

Network Configuration Example

Configuring Multichassis Link Aggregation on a QFX Series Switch



Published: 2014-01-10

Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, California 94089
USA
408-745-2000
www.juniper.net

Juniper Networks, Junos, Steel-Belted Radius, NetScreen, and ScreenOS are registered trademarks of Juniper Networks, Inc. in the United States and other countries. The Juniper Networks Logo, the Junos logo, and JunosE are trademarks of Juniper Networks, Inc. All other trademarks, service marks, registered trademarks, or registered service marks are the property of their respective owners.

Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

Network Configuration Example Configuring Multichassis Link Aggregation on a QFX Series Switch
NCE0064
Copyright © 2014, Juniper Networks, Inc.
All rights reserved.

The information in this document is current as of the date on the title page.

YEAR 2000 NOTICE

Juniper Networks hardware and software products are Year 2000 compliant. Junos OS has no known time-related limitations through the year 2038. However, the NTP application is known to have some difficulty in the year 2036.

END USER LICENSE AGREEMENT

The Juniper Networks product that is the subject of this technical documentation consists of (or is intended for use with) Juniper Networks software. Use of such software is subject to the terms and conditions of the End User License Agreement ("EULA") posted at <http://www.juniper.net/support/eula.html>. By downloading, installing or using such software, you agree to the terms and conditions of that EULA.

Table of Contents

Introduction	1
Advantages of Using Multichassis Link Aggregation Groups	1
Understanding Multichassis Link Aggregation	1
Active-Active Mode	3
ICCP and ICL-PL	3
Failure Handling	3
Multichassis Link Protection	4
MC-LAG Packet Forwarding	4
Layer 3 Routing	4
Spanning Tree Protocol (STP) Guidelines	5
MC-LAG Upgrade Guidelines	5
Layer 2 Unicast Features Supported	6
Layer 2 Multicast Features Supported	6
IGMP Snooping on an Active-Active MC-LAG	6
Layer 3 Unicast Features Supported	7
VRRP Active-Standby Support	7
Routed VLAN Interface (RVI) MAC Address Synchronization	8
Address Resolution Protocol (ARP)	8
DHCP Relay with Option 82	8
Private VLAN (PVLAN)	9
Layer 3 Multicast	9
PIM Operation With Normal Mode DR Election	9
PIM Operation with Dual-DR Mode	10
Configuration Guidelines and Caveats	10
Example: Configuring Multichassis Link Aggregation	11
Configuring Multichassis Link Aggregation	26

Introduction

This document describes multichassis link aggregation groups (MC-LAGs) between QFX Series switches, identifies the advantages of enabling MC-LAGs, and provides a step-by-step procedure for configuring MC-LAGs.

Advantages of Using Multichassis Link Aggregation Groups

A Multichassis Link Aggregation Group (MC-LAG) reduces operational expenses by providing active-active links with a Link Aggregation Group (LAG), eliminates the need for the Spanning Tree Protocol (STP), and provides faster layer 2 convergence upon link and device failures.

An MC-LAG adds node-level redundancy to the normal link-level redundancy that a LAG provides. An MC-LAG improves network resiliency, which reduces network down time as well as expenses.

In data centers, it is desirable for servers to have redundant connections to the network. You probably want active-active connections along with links from any server to at least two separate switches.

An MC-LAG allows you to bond two or more physical links into a logical link between two switches or between a server and a switch, which improves network efficiency. An MC-LAG enables you to load balance traffic on multiple physical links. If a link fails, the traffic can be forwarded through the other available link and the logical aggregated link remains in the UP state.

Related Documentation

- [Understanding Multichassis Link Aggregation on page 1](#)
- [Example: Configuring Multichassis Link Aggregation on page 11](#)
- [Configuring Multichassis Link Aggregation on page 26](#)

Understanding Multichassis Link Aggregation

Layer 2 networks are increasing in scale mainly because of technologies such as virtualization. Protocol and control mechanisms that limit the disastrous effects of a topology loop in the network are necessary. Spanning Tree Protocol (STP) is the primary solution to this problem because it provides a loop-free Layer 2 environment. STP has gone through a number of enhancements and extensions, and although it scales to very large network environments, it still only provides one active path from one device to another, regardless of how many actual connections might exist in the network. Although STP is a robust and scalable solution to redundancy in a Layer 2 network, the single logical link creates two problems: At least half of the available system bandwidth is off-limits to data traffic, and network topology changes occur. The Rapid Spanning Tree Protocol (RSTP) reduces the overhead of the rediscovery process and allows a Layer 2 network to reconverge faster, but the delay is still high.

Link aggregation (IEEE 802.3ad) solves some of these problems by enabling users to use more than one link connection between switches. All physical connections are

considered one logical connection. The problem with standard link aggregation is that the connections are point to point.

Multichassis link aggregation groups (MC-LAGs) enable a client device to form a logical LAG interface between two MC-LAG peers (QFX3500 and QFX3600 devices). An MC-LAG provides redundancy and load balancing between the two MC-LAG peers, multihoming support, and a loop-free Layer 2 network without running the Spanning Tree Protocol (STP).

On one end of an MC-LAG, there is an MC-LAG client device, such as a server, that has one or more physical links in a link aggregation group (LAG). This client device does not need to have an MC-LAG configured. On the other side of the MC-LAG, there are two MC-LAG peers. Each of the MC-LAG peers has one or more physical links connected to a single client device.

The MC-LAG peers use Interchassis Control Protocol (ICCP) to exchange control information and coordinate with each other to ensure that data traffic is forwarded properly.

Link Aggregation Control Protocol (LACP) is a subcomponent of the IEEE 802.3ad standard. LACP is used to discover multiple links from a client device connected to an MC-LAG peer. LACP must be configured on all member links for an MC-LAG to work correctly.

See [Table 1 on page 3](#) for information about ICCP failure scenarios.

The following sections provide an overview of the terms and features associated with MC-LAG:

- [Active-Active Mode on page 3](#)
- [ICCP and ICL-PL on page 3](#)
- [Failure Handling on page 3](#)
- [Multichassis Link Protection on page 4](#)
- [MC-LAG Packet Forwarding on page 4](#)
- [Layer 3 Routing on page 4](#)
- [Spanning Tree Protocol \(STP\) Guidelines on page 5](#)
- [MC-LAG Upgrade Guidelines on page 5](#)
- [Layer 2 Unicast Features Supported on page 6](#)
- [Layer 2 Multicast Features Supported on page 6](#)
- [IGMP Snooping on an Active-Active MC-LAG on page 6](#)
- [Layer 3 Unicast Features Supported on page 7](#)
- [VRRP Active-Standby Support on page 7](#)
- [Routed VLAN Interface \(RVI\) MAC Address Synchronization on page 8](#)
- [Address Resolution Protocol \(ARP\) on page 8](#)
- [DHCP Relay with Option 82 on page 8](#)

- [Private VLAN \(PVLAN\) on page 9](#)
- [Layer 3 Multicast on page 9](#)

Active-Active Mode

In active-active mode, all member links are active on the MC-LAG. In this mode, MAC addresses learned on one MC-LAG peer are propagated to the other MC-LAG peer. Active-active mode is the only mode supported at this time.

ICCP and ICL-PL

ICCP replicates control traffic and forwarding states across the MC-LAG peers and communicates the operational state of the MC-LAG members. Because ICCP uses TCP/IP to communicate between the peers, the two peers must be connected to each other. ICCP messages exchange MC-LAG configuration parameters and ensure that both peers use the correct LACP parameters.

The interchassis link-protection link (ICL-PL) provides redundancy when a link failure (for example, an MC-LAG trunk failure) occurs on one of the active links. The ICL-PL can be either a 10-Gigabit Ethernet interface or an aggregated Ethernet interface. You can configure only one ICL-PL between the two peers, although you can configure multiple MC-LAGs between them.

Failure Handling

Configuring ICCP adjacency over aggregated links mitigates the possibility of a split-brain state. A split brain state occurs when the ICL-PL configured between the MC-LAG peers goes down. To work around this problem, enable backup liveness detection. With backup liveness enabled, the MC-LAG peers can communicate through the keepalive link.

During a split-brain state, the standby peer brings down local members in the MC-LAG links by changing the LACP system ID. When the ICCP connection is active, both of the MC-LAG peers use the configured LACP system ID. If the LACP system ID is changed during failures, the server that is connected over the MC-LAG removes these links from the aggregated Ethernet bundle.

When the ICL-PL is operationally down and the ICCP connection is active, the LACP state of the links with status control configured as standby is set to the standby state. When the LACP state of the links is changed to standby, the server that is connected over the MC-LAG makes these links inactive and does not use them for sending data.

[Table 1 on page 3](#) describes the different ICCP failure scenarios. The dash means that the item is not applicable.

Table 1: ICCP Failure Scenarios

ICCP Connection Status	ICL-PL Status	Backup Liveness Peer Status	Action on Multichassis Aggregated Ethernet (MC-AE) Interface with Status Set to Standby
Down	Down or Up	Not configured	LACP system ID is changed to default value.
Down	Down or Up	Active	LACP system ID is changed to default value.

Table 1: ICCP Failure Scenarios (*continued*)

ICCP Connection Status	ICL-PL Status	Backup Liveness Peer Status	Action on Multichassis Aggregated Ethernet (MC-AE) Interface with Status Set to Standby
Down	Down or Up	Inactive	No change in LACP system ID.
Up	Down	–	LACP state is set to standby. MUX state moves to waiting state.

Split-brain states bring down the MC-LAG link completely if the primary peer members are also down for other reasons. Recovery from the split-brain state occurs automatically when the ICCP adjacency comes up between the MC-LAG peers.

Multichassis Link Protection

Multichassis link protection provides link protection between the two MC-LAG peers hosting an MC-LAG. If the ICCP connection is up and the ICL-PL comes up, the peer configured as standby brings up the multichassis aggregated Ethernet (MC-AE) interfaces shared with the peer. Multichassis protection must be configured on each MC-LAG peer that is hosting an MC-LAG.

MC-LAG Packet Forwarding

To prevent the server from receiving multiple copies from both of the MC-LAG peers, a block mask is used to prevent forwarding of traffic received on the ICL-PL toward the MC-AE interface. Preventing forwarding of traffic received on the ICL-PL interface toward the MC-AE interface ensures that traffic received on MC-LAG links is not forwarded back to the same link on the other peer. The forwarding block mask for a given MC-LAG link is cleared if all of the local members of the MC-LAG link go down on the peer. To achieve faster convergence, if all local members of the MC-LAG link are down, outbound traffic on the MC-LAG is redirected to the ICL-PL interface on the data plane.

Layer 3 Routing

To provide Layer 3 routing functions to downstream clients, configure the same gateway address on both MC-LAG network peers. To upstream routers, the MC-LAG network peers could be viewed as either equal-cost multipath (ECMP) or two routes with different preference values.

Junos OS supports active-active MC-LAGs by using Virtual Router Redundancy Protocol (VRRP) over routed VLAN interfaces (RVIs). Junos OS also supports active-active MC-LAGs by using RVI MAC address synchronization. You must configure the RVI using the same IP address across MC-LAG peers. RVI MAC synchronization is supported on 32-bit interfaces and interoperates with earlier MPC and MIC releases.

Spanning Tree Protocol (STP) Guidelines

- Enable STP globally.

STP might detect local miswiring loops within the peer or across MC-LAG peers.

STP might not detect network loops introduced by MC-LAG peers.

- Disable STP on ICL-PL links; otherwise, it might block ICL-PL ports and disable protection.
- Disable STP on interfaces that are connected to aggregation switches.
- Do not enable bridge protocol data unit (BPDU) block on interfaces connected to aggregation switches.

For more information about BPDU block, see Understanding BPDU Protection for STP, RSTP, and MSTP.

MC-LAG Upgrade Guidelines

Upgrade the MC-LAG peers according to the following guidelines. See Upgrading Software on QFX3500 and QFX3600 Standalone Switches for exact details about how to perform a software upgrade.



NOTE: After a reboot, the MC-AE interfaces come up immediately and might start receiving packets from the server. If routing protocols are enabled, and the routing adjacencies have not been formed, packets might be dropped.

To prevent this scenario, issue the `set interfaces interface-name aggregated-ether-options mc-ae init-delay-time time` to set a time by which the routing adjacencies are formed.

1. Make sure that both of the MC-LAG peers (node1 and node2) are in the active-active state using the following command on any one of the MC-LAG peers:

```
user@switch> show interfaces mc-ae id 1
Member Link           : ae0
Current State Machine's State: mcae active state
Local Status          : active<<<<<<<<
Local State           : up
Peer Status           : active<<<<<<<<
Peer State            : up
  Logical Interface    : ae0.0
  Topology Type        : bridge
  Local State         : up
  Peer State          : up
  Peer Ip/MCP/State    : 20.1.1.2 ae2.0 up
```

2. Upgrade node1 of the MC-LAG.

When node1 is upgraded it is rebooted, and all traffic is sent across the available LAG interfaces of node2, which is still up. The amount of traffic lost depends on how quickly the neighbor devices detect the link loss and rehash the flows of the LAG.

3. Verify that node1 is running the software you just installed. Issue the **show version** command.
4. Make sure that both nodes of the MC-LAG (node1 and node2) are in the active-active state after the reboot of node1.
5. Upgrade node2 of the MC-LAG.
Repeat step 1 through step 3 to upgrade node2.

Layer 2 Unicast Features Supported

The following Layer 2 unicast features are supported:

- L2 unicast: learning and aging
 - Learned MAC addresses are propagated across MC-LAG peers for all of the VLANs that are spawned across the peers.
 - Aging of MAC addresses occurs when the MAC address is not seen on both of the peers.
 - MAC learning is disabled on the ICL-PL automatically.
 - MAC addresses learned on single-homed links are propagated across all of the VLANs that have MC-LAG links as members.

Layer 2 Multicast Features Supported

The following Layer 2 multicast features are supported:

- L2 multicast: unknown unicast and IGMP snooping
 - Flooding happens on all links across peers if both peers have virtual LAN membership. Only one of the peers forwards traffic on a given MC-LAG link.
 - Known and unknown multicast packets are forwarded across the peers by adding the ICL-PL port as a multicast router port.
 - IGMP membership learned on MC-LAG links is propagated across peers.
 - During an MC-LAG peer reboot, known multicast traffic is flooded until the IGMP snooping state is synced with the peer.

IGMP Snooping on an Active-Active MC-LAG

IGMP snooping controls multicast traffic in a switched network. When IGMP snooping is not enabled, the Layer 2 device broadcasts multicast traffic out of all of its ports, even if the hosts on the network do not want the multicast traffic. With IGMP snooping enabled, a Layer 2 device monitors the IGMP join and leave messages sent from each connected host to a multicast router. This enables the Layer 2 device to keep track of the multicast groups and associated member ports. The Layer 2 device uses this information to make intelligent decisions and to forward multicast traffic to only the intended destination hosts. IGMP uses Protocol Independent Multicast (PIM) to route the multicast traffic. PIM uses distribution trees to determine which traffic is forwarded.

In an active-active MC-LAG configuration, IGMP snooping replicates the Layer 2 multicast routes so that each MC-LAG peer has the same routes. If a device is connected to an MC-LAG peer by way of a single-homed interface, IGMP snooping replicates join message to its IGMP snooping peer. If a multicast source is connected to an MC-LAG by way of a Layer 3 device, the Layer 3 device passes this information to the RVI that is configured on the MC-LAG. The first hop DR is responsible for sending the register and register-stop messages for the multicast group. The last hop DR is responsible for sending PIM join and leave messages toward the rendezvous point and source for the multicast group. The routing device with the smallest preference metric forwards traffic on transit LANs.

Configure the ICL-PL interface as a router-facing interface. For the scenario in which traffic arrives by way of a Layer 3 interface, PIM and IGMP must be enabled on the RVI interface configured on the MC-LAG peers.

Layer 3 Unicast Features Supported

The following Layer 3 unicast features are supported:

- VRRP active-standby support enables Layer 3 routing over MC-AE interfaces.
- Routed VLAN interface (RVI) MAC address synchronization enables MC-LAG peers to forward Layer 3 packets arriving on MC-AE interfaces with either its own RVI MAC address or its peer's RVI MAC address.
- Address Resolution Protocol (ARP) synchronization enables ARP resolution on both of the MC-LAG peers.
- DHCP Relay with option 82 enables option 82 on the MC-LAG peers. Option 82 provides information about the network location of DHCP clients. The DHCP server uses this information to implement IP addresses or other parameters for the client.

VRRP Active-Standby Support

VRRP in active-standby mode enables Layer 3 routing over the MC-AE interfaces on the MC-LAG peers. In this mode, the MC-LAG peers act as virtual routers. The virtual routers share the virtual IP address that corresponds to the default route configured on the host or server connected to the MC-LAG. This virtual IP address, known as a routed VLAN interface (RVI), maps to either of the VRRP MAC addresses or the logical interfaces of the MC-LAG peers. The host or server uses the VRRP MAC address to send any Layer 3 upstream packets. At any time, one of the VRRP routers is the master (active), and the other is a backup (standby). Both VRRP active and VRRP backup routers forward Layer 3 traffic arriving on the MC-AE interface. If the master router fails, all the traffic shifts to the MC-AE link on the backup router.



NOTE: You must configure VRRP on both MC-LAG peers in order for both the active and standby members to accept and route packets. Additionally, configure the VRRP backup router to send and receive ARP requests.

Routing protocols run on the primary IP address of the RVI, and both of the MC-LAG peers run routing protocols independently. The routing protocols use the primary IP address of the RVI and the RVI MAC address to communicate with the MC-LAG peers. The RVI MAC address of each MC-LAG peer is replicated on the other MC-LAG peer and is installed as a MAC address that has been learned on the ICL-PL.

Routed VLAN Interface (RVI) MAC Address Synchronization

Routed VLAN interface (RVI) MAC address synchronization enables MC-LAG peers to forward Layer 3 packets arriving on MC-AE interfaces with either its own RVI MAC address or its peer's RVI MAC address. Each MC-LAG peer installs its own RVI MAC address as well as the peer's RVI MAC address in the hardware. Each MC-LAG peer treats the packet as if it were its own packet. If RVI MAC address synchronization is not enabled, the RVI MAC address is installed on the MC-LAG peer as if it was learned on the ICL-PL.



NOTE: If you need routing capability, configure both VRRP and routing protocols on each MC-LAG peer.

Control packets destined for a particular MC-LAG peer that arrive on an MC-AE interface of its MC-LAG peer are not forwarded on the ICL-PL interface. Additionally, using the gateway IP address as a source address when you issue either a ping, traceroute, telnet, or FTP request is not supported.

To enable RVI MAC address synchronization, issue the **set vlan *vlan-name* l3_interface *rvi-name* mcae-mac-synchronize** on each MC-LAG peer. Configure the same IP address on both MC-LAG peers. This IP address is used as the default gateway for the MC-LAG servers or hosts.

Address Resolution Protocol (ARP)

Address Resolution Protocol (ARP) maps IP addresses to MAC addresses. Without synchronization, if one MC-LAG peer sends an ARP request, and the other MC-LAG peer receives the response, ARP resolution is not successful. With synchronization, the MC-LAG peers synchronize the ARP resolutions by sniffing the packet at the MC-LAG peer receiving the ARP response and replicating this to the other MC-LAG peer. This ensures that the entries in ARP tables on the MC-LAG peers are consistent.

When one of the MC-LAG peers restarts, the ARP destinations on its MC-LAG peer are synchronized. Because the ARP destinations are already resolved, its MC-LAG peer can forward Layer 3 packets out of the MC-AE interface.

DHCP Relay with Option 82

DHCP relay with option 82 provides information about the network location of DHCP clients. The DHCP server uses this information to implement IP addresses or other parameters for the client. With DHCP relay enabled, DHCP request packets might take the path to the DHCP server through either of the MC-LAG peers. Because the MC-LAG peers have different host names, chassis MAC addresses, and interface names, you need to observe these requirements when you configure DHCP relay with option 82:

-
- Use the interface description instead of the interface name.
 - Do not use the hostname as part of the circuit ID or remote ID strings.
 - Do not use the chassis MAC address as part of the remote ID string.
 - Do not enable the vendor ID.
 - If the ICL-PL interface receives DHCP request packets, the packets are dropped to avoid duplicate packets in the network.

A counter called *Due to received on ICL interface* has been added to the **show helper statistics** command, which tracks the packets that the ICL-PL interface drops.

An example of the CLI output follows:

```
user@switch> show helper statistics
BOOTP:
  Received packets: 6
  Forwarded packets: 0
  Dropped packets: 6
    Due to no interface in DHCP Relay database: 0
    Due to no matching routing instance: 0
    Due to an error during packet read: 0
    Due to an error during packet send: 0
    Due to invalid server address: 0
    Due to no valid local address: 0
    Due to no route to server/client: 0
    Due to received on ICL interface: 6
```

The output shows that six packets received on the ICL-PL interface have been dropped.

Private VLAN (PVLAN)

Private VLANs allow you to split a broadcast domain into multiple isolated broadcast subdomains, essentially putting a VLAN inside of a VLAN. A PVLAN can span multiple peers on an MC-LAG.

When configuring a PVLAN, you must configure the ICL-PL interface as the PVLAN trunk interface for the PVLAN. This is essential for traffic to be switched to the required primary and secondary ports of the PVLAN across the MC-LAG peers.

Layer 3 Multicast

- [PIM Operation With Normal Mode DR Election on page 9](#)
- [PIM Operation with Dual-DR Mode on page 10](#)
- [Configuration Guidelines and Caveats on page 10](#)

Protocol Independent Multicast (PIM) and Internet Group Management Protocol (IGMP) provide support for Layer 3 multicast. In addition to the standard mode of PIM operation, there is a special mode called PIM dual DR (designated router). PIM dual DR minimizes traffic loss in case of failures.

PIM Operation With Normal Mode DR Election

In normal mode DR election, the RVI interfaces on both of the MC-LAG peers are configured with PIM enabled. In this mode, one of the MC-LAG peers becomes the DR

through the PIM DR election mechanism. The elected DR maintains the rendezvous-point tree (RPT) and shortest-path tree (SPT) so it can receive data from the source device. The elected DR participates in periodic PIM join and prune activities toward the rendezvous point (RP) or the source.

The trigger for initiating these join and prune activities is the IGMP membership reports that are received from interested receivers. IGMP reports received over MC-AE interfaces (potentially hashing on either of the MC-LAG peers) and single-homed links are synchronized to the MC-LAG peer through ICCP.

Both MC-LAG peers receive traffic on their incoming interface (IIF). The non-DR receives traffic by way of the ICL-PL interface, which acts as a multicast router (mrouter) interface.

If the DR fails, the non-DR has to build the entire forwarding tree (RPT and SPT), which can cause multicast traffic loss.

PIM Operation with Dual-DR Mode

In this mode, both of MC-LAG peers act as DRs (active and backup) and send periodic join and prune messages upstream towards the RP, or source, and eventually join the RPT or SPT.

The primary MC-LAG peer forwards the multicast traffic to the receiver devices even if the standby MC-LAG peer has a smaller preference metric.

The standby MC-LAG peer also joins the forwarding tree and receives the multicast data. The standby MC-LAG peer drops the data because it has an empty outgoing interface list (OIL). When the standby MC-LAG peer detects the primary MC-LAG peer failure, it adds the receiver VLAN to the OIL, and starts to forward the multicast traffic

To enable a multicast dual DR, issue the **set protocols pim interface interface-name dual-dr** command on the VLAN interfaces of each MC-LAG peer.

Configuration Guidelines and Caveats

- Configure the IP address on the active MC-LAG peer with a high IP address or a high DR priority. To ensure that the active MC-LAG peer retains the DR membership designation if PIM neighborship with the peer goes down.
- Using Bidirectional Forwarding Detection (BFD) and RVI MAC synchronization together is not supported because ARP fails.
- When using RVI MAC synchronization, make sure that you configure the primary IP address on both MC-LAG peers. Doing this ensures that both MC-LAG peers cannot become assert winners.
- The number of BFD sessions on RVIs with PIM enabled is restricted to 100. Also, if you have more than 100 RVIs configured, do not configure BFD, and make sure that the hello interval is 2 seconds.

Related Documentation

- [Advantages of Using Multichassis Link Aggregation Groups on page 1](#)
- [Configuring Multichassis Link Aggregation on page 26](#)

-
- [Example: Configuring Multichassis Link Aggregation on page 11](#)

Example: Configuring Multichassis Link Aggregation

Multichassis link aggregation groups (MC-LAGs) enable a client device to form a logical LAG interface between two switches. An MC-LAG provides redundancy and load balancing between the two switches, multihoming support, and a loop-free Layer 2 network without running Spanning Tree Protocol (STP).

The peers in an MC-LAG use an interchassis control link-protection link (ICL-PL) to replicate forwarding information across the peers. The Interchassis Control Protocol (ICCP) exchanges the control information between two MC-LAG switches. Additionally, ICCP propagates the operational state of MC-LAG members through the ICL-PL.

On one end of an MC-LAG is an MC-LAG client device, such as a server, that has one or more physical links in a link aggregation group (LAG). This client device does not need to detect the MC-LAG. On the other side of an MC-LAG are two MC-LAG switches. Each of the switches has one or more physical links connected to a single client device. The switches coordinate with each other to ensure that data traffic is forwarded properly.

- [Requirements on page 11](#)
- [Overview on page 12](#)
- [Configuration on page 13](#)
- [Troubleshooting on page 25](#)

Requirements

This example uses the following hardware and software components:

- Junos OS Release 12.2 or later for the QFX Series
- Two switches



NOTE: This configuration example has been tested using the software release listed and is assumed to work on all later releases.

Before you configure an MC-LAG, be sure that you understand how to:

- Configure aggregated Ethernet interfaces on a switch. See [Example: Configuring Link Aggregation Between a QFX Series Product and an Aggregation Switch](#).
- Configure the Link Aggregation Control Protocol (LACP) on aggregated Ethernet interfaces on a switch. See [Example: Configuring Link Aggregation with LACP Between a QFX Series Product and an Aggregation Switch](#).

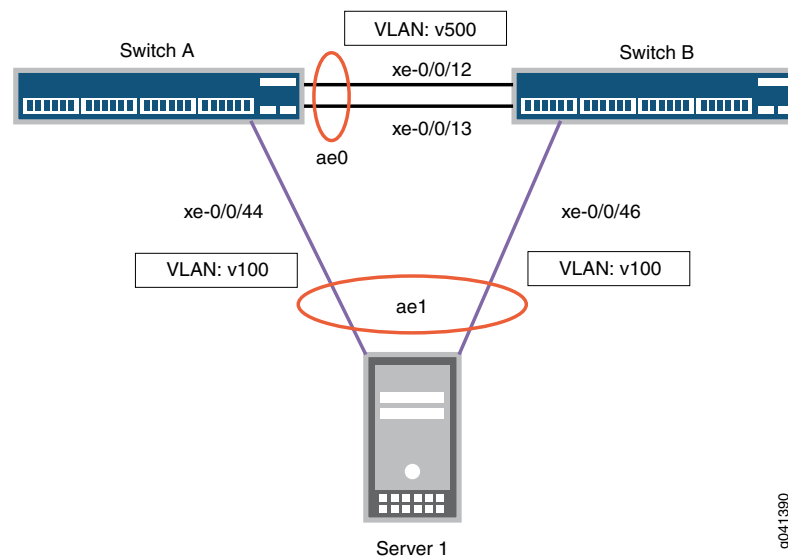
Overview

In this example, you configure an MC-LAG across two switches, consisting of two aggregated Ethernet interfaces, an interchassis control link-protection link (ICL-PL), multichassis protection link for the ICL-PL, ICCP for the peers hosting the MC-LAG, and Layer 3 connectivity between MC-LAG peers. Layer 3 connectivity is required for ICCP.

Topology

The topology used in this example consists of two switches hosting an MC-LAG. The two switches are connected to a server. [Figure 1 on page 12](#) shows the topology of this example.

Figure 1: Configuring a Multichassis LAG Between Switch A and Switch B



[Table 2 on page 12](#) details the topology used in this configuration example.

Table 2: Components of the Topology for Configuring a Multichassis LAG Between Two Switches

Hostname	Base Hardware	Multichassis Link Aggregation Group
Switch A	QFX3500 switch or QFX3600 switch	ae0 is configured as an aggregated Ethernet interface, and is used as an ICL-PL. The following interfaces are part of ae0 : xe-0/0/12 and xe-0/0/13 on Switch A and xe-0/0/12 and xe-0/0/13 on Switch B. ae1 is configured as an MC-LAG, and the following two interfaces are part of ae1 : xe-0/0/44 on Switch A and xe-0/0/46 on Switch B. .
Switch B	QFX3500 switch or QFX3600 switch	

Configuration

CLI Quick Configuration To quickly configure this example, copy the following commands, paste them in a text file, remove any line breaks, change any details necessary to match your network configuration, and paste the commands into the CLI at the **[edit]** hierarchy level of Switch A.

```
set chassis aggregated-devices ethernet device-count 2
set interfaces xe-0/0/12 ether-options 802.3ad ae0
set interfaces xe-0/0/13 ether-options 802.3ad ae0
set interfaces xe-0/0/44 ether-options 802.3ad ae1
set interfaces ae0 unit 0 family ethernet-switching port-mode trunk
set interfaces ae0 unit 0 family ethernet-switching vlan members v500
set interfaces ae1 aggregated-ether-options lacp active
set interfaces ae1 aggregated-ether-options lacp system-id 00:01:02:03:04:05
set interfaces ae1 aggregated-ether-options lacp admin-key 3
set interfaces ae1 aggregated-ether-options mc-ae mc-ae-id 3
set interfaces ae1 aggregated-ether-options mc-ae chassis-id 0
set interfaces ae1 aggregated-ether-options mc-ae mode active-active
set interfaces ae1 aggregated-ether-options mc-ae status-control active
set ae1 aggregated-ether-options mc-ae init-delay-time 240
set interfaces ae1 unit 0 family ethernet-switching port-mode trunk
set interfaces ae1 unit 0 family ethernet-switching vlan members v100
set interfaces vlan unit 500 family inet address 3.3.3.2/24
set vlans v100 vlan-id 100
set vlans v500 vlan-id 500
set vlans v500 l3-interface vlan.500
set protocols iccp peer 3.3.3.1 session-establishment-hold-time 50
set protocols iccp peer 3.3.3.1 backup-liveness-detection backup-peer-ip 10.207.64.233
set protocols iccp peer 3.3.3.1 liveness-detection minimum-receive-interval 60
set protocols iccp peer 3.3.3.1 liveness-detection transmit-interval minimum-interval 60
set protocols rstp interface ae0.0 disable
set protocols rstp interface ae1.0 edge
set protocols rstp interface all mode point-to-point
set protocols rstp bpdu-block-on-edge
set multi-chassis multi-chassis-protection 3.3.3.1 interface ae0
```

To quickly configure this example, copy the following commands, paste them in a text file, remove any line breaks, change any details necessary to match your network configuration, and paste the commands into the CLI at the **[edit]** hierarchy level of Switch B.

```
set chassis aggregated-devices ethernet device-count 2
set interfaces xe-0/0/12 ether-options 802.3ad ae0
set interfaces xe-0/0/13 ether-options 802.3ad ae0
set interfaces xe-0/0/46 ether-options 802.3ad ae1
set interfaces ae0 unit 0 family ethernet-switching port-mode trunk
set interfaces ae0 unit 0 family ethernet-switching vlan members v500
set interfaces ae1 aggregated-ether-options lacp active
set interfaces ae1 aggregated-ether-options lacp system-id 00:01:02:03:04:05
set interfaces ae1 aggregated-ether-options lacp admin-key 3
set interfaces ae1 aggregated-ether-options mc-ae mc-ae-id 3
set interfaces ae1 aggregated-ether-options mc-ae chassis-id 1
set interfaces ae1 aggregated-ether-options mc-ae mode active-active
set interfaces ae1 aggregated-ether-options mc-ae status-control standby
set ae1 aggregated-ether-options mc-ae init-delay-time 240
set interfaces ae1 unit 0 family ethernet-switching port-mode trunk
set interfaces ae1 unit 0 family ethernet-switching vlan members v100
```

```
set interfaces vlan unit 500 family inet address 3.3.3.1/24
set vlans v100 vlan-id 100
set vlans v500 vlan-id 500
set vlans v500 l3-interface vlan.500
set protocols iccp local-ip-addr 3.3.3.1
set protocols iccp peer 3.3.3.2 session-establishment-hold-time 50
set protocols iccp peer 3.3.3.2 backup-liveness-detection backup-peer-ip 10.207.64.233
set protocols iccp peer 3.3.3.2 liveness-detection minimum-receive-interval 60
set protocols iccp peer 3.3.3.2 liveness-detection transmit-interval minimum-interval 60
set protocols rstp interface ae0.0 disable
set protocols rstp interface ae1.0 edge
set protocols rstp interface all mode point-to-point
set protocols rstp bpdu-block-on-edge
set multi-chassis multi-chassis-protection 3.3.3.2 interface ae0
```

Configuring MC-LAG on Two Switches

Step-by-Step Procedure

To enable multichassis protection link between MC-LAG peers:

The following example requires you to navigate various levels in the configuration hierarchy. For information about navigating the CLI, see *Using the CLI Editor in Configuration Mode* in the *CLI User Guide*.

1. Configure the number of LAGs on both Switch A and Switch B.

```
[edit chassis]
user@switch# set aggregated-devices ethernet device-count 2
```
2. Add member interfaces to the aggregated Ethernet interfaces on both Switch A and Switch B.

```
[edit interfaces]
user@switch# set xe-0/0/12 ether-options 802.3ad ae0
[edit interfaces]
user@switch# set xe-0/0/13 ether-options 802.3ad ae0
[edit interfaces]
user@switch# set xe-0/0/44 ether-options 802.3ad ae1
[edit interfaces]
user@switch# set xe-0/0/46 ether-options 802.3ad ae1
```

3. Configure a trunk interface between Switch A and Switch B.

```
[edit interfaces]
user@switch# set ae0 unit 0 family ethernet-switching port-mode trunk
```

4. Configure a multichassis protection link between Switch A and Switch B.

Switch A:

```
[edit]
user@switch# set multi-chassis multi-chassis-protection 3.3.3.2 interface ae0
```

Switch B:

```
[edit]
user@switch# set multi-chassis multi-chassis-protection 3.3.3.1 interface ae0
```

Step-by-Step Procedure

To enable ICCP:

The following example requires you to navigate various levels in the configuration hierarchy. For information about navigating the CLI, see *Using the CLI Editor in Configuration Mode* in the *CLI User Guide*.

1. Configure the local IP address to be in the ICCP connection on Switch A and Switch B.

Switch A:

```
[edit protocols]
user@switch# set iccp local-ip-addr 3.3.3.2
```

Switch B:

```
[edit protocols]
user@switch# set iccp local-ip-addr 3.3.3.1
```

2. Configure the peer IP address and minimum receive interval for a (BFD) session for ICCP on Switch A and Switch B.

Switch A:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.1 liveness-detection minimum-receive-interval 60
```

Switch B:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 liveness-detection minimum-receive-interval 60
```

3. Configure the peer IP address and minimum transmit interval for Bidirectional Forwarding Detection (BFD) session for ICCP on Switch A and Switch B.

Switch A:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.1 liveness-detection transmit-interval minimum-interval 60
```

Switch B:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 liveness-detection transmit-interval minimum-interval 60
```

4. (Optional) Configure the time during which an ICCP connection must succeed between MC-LAG peers on Switch A and Switch B.



NOTE: Configuring session establishment hold time helps in faster ICCP connection establishment. The recommended value is 50 seconds.

Switch A:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.1 session-establishment-hold-time 50
```

Switch B:

```
[edit protocols]
```

- user@switch# set iccp peer 3.3.3.2 session-establishment-hold-time 50
- (Optional) Configure the backup IP address to be used for backup liveness detection on both Switch A and Switch B.



NOTE: By default, backup liveness detection is not enabled. Configuring a backup IP address helps achieve sub-second traffic loss during a MC-LAG peer reboot.

Switch A:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.1 backup-liveness-detection backup-peer-ip 10.207.64.233
```

Switch B:

- ```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 backup-liveness-detection backup-peer-ip 10.207.64.232
```
- Configure Layer 3 connectivity between the MC-LAG peers on both Switch A and Switch B.

```
[edit vlans]
user@switch# set v500 vlan-id 500
[edit vlans]
user@switch# set v500 l3-interface vlan.500
[edit interfaces]
user@switch# set ae0 unit 0 family ethernet-switching port-mode trunk vlan members v500
```

#### Step-by-Step Procedure

To enable the MC-LAG interface:

The following example requires you to navigate various levels in the configuration hierarchy. For information about navigating the CLI, see *Using the CLI Editor in Configuration Mode* in the *CLI User Guide*.

- Enable LACP on the MC-LAG interface on Switch A and Switch B.



**NOTE:** At least one end needs to be active. The other end can be either active or passive.

- ```
[edit interfaces]
user@switch# set ae1 aggregated-ether-options lacp active
```
- Specify the same multichassis aggregated Ethernet identification number on both MC-LAG peers on Switch A and Switch B.

- ```
[edit interfaces]
user@switch# set ae1 aggregated-ether-options mc-ae mc-ae-id 3
```
- Specify a unique chassis ID for the MC-LAG on the MC-LAG peers on Switch A and Switch B.

Switch A:

```
[edit interfaces]
```

---

```
user@switch# set ae1 aggregated-ether-options mc-ae chassis-id 0
```

Switch B:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae chassis-id 1
```

4. Specify the operating mode of the MC-LAG on both Switch A and Switch B.



**NOTE:** Only active-active mode is supported at this time.

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae mode active-active
```

5. Specify the status control for MC-LAG on Switch A and Switch B.



**NOTE:** You must configure status control on both Switch A and Switch B hosting the MC-LAG. If one peer is in active mode, the other must be in standby mode.

Switch A:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae status-control active
```

Switch B:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae status-control standby
```

6. Specify the number of seconds by which the bring-up of the MC-AE interface should be deferred after you reboot Switch A and Switch B.



**NOTE:** The recommended value for maximum VLAN configuration (for example, 4,000 VLANs) is 240 seconds. If IGMP snooping is enabled on all of the VLANs, the recommended value is 420 seconds.

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae init-delay-time 240
```

7. Specify the same LACP system ID for the MC-LAG on Switch A and Switch B.

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options lacp system-ID 00:01:02:03:04:05
```

8. Specify the same LACP administration key on both Switch A and Switch B.

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options lacp admin-key 3
```

9. Enable a VLAN on the MC-LAG on Switch A and Switch B.

```
[edit interfaces]
```

```
user@switch# set ae1 unit 0 family ethernet-switching port-mode trunk
```

```
[edit]
```

```
user@switch# set vlans v100 vlan-id 100
```

```
[edit interfaces]
```

- ```

user@switch# set ae1 unit 0 family ethernet-switching vlan members v100

```
10. (Optional) Enable a private VLAN on the MC-LAG on Switch A and Switch B.


```

[edit]
user@switch# set vlans v1an100 pvlan isolation-vlan-id 200
extend-secondary-vlan-id
[edit]
user@switch# set vlans v1an100 interface ae0.0 pvlan-trunk

```

Step-by-Step Procedure

To enable RSTP:

The following example requires you to navigate various levels in the configuration hierarchy. For information about navigating the CLI, see *Using the CLI Editor in Configuration Mode* in the *CLI User Guide*.

1. Enable RSTP globally on all interfaces on Switch A and Switch B.


```

[edit]
user@switch# set protocols rstp interface all mode point-to-point

```
2. Disable RSTP on the ICL-PL interfaces on Switch A and Switch B:


```

[edit]
user@switch# set protocols rstp interface ae0.0 disable

```
3. Configure the MC-LAG interfaces as edge ports on Switch A and Switch B.



NOTE: The ae1 interface is a downstream interface. This is why RSTP and bpd-block-on-edge need to be configured.

- ```

[edit]
user@switch# set protocols rstp interface ae1.0 edge

```
4. Enable BPDU blocking on all interfaces except for the ICL-PL interfaces on Switch A and Switch B.



**NOTE:** The ae1 interface is a downstream interface. This is why RSTP and bpd-block-on-edge need to be configured.

```

[edit]
user@switch# set protocols rstp bpd-block-on-edge

```

### Verification

To verify that the MC-LAG group has been created and is working properly, perform these tasks:

- [Verifying That ICCP Is Working on Switch A on page 19](#)
- [Verifying That ICCP Is Working on Switch B on page 19](#)
- [Verifying That LACP Is Active on Switch A on page 19](#)
- [Verifying That LACP Is Active on Switch B on page 20](#)
- [Verifying That the MC-AE and ICL-PL Interfaces Are Up on Switch A on page 20](#)

- [Verifying That the MC-AE and ICL-PL Interfaces Are Up on Switch B on page 20](#)
- [Verifying that MAC Learning Is Occurring on Switch A on page 21](#)
- [Verifying that MAC Learning Is Occurring on Switch B on page 21](#)

#### ***Verifying That ICCP Is Working on Switch A***

**Purpose** Verify that ICCP is running on Switch A.

**Action** [edit]  
 user@switch# **show iccp**  
 Redundancy Group Information for peer 3.3.3.1  
     TCP Connection : Established  
     Liveliness Detection : Up  
  
 Client Application: MCSNOOPD  
  
 Client Application: eswd

**Meaning** This output shows that the TCP connection between the peers hosting the MC-LAG is up, liveness detection is up, and MCSNOOPD and ESWD client applications are running.

#### ***Verifying That ICCP Is Working on Switch B***

**Purpose** Verify that ICCP is running on Switch B.

**Action** **show iccp**  
  
 [edit]  
 user@switch# **show iccp**  
 Redundancy Group Information for peer 3.3.3.2  
     TCP Connection : Established  
     Liveliness Detection : Up  
  
 Client Application: MCSNOOPD  
  
 Client Application: eswd

**Meaning** This output shows that the TCP connection between the peers hosting the MC-LAG is up, liveness detection is up, and MCSNOOPD and ESWD client applications are running.

#### ***Verifying That LACP Is Active on Switch A***

**Purpose** Verify that LACP is active on Switch A.

**Action** [edit]  
 user@switch# **show lacp interfaces**  
 Aggregated interface: ae1  

|             |         |     |     |      |     |     |      |         |          |
|-------------|---------|-----|-----|------|-----|-----|------|---------|----------|
| LACP state: | Role    | Exp | Def | Dist | Col | Syn | Aggr | Timeout | Activity |
| xe-0/0/46   | Actor   | No  | No  | Yes  | Yes | Yes | Yes  | Fast    | Active   |
| xe-0/0/46   | Partner | No  | No  | Yes  | Yes | Yes | Yes  | Fast    | Active   |

|                |               |                |                         |
|----------------|---------------|----------------|-------------------------|
| LACP protocol: | Receive State | Transmit State | Mux State               |
| xe-0/0/46      | Current       | Fast periodic  | Collecting distributing |

**Meaning** This output shows that Switch A is participating in LACP negotiation.

**Verifying That LACP Is Active on Switch B**

**Purpose** Verify that LACP is active on Switch B

**Action** [edit]

```
user@switch# show lacp interfaces
```

```
Aggregated interface: ae1
```

| LACP state: | Role    | Exp | Def | Dist | Col | Syn | Aggr | Timeout | Activity |
|-------------|---------|-----|-----|------|-----|-----|------|---------|----------|
| xe-0/0/44   | Actor   | No  | No  | Yes  | Yes | Yes | Yes  | Fast    | Active   |
| xe-0/0/44   | Partner | No  | No  | Yes  | Yes | Yes | Yes  | Fast    | Active   |

| LACP protocol: | Receive State | Transmit State | Mux State               |
|----------------|---------------|----------------|-------------------------|
| xe-0/0/44      | Current       | Fast periodic  | Collecting distributing |

**Meaning** This output shows that Switch B is participating in LACP negotiation.

**Verifying That the MC-AE and ICL-PL Interfaces Are Up on Switch A**

**Purpose** Verify that the MC-AE and ICL-PL interfaces are up on Switch A.

**Action** [edit]

```
user@switch# show interfaces mc-ae
```

```
Member Link : ae1
```

```
Current State Machine's State: mcae active state
```

```
Local Status : active
```

```
Local State : up
```

```
Peer Status : active
```

```
Peer State : up
```

```
Logical Interface : ae1.0
```

```
Topology Type : bridge
```

```
Local State : up
```

```
Peer State : up
```

```
Peer Ip/MCP/State : 3.3.3.1 ae0.0 up
```

**Meaning** This output shows that the MC-AE interface on Switch A is up and active.

**Verifying That the MC-AE and ICL-PL Interfaces Are Up on Switch B**

**Purpose** Verify that the MC-AE and ICL-PL interfaces are up on Switch B.



---

**Action** [edit]  
user@switch# **show interfaces mc-ae**  
Member Link : ae1  
Current State Machine's State: mcae active state  
Local Status : active  
Local State : up  
Peer Status : active  
Peer State : up  
Logical Interface : ae1.0  
Topology Type : bridge  
Local State : up  
Peer State : up  
Peer Ip/MCP/State : 3.3.3.2 ae0.0 up

**Meaning** This output shows that the MC-AE interface on Switch B is up and active.

#### *Verifying that MAC Learning Is Occurring on Switch A*

**Purpose** Verify that MAC learning is working on Switch A.

**Action** [edit]  
user@switch# **show ethernet-switching table**  
Ethernet-switching table: 10 entries, 4 learned, 0 persistent entries

| VLAN | MAC address       | Type     | Age | Interfaces    |
|------|-------------------|----------|-----|---------------|
| V100 | *                 | Flood    |     | - All-members |
| V100 | 00:10:94:00:00:05 | Learn(L) | 33  | ae0.0 (MCAE)  |

**Meaning** The output shows four learned MAC addresses entries.

#### *Verifying that MAC Learning Is Occurring on Switch B*

**Purpose** Verify that MAC learning is working on Switch B.

**Action** [edit]  
user@switch# **show ethernet-switching table**  
Ethernet-switching table: 10 entries, 4 learned, 0 persistent entries

| VLAN | MAC address       | Type     | Age | Interfaces    |
|------|-------------------|----------|-----|---------------|
| V100 | *                 | Flood    |     | - All-members |
| V100 | 00:10:94:00:00:05 | Learn(L) | 33  | ae0.0 (MCAE)  |

**Meaning** The output shows four learned MAC addresses entries.

### Results

---

Display the results of the configuration on Switch A.

```
chassis {
 aggregated-devices {
 ethernet {
 device-count 2;
 }
 }
}
```

```
interfaces {
 xe-0/0/12 {
 ether-options {
 802.3ad ae0;
 }
 }
 xe-0/0/13 {
 ether-options {
 802.3ad ae0;
 }
 }
 xe-0/0/44 {
 ether-options {
 802.3ad ae1;
 }
 }
}
ae0 {
 unit 0 {
 family ethernet-switching {
 port-mode trunk;
 vlan {
 members v500;
 }
 }
 }
}
ae1 {
 aggregated-ether-options {
 lacp {
 active;
 system-id 00:01:02:03:04:05;
 admin-key 3;
 }
 mc-ae {
 mc-ae-id 3;
 chassis-id 0;
 mode active-active;
 status-control active;
 init-delay-time 240
 }
 }
 unit 0 {
 family ethernet-switching {
 port-mode trunk;
 vlan {
 members v100;
 }
 }
 }
}
vlan {
 unit 500 {
 family inet {
 address 3.3.3.2/24;
 }
 }
}
```

---

```

 }
 }
 protocols {
 iccp {
 local-ip-addr 3.3.3.2;
 peer 3.3.3.1 {
 session-establishment-hold-time 50;
 backup-liveness-detection {
 backup-peer-ip 10.207.64.233;
 }
 liveness-detection {
 minimum-receive-interval 60;
 transmit-interval {
 minimum-interval 60;
 }
 }
 }
 }
 }
 rstp {
 interface ae0.0 {
 disable;
 }
 interface ae1.0 {
 edge;
 }
 interface all {
 mode point-to-point;
 }
 bpdu-block-on-edge;
 }
}
multi-chassis {
 multi-chassis-protection 3.3.3.1 {
 interface ae0;
 }
}
vllans {
 v100 {
 vlan-id 100;
 }
 v500 {
 vlan-id 500;
 l3-interface vlan.500;
 }
}
}

```

Display the results of the configuration on Switch B.

```

chassis {
 aggregated-devices {
 ethernet {
 device-count 2;
 }
 }
}
interfaces {

```

```
xe-0/0/12 {
 ether-options {
 802.3ad ae0;
 }
}
xe-0/0/13 {
 ether-options {
 802.3ad ae0;
 }
}
xe-0/0/46 {
 ether-options {
 802.3ad ae1;
 }
}
ae0 {
 unit 0 {
 family ethernet-switching {
 port-mode trunk;
 vlan {
 members v500;
 }
 }
 }
}
ae1 {
 aggregated-ether-options {
 lacp {
 active;
 system-id 00:01:02:03:04:05;
 admin-key 3;
 }
 mc-ae {
 mc-ae-id 3;
 chassis-id 1;
 mode active-active;
 status-control standby;
 init-delay-time 240
 }
 }
 unit 0 {
 family ethernet-switching {
 port-mode trunk;
 vlan {
 members v100;
 }
 }
 }
}
vlan {
 unit 500 {
 family inet {
 address 3.3.3.1/24;
 }
 }
}
```

```

}
protocols {
 iccp {
 local-ip-addr 3.3.3.1;
 peer 3.3.3.2 {
 session-establishment-hold-time 50;
 backup-liveness-detection {
 backup-peer-ip 10.207.64.233;
 }
 liveness-detection {
 minimum-receive-interval 60;
 transmit-interval {
 minimum-interval 60;
 }
 }
 }
 }
}
rstp {
 interface ae0.0 {
 disable;
 }
 interface ae1.0 {
 edge;
 }
 interface all {
 mode point-to-point;
 }
 bpdu-block-on-edge;
}
}
multi-chassis {
 multi-chassis-protection 3.3.3.2 {
 interface ae0;
 }
}
vllans {
 v100 {
 vlan-id 100;
 }
 v500 {
 vlan-id 500;
 l3-interface vlan.500;
 }
}
}

```

## Troubleshooting

### Troubleshooting a LAG That Is Down

**Problem** The `show interfaces terse` command shows that the MC-LAG is **down**

**Solution** Check the following:

- Verify that there is no configuration mismatch.
- Verify that all member ports are up.
- Verify that the MC-LAG is part of family Ethernet switching (Layer 2 LAG).
- Verify that the MC-LAG member is connected to the correct MC-LAG member at the other end.

**Related Documentation**

- *Understanding Multichassis Link Aggregation*
- *Configuring Multichassis Link Aggregation*

---

## Configuring Multichassis Link Aggregation

Multichassis link aggregation groups (MC-LAGs) enable a client device to form a logical LAG interface between two switches. An MC-LAG provides redundancy and load balancing between the two switches, multihoming support, and a loop-free Layer 2 network without running the Spanning Tree Protocol (STP). At this time, MC-LAGs support only Layer 2 features.

The MC-LAG switches use the Interchassis Control Protocol (ICCP) to exchange the control information between two MC-LAG switches.

On one end of an MC-LAG is an MC-LAG client device, such as a server, that has one or more physical links in a link aggregation group (LAG). This client device does not need to detect the MC-LAG. On the other side of MC-LAG are two MC-LAG switches. Each of the switches has one or more physical links connected to a single client device. The switches coordinate with each other to ensure that data traffic is forwarded properly.



**NOTE:** An interface with an already configured IP address cannot form part of the aggregated Ethernet interface or multichassis aggregated Ethernet interface group.

Perform the following steps on each switch that is hosting an MC-LAG:

1. Specify the same multichassis aggregated Ethernet identification number for the MC-LAG that the aggregated Ethernet interface belongs to on each switch.

```
[edit interfaces]
user@switch# set aeX aggregated-ether-options mc-ae mc-ae-id number
```

For example:

```
[edit interfaces]
user@switch# set ae1 aggregated-ether-options mc-ae mc-ae-id 3
```

2. Specify a unique chassis ID for the MC-LAG that the aggregated Ethernet interface belongs to on each switch.

```
[edit interfaces]
user@switch# set aeX aggregated-ether-options mc-ae chassis-id number
```

For example:

---

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae chassis-id 0
```

3. Specify the mode of the MC-LAG the aggregated Ethernet interface belongs to.



**NOTE:** Only active-active mode is supported at this time.

```
[edit interfaces]
```

```
user@switch# set aeX aggregated-ether-options mc-ae mode mode
```

For example:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae mode active-active
```

4. Specify whether the aggregated Ethernet interface participating in the MC-LAG is primary or secondary. Primary is **active**, and secondary is **standby**.



**NOTE:** You must configure status control on both switches hosting the MC-LAG. If one switch is in active mode, the other must be in standby mode.

```
[edit interfaces]
```

```
user@switch# set aeX aggregated-ether-options mc-ae status-control (active | standby)
```

For example:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options mc-ae status-control active
```

5. Specify the same LACP system ID on each switch.

```
[edit interfaces]
```

```
user@switch# set aeX aggregated-ether-options lacp system-id mac-address
```

For example:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options lacp system-id 00:01:02:03:04:05
```

6. Specify the same LACP administration key on each switch.

```
[edit interfaces]
```

```
user@switch# set aeX aggregated-ether-options lacp admin-key number
```

For example:

```
[edit interfaces]
```

```
user@switch# set ae1 aggregated-ether-options lacp admin-key 3
```

7. Configure ICCP by doing the following on each switch hosting the MC-LAG:

- a. Configure the local IP address to be used by all switches hosting the MC-LAG.

```
[edit protocols]
```

```
user@switch# set iccp local-ip-addr local-ip-address
```

For example:

```
[edit protocols]
```

```
user@switch# set iccp local-ip-addr 3.3.3.1
```

- b. (Optional) Configure the IP address of the switch and the time during which an ICCP connection must succeed between the switches hosting the MC-LAG.

Configured session establishment hold time results in faster ICCP connection establishment. The recommended value is 50 seconds.

```
[edit protocols]
user@switch# set iccp peer peer-ip-address session-establishment-hold-time seconds
```

For example:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 session-establishment-hold-time 50
```

- c. (Optional) Configure the IP address to be used for backup liveness detection:



**NOTE:** By default, backup liveness detection is not enabled. Configure backup liveness detection if you require minimal traffic loss during a reboot. Backup liveness detection helps achieve sub-second traffic loss during an MC-LAG reboot.

```
[edit protocols]
user@switch# set iccp peer peer-ip-address backup-liveness-detection backup-peer-ip ip-address
```

For example:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 backup-liveness-detection backup-peer-ip 10.207.64.232
```

- d. Configure the minimum interval at which the switch must receive a reply from the other switch with which it has established a Bidirectional Forwarding Detection (BFD) session.



**NOTE:** Configuring the minimum receive interval is required to enable BFD.

```
[edit protocols]
user@switch# set iccp peer peer-ip-address liveness-detection minimum-receive-interval seconds
```

For example:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 liveness-detection minimum-receive-interval 60
```

- e. Configure the minimum transmit interval during which a switch must receive a reply from a switch with which it has established a BFD session.

```
[edit protocols]
user@switch# set iccp peer peer-ip-address liveness-detection transmit-interval minimum-interval seconds
```

For example:

```
[edit protocols]
user@switch# set iccp peer 3.3.3.2 liveness-detection transmit-interval minimum-interval 60
```

8. Configure a multichassis protection link between the switches.

```
[edit]
```



---

```
user@switch# set multi-chassis multi-chassis-protection peer-ip-address interface
interface-name
```

For example:

```
[edit protocols]
user@switch# set multi-chassis multi-chassis-protection 3.3.3.1 interface ae0
```

9. Enable RSTP globally on all interfaces.

```
[edit]
user@switch# set protocols rstp interface all mode point-to-point
```

10. Disable RSTP on the ICL-PL interfaces on both switches.

```
[edit]
user@switch# set protocols rstp interface interface-name disable
```

For example:

```
[edit]
user@switch# set protocols rstp interface ae0.0 disable
```

11. Configure the MC-LAG interfaces as edge ports on both switches.

```
set protocols rstp interface interface-name
```

For example:

```
[edit]
user@switch# set protocols rstp interface ae1.0
```

12. Enable BPDU block on all interfaces except for the ICL-PL interfaces on both switches.

```
[edit]
user@switch# set protocols rstp bpdu-block-on-edge
```

For example:

```
[edit]
user@switch# set protocols rstp bpdu-block-on-edge
```

#### Related Documentation

- [Advantages of Using Multichassis Link Aggregation Groups on page 1](#)
- [Understanding Multichassis Link Aggregation on page 1](#)
- [Example: Configuring Multichassis Link Aggregation on page 11](#)

