanning-tree brief
Spanning-tree Mode: Spanning Tree Protocol
Root ID Priority 4096
Address 768e.f805.5800
Hello Time 8, Max Age 25, Forward Delay 20
Bridge ID Priority 4096
Address 768e.f805.5800
Hello Time 8, Max Age 25, Forward Delay 20
Interface      Role      Sts      Cost      Prio      Link-type      Edge
-----
Eth 0/1        DES      FWD      2000      128      P2P            No
Eth 0/2        DES      FWD      2000      128      P2P            No
Eth 0/12       DES      FWD      2000      128      P2P            No
```

Observe that the settings comply with the formula set out in the STP parameter configuration section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case $38 \geq 25 \geq 18$.

15. Save the configuration.

```
device# copy running-config startup-config
```

Basic STP configuration example

```
device# configure terminal
device(config)# protocol spanning-tree stp
device(config-stp)# description stp1
device(conf-stp)# bridge-priority 28672
device(config-stp)# forward-delay 20
device(config-stp)# max-age 25
device(config-stp)# hello-time 8
device(config-stp)# error-disable-timeout enable
device(config-stp)# error-disable-timeout interval 60
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# spanning-tree portfast
device(conf-if-eth-0/1)# exit
device(config)# interface ethernet 0/2
device(conf-if-eth-0/2)# spanning-tree portfast
device(conf-if-eth-0/2)# exit
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# spanning-tree priority 1
device(conf-if-eth-0/1)# exit
device(config)# interface ethernet 0/12
device(conf-if-eth-0/12)# no shutdown
device(conf-if-eth-0/12)# spanning-tree guard root
device(conf-if-eth-0/12)# end
device# show spanning-tree brief
device# copy running-config startup-config
```


Re-enabling an error-disabled port automatically

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter STP configuration mode.

```
device(config)# protocol spanning-tree stp
```

3. Enable the error-disable-timeout timer.

```
device(config-stp)# error-disable-timeout enable
```

4. Set an interval after which port shall be enabled.

```
device(config-stp)# error-disable-timeout interval 60
```

The interval range is from 0 to 1000000 seconds, the default is 300 seconds.

5. Return to privileged EXEC mode.

```
device(config-stp)# end
```

6. Verify the configuration.

```
device# show spanning-tree
Spanning-tree Mode: Spanning Tree Protocol

Root Id: 8000.768e.f805.5800 (self)
Bridge Id: 8000.768e.f805.5800

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0

Bpdu-guard errdisable timeout: enabled
Bpdu-guard errdisable timeout interval: 60 sec
```

Automatically re-enable an error-disabled port configuration example

```
device# configure terminal
device(config)# protocol spanning-tree stp
device(config-stp)# error-disable-timeout enable
device(config-stp)# error-disable-timeout interval 60
device(config-stp)# end
device# show spanning-tree
```

Clearing spanning tree counters

Procedure

1. Clear spanning tree counters on all interfaces.

```
device# clear spanning-tree counter
```

2. Clear spanning tree counters on a specified Ethernet interface.

```
device# clear spanning-tree counter interface ethernet 0/3
```

3. Clear spanning tree counters on a specified port channel interface.

```
device# clear spanning-tree counter interface port-channel 12
```

Port channel interface numbers range from 1 through 64.

Clearing spanning tree-detected protocols

About This Task

These commands force a spanning tree renegotiation with neighboring devices on either all interfaces or on a specified interface.

Procedure

1. Restart the spanning tree migration process on all interfaces.

```
device# clear spanning-tree detected-protocols
```

2. Restart the spanning tree migration process on a specific Ethernet interface.

```
device# clear spanning-tree detected-protocols interface ethernet 0/3
```

3. Restart the spanning tree migration process on a specific port channel interface.

```
device# clear spanning-tree detected-protocols port-channel 12
```

Port channel interface numbers range from 1 through 64.

Shutting down STP

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Shut down STP.

- Shut down STP globally and return to privileged EXEC mode.

```
device(config)# protocol spanning-tree stp
device(config-stp)# shutdown
device(config-stp)# end
```

- Shut down STP on a specific VLAN and return to privileged EXEC mode.

```
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-vlan-10)# end
```

3. Verify the configuration.

```
device# show spanning-tree
device#
```

4. Save the running configuration to the startup configuration.

```
device# copy running-config startup-config
```

Shut down STP configuration example

```
device# configure terminal
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-stp)# end
device# show spanning-tree
device# copy running-config startup-config
```



Note

Shutting down STP on a VLAN is used in this example.



802.1w Rapid Spanning Tree Protocol

[Rapid Spanning Tree Protocol overview](#) on page 140
[Configuring RSTP](#) on page 142

Rapid Spanning Tree Protocol overview

The STP (802.1d) standard was designed at a time when recovering connectivity after an outage within a minute or so was considered adequate performance. With the advent of Layer 3 switching in LAN environments, bridging competes with routed solutions where protocols such as OSPF are able to provide an alternate path in less time.

The RSTP can be seen as evolution of STP standard. It provides rapid convergence of connectivity following the failure of bridge, a bridge port or a LAN. It provides rapid convergence of edge ports, new root ports and port connected through point-to-point links. The port, which qualifies for fast convergence, is derived from the duplex mode of a port. A port operating in full-duplex will be assumed to be point-to-point, while a half-duplex port will be considered as a shared port by default. This automatic setting can be overridden by explicit configuration.

RSTP is designed to be compatible and interoperate with the STP. However, the benefit of the RSTP fast convergence is lost when interacting with legacy STP (802.1d) bridges since the RSTP downgrades itself to the STP when it detects a connection to a legacy bridge.

The states for every Layer 2 interface running the RSTP are as follows:

State	Action
Learning	The interface prepares to participate in frame forwarding.
Forwarding	The interface forwards frames.
Discarding	The interface discards frames. Ports in the discarding state do not take part in the active topology and do not learn MAC addresses. Note: The STP disabled, blocking, and listening states are merged into the RSTP discarding state.

The RSTP port roles for the interface are also different. The RSTP differentiates explicitly between the state of the port and the role it plays in the topology. The RSTP uses the root port and designated port roles defined in the STP, but splits the blocked port role into backup port and alternate port roles:

Backup port	Provides a backup for the designated port and can only exist where two or more ports of the switch are connected to the same LAN; the LAN where the bridge serves as a designated switch.
Alternate port	Serves as an alternate port for the root port providing a redundant path towards the root bridge.

Only the root port and the designated ports are part of the active topology; the alternate and backup ports do not participate in it. When the network is stable, the root and the designated ports are in the forwarding state, while the alternate and backup ports are in the discarding state. When there is a topology change, the new RSTP port roles allow a faster transition of an alternate port into the forwarding state.

For more information about spanning trees, see the introductory sections in the [802.1d Spanning Tree Protocol](#) chapter.

RSTP parameters

The parameters you would normally set when you configure STP are applicable to RSTP. Before you configure RSTP see the STP parameters sections for descriptions of the bridge parameters, the error disable timeout parameter and the port channel path cost parameter.

There is one parameter that can be configured in RSTP that is not available in STP; the transmit hold count. This parameter configures the BPDU burst size by specifying the maximum number of BPDUs transmitted per second for before pausing for 1 second. The range is 1 through 10 while the default is 6. See the section Enabling RSTP and configuring RSTP parameters for the procedure to configure this parameter.

The edge port and auto edge features can be enabled in RSTP as well. See the section Edge port and automatic edge detection and the section Configuring RSTP on an interface for descriptions of these features and how they are configured.

Edge port and automatic edge detection

From an interface, you can configure a device to automatically identify the edge port. The port can become an edge port if no BPDU is received. By default, automatic edge detection is disabled.

Follow these guidelines to configure a port as an edge port:

- When edge port is enabled, the port still participates in a spanning tree.
- A port can become an edge port if no BPDU is received.

- When an edge port receives a BPDU, it becomes a normal spanning tree port and is no longer an edge port.
- Because ports that are directly connected to end stations cannot create bridging loops in the network, edge ports transition directly to the forwarding state and skip the listening and learning states.

**Note**

If BPDUs are received on a port fast enabled interface, the interface loses the edge port status unless it receives a **shutdown** or **no shutdown** command.

Configuring RSTP

Enabling and configuring RSTP globally

About This Task

See the section STP parameters for parameters applicable to all STP variants.

You can enable RSTP or RSTP with one or more parameters enabled. The parameters can be enabled or changed individually by entering the commands in steps 1 and 2, running the parameter command, verifying the result, and then saving the configuration.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable RSTP.

```
device(config)# protocol spanning-tree rstp
```

You can shut down RSTP by entering the **shutdown** command.

3. Designate the root device.

```
device(conf-rstp)# bridge-priority 28582
```

The range is 0 through 61440 and the priority values can be set only in increments of 4096.

You can shut down RSTP by entering the **shutdown** command when in RSTP configuration mode.

4. Configure the bridge forward delay value.

```
device(conf-rstp)# forward-delay 15
```

5. Configure the bridge maximum aging time value.

```
device(conf-rstp)# max-age 20
```

6. Enable the error disable timeout timer.

- a. Enable the timer.

```
device(conf-rstp)# error-disable-timeout enable
```

- b. Configure the error disable timeout interval value.

```
device(conf-rstp)# error-disable-timeout interval 60
```

7. Configure the port-channel path cost.

```
device(conf-rstp)# port-channel path-cost custom
```

8. Configure the bridge hello-time value.

```
device(conf-rstp)# hello-time 2
```

9. Specify the transmit hold count.

```
device(config-rstp)# transmit-holdcount 5
```

This command configures the maximum number of BPDUs transmitted per second.

10. Return to privileged exec mode.

```
device(conf-rstp)# end
```

11. Save the configuration.

```
device# copy running-config startup-config
```

Enabling RSTP and configuring RSTP parameters example

Enabling and configuring RSTP on an interface

About This Task

You can configure the parameters individually on an interface by doing the following:

1. Entering the commands in Steps 1 through 3.
2. Specifying additional parameters, as appropriate.
3. Verifying the result.
4. Saving the configuration.

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enter interface subtype configuration mode.

```
device(config)# interface ethernet 0/10
```

3. Enable the interface.

```
device(conf-if-eth-0/10)# no shutdown
```

To disable the spanning tree on the interface you use the **spanning-tree shutdown** command.

- Specify the port priority on the interface.

```
device(conf-if-eth-0/10)# spanning-tree priority 128
```

The range is from 0 through 240 in increments of 16. The default value is 128.

- Specify the path cost on the interface.

```
device(conf-if-eth-0/10)# spanning-tree cost 20000000
```

The lower the path cost means a greater chance that the interface becomes the root port. The range is 1 through 200000000. The default path cost is assigned as per the port speed.

- Enable edge port.

```
device(conf-if-eth-0/10)# spanning-tree edgeport
```

If BPDUs are received on a port fast enabled interface, the interface loses the edge port status unless it receives a **shutdown** or **no shutdown** command.

- Enable BPDU guard on the interface.

```
device(conf-if-eth-0/10)# spanning-tree edgeport bpdu-guard
```

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

- Enable automatic edge detection on the interface.

```
device(conf-if-eth-0/10)# spanning-tree autoedge
```

You use this command to automatically identify the edge port. A port becomes an edge port if it receives no BPDUs. By default, automatic edge detection is disabled.

- Enable root guard on the interface.

```
device(conf-if-eth-0/10)# spanning-tree guard root
```

Root guard protects the root bridge from malicious attacks and unintentional misconfigurations where a bridge device that is not intended to be the root bridge becomes the root bridge.

- Specify a link type on the interface.

```
device(conf-if-eth-0/10)# spanning-tree link-type point-to-point
```



Note

The link type is explicitly configured as **point-to-point** rather than **shared**.

- Return to privileged EXEC mode.

```
device(conf-if-eth-0/10)# end
```

- Verify the configuration.

```
device# show spanning-tree
```

```
Spanning-tree Mode: Rapid Spanning Tree Protocol
```



```

Root Id: 8000.01e0.5200.0180 (self)
Bridge Id: 8000.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0

Bpdu-guard errdisable timeout: enabled
Bpdu-guard errdisable timeout interval: 60 sec

Port Eth 0/10 enabled
  Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Spanning Tree Protocol - Received None - Sent STP
  Edgeport: on; AutoEdge: yes; AdminEdge: no; EdgeDelay: 3 sec
  Configured Root guard: on; Operational Root guard: on
  Bpdu-guard: on
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

```

Interface	Role	Sts	Cost	Prio	Link-type	Edge
Eth 0/10	DES	FWD	20000000	128	P2P	No

The **forward-delay**, **hello-time**, and **max-age** parameters are set globally, not on the interface.

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

13. Save the configuration.

```
device# copy running-config startup-config
```

RSTP on an interface configuration example

```

device# configure terminal
device(config)# interface ethernet 0/10
device(conf-if-eth-0/10)# no spanning-tree shutdown
device(conf-if-eth-0/10)# spanning-tree priority 128
device(conf-if-eth-0/10)# spanning-tree cost 20000000
device(conf-if-eth-0/10)# spanning-tree edgeport
device(conf-if-eth-0/10)# spanning-tree edgeport bpdu-guard
device(conf-if-eth-0/10)# spanning-tree autoedge
device(conf-if-eth-0/10)# spanning-tree guard root
device(conf-if-eth-0/10)# spanning-tree link-type point-to-point
device(conf-if-eth-0/10)# end
device# show spanning-tree
device# copy running-config startup-config

```

Configuring basic RSTP parameters

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable RSTP.

```
device(config)# protocol spanning-tree rstp
```

3. Designate the root device.

```
device(conf-rstp)# bridge-priority 28582
```

4. Enable the error disable timeout timer value.

```
device(conf-rstp)# error-disable-timeout enable
```

5. Configure the error-disable-timeout interval value.

```
device(conf-rstp)# error-disable-timeout interval 60
```

6. Enable edge port on switch ports.

- a. Enter interface subtype configuration mode for the switchport.

```
device(conf-rstp)# interface ethernet 0/10
```

- b. Enable edge port.

```
device(conf-if-eth-0/10)# spanning-tree edge-port
```

- c. Return to global configuration mode.

```
device(conf-if-eth-0/10)# exit
```

- d. Repeat the above steps for all ports that connect to a workstation or PC.

7. Specify port priorities.

- a. Enter interface subtype configuration mode.

```
device(config)# interface ethernet 0/11
```

- b. Configure the port priority.

```
device(conf-if-eth-0/11)# spanning-tree priority 1
```

- c. Return to global configuration mode.

```
device(conf-if-eth-0/11)# exit
```

8. Enable the guard root feature.

- a. Enter interface configuration mode.

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

- b. Configure the port priority.

```
device(conf-if-eth-0/1)# spanning-tree guard root
```

- c. Return to privileged EXEC mode.

```
device(conf-if-eth-0/1)# exit
```

9. Verify the configuration.

```
device# show spanning-tree

Spanning-tree Mode: Rapid Spanning Tree Protocol

Root Id: 4096.01e0.5200.0180 (self)
Bridge Id: 4096.01e0.5200.0180
Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0
Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec
switch# show spanning-tree brief
Spanning-tree Mode: Rapid Spanning Tree Protocol
Root ID Priority 4096
Address 768e.f805.5800
Hello Time 2, Max Age 20, Forward Delay 15
Bridge ID Priority 4096
Address 768e.f805.5800
```

Interface	Role	Sts	Cost	Prio	Link-type	Edge
Eth 0/1	DES	FWD	2000	128	P2P	No
Eth 0/10	DES	FWD	2000	128	P2P	No
Eth 0/11	DES	FWD	2000	128	P2P	No

Observe that the settings comply with the formula set out in the STP parameters section, as follows:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

10. Save the configuration.

```
device# copy running-config startup-config
```

Basic RSTP configuration example

```
device# configure terminal
device(config)# protocol spanning-tree rstp
device(conf-rstp)# bridge-priority 28582
device(conf-rstp)# error-disable-timeout enable
device(conf-rstp)# error-disable-timeout interval 60
device(conf-rstp)# interface ethernet 0/10
device(conf-if-eth-0/10)# spanning-tree edge-port
device(conf-if-eth-0/10)# exit
device(config)# interface ethernet 0/11
device(conf-if-eth-0/11)# spanning-tree priority 1
device(conf-if-eth-0/11)# exit
device(config)# interface ethernet 0/1
device(conf-if-eth-0/1)# spanning-tree guard root
device(conf-if-eth-0/1)# exit
```

```
device# show spanning-tree
device# copy running-config startup-config
```

Clearing spanning tree counters

Procedure

1. Clear spanning tree counters on all interfaces.

```
device# clear spanning-tree counter
```

2. Clear spanning tree counters on a specified Ethernet interface.

```
device# clear spanning-tree counter interface ethernet 0/3
```

3. Clear spanning tree counters on a specified port channel interface.

```
device# clear spanning-tree counter interface port-channel 12
```

Port channel interface numbers range from 1 through 64.

Clearing spanning tree-detected protocols

About This Task

These commands force a spanning tree renegotiation with neighboring devices on either all interfaces or on a specified interface.

Procedure

1. Restart the spanning tree migration process on all interfaces.

```
device# clear spanning-tree detected-protocols
```

2. Restart the spanning tree migration process on a specific Ethernet interface.

```
device# clear spanning-tree detected-protocols interface ethernet 0/3
```

3. Restart the spanning tree migration process on a specific port channel interface.

```
device# clear spanning-tree detected-protocols port-channel 12
```

Port channel interface numbers range from 1 through 64.

Shutting down RSTP

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Shut down RSTP.

- Shut down STP globally and return to privileged EXEC mode.

```
device(config)# protocol spanning-tree rstp
```

```
device(conf-rstp)# shutdown
device(conf-rstp)# end
```

- Shut down RSTP on a specific VLAN and return to privileged EXEC mode.

```
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-vlan-10)# end
```

3. Verify the configuration.

```
device# show spanning-tree
device#
```

4. Save the configuration.

```
device# copy running-config startup-config
```



Per-VLAN Spanning Tree+ and Rapid Per-VLAN Spanning Tree+

[PVST+ and R-PVST+ overview on page 150](#)

[Configuring PVST+ and R-PVST+ on page 152](#)

PVST+ and R-PVST+ overview

Both the STP and the RSTP build a single logical topology. A typical network has multiple VLANs. A single logical topology does not efficiently utilize the availability of redundant paths for multiple VLANs. A single logical topology does not efficiently utilize the availability of redundant paths for multiple VLANs. If a port is set to the blocked state or the discarding state for one VLAN (under the STP or the RSTP), it is the same for all other VLANs. PVST+ builds on the STP on each VLAN, and R-PVST+ builds on the RSTP on each VLAN.

PVST+ R-PVST+ provide interoperability with Cisco PVST and R-PVST and other vendor switches which implement Cisco PVST or R-PVST. the PVST+ and R-PVST+ implementations are extensions to PVST and R-PVST, which can interoperate with an STP topology, including MSTP (CIST), on Extreme and other vendor devices sending untagged IEEE BPDUs.

PVST+ and R-PVST+ guidelines and restrictions

Consider the following when configuring PVST+ and R-PVST+:

- Extreme supports PVST+ and R-PVST+ only. The PVST and R-PVST protocols are proprietary to Cisco and are not supported.
- A port native VLAN is the native VLAN ID associated with a trunk port on an Extreme switch. This VLAN ID is associated with all untagged packets on the port. The default native VLAN ID for a trunk port is 1.
- IEEE compliant switches run just one instance of STP protocol shared by all VLANs, creating a Mono Spanning Tree (MST). A group of such switches running a single spanning tree forms an MST region.
- You can configure up to 128 PVST+ or R-PVST+ instances. If you have more than 128 VLANs configured on the switch and enable PVST then the first 128 VLANs are PVST/+ or R-PVST+ enabled.
- In PVST/+ or R-PVST+ mode, when you are connected to a Cisco or MLX switch, the Cisco proprietary MAC address to which the BPDUs are sent/processed must be explicitly configured on a per-port basis.
- In PVST/+ or R-PVST+ mode, when you connect to a Cisco switch using a trunk port, the Extreme switch must have a native VLAN configured on the trunk port (same configuration as on the other side).

- A Common Spanning Tree (CST) is the single spanning tree instance used by Extreme switches to interoperate with 802.1q bridges. This spanning tree instance stretches across the entire network domain (including PVST, PVST+ and 802.1q regions). It is associated with VLAN 1 on the Extreme switch.
- In order to interact with STP and IEEE 802.1q trunk, PVST evolved to PVST+ to interoperate with STP topology by STP BPDU on the native or default VLAN.
- A group of switches running PVST+ is called a PVST+ region.

For more information about spanning trees, see the introductory sections in the Spanning Tree Protocol chapter.

PVST+ and R-PVST+ parameters

The parameters you would normally set when you configure STP are applicable to PVST+ and R-PVST+. Before you configure PVST+ or R-PVST+ parameters see the sections in the Standing Tree Protocol chapter explaining bridge parameters, the error disable timeout parameter and the port channel path cost parameter.

There is one parameter that can be configured in R-PVST+ that is not available in STP or PVST+; the transmit hold count. This parameter configures the BPDU burst size by specifying the maximum number of BPDUs transmitted per second for before pausing for 1 second. The range is 1 through 10 while the default is 6. See the section Configuring R-PVST+ for the procedure to configure this parameter.

PortFast

Consider the following when configuring PortFast:

- Do not enable PortFast on ports that connect to other devices.
- PortFast only needs to be enabled on ports that connect to workstations or PCs. Repeat this configuration for every port connected to workstations or PCs.
- Enabling PortFast on ports can cause temporary bridging loops, in both trunking and nontrunking mode.
- If BPDUs are received on a PortFast-enabled interface, the interface loses the edge port status unless it receives a **shutdown/no shutdown** command.
- PortFast immediately puts the interface into the forwarding state without having to wait for the standard forward time.

Edge port and automatic edge detection

From an interface, you can configure a device to automatically identify the edge port. The port can become an edge port if no BPDU is received. By default, automatic edge detection is disabled.

Follow these guidelines to configure a port as an edge port:

- When edge port is enabled, the port still participates in a spanning tree.
- A port can become an edge port if no BPDU is received.

- When an edge port receives a BPDU, it becomes a normal spanning tree port and is no longer an edge port.
- Because ports that are directly connected to end stations cannot create bridging loops in the network, edge ports transition directly to the forwarding state and skip the listening and learning states.

**Note**

If BPDUs are received on a port fast enabled interface, the interface loses the edge port status unless it receives a **shutdown** or **no shutdown** command.

Configuring PVST+ and R-PVST+

Enabling and configuring PVST+ globally

About This Task

You can enable PVST+ with one or more parameters configured. The parameters can be configured or changed individually by entering the commands in steps 1 and 2, running the parameter command, verifying the result, and then saving the configuration.

For more information about spanning trees and spanning tree parameters, see the introductory sections in the Spanning Tree Protocol chapter.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable PVST+.

```
device(config)# protocol spanning-tree pvst
```

3. Configure the bridge priority for the common instance.

```
device(config-pvst)# bridge-priority 4096
```

Valid values range from 0 through 61440 in increments of 4096. Assigning a lower priority value indicates that the bridge might become root.

You can shut down PVST+ by entering the **shutdown** command when in PVST configuration mode.

4. Configure the forward delay parameter.

```
device(config-pvst)# forward-delay 11
```

5. Configure the hello time parameter.

```
device(config-pvst)# hello-time 2
```

6. Configure the maximum age parameter.

```
device(config-pvst)# max-age 7
```


- Return to privileged exec mode.

```
device(config-pvst)# end
```

- Verify the configuration.

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $20 \geq 7 \geq 6$.

- Save the configuration.

```
device# copy running-config startup-config
```

PVST+ configuration example

```
device# configure terminal
device(config)# protocol spanning-tree pvst
device(config-pvst)# bridge-priority 4096
device(config-pvst)# forward-delay 11
device(config-pvst)# hello-time 2
device(config-pvst)# max-age 7
device(config-pvst)# end
device# show spanning-tree brief
device# copy running-config startup-config
```

What to Do Next

For more information about configuring PVST+ parameters, see [STP parameter configuration](#). PVST+, R-PVST+, and other types of spanning trees share many tasks with STP.

Enabling and configuring PVST+ on an interface

About This Task

The ports and parameters can be configured individually on a system by:

- Entering the commands in steps 1, and 2
- Running the relevant addition steps and parameter commands
- Verifying the result
- Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

- Enter global configuration mode.

```
device# configure terminal
```

- Enable PVST+.

```
device(config)# protocol spanning-tree pvst
```

3. Enter interface configuration mode.

```
device(config-pvst)# interface ethernet 0/3
```

4. Enable spanning tree on the interface.

```
device(conf-if-eth-0/3)# no spanning-tree shutdown
```

5. Configure the interface link type.

```
device(conf-if-eth-0/3)# spanning-tree link-type point-to-point
```

6. Specify the port priority to influence the selection of root or designated ports.

```
device(conf-if-eth-0/3)# spanning-tree priority 64
```

The range is from 0 through 240 in increments of 16. The default value is 128.

7. Configure the path cost for spanning tree calculations on the interface.

```
device(conf-if-eth-0/3)# spanning-tree cost 10000
```

The lower the path cost means a greater chance that the interface becomes the root port. The range is 1 through 200000000. The default path cost is assigned as per the port speed.

8. Configure the path cost for spanning tree calculations a specific VLAN.

```
device(conf-if-eth-0/3)# spanning-tree vlan 10 cost 10000
```

The lower the path cost means a greater chance that the interface becomes the root port. The range is 1 through 200000000. The default path cost is assigned as per the port speed.

9. Enable root guard on the interface.

```
device(conf-if-eth-0/3)# spanning-tree guard root
```

Root guard protects the root bridge from malicious attacks and unintentional misconfigurations where a bridge device that is not intended to be the root bridge becomes the root bridge.

10. Enable BPDU guard on the interface.

```
device(conf-if-eth-0/3)# spanning-tree port-fast bpdu-guard
```

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

11. Return to privileged EXEC mode.

```
device(conf-if-eth-0/3)# exit
```

12. Verify the configuration.

```
device# show spanning-tree brief
```

```
Spanning-tree Mode: PVST Protocol
```

```
Root ID      Priority 4096
Address 768e.f805.5800
Hello Time 8, Max Age 25, Forward Delay 20
```

```

Bridge ID      Priority 4096
               Address 768e.f805.5800
               Hello Time 8, Max Age 25, Forward Delay 20

```

Interface	Role	Sts	Cost	Prio	Link-type	Edge
Eth 0/3	DES	FWD	200000	64	P2P	No

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case :38 ≥ 25 ≥ 18.

13. Save the configuration.

```
device# copy running-config startup-config
```

PVST+ on an interface configuration example

```

device# configure terminal
device(config)# protocol spanning-tree pvst
device(conf-pvst)# interface ethernet 0/3
device(conf-if-eth-0/3)# no spanning-tree shutdown
device(conf-if-eth-0/3)# spanning-tree link-type point-to-point
device(conf-if-eth-0/3)# spanning-tree priority 64
device(conf-if-eth-0/3)# spanning-tree cost 10000
device(conf-if-eth-0/3)# spanning-tree vlan 10 cost 10000
device(conf-if-eth-0/3)# spanning-tree guard root
device(conf-if-eth-0/3)# spanning-tree port-fast bpdu-guard
device(conf-if-eth-0/3)# exit
device# show spanning-tree
device# copy running-config startup-config

```

Enabling and configuring PVST+ on a system

About This Task

The ports and parameters can be configured individually on a system by:

1. Entering the commands in steps 1, and 2
2. Running the relevant addition steps and parameter commands
3. Verifying the result
4. Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable PVST+.

```
device(config)# protocol spanning-tree pvst
```

3. Configure the bridge priority for the common instance.

```
device(config-pvst)# bridge-priority 4096
```

Valid values range from 0 through 61440 in multiples of 4096. Assigning a lower priority value indicates that the bridge might become root.

4. Configure the forward delay parameter.

```
device(config-pvst)# forward-delay 15
```

5. Configure the hello time parameter.

```
device(config-pvst)# hello-time 2
```

6. Configure the maximum age parameter.

```
device(config-pvst)# max-age 20
```

7. Add VLANs.

- a. Configure VLAN 100 with a priority of 0.

```
device(config-pvst)# vlan 100 priority 0
```

The bridge priority is configured in multiples of 4096.

- b. Configure VLAN 201 with a priority of 12288.

```
device(config-pvst)# vlan 201 priority 12288
```

- c. Configure VLAN 301 with a priority of 20480.

```
device(config-pvst)# vlan 301 priority 20480
```

8. Set the switching characteristics for interface 0/3.

- a. Enter interface configuration mode.

```
device(config)# interface ethernet 0/3
```

- b. Set the switching characteristics of the interface.

```
device(conf-if-eth-0/3)# switchport
```

- c. Set the interface mode to trunk.

```
device(conf-if-eth-0/3)# switchport mode trunk
```

- d. Add VLAN 100 as a member VLAN.

```
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 100
```

- e. Add VLAN 201 as a member VLAN.

```
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 201
```

- f. Add VLAN 301 as a member VLAN.

```
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 301
```

- g. Enable spanning tree on the interface.

```
device(conf-if-eth-0/3)# no spanning-tree shutdown
```

- h. Return to privileged EXEC mode.

```
device(conf-if-eth-0/3)# exit
```

9. Set the switching characteristics for interface 0/4.

- a. Enter interface configuration mode.

```
device(config)# interface ethernet 0/4
```

- b. Set the switching characteristics of the interface.

```
device(conf-if-eth-0/4)# switchport
```

- c. Set the interface mode to trunk.

```
device(conf-if-eth-0/4)# switchport mode trunk
```

- d. Add VLAN 100 as a member VLAN.

```
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 100
```

- e. Add VLAN 201 as a member VLAN.

```
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 201
```

- f. Add VLAN 301 as a member VLAN.

```
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 301
```

- g. Enable spanning tree on the interface.

```
device(conf-if-eth-0/4)# no spanning-tree shutdown
```

- h. Return to privileged EXEC mode.

```
device(conf-if-eth-0/4)# exit
```

10. To interoperate with switches other than VDX switches in PVST+ mode, you must configure the interface that is connected to that switch.

- a. Enter interface configuration mode for the port that interoperates with a VDX device.

```
device(config)# interface ethernet 0/12
```

- b. Specify the MAC address for the device.

```
device(conf-if-eth-0/12)# spanning-tree bpdu-mac 0100.0ccc.cccd
```

- c. Enable spanning tree on the interface.

```
device(conf-if-eth-0/12)# no spanning-tree shutdown
```

- d. Return to privileged EXEC mode.

```
device(conf-if-eth-0/12)# end
```

11. Verify the configuration.

```
device# show spanning-tree

VLAN 1

Spanning-tree Mode: PVST Protocol

Root Id: 0001.01e0.5200.0180 (self)
Bridge Id: 0001.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
  Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
  Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

VLAN 100

Spanning-tree Mode: PVST Protocol

Root Id: 0064.01e0.5200.0180 (self)
Bridge Id: 0064.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0
```

```
Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
  Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
  Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

VLAN 201

Spanning-tree Mode: PVST Protocol

Root Id: 30c9.01e0.5200.0180 (self)
Bridge Id: 30c9.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
  Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
  Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
```

```

Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

VLAN 301

Spanning-tree Mode: PVST Protocol

Root Id: 512d.01e0.5200.0180 (self)
Bridge Id: 512d.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
Designated Path Cost: 0
Configured Path Cost: 20000000
Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
Designated Path Cost: 0
Configured Path Cost: 20000000
Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

```

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

12. Save the configuration.

```
device# copy running-config startup-config
```


Enable PVST+ on a system configuration example

```
device# configure terminal
device(config)# protocol spanning-tree pvst
device(config-pvst)# bridge-priority 4096
device(config-pvst)# forward-delay 15
device(config-pvst)# hello-time 2
device(config-pvst)# max-age 20
device(config-pvst)# vlan 100 priority 0
device(config-pvst)# vlan 201 priority 12288
device(config-pvst)# vlan 301 priority 20480
device(config-pvst)# interface ethernet 0/3
device(conf-if-eth-0/3)# switchport
device(conf-if-eth-0/3)# switchport mode trunk
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 100
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 201
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 301
device(conf-if-eth-0/3)# no spanning-tree shutdown
device(conf-if-eth-0/3)# exit
device(config)# interface ethernet 0/4
device(conf-if-eth-0/4)# switchport
device(conf-if-eth-0/4)# switchport mode trunk
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 100
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 201
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 301
device(conf-if-eth-0/4)# no spanning-tree shutdown
device(conf-if-eth-0/4)# end
device# show spanning-tree
device# copy running-config startup-config
```

Enabling and configuring R-PVST+ globally

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable R-PVST+.

```
device(config)# protocol spanning-tree rpvst
```

3. Configure the bridge priority for the common instance.

```
device(config-rpvst)# bridge-priority 4096
```

Valid priority values range from 0 through 61440 in multiples of 4096. Assigning a lower priority value indicates that the bridge might become root.

4. Configure the forward delay parameter.

```
device(config-rpvst)# forward-delay 20
```

5. Configure the hello time parameter.

```
device(config-rpvst)# hello-time 22
```

6. Configure the maximum age parameter.

```
device(config-rpvst)# max-age 8
```

- Set the transmit hold count for the bridge.

```
device(config-rpvst)# transmit-holdcount 9
```

This command configures the maximum number of BPDUs transmitted per second before pausing for 1 second. The range is 1 through 10. The default is 6.

- Return to privileged exec mode.

```
device(config-rpvst)# end
```

- Verify the configuration.

```
device# show spanning-tree brief
VLAN 1

Spanning-tree Mode: Rapid PVST Protocol

    Root ID          Priority 4096
                Address 01e0.5200.0180
                Hello Time 2, Max Age 7, Forward Delay 11

    Bridge ID        Priority 32769
                Address 01e0.5200.0180
                Hello Time 8, Max Age 22, Forward Delay 20, Tx-HoldCount 9
                Migrate Time 3 sec

Interface      Role  Sts  Cost      Prio  Link-type      Edge
-----
```

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $20 \geq 7 \geq 6$.

- Save the configuration.

```
device# copy running-config startup-config
```

R-PVST+ configuration example

```
device# configure terminal
device(config)# protocol spanning-tree rpvst
device(config-rpvst)# bridge-priority 4096
device(config-rpvst)# forward-delay 20
device(config-rpvst)# hello-time 22
device(config-rpvst)# max-age 8
device(config-rpvst)# transmit-holdcount 9
device(config-rpvst)# end
device# show spanning-tree brief
device# copy running-config startup-config
```

What to Do Next

For more information about configuring parameters, see the section STP parameter configuration.

Enabling and configuring R-PVST+ on an interface

About This Task

The ports and parameters can be configured individually on a system by:

1. Entering the commands in steps 1-3
2. Running the relevant addition steps and parameter commands
3. Verifying the result
4. Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable R-PVST+.

```
device(config)# protocol spanning-tree rpvst
```

3. Enter interface configuration mode.

```
device(config-rpvst)# interface ethernet 0/3
```

4. Enable the spanning tree on the interface.

```
device(conf-if-eth-0/3)# no spanning-tree shutdown
```

5. Configure the interface link type.

```
device(conf-if-eth-0/3)# spanning-tree link-type point-to-point
```

6. Specify the port priority to influence the selection of root or designated ports.

```
device(conf-if-eth-0/3)# spanning-tree priority 64
```

The range of priority values is from 0 through 240 in multiples of 16. The default value is 128.

7. Configure the path cost for spanning tree calculations on the interface.

```
device(conf-if-eth-0/3)# spanning-tree cost 200000
```

The lower the path cost means a greater chance that the interface becomes the root port. The range is 1 through 200000000. The default path cost is assigned as per the port speed.

8. Configure the path cost for spanning tree calculations a specific VLAN.

```
device(conf-if-eth-0/3)# spanning-tree vlan 10 cost 10000
```

The lower the path cost means a greater chance that the interface becomes the root port. The range is 1 through 200000000. The default path cost is assigned as per the port speed.

9. Enable automatic edge detection on the interface.

```
device(conf-if-eth-0/3)# spanning-tree autoedge
```

You use this command to automatically identify the edge port. A port becomes an edge port if it receives no BPDUs. By default, automatic edge detection is disabled.

10. Enable root guard on the interface.

```
device(conf-if-eth-0/3)# spanning-tree guard root
```

Root guard protects the root bridge from malicious attacks and unintentional misconfigurations where a bridge device that is not intended to be the root bridge becomes the root bridge.

11. Enable the spanning tree on the edge port.

```
device(conf-if-eth-0/3)# spanning-tree edgeport
```

If BPDUs are received on a port fast enabled interface, the interface loses the edge port status unless it receives a **shutdown** or **no shutdown** command.

12. Enable BPDU guard on the interface.

```
device(conf-if-eth-0/3)# spanning-tree edgeport bpdu-guard
```

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

13. Return to privileged EXEC mode.

```
device(conf-if-eth-0/3)# exit
```

14. Verify the configuration.

```
device# show spanning-tree brief

Spanning-tree Mode: Rapid PVST Protocol

    Root ID          Priority 4096
                   Address 768e.f805.5800
                   Hello Time 8, Max Age 25, Forward Delay 20

    Bridge ID        Priority 4096
                   Address 768e.f805.5800
                   Hello Time 8, Max Age 25, Forward Delay 20

Interface   Role  Sts  Cost      Prio  Link-type  Edge
-----
Eth 0/3     DES   FWD  200000    128   P2P        No
```

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $38 \geq 25 \geq 18$.

15. Save the configuration.

```
device# copy running-config startup-config
```

R-PVST+ on an interface configuration example

```
device# configure terminal
device(config)# protocol spanning-tree rpvst
device(config-rpvst)# interface ethernet 0/3
device(conf-if-eth-0/3)# no spanning-tree shutdown
device(conf-if-eth-0/3)# spanning-tree link-type point-to-point
device(conf-if-eth-0/3)# spanning-tree priority 64
device(conf-if-eth-0/3)# spanning-tree cost 200000
device(conf-if-eth-0/3)# spanning-tree vlan 10 cost 10000
device(conf-if-eth-0/3)# spanning-tree autoedge
device(conf-if-eth-0/3)# spanning-tree guard root
device(conf-if-eth-0/3)# spanning-tree edgeport
device(conf-if-eth-0/3)# spanning-tree edgeport bpdu-guard
device(conf-if-eth-0/3)# exit
device# show spanning-tree
device# copy running-config startup-config
```

Enabling and configuring R-PVST+ on a system

About This Task

The ports and parameters can be configured individually by:

1. Entering the commands in steps 1 and 2
2. Running the relevant addition steps and parameter commands
3. Verifying the result
4. Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable R-PVST+.

```
device(config)# protocol spanning-tree rpvst
```

You can shut down R-PVST+ by entering the **shutdown** command when in `rpvst` configuration mode.

3. Configure the bridge priority for the common instance.

```
device(config-rpvst)# bridge-priority 4096
```

Valid values range from 0 through 61440 in increments of 4096. Assigning a lower priority value indicates that the bridge might become root.

4. Configure the forward delay parameter.

```
device(config-rpvst)# forward-delay 20
```

5. Configure the hello time parameter.

```
device(config-rpvst)# hello-time 8
```

6. Configure the maximum age parameter.

```
device(config-rpvst)# max-age 22
```

7. Specify the transmit hold count.

```
device(config-rpvst)# transmit-holdcount 5
```

This command configures the maximum number of BPDUs transmitted per second. The range of values is 1 through 10.

8. Configure VLANs.

- a. Configure VLAN 100 with a priority of 0.

```
device(config-rpvst)# vlan 100 priority 0
```

Valid priority values range from 0 through 61440 in multiples of 4096.

- b. Configure VLAN 201 with a priority of 12288.

```
device(config-rpvst)# vlan 201 priority 12288
```

- c. Configure VLAN 301 with a priority of 20480.

```
device(config-rpvst)# vlan 301 priority 20480
```

9. Set the switching characteristics for interface 0/3.

- a. Enter interface configuration mode.

```
device(config-rpvst)# interface ethernet 0/3
```

- b. Set the switching characteristics of the interface.

```
device(conf-if-eth-0/3)# switchport
```

- c. Set the interface mode to trunk.

```
device(conf-if-eth-0/3)# switchport mode trunk
```

- d. Add VLAN 100 as a member VLAN.

```
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 100
```

- e. Add VLAN 201 as a member VLAN.

```
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 201
```

- f. Add VLAN 301 as a member VLAN.

```
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 301
```

- g. Enable spanning tree on the interface.

```
device(conf-if-eth-0/3)# no spanning-tree shutdown
```

- h. Return to privileged EXEC mode.

```
device(conf-if-eth-0/3)# exit
```

10. Set the switching characteristics for interface 0/4.

- a. Enter interface configuration mode.

```
device(config-rpvst)# interface ethernet 0/4
```

- b. Set the switching characteristics of the interface.

```
device(conf-if-eth-0/4)# switchport
```

- c. Set the interface mode to trunk.

```
device(conf-if-eth-0/4)# switchport mode trunk
```

- d. Add VLAN 100 as a member VLAN.

```
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 100
```

- e. Add VLAN 201 as a member VLAN.

```
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 201
```

- f. Add VLAN 301 as a member VLAN.

```
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 301
```

- g. Enable spanning tree on the interface.

```
device(conf-if-eth-0/4)# no spanning-tree shutdown
```

- h. Return to privileged EXEC mode.

```
device(conf-if-eth-0/4)# exit
```

11. To interoperate with switches other than VDX switches in R-PVST+ mode, you must configure the interface that is connected to that switch.

- a. Enter interface configuration mode for the port that interoperates with a VDX switch.

```
device(config)# interface ethernet 0/12
```

- b. Specify the MAC address for the device.

```
device(conf-if-eth-0/12)# spanning-tree bpdu-mac 0100.0ccc.cccd
```

- c. Enable spanning tree on the interface.

```
device(conf-if-eth-0/12)# no spanning-tree shutdown
```

- d. Return to privileged EXEC mode.

```
device(conf-if-eth-0/12)# end
```

12. Verify the configuration.

```
device# show spanning-tree

VLAN 1

Spanning-tree Mode: Rapid PVST Protocol

Root Id: 0001.01e0.5200.0180 (self)
Bridge Id: 0001.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 20; Hello Time: 8; Max Age: 22; Max-hops: 20
Tx-HoldCount 5
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
  Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
  Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

VLAN 100

Spanning-tree Mode: Rapid PVST Protocol

Root Id: 0064.01e0.5200.0180 (self)
Bridge Id: 0064.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 20; Hello Time: 8; Max Age: 22; Max-hops: 20
Tx-HoldCount 5
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
  Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
```



```
Designated Path Cost: 0
Configured Path Cost: 20000000
Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
Designated Path Cost: 0
Configured Path Cost: 20000000
Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

VLAN 201

Spanning-tree Mode: Rapid PVST Protocol

Root Id: 30c9.01e0.5200.0180 (self)
Bridge Id: 30c9.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 20; Hello Time: 8; Max Age: 22; Max-hops: 20
Tx-HoldCount 5
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
Designated Path Cost: 0
Configured Path Cost: 20000000
Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
Designated Path Cost: 0
Configured Path Cost: 20000000
Designated Port Id: 0; Port Priority: 128
Designated Bridge: 0000.0000.0000.0000
Number of forward-transitions: 0
Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
Portfast: off
```

```

Configured Root guard: off; Operational Root guard: off
Bpdu-guard: off
Link-type: point-to-point
Received BPDUs: 0; Sent BPDUs: 0

VLAN 301

Spanning-tree Mode: Rapid PVST Protocol

Root Id: 512d.01e0.5200.0180 (self)
Bridge Id: 512d.01e0.5200.0180

Root Bridge Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 20; Hello Time: 8; Max Age: 22; Max-hops: 20
Tx-HoldCount 5
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec

Port Et 0/3 enabled
  Ifindex: 201351168; Id: 8001; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

Port Et 0/4 enabled
  Ifindex: 201359360; Id: 8002; Role: Disabled; State: Disabled
  Designated Path Cost: 0
  Configured Path Cost: 20000000
  Designated Port Id: 0; Port Priority: 128
  Designated Bridge: 0000.0000.0000.0000
  Number of forward-transitions: 0
  Version: Per-VLAN Spanning Tree Protocol - Received None - Sent STP
  Portfast: off
  Configured Root guard: off; Operational Root guard: off
  Bpdu-guard: off
  Link-type: point-to-point
  Received BPDUs: 0; Sent BPDUs: 0

```

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

13. Save the configuration.

```
device# copy running-config startup-config
```

Enable R-PVST+ on a system configuration example

```
device# configure terminal
device(config)# protocol spanning-tree rpvst
```

```
device(config-rpvst)# bridge-priority 4096
device(config-rpvst)# forward-delay 20
device(config-rpvst)# hello-time 8
device(config-rpvst)# max-age 22
device(config-rpvst)# transmit-holdcount 5
device(config-rpvst)# vlan 100 priority 0
device(config-rpvst)# vlan 201 priority 12288
device(config-rpvst)# vlan 301 priority 20480
device(config-rpvst)# interface ethernet 0/3
device(conf-if-eth-0/3)# switchport
device(conf-if-eth-0/3)# switchport mode trunk
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 100
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 201
device(conf-if-eth-0/3)# switchport trunk allowed vlan add 301
device(conf-if-eth-0/3)# no spanning-tree shutdown
device(conf-if-eth-0/3)# exit
device(config)# interface ethernet 0/4
device(conf-if-eth-0/4)# switchport
device(conf-if-eth-0/4)# switchport mode trunk
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 100
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 201
device(conf-if-eth-0/4)# switchport trunk allowed vlan add 301
device(conf-if-eth-0/4)# no spanning-tree shutdown
device(conf-if-eth-0/4)# end
device# show spanning-tree
device# copy running-config startup-config
```

Clearing spanning tree counters

Procedure

1. Clear spanning tree counters on all interfaces.

```
device# clear spanning-tree counter
```

2. Clear spanning tree counters on a specified Ethernet interface.

```
device# clear spanning-tree counter interface ethernet 0/3
```

3. Clear spanning tree counters on a specified port channel interface.

```
device# clear spanning-tree counter interface port-channel 12
```

Port channel interface numbers range from 1 through 64.

Clearing spanning tree-detected protocols

About This Task

These commands force a spanning tree renegotiation with neighboring devices on either all interfaces or on a specified interface.

Procedure

1. Restart the spanning tree migration process on all interfaces.

```
device# clear spanning-tree detected-protocols
```

- Restart the spanning tree migration process on a specific Ethernet interface.

```
device# clear spanning-tree detected-protocols interface ethernet 0/3
```

- Restart the spanning tree migration process on a specific port channel interface.

```
device# clear spanning-tree detected-protocols port-channel 12
```

Port channel interface numbers range from 1 through 64.

Shutting down PVST+ or R-PVST+

Procedure

- Enter global configuration mode.

```
device# configure terminal
```

- Shut down PVST+ or R-PVST+.

- Shut down PVST+ or R-PVST+ globally and return to privileged EXEC mode.

```
device(config)# protocol spanning-tree pvst
device(config-pvst)# shutdown
device(config-pvst)# end
```

- Shut down PVST+ or R-PVST+ on a specific interface and return to privileged EXEC mode.

```
device(config)# interface ethernet 0/2
device(config-if-eth-0/2)# spanning-tree shutdown
device(config-if-eth-0/2)# end
```

- Shut down PVST+ or R-PVST+ on a specific VLAN. and return to privileged EXEC mode.

```
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-vlan-10)# end
```

- Verify the configuration.

```
device# show spanning-tree
device#
```

- Save the running configuration to the startup configuration.

```
device# copy running-config startup-config
```

Shut down PVST+ or R-PVST+ configuration example

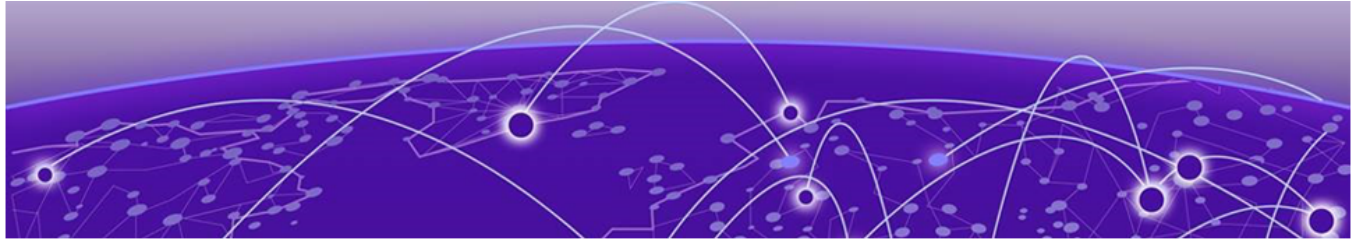
```
device# configure terminal
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-vlan-10)# end
```

```
device# show spanning-tree  
device# copy running-config startup-config
```



Note

Shutting down PVST+ on a VLAN is used in this example.



802.1s Multiple Spanning Tree Protocol

[MSTP overview on page 174](#)

[MSTP global level parameters on page 176](#)

[MSTP interface level parameters on page 176](#)

[Configuring MSTP on page 177](#)

MSTP overview

MSTP uses RSTP to group VLANs into separate spanning-tree instance. Each instance has its own spanning-tree topology independent of other spanning tree instances, which allows multiple forwarding paths, permits load balancing, and facilitates the movement of data traffic. A failure in one instance does not affect other instances. By enabling the MSTP, you are able to more effectively utilize the physical resources present in the network and achieve better load balancing of VLAN traffic.

The MSTP evolved as a compromise between the two extremes of the RSTP and R-PVST+, it was standardized as IEEE 802.1s and later incorporated into the IEEE 802.1Q-2003 standard. The MSTP configures a meshed topology into a loop-free, tree-like topology. When the link on a bridge port goes up, an MSTP calculation occurs on that port. The result of the calculation is the transition of the port into either a forwarding or blocking state. The result depends on the position of the port in the network and the MSTP parameters. All the data frames are forwarded over the spanning tree topology calculated by the protocol.



Note

Multiple switches must be configured consistently with the same MSTP configuration to participate in multiple spanning tree instances. A group of interconnected switches that have the same MSTP configuration is called an MSTP region.

MSTP is backward compatible with the STP and the RSTP.

Common Spanning Tree (CST)

The single Spanning Tree instance used by the Extreme device, and other vendor devices to interoperate with MSTP bridges. This spanning tree instance stretches across the entire network domain (including PVST, PVST+ and MSTP regions). It is associated with VLAN 1 on the Extreme device.

Internal Spanning Tree (IST)

An MSTP bridge must handle at least these two instances: one IST and one or more MSTIs (Multiple Spanning Tree Instances). Within each MST region, the MSTP maintains multiple spanning-tree

instances. Instance 0 is a special instance known as IST, which extends CST inside the MST region. IST always exists if the device runs MSTP. Besides IST, this implementation supports up to 31 MSTIs.

Common Internal Spanning Tree (CIST)

The single spanning tree calculated by STP (including PVST+) and RSTP (including R-PVST+) and the logical continuation of that connectivity through MSTP bridges and regions, calculated by MSTP to ensure that all LANs in the bridged LAN are simply and fully connected

Multiple Spanning Tree Instance (MSTI)

One of a number of spanning trees calculated by the MSTP within an MST Region, to provide a simply and fully connected active topology for frames classified as belonging to a VLAN that is mapped to the MSTI by the MST configuration table used by the MST bridges of that MST region.

The Extreme implementation supports up to 32 spanning tree instances in an MSTP enabled bridge that can support up to 32 different Layer 2 topologies. The spanning tree algorithm used by the MSTP is the RSTP, which provides quick convergence.

By default all configured VLANs including the default VLAN are assigned to and derive port states from CIST until explicitly assigned to MSTIs.

MST regions

MST regions are clusters of bridges that run multiple instances of the MSTP protocol. Multiple bridges detect that they are in the same region by exchanging their configuration (instance to VLAN mapping), name, and revision-level. Therefore, if you need to have two bridges in the same region, the two bridges must have identical configurations, names, and revision-levels. Also, one or more VLANs can be mapped to one MST instance (IST or MSTI) but a VLAN cannot be mapped to multiple MSTP instances

MSTP regions

MSTP introduces a hierarchical way of managing device domains using regions. Devices that share common MSTP configuration attributes belong to a region. The MSTP configuration determines the MSTP region where each device resides. The common MSTP configuration attributes are as follows:

- Alphanumeric configuration name (32 bytes)
- Configuration revision number (2 bytes)
- 4096-element table that maps each of the VLANs to an MSTP instance

Region boundaries are determined by the above attributes. An MSTI is an RSTP instance that operates inside an MSTP region and determines the active topology for the set of VLANs mapping to that instance. Every region has a CIST that forms a single spanning tree instance which includes all the devices in the region. The difference between the CIST instance and the MSTP instance is that the CIST instance operates across the MSTP region and forms a loop-free topology across regions, while the MSTP instance operates only within a region. The CIST instance can operate using the RSTP only if all the devices across the regions support the RSTP. However, if any of the devices operate using the STP, the CIST instance reverts to the STP.

Each region is viewed logically as a single STP or a single RSTP bridge to other regions.

**Note**

Extreme supports 32 MSTP instances and one MSTP region.

For more information about spanning trees, see the introductory sections in the Spanning Tree Protocol chapter.

MSTP global level parameters

To configure a switch for MSTP, first you set the region name and the revision on each switch that is being configured for MSTP. You must then create an MSTP Instance and assign an ID. VLANs are then assigned to MSTP instances. These instances must be configured on all switches that interoperate with the same VLAN assignments.

Each of the steps used to configure and operate MSTP are described in the following:

**Note**

The MSTP Region and Revision global parameters are enabled for interface level parameters as described below.

- Set the MSTP region name — Each switch that is running MSTP is configured with a name. It applies to the switch which can have many different VLANs that can belong to many different MSTP regions. The default MSTP name is "NULL".
- Set the MSTP revision number — Each switch that is running MSTP is configured with a revision number. It applies to the switch, which can have many different VLANs that can belong to many different MSTP regions.
- Enabling and disabling Cisco interoperability — While in MSTP mode, use the **cisco-interoperability** command to enable or disable the ability to interoperate with certain legacy Cisco switches. If Cisco interoperability is required on any switch in the network, then all switches in the network must be compatible, and therefore enabled by means of this command. By default the Cisco interoperability is disabled.
- The parameters you would normally set when you configure STP are applicable to MSTP. Before you configure MSTP parameters see the sections explaining bridge parameters, the error disable timeout parameter and the port-channel path cost parameter in the STP section of this guide.

MSTP interface level parameters

Edge port and automatic edge detection

From an interface, you can configure a device to automatically identify the edge port. The port can become an edge port if no BPDU is received. By default, automatic edge detection is disabled.

Follow these guidelines to configure a port as an edge port:

- When edge port is enabled, the port still participates in a spanning tree.
- A port can become an edge port if no BPDU is received.

- When an edge port receives a BPDU, it becomes a normal spanning tree port and is no longer an edge port.
- Because ports that are directly connected to end stations cannot create bridging loops in the network, edge ports transition directly to the forwarding state and skip the listening and learning states.

**Note**

If BPDUs are received on a port fast enabled interface, the interface loses the edge port status unless it receives a **shutdown** or **no shutdown** command.

BPDU guard

In a valid configuration, edge port-configured interfaces do not receive BPDUs. If an edge port-configured interface receives a BPDU, an invalid configuration exists, such as the connection of an unauthorized device. The BPDU Guard provides a secure response to invalid configurations because the administrator must manually put the interface back in service.

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

In some instances, it is unnecessary for a connected device, such as an end station, to initiate or participate in an STP topology change. In this case, you can enable the STP BPDU guard feature on the Extreme device port to which the end station is connected. The STP BPDU guard shuts down the port and puts it into an "error disabled" state. This disables the connected device's ability to initiate or participate in an STP topology. A log message is then generated for a BPDU guard violation, and a message is displayed to warn the network administrator of an invalid configuration.

The BPDU Guard provides a secure response to invalid configurations because the administrator must manually put the interface back in service with the **no shutdown** command if error disable recovery is not enabled by enabling the **errdisable-timeout** command. The interface can also be automatically configured to be enabled after a timeout. However, if the offending BPDUs are still being received, the port is disabled again.

Expected behavior in an interface context

When BPDU Guard is enabled on an interface, the device is expected to put the interface in Error Disabled state when BPDU is received on the port when edge-port and BPDU guard is enabled on the switch interface. When the port ceases to receive the BPDUs, it does not automatically switch to edge port mode, you must configure **error disable timeout** or **no shutdown** on the port to move the port back into edge port mode.

Configuring MSTP

Enabling and configuring MSTP globally

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable MSTP.

```
device(config)# protocol spanning-tree mstp
```

This command creates a context for MSTP. MSTP is automatically enabled. All MSTP specific CLI commands can be issued only from this context. Entering **no protocol spanning-tree mstp** deletes the context and all the configurations defined within the context.

3. Specify the region name.

```
device(config-mstp)# region kerry
```

4. Specify the revision number.

```
device(config-mstp)# revision 1
```

5. Configure an optional description of the MSTP instance.

```
device(config-mstp)# description kerry switches
```

6. Specify the maximum hops for a BPDU to prevent the messages from looping indefinitely on the interface.

```
device(config-mstp)# max-hops 25
```

Setting this parameter prevents messages from looping indefinitely on the interface. The range is 1 through 40 hops while the default is 20 .

7. Map VLANs to MSTP instances and set the instance priority.

- a. Map VLANs 7 and 8 to instance 1.

```
device(config-mstp)# instance 1 vlan 7,8
```

- b. Map VLANs 21, 22, and 23 to instance 2.

```
device(config-mstp)# instance 2 vlan 21-23
```

- c. Set the priority of instance 1.

```
device(config-mstp)# instance 1 priority 4096
```

This command can be used only after the VLAN is created. VLAN instance mapping is removed from the configuration if the underlying VLANs are deleted.

8. Configure a bridge priority for the CIST bridge.

```
device(config-mstp)# bridge-priority 4096
```

The range is 0 through 61440 in increments of 4096. The default is 32768.

9. Set the error disable parameters.
 - a. Enable the timer to bring the port out of error disable state.

```
device(config-mstp)# error-disable-timeout enable
```

- b. Specify the time in seconds it takes for an interface to time out..

```
device(config-mstp)# error-disable-timeout interval 60
```

The range is from 10 to 1000000 seconds with a default of 300 seconds.

10. Configure forward delay.
 - a. Specify the bridge forward delay.

```
device(config-mstp)# forward-delay 15
```

This command allows you to specify how long an interface remains in the listening and learning states before it begins forwarding. This command affects all MSTP instances. The range of values is from 4 to 30 seconds with a default of 15 seconds.

11. Configure hello time.

```
device(config-mstp)# hello-time 2
```

The hello time determines how often the switch interface broadcasts hello BPDUs to other devices. The range is from 1 through 10 seconds with a default of 2 seconds.

12. Configure the maximum age.

```
device(config-mstp)# max-age 20
```

You must set the **max-age** so that it is greater than the **hello-time**. The range is 6 through 40 seconds with a default of 20 seconds.

13. Specify the port-channel path cost.

```
device(config-mstp)# port-channel path-cost custom
```

This command allows you to control the path cost of a port channel according to bandwidth.

14. Specify the transmit hold count.

```
device(config-mstp)# transmit-holdcount 5
```

The transmit hold count is used to limit the maximum number of MSTP BPDUs that the bridge can transmit on a port before pausing for 1 second. The range is from 1 to 10 seconds with a default of 6 seconds.

15. Configure Cisco interoperability.

```
device(config-mstp)# cisco-interoperability enable
```

This command enables the ability to interoperate with certain legacy Cisco switches. The default is Cisco interoperability is disabled.

16. Return to privileged exec mode.

```
device(config-mstp)# end
```

17. Verify the configuration. The following is an example configuration.

```

device# show spanning-tree mst-config

Spanning-tree Mode: Multiple Spanning Tree Protocol

CIST Root Id: 8000.001b.ed9f.1700
CIST Bridge Id: 8000.768e.f80a.6800
CIST Reg Root Id: 8000.001b.ed9f.1700

CIST Root Path Cost: 0; CIST Root Port: Eth 1/2
CIST Root Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 19
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20;
Tx-HoldCount: 6
Number of topology change(s): 139; Last change occurred 00:03:36 ago on Eth 1/2

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec
Migrate Time: 3 sec

Name          : kerry
Revision Level : 1
Digest        : 0x9357EBB7A8D74DD5FEF4F2BAB50531AA

Instance      VLAN
-----
0:            1
1:            7,8
2:            21-23

```

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

18. Save the configuration.

```

device# copy running-config startup-config

```

MSTP configuration example

```

device# configure terminal
device(config)# protocol spanning-tree mstp
device(config-mstp)# region kerry
device(config-mstp)# revision 1
device(config-mstp)# description kerry switches
device(config-mstp)# max-hops 20
device(config-mstp)# instance 1 vlan 7,8
device(config-mstp)# instance 2 vlan 21-23
device(config-mstp)# instance 1 priority 4096
device(config-mstp)# bridge-priority 4096
device(config-mstp)# error-disable-timeout enable
device(config-mstp)# error-disable-timeout interval 60
device(config-mstp)# forward-delay 16
device(config-mstp)# hello-time 5
device(config-mstp)# max-age 16
device(config-mstp)# port-channel path-cost custom
device(config-mstp)# transmit-holdcount 5
device(config-mstp)# cisco-interopability enable
device(config-mstp)# end

```

```
device# show spanning-tree mst
device# copy running-config startup-config
```

Enabling and configuring MSTP on an interface

About This Task

The parameters can be configured individually on an interface by:

1. Entering the commands in Steps 1 through Step 3 for the target interface
2. Running the relevant parameter command
3. Verifying the result
4. Saving the configuration

For detailed descriptions of the parameters and features, see the sections STP parameters and STP features.

Procedure

1. Enter configuration mode.

```
device# configure terminal
```

2. Enable MSTP.

```
device(config)# protocol spanning-tree mstp
```

3. Enter interface configuration mode.

```
device(config-mstp)# interface ethernet 0/5
```

4. Enable the interface.

```
device(conf-if-eth-0/5)# no shutdown
```

5. Configure the restricted role feature for the port.

```
device(conf-if-eth-0/5)# spanning-tree restricted-role
```

This command keeps a port from becoming a root.

6. Restrict topology change notifications (TCN) BPDUs for an MSTP instance.

```
device(conf-if-eth-0/5)# spanning-tree instance 5 restricted-tcn
```

This prevents the port from propagating received TCNs and topology changes originating from a bridge, external to the core network, to other ports.

7. Enable auto detection of an MSTP edge port.

```
device(conf-if-eth-0/5)# spanning-tree autoedge
```

Enabling this feature allows the system to automatically identify the edge port. The port can become an edge port if no BPDU is received. By default, automatic edge detection is disabled.

8.

```
device(conf-if-eth-0/5)# spanning-tree edgeport
```

Enabling edge port allows the port to quickly transition to the forwarding state. By default, automatic edge detection is disabled.

9. Enable BPDU guard on the port

```
device(conf-if-eth-0/5)# spanning-tree edgeport bpdu-guard
```

BPDU guard removes a node that reflects BPDUs back in the network. It enforces the STP domain borders and keeps the active topology predictable by not allowing any network devices behind a BPDU guard-enabled port to participate in STP.

10. Set the path cost of a port.

```
device(conf-if-eth-0/5)# spanning-tree cost 200000
```

The path cost range is from 1 to 200000000. Leaving the default adjusts path cost relative to changes in the bandwidth. A lower path cost indicates greater likelihood of becoming root port.

11. Configure the link type.

```
device(conf-if-eth-0/5)# spanning-tree link-type point-to-point
```

The options are **point-to-point** or **shared**.

12. Enable port priority.

```
device(conf-if-eth-0/5)# spanning-tree priority 128
```

The range is from 0 to 240 in increments of 16 with a default of 32. A lower priority indicates greater likelihood of becoming root port.

13. Return to privileged exec mode.

```
device(conf-if-eth-0/5)# end
```

14. Verify the configuration.

```
device# show spanning-tree interface ethernet 0/5

Spanning-tree Mode: Multiple Spanning Tree Protocol

Root Id: 8000.001b.ed9f.1700
Bridge Id: 8000.01e0.5200.011d

Port Eth 0/5 enabled
  Ifindex: 411271175; Id: 8002; Role: Designated; State: Forwarding
  Designated External Path Cost: 0; Internal Path Cost: 2000000
  Configured Path Cost: 200000
  Designated Port Id: 8002; Port Priority: 128
  Designated Bridge: 8000.01e0.5200.011d
  Number of forward-transitions: 1
  Version: Multiple Spanning Tree Protocol - Received MSTP - Sent MSTP
  Edgeport: yes; AutoEdge: yes; AdminEdge: no; EdgeDelay: 3 sec
  Restricted-role is enabled
  Restricted-tcn is enabled
  Boundary: no
  Bpdu-guard: on
  Link-type: point-to-point
  Received BPDUs: 86; Sent BPDUs: 1654
```

15. Save the configuration.

```
device# copy running-config startup-config
```

Enable MSTP on an interface configuration example

```
device# configure terminal
device(config)# protocol spanning-tree mstp
device(config-mstp)# interface ethernet 0/5
device(conf-if-eth-0/5)# no shutdown
device(conf-if-eth-0/5)# spanning-tree restricted-role
device(conf-if-eth-0/5)# spanning-tree instance 5 restricted-tcn
device(conf-if-eth-0/5)# spanning-tree autoedge
device(conf-if-eth-0/5)# spanning-tree edgeport
device(conf-if-eth-0/5)# spanning-tree edgeport bpdu-guard
device(conf-if-eth-0/5)# spanning-tree cost 200000
device(conf-if-eth-0/5)# spanning-tree link-type point-to-point
device(conf-if-eth-0/5)# spanning-tree priority 128
device(conf-if-eth-0/5)# end
device# show spanning-tree interface ethernet 0/5
device# copy running-config startup-config
```

Enabling MSTP on a VLAN

Procedure

1. Enter configuration mode.

```
device# configure terminal
```

2. Enter the protocol command to enable MSTP configuration.

```
device(config)# protocol spanning-tree mstp
```

3. Map a VLAN to an MSTP instance.

```
device(config-mstp)# instance 5 vlan 300
```

4. Return to privileged EXEC mode.

```
device(config-mstp)# end
```

5. Verify the configuration.

```
device# show spanning-tree mst

Spanning-tree Mode: Multiple Spanning Tree Protocol

CIST Root Id: 8000.609c.9f5d.4800 (self)
CIST Bridge Id: 8000.609c.9f5d.4800
CIST Reg Root Id: 8000.609c.9f5d.4800 (self)

CIST Root Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20;
Tx-HoldCount: 6
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec
```

```

Migrate Time: 3 sec

Name          : NULL
Revision Level : 0
Digest        : 0xD5FF4C3F6C18E2F27AF3A8300297ABAA

Instance      VLAN
-----      ----
0:            1
5:            100

```

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

6. Save the configuration.

```
device# copy running-config startup-config
```

Enable spanning tree on a VLAN configuration example

```

device# configure terminal
device(config)# protocol spanning-tree mstp
device(config-mstp)# instance 5 vlan 300
device(config-mstp)# end
device# show spanning-tree mst
device# copy running-config startup-config

```

Configuring basic MSTP parameters

Procedure

1. Enter global configuration mode.

```
device# configure terminal
```

2. Enable MSTP.

```
device(config)# protocol spanning-tree mstp
```

3. Specify the region name.

```
device(config-mstp)# region connemara
```

4. Specify the revision number.

```
device(config-mstp)# revision 1
```

5. Map MSTP instances to VLANs.

- a. Map instance 1 to VLANs 2 and 3.

```
device(config-mstp)# instance 1 vlan 2,3
```

- b. Map instance 2 to VLANs 4, 5, and 6.

```
device(config-mstp)# instance 2 vlan 4-6
```


6. Set a priority for an instance.

```
device(conf-Mstp)# instance 1 priority 28672
```

The priority ranges from 0 through 61440 and the value must be in multiples of 4096.

7. Specify the maximum hops for a BPDU.

```
device(conf-Mstp)# max-hops 25
```

This prevents the messages from looping indefinitely on an interface

8. Return to privileged EXEC mode.

```
device(conf-Mstp)# end
```

9. Verify the configuration.

```
device# show spanning-tree mst

Spanning-tree Mode: Multiple Spanning Tree Protocol

CIST Root Id: 8000.609c.9f5d.4800 (self)
CIST Bridge Id: 8000.609c.9f5d.4800
CIST Reg Root Id: 8000.609c.9f5d.4800 (self)

CIST Root Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 20
Configured Forward Delay: 15; Hello Time: 2; Max Age: 20; Max-hops: 25;
Tx-HoldCount: 6
Number of topology change(s): 0

Bpdu-guard errdisable timeout: disabled
Bpdu-guard errdisable timeout interval: 300 sec
Migrate Time: 3 sec

Name          : connemara
Revision Level : 1
Digest        : 0xD5FF4C3F6C18E2F27AF3A8300297ABAA

Instance      VLAN
-----      ----
0:            1,7,8,9
1:            2,3
2:            4-6
```



Note

Observe that the settings comply with the formula set out in the STP parameters section, as:

$$(2 \times (\text{forward delay} - 1)) \geq \text{maximum age} \geq (2 \times (\text{hello time} + 1))$$

or in this case: $28 \geq 20 \geq 6$.

```
device# show running-config | begin spanning-tree
protocol spanning-tree mstp
instance 1 vlan 2,3
instance 1 priority 28672
instance 2 vlan 4-6
region connemars
revision 1
max-hops 25
```

```
!
...
```

10. Save the configuration

```
device# copy running-config startup-config
```

Basic MSTP configuration example

```
device# configure terminal
device(config)# protocol spanning-tree mstp
device(config-mstp)# region connemara
device(config-mstp)# revision 1
device(config-mstp)# instance 1 vlan 2,3
device(config-mstp)# instance 2 vlan 4-6
device(config-mstp)# instance 1 priority 28582
device(config-mstp)# max-hops 25
device(config-mstp)# end
device# show spanning-tree mst
device# copy running-config startup-config
```

Clearing spanning tree counters

Procedure

1. Clear spanning tree counters on all interfaces.

```
device# clear spanning-tree counter
```

2. Clear spanning tree counters on a specified Ethernet interface.

```
device# clear spanning-tree counter interface ethernet 0/3
```

3. Clear spanning tree counters on a specified port channel interface.

```
device# clear spanning-tree counter interface port-channel 12
```

Port channel interface numbers range from 1 through 64.

Clearing spanning tree-detected protocols

About This Task

These commands force a spanning tree renegotiation with neighboring devices on either all interfaces or on a specified interface.

Procedure

1. Restart the spanning tree migration process on all interfaces.

```
device# clear spanning-tree detected-protocols
```

2. Restart the spanning tree migration process on a specific Ethernet interface.

```
device# clear spanning-tree detected-protocols interface ethernet 0/3
```

- Restart the spanning tree migration process on a specific port channel interface.

```
device# clear spanning-tree detected-protocols port-channel 12
```

Port channel interface numbers range from 1 through 64.

Shutting down MSTP

Procedure

- Enter global configuration mode.

```
device# configure terminal
```

- Shut down MSTP.

- Shut down MSTP globally and return to privileged EXEC mode.

```
device(config)# protocol spanning-tree mstp
device(config-mstp)# shutdown
device(config-mstp)# end
```

- Shut down MSTP on a specific interface and return to privileged EXEC mode.

```
device(config)# interface ethernet 0/2
device(config-if-eth-0/2)# spanning-tree shutdown
device(config-if-eth-0/2)# end
```

- Shut down MSTP on a specific VLAN and return to privileged EXEC mode.

```
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-vlan-10)# end
```

- Verify the configuration.

```
device# show spanning-tree
device#
```

- Save the running configuration to the startup configuration.

```
device# copy running-config startup-config
```

Shut down MSTP configuration example

```
device# configure terminal
device(config)# vlan 10
device(config-vlan-10)# spanning-tree shutdown
device(config-stp)# end
device# show spanning-tree
device# copy running-config startup-config
```



Note

Shutting down MSTP on a VLAN is used in this example.



Topology Groups

[Topology groups on page 188](#)

[Master VLAN, member VLANs, and bridge-domains on page 188](#)

[Control ports and free ports on page 189](#)

[Configuration considerations on page 189](#)

[Configuring a topology group on page 190](#)

[Displaying topology group information on page 192](#)

Topology groups

A topology group is a named set of VLANs and bridge-domains that share a Layer 2 control protocol. Topology groups simplify configuration and enhance scalability of Layer 2 protocols by allowing you to run a single instance of a Layer 2 protocol on multiple VLANs and bridge-domains. One instance of the Layer 2 protocol controls all the VLANs and bridge-domains.

You can use topology groups with the following Layer 2 protocols:

- Per VLAN Spanning Tree (PVST+)
- Rapid per VLAN Spanning tree (R-PVST+)

Master VLAN, member VLANs, and bridge-domains

Each topology group contains a master VLAN and can contain one or more member VLANs and bridge-domains. A definition for each of these VLAN types follows:

- **Master VLAN**—The master VLAN contains the configuration information for the Layer 2 protocol. For example, if you plan to use the topology group for Rapid per VLAN Spanning tree (R-PVST), the topology group's master VLAN contains the R-PVST configuration information.
- **Member VLANs**—The member VLANs are additional VLANs that share ports with the master VLAN. The Layer 2 protocol settings for the ports in the master VLAN apply to the same ports in the member VLANs. A change to the master VLAN's Layer 2 protocol configuration or Layer 2 topology affects all the member VLANs. Member VLANs do not independently run a Layer 2 protocol.
- **Member bridge domains**—The member bridge domains are similar to VLANs that share ports with the master VLAN. The Layer 2 protocol settings for the ports in the master VLAN apply to the same ports in the bridge domains. A change to the master VLAN's Layer 2 protocol configuration or Layer 2 topology affects all the bridge domains. Bridge domains do not independently run a Layer 2 protocol. In a bridge domain, a single port can have multiple logical interfaces. In this scenario, all the logical interfaces on that port (and bridge domain) will follow the state of master VLAN port.

When a Layer 2 topology change occurs, resulting in a change of port state in the master VLAN, the same port state is applied to all the member VLANs and bridge-domains belonging to the topology group on that port. For example, if you configure a topology group whose master VLAN contains ports 1/1 and 1/2, a Layer 2 state change on port 1/1 applies to port 1/1 in all the member VLANs and bridge-domains that contain that port. However, the state change does not affect port 1/1 in VLANs that are not members of the topology group.

Control ports and free ports

A port in a topology group can be a control port or a free port:

- A **control port** is a port in the master VLAN and, therefore, is controlled by the Layer 2 protocol configured in the master VLAN. The same port in all the member VLANs and bridge-domains is controlled by the master VLAN's Layer 2 protocol. Each member VLAN and bridge-domain must contain all of the control ports. All other ports in the member VLAN and bridge-domain are "free ports."
- **Free ports** are not controlled by the master VLAN's Layer 2 protocol. The master VLAN can contain free ports. (In this case, the Layer 2 protocol is disabled on those ports.) In addition, any ports in the member VLANs and bridge-domains that are not also in the master VLAN are free ports.



Note

Because free ports are not controlled by the master port's Layer 2 protocol, they are always in the forwarding state.

Configuration considerations

The configuration considerations are as follows:

- You can configure up to 128 topology groups. A VLAN or bridge-domain cannot be controlled by more than one topology group. You can configure up to 4K VLANs or bridge-domains as members of a topology group.
- The topology group must contain a master VLAN. The group can also contain individual member VLANs and/or member bridge-domains. You must configure the member VLANs or member bridge-domains before adding them to the topology group. Bridge-domains cannot be configured as a master VLAN.
- You cannot delete a master VLAN from the topology group when the member VLANs or bridge-domains are in the topology group.
- The control port membership must match the master VLAN when adding a member VLAN or member bridge-domain.
- If a VLAN enabled with the PVST+ or R-PVST+ protocol is added as a member VLAN of a topology group, the protocol is disabled. The member VLAN is added to the topology group. If the VLAN is removed from the topology group, the protocol is disabled, and you must enable the protocol if required.
- Enabling STP on an interface is only allowed if both master VLAN and member VLAN or bridge-domains are configured on the interface across all topology groups.
- You cannot remove the master VLAN or member VLAN or bridge-domains from an STP-enabled interface.
- Topology group configuration is allowed only with PVST+ and R-PVST+ spanning tree configurations.

Configuring a topology group

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **topology-group** command to create a topology group at the global configuration level.

```
device(config)# topology-group 1
device(conf-topo-group-1)#
```



Note

The **no topology-group** command deletes an existing topology group.

Configuring a master VLAN

Before You Begin

Before configuring a master VLAN, you should have configured a topology group.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **topology-group** command to create a topology group at the global configuration level.

```
device(config)# topology-group 1
device(conf-topo-group-1)#
```

3. Enter the **master-vlan** command to configure a master VLAN in the topology group.

```
device(conf-topo-group-1)# master-vlan 100
```



Note

The **no master-vlan** command removes an existing master VLAN from the topology group.

Adding member VLANs

Before You Begin

Before adding a member VLAN, you should have created a topology group and configured the master VLAN for that group. The VLAN should not be part of any other topology group. All control ports of master VLAN must also be configured for the member VLAN.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **topology-group** command to create a topology group at the global configuration level.

```
device(config)# topology-group 1
device(conf-topo-group-1)#
```

3. Enter the **master-vlan** command to configure a master VLAN in the topology group.

```
device(conf-topo-group-1)# master-vlan 100
```

4. Enter the **member-vlan** command to add member VLANs to the topology group.

```
device(conf-topo-group-1)# member-vlan add 200-201
```



Note

The **member-vlan remove** command removes an existing member VLAN from the topology group.

```
device(conf-topo-group-1)# member-vlan remove 200
```

Adding member bridge-domains

Before You Begin

Before adding a bridge domain, you should have created a topology group and configured the master VLAN for that group. The bridge-domain should not be part of any other topology group. All control ports of master VLAN must also be configured for the member bridge-domain.

Procedure

1. Enter the **configure terminal** command to access global configuration mode.

```
device# configure terminal
device(config)#
```

2. Enter the **topology-group** command to create a topology group at the global configuration level.

```
device(config)# topology-group 1
device(conf-topo-group-1)#
```

3. Enter the **master-vlan** command to configure a master VLAN in the topology group.

```
device(conf-topo-group-1)# master-vlan 100
```

4. Enter the **member-bridge-domain** command to add member bridge-domains to the topology group.

```
device(conf-topo-group-1)# member-bridge-domain add 300
```



Note

The **member-bridge-domain remove** command removes an existing member bridge-domain from the topology group.

```
device(conf-topo-group-1)# member-bridge-domain remove 1
```

Example

The example adds 300 as member bridge-domain to the topology group.

```
device# configure terminal
device(config)# topology-group 1
device(conf-topo-group-1)# master-vlan 100
device(conf-topo-group-1)# member-bridge-domain add 300
```

Replacing a master VLAN

To avoid temporary loops when the master VLAN is replaced by another VLAN, the following recommendation is made:

- Control ports for both the old and the new master VLAN must match.
- The new master VLAN and the old master VLAN must have same ports in the blocking state to avoid the possibility of temporary loops.

If the recommendation is not followed, and a new master VLAN is configured with a different convergence, the configuration is still accepted.



Note

The master VLAN replacement is accepted if both the old and the new master VLANs are spanning-tree disabled.

Displaying topology group information

Before You Begin

Before displaying the topology group information, you should have configured a topology group and defined the master VLAN.

Procedure

Enter the **show topology-group** command to display the group information.

```
device# show topology-group 1
Topology Group 1
=====
Master VLAN : 100
L2 Protocol: R-PVST
Member VLANs : 200 300
Member Bridge-domains: 10
```



```
Control Ports : eth 2/1, eth 2/2, po10
Free Ports : VLAN: 200 -eth 2/3, po11
Bridge-domain: 10 -eth 2/3.20, po11.10
```

The example displays information about topology group 1.

The **show running-config** command displays topology group configurations.

```
device# show running-config
topology-group 1
  master-vlan 100
  member-vlan add 200 300
  member-bridge-domain add 10
```



Loop Detection

[LD protocol overview](#) on page 194

[LD use cases](#) on page 200

[Configuring LD protocol](#) on page 202

[Loop detection for VLAN](#) on page 205

LD protocol overview

Layer 2 networks rely on learning and flooding to build their forwarding databases. Because of the flooding nature of these networks, any loops can be disastrous as they cause broadcast storms.



Important

The LD feature should be used only as a tool to detect loops in the network. It should not be used to replace other Layer 2 protocols such as STP.

This feature provides support for the following:

- Strict and loose modes
- Multi-Chassis Trunk (MCT)
- Breakout ports
- EPVN VLAN tunnels

LD protocol data units (PDUs) are initiated and received on the native device. Loop detection and action on the port state is also done on the same native device. Intermediate devices in the network must be capable of flooding unknown Layer 2 unicast PDUs on the VLAN through which they are received.

Strict mode

In what is referred to as *strict mode*, LD is configured on an interface. If the LD PDU is sent on an interface and received on the same interface, that port is shut down by LD. Strict mode overcomes specific hardware issues that cause packets to be echoed back to the input port. The following figure illustrates strict mode.

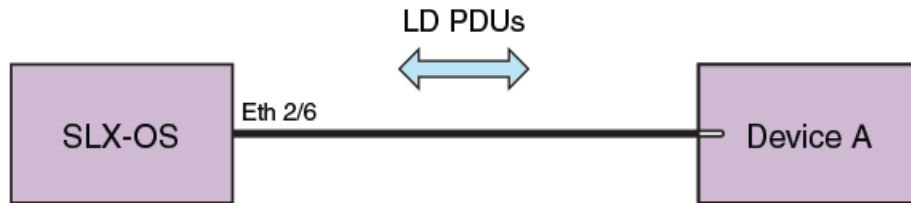


Figure 17: Strict mode

If the user provides a VLAN, then the PDUs are tagged accordingly. Otherwise PDUs are sent untagged. With a LAG, PDUs are sent out on the port-channel interface. If Device A has a loop (for example, a LAG is not configured), then the PDU is flooded back to SLX-OS, which detects the loop. In case of a loop, the port-channel interface is shut down. The following figure illustrates LD on a LAG.

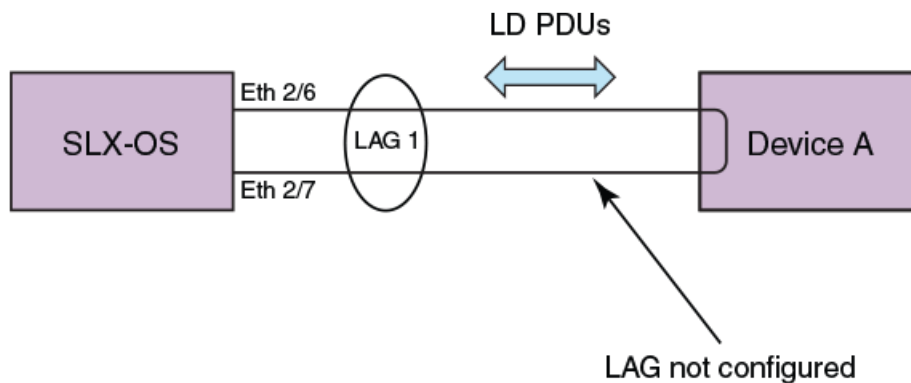


Figure 18: LD on a LAG

LD supports 256 instances of strict mode.

Loose mode

In what is referred to as *loose mode*, LD is configured on a VLAN. If a VLAN in the device receives an LD PDU that originated from the same device on that VLAN, this is considered to be a loop and the receiving port is shut down. In loose mode, LD works at the VLAN level and takes action at the logical interface (LIF) level. The following figure illustrates loose mode, with LD on a VLAN.

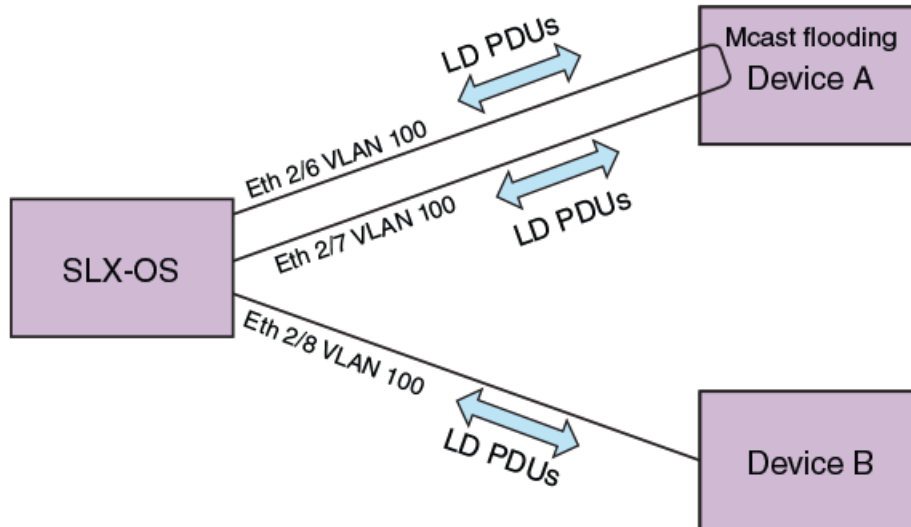


Figure 19: Loose mode: LD on a VLAN

SLX-OS generates the LD PDUs on the VLAN. If Device A has a loop, PDUs are flooded back to SLX-OS, which detects the loop. SLX-OS then shuts down the receiving LIF of the port on the VLAN.

LD supports 256 instances of loose mode, which means that it can be enabled on 256 VLANs.

LD PDU format

The following figure illustrates the format of the LD PDU in bytes.

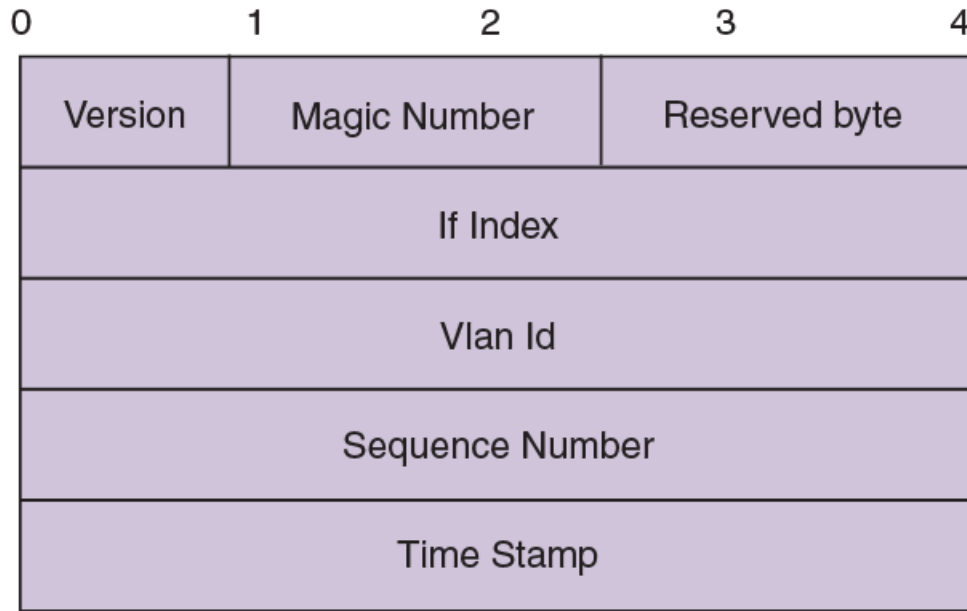


Figure 20: LD PDU format in bytes

Parameter	Definition
Version	LD protocol version (1 by default)
Magic Number	0x13EF; used to differentiate between LD multicast PDUs and other multicast PDUs
Reserved byte	For future use
If Index	Index of the source port; populated only in strict mode
Vlan Id	VLAN ID
Sequence Number	Reserved for future enhancements
Time Stamp	Reserved for future enhancements

LD PDU transmission

Each LD-enabled interface or VLAN on a device continually transmits Layer 2 LD PDUs at a 1-second default hello-timer interval, with the destination MAC address as the multicast address. The multicast MAC address is derived from the system MAC address of the device with the multicast bit (8) and the local bit (7) set.

For example, if the MAC address is 00E0.5200.1800, then the multicast MAC address is 03E0.5200.1800. In the case of a LAG port-channel, LD PDUs are sent out one of the ports of the LAG as chosen by hardware.

LD PDU reception

When the LD PDU is received and is generated by the same device, the PDU is processed. If the PDU is generated by another device, then the PDU is flooded.

If a port is already blocked by any other Layer 2 protocol such as STP, then the LD PDUs are neither sent for LD processing nor flooded on that port.

LD parameters

This section discusses the various global protocol-level, interface level, and VLAN-level parameters that are used to control and process LD PDUs.

Protocol level

hello-interval

hello-interval is the rate at which the LD PDUs are transmitted by an LD-enabled interface or VLAN, which is 1000 milliseconds by default. Lowering the hello-interval below the default increases the PDU transmission rate, providing faster loop detection and also removing transient loops that last less than one second. On the other hand, increasing the interval above the default (for example, to 100 milliseconds) can increase the steady-state CPU load.

shutdown-time

shutdown-time is the duration after which an interface that is shut down by LD is automatically reenabled. The range is from 0 through 1440 minutes. The default is 0 minutes, which means that the interface is not automatically reenabled.



Important

Changing this value can cause repeated interface flapping when a loop is persistent in the network.

raslog-duration

raslog-duration is the interval between RASLog messages when a port is shut down by LD to prevent flooding of these messages. The range is from 10 through 1440 minutes. The default is 10.

Interface level

In strict mode, the parameters in this section are configurable at the interface level, and the configuration is specific to an interface. The following figure illustrates strict mode configuration.

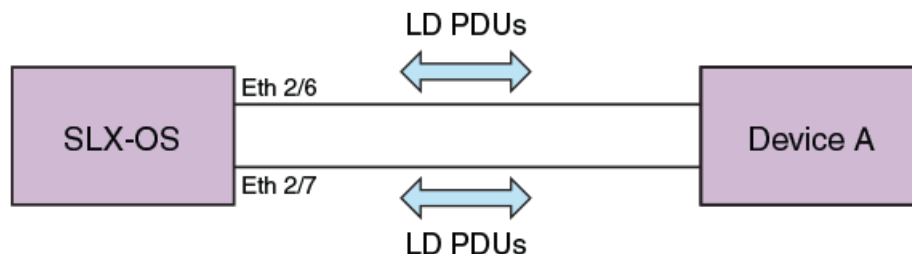


Figure 21: Strict mode configuration

shutdown-disable

By default, the device shuts down the interface if a loop is detected. Configuring **shutdown-disable** means that the interface shutdown is disabled and LD never brings down such interface. If a loop is

already detected by LD and the port is in shutdown state, then configuring **shutdown-disable** is not effective until the port is back up.

vlan-association

Although user can enable LD on an interface without specifying a VLAN, the **vlan-association** keyword is used to specify a VLAN associated with the interface.

VLAN level

In loose mode, the user can configure LD under a VLAN. In this case, LD PDUs are flooded on the VLAN. The following figure illustrates loose mode configuration.

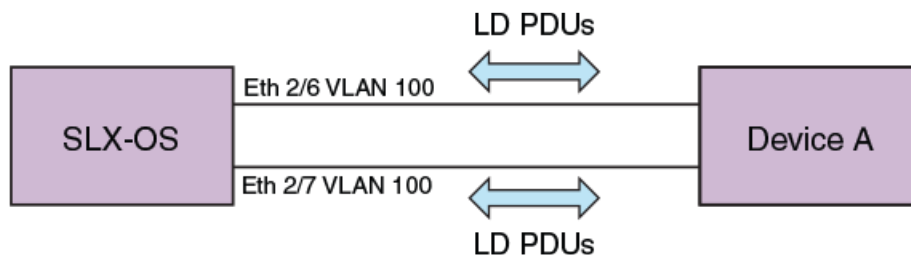


Figure 22: Loose mode configuration

LD PDU processing

As long as LD PDUs are not received, there is no loop. If an LD PDU is received, then there is a loop that is present in the network.

If the if-index field in the received LD PDU is valid, then it is considered to be operating in strict mode. If the port on which the LD PDU was received is same as one encoded in the PDU (with a match for VLAN ID if a VLAN is associated), the port is shut down. For an MCT, if a strict mode LD PDU is received on an ICL interface, and the PDU is originated by another interface, then the ICL interface is not shut down. Instead, the sender interface is shut down. In addition, for strict mode the required interfaces should be configured with LD, or else the PDUs will not get processed

If the if-index field in the received LD PDU is invalid, then it is considered to be operating in loose mode. Based on VLAN ID information present in the received LD PDU, the receiving LIF is shut down. If the receiving interface is an MCT ICL interface, the LD PDU is dropped.

In the case of an LD-enabled LAG (port-channel) interface, if the sent LD PDU is received on the port-channel, then the port-channel interface is shut down.

If the **shutdown-disable** option is configured for the particular interface, then the port drops the received PDU without processing it.

The re-enablement of the LD shut down port depends on the **shutdown-time** configuration. For manual recovery, either flap the interface, by means of the **shutdown** and **no shutdown** commands, or clear the loop by means of the **clear loop-detection** command.

Support for EPVN VLAN tunnels

LD loose mode is used to support a shutdown at the attachment circuit (AC) logical interface (LIF) level instead of at the physical port level. See [Loop detection for EVPN VLAN](#) and configuration examples.

Configuration considerations

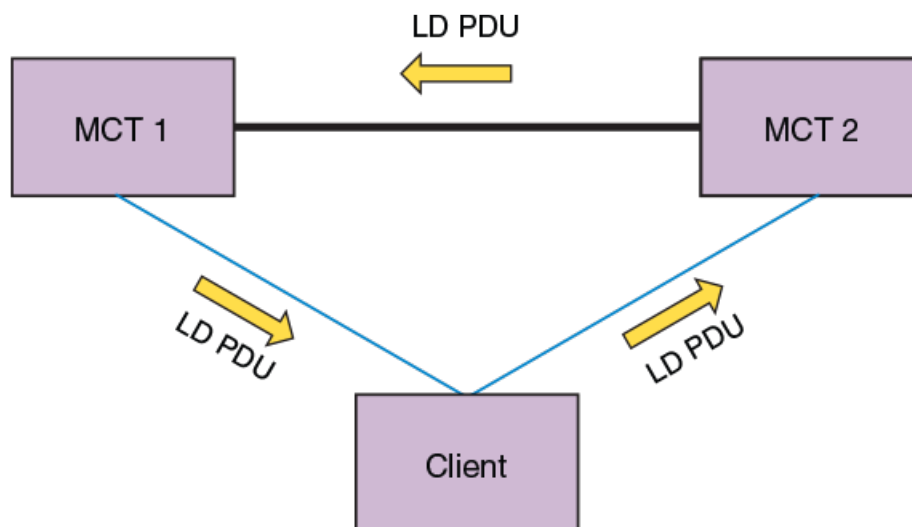
On an external switch that is unaware of LD or where LD is not configured, there may be some ACL rules applied to interfaces to permit traffic from known MAC addresses, and at the last of these rules there is an ACL deny-any rule to block all unknown MAC addresses. If this interface is part of a loop, LD enabled on SLX-OS will not be able to detect and break the loop.

LD use cases

In an MCT configuration, LD runs independently on both nodes. With loose mode the user must enable loop detection for the same VLANs on both nodes in the MCT cluster. MCT strict mode and loose mode use cases are detailed below.

MCT strict mode

Strict mode LD is enabled on the MCT 1 cluster client edge port (CCEP) interface that connects to the Client.



1. MCT 1 generates LD PDUs.
2. If the Client has the LAG interface configured to support LD, the Client drops the PDUs and there is no loop.
3. If there is a misconfiguration, the Client floods the PDUs, reaching MCT 2.
4. MCT 1 then identifies the interface information encoded in the PDUs, shutting down the interface on which the packets were generated.

Figure 23: MCT strict mode

MCT loose mode

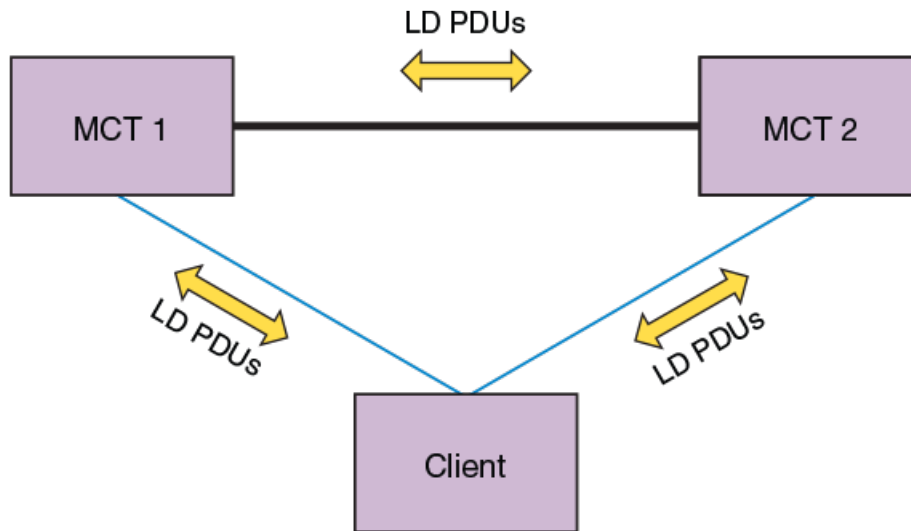


Figure 24: MCT loose mode: Use case 1

Use case 1: LD enabled on VLAN x on MCT 1

1. MCT 1 sends LD PDUs on VLAN x on all the interfaces that are part of the CCEP, client edge port (CEP), and ICL interface.
2. If the Client has LD configured on the LAG interface, then it drops the PDUs and no loop exists. If there is a misconfiguration, the Client floods the PDUs and they reach MCT 2.
3. MCT 2 floods the PDUs back to MCT 1, where the loop is detected. With loose mode no information about the interface that transmitted the PDU is encoded in the PDU, so normally the receiving interface is shut down. Because in this case the PDU is received on the ICL interface, that interface is not shut down.
4. MCT 1 receives the loop detection PDUs on the CCEP interface as well, as the packets were flooded in the VLAN in the following sequence: MCT 1 > MCT 2 > Client > MCT 1. In this case the receiving CCEP is shut down to break the loop. For MCT 2 to forward the PDUs in this case it must be the designated forwarder (DF) for that VLAN.

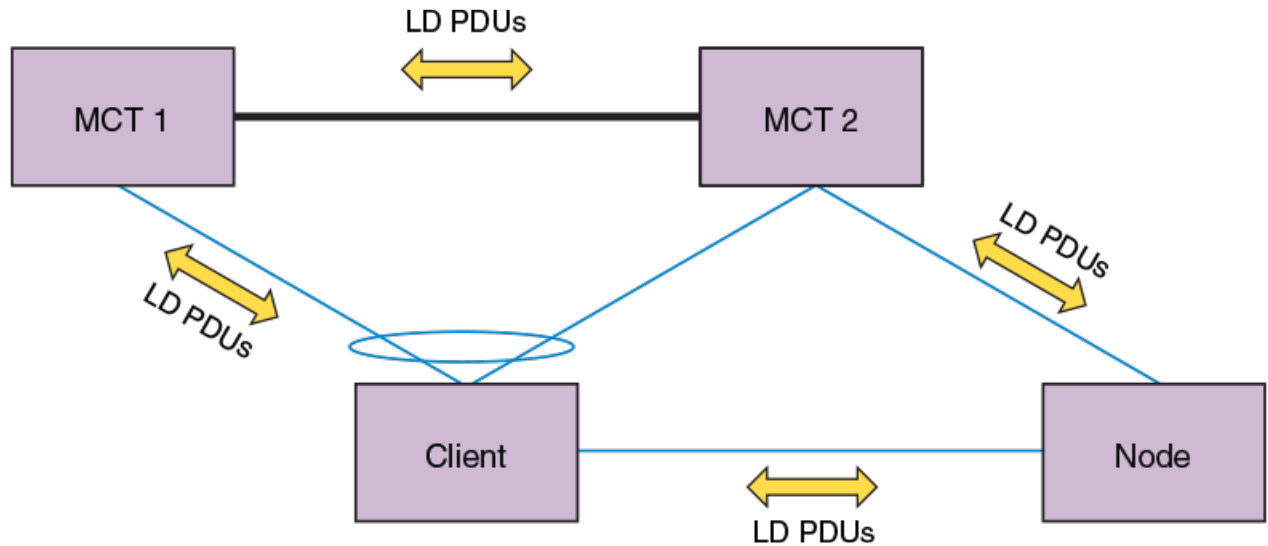


Figure 25: MCT loose mode: Use case 2

Use case 2: LD enabled on VLAN x on MCT 1 and MCT 2

1. Both MCT 1 and MCT 2 will flood the PDUs in VLAN x on all the interfaces that are part of the CCEP, CEP, and ICL interface.
2. Assuming PDUs from MCT 1 take the path MCT 1 > MCT 2 > Node > Client > MCT 1, then the receiving CCEP interface is shut down. For MCT 2 to forward the PDUs in this case, it must be the DF for that VLAN.
3. Assuming PDUs from MCT 2 take the path MCT 2 > MCT 1 > Client > Node > MCT 2, then the receiving CEP interface is shut down
4. If PDUs from MCT 2 take the path MCT 2 > Node > Client > MCT 2, then the receiving CCEP interface is shut down.
5. Multiple interfaces can be shut down in this case, depending on the sequence in which loops are detected.
6. In addition, to avoid CCEP interfaces from being shut down over a CEP interface, the user can configure a CCEP port not to be shut down.

Configuring LD protocol

Procedure

1. Enter global configuration mode.


```
device# configure terminal
```
2. Enter the **protocol loop-detection** command to enable loop detection, enter Protocol Loop Detect configuration mode, and configure a variety of global options.


```
device(config)# protocol loop-detection
```
3. (Optional) Enter the **hello-interval** command to change the hello interval from the default.


```
device(config-loop-detect)# hello-interval 2000
```
4. (Optional) Enter the **shutdown-time** command to change from the default the interval after which an interface that is shut down by loop detection (LD) protocol is automatically reenabled.


```
device(config-loop-detect)# shutdown-time 20
```

5. (Optional) Enter the **raslog-duration** command to change from default the interval between RASLog messages that are sent when a port is disabled by the loop detection (LD) protocol.

```
device(config-loop-detect)# raslog-duration 20
```

6. Enable LD at the interface level.
 - a. In global configuration mode, specify an interface (either an Ethernet interface or a port-channel interface).

```
device(config)# interface ethernet 2/6
```

- b. In interface subtype configuration mode, enter the **loop-detection** command.

```
device(conf-if-eth-2/6)# loop-detection
```

7. Enable LD at the VLAN level.
 - a. In global configuration mode, create a VLAN.

```
device(config)# vlan 5
```

- b. In VLAN configuration mode, enter the **loop-detection** command.

```
device(config-vlan-5)# loop-detection
```

8. Associate the VLAN with an interface.
 - a. In global configuration mode, specify an interface (either an Ethernet interface or a port-channel interface).

```
device(config)# interface ethernet 2/6
```

- b. In interface subtype configuration mode, enter the **loop-detection vlan** command and specify a VLAN. (The VLAN must already be created.)

```
device(conf-if-eth-2/6)# loop-detection vlan 5
```

9. (Optional) Disable the shutting down of an interface (Ethernet or port-channel) as a result of the loop detection (LD) protocol.
 - a. In global configuration mode, specify an interface (either an Ethernet interface or a port-channel interface).

```
device(config)# interface ethernet 2/6
```

- b. In interface subtype configuration mode, enter the **loop-detection shutdown-disable** command.

```
device(conf-if-eth-2/6)# loop-detection shutdown-disable
```

10. (Optional) Disable the shutting down of an interface (Ethernet or port-channel) as a result of the loop detection (LD) protocol.
 - a. In global configuration mode, specify an interface (either an Ethernet interface or a port-channel interface).

```
device(config)# interface ethernet 2/6
```

- b. In interface subtype configuration mode, enter the **loop-detection shutdown-disable** command.

```
device(conf-if-eth-2/6)# loop-detection shutdown-disable
```

11. Confirm the LD configuration, using the **show loop-detection** command with a variety of options.
 - a. To display LD information at the system level, enter the **show loop-detection** command as in the following example.

```
device# show loop-detection
Strict Mode:
-----

Number of loop-detection instances enabled: 1

Interface: eth 2/6
  Enabled on VLANs: 100
  Shutdown Disable: No
  Interface status: UP
  Auto enable in: Never

Packet Statistics:
vlan          sent          rcvd          disable-count
100           100           0             0

Loose Mode:
-----

Number of LD instances: 2
Disabled Ports:          2/7

Packet Statistics:
vlan          sent          rcvd          disable-count
100           100           0             0
```

- b. To display ports disabled by LD, enter the **show loop-detection disabled-ports** command as in the following example.

```
device# show loop-detection disabled-ports
Ports disabled by loop detection
-----

port          age (min)          disable cause
2/6           5                  Disabled by Self
```

- c. To display global LD configuration values, enter the **show loop-detection globals** command.

```
device# show loop-detection globals
Loop Detection:          Disabled
Shutdown-time (minutes): 0
Hello-time (msec):      1000
Raslog-duration (minutes): 10
```

12. Use the **clear loop-detection** command in privileged EXEC mode with a variety of options to reenble ports that were disabled by LD and clear the LD statistics.
 - a. To enable LD-disabled ports and clear LD statistics on all interfaces, enter the **clear loop-detection** command.

```
device# clear loop-detection
```

- b. To enable LD-disabled ports and clear LD statistics on an Ethernet interface, enter the **clear loop-detection interface ethernet** command.

```
device# clear loop-detection interface ethernet 2/6
```

- c. To enable LD-disabled ports and clear LD statistics on a port-channel interface, enter the **clear loop-detection interface port-channel** command.

```
device# clear loop-detection interface port-channel 20
```

- d. To enable LD-disabled ports and clear LD statistics on a VLAN, enter the **clear loop-detection interface vlan** command.

```
device# clear loop-detection interface vlan 10
```

Loop detection for VLAN



Note

Loop detection is not supported for bridge domains (BDs).

When a loop is detected on a VLAN and port, only the LIF of the VLAN on the port is shut down, but the physical port still remains up and other VLANs on the port are not affected.

The existing LD loose mode configuration commands support loop detection for VLAN tunnels. In LD loose mode, if the VLAN is mapped to VLAN tunnels and LD is enabled, VLAN tunnels loop detection is supported. Up to 256 LD loose mode instances can be configured.

If a loop is detected from a VLAN tunnel, the following actions can take place:

- A RASLog is sent and the tunnel VNI LIFs on which the loop is detected is shut down.
- A RASLog is sent but the tunnel LIF is not shut down.

Configuring loop detection for VLAN

The following example enables loop detection on a VLAN and enters Protocol Loop Detection configuration mode.

```
device# configure terminal
device(config)# vlan 5
device(config-vlan-5)# loop-detection
device(config-loop-detect)#
```

The following example enables ports associated with VLAN 8 and clears LD statistics for that VLAN.

```
device# clear loop-detection vlan 8
```

The following example displays LD configuration values, including logical interfaces (LIFs), for a VLAN tunnel.

```
device# show loop-detection vlan 20
Number of LD instances: 1
LIF (Logical Interface) Disabled on Ports: eth2/2

Packet Statistics:
vlan      sent      rcvd
20        119       1
```

The following example displays LD configuration values for a VLAN tunnel if LD shutdown is disabled.

```
device# show loop-detection vlan 20
Number of LD instances: 1
LIF (Logical Interface) ShutDown is disabled for VLAN 20
```

```
Packet Statistics:
vlan          sent          rcvd
20            10             10
```

The following example disables the shutdown of a VLAN VLAN tunnel.

```
device# configure terminal
device(config)# vlan 20
device(config-vlan-20)# loop-detection-shutdown-disable
```

You can control the auto-enable behavior of an LD-disabled logical interface (LIF), by using the **shutdown-time** command in Protocol Loopback Detection configuration mode. The following example specifies a shutdown time of 1 minute.

```
device# configure terminal
device(config)# loop-detection
device(config-loop-detect)# shutdown-time 1
2017/10/20-16:04:48, [ELD-1005], 3749, M2 | Active | DCE, INFO, SLX, Loop is detected on Ethernet 2/2
VLAN 20, the LIF (logical interface) is shutdown.
2017/10/20-16:05:46, [ELD-1007], 3750, M2 | Active | DCE, INFO, SLX, Loop detection disabled LIF
(Logical interface) on Ethernet 2/2 VLAN 20 is auto-enabled.
```

By default the shutdown time is 0, which means that an LD-disabled LIF is never auto-enabled. If the shutdown time is configured with a nonzero value, the LD-disabled LIF is auto-enabled following the specified shutdown time.

To enable LD-disabled ports and clear LD statistics on all interfaces, use the **clear loop-detection** command as in the following example.

```
device# clear loop-detection
```