JUNIPER NETWORKS

Network Configuration Example

Configuring Multichassis Link Aggregation on a QFX Series Switch

Release
12.3

Published: 2012-11-14

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

Network Configuration Example Configuring Multichassis Link Aggregation on a QFX Series Switch
Release 12.3

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

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 6
- Configuring Multichassis Link Aggregation on page 21

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 topologies changes. 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 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). In Junos OS Release 12.2, MC-LAGs support only Layer 2 features. Junos OS Release 12.2 does not support Layer 3 features.

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 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 the Interchassis Control Protocol (ICCP) to exchange control information and coordinate with each other to ensure that data traffic is forwarded properly.

The Link Aggregation Contol 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 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

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 between MC-LAG peers. Additionally, enabling backup liveness detection allows the peers to detect each other through the keepalive link in the event of a split-brain state.

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 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.

NOTE: The en dash means that this 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.
Down	Down or Up	Inactive	No change in LACP system ID.
Up	Down	-	THe 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. The split-brain state is automatically recovered when the ICCP adjacency comes up between the MC-LAG peers.

Multichassis Link Protection

Mutlichassis 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.

Layer 2 Features Supported

The following Layer 2 features are supported:

NOTE: **Layer 3 features are not supported at this time.**

- 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 happens 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.
- 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.

Spanning Tree Protocol (STP) Guidelines

- Enable STP globally.

  STP might detect local mis-wiring 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 the 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 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 toward the MC-AE interface ensures that a split-horizon for 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.

MC-LAG Upgrade Guidelines

Upgrade the MC-LAG peers according to the following guidelines. See Upgrading Software on a QFX3500 Switch for exact details on how to perform a software upgrade on a QFX3500 switch.

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

```
user@switch> show interfaces mc-ae id 1
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
```

- Upgrade node1 of the MC-LAG.

  Node 1 being upgraded gets rebooted and all traffic is sent across the available LAG interfaces of node 2, which is still UP. Any traffic loss is attributed to how quickly the neighbor devices detect the link loss and rehash the flows of the LAG.

- Verify that node1 is running the software you just installed. Issue the **show version** command.
- Make sure that both nodes of the MC-LAG (node 1 and node 2) are in active-active state after the reboot of node 1.
- Upgrade node 2 of the MC-LAG.

Node2 being upgraded gets rebooted and all traffic is sent across the available LAG interfaces of node 1, which is still UP. Any traffic loss is attributed to how quickly the neighbor devices detect the link loss and rehash the flows of the LAG.

- Verify that node 2 is running the software you just installed. Issue the **show version** command.
- Make sure that both nodes of the MC-LAG (node 1 and node 2) are in active-active state after the node 2 reboot.

Related Documentation

- Configuring Link Aggregation
- Example: Configuring Link Aggregation Between a QFX Series Product and an Aggregation Switch
- Example: Configuring Multichassis Link Aggregation on page 6
- Verifying the Status of a LAG Interface
- Junos® OS Network Interfaces

Example: Configuring Multichassis Link Aggregation

Multichassis link aggregation groups (MC-LAGs) enable a client device to form a logical LAG interface between two QFX3500 switches. An MC-LAG provides redundancy and load balancing between the two QFX3500 switches, multihoming support, and a loop-free Layer 2 network without running Spanning Tree Protocol (STP). MC-LAGs support only Layer 2 forwarding on Junos OS Release 12.2.

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 QFX3500 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 QFX3500 switches. Each of the QFX3500 switches has one or more physical links connected to a single client device. The QFX3500 switches coordinate with each other to ensure that data traffic is forwarded properly.

Requirements

This example uses the following hardware and software components:

- Junos OS Release 12.2 or later for the QFX Series
- Two QFX3500 switches

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

- Configure aggregated Ethernet interfaces on a QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 switches hosting an MC-LAG. The two QFX3500 switches are connected to a server. Figure 1 on page 7 shows the topology of this example.

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

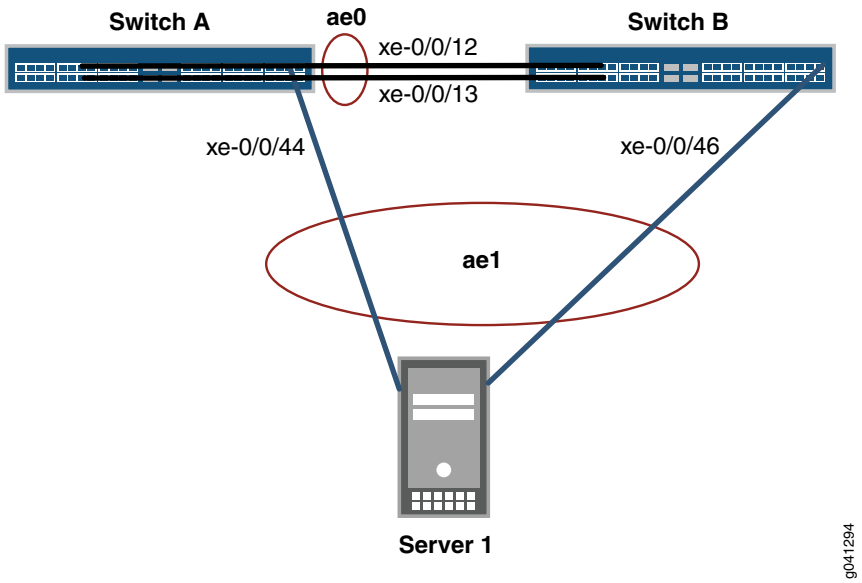

Table 2 on page 8 details the topology used in this configuration example.

Table 2: Components of the Topology for Configuring a Multichassis LAG Between Two QFX3500 Switches

Hostname	Base Hardware	Multichassis Link Aggregation Group
Switch A Switch B	QFX3500 switch QFX3500 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**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. .

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 local-ip-addr 3.3.3.2
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 QFX3500 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
```

5. (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
```

6. 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.

1. 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
   ```

2. 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
   ```

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
```

Step-by-Step Procedure

To enabled 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.

```
[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.

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

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;
    }
  }
  vlans {
    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;
      }
    }
    vlans {
      v100 {
        vlan-id 100;
      }
      v500 {
        vlan-id 500;
        l3-interface vlan.500;
      }
    }
```

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

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 and Switch B

Purpose Verify that MAC learning is working on Switch A and 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
  v222              *                 Flood          - All-members
  v222              00:00:5e:00:01:01 Static         - Router
  v222              00:10:94:00:00:05 Learn(L)       33 ae0.0 (MCAE)
  v222              84:18:88:df:ac:ae Learn(R)        0 ae2.0
```

Meaning The output shows four learned MAC addresses entries.

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 on page 1
- Configuring Multichassis Link Aggregation on page 21

Configuring Multichassis Link Aggregation

Multichassis link aggregation groups (MC-LAGs) enable a client device to form a logical LAG interface between two QFX3500 switches. An MC-LAG provides redundancy and load balancing between the two QFX3500 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 QFX3500 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 QFX3500 switches. Each of the QFX3500 switches has one or more physical links connected to a single client device. The QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 switch hosting the MC-LAG:

   a. Configure the local IP address to be used by all QFX3500 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 QFX3500 switch and the time during which an ICCP connection must succeed between the QFX3500 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 QFX3500 switch 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 QFX3500 switch must receive a reply from the other QFX3500 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 iccppeer 3.3.3.2 liveness-detection minimum-receive-interval 60
```

e. Configure the minimum transmit interval during which a QFX3500 switch must receive a reply from a QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 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 QFX3500 switchs.

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

    For example:

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

Related Documentation

- Understanding Multichassis Link Aggregation on page 1
- Example: Configuring Multichassis Link Aggregation on page 6