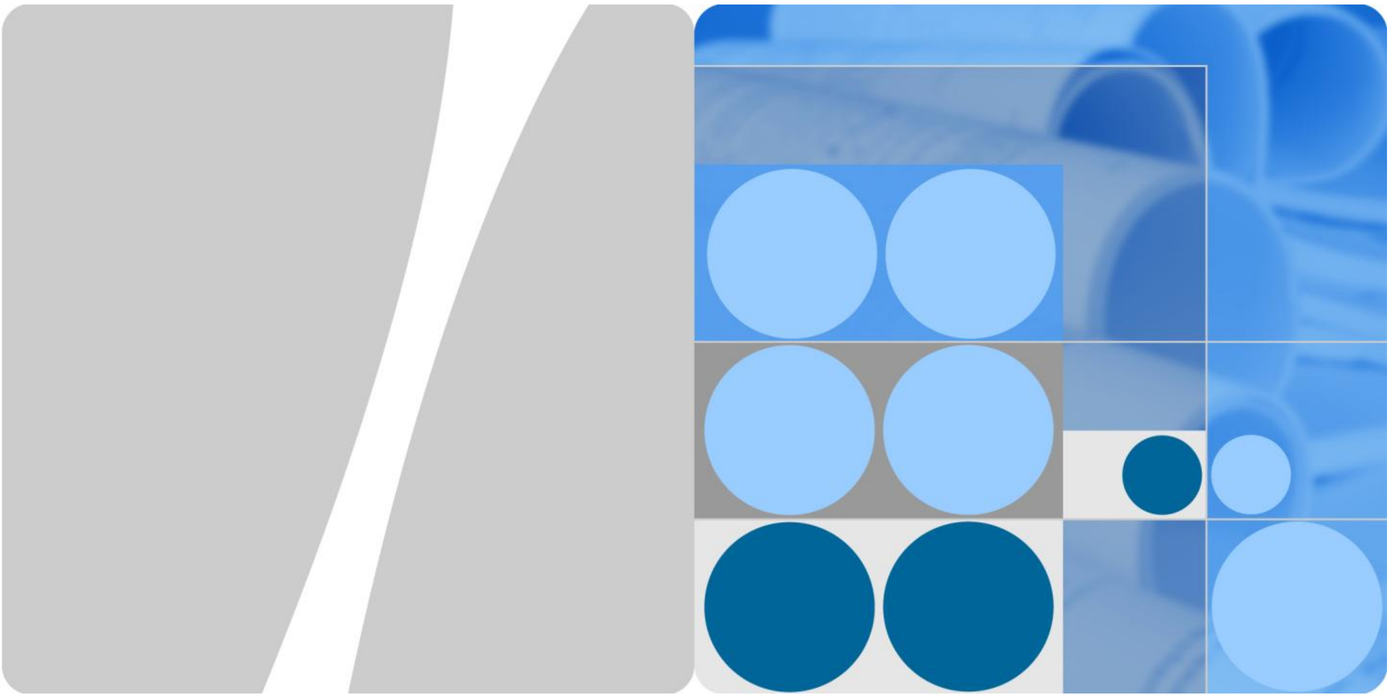


Huawei E9000 Server
V100R001

Network Technology White Paper

Issue	**03**
Date	**2017-05-11**

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions

HUAWEI and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://www.huawei.com

Email: support@huawei.com

Contents

1 About This Document

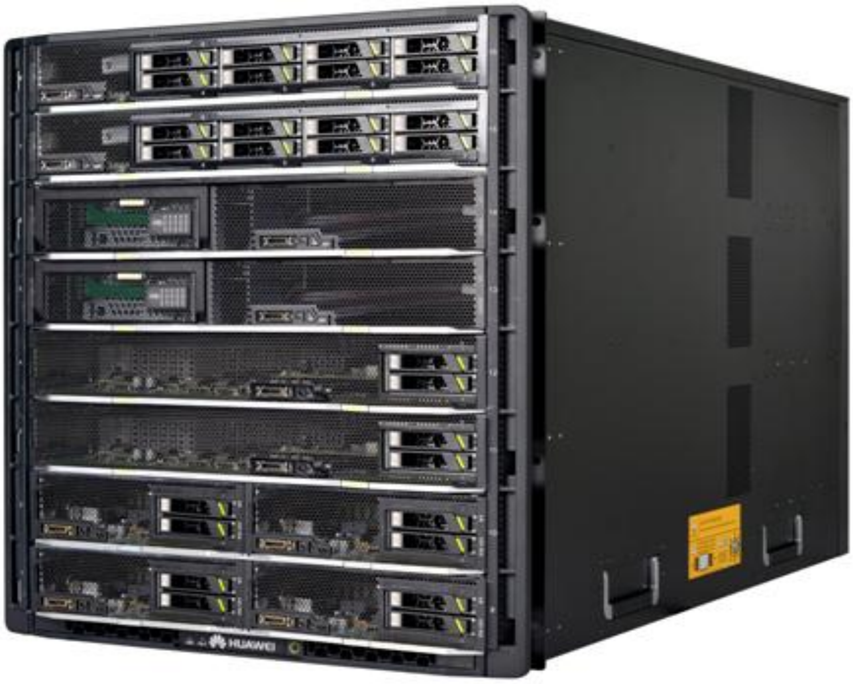
This document describes network technologies used by Huawei E9000 servers and typical networking and configurations of E9000 servers when connected to switches of different vendors in various application scenarios. This document helps readers to quickly understand E9000 network features and improve deployment efficiency.

The network features and networking described in this document are for reference only. For details about E9000 network features, see the *Switch Module White Paper*, *Configuration Guide*, and *Command Reference* of the E9000 server.

2 E9000 Overview

The Huawei E9000 converged architecture blade server (E9000 for short) is a new-generation powerful infrastructure platform that integrates computing, storage, switching, and management. It delivers high computing platform availability, computing density, and energy efficiency, optimal backplane bandwidth, low network latency, and intelligent management and control. It also allows elastic configuration and flexible expansion of computing and storage resources and application acceleration. An E9000 can be configured with eight full-width blades or 16 half-width blades. It allows mixed configuration of half-width and full-width blades to meet different service requirements. Each blade provides two or four processors and 48 DIMM slots, supporting a maximum of 3 TB memory capacity.

Figure 2-1 E9000 chassis appearance



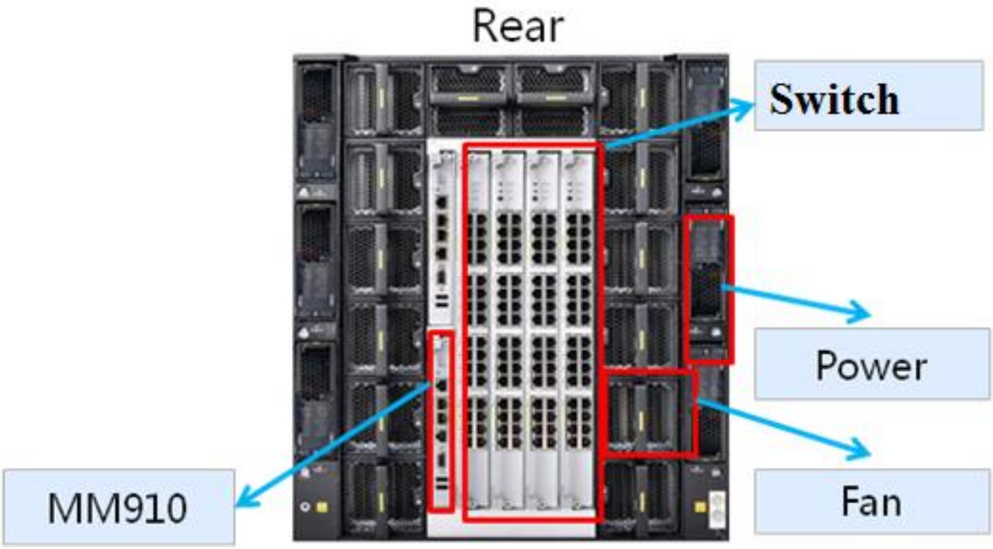
The E9000 supports flexible computing and I/O expansion. The computing blades support Intel processors of the next three generations. The E9000 supports a wide range of network ports, including ports from the mainstream GE/10GE to 40GE and to the future 100GE, as well as Fibre Channel (FC) ports and InifiBand (IB) ports. The E9000 houses four switch modules in the rear of the chassis to support Ethernet, FC, and IB switching networks. The switch modules provide high data switching capability and rich network features. Each computing blade provides standard PCIe slots, and a full-width blade supports a maximum of four PCIe cards.

2.1 E9000 Server Network

2.1.1 CX Switch Modules

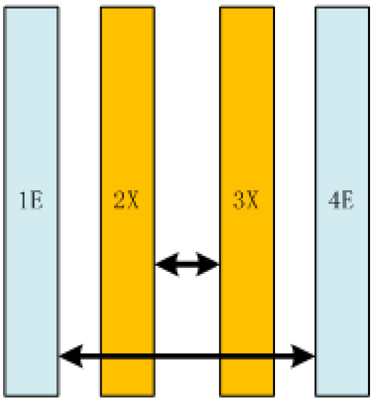
The E9000 provides four slots in the rear of the chassis for installing pass-through and switch modules that support GE, 10GE, and 40 GE ports, 8 Gbit/s and 16 Gbit/s FC ports, and 40 Gbit/s and 56 Gbit/s IB ports, as well as evolution toward 100 Gbit/s Ethernet and IB ports. The switch modules provide powerful data switching capabilities.

Figure 2-2 Installation position of CX switch modules



The four switch module slots are numbered 1E, 2X, 3X, and 4E from left to right. The backplane provides eight pairs of links to connect switch modules in slots 1E and 4E and switch modules in slots 2X and 3X. The switch modules can be stacked or cascaded by links through 10GE or 40GE ports. Figure 2-3 shows installation slots of the switch modules.

Figure 2-3 Slot names and connections between switch modules over the backplane



If switches in slots 1E and 4E need to be stacked and the Eth-Trunk interface is enabled (by default) to forward local traffic preferentially, the ports on the backplane can provide the bandwidth required for stacking. No stacking cable is required.

NOTE

The backplane does not provide links for interconnecting the CX915 and CX111 switch modules. If these two modules need to be stacked, optical or electric cables are required to connect the 10GE ports on the panel.

2.1.1.1 Mezzanine Cards and Switch Modules

The E9000 supports switch modules capable of Ethernet, FC, and IB switching. These switch modules can be used with different mezzanine cards to build different types of networks with multiple planes at different rates. Table 2-1 lists the switch modules and compatible mezzanine cards.

Table 2-1 Mapping between mezzanine cards and switch modules

Switch Module Type	Switch Module		Mezzanine Card	
GE switch module	CX110	Downlink: 16 x (2 x GE)+2 x 40GE Uplink: 12 x GE + 4 x 10GE	MZ110	4 x GE
	CX111	Downlink: 16 x (2 x GE) Uplink: 12 x GE + 4 x 10GE	MZ110	4 x GE
Converged switch module	CX310	Downlink: 16 x (2 x 10GE) + 40GE Uplink: 16 x 10GE	MZ510	2 x 10GE converged network adapter (CNA)
			MZ512	2 x (2 x 10GE) CNA
	CX311	Downlink: 16 x (2 x 10GE) + 40GE Uplink: 16 x 10GE + 8 x 8G FC	MZ510	2 x 10GE CNA
			MZ512	2 x (2 x 10GE)

				CNA
	CX320	Downlink: 16 x (2 x 10GE) + 40GE Uplink: 8 x 10GE + 2 x 40GE + 2 x (4 x 8G FC/4 x 10GE)	MZ510	2 x 10GE CNA
			MZ512	2 x (2 x 10GE) CNA
	CX710	Downlink: 16 x 40GE Uplink: 8 x 40GE	MZ710	2*40GE
Multi-plane switch module	CX911	Ethernet switching plane: Downlink: 16 x (2 x 10GE) + 2 x 40GE Uplink: 16 x 10GE FC switching plane (QLogic): Downlink: 16 x 8G FC Uplink: 8 x 8G FC	MZ910	2 x 10GE + 2 x 8G FC/10G Fibre Channel over Ethernet (FCoE)
	CX912	Ethernet switching plane: Downlink: 16 x (2 x 10GE) + 2 x 40GE Uplink: 16 x 10GE FC switching plane (Brocade): Downlink: 16 x 8G FC Uplink: 8 x 8G FC	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
	CX915	Ethernet switching plane: Downlink: 16 x (2 x GE) Uplink: 12 x GE + 4 x 10GE FC switching plane (QLogic): Downlink: 16 x 8G FC Uplink: 8 x 8G FC	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
	CX920	10GE switching plane: Downlink: 16 x 10GE + 1 x 40GE Uplink: 8 x 10GE 40GE switching plane: Downlink: 16 x 40GE + 2 x 40GE Uplink: 8 x 40GE	MZ710	2*40GE
FC switch module	CX210	FC switching plane (Brocade): Downlink: 16 x 8G FC Uplink: 8 x 8G FC	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
InfiniBand switch module	CX611	QDR/FDR InfiniBand switch module: Downlink: 16 x 4X QDR/FDR	MZ610	2 x 40G QDR
			MZ611	2 x 56G FDR

		Uplink: 18 x QDR/FDR QSFP+		
Pass-through module	CX116	Downlink: 16 x (2 x GE) Uplink: 16 x (2 x GE)	MZ110	4 x GE
	CX317	Downlink: 16 x (2 x 10GE) Uplink: 16 x (2 x 10GE)	MZ510	2 x 10GE CNA
			MZ512	2 x (2 x 10GE) CNA
	CX318	Downlink: 16 x (2 x 10GE) Uplink: 16 x (2 x 10GE)	MZ510	2 x 10GE CNA
			MZ512	2 x (2 x 10GE) CNA
			MZ310	2*10GE
			MZ312	2 x (2 x 10GE)

NOTE

If the MZ910 is used with a CX210 switch module, the two 10GE ports cannot be used. (The CX210 does not provide an Ethernet switching plane).

If the MZ910 is used with a switch module that provides the QLogic FC switching plane, ports in slots 1 to 12 are FCoE ports and ports in slots 13 to 16 are FC ports. If the MZ910 is used with a switch module that provides the Brocade FC switching plane, ports in slots 1 to 16 are all FC ports.

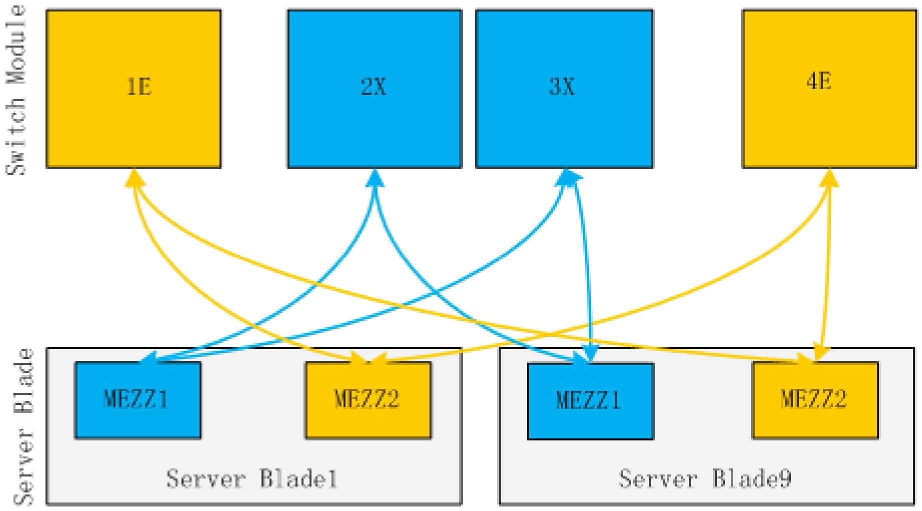
If the MZ910 is used with a CX915 switch module, the rate of 10GE ports is automatically reduced to fit for GE ports.

2.1.1.2 Connections Between Mezz Cards and Switch Modules

Each half-width blade in the E9000 provides one or two Mezz cards, and each full-width blade in the E9000 provides one, two, or four Mezz cards.

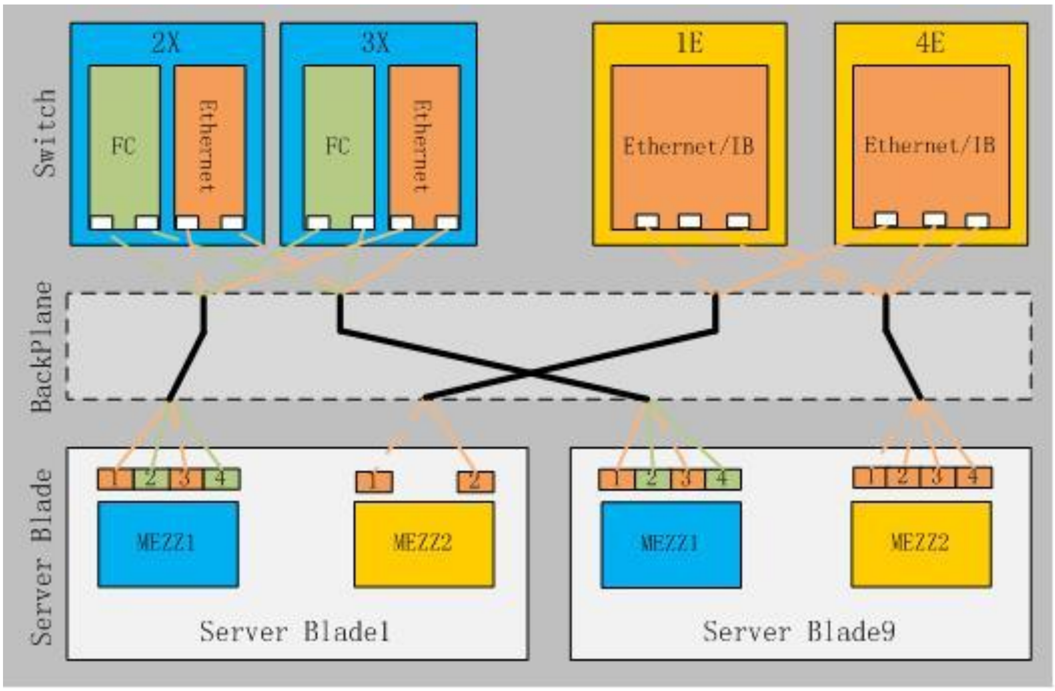
A Mezz card connects a blade and a switch module in a slot at the rear of the chassis through the backplane. The type of network (Ethernet, FC, and InifiBand) connected to the blade and the number of ports vary with the Mezz card, which must be matched with the switch module. Figure 2-4 shows the connections between Mezz cards and switch modules.

Figure 2-4 Connections between Mezz cards and switch modules



Mezz 1 in a half-width blade connects to I/O modules in slots 2X and 3X, and Mezz 2 in a half-width blade connects to switch modules in slots 1E and 4E. Mezz 1 and Mezz 3 (optional) in full-width blades connect to switch modules in slots 2X and 3X, and Mezz 2 and Mezz 4 (optional) in full-width blades connect to switch modules in slots 1E and 4E. Figure 2-5 shows the connections between switch modules and Mezz card ports.

Figure 2-5 Connections between switch modules and Mezz card ports

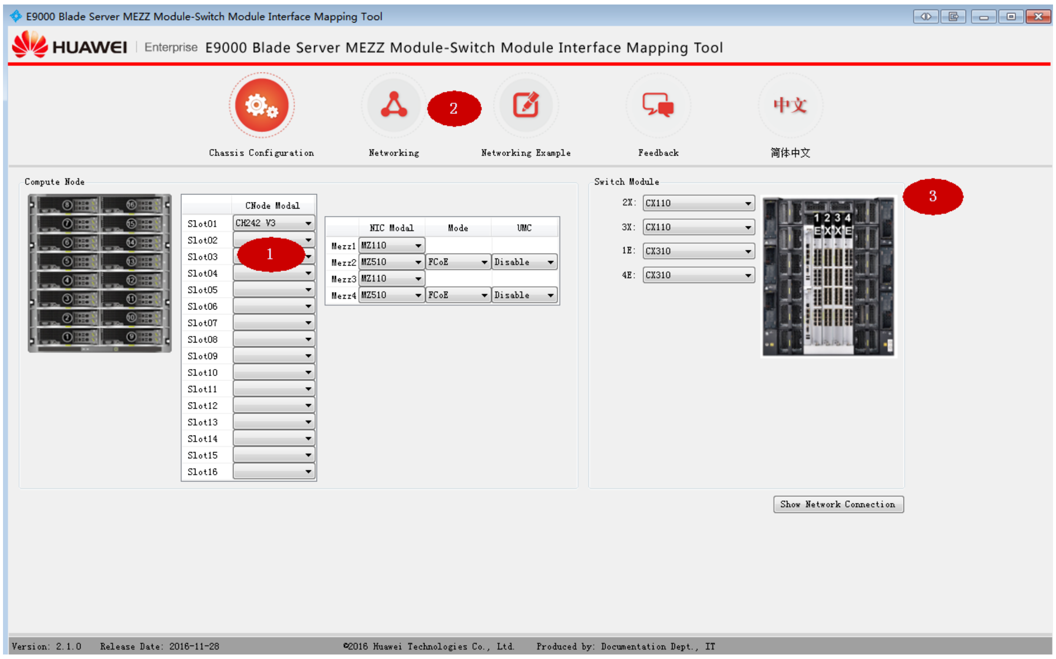


A multiple-plane Mezz card can be in the position of Mezz 1 or Mezz 2, connected to switch modules in slots 2X and 3X or in slots 1E and 4E. In addition to multiple planes, a multiple-plane Mezz card provides two or four ports. A 4-port Mezz card is connected to two ports of each switch module.

2.1.1.3 Networking Assistant Tool

A networking assistant tool is provided to help users to query the mapping between the ports on Mezz cards and switch modules and determine the network interface card (NIC), such as, ethx or vmnicx, to be used on the host operating system (OS). Figure 2-6 shows the operation interface of the tool.

Figure 2-6 Networking assistant tool



	CNode Modal
Slot01	CH242 V3
Slot02	
Slot03	
Slot04	
Slot05	
Slot06	
Slot07	
Slot08	
Slot09	
Slot10	
Slot11	
Slot12	
Slot13	
Slot14	
Slot15	
Slot16	

	NIC Modal	Mode	UMC
Mezz1	MZ110		
Mezz2	MZ510	FCoE	Disable
Mezz3	MZ110		
Mezz4	MZ510	FCoE	Disable

2X: CX110
3X: CX110
1E: CX310
4E: CX310

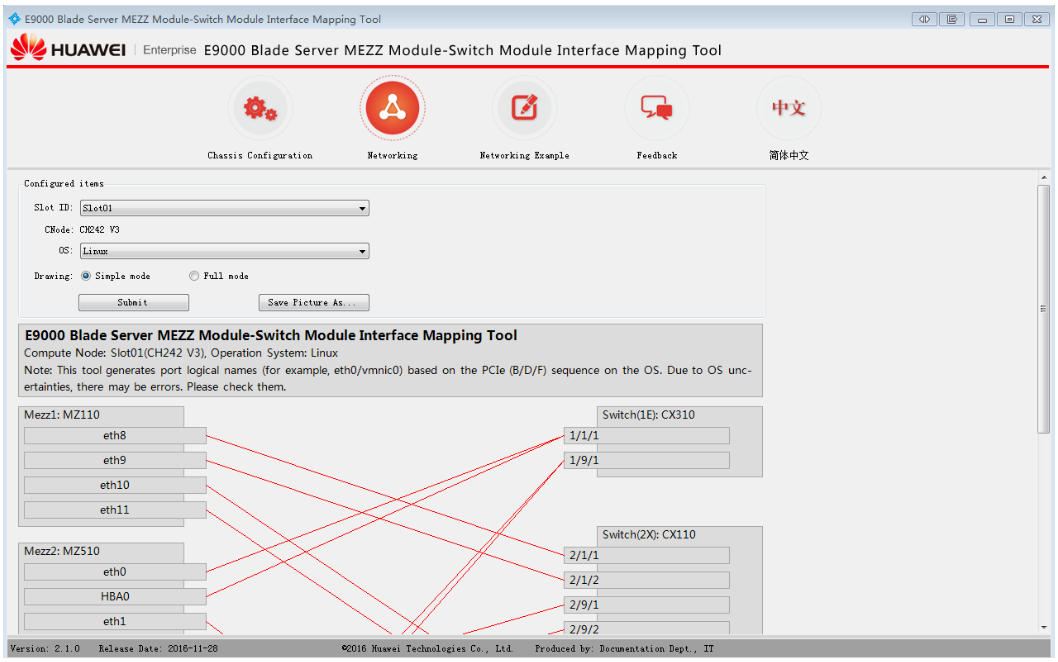
Version: 2.1.0 Release Date: 2016-11-28 ©2016 Huawei Technologies Co., Ltd. Produced by: Documentation Dept., IT

Use the tool as follows:

1. Choose the blade type and the corresponding Mezz card type. If CNAs are used, configure the daughter board attributes.
2. Choose the switch module type.
3. Click **Show Network Connection**.

The mapping between the ports on CNAs and switch modules is displayed as shown in Figure 2-7.

Figure 2-7 Network connections



E9000 Blade Server MEZZ Module-Switch Module Interface Mapping Tool
HUAWEI | Enterprise E9000 Blade Server MEZZ Module-Switch Module Interface Mapping Tool
Chassis Configuration
Networking
Networking Example
Feedback
中文
简体中文

Configured items

Slot ID: Slot01

CNode: CH242 V3

OS: Linux

Drawing: ● Simple mode ○ Full mode

Submit
Save Picture As...

E9000 Blade Server MEZZ Module-Switch Module Interface Mapping Tool

Compute Node: Slot01(CH242 V3), Operation System: Linux

Note: This tool generates port logical names (for example, eth0/vmnic0) based on the PCIe (B/D/F) sequence on the OS. Due to OS uncertainties, there may be errors. Please check them.

Version: 2.1.0 Release Date: 2016-11-28 ©2016 Huawei Technologies Co., Ltd. Produced by: Documentation Dept., IT

3 Network Technologies

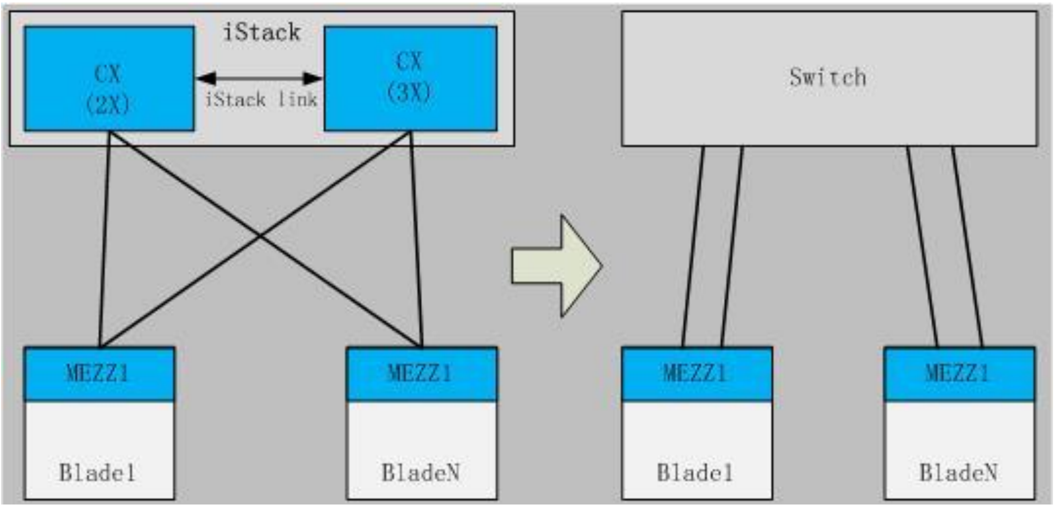
This section describes common technologies used for networking of the E9000 and provides some configuration instances. The networking technologies include iStack, Link Aggregation Control Protocol (LACP), NIC teaming, Data Center Bridging (DCB), FCoE, Smart Link, and Monitor Link.

3.1 iStack

3.1.1 iStack Overview

An E9000 chassis can house two or four switch modules, which are similar to two or four conventional switches. A Mezz card on the E9000 provides two or four ports to connect to two switch modules separately. Users need to log in to the two switch modules over SSH or a serial port to perform configuration, which increases O&M workloads. The two switch modules need to be configured with reliability features separately to ensure network reliability. For example, if an uplink port of a switch module is faulty, the fault needs to be sent to the module bound with the NIC on the OS using technologies, such as Monitor Link, to trigger a link switchover. iStack allows multiple switch modules to be virtualized and stacked as one logical device without changing the physical network topology. This technology simplifies network structure and network protocol deployment, and improves network reliability and manageability. Figure 3-1 illustrates the iStack technology.

Figure 3-1 Logical structure of a stack system



3.1.2 Technical Advantages

3.1.2.1 Simplified Configuration and Management

iStack allows multiple physical devices to be presented as one logical device. Users can log in from any member device and uniformly configure and manage all the member devices.

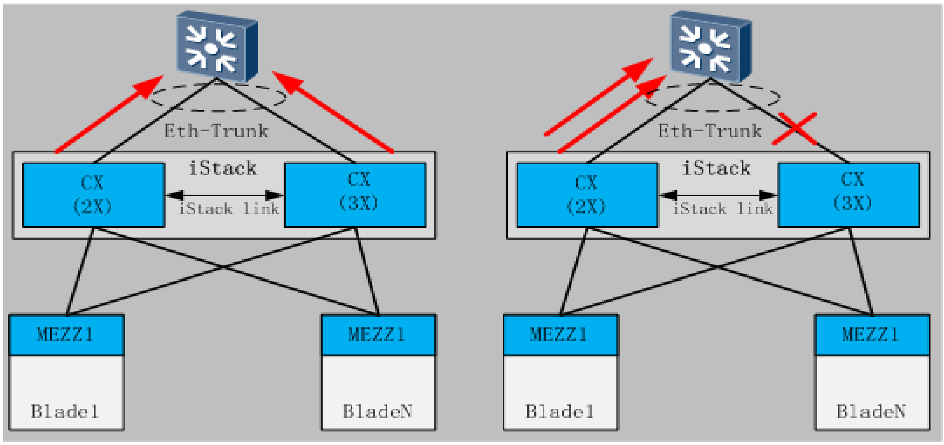
3.1.2.2 Control Planes in 1+1 Redundancy

After stacked, the switch modules set up control panels in 1+1 backup. Normally, the master switch processes services, and the standby switch functions as backup of the master switch and synchronizes data with it. If the master switch fails, the standby switch takes over service processing immediately and a new standby switch is selected from the slave switches.

3.1.2.3 Link Backup

iStack supports link aggregation among multiple switch modules. That is, uplink ports of multiple switch modules can be added to an Eth-Trunk group to improve the uplink bandwidth and network reliability. With iStack, users do not need to configure fault association technologies.

Figure 3-2 Link backup of a stack system



3.1.3 Basic Concepts

3.1.3.1 Role

Switch modules that have joined a stack are member switches. Each member switch in a stack plays one of the following roles based on their functions:

1. **Master switch**: manages the entire stack. There is only one master switch in a stack.
2. **Standby switch**: a backup of the master switch. There is only one standby switch in a stack.
3. **Slave switch**: any switch except the master switch in a stack. The standby switch is selected from slave switches.

3.1.3.2 Stack Domain

A stack domain is a set of switches connected using stack links. Only switches that share the same domain can be stacked together.

3.1.3.3 Stack Member ID

A member ID uniquely identifies a switch in a stack. The member ID is the slot number of the switch module. By default, the stack member IDs of switch modules in slots 1E, 2X, 3X, and 4E are 1, 2, 3, and 4 respectively. The stack member IDs can be modified by running the **stack** command.

3.1.3.4 Stack Priority

The stack priority determines the role of a member switch in a role election. A larger value indicates a higher priority and higher probability that the member switch is elected as the master switch.

3.1.3.5 Physical Stack Member Port

A physical stack member port is used for stack connection. It forwards service packets or stack protocol packets between member switches.

3.1.3.6 Stack Port

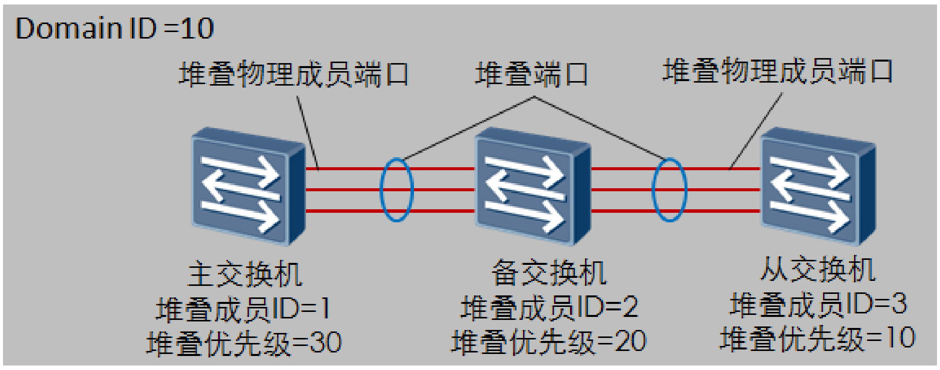
A stack port is a logical port dedicated for stacking and must be bound with a physical member port. Each member switch in a stack provides two stack ports: stack-port n/1 and stack-port n/2. "n" indicates the stack member ID of a member switch.

3.1.4 Basic Principles

3.1.4.1 Stack Setup

A CX switch module can directly connect to a 10GE/40GE link through the backplane or use a 10GE port on the panel as a physical stack port. The 10GE and 40GE ports cannot be used as the stack port at the same time. Multiple physical stack member ports can be bound to a stack port to improve the link bandwidth and reliability. The Eth-Trunk interface can be enabled to forward local traffic preferentially to reduce the requirements for stack port bandwidths. Figure 3-3 shows how a stack is set up.

Figure 3-3 Stack setup



A stack consists of multiple member switches, each of which has a specific role. When a stack is created, the member switches send stack competition packets to each other to select the master switch. After a master switch is selected, the remaining switches function as slave switches. The master switch is selected based on the following rules in sequence:

1. **Running status**: The switch that is in proper running state becomes the master switch.
2. **Stack priority**: The switch with the highest stack priority becomes the master switch.
3. **Software version**: The switch running the latest software version becomes the master switch.
4. **MAC address**: The switch with a smaller MAC address becomes the master switch.

The master switch collects stack member information, works out the stack topology, and synchronizes the topology information to all member switches. If the stack ID of a slave switch conflicts with an existing stack ID, the switch repeatedly restarts. If the master and slave switches use different software versions, the software version of the master switch will be synchronized to the slave switch, which then restarts and joins the stack.

The master switch selects a standby switch from slave switches as its backup. When the master switch fails, the standby switch takes over all services from the master switch. The master switch compares the following conditions of the member switches in sequence until a standby switch is selected:

1. **Stack priority**: The switch with the highest stack priority becomes the standby switch.

2. **MAC address**: The switch with a smaller MAC address becomes the standby switch.

Before a stack is set up, each switch is an independent entity and has its own IP address. Users need to manage the switches separately. After a stack is set up, the switches in the stack form a logical entity, and users can use a single IP address to manage and maintain the switches uniformly. The IP address and MAC address of a stack is the IP address and MAC address of the master switch when the stack is set up for the first time.

3.1.4.2 Removing Stack Members

A member switch leaves a stack after it is disconnected from the stack. The impact on the system varies with the role of the member switch that leaves a stack:

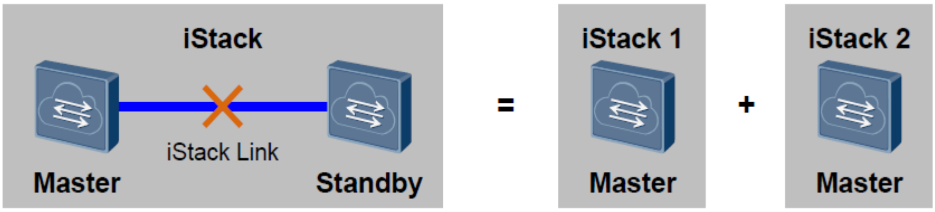
1. If the master switch leaves a stack, the standby switch becomes the new master switch, updates the stack topology, and specifies a new standby switch.

   If the standby switch leaves a stack, the master switch updates the stack topology and specifies a new standby switch.
2. If a slave switch leaves a stack, the master switch updates the stack topology.
3. If both the master and standby switches leave a stack, all slave switches restart and set up a new stack.

3.1.4.3 Stack Split

Stack split occurs when the stacking cable is faulty and a stack splits into multiple stacks. After a stack splits, multiple stacks with the same configuration may be generated. Conflicts between IP addresses and MAC addresses cause network faults. Figure 3-4 shows the stack split.

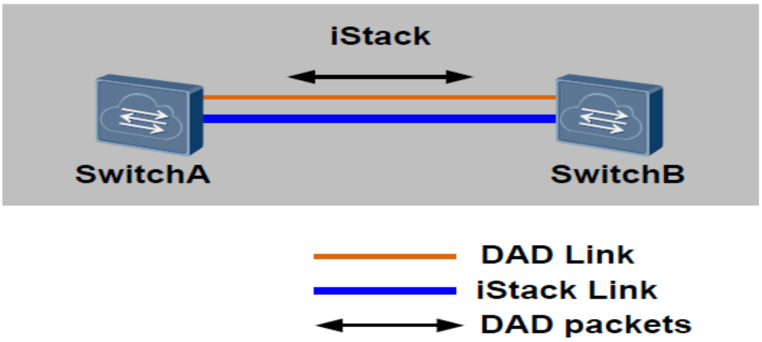
Figure 3-4 Stack split



3.1.4.4 DAD

Dual-active detection (DAD) is a protocol used to detect a stack split and dual-active situations, handle conflicts, and take recovery actions to minimize impact of a stack split on services. A DAD link directly connecting the stacked switches can detect dual-active switches, as shown in Figure 3-5.

Figure 3-5 DAD



After a stack splits, the stacks exchange DAD competition packets and compare information in the received DAD competition packet with local information based on the DAD competition rules. If the active stack is determined, switches in the stack remain in active state and continue forwarding service packets. The switches in another stack shut down all service ports except the reserved ones, enter the recovery state, and stop forwarding service packets. The active stack is determined based on the following DAD competition rules in sequence:

1. **Stack priority**: The switch with the highest stack priority becomes the master switch.
2. **MAC address**: The switch with a smaller MAC address becomes the master switch.

When the faulty stack links recover, stacks in recovery state restart and the shutdown ports change to the **Up** state. The entire stack system then recovers.

3.1.4.5 Fast Upgrade

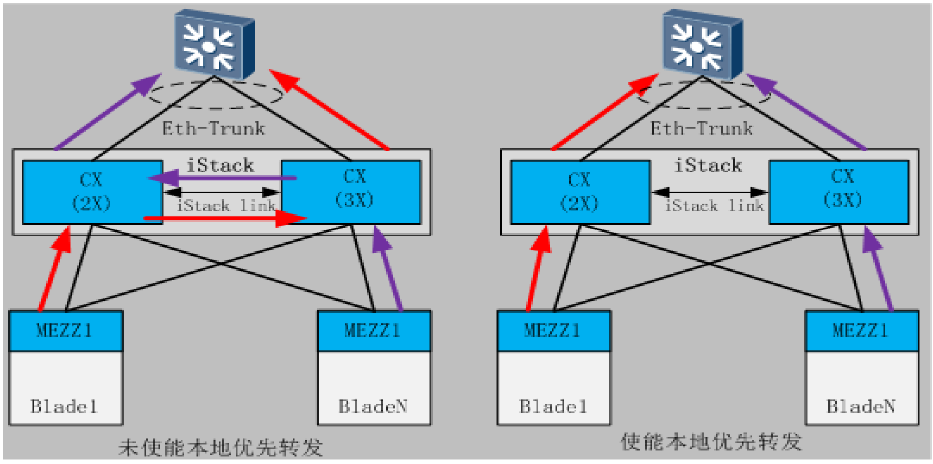
Fast upgrade allows the member switches in a stack to be upgraded without interrupting services. This feature minimizes the impact of upgrades on services. During a fast stack upgrade, the standby switch restarts with the new version first. The master switch forwards traffic in this period.

- If the upgrade fails, the standby switch restarts and rolls back to the source version.
- If the upgrade is successful, the standby switch becomes the new master switch and starts to forward data traffic. Then the original master switch restarts with the new version. After the original master switch is upgraded, it serves as the standby switch. The fast stack upgrade command is **stack upgrade fast**.

3.1.5 Local Preferential Forwarding

In a stack system, after an Eth-Trunk interface across switch modules is configured, traffic is sent through the link selected based on the configured routing. Because Eth-Trunk consists of ports on different switch modules, some traffic is forwarded among switch modules. Enabling the local forwarding for the Eth-Trunk interface can reduce the traffic forwarded among switch modules, as shown in Figure 3-6.

Figure 3-6 Local preferential forwarding

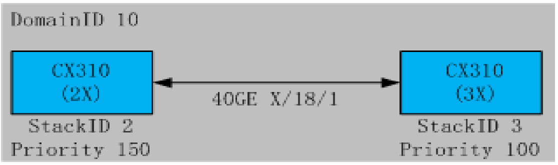


After local preferential forwarding is enabled, outgoing traffic is preferentially forwarded through the ports receiving the traffic. If the number of active links of a switch module in the Eth-Trunk interface is fewer than the minimum number of active links on the switch module, local preferential forwarding will be automatically disabled. Traffic will be sent through a link selected from the Eth-Trunk member interfaces. The local preferential forwarding function is enabled by default after an Eth-Trunk interface is created.

3.1.6 Configuration Instance

The CX310s in slots 2X and 3X are stacked using a 40GE link provided by the backplane, as shown in Figure 3-7.

Figure 3-7 CX310 stacking network



# Run the **reset saved-configuration** command to restore the default configuration of the switch module in slot 2X and restart the switch module.

```
<HUAWEI>reset saved-configuration
 The action will delete the saved configuration in the device.    The configuration
will be erased to reconfigure.Continue? [Y/N]:Y
Warning: Now clearing the configuration in the device.......
 begin synchronize configuration to SMM ...
 slot 2: upload configuration to SMM successfully.

 Info: Succeeded in clearing the configuration in the device.
 <HUAWEI>reboot fast
```

# Run the **reset saved-configuration** command to restore the default configuration of the switch module in slot 3X and restart the switch module.

```
<HUAWEI>reset saved-configuration
 The action will delete the saved configuration in the device.    The configuration
will be erased to reconfigure.Continue? [Y/N]:Y
Warning: Now clearing the configuration in the device.
 begin synchronize configuration to SMM ...
 slot 3: upload configuration to SMM successfully.

 Info: Succeeded in clearing the configuration in the device.
 <HUAWEI>reboot fast
```

# Set the default stack member ID to **2**, domain ID to **10**, and priority to **150** for the CX310 in slot 2X.

```
<HUAWEI> system-view
[~HUAWEI] sysname CX310_2
[*HUAWEI] commit
[~CX310_2] stack
[*CX310_2-stack] stack member priority 150
[*CX310_2-stack] stack member 2 domain 10
[*CX310_2-stack] quit
[*CX310_2] commit
```

# Add service port 40GE 2/18/1 of the CX310_2 to stack port 2/1.

```
[~CX310_2] interface 40GE 2/18/1
[*CX310_2-40GE2/18/1] port mode stack
[*CX310_2-40GE2/18/1] quit
[*CX310_2] commit
[~CX310_2] interface stack-port 2/1
[*CX310_2-Stack-Port2/1] port member-group interface 40GE 2/18/1
[*CX310_2-Stack-Port2/1] quit
[*CX310_2] commit
```

# Set the default stack member ID to **3**, domain ID to **10**, and priority to **100** for the CS310 in slot 3X.

```
<HUAWEI> system-view
[~HUAWEI] sysname CX310_3
[*HUAWEI] commit
[~CX310_3] stack
[*CX310_3-stack] stack priority 100
[*CX310_3-stack] stack member 3 domain 10
[*CX310_3-stack] quit
[*CX310_3] commit
```

# Add service port 40GE 3/18/1 of the CX310 in slot 3X to stack port 3/1.

```
[~CX310_3] interface 40GE 3/18/1
[*CX310_3-40GE3/18/1] port mode stack
[*CX310_3-40GE3/18/1] quit
[*CX310_3] commit
[~CX310_3] interface stack-port 3/1
[*CX310_3-Stack-Port3/1] port member-group interface 40GE 3/18/1
[*CX310_3-Stack-Port3/1] quit
[*CX310_3] commit
```

# Enable the 40GE ports interconnecting slots 2X and 3X.

```
[~CX310_2] interface 40GE 2/18/1
[*CX310_2-40GE2/18/1] undo shutdown
[*CX310_2-40GE2/18/1] quit
[*CX310_2] commit
[~CX310_2] quit
 <CX310_2> save
[~CX310_3] interface 40GE 3/18/1
[*CX310_3-40GE3/18/1] undo shutdown
[*CX310_3-40GE3/18/1] quit
[*CX310_3] commit
[~CX310_3] quit
 <CX310_3> save
```

NOTE

The switch modules in slots 2X and 3X must be configured in the same stack domain. After a low-priority switch module (in slot 3X) is configured, the standby switch restarts automatically and a stack system is created. If stack system configuration is correct, run the **save** command to save the configuration. (If the switch module in slot 2X is not the master switch for the first master competition, run the **reboot** command to restart the stack system. Then, the system will select the switch module in slot 2X as the master switch based on the priority.)

# Change the system name and view the stack system information.

```
[CX310_2] sysname CX310_C
[~CX310_2] commit
[~CX310_C] display stack
--------------------------------------------------------------------
MemberID  Role    MAC           Priority  Device Type  Bay/Chassis
--------------------------------------------------------------------
2        Master   0004-9f31-d540  150      CX310       2X/1
3        Standby  0004-9f62-1f80  100      CX310       3X/1
--------------------------------------------------------------------
[~CX310_C] quit
<CX310_C> save
```

3.2 LACP

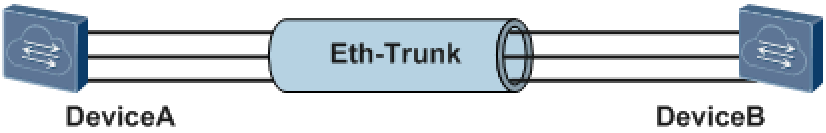
3.2.1 LACP Overview

LACP allows multiple physical ports to be bound as a logical interface to increase link bandwidths without upgrading hardware. In addition, the link backup mechanism of LACP provides higher link transmission reliability. LACP has the following advantages:

1. **Increased bandwidth**: The bandwidth of a link aggregation interface (LAI) is the sum of bandwidths of its member interfaces. The maximum number of aggregation members is 16.
2. **High reliability**: If an active link fails, traffic along the link is switched to another active links, improving the reliability of LAIs.
3. **Load balancing**: In a link aggregation group (LAG), traffic is evenly distributed among the active member links.

The LAI of CX switch modules is named Eth-Trunk. In an Eth-Trunk, all the member ports must be of the same type and use the default configuration. Figure 3-8 shows the LAI of CX switch modules.

Figure 3-8 LAI



3.2.2 Basic Concepts

3.2.2.1 Link Aggregation, LAG and LAI

Link aggregation allows multiple physical ports to be combined as a logical interface to increase the bandwidth and reliability. A LAG is a logical link consisting of multiple Ethernet links bound together. Each LAG corresponds to a unique logical interface, which is called an aggregation interface or Eth-Trunk interface.

3.2.2.2 Member Interfaces and Links

Physical ports that constitute an Eth-Trunk interface are member interfaces. A link corresponding to a member interface is called a member link.

3.2.2.3 Active/Inactive Interfaces and Links

The member interfaces in a LAP are classified into active and inactive interfaces. An active interface has data transmitted. An inactive interface has no data transmitted. Links connected to active interfaces are called active links, and links connected to inactive interfaces are called inactive links.

3.2.2.4 Upper Threshold of the Active Interface Quantity

The objective of setting the upper threshold of the active interface quantity is to improve network reliability while maintaining sufficient bandwidth. If the number of active links reaches the upper threshold, the number of active interfaces in an Eth-Trunk remains the same even if more member interfaces are added into the Eth-Trunk. The additional member links are set to **Down** and server as backup links.

For example, an Eth-Trunk has eight links, and each link provides a bandwidth of 1 Gbit/s. If the maximum bandwidth required is 5 Gbit/s, you can set the upper threshold to 5. As a result, the other three member links automatically enter the backup state to improve the network reliability.

3.2.2.5 Lower Threshold of the Active Interface Quantity

The objective of setting the lower threshold of the active interface quantity is to ensure the minimum bandwidth. When the number of active links is lower than the lower threshold, the Eth-Trunk interface goes **Down**.

For example, if each physical link provides a bandwidth of 1 Gbit/s and the minimum bandwidth required is 2 Gbit/s, you can set the lower threshold to 2 or a larger value.

3.2.2.6 Minimum Number of Local Active Links

Local preferential forwarding is enabled for the Eth-Trunk interface by default, which prevents traffic from being forwarded among switch modules. The minimum number of local active links is set to ensure the forwarding bandwidth. If the number of active links of each switch module in the Eth-Trunk is smaller than the minimum number of local active links, local preferential forwarding will be disabled automatically.

3.2.3 Link Aggregation Load Balancing Mode

To improve bandwidth utilization, traffic from the Eth-Trunk is sent through different physical member links to achieve load balancing. The CX switch modules support the following load-sharing modes:

1. **dst-ip**: load balancing based on the destination IP address. In this mode, the system obtains the specified four-bit value from each of the destination IP address and the TCP or UDP port number in outbound packets to perform the Exclusive-OR calculation, and then selects an outbound interface from the Eth-Trunk table based on the calculation result.

   dst-mac: load balancing based on the destination MAC address. In this mode, the system obtains the specified four-bit value from each of the destination MAC address, VLAN ID, Ethernet type, and inbound interface information to perform the Exclusive-OR calculation, and then selects an outbound interface from the Eth-Trunk table based on the calculation result.
2. **src-ip**: load balancing based on the source IP address. In this mode, the system obtains the specified four-bit value from each of the the source IP address and the TCP or UDP port number in inbound packets to perform the Exclusive-OR calculation, and then selects an outbound interface from the Eth-Trunk table based on the calculation result.
3. **src-mac**: load balancing based on the source MAC address. In this mode, the system obtains the specified four-bit value from each of the source MAC address, VLAN ID, Ethernet type, and inbound interface information to perform the Exclusive-OR calculation, and then selects an outbound interface from the Eth-Trunk table based on the calculation result.
4. **src-dst-ip**: load balancing based on the Exclusive-OR result of the source and destination IP address. In this mode, the system performs the Exclusive-OR calculation between the Exclusive-OR results of the dst-ip and src-ip modes, and then selects an outbound interface from the Eth-Trunk table based on the calculation result.
5. **src-dst-mac**: load balancing based on the Exclusive-OR result of the source MAC address and destination MAC address. In this mode, the system obtains the specified four-bit value from each of the source MAC address, destination MAC address, VLAN ID, Ethernet type, and inbound interface information to perform the Exclusive-OR calculation, and then selects an outbound interface from the Eth-Trunk table based on the calculation result.
6. **enhanced**: The system selects outbound interfaces for different packets based on an enhanced load-balancing profile.

The enhanced mode is the default load balancing mode. The load balancing modes defined in the profile vary with the packet type. Table 3-1 lists the profile.

Table 3-1 Default configuration of the enhanced mode

Inbound Packet Type	Default Load-balancing Mode	Configurable Load-balancing Mode	Remarks
IPv4 packets	src-ip+dst-ip+l4-src-port+l4-dst-port	**src-ip/dst-ip//l4-src-port/l4-dst-port/protocol**	The load-sharing mode varies with the packet type. It is irrelevant to
IPv6 packets	src-ip+dst-ip+l4-src-	**src-ip/dst-ip//l4-src-port/l4-**	

	port+l4-dst-port	**dst-port/protocol**	the packet forwarding process. The system identifies the type of the packet carried in an Ethernet frame. For example, if an IPv4 packet is identified, the load-balancing mode configured for IPv4 packets will be applied even if Layer 2 forwarding is required. If the packet is not an IPv4, IPv6, or MPLS packet, the system applies the L2-specific load-balancing modes (**src-mac**, **dst-mac**, **src-interface**, and **eth-type**).
MPLS packets	top-label+2nd-label	**top-label/2nd-label/dst-ip/src-ip**	
Other Layer 2 (L2) packets	src-mac+dst-mac	**src-mac/dst-mac/src-interface/eth-type**	
Trill packets	Ingress nodes: Use the inner **src-mac** and **dst-mac** modes for L2 packets. Use the **src-ip**, **dst-ip**, **l4-src-port**, and **l4-dst-port** modes for L3 packets.	**src-mac/dst-mac/src-ip/dst-ip/src-interface/l4-src-port/l4-dst-port/protocol/eth-type**	Only when **load-balance enhanced profile** *profile-name* is used, users can set load balancing for trill packets on transit and egress nodes.
	Transit/Egress nodes: Use the inner **src-mac** and **dst-mac** modes for L2 packets. Use the **src-ip**, **dst-ip**, **l4-src-port**, and **l4-dst-port** modes for L3 packets.	Cannot be configured.	

3.2.4 Link Aggregation Working Modes

3.2.4.1 Manual Link Aggregation

The Eth-Trunk is manually created, and member interfaces are manually added the Eth-Trunk. No LACP protocol is involved. In this mode, all active links work in load-sharing mode to forward data. If an active link is faulty, the remaining active links evenly share the traffic.

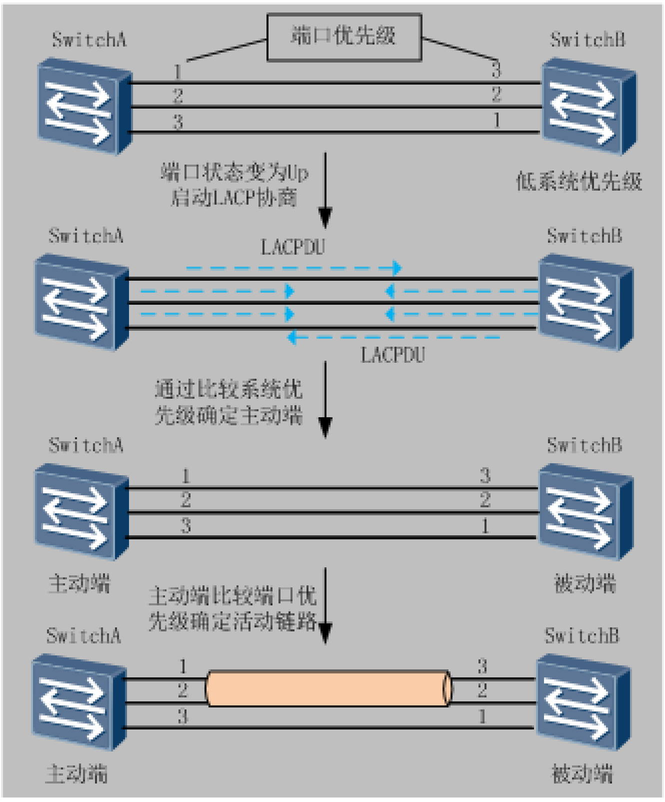
3.2.4.2 LACP Link Aggregation

The LACP protocol defined by Institute of Electrical and Electronics Engineers (IEEE) 802.3ad implements dynamic link aggregation and de-aggregation. LACP communicates with the peer ends using the link aggregation control protocol data units (LACPDUs). After member interfaces are added to the Eth-Trunk interface, these interfaces send LACPDUs to inform peer ends of their system priorities, MAC addresses, interface priorities, interface numbers, and operation keys (used to determine whether the peer ends are in the same LAG and interface bandwidths are the same). After receiving the information, the peer end compares the information with the information stored on its interface and selects the interfaces that can be aggregated. Then, the devices at both ends determine the active interfaces and links to be used. IEEE802.3ad defines the following two priorities:

1. **System LACP priority**: A smaller value indicates a higher priority. The end with a higher priority is the active end and selects the active port. The end with a lower priority is the passive end and uses the active links selected by the active end.
2. **Interface LACP priority**: indicates the priority of interfaces in the same Eth-Trunk. A smaller value indicates a higher priority. An interface with a higher priority increases its likelihood to be selected as an active interface.

After an interface is added to the Eth-Trunk, the interface status changes from **Down** to **Up** and LACP protocol negotiation starts. Figure 3-9 shows the LACP process.

Figure 3-9 LACP Link Aggregation



LACP has two modes: lacp static and lacp dynamic. They handle link negotiation failures in different ways. In lacp static mode, the Eth-Trunk interface becomes **Down** and cannot forward data after the LACP negotiation fails. In lacp dynamic mode, the Eth-Trunk interface becomes **Down** after the LACP negotiation fails, but member interfaces inherit Eth-Trunk VLAN attributes and change to **Indep** state to independently perform L2 data forwarding.

3.2.5 Comparison Between Huawei Eth-Trunk Interfaces and Cisco Port Channels

Huawei Eth-Trunk interfaces and Cisco port channels support the manual link aggregation and LACP modes. Table 3-2 lists the mode mappings.

Table 3-2 Comparison of working modes between Huawei Eth-Trunk interface and Cisco port channels

	Eth-Trunk	Port Channel
mode manual	mode manual	channel-group 1 mode on
lacp	mode lacp static	channel-group 1 mode active
	mode lacp static	channel-group 1 mode passive

Huawei Eth-Trunk determines the active and passive ends based on LACP priorities, that is, the end with a higher priority is the active end and the end with a lower priority is the passive end.

3.3 NIC Teaming

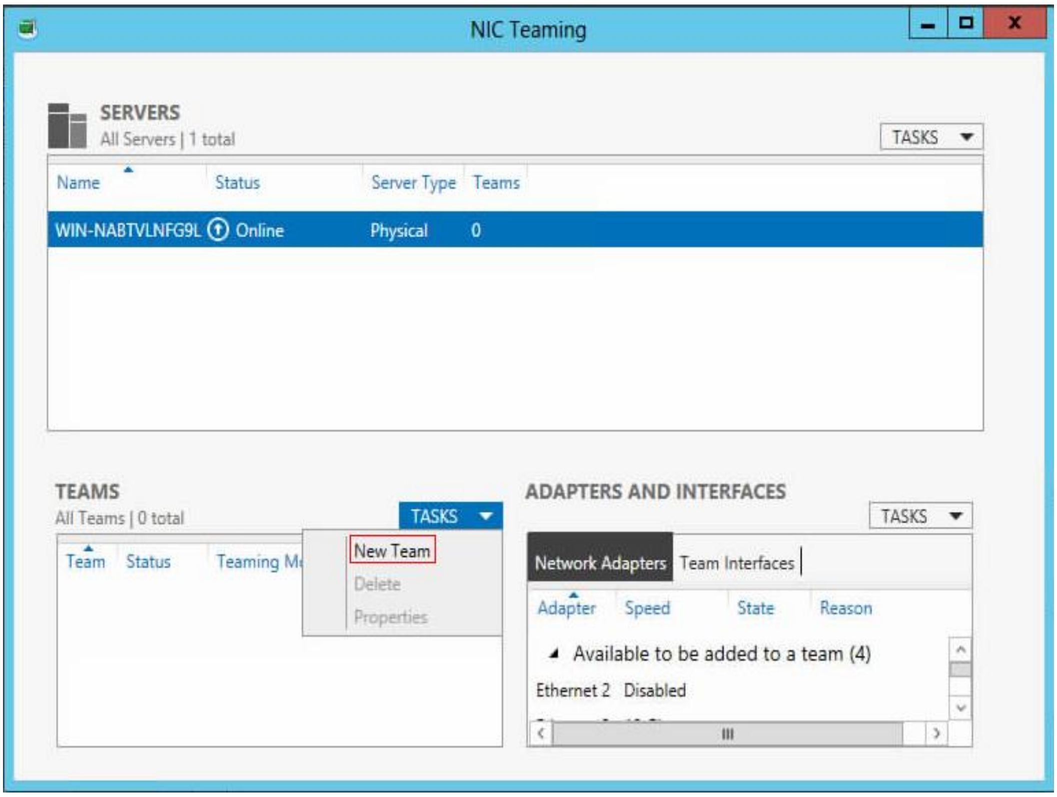
3.3.1 Overview

NIC Teaming allows multiple physical NICs on an Ethernet server to be bound as a virtual NIC using software. This server has only one NIC presented to the external network and only one network connection for any application in the network. After NICs are bonded, data can be sent in load-sharing or active/standby mode. If one link fails, a traffic switchover or active/standby switchover is performed to ensure server network reliability.

3.3.2 NIC Teaming in Windows

NIC Teaming, also known as load balancing and failover (LBFO), allows multiple network adapters on a server to be grouped as a team for bandwidth aggregation and traffic failover to maintain connectivity in the event of a network component failure. It is supported by Windows Server 2012 and later versions, including Server Core and Full GUI. Versions earlier than Windows Server 2012 usually adopts NIC binding tools built in NICs. Intel, Emulex, and Broadcom all provide related GUI configuration tools.

Figure 3-10 Creating a NIC Team



NIC Teaming

SERVERS
All Servers | 1 total

Name	Status	Server Type	Teams
WIN-NABTVLNFG9L	Online	Physical	0

TEAMS
All Teams | 0 total

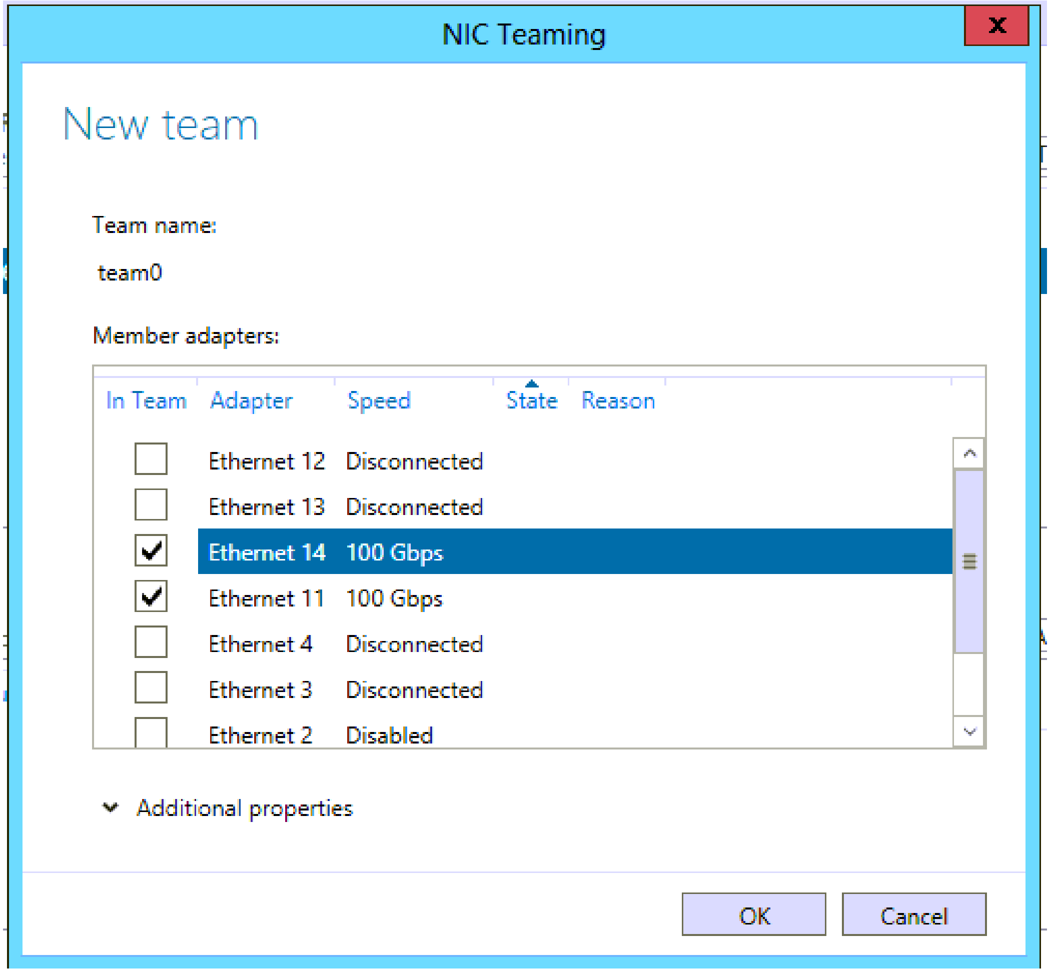
Team	Status	Teaming M

ADAPTERS AND INTERFACES

Adapter	Speed	State	Reason
Available to be added to a team (4)			
Ethernet 2	Disabled		

On the GUI, enter a team name, select NICs to be bound, and set the working mode of NIC teaming.

Figure 3-11 Selecting NICs to be bound



Team name:

team0

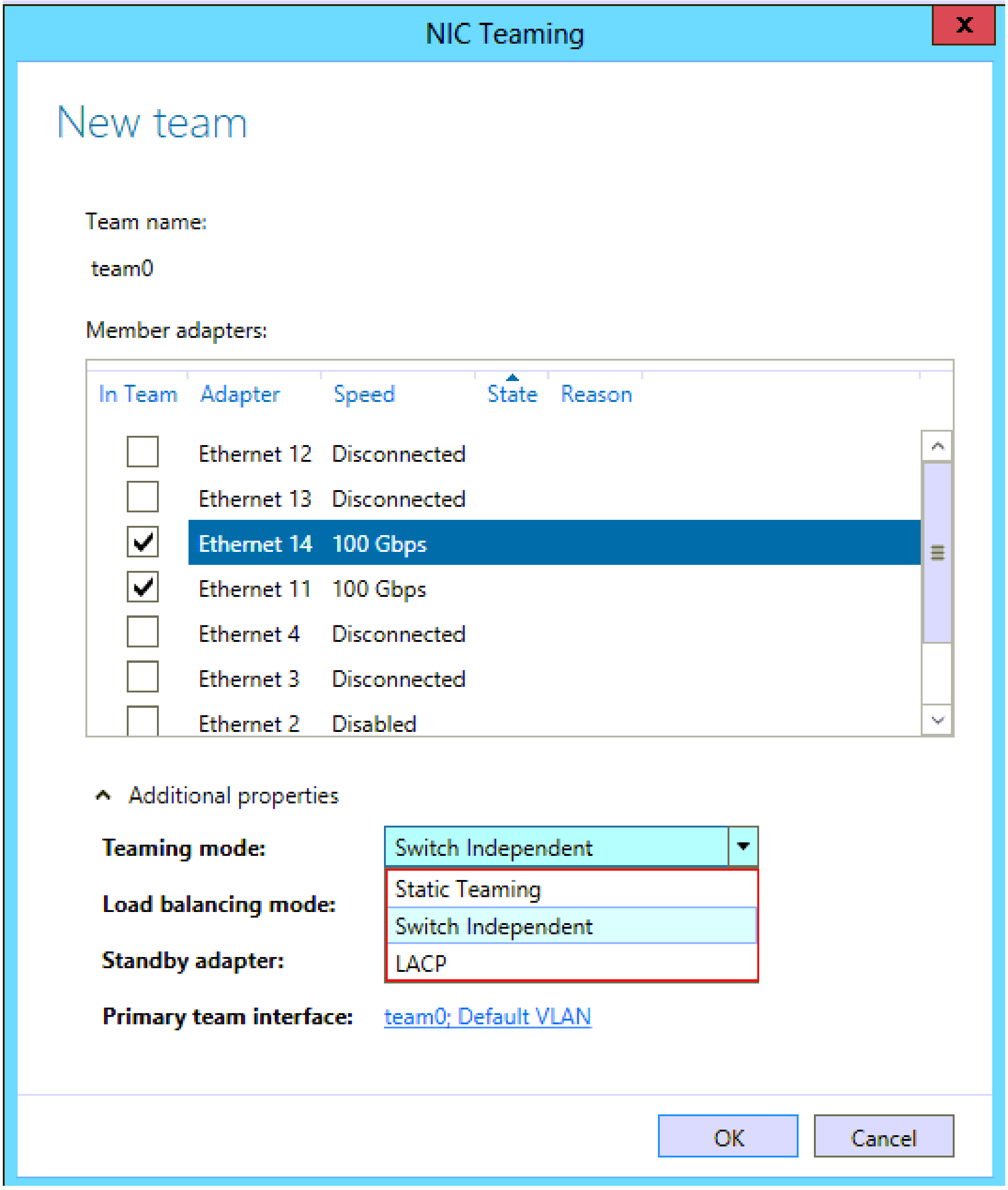
Member adapters:

In Team	Adapter	Speed	State	Reason
□	Ethernet 12	Disconnected		
□	Ethernet 13	Disconnected		
☑	Ethernet 14	100 Gbps		
☑	Ethernet 11	100 Gbps		
□	Ethernet 4	Disconnected		
□	Ethernet 3	Disconnected		
□	Ethernet 2	Disabled		

Working modes of NIC teaming are as follows:

1. **Static group**: manual load-balancing mode. Add the ports of the switch modules to the same Eth-Trunk interface.
2. **Independent switches**: work in active/standby mode. The switches do not need to be configured.
3. **LACA dynamic combination (LACP)**: After link negotiation over LACP is successful, each link periodically sends heartbeat negotiation messages to improve link reliability. In addition, the corresponding ports on a switch must be added to the Eth-Trunk and set to the static LACP mode.

Figure 3-12 Setting the working mode of NIC teaming



New team

Team name:
team0

Member adapters:

In Team	Adapter	Speed	State	Reason
☐	Ethernet 12	Disconnected		
☐	Ethernet 13	Disconnected		
☑	Ethernet 14	100 Gbps		
☑	Ethernet 11	100 Gbps		
☐	Ethernet 4	Disconnected		
☐	Ethernet 3	Disconnected		
☐	Ethernet 2	Disabled		

Additional properties

Teaming mode: Switch Independent

Load balancing mode:

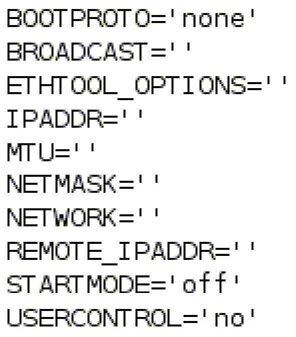
Standby adapter:

Primary team interface: team0; Default VLAN

3.3.3 Bonding in Linux

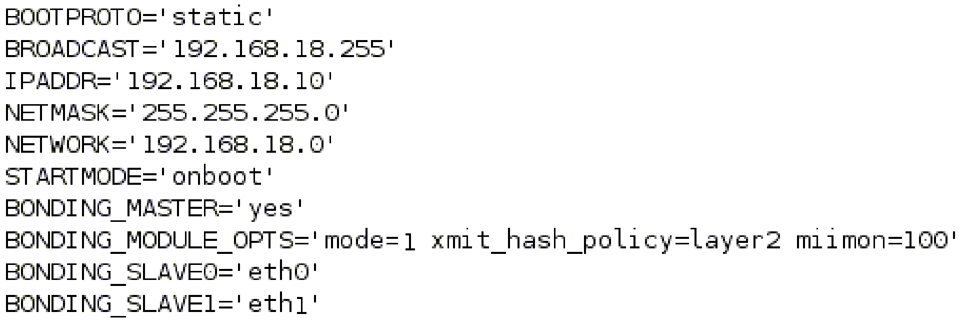
The NIC teaming function provided by Linux is called bonding, which can be configured through a configuration file or by using a GUI configuration tool. The directory of the configuration file varies according to the system, for example, the configuration file on SUSE Linux Enterprise 11 is in the **/etc/sysconfig/network/** directory) or by using a GUI configuration tool. NIC Bonding supports multiple working modes, such as the manual load balancing, active/standby, and LACP modes. For example, to configure NIC bonding on SUSE Linux Enterprise 11, perform the following steps:

1. Go to **/etc/sysconfig/network/**, create the configuration file **ifcfg-eth0/ifcfg-eth1** for NIC eth0/eth1 respectively, and add the following information to the file (if the file exists, modify it as follows):



```
BOOTPROTO='none'
BROADCAST=''
ETHTOOL_OPTIONS=''
IPADDR=''
MTU=''
NETMASK=''
NETWORK=''
REMOTE_IPADDR=''
STARTMODE='off'
USERCONTROL='no'
```

2. Go to **/etc/sysconfig/network**, create the file **ifcfg-bonding0**, and add the following information to the file:



```
BOOTPROTO='static'
BROADCAST='192.168.18.255'
IPADDR='192.168.18.10'
NETMASK='255.255.255.0'
NETWORK='192.168.18.0'
STARTMODE='onboot'
BONDING_MASTER='yes'
BONDING_MODULE_OPTS='mode=1 xmit_hash_policy=layer2 miimon=100'
BONDING_SLAVE0='eth0'
BONDING_SLAVE1='eth1'
```

   STARTMODE='onboot' indicates that bonding 0 automatically takes effect during the system startup. **mode=1** indicates that the bound network ports work in active/standby mode. **miimon=100** indicates that the link status is checked every 100 ms.

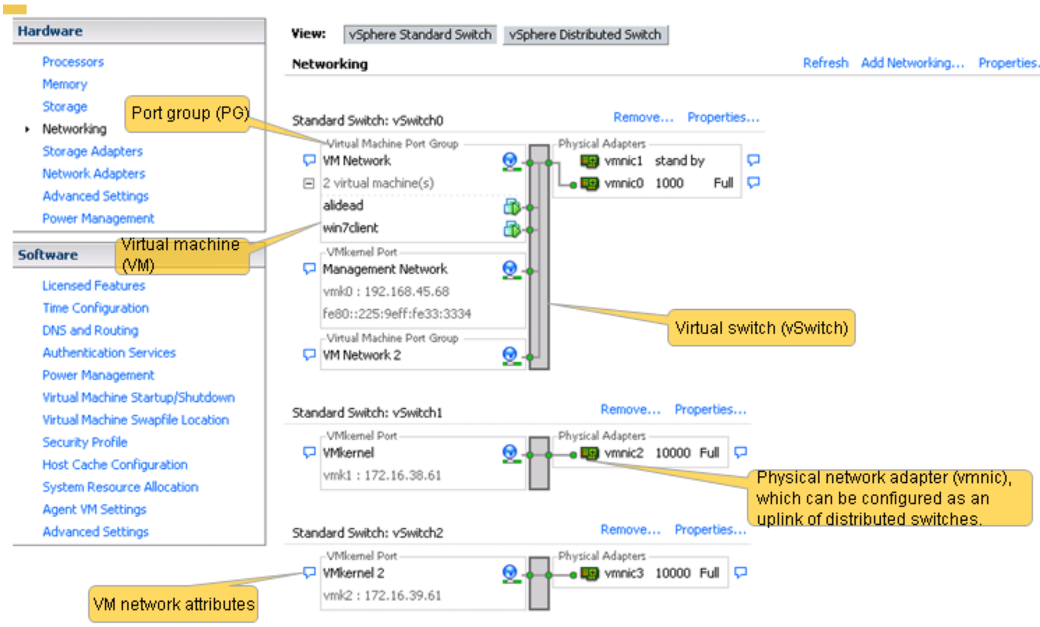
3. Run **/etc/init.d/network restart** to make the bonding configuration to take effect.

3.3.4 NIC Teaming in vSphere

3.3.4.1 vSphere Virtual Network Components

Figure 3-13 shows all components on the vSphere virtual network.

Figure 3-13 vSphere virtual network components



Standard Switch: vSwitch0

Virtual Machine Port Group: VM Network — 2 virtual machine(s): alidead, win7client

VMkernel Port: Management Network — vmk0 : 192.168.45.68, fe80::225:9eff:fe33:3334

Virtual Machine Port Group: VM Network 2

Physical Adapters: vmnic1 stand by; vmnic0 1000 Full

Standard Switch: vSwitch1

VMkernel Port: VMkernel — vmk1 : 172.16.38.61

Physical Adapters: vmnic2 10000 Full

Standard Switch: vSwitch2

VMkernel Port: VMkernel 2 — vmk2 : 172.16.39.61

Physical Adapters: vmnic3 10000 Full

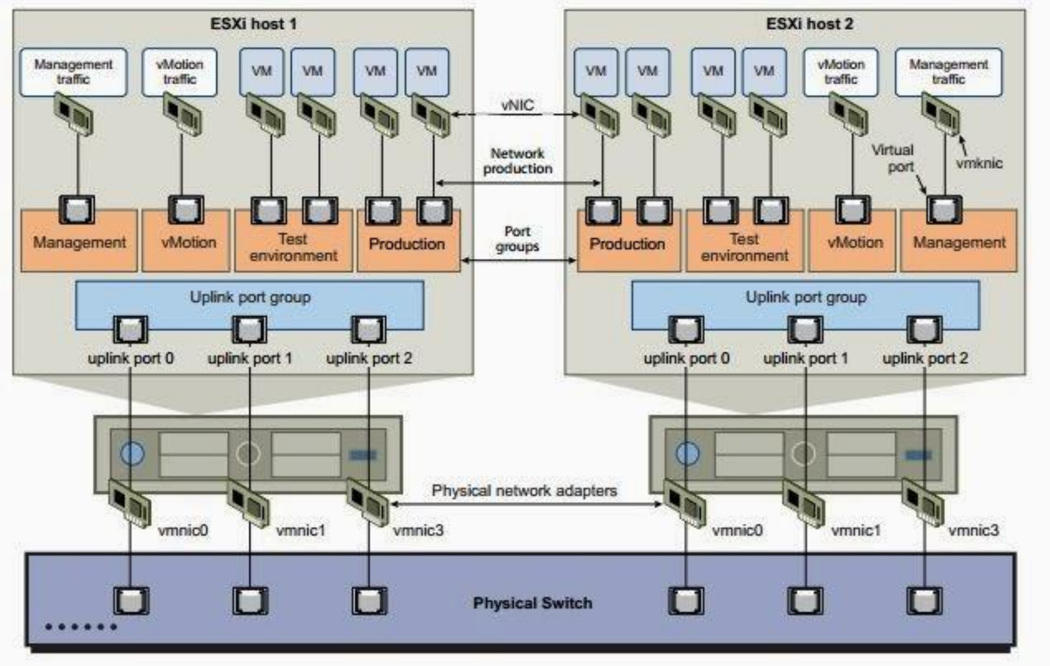
3.3.4.2 vSphere Virtual Switches

vSphere provides two types of virtual switches, that is, virtual standard switches (VSSs) and virtual distributed switches (VDSs).

VSS

The VSS running mode is similar to that of a physical Ethernet switch. The VSS detects VMs that logically connect to its virtual ports and forwards traffic to the correct VM based on the detection result. Physical Ethernet adapters (also called uplink adapters) can be used to connect a virtual network and a physical network so as to connect a vSphere standard switch to a physical switch. This connection is similar to the connection between physical switches for creating a large-scale network. Although similar to physical switches, vSphere VSSs do not provide some advanced functions of physical switches. Figure 3-14 shows the architecture of vSphere VSSs.

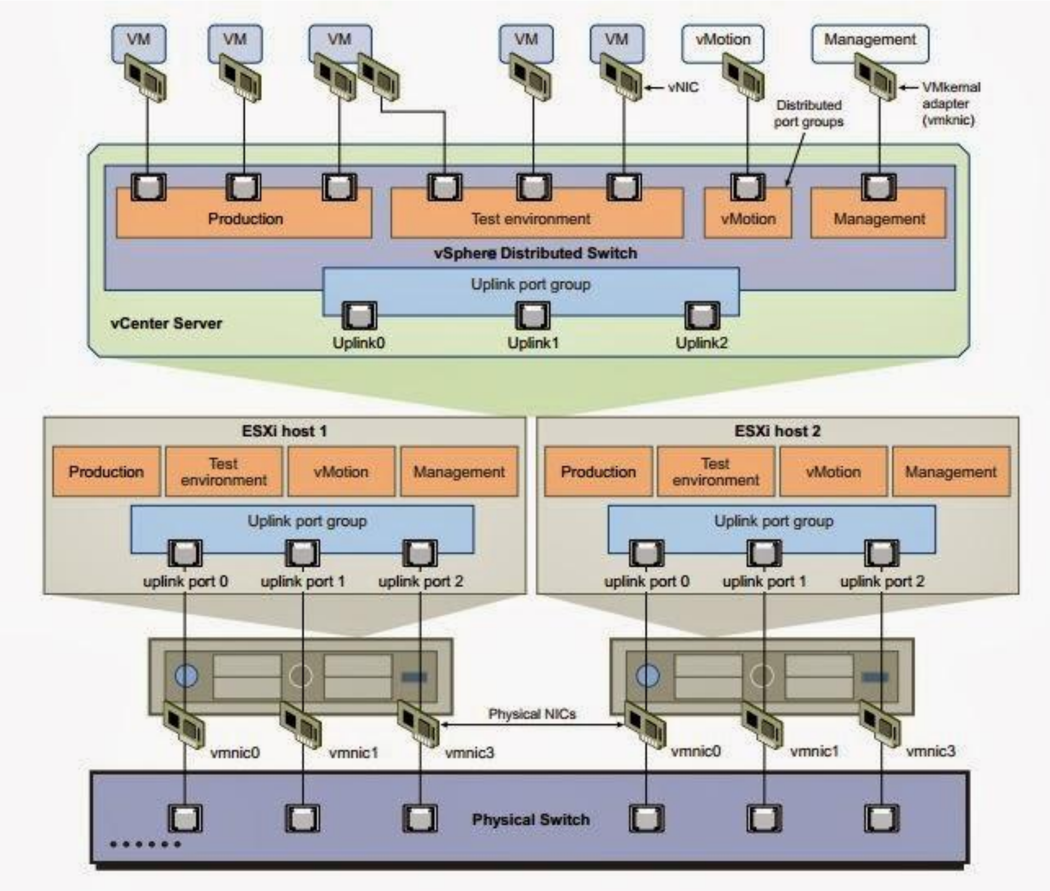
Figure 3-14 Architecture of vSphere VSSs



VDS

The VDS can be used as a single switch for all associated hosts in a data center, to provide centralized deployment, management, and monitoring of the virtual network. The administrator can configure a vSphere distributed switch on a vCenter server. This configuration will be sent to all hosts associated with the switch, which allows consistent network configuration when VMs are migrated across hosts. Figure 3-14 shows the architecture of vSphere VDSs.

Figure 3-15 Architecture of vSphere VDSs



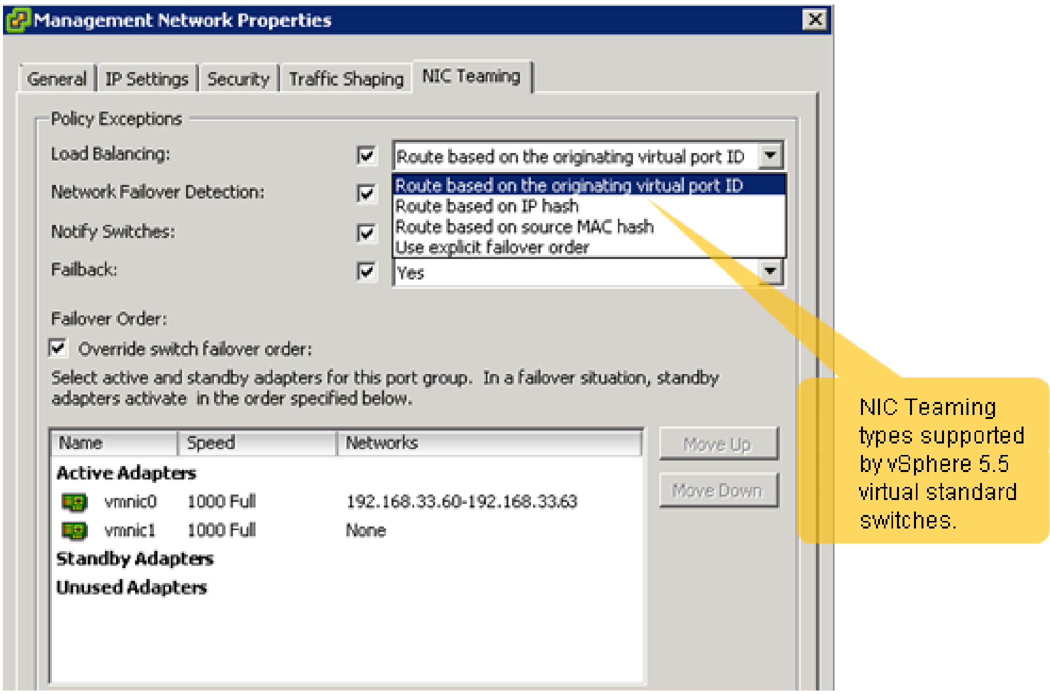
Same as the physical network, the virtual network also needs to improve the network connection reliability and bandwidth of a single network port. NIC Teaming is provided for the networking in vSphere virtual environments. vSphere NIC Teaming allows the traffic between the physical and virtual networks to be shared by some or all members and implements switchovers when a hardware fault or network interruption occurs. Table 3-3 describes the four types of NIC Teaming provided by VMware vSphere 5.5 VSSs.

Table 3-3 NIC Teaming types

Item	Description
Route based on the originating virtual port ID	Choose an uplink based on the virtual port through which traffic is transmitted to the virtual switch.
Route based on IP hash	Choose an uplink based on a hash of the source and destination IP addresses of each packet.
Route based on source MAC hash	Choose an uplink based on a hash of the source Ethernet MAC address.
Use explicit failover order	Always use the highest-order uplink from the list of active adapters to pass failover detection criteria.

The NIC Teaming types can be found in the **Load Balancing** drop-down list shown in Figure 3-16.

Figure 3-16 NIC Teaming types



Management Network Properties

General | IP Settings | Security | Traffic Shaping | NIC Teaming

Policy Exceptions

Load Balancing: ☑ Route based on the originating virtual port ID

- Route based on the originating virtual port ID
- Route based on IP hash
- Route based on source MAC hash
- Use explicit failover order

Network Failover Detection: ☑

Notify Switches: ☑

Failback: ☑ Yes

Failover Order:

☑ Override switch failover order:

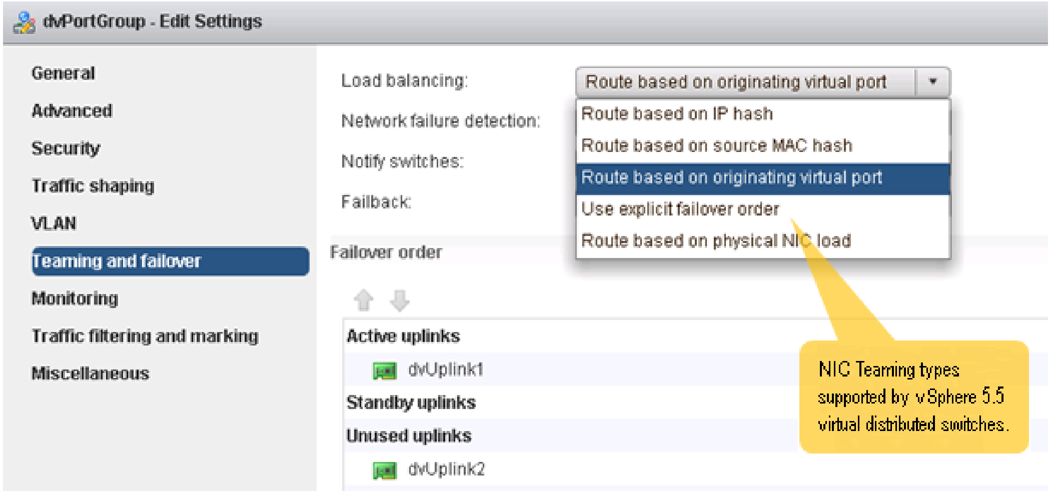
Select active and standby adapters for this port group. In a failover situation, standby adapters activate in the order specified below.

Name	Speed	Networks
Active Adapters		
vmnic0	1000 Full	192.168.33.60-192.168.33.63
vmnic1	1000 Full	None
Standby Adapters		
Unused Adapters		

NIC Teaming types supported by vSphere 5.5 virtual standard switches.

In addition to NIC Teaming types supported by VSSs, NIC Teaming provided by port groups of VMware vSphere 5.5 VDSs also includes Route based on physical NIC load, as shown in Figure 3-17. Only vSphere Enterprise Plus supports distributed switches. vSphere distributed switches support LACP, as shown in Figure 3-17 and Figure 3-18.

Figure 3-17 vSphere distributed switches supporting LACP 1



dvPortGroup - Edit Settings

Load balancing: Route based on originating virtual port

- Route based on IP hash
- Route based on source MAC hash
- Route based on originating virtual port
- Use explicit failover order
- Route based on physical NIC load

Network failure detection:

Notify switches:

Failback:

Failover order

Active uplinks

dvUplink1

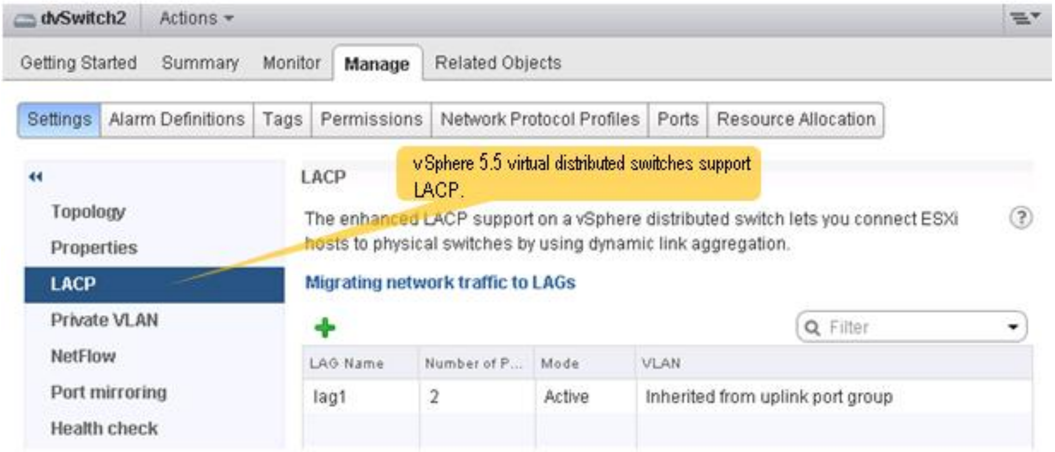
Standby uplinks

Unused uplinks

dvUplink2

NIC Teaming types supported by vSphere 5.5 virtual distributed switches.

Figure 3-18 vSphere distributed switches supporting LACP 2



LACP

vSphere 5.5 virtual distributed switches support LACP.

The enhanced LACP support on a vSphere distributed switch lets you connect ESXi hosts to physical switches by using dynamic link aggregation.

Migrating network traffic to LAGs

LAG Name	Number of P...	Mode	VLAN
lag1	2	Active	Inherited from uplink port group

For load balancing algorithm, the components for performing hash functions in the physical environment include the source and destination IP addresses, TCP/UDP port numbers, and source and destination MAC addresses. In LACP load balancing mode, virtual switches and physical switches negotiate with each other to determine the forwarding policy. In the vSphere virtual environment, components for performing hash functions include the source and destination IP addresses, source MAC address, and switch port numbers.

3.3.5 Mapping Between NIC Teaming and Switch Modules

The working mode of NIC Teaming varies depends on the switch module configuration. Some modes take effect only after switch modules are correctly configured. The main working modes of NIC Teaming are as follows:

1. **Switch Independent**: The working mode of NIC Teaming does not depend on a switch. NIC Teaming works in independent mode. For example, **mode = 1/5/6** of Linux bonding, **Route based on the originating virtual port ID**, **Route based on source MAC hash**, and **Use explicit failover order** in VMware, and **Switch Independent** mode in Windows.
2. **Manual load balancing**: In manual load balancing mode, NIC Teaming chooses active links by using the hash algorithm. Link selection is independent from the LACP negotiation results. For example, **mode = 0/2/3** of Linux bonding, **Route based on IP hash** in VMware, and **Static teaming** in Windows.
3. **LACP load balancing**: NICs and switch modules on a server negotiate with each other about NIC Teaming through LACP. Physical links change to active links after the LACP negotiation is successful.

Table 3-4 Mapping between working modes of NIC Teaming and switch modules

Operating System	Mode of NIC Teaming	Mode of Eth-Trunk
Linux	balance-rr or 0 (default) balance-xor or 2 broadcast or 3	mode manual
	active-backup or 1 (recommended)	NA

Operating System	Mode of NIC Teaming	Mode of Eth-Trunk
	balance-tlb or 5 balance-alb or 6	
	802.3ad (LACP) or 4 (recommended)	mode lacp static
VMware	Route based on the originating virtual port ID (recommended) Route based on source MAC hash Use explicit failover order	NA
	Route based on the source and destination IP hash	mode manual
	LACP (recommended)	mode lacp static
Windows	Independent switches	NA
	Static group	mode manual
	LACP	mode lacp static

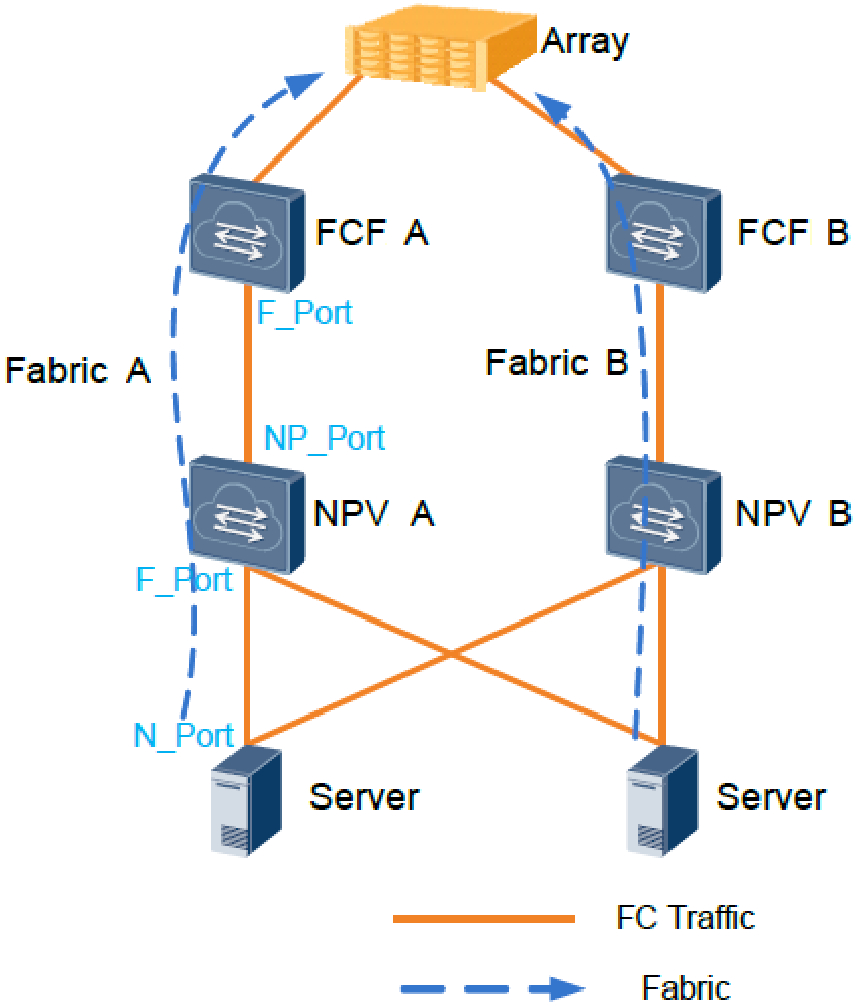
3.4 FC Technology

This chapter describes the working principles of the FC technology.

3.4.1 Basic Concepts

As shown in Figure 3-19, FC involves the following concepts: Fabric, FCF, NPV, WWN, FC_ID, Domain_ID, Area_ID, Port_ID, FC-MAP, Zone, and port roles.

Figure 3-19 FC networking



- Fabric

A fabric is the network topology where servers and storage devices are interconnected through one or more switches.

- FCF

A Fibre Channel Forwarder (FCF) is a switch that supports both FCoE and FC protocol stacks and is used to connect to a SAN or LAN. An FCF forwards FCoE packets and encapsulates or decapsulates them

- NPV

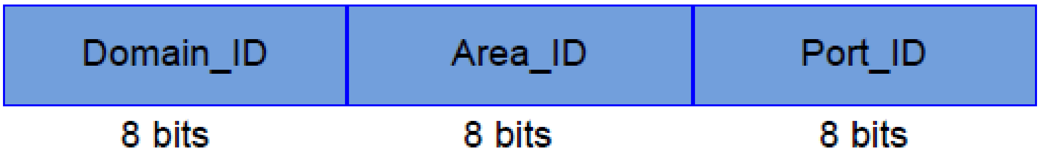
An N-Port Virtualization (NPV) switch is at the edge of a fabric network and between ENodes and FCFs, forwarding traffic from node devices to FCF switches.

- WWN

A World Wide Name (WWN) identifies an entity in a fabric network. A WWN is either a World Wide Node Name (WWNN) that identifies a node device or a World Wide Port Name (WWPN) that identifies a device port. Each entity in a SAN is assigned with a WWN before the entity is delivered from the factory.

- FC ID

Figure 3-20 FC_ID format



An FC_ID is an FC address. In a SAN, the FC protocol accesses entities by using their FC addresses. An FC address uniquely identifies an N-Port of a node device.

- Domain_ID

A domain ID in a SAN uniquely identifies an FC switch. Routing and forwarding among FC switches are based on domain IDs.

- Area_ID

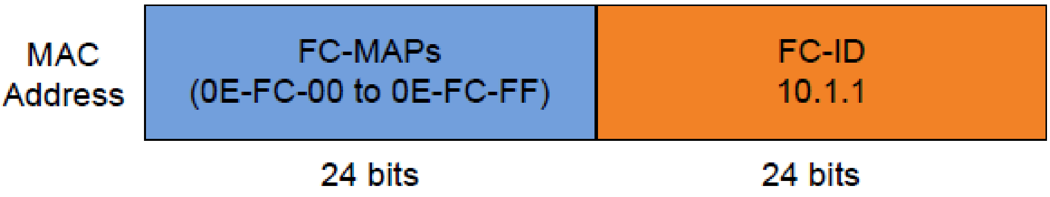
One or more N_Ports of a node device can be assigned to an area, which is identified by an area ID.

- Port_ID

A port ID identifies an N_Port.

- FC-MAP

Figure 3-21 Fabric-provided MAC address (FPMA)



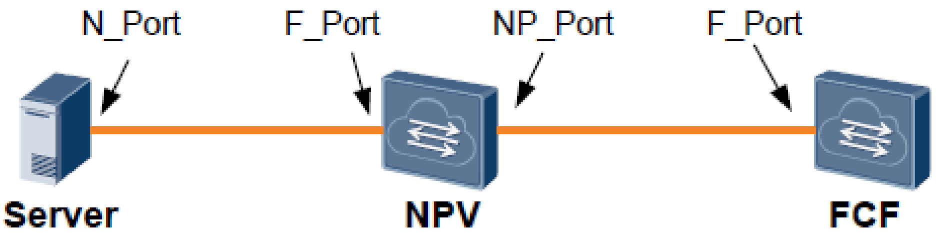
FCoE frames are forwarded by using locally unique MAC addresses (unique only in the local Ethernet subnetwork). FCFs assign locally unique MAC addresses to ENodes or ENodes specify their own locally unique MAC addresses and inform FCFs. In FPMA mode, FCFs assigned locally uniquely MAC addresses to ENodes. An FPMA is an FC ID with a 24-bit FCoE MAC address prefix (FC-MAP).

- Zone

N_Ports are added to different zones so that the N_Ports are isolated. A zone set is a set of zones. It is a logical control unit between zones and instances and simplifies configurations. Each instance can have only one activated zone set.

- Port roles

Figure 3-22 Port roles



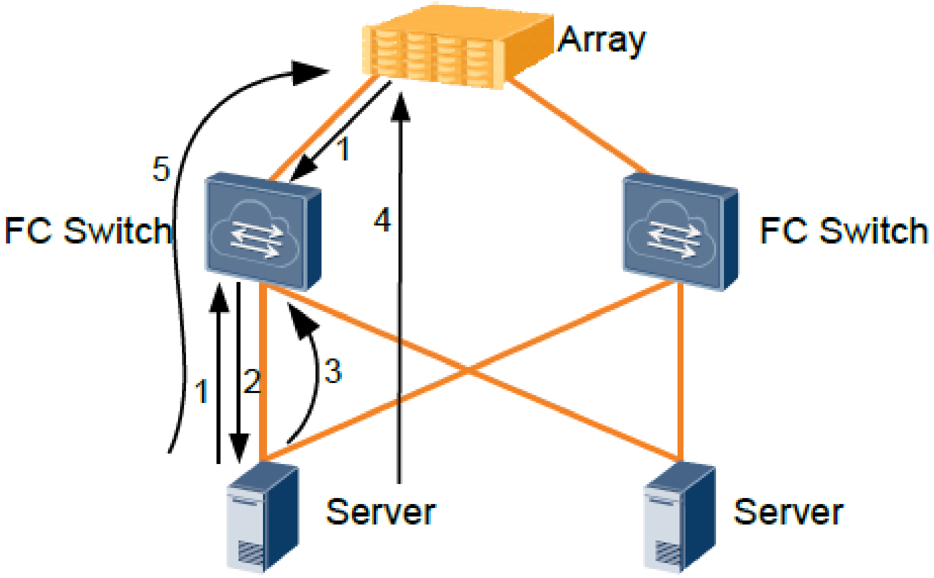
In a traditional FC network, FC devices interact with each other through FC ports. FC ports include N-Ports, F-Ports, and NP-Ports.

1. Node port (N_Port): indicates a port on an FC host (server or storage device) and connects to an FC switch.
2. Fabric port (F_Port): indicates a port on an FC switch and connects to an FC host, enabling the FC host to access the fabric.
3. N_Port Proxy (NP_Port): indicates a port on an NPV switch and connects to an FCF switch.

3.4.2 Working Principles

A fabric network formed by FC switches provides data transmission services. The following introduces the FC SAN communication process by describing how a server accesses a storage array, as shown in Figure 3-23.

Figure 3-23 FC SAN communication process



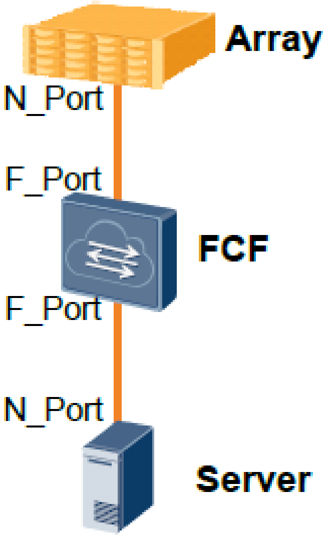
1. When servers or storage arrays go online in the fabric network, they request FC switches to provide services and register with FC switches through fabric login (FLOGI) packets.
2. FC switches allocate FC addresses to servers and storage devices.

3. Servers and storage devices send name service registration requests to FC switches, which create and maintain the mapping table between FC addresses and WWNs.
4. A server sends session requests to a target node through a port login (PLOGI).
5. After the session is established between a server and a storage device through a PLOGI, FC data can be transmitted. An FC switch determines the route and forwards data based on FC addresses of the server and the storage device.

3.4.3 FCF

A Fibre Channel Forwarder (FCF) switch supports both FCoE and FC protocol stacks for connecting to SAN and LAN environments. In an FC SAN, an FCF is mainly used for transmitting FC data. An FCF forwards FCoE packets and encapsulates or decapsulates them.

Figure 3-24 FCF network

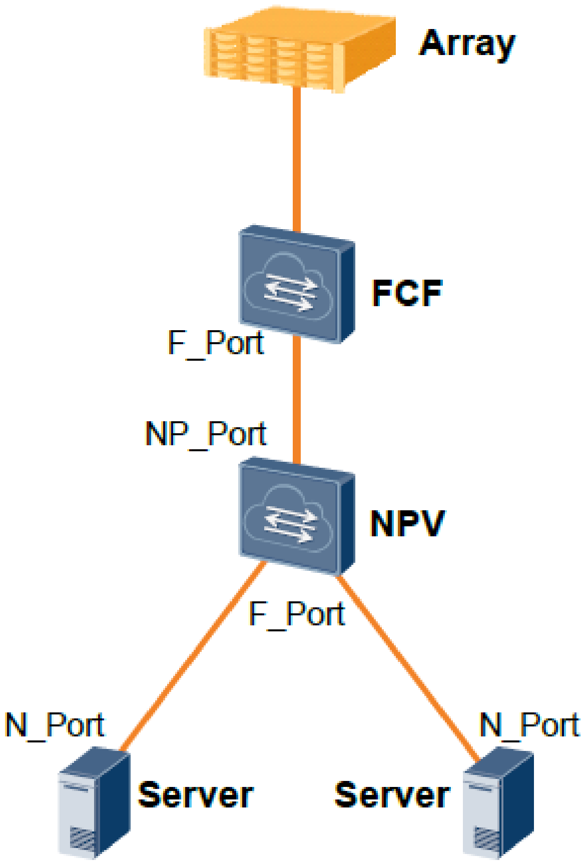


As shown in Figure 3-24, F_Ports of the FCF directly connect to N_Ports of a server and a storage array. Each FCF switch has a domain ID. Each FC SAN supports a maximum of 239 domain IDs. Therefore, each FC SAN can contain a maximum of 239 FCF switches.

3.4.4 NPV

A SAN has high demands for edge switches directly connected to node devices. N-Port virtualization (NPV) switches do not occupy domain IDs and enable a SAN to exceed the limit of 239 edge switches.

Figure 3-25 NPV network



As shown in Figure 3-25, an NPV switch is located at the edge of a fabric network and between node devices and an FCF. The NPV switch use F_Ports to connect to N_Ports of node devices and use NP_Ports to connect to F_Ports of the FCF switch. As a result, the node devices connect to the fabric network through the NPV switch, which forwards traffic from all node devices to the core switch.

For a node device, the NPV switch is an FCF switch that provides F_Ports. For an FCF switch, an NPV switch is a node device that provides N_Ports.

3.4.5 Zone

In an FCoE network, users can use zones to control access between node devices to improve network security.

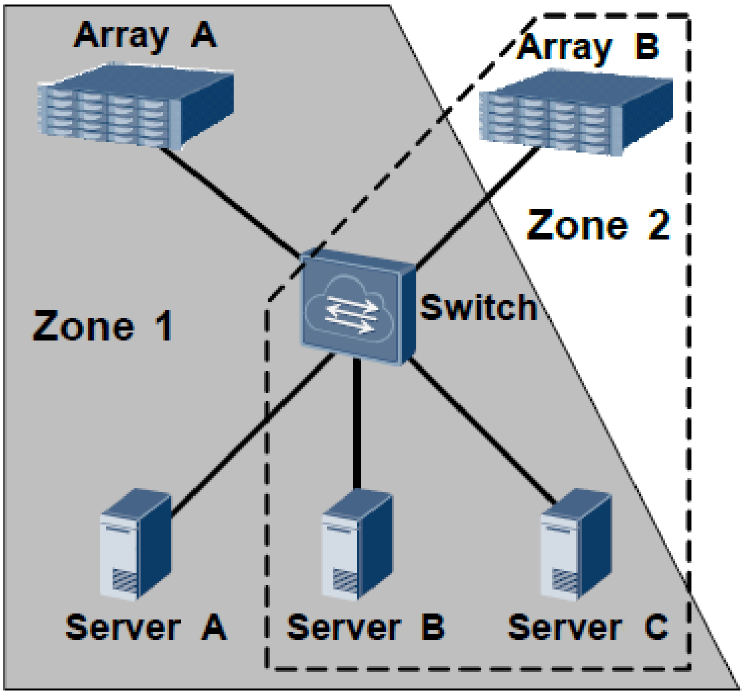
- Zone

A zone contains multiple zone members. A node device can join different zones at the same time. Node devices in the same zone can access each other. Node devices in different zones cannot access each other. A zone member can be defined in the following ways:

1. Zone alias: After a zone alias joins a zone, the members in the zone alias also join the zone.
2. FC_ID: indicates an FC address. Node devices in an FC network access each other through FC addresses.

3. FCoE interface: Node devices in an FCoE network interact with each other through FCoE interfaces.
4. WWNN (World Wide Node Name): A WWNN is a 64-bit address used to identify a node device in an FC network.
5. WWPN (World Wide Port Name): A WWPN is a 64-bit address used to identify a port of a node device in an FC network.

Figure 3-26 Zone network



As shown in Figure 3-26, users can control access between node devices by adding node devices to different zones. For example, array B can only interact with Server B and Server C, not Server A.

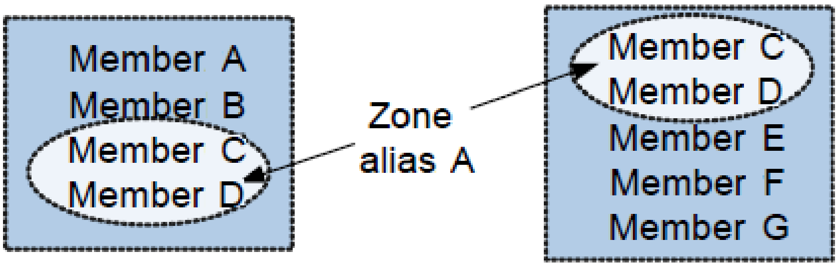
- Zone set

A zone set contain multiple zones. A zone can join different zone sets at the same time. Zones in a zone set are valid only after the zone set is activated. Each instance can have only one activated zone set.

- Zone alias

Applying zone aliases to zone configurations simplifies the configurations. If multiple zone members need to join multiple zones, you can add the zone members to a zone alias and then add the zone alias to a zone as a zone member. This avoids adding multiple zone members repeatedly.

Figure 3-27 Zone alias



As shown in Figure 3-27, if Members C and D both need to join Zones A and B, you can add Members C and D to Zone Alias A, and then add Zone Alias A to Zones A and B. This simplifies configurations.

3.5 DCB

3.5.1 DCB Overview

Data Center Bridging (DCB) is a set of enhancements to Ethernet for use in data center environments. It is defined by the IEEE 802.1 working group. DCB is used to build lossless Ethernet to meet the quality of service (QoS) requirements of a converged data center network. Table 3-5 describes three features defined by DCB, that is, PFC, ETS, and DCBX.

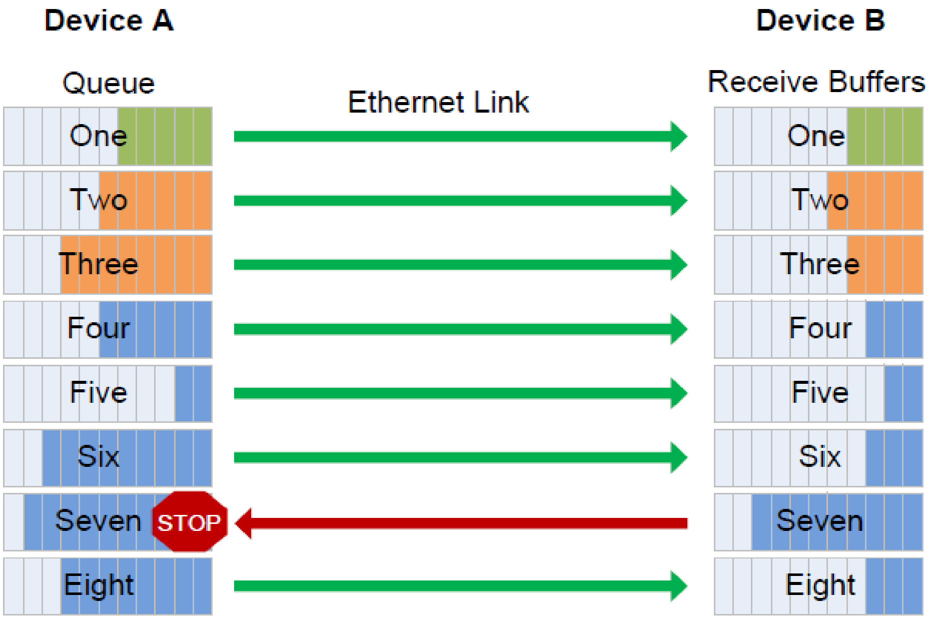
Table 3-5 Features defined by DCB

Feature	Description
Priority-based Flow control (PFC)	Implements priority-based flow control on a shared link.
Enhanced transmission selection (ETS)	Implements priority-based bandwidth control on a shared link.
Data Center Bridging Exchange (DCBX) protocol	Provides auto-negotiation of PFC/ETS parameters between Ethernet devices.

3.5.2 PFC

PFC is also called Per Priority Pause or Class Based Flow Control (CBFC). It is an enhancement to the Ethernet Pause mechanism. PFC is a priority-based flow control mechanism. As shown in Figure 3-28, the transmit interface of Device A is divided into eight queues of different priorities. The receive interface of Device B is divided into eight buffers. The eight queues and eights buffers are in one-to-one correspondence. When a receive buffer on Device B is to be congested, Device B sends a STOP signal to Device A. Device A stops sending packets in the corresponding priority queue when receiving the STOP signal.

Figure 3-28 PFC working principle

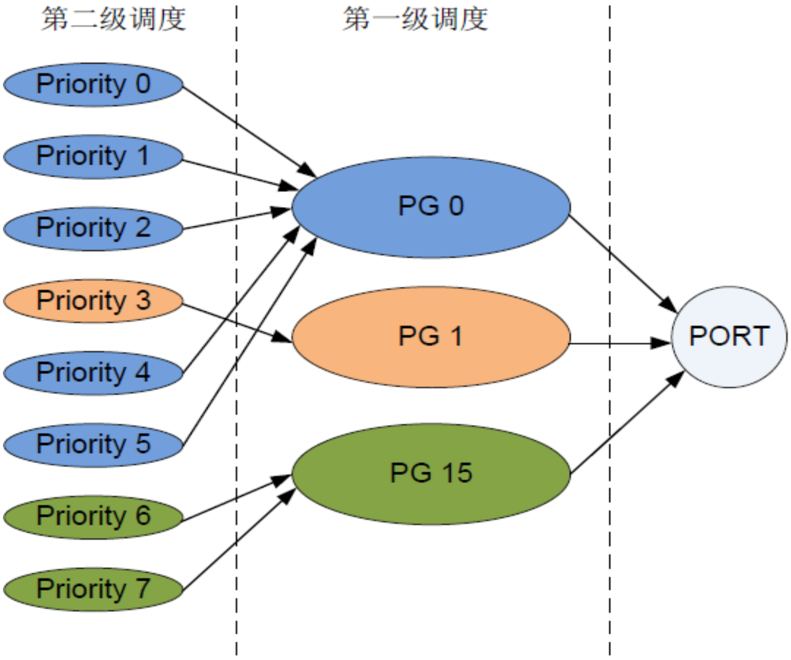


PFC allows traffic in one or multiple queues to be stopped, which does not affect data exchange on the entire interface. Data transmission in each queue can be separately stopped or resumed without affecting other queues. This feature enables various types of traffic to be transmitted on the same link. The system does not apply the backpressure mechanism to the priority queues with PFC disabled and directly discards packets in these queues when congestion occurs.

3.5.3 ETS

The converged data center network bears three types of traffic: inter-process communication (IPC) traffic, local area network (LAN) traffic, and storage area network (SAN) traffic. The converged network has high QoS requirements. The traditional QoS cannot meet requirements of the converged network, whereas Enhanced Transmission Selection (ETS) uses hierarchical scheduling to guarantee QoS on the converged network. ETS provides two levels of scheduling: scheduling based on the priority group (PG) and scheduling based on the priority. Figure 3-29 illustrates how ETS works. On an interface, PG-based scheduling is performed first, and then priority-based scheduling is performed.

Figure 3-29 ETS scheduling model

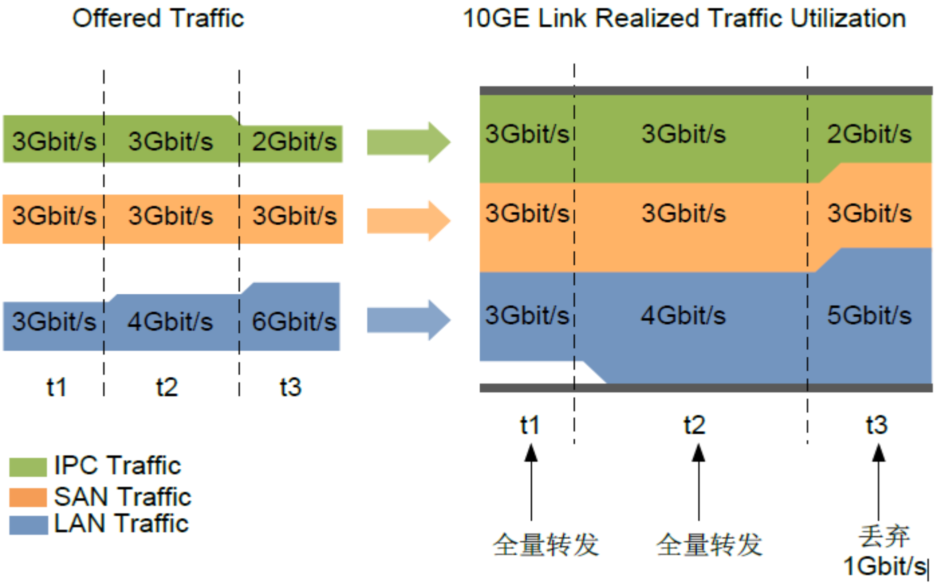


A PG is a group of priority queues with the same scheduling attributes. Users can add queues with different priorities to a PG. PG-based scheduling is called level-1 scheduling. ETS defines three PGs: PG0 for LAN traffic, PG1 for SAN traffic, and PG15 for IPC traffic.

As defined by ETS, PG0, PG1, and PG15 use priority queue (PQ)+Deficit Round Robin (DRR). PG15 uses PQ to schedule delay-sensitive IPC traffic. PG0 and PG1 use DRR. In addition, bandwidth can be allocated to PGs based on actual networking.

As shown in Figure 3-30, the queue with priority 3 carries FCoE traffic and is added to the SAN group (PG1). Queues with priorities 0, 1, 2, 4, and 5 carry LAN traffic and are added to the LAN group (PG0). The queue with priority 7 carries IPC traffic and is added to the IPC group (PG15). The total bandwidth of the interface is 10 Gbit/s. Each of PG1 and PG0 is assigned 50% of the total bandwidth, that is, 5 Gbit/s.

Figure 3-30 Configure the PG bandwidth



At t1 and t2, all traffic can be forwarded because the total traffic on the interface is within the interface bandwidth. At t3, the total traffic exceeds the interface bandwidth and LAN traffic exceeds the given bandwidth. At this time, LAN traffic is scheduled based on ETS parameters and 1 Gbit/s LAN traffic is discarded.

ETS also provides PG-based traffic shaping. The traffic shaping mechanism limits traffic bursts in a PG to ensure that traffic in this group is sent out at an even rate.

In addition to PG-based scheduling, ETS also provides priority-based scheduling, that is, level-2 scheduling, for queues in the same PG. Queues in the same PG support queue congestion management, queue shaping, and queue congestion avoidance.

3.5.4 DCBX

To implement lossless Ethernet on a converged data center network, both ends of an FCoE link must have the same PFC and ETS parameter settings. Manual configuration of PFC and ETS parameters may increase administrator's workloads and cause configuration errors. DCBX, a link discovery protocol, enables devices at both ends of a link to discover and exchange DCB configurations. This greatly reduces administrator's workloads. DCBX provides the following functions:

1. Detects the DCB configurations of the peer device.
2. Detects the DCB configuration errors of the peer device.
3. Configures DCB parameters for the peer device.

DCBX enables DCB devices at both ends to exchange the following DCB configurations:

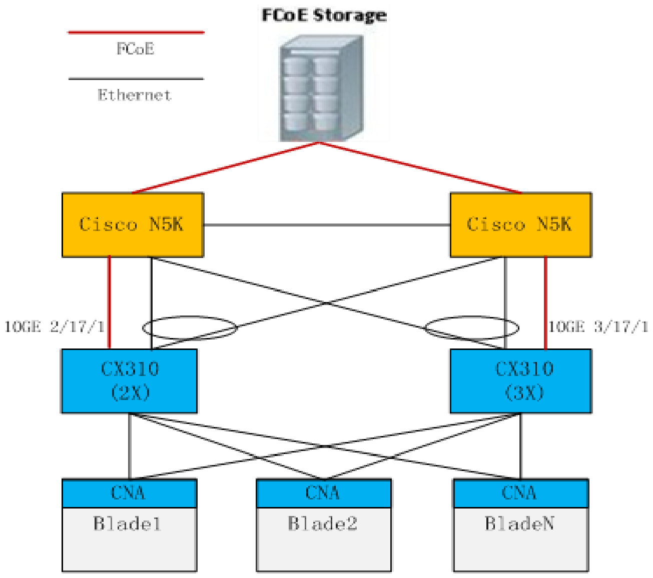
1. ETS PG information
2. PFC

DCBX encapsulates DCB configurations into Link Layer Discovery Protocol (LLDP) type-length-values (TLVs) so that devices at both ends of a link can exchange DCB configurations.

3.5.5 Configuration Instance

CNAs on a blade server connect to an FCoE storage device through CX310s and Cisco N5K switches, as shown in Figure 3-31. Configure DCB features for the FCoE links between the CX310s and N5K switches. This section provides only the configuration procedure of the CX310 in slot 2X. The configuration procedure of the CX310 in slot 3X is the same.

Figure 3-31 DCB application networking



# Configure DCBX functions.

```
[~CX310_2] lldp enable
[*CX310_2] interface 10GE 2/17/1
[*CX310_2-10GE2/17/1] lldp tlv-enable dcbx
[*CX310_2-10GE2/17/1] quit
```

# Enable the PFC function of the interface.

```
[*CX310_2] interface 10GE 2/17/1
[*CX310_2-10GE2/17/1] dcb pfc enable mode auto
[*CX310_2-10GE2/17/1] quit
```

# Create an ETS profile and add queue 3 to PG1 (by default), and other queues to PG0. PG15 is empty.

```
[*CX310_2] dcb ets-profile ets1
[*CX310_2-ets-ets1] priority-group 0 queue 0 to 2 4 to 7
```

# Configure PG-based flow control and set DRR weights of PG0 and PG1 to 60% and 40% respectively.

```
[*CX310_2-ets-ets1] priority-group 0 drr weight 60
[*CX310_2-ets-ets1] quit
```

# Apply the ETS Profile to the FCoE interface.

```
[*CX310_2] interface 10GE 2/17/1
[*CX310_2-10GE2/17/1] dcb ets enable ets1
```

```
[*CX310_2-10GE2/17/1] quit
[*CX310_2] commit
```

3.6 FCoE

3.6.1 FCoE Overview

The LAN and SAN in a traditional data center are deployed and maintained independently. The LAN implements communication between servers and between servers and clients, and the SAN implements communication between servers and storage devices. As the number of servers dramatically increases with rapid development of data centers, independent deployment of LANs and SANs results in the following problems:

1. **Complicated networking**: Independent LANs and SANs result in poor flexibility in service deployment and network expansion, and high management costs.
2. **Low energy efficiency**: Four to six NICs must be installed on each server, including NICs connected to LANs and host bus adapters (HBAs) connected to SANs. The use of diversified NICs on servers increases the power consumption and cooling cost for data centers.

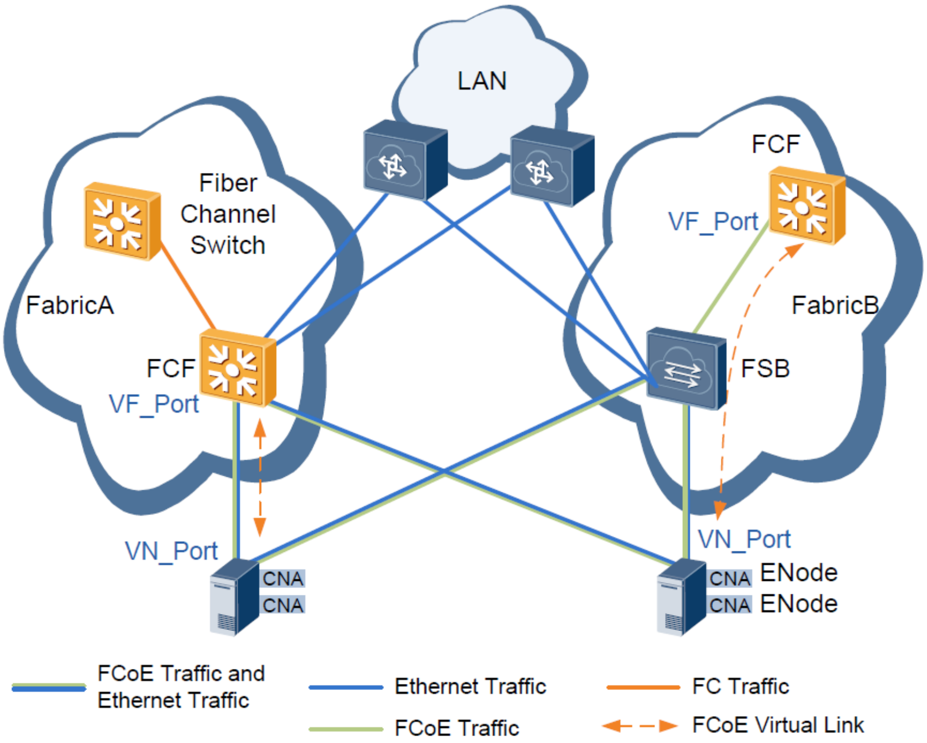
FCoE and DCB are developed to resolve the preceding problems:

1. FCoE encapsulates FC frames into Ethernet frames, allowing LANs and SANs to share network resources. With FCoE technology, LAN and SAN networks can be converged.
2. DCB builds a lossless Ethernet network on a data center network. This technology enables traditional Ethernet to implement congestion control as on the FC SAN and ensures transmission quality for FCoE convergence services.

3.6.2 Basic Concepts

Figure 3-32 shows FCoE networking. The basic concepts of FCoE include ENode, FCoE Forwarder (FCF), Fabric, FCoE virtual link, FCoE Initialization Protocol (FIP), FIP Snooping Bridge (FSB), port roles, and FCoE VLAN.

Figure 3-32 FCoE networking



3.6.2.1 ENode

An ENode is a CNA that supports FCoE and FC. A traditional server houses two network adapters: an NIC connected to a LAN and an HBA connected to a SAN. The CNA provides both NIC and HBA functions. It can forward Ethernet data, process FCoE packets in upper layers, and encapsulate and decapsulate FCoE frames.

3.6.2.2 FCoE Virtual Link

An FCoE virtual link is a point-to-point logical link between FCoE devices, for example, between an ENode and an FCF. The connection between an ENode and FCF is not point-to-point when the ENode and FCF are connected through a lossless Ethernet network. The FCoE virtual link is used to solve this problem.

3.6.2.3 FIP

FIP is a L2 protocol that discovers FC terminals on an FCoE network, implements fabric login, and establishes FCoE virtual links. An ENode can log in to the fabric over FIP to communicate with the target FC device. FIP can also maintain FCoE virtual links.

3.6.2.4 FCoE VLAN

FCoE frames are forwarded in specified VLANs as defined in FC-BB-5. In the FC protocol stack, FC devices support multiple virtual storage area networks (VSANs), which are similar to Ethernet VLANs. FC traffic in different VSANs is identified by FCoE VLANs during FCoE encapsulation. An FCoE virtual link corresponds to one FCoE VLAN. An FCoE VLAN bears only FCoE traffic and does not bear any Ethernet traffic, such as IP traffic.

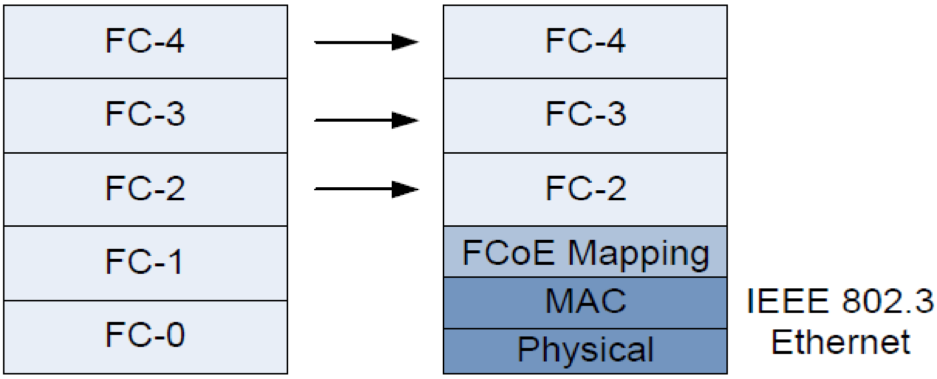
3.6.3 FCoE Packet Format

In the traditional FC protocol, the FC protocol stack is divided into five layers:

1. FC-0: defines the bearer medium type.
2. FC-1: defines the frame coding and decoding mode.
3. FC-2: defines the frame division protocol and flow control mechanism.
4. FC-3: defines general services.
5. FC-4: defines the mapping from upper-layer protocols to FC.

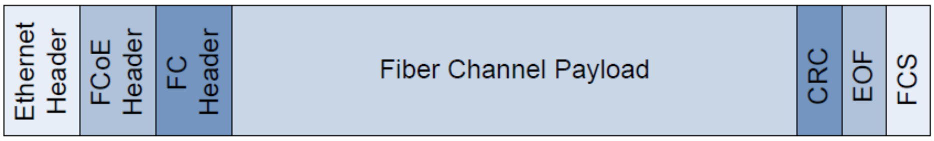
FC-0 and FC-1 in the FCoE protocol stack map physical and MAC layers in IEEE 802.3 Ethernet respectively. The FCoE protocol stack adds FCoE mapping as an adaptation layer between the upper-layer FC protocol stack and lower-layer Ethernet protocol stack, as shown in Figure 3-33.

Figure 3-33 FCoE packet format



FCoE encapsulates an FC frame into an Ethernet frame. Figure 3-34 shows FCoE frame encapsulation.

Figure 3-34 FCoE packet protocol



1. The Ethernet Header specifies the packet source, destination MAC address, Ethernet frame type, and FCoE VLAN.
2. The FCoE Header specifies the FCoE frame version number and flow control information.
3. Similar to a traditional FC frame, the FC Header specifies the source and destination addresses of an FC frame.

3.6.4 FIP

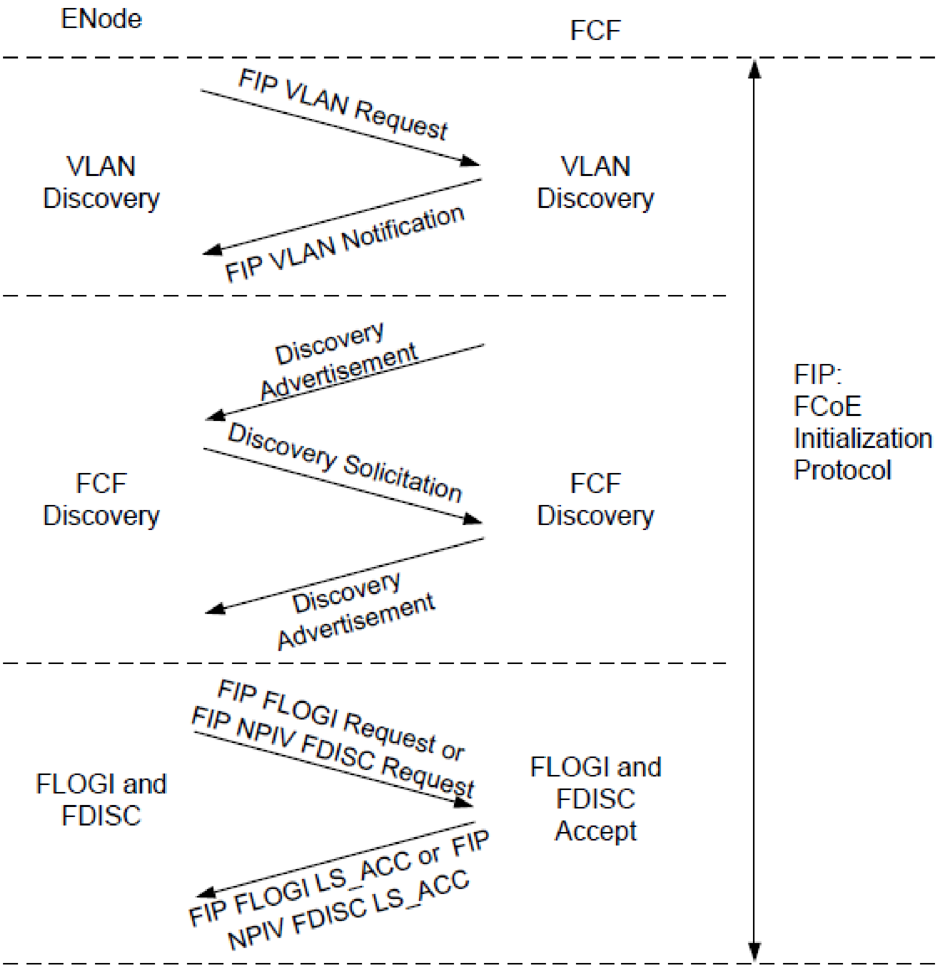
FIP, an FCoE control protocol, establishes and maintains FCoE virtual links between FCoE devices, for example, between ENodes and FCFs. In the process of creating a virtual link:

1. FIP discovers an FCoE VLAN and the FCoE virtual interface of the remote device.
2. FIP completes initialization tasks, such as fabric login (FLOGI) and fabric discovery (FDISC), for the FCoE virtual link.

After an FCoE virtual link is set up, FIP maintains the FCoE virtual link in the following way:

1. Periodically checks whether FCoE virtual interfaces at both ends of the FCoE virtual link are reachable.
2. Remove the FCoE virtual link through fabric logout (FLOGO).

The following figure shows the process of setting up an FCoE virtual link between an ENode and FCF. The ENode and FCF exchange FIP frames to establish the FCoE virtual link. After the FCoE virtual link is set up, FCoE frames are transmitted over the link.



An FCoE virtual link is set up through three phases: FIP VLAN discovery, FIP FCF discovery, and FIP FLOGI and FDISC. The FIP FLOGI and FDISC process is similar to the FLOGI and FDISC process defined in traditional FC protocol.

3.6.4.1 FIP VLAN Discovery

FIP VLAN discovery discovers the FCoE VLANs that will transmit FCoE frames. In this phase, an ENode discovers all potential FCoE VLANs but does not select an FCF. The FIP VLAN discovery process is as follows:

1. An ENode sends a FIP VLAN discovery request packet (FIP VLAN request) to a multicast MAC address called All-FCF-MAC 01-10-18-01-00-02. All FCFs listen on packets destined for this MAC address.
2. All FCFs report one or more FCoE VLANs to the ENode through a common VLAN. The FCoE VLANs are available for the ENode's VN_Port login. FIP VLAN discovery is an optional phase as defined in FC-BB-5. An FCoE VLAN can be manually configured by an administrator, or dynamically discovered using FIP VLAN discovery.

3.6.4.2 FIP FCF Discovery

FIP FCF discovery is used by ENodes to discover FCFs that allow logins. The FIP FCF discovery process is as follows:

1. Each FCF periodically sends FIP FCF discovery advertisement messages in each configured FCoE VLAN. The advertisement messages are destined for the multicast MAC address All-ENode-MAC 01-10-18-01-00-01 to which all ENodes can listen. The FIP FCF discovery advertisement messages contain the FCF MAC address and FCoE virtual link parameters, such as the FCF priority and the timeout interval of FIP packets.
2. The ENode obtains FCF information from the received discovery advertisement messages, selects an FCF with the highest priority, and sends a unicast FIP FCF discovery solicitation message to the selected FCF.
3. After receiving the discovery solicitation message, the FCF sends a unicast discovery advertisement message, allowing the ENode to log in.

In addition to receiving discovery advertisement messages periodically, ENodes newly joining a network do not need to wait for the messages from all FCFs. FC-BB-5 allows ENodes to send FIP FCF discovery solicitation messages to the multicast MAC address All-FCF-MAC. FCFs that receive the solicitation messages send a unicast FIP FCF discovery advertisement message to the requesting ENode. Based on the received Advertisement messages, the ENode selects an FCF with high priority to set up a virtual link with its VN_Port.

3.6.4.3 FIP FLOGI and FDISC

After discovering all FCFs and selecting one for login, the ENode notifies the selected FCF to set up a virtual link with its VF_Port. Then FCoE frames can be exchanged on the established FCoE virtual link. FIP FLOGI and FIP FDISC packets are unicast packets, similar to the FC FLOGI and FDISC packets that they replace. The unicast packets assign MAC address to the ENode so that it can log in to the fabric. FIP FDISC is similar to FIP FLOGI. The difference is that FIP FLOGI refers to the procedure for setting up a virtual link when ENode logs in to the fabric for the first time, and FIP FDISC refers to the procedure for setting up a virtual link for each virtual machine (VM) when multiple VMs exist on an ENode. Take FIP FLOGI as an example and the FIP FLOGI process is as follows:

1. An ENode sends an FIP FLOGI request to an FCF.
2. The FCF allocates a locally unique fabric provided MAC address (FPMA) or a server provided MAC address (SPMA) to the ENode.

3.6.5 FCoE Virtual Link Maintenance

On the traditional FC network, FC can immediately detect faults on a physical link. In FCoE, FC cannot immediately detect faults on a physical link because Ethernet encapsulation is used. FIP provides a Keepalive mechanism to solve the problem. FCoE monitors FCoE virtual links as follows:

1. The ENode periodically sends FIP Keepalive packets to the FCF. If the FCF does not receive FIP Keepalive packets within 2.5 times the keepalive interval, the FCF considers the FCoE virtual link faulty and terminates the FCoE virtual link.
2. The FCF periodically sends multicast discovery advertisement messages to the destination MAC address ALL-ENode-MAC to all ENodes. If the ENode does not receive multicast discovery advertisement messages within 2.5 times the keepalive interval, the ENode considers the FCoE virtual link faulty and terminates the FCoE virtual link.

If the FCF does not receive FIP keepalive packets from the ENode, the FCF sends an FIP clear virtual link message to the ENode to clear the FCoE virtual link. When the ENode logs out, it sends a Fabric Logout request message to the FCF to terminate the virtual link.

3.6.6 FIP Snooping

On an FC network, FC switches are considered as trusted devices. Other FC devices, such as ENodes, must log in to an FC switch before they can connect to the FC network. The FCF switch then assigns addresses to the FC devices. FC links are point-to-point, and an FC switch can completely control traffic received and sent by FC devices. Therefore, FC switches enable FC devices to exchange packets using the specified addresses and protect FC devices against malicious attacks.

When an FCoE switch is deployed between an ENode and an FCF, FCoE frames are forwarded on the FCoE switch based on the Ethernet protocol because FCoE switches do not support FC. In this case, FCoE frames may not be destined for the FCF, and the point-to-point connection between the ENode and FCF is terminated. To achieve equivalent robustness as an FC network, the FCoE switch must forward FCoE traffic from all ENodes to the FCF. FIP snooping obtains FCoE virtual link information by listening to FIP packets, controls the setup of FCoE virtual links, and defends against malicious attacks.

The FCoE switch running FIP snooping is called an FSB. The 10GE switch modules of the E9000 support FIP Snooping.

3.6.7 Configuration Instance

# Create an FCoE instance.

```
[~CX310_2] fcoe FSB
[*CX310_2-fcoe-FSB] vlan 2094
[*CX310_2-fcoe-FSB] quit
[*CX310_2] commit
```

# Configure port roles.

```
[~CX310_2-10GE2/17/1] fcoe role vnp
[*CX310_2-10GE2/17/1] quit
[*CX310_2] commit
```

3.7 Smart Link

3.7.1 Background

Dual-uplink networking is commonly used to connect E9000 servers and the existing network system to ensure network reliability. Figure 3-35 and Figure 3-36 show two types of dual-uplink networking.

Figure 3-35 CX switch module dual-uplink networking 1

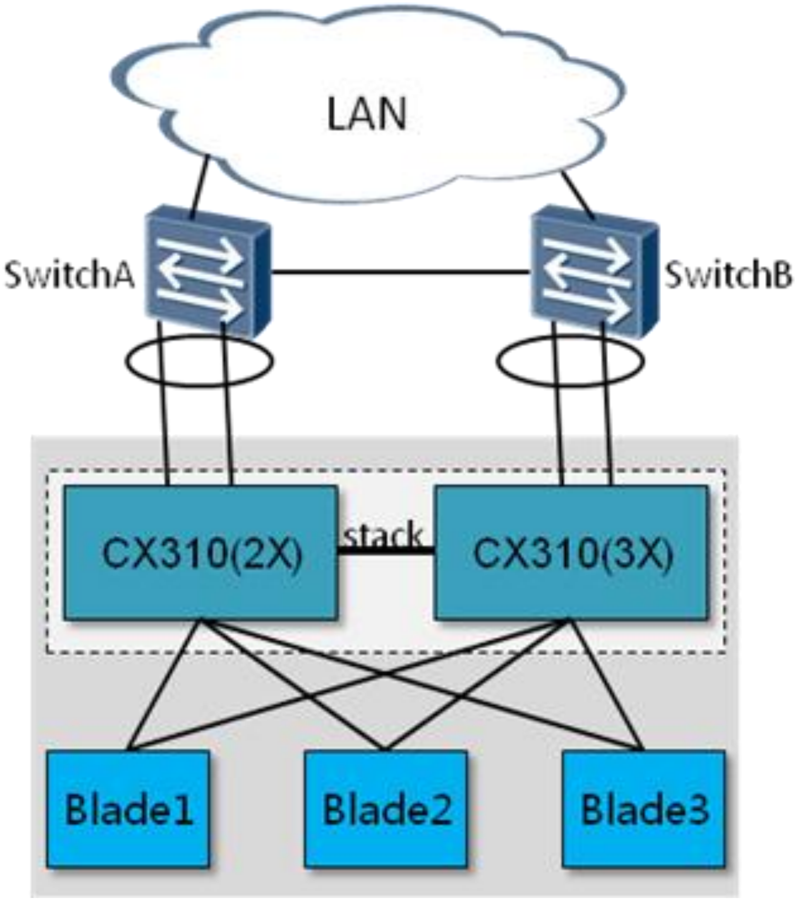
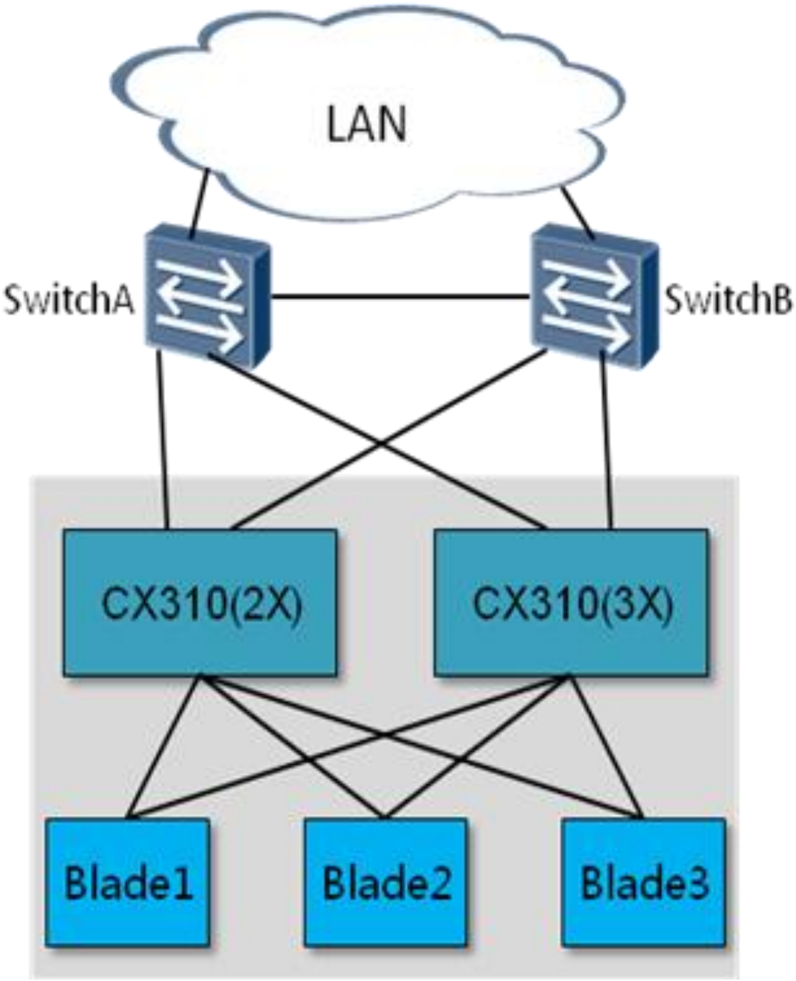


Figure 3-36 CX switch module dual-uplink networking 2



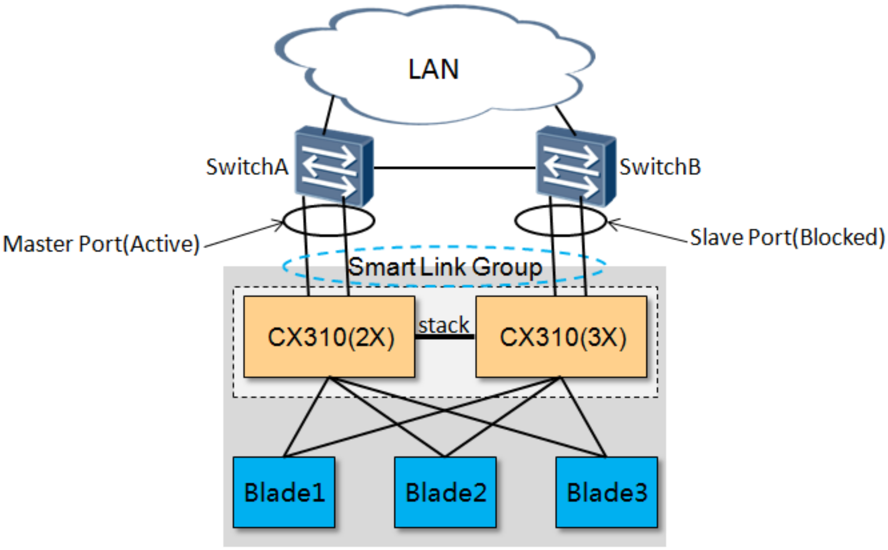
The dual-uplink networking, however, creates a loop between the cascaded access switches A and B and switch modules in slots 2X and 3X of the E9000, which may cause network broadcast storms. Generally, Spanning Tree Protocol (STP) can be used to prevent loops. However, STP convergence is long and a large amount of traffic is lost during the convergence. Therefore, STP cannot be applied to the network that demands short convergence time. In addition, STP cannot be directly used for Cisco network devices due to Cisco proprietary protocol Per-VLAN Spanning Tree Plus (PVST+). To address these issues, Huawei has proposed the Smart Link solution.

3.7.2 Smart Link Basic Concepts

3.7.2.1 Smart Link Group

A Smart Link group consists of two member interfaces. One is the master port, and the other one is the slave port. Typically, one port is active, and the other one is blocked and in the standby state. When a link fault occurs on the port in the active state, the Smart Link group automatically blocks the port, and the previous blocked port in the standby state switches to the active state. Link faults mainly refer to that a port becomes **Down** or an Ethernet operation, administration, and maintenance (OAM) link fault occurs.

Figure 3-37 Smart Link group



3.7.2.2 Master Port

The master port is a port role in a Smart Link group specified by using the command line. It can be an Ethernet interface (an electrical or optical interface) or aggregation interface.

3.7.2.3 Slave Port

The slave port is another port role in a Smart Link group specified by using the command line. It can be an Ethernet interface (an electrical or optical interface) or aggregation interface. The link of the slave port is also known as the standby link.

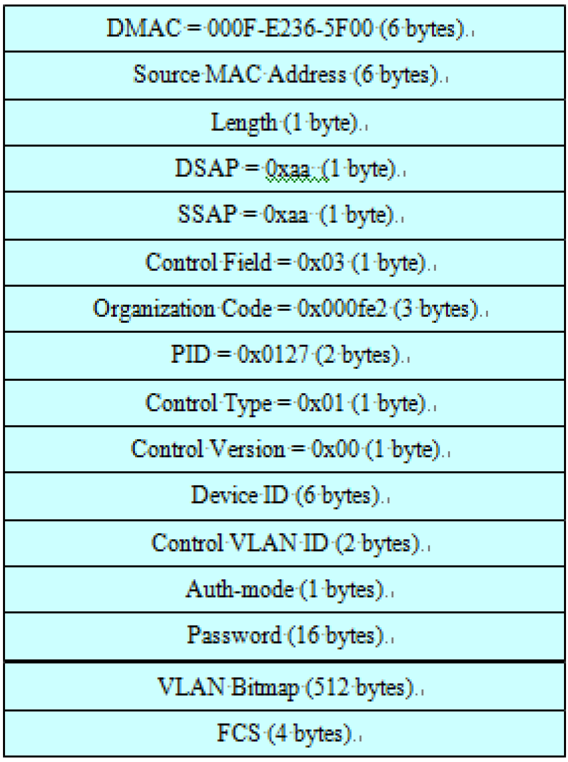
3.7.2.4 Control VLAN

Control VLANs consist of transmit control VLANs and receive control VLANs. A transmit control VLAN is used by a Smart Link group to broadcast Flush packets. A receive control VLAN is used by upstream devices to receive and process Flush packets.

3.7.2.5 Flush Packet

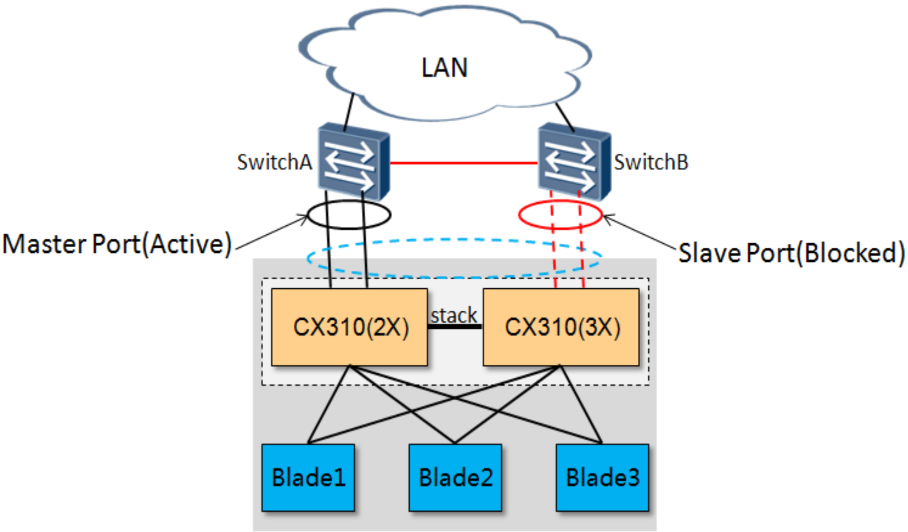
When a failover occurs between links of a Smart Link group, the original forwarding entries no longer apply to the new topology. All MAC address forwarding entries and ARP entries on the network need to be updated. The Smart Link group notifies other devices to refresh the entries by sending Flush packets.

Flush packets are encapsulated using IEEE 802.3, including information fields, such as the destination MAC address, source MAC address, control VLAN ID, and VLAN bitmap. Shows the Flush packet format.



DMAC = 000F-E236-5F00 (6 bytes)
Source MAC Address (6 bytes)
Length (1 byte)
DSAP = 0xaa (1 byte)
SSAP = 0xaa (1 byte)
Control Field = 0x03 (1 byte)
Organization Code = 0x000fe2 (3 bytes)
PID = 0x0127 (2 bytes)
Control Type = 0x01 (1 byte)
Control Version = 0x00 (1 byte)
Device ID (6 bytes)
Control VLAN ID (2 bytes)
Auth-mode (1 bytes)
Password (16 bytes)
VLAN Bitmap (512 bytes)
FCS (4 bytes)

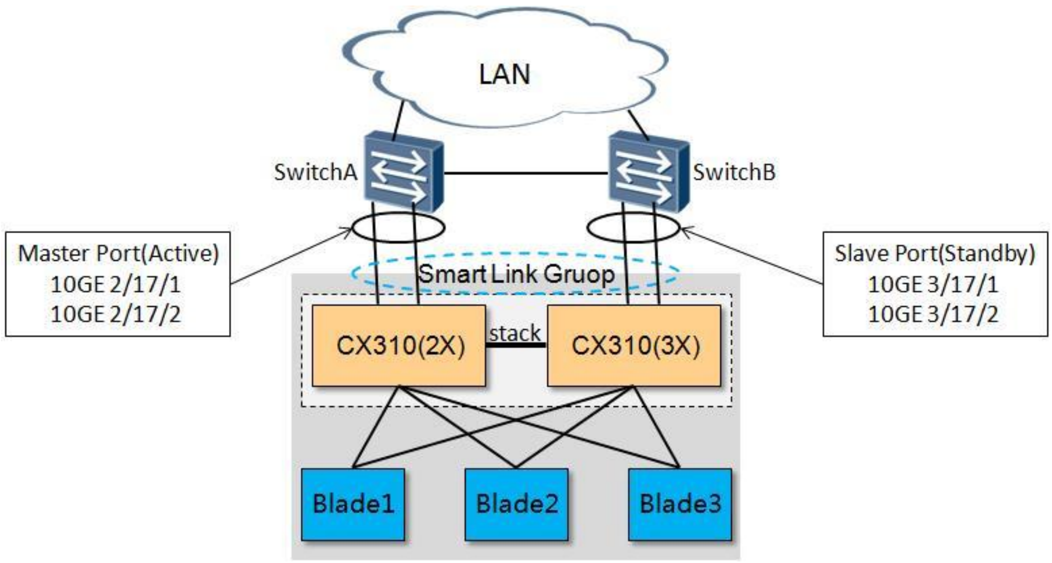
3.7.3 Smart Link Working Mechanism



When all links and devices are running properly, the links between the switch module in slot 2X and switch A are active links and forward service data. The links between the switch module in slot 3X and switch B are blocked. When switch A is faulty (a VRRP switchover is required) or the links connected to the switch module in slot 2X are faulty, the links between the switch module in slot 3X and switch B change to the active state and forward data. That is, data is forwarded along the red lines shown in the figure.

As Flush packets are Huawei proprietary protocol packets, only when upstream devices are Huawei switches and support Smart Link features, traffic switchovers can be performed through Flush packets. Otherwise, a forwarding path switchover is performed using the uplink and downlink traffic, or the system can only wait until MAC address entries are aged.

3.7.4 Configuration Instance

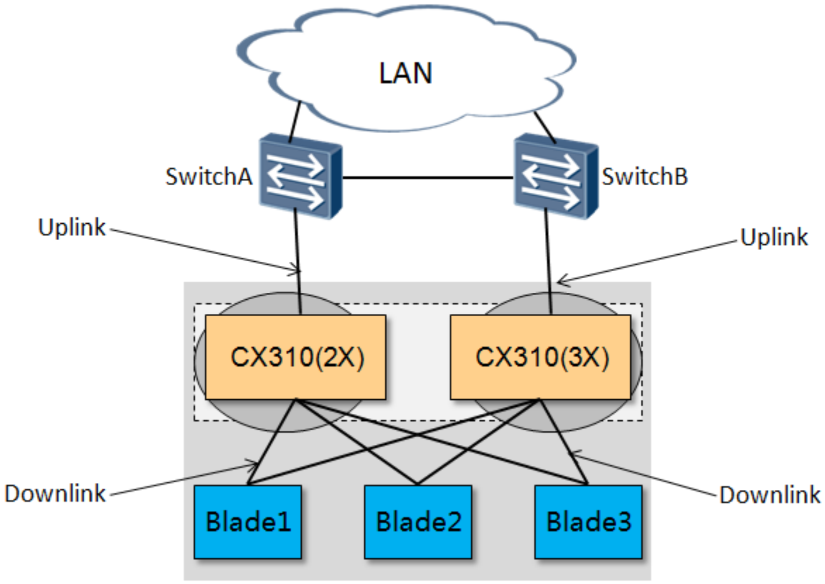


```
# Cascade the CX310s in slots 2X and 3X. (For details, see section 3.1.6
&quot;Configuration Instance&quot;.)

# Create an Eth-Trunk interface, add the first and second ports to the Eth-Trunk, and
disable STP.
[~CX310_C] interface Eth-Trunk 2
[*CX310_C-Eth-Trunk2] mode lacp-static
[*CX310_C-Eth-Trunk2] stp disable
[*CX310_C-Eth-Trunk2] trunkport 10GE 2/17/1 to 2/17/2
[*CX310_C-Eth-Trunk2] quit
[*CX310_C] interface Eth-Trunk 3
[*CX310_C-Eth-Trunk3] mode lacp-static
[*CX310_C-Eth-Trunk3] stp disable
[*CX310_C-Eth-Trunk3] trunkport 10GE 3/17/1 to 3/17/2
[*CX310_C-Eth-Trunk3] quit
[*CX310_C] commit

# Create a Smart Link group, and set Eth-Trunk 2 as the master port and Eth-Trunk 3
as the slave port.
[~CX310_C] smart-link group 1
[*CX310_C-smlk-group1] port Eth-Trunk 2 master
[*CX310_C-smlk-group1] port Eth-Trunk 3 slave
[*CX310_C-smlk-group1 quit
[*CX310_C] commit
```

3.8 Monitor Link

3.8.1 Monitor Link Overview

A Monitor Link group consists of one or more uplink and downlink ports. The downlink port status changes based on the status of uplink ports.



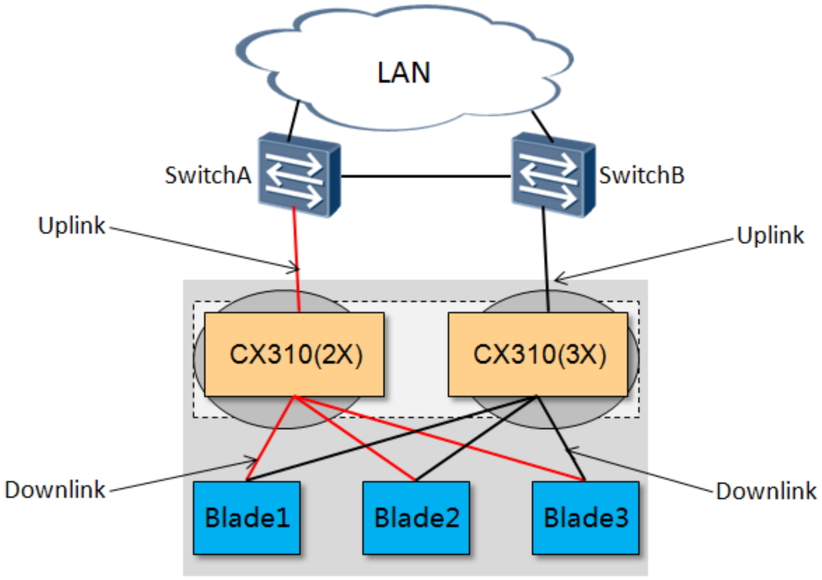
3.8.1.1 Uplink Ports

An uplink port is a monitored object specified by using the command line. The uplink port of a Monitor Link group can be an Ethernet port (an electrical or optical port), aggregation port, or a Smart Link group. If multiple ports are configured as uplink ports of a Monitor Link group, the Monitor Link group status is **Up** if any of the ports is forwarding data. If all uplink ports are faulty, the Monitor Link group status changes to **Down** and all downlink ports will be shut down.

3.8.1.2 Downlink Ports

The downlink port, another port role specified by using the command line, monitors the uplink port in the same Monitor Link group. The downlink port of a Monitor Link group can be an Ethernet port (an electrical or optical port) or aggregation port.

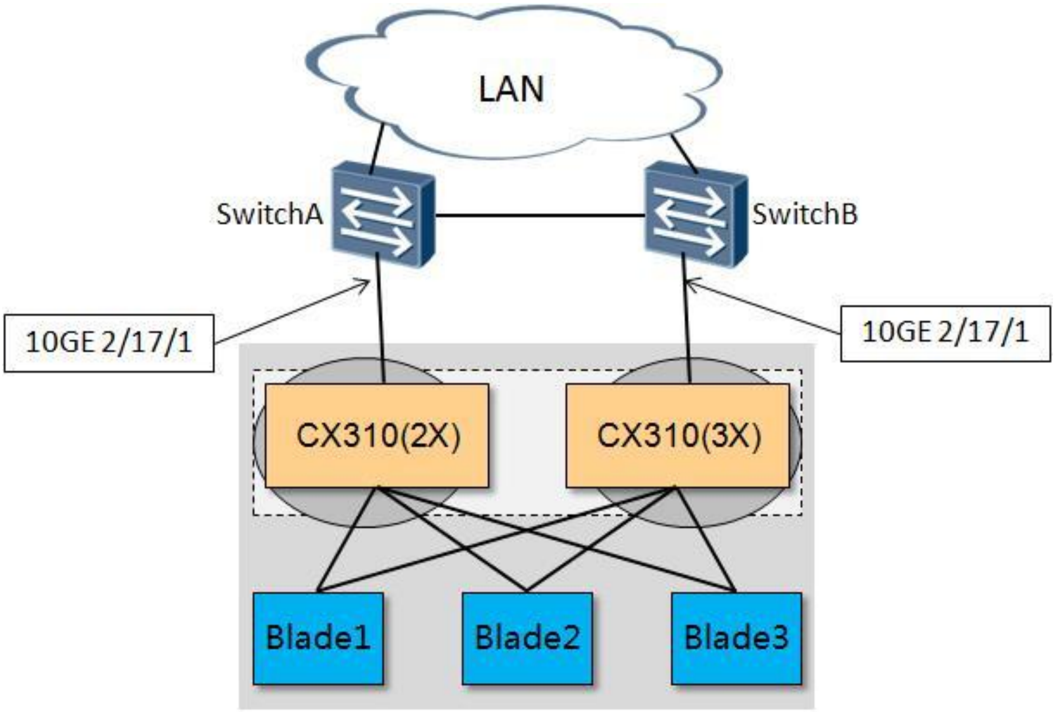
3.8.2 Monitor Link Working Mechanism



As shown in the preceding figure, the CX310s in slots 2X and 3X are configured with one Monitor Link group respectively, to monitor uplinks to switch A and switch B. Blades 1, 2, and 3 are connected to two 10GE ports of CX310s in slots 2X and 3X separately, working in active/standby mode (or in VM-based load-sharing mode). If the link between the CX310 in slot 3X and switch B is faulty, the Monitor Link group in the CX310 shuts down all downlink ports of this group (ports connected to the blade servers). When the blade servers detect the port faults, service data is switched to the link between the CX310 in slot 2X and switch A (through the red lines shown in the figure) to synchronize uplink status with downlink status.

3.8.3 Configuration Instance

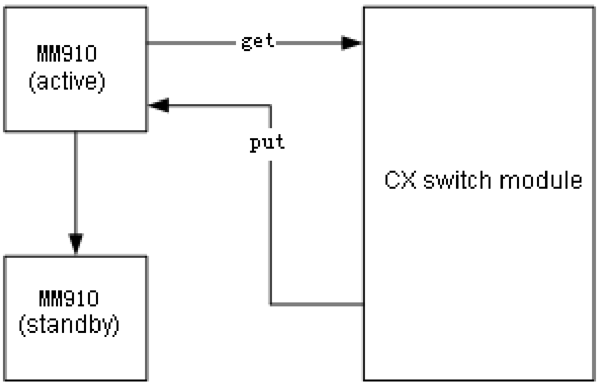
The CX310s switch modules in slots 2X and 3X connect to access switches A and B respectively. The CX310s work independently. Create a Monitor Link group to monitor the links to switch A and switch B, respectively. Set the ports connected to blades 1, 2, and 3 as the downlink ports of the Monitor Link groups.



```
# Create a Monitor Link group, and add uplink and downlink ports to the group. (The
configuration is the same for the CX310s in slots 2X and 3X.)
[~CX310_2] interface 10GE2/17/1
[~CX310_2-10GE2/17/1] stp disable
[*CX310_2-10GE2/17/1] quit
[*CX310_2] monitor-link group 1
[*CX310_2-mtlk-group1] port 10GE2/17/1 uplink
[*CX310_2-mtlk-group1] port 10GE2/1/1 downlink 1
[*CX310_2-mtlk-group1] port 10GE2/2/1 downlink 2
[*CX310_2-mtlk-group1] port 10GE2/3/1 downlink 3
[*CX310_2-mtlk-group1] quit
[*CX310_2] commit
```

3.9 Configuration Restoration

When a newly installed switch module starts, it obtains the configuration file of the switch module from the active MM910. After the switch module is started, it uses the configuration file to update configuration information. This simplifies configuration during blade hardware replacement. The following figure shows how the configuration file is transmitted between modules.



If the switch module fails to obtain the configuration file from the active MM910 during the startup process, an error message is displayed over the serial port (or SOL).

4 E9000 Networking

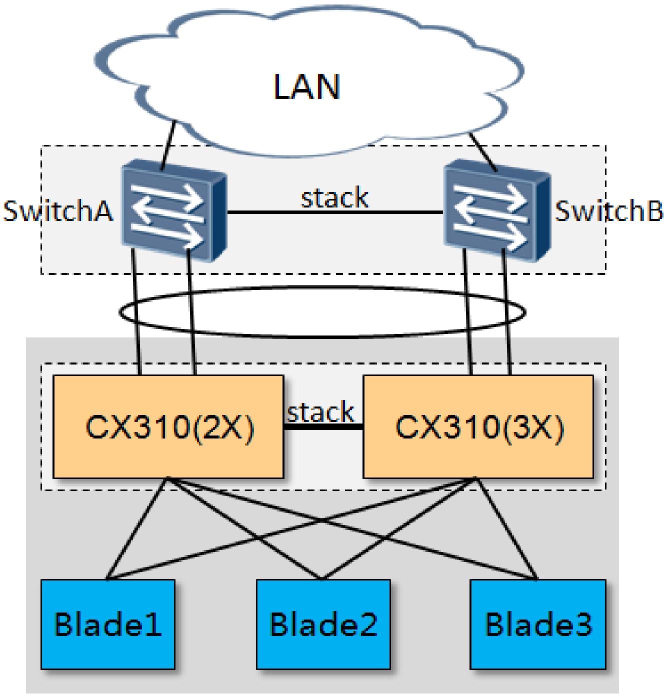
This section describes common networking modes of the E9000 in the actual application process, including the Ethernet, FC, and FCoE converged networking, and Cisco switch proprietary protocol networking scheme.

4.1 Ethernet Networking

4.1.1 Stack Networking

The switch modules in slots 2X and 3X or in slots 1E and 4E are stacked as a logical switch. The switch modules in slots 2X and 3X or in slots 1E and 4E must be of the same type. The ports on the switch modules are connected to switch A and switch B that are also stacked. The links between the two logical switches are configured with (manual or static LACP) link aggregation. Blade servers connect to switch module NICs, which can be bound in active/standby or load-sharing mode as required. In this way, the networking between one server and two switches is simplified. #EN-US_TOPIC_0009789517/fig5625303010439 shows the networking.

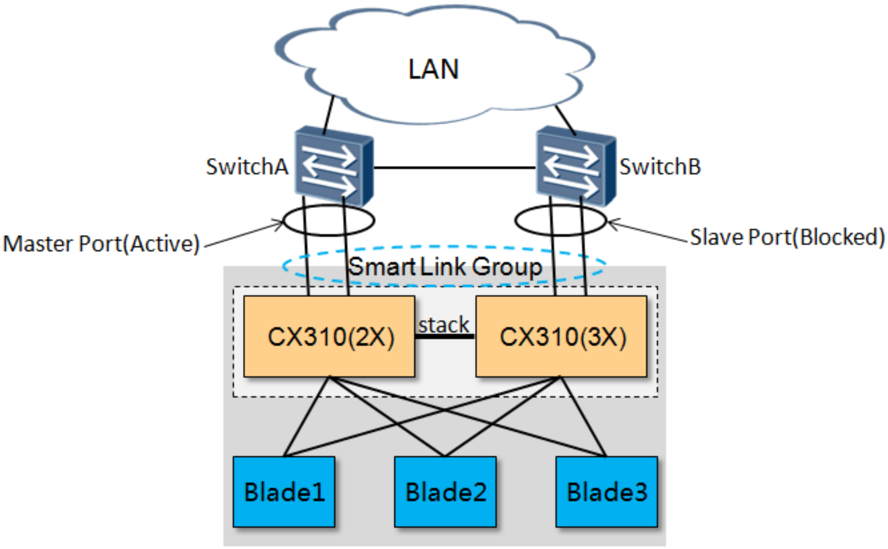
Figure 4-1 Stack networking



4.1.2 Smart Link Networking

In stack networking, the uplink switches A and B must also be stacked. In actual networking, however, the existing switches may not support stacking. In this case, you can use Smart Link, which allows the E9000 to connect to the existing network without any modification. Figure 4-2 shows the Smart Link networking architecture.

Figure 4-2 Smart Link networking architecture

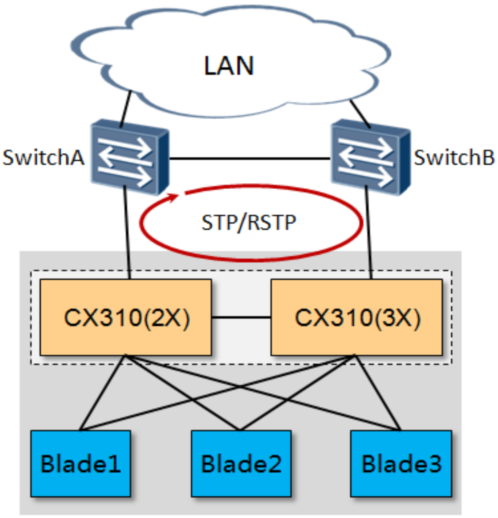


For details about the Smart Link networking, see section 3.6. Smart Link prevents a loop between multiple switches by blocking the standby links. It also eliminates the modification on the access switches A and B.

4.1.3 STP/RSTP Networking

Activate the 40GE link between switch modules in slots 2X and 3X as a cascading link. Connect the switch modules to uplink access switches A and B. Enable STP or Rapid Spanning Tree Protocol (RSTP) on uplink ports and cascading ports to prevent a loop between the switch modules and access switches.

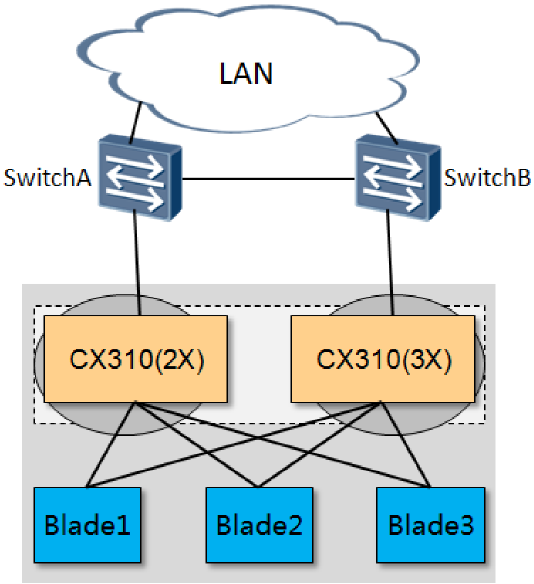
Figure 4-3 STP/RSTP networking



4.1.4 Monitor Link Networking

Deactivate the 40GE link between switch modules in slots 2X and 3X. (The 40GE link is disabled by default.) Create a Monitor Link group for each switch module. The physical or aggregation links between the switch modules and switch A B are uplinks. The links between the switch modules and blade servers are downlinks. Configure NIC teaming for the blades. If the uplinks in a Monitor Link group are unavailable, the downlinks are blocked. The NIC switches over services to the other Monitor Link group. Figure 4-4 shows the Monitor Link networking architecture.

Figure 4-4 Monitor Link networking



4.2 Networking with Cisco Switches (PVST+)

4.2.1 Cisco PVST+ Protocol

Cisco switches support the following spanning tree protocols: Per-VLAN Spanning Tree (PVST), PVST+, Rapid-PVST+, Multiple Instance Spanning Tree Protocol (MISTP), and Multiple Spanning Tree (MST). Table 4-1 lists protocols supported by different series of switches.

Table 4-1 Protocols supported by Catalyst and Nexus series switches

	PVST+	Rapid-PVST+	MST
Catalyst series switches (IOS 12.2 and later)	Y	Y	Y
Nexus series switches	N	Y	Y

PVST can be considered as a common STP running in each VLAN. Each VLAN has an independent STP status and a spanning tree calculated. PVST series protocols cannot interact with IEEE standard STP/RSTP/MSTP series protocols. Table 4-2 lists differences between PVST and STP/RSTP/MSTP frames.

Table 4-2 Differences between PVST and STP/RSTP/MSTP frames

	STP/RSTP/MSTP	PVST Series Protocols
The Ethernet frame header carries VLAN ID	N	Y

	STP/RSTP/MSTP	PVST Series Protocols
Destination MAC address	01-80-C2-00-00-00. This MAC address is the bridge protocol data unit (BPDU) MAC address defined by IEEE standard 802.1D spanning tree protocols.	01-00-0C-CC-CC-CD. This MAC address is the BPDU MAC address defined by Cisco.
Protocol frame format	Frame format defined by IEEE 802.1D/ 802.1W/802.1S	Cisco-defined frame format

Cisco develops PVST+ based on PVST and Rapid-PVST+ based on PVST+. Table 4-3 describes the improvements.

Table 4-3 PVST+ evolution

Protocol	Improvement
PVST+	Enables interworking with standard STPs. Adds PortFast, UplinkFast, and BackboneFast functions.
Rapid-PVST+	Uses the RSTP mechanism to implement rapid migration.

Table 4-4 describes how interworking between PVST+ and standard STPs is implemented on different types of ports.

Table 4-4 PVST+ interoperability with standard STPs

Port Type	PVST+ Interoperability with Standard STPs
Access	PVST+ allows only the BPDUs in standard STP/RSTP format to be sent over the Access port.
Trunk	The default VLAN (VLAN 1) allows two types of packets: BPDU in standard STP/RSTP format and private PVST BPDUs without tags. Private PVST BPDUs (with the destination MAC address of 01-00-0C-CC-CC-CD) are sent over other VLANs allowed.

NOTE

If the Trunk port does not allow the packets from the default VLAN to pass through, the port does not transmit standard STPs/RSTP BPDUs or private PVST BPDUs without tags.

4.2.2 Processing of PVST+ BPDUs

By default, the CX switch modules consider Cisco PVST BPDUs (the destination MAC address is **01-00-0C-CC-CC-CD**) as unknown multicast packets and broadcast the packets. The **mac-address bpdu** command can be used to set the BPDU MAC address to **01-00-0C-CC-CC-CD**. After that, the system discards PVST BPDUs.

4.2.3 Standard MSTP

MSTP is a new spanning tree protocol defined by IEEE 802.1s. Compared with STP and RSTP, MSTP has the following advantages:

1. MSTP divides a switching network into multiple domains. Each domain has multiple spanning trees that are independent from each other. MSTP uses a Common and Internal Spanning Tree (CIST) to prevent loops in the entire network topology.
2. MSTP maps multiple VLANs to one instance to reduce communication overheads and conserve resources. The topology of each MSTP instance is calculated independently (each instance has an independent spanning tree). The Traffic from different VLAN is evenly distributed by the instances.
3. MSTP provides a fast port transition mechanism similar to that used in RSTP.
4. MSTP is compatible with STP and RSTP.

MSTP and RSTP can recognize BPDUs of each other. STP cannot identify MSTP BPDUs. To implement smooth networking, MSTP is compatible with STP and provides STP-compatible mode, RSTP mode, and MSTP mode.

1. In STP-compatible mode, each port sends STP BPDUs.
2. In RSTP mode, each port sends RSTP BPDUs. When a port is connected to a device running STP, the port transits to the STP-compatible mode automatically.
3. In MSTP mode, each port sends MSTP BPDUs. When a port is connected to a device running STP, the port transits to the STP-compatible mode automatically.

A device working in RSTP or MSTP mode can transit to the STP-compatible mode automatically, but a device working in STP-compatible mode cannot transit to the RSTP or MSTP mode automatically. To change the working mode from STP-compatible to RSTP or MSTP, perform the mCheck operation. If a port of an MSTP or RSTP device on a switching network is connected to an STP device, the port automatically transits to the STP-compatible mode. If the STP device is removed, the port cannot automatically transit back to MSTP or RSTP mode. You can perform the mCheck operation to forcibly transit the port to MSTP or RSTP mode. In STP-compatible or RSTP mode, multiple instances can be configured and the state of each port of MSTPs is consistent with that of the CIST. To reduce loads on CPUs, do not configure multiple instances in STP-compatible or RSTP mode.

4.2.4 Difference Between Cisco and Huawei MSTPs

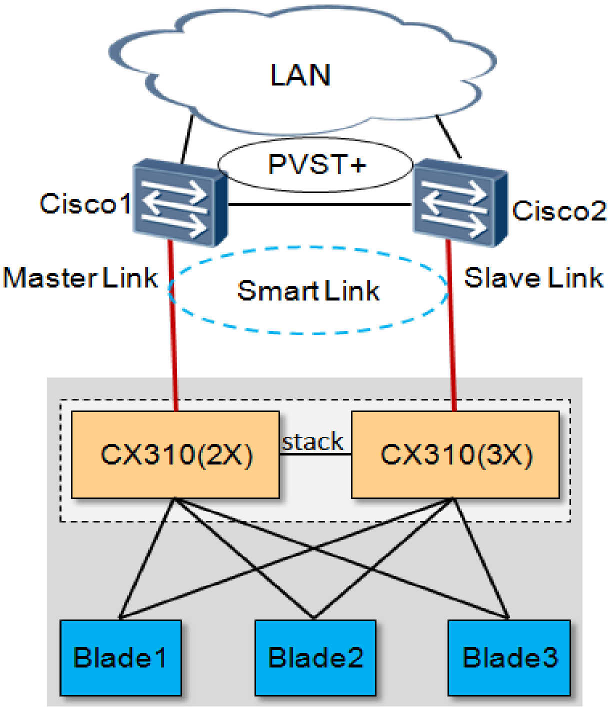
Cisco MST is a standard MSTP protocol. MST BPDUs use the standard format defined by IEEE. Huawei and Cisco switches use different keys to generate MSTP digests in BPDUs. By default, MSTP and Cisco MST can implement only inter-domain interoperation because Huawei and Cisco switches generate different digests. To implement communication between MSTP and Cisco MST within an MPST domain, enable the digest snooping function for the ports on the Huawei and Cisco switches.

4.2.5 Interconnection Scheme

4.2.5.1 Smart Link Group Interconnecting with Cisco PVST+ Network

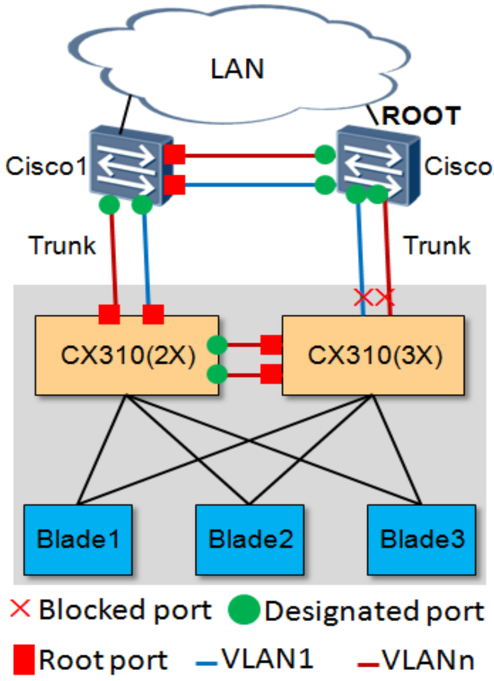
Stack the switch modules in slots 2X and 3X or in slots 1E and 4E. Connect the stacked switch modules to the Cisco PVST+ network through two uplinks. Configure the two uplinks as one Smart Link group, enable one link (Master) to forward data, and block the other link (Slave). All packets, including Cisco PVST BPDUs, received on the slave link are discarded. Loops are prevented in this way. Service data sent from blade servers may pass through the stack link between slots 2X and 3X. Therefore, configure sufficient bandwidths for the stack links.

Figure 4-5 Smart Link group interconnecting with Cisco PVST+ network



4.2.5.2 IEEE Standard Protocol (Root Bridge on the Cisco PVST+ Network Side)

Figure 4-6 Interconnecting with a Cisco PVST+ network through MSTP

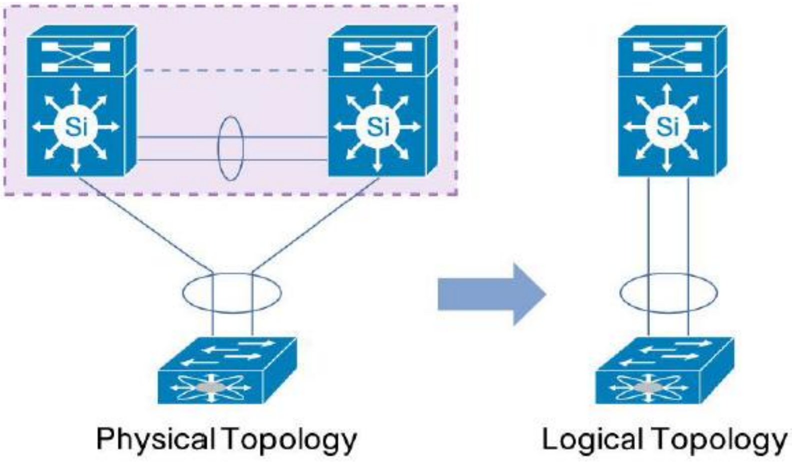


CX switch modules are connected to the Cisco PVST+ network over MSTP, and interconnection ports automatically switch to the RSTP mode. To ensure that root ports are on the Cisco PVST+ network side, the CX switch modules and Cisco switches must be configured with proper cost values and priorities (the bridge priority for VLAN 1 should be higher than that of Huawei CST). Ensure that root bridges are on the Cisco switches and blocked ports of VLAN 1 are on the CX switch ports. The CX switch modules also block other VLAN packets. Therefore, block points of other VLANs of the Cisco PVST+ network are on the same port of the CX switch module.

4.3 Cisco vPC Interoperability

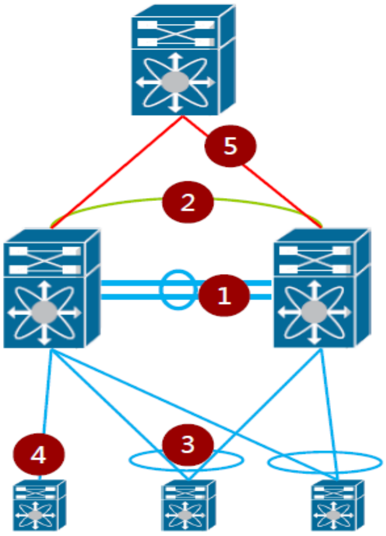
Virtual Port Channel (vPC) implements network virtualization of Cisco Nexus series data center switches, allowing multiple physical links of downstream devices to connect to two different Nexus switches by using link aggregation. Logically, the two physical switches are presented as one logical switch.

Figure 4-7 vPC network topology



vPC uses two independent control planes, whereas the VSS (stack) technology uses one control plane. Figure 4-8 shows the functional components involved in vPC.

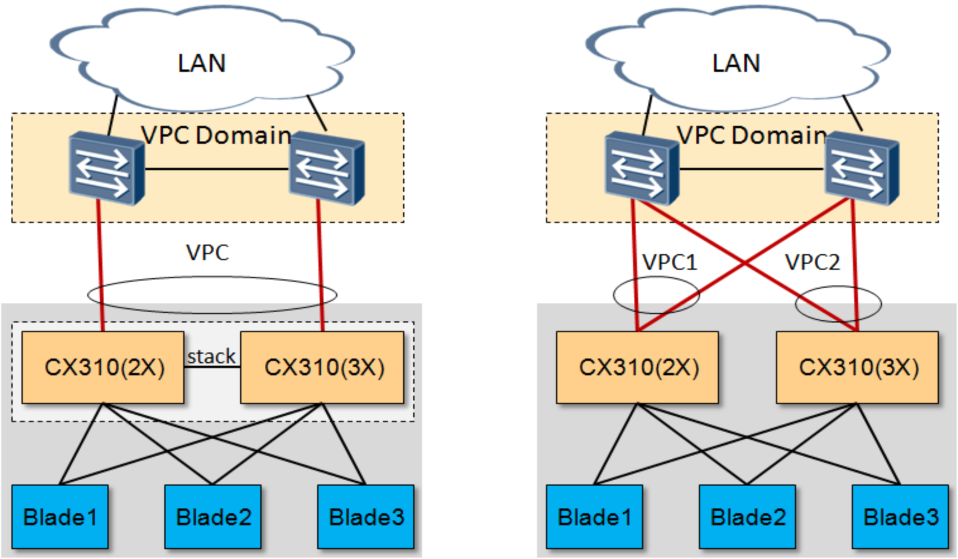
Figure 4-8 vPC functional components



- vPC peer link: synchronizes vPC device status.
- vPC peer-keepalive link: checks whether the peer vPC device is available.
- vPC member port: accesses vPC to link aggregation ports, as long as the connected device supports manual or static LACP link aggregation.
- vPC member port: accesses vPC to a single link aggregation port.
- Router: an upstream router. Independent links connect to two vPC switches, and the router selects a path over ECMP.

If the E9000 connects to the vPC network, CX switch modules are used as downstream devices, which connect to the vPC domain through L2 link aggregation ports. Figure 4-9 shows two access scenarios.

Figure 4-9 Connecting CX switch modules to the vPC domain



Stack the two CX switch modules as a logical switch and connect the switch modules to the vPC domain through the Eth-Trunk across switch modules. Alternatively, connect each of the switch module to the vPC domain through an Eth-Trunk.

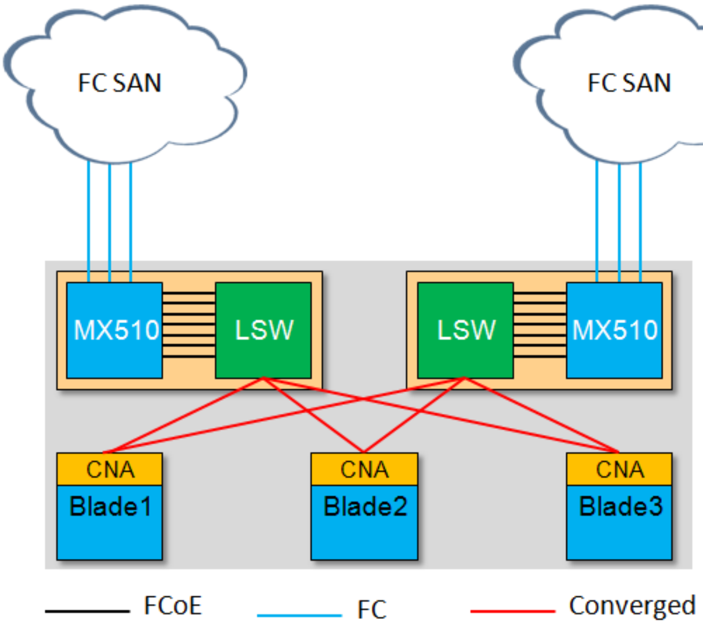
4.4 FCoE Converged Networking

4.4.1 CX311 FCoE Converged Networking

4.4.1.1 CX311 Switch Module

The CX311 converged switch module provides sixteen 10GE ports and eight 8G FC ports on the panel to connect to Ethernet devices, FC storage devices, or FC switches. The CX311 has a built-in Qlogic FCoE switch module MX510. Figure 4-10 shows the networking of CX311 converged switch modules.

Figure 4-10 CX311 converged switch module networking



The CX311 10GE Ethernet switching chips (LSWs) connect to MX510s through eight 10GE ports. FCoE traffic sent from CNAs is forwarded to the MX510s through the eight 10GE ports. The MX510s implement the conversion between FCoE and FC ports and externally connect to FC storage or switches.

The MX510 can work in Transparent (NPV) or Full Fabric (by default) mode. You can change the working mode on the MX510 CLI. Table 4-5 lists recommended working modes of the MX510 when it is connected to different devices.

Table 4-5 Working modes of the MX510

Connected Device	Model	MX510 Working Mode
FC storage device (direct connection)	NA	Full fabric
FC switch	MDS series: MDS9100/MDS9500	Transparent
	QLogic 3000/QLogic 5000/QLogic 9000	Full fabric
	FC switch: Brocade 300/Brocade 5000/Brocade 7500	Transparent

Connect the FC switches from different vendors in NPV mode to reduce interconnection risks.

4.4.1.2 Default Configuration of the CX311

The default working mode of the MX510 is full fabric. The MX510 in default working mode can directly connect to FC storage devices. FCoE features take effect only when FSB and security-related features are configured for the 10GE switching plane of the CX311. Table 4-6 describes the default FCoE configuration on the 10GE switching plane of the CX311.

Table 4-6 Default configuration

Configuration Task	Default Configuration	Command
Create an FCoE VLAN.	Create FCoE VLAN 1002, and add ports for connecting to blade servers and ports for connecting to MX510s to the VLAN. Set the former port type to Hybrid, and the latter to Trunk.	[~CX311_C] fcoe FCOE-1002 [~CX311_C-fcoe-FCOE-1002] vlan 1002
Configure port roles.	VNP (ENode-facing by default)	# Configure roles for the ports connecting to MX510s. [~CX311_C] interface 10ge 2/20/3 [~CX311_C-10GE2/20/3] fcoe role vnp
Configure DCB features.	• Enable the dcbx protocol on the ports for connecting to blade servers and ports for connecting to MX510s. The dcbx version is Intel-DCBX. • Set ETS parameters. The DRR weights of PG0 and PG1 are both 50. Queues 0, 1, 2, 4, 5, 6, and 7 belong to PG0, and queue 3 belongs to PG1. • Prevent the ports connecting to MX510s from advertising three TLVs, basic, dot1, and dot3.	# Enable the dcbx protocol. [~CX311_C-10GE2/1/1] lldp tlv-enable dcbx # Enable pfc (ports connecting to MX510s are in manual mode, and other ports are in auto mode). [~CX311_C-10GE2/1/1] dcb pfc enable mode auto [~CX311_C-10GE2/20/1] dcb pfc enable mode manual # Configure the ETS queue and bandwidth profile. [~CX311] dcb ets-profile DCBX [*CX311-ets-DCBX] priority-group 0 queue 0 to 2 4 to 7 [*CX311_C-10GE3/1/1] dcb ets enable DCBX # Set the DCBX version to Intel. [*CX311_C-10GE2/1/1] dcb compliance intel-oui # Prevent the ports connecting to MX510s from advertising some TLVs. [*CX311_C-10GE2/20/1] lldp tlv-disable basic-tlv all [*CX311_C-10GE2/20/1] lldp tlv-disable dot1-tlv all [*CX311_C-10GE2/20/1] lldp tlv-disable

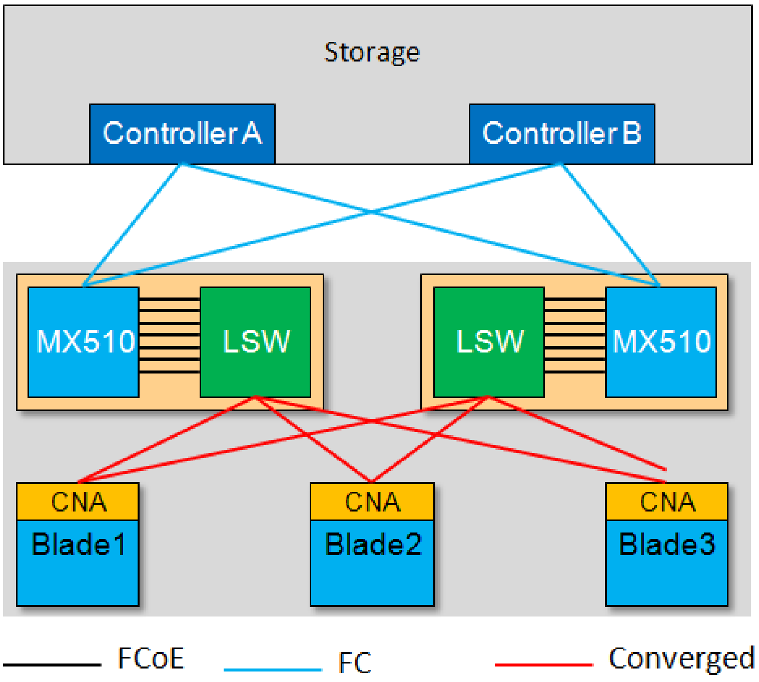
Configuration Task	Default Configuration	Command
		dot3-tlv all
Configure security features.	• Add all the ports connecting to MX510s to a port group to prevent mutual interference between Ethernet packets. Suppress outbound broadcast, unknown unicast, and multicast packets on all ports. (Only allow port */20/1 to send multicast packets with destination MAC addresses of 0110-1801-0002 and traffic less than or equal to 2 Mbps). Of all outbound packets, only FIP and FCoE packets are allowed. • Prevent panel ports from sending or receiving FIP and FCoE packets. • Prevent the ports connecting to MM910s and the 40GE ports connecting boards from sending or receiving FIP and FCoE packets.	# Apply the traffic policy FCOE-p1 to the ports connecting to the MX510s. [*CX311_C-10GE2/20/1] port-isolate enable group 1 [*CX311_C-10GE2/20/1] stp disable [*CX311_C-10GE2/20/1] storm suppression broadcast block outbound [*CX311_C-10GE2/20/1] storm suppression unknown-unicast block outbound [*CX311_C-10GE2/20/1] traffic-policy FCOE-p1 outbound # Apply the traffic policy FCOE-p11 to panel ports. [*CX311_C-10GE2/17/1] traffic-policy FCOE-p11 inbound [*CX311_C-10GE2/17/1] traffic-policy FCOE-p11 outbound Note: If the panel ports are added to the Eth-Trunk, apply the traffic policy FCOE-p11 to the Eth-Trunk interface.
Configure the MX510.	Create FCoE VLAN 1002 and add to the eight ports for connecting LSWs to MX510s to the VLAN.	FCoE_GW (admin-config): admin> vlan 1002 create FCoE_GW(admin-config): admin>vlan 1002 add port 12 FCoE_GW(admin-config): admin>vlan 1002 add port 13 FCoE_GW(admin-config):admin> vlan 1002 add port 14 FCoE_GW(admin-config): admin>vlan 1002 add port 15 FCoE_GW(admin-config): admin>vlan 1002 add port 16

Configuration Task	Default Configuration	Command
		FCoE_GW(admin-config): admin>vlan 1002 add port 17 FCoE_GW(admin-config): admin>vlan 1002 add port 18 FCoE_GW(admin-config): admin>vlan 1002 add port 19

The CX311 of default configurations (non-stacking) can directly connect to FC storage devices. To connect the CX311 to FC switches, change the MX510 working mode to Transparent or full fabric based on the type of the connected FC switch.

4.4.1.3 Connecting MX510s to Storage Devices

Figure 4-11 Connecting MX510s to storage devices

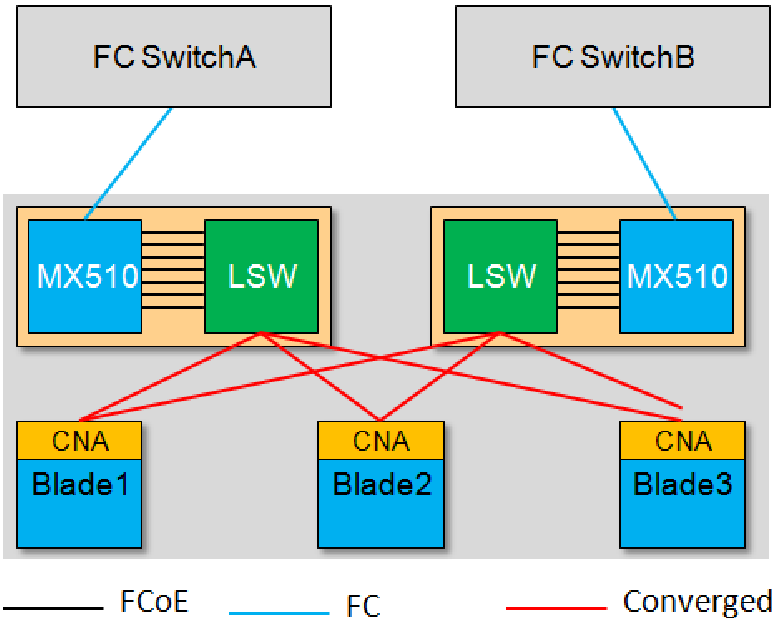


The MX510 in full fabric mode can directly connect to storage devices. Connect the FC ports on the switch module panel to both the controllers A and B of the storage device to ensure system reliability and storage performance.

4.4.1.4 Connecting MX510s to FC Switches

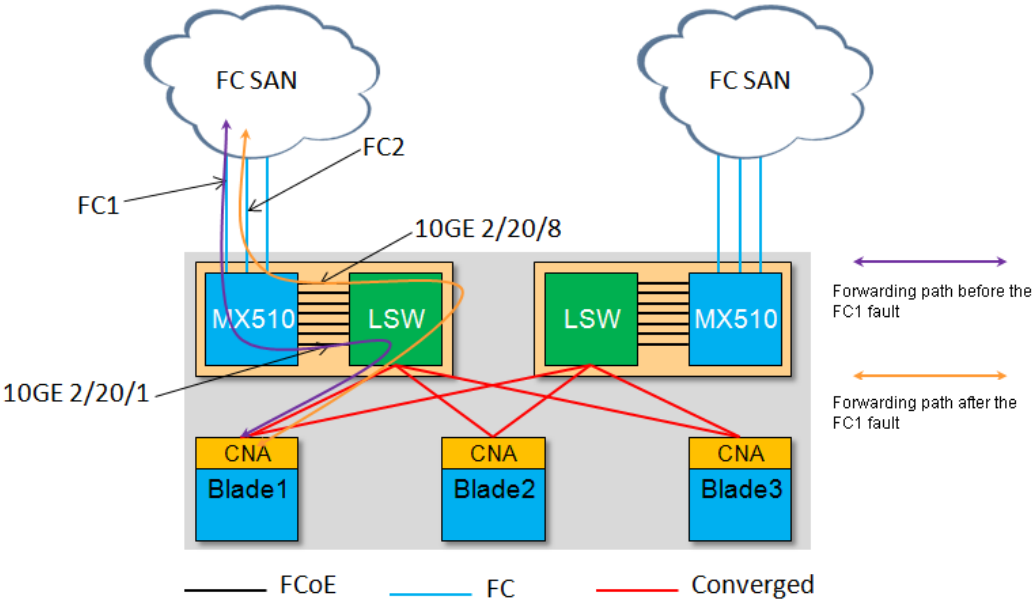
Connect each MX510 to one external FC switch through one or more FC ports on the switch module panel. Do not connect one MX510 to two FC switches. It may cause multiple paths to connect to one switch.

Figure 4-12 Connecting MX510s to FC switches



4.4.1.5 MX510 Link Load Balancing and Failover

Figure 4-13 MX510 link load balancing and failover



A CNA of a blade server provides at least two ports to connect to the switch modules in slots 2X and 3X or in slots 1E and 4E. The two ports perform the FIP VLAN discovery, FIP FCF discovery, and FIP FLOGI operations on the connected MX510 in sequence. When receiving a CNA registration request, the MX510 randomly selects a port from the eight 10GE ports that connect the MX510 to the LSW as the FCoE session port. After that, all FCoE packets are sent and received through this port. If the port is faulty, FIP Keepalive fails and the FCoE session is invalid. The CNA repeats the FIP VLAN discovery, FIP FCF discovery, and FIP FLOGI operations operations to enable the MX510 to assign a new port. After the port is assigned, a new FCoE session is set up automatically.

Similarly, the MX510 connects to an FC switch through multiple FC links. During FIP login, the MX510 randomly selects an FC link as the FCoE session transmission link. If the FC link is faulty, the CNA registers again and the MX510 reselects an FC link randomly to implement a failover. (If #EN-US_TOPIC_0009789413/fig13891893121951FC1 fails, a new port will be randomly selected from the panel FC ports and 10GE ports between LSWs and MX510s.)

4.4.1.6 CX311 in Stacking Mode

After the CX311s in slots 2X and 3X or in slots 1E and 4E are stacked, the default configuration of the CX311s changes. Only the master CX311 in the initial setup of the stack retains the default FCoE configuration. The standby CX311 does not keep the default FCoE configuration. Therefore, after the stack is set up, you must add the FCoE configuration. For example, the CX311 in slot 2X is the master module. Table 4-7 lists the FCoE configuration to be added for the stack.

Table 4-7 FCoE configuration to be added

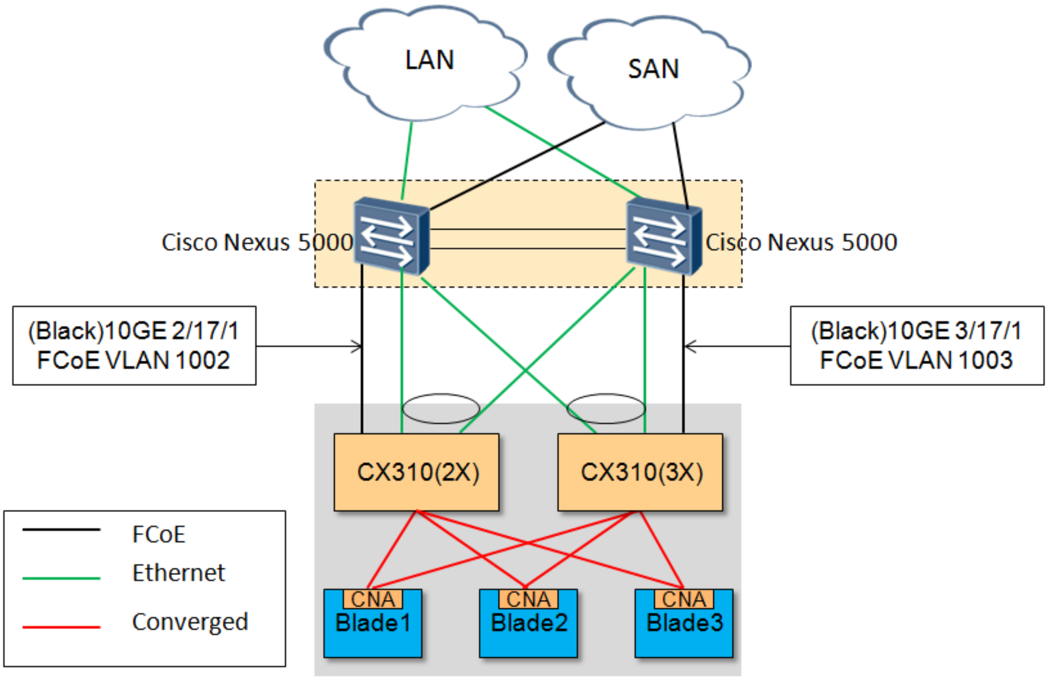
Configuration Task	Configuration Items	Main Command
Create an FCoE VLAN.	Create FCoE VLAN 1003.	[~CX311_C] fcoe FCOE-1003 [~CX311_C-fcoe- FCOE-1003] vlan 1003
Configure port roles for the MX510.	Set the port role of the CX311 in slot 3X for connecting to MX510s to VNP.	Same as the default configuration.
Configure DCB features.	Configure ports of the CX311 in slot 3X for connecting to MX510s and blade servers. Configuration items are the same.	Same as the default configuration.
Configure the MX510.	Create FCoE VLAN 1003, delete the eight ports connected to LSWs from FCoE VLAN 1002, and add the ports to FCoE VLAN 1003.	FCoE_GW(admin-config): admin> vlan 1002 remove port 12 FCoE_GW(admin-config): admin> vlan 1002 remove port 13 FCoE_GW(admin-config): admin> vlan 1002 remove port 14 FCoE_GW(admin-config): admin> vlan 1002 remove port 15 FCoE_GW(admin-config): admin> vlan 1002 remove port 16 FCoE_GW(admin-config): admin> vlan 1002 remove port 17 FCoE_GW(admin-config): admin> vlan 1002 remove port 18 FCoE_GW(admin-config):

Configuration Task	Configuration Items	Main Command
		admin> vlan 1002 remove port 19 FCoE_GW(admin-config): admin> vlan 1003 create FCoE_GW(admin-config): admin> vlan 1003 add port 12 FCoE_GW(admin-config): admin> vlan 1003 add port 13 FCoE_GW(admin-config): admin> vlan 1003 add port 14 FCoE_GW(admin-config): admin> vlan 1003 add port 15 FCoE_GW(admin-config): admin> vlan 1003 add port 16 FCoE_GW(admin-config): admin> vlan 1003 add port 17 FCoE_GW(admin-config): admin> vlan 1003 add port 18 FCoE_GW(admin-config): admin> vlan 1003 add port 19

4.4.2 Connecting CX310s to Cisco Nexus 5000 Series Switches

The CX310 converged switch module supports FSB and provides sixteen 10GE ports on the panel. The ports can be connected to external FCoE switches, which provide FC ports to connect to FC storage devices or FC switches. Figure 4-14 shows the connection between CX310s and Cisco Nexus 5000 series FCoE switches.

Figure 4-14 Connecting CX310s to Cisco Nexus 5000 series switches



The CX310s are not stacked. (If CX310s are stacked, you only need to configure the ETS template and LLDP once. Port configurations in slots 2X and 3X are the same except for the FCoE VLAN.) This section uses the CX310 in slot 2X as an example to describe how to configure the connection between the CX310 switch module and Cisco Nexus 5000 switch.

Table 4-8 Configuration items

Configuration Item	Description	Description
Create an FCoE VLAN.	Create FCoE VLAN 1002 and add ports for connecting to blade servers and FCoE ports on the panel to the VLAN. The port type is Hybrid.	[~CX310_2] fcoe FCOE-1002 [*CX310_2-fcoe-FSB] vlan 1002 [*CX310_2] interface 10GE 2/1/1 [*CX310_2-10GE2/1/1] port link-type hybrid [*CX310_2-10GE2/1/1] port hybrid tagged vlan 1002 [*CX310_2-10GE2/1/1] quit [*CX310_2] interface 10GE 2/2/1 [*CX310_2-10GE2/2/1] port link-type hybrid [*CX310_2-10GE2/2/1] port hybrid tagged vlan 1002 [*CX310_2-10GE2/2/1] quit [*CX310_2] interface 10GE 2/3/1 [*CX310_2-10GE2/3/1] port link-type hybrid [*CX310_2-10GE2/3/1] port hybrid tagged vlan 1002 [*CX310_2-10GE2/3/1] quit [*CX310_2] interface 10GE 2/17/1

Configuration Item	Description	Description
		[*CX310_2-10GE2/17/1] port link-type hybrid [*CX310_2-10GE2/17/1] port hybrid tagged vlan 1002 [*CX310_2-10GE2/17/1] quit [*CX310_2] commit
Configure port roles.	Set the type of the ports on the panel to VNP.	[*CX310_2] interface 10GE 2/17/1 [*CX310_2-10GE2/17/1] fcoe role vnp
Configure DCB features.	• Enable LLDP and PFC. • Configure the ETS queue bandwidth control template and DCBX version. • Prevent the ports connecting to Cisco Nexus 5000 switches from advertising some TLVs.	# Enable LLDP and PFC. [*CX310_2] lldp enable [*CX310_2] interface 10GE 2/1/1 [*CX310_2-10GE2/1/1] lldp tlv-enable dcbx [*CX310_2-10GE2/1/1] dcb pfc enable mode manual [*CX310_2-10GE2/1/1] quit [*CX310_2] interface 10GE 2/2/1 [*CX310_2-10GE2/2/1] lldp tlv-enable dcbx [*CX310_2-10GE2/2/1] dcb pfc enable mode manual [*CX310_2-10GE2/2/1] quit [*CX310_2] interface 10GE 2/3/1 [*CX310_2-10GE2/3/1] lldp tlv-enable dcbx [*CX310_2-10GE2/3/1] dcb pfc enable mode manual [*CX310_2-10GE2/3/1] quit [*CX310_2] interface 10GE 2/17/1 [*CX310_2-10GE2/17/1] lldp tlv-enable dcbx [*CX310_2-10GE2/17/1] dcb pfc enable mode manual [*CX310_2-10GE2/17/1] quit [*CX310_2] commit # Configure ETS parameters (ports connecting CNAs and Cisco Nexus 5000 switches) and set the DCBX version to Intel DCBX. [~CX310_2] dcb ets-profile DCBX [*CX310_2-ets-DCBX] priority-group 0 queue 0 to 2 4 to 7 [*CX310_2-ets-DCBX] quit [*CX310_2] interface 10GE 2/1/1 [*CX310_2-10GE2/1/1] dcb ets enable DCBX [*CX310_2-10GE2/1/1] dcb compliance intel-oui [*CX310_2-10GE2/1/1] quit [*CX310_2] interface 10GE 2/2/1 [*CX310_2-10GE2/2/1] dcb ets enable DCBX

Configuration Item	Description	Description
		[*CX310_2-10GE2/2/1] dcb compliance intel-oui [*CX310_2-10GE2/2/1] quit [*CX310_2] interface 10GE 2/3/1 [*CX310_2-10GE2/3/1] dcb ets enable DCBX [*CX310_2-10GE2/3/1] dcb compliance intel-oui [*CX310_2-10GE2/3/1] quit [*CX310_2] interface 10GE 2/17/1 [*CX310_2-10GE2/17/1] dcb ets enable DCBX [*CX310_2-10GE2/17/1] dcb compliance intel-oui [*CX310_2-10GE2/17/1] quit [*CX310_2] commit # Prevent the ports connecting to Cisco Nexus 5000 switches from advertising some TLVs. [~CX310_2-10GE2/17/1] lldp tlv-disable basic-tlv all [*CX310_2-10GE2/17/1] lldp tlv-disable dot1-tlv all [*CX310_2-10GE2/17/1] lldp tlv-disable dot3-tlv all [*CX310_2-10GE2/17/1] quit [*CX310_2] commit

4.4.3 Connecting CX310s to Brocade VDX6700 Series Switches

The connection between CX310s and Brocade VDX6700 series switches over FCoE links is the same as that between CX310s and Cisco Nexus 5000 series switches, except the mapping between the ETS queues and PGs. Table 4-9 describes ETS configurations.

Table 4-9 Configuration items

Configuration Item	Description	Command
Configure DCB features.	Add an ETS profile DCBXF to configure switch ports connected to the Brocade VDX6700 series switches.	# Configure ETS parameters (ports connecting to Brocade VDX6700 series switches) and the DCBXF version. [*CX310_2] dcb ets-profile DCBXF [*CX310_2-ets-DCBXF] priority-group 0 queue 0 to 2 4 to 6 [*CX310_2-ets-DCBXF] priority-group 15 queue 7 [*CX310_2-ets-DCBXF] quit [*CX310_2] interface 10GE 2/17/1 [*CX310_2-10GE2/17/1] dcb ets enable DCBXF [*CX310_2-10GE2/17/1] dcb compliance intel-oui [*CX310_2-10GE2/17/1] quit

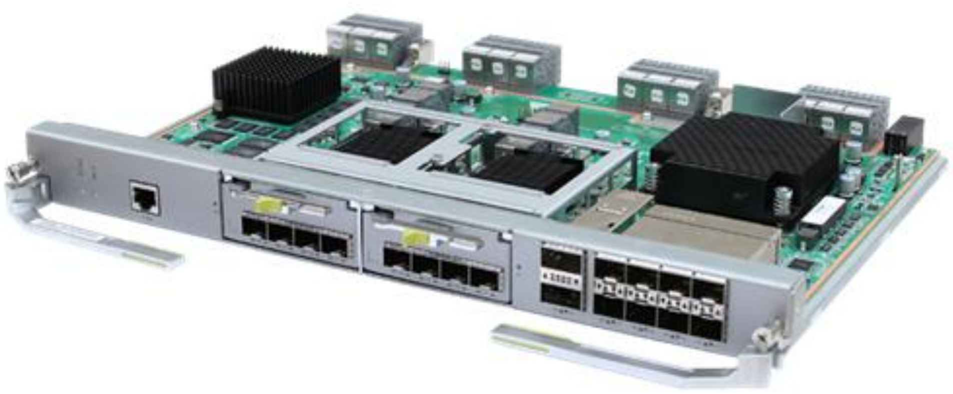
Configuration Item	Description	Command
		[*CX310_2] commit

4.4.4 CX320 Converged Networking

4.4.4.1 CX320 Converged Switch Module

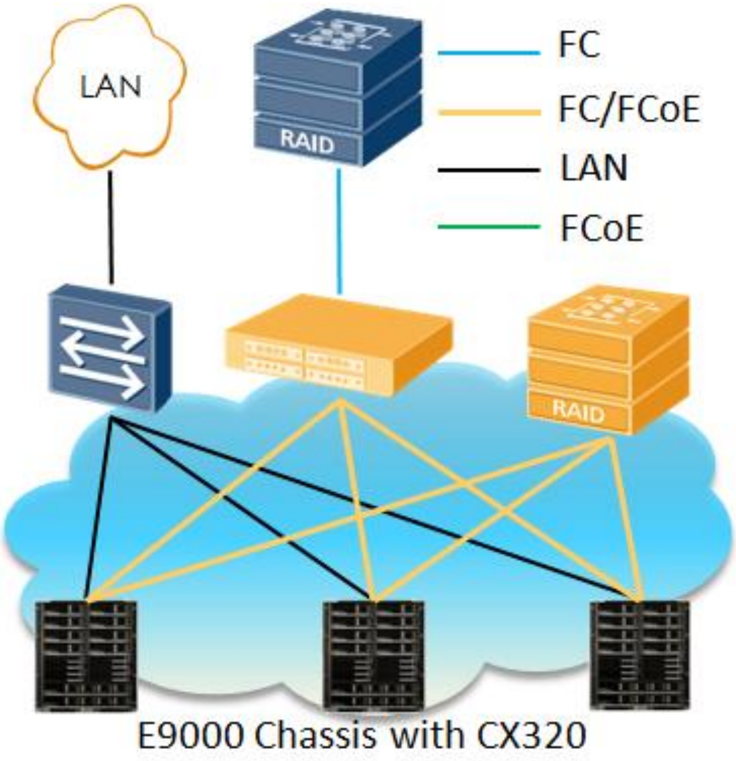
The CX320 converged switch module provides fixed ports on the panel: 8 x GE + 2 x 40GE. It also supports two flexible cards, such as the MX517 card, which provides four SFP+ Unified ports that can be configured as four 10GE or 8G FC ports. Figure 4-15 shows the CX320.

Figure 4-15 CX320 switch module



The CX320 supports three FCoE networking modes: FSB, FCF, and NPV. It is the best choice for converged networks of enterprises and carriers. The CX320 can be configured with an FCF instance to connect to FC or FCoE storage devices or configured with an NPV instance to connect to FC or FCoE switches, so that various converged networking requirements are met.

Figure 4-16 CX320 converged network



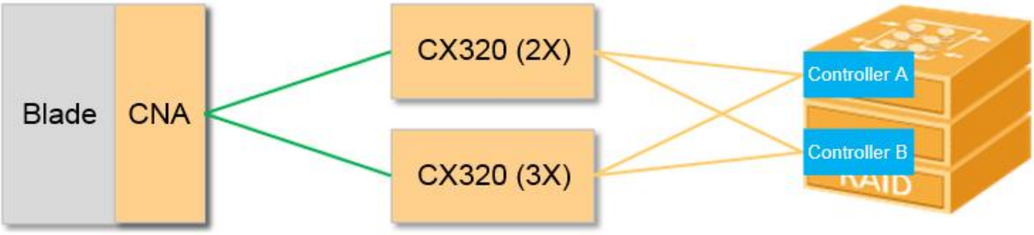
The CX320 switch modules can be installed in slots 2X and 3X or slots 1E and 4E and work with 2-port or 4-port CNAs to meet various converged application requirements. A shown in Figure 4-16, CX320 switch modules are installed in E9000 servers to implement a converged network inside the E9000 chassis and convergence of the external LAN and SAN. The MX517 flexible cards provide 8G FC ports to connect to FC switches or storage so that the existing SAN resources are fully utilized. With flexible cards, the CX320 supports evolution to 16G FC, 32G FC, and 25GE to fully protect customers' investments.

4.4.4.2 FCF Networking

- FCF network

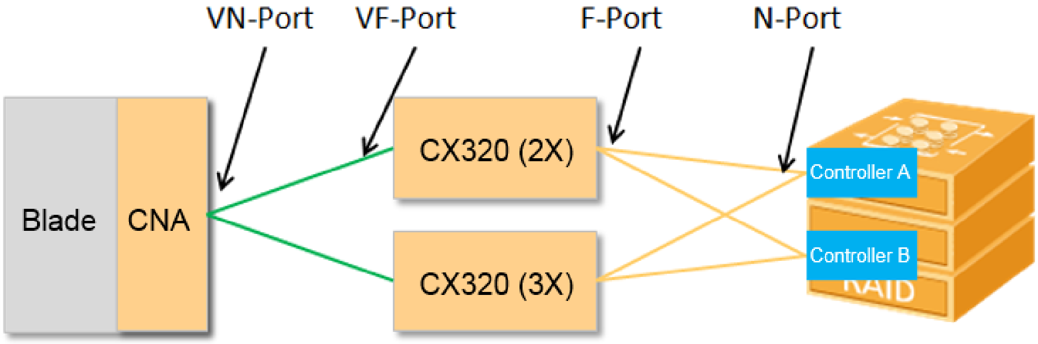
The CX320 switch modules function as FCF switches and can connect to FC and FCoE storage devices at the same time, as shown in Figure 4-17.

Figure 4-17 FCF network



The CNAs on compute nodes connect to the CX320 switch modules in slots 2X and 3X or slots 1E and 4E. Each CX320 connects to both storage controllers of the FC or FCoE storage. If one CX320 fails, the other one is still connected to the two controllers so that the number of storage controllers and storage performance are not reduced. After the CX320 switch modules are configured with FCF instances and connected to FC storage, the port roles are as shown in Figure 4-18.

Figure 4-18 Port roles in an FCF network

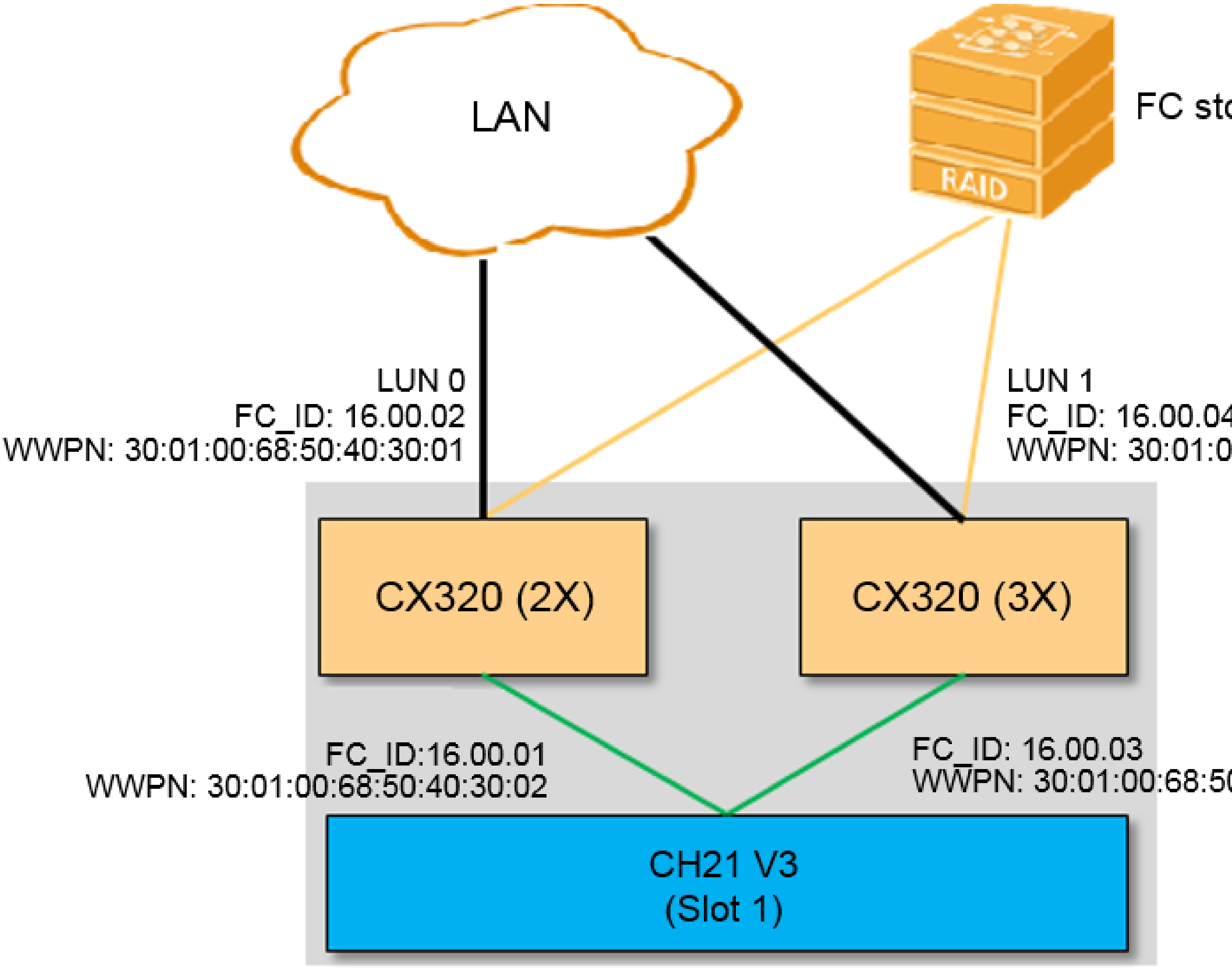


In the figure, the CX320 switch modules use VF_Ports connect to CNAs of compute nodes and use F_Ports to connect to storage devices.

- FCF network configuration example

As shown in Figure 4-19, CX320 switch modules are installed in slots 2X and 3X; an MX517 is installed in flexible card slot 1 of each CX320; the compute node is installed in slot 1; An MZ510 is installed on the compute node; each CX320 uses one 10GE link to connect to the LAN and one FC link to connect to FC storage so that network reliability is ensured.

Figure 4-19 FCF network configuration example



Step 1 Plan the configuration process.

1. Configure the MZ510 working mode to NIC + FCoE.
2. Configure the ports of the flexible card as FC ports.
3. Create an FCF instance, specify an FCoE VLAN, and add FC and FCoE ports to the FCF instance.

Step 2 Prepare data.

Prepare the following data for this example.

Table 4-10 Device information

Device	Port No.	WWPN/VLAN	FC_ID	Connected To

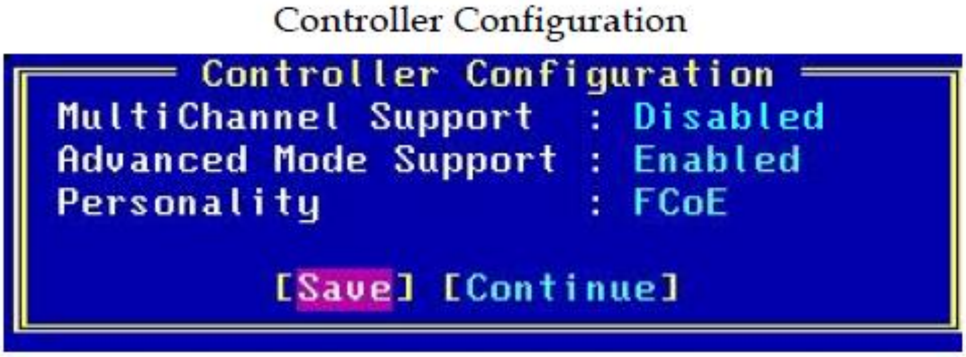
Device	Port No.	WWPN/VLAN	FC_ID	Connected To
CX320 (2X)	FC 2/21/1	-	-	FC Storage
	10GE 2/1/1	2094, 2093	-	CH121 V3
	10GE 2/17/1	2093	-	LAN
CX320 (3X)	FC 3/21/1	-	-	FC Storage
	10GE 3/1/1	2094, 2093	-	CH121 V3
	10GE 3/17/1	2093		LAN
CH121 V3	-	30:00:00:68:50:40:30:02	16.00.01	CX320(2X)
	-	30:00:00:68:50:40:30:04	16.00.03	CX320 (3X)
FC Storage		30:00:00:68:50:40:30:01	16.00.02	CX320 (2X)
		30:00:00:68:50:40:30:03	16.00.04	CX320 (3X)

Step 3 Perform the configuration.

Configure the MX510.

1. During the BIOS startup, press **Ctrl+P** when "Press <Ctrl><P> for PXESelect (TM) Utility" is displayed. The **Controller Configuration** screen shown in Figure 4-20 is displayed.

Figure 4-20 Configuring the MZ510



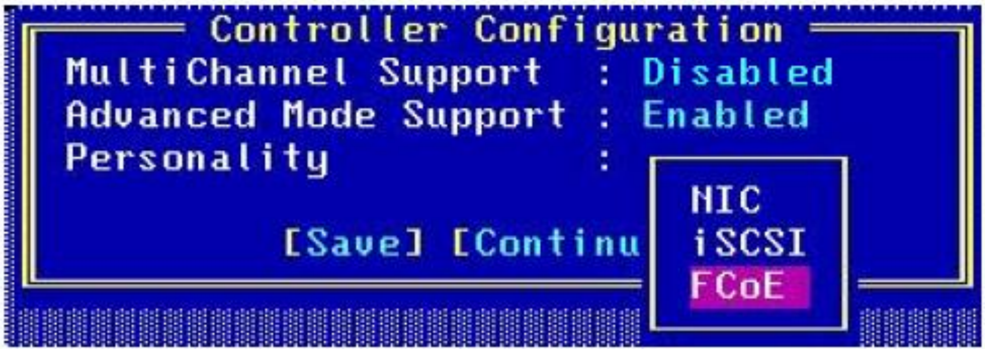
Controller Configuration

Controller Configuration

MultiChannel Support : Disabled
Advanced Mode Support : Enabled
Personality : FCoE

2. Select **Personality** by pressing arrow keys and press **Enter**. The **Personality** options shown in Figure 4-21 are displayed.

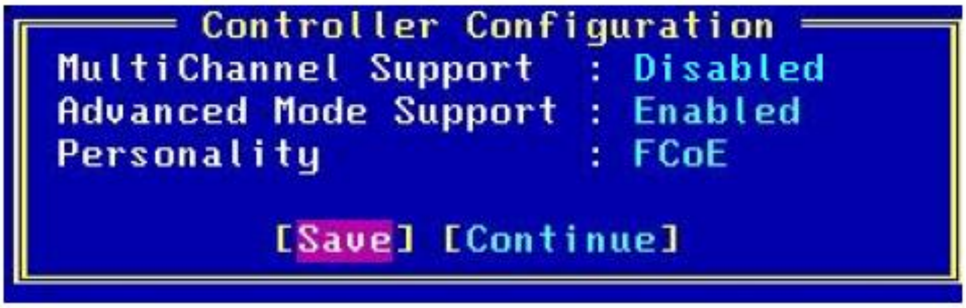
Figure 4-21 Setting the NIC working mode



```
Controller Configuration
MultiChannel Support   : Disabled
Advanced Mode Support  : Enabled
Personality            :
```

3. Select **FCoE** by pressing arrow keys and press **Enter**. Then select **Save** by pressing arrow keys and press **Enter**, as shown in Figure 4-22.

Figure 4-22 Saving the configuration



```
Controller Configuration
MultiChannel Support   : Disabled
Advanced Mode Support  : Enabled
Personality            : FCoE
```

Configure FC ports.

# Before configuring FC ports, insert FC optical modules into the ports. If the Link indicator is steady green, the port is up.

# Configure port 10GE 2/21/1 of the CX320 in slot 2X as an FC port.

```
<HUAWEI> system-view
[~HUAWEI] sysname CX320_2X
[*HUAWEI] commit
[~CX320_2X] port mode fc 10GE 2/21/1
Warning: This operation will cause all the other configurations on the port to be lost.
Continue?[Y/N]:Y
[*CX320_2X] commit
```

# Check the FC port status.

```
[~CX320_2X] display interface fc 2/21/1
FC2/21/1 current state : UP (ifindex:144)
Line protocol current state : UP
```

......

Create an FCF instance.

```
# Create VLAN 2093 for Ethernet services.

[~CX320_2X] vlan 2093

# For the CX320 in slot 2X, create an FCF instance fcf1 and specify VLAN 2094 as the FCoE VLAN of the instance.

[*CX320_2X] fcoe fcf1 fcf

[*CX320_2X-fcoe-fcf-fcf1] vlan 2094

[*CX320_2X] quit

# Create FCoE port 1.

[*CX320_2X] interface fcoe-port 1

[*CX320_2X-FCoE-Port1] quit

# Set the ports connected to CH121 compute nodes as hybrid ports and add the FCoE and Ethernet VLANs.

[*CX320_2X] interface 10GE 2/1/1

[*CX320_2X-10GE2/1/1] port link-type hybrid

[*CX320_2X-10GE2/1/1] port hybrid tagged vlan 2094

[*CX320_2X-10GE2/1/1] port hybrid tagged vlan 2093

[*CX320_2X-10GE2/1/1] fcoe-port 1

[*CX320_2X-10GE2/1/1] quit

# Set ports connected to the LAN as access ports and add the Ethernet VLAN.

[*CX320_2X] interface 10GE 2/17/1

[*CX320_2X-10GE2/17/1] port link-type access

[*CX320_2X-10GE2/17/1] port default vlan 2093

# Add port FC 2/21/1 of the CX320 in slot 2X to the fcf1 instance.

[*CX320_2X] fcoe fcf1

[*CX320_2X-fcoe-fcf-fcf1] member interface fc 2/21/1

[*CX320_2X-fcoe-fcf-fcf1] member interface fcoe-port 1

[*CX320_2X] commit
```

Verify the configuration.

```
# View registration information of servers and storage.

[~CX320_2X] display fcoe name-server brief

The Name-Server Information:

-------------------------------------------------------------------------------
```

```
Interface              FC-ID              WWPN
-----------------------------------------------------------------------------
FCoE-Port1          16.00.01          30:00:00:68:50:40:30:02
FC2/21/1              16.00.02          30:00:00:68:50:40:30:01
-----------------------------------------------------------------------------
Total: 2
```

Configure CX320_3X.

The configuration procedure is the same as that of the CX320 in slot 2X, except that the FC port numbers and the numbers of the ports connected to compute nodes.

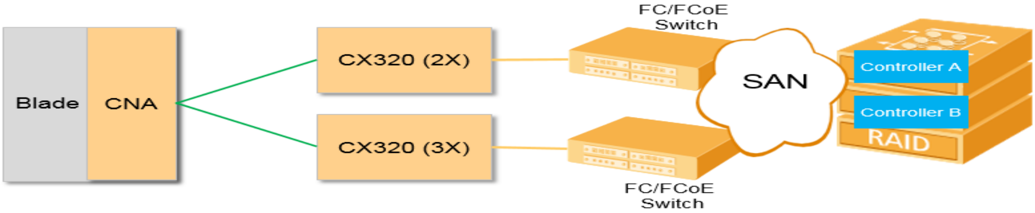
----End

4.4.4.3 NPV Networking

- NPV network

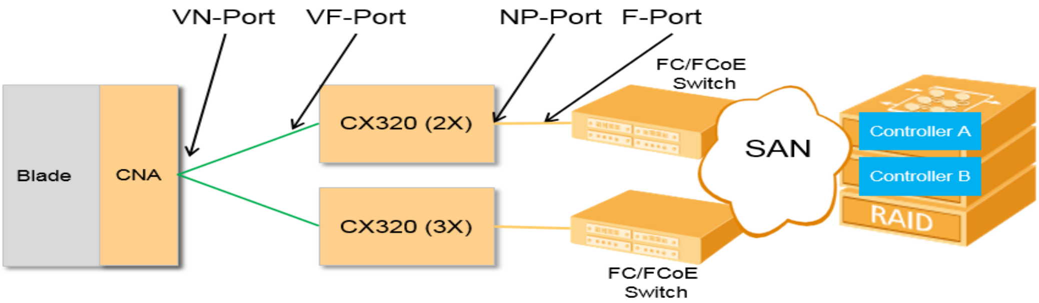
Functioning as an NPV switch, the CX320 is located at the edge of the SAN fabric network and between node devices and FCF switches, as shown in Figure 4-23.

Figure 4-23 NPV network



The CNA on a compute node connects to the CX320 switch modules in slots 2X and 3X or slots 1E and 4E. Each CX320 connects to an FC or FCoE switch, which is connected to storage through a SAN. After the CX320 switch modules are configured with FCF instances and connected to FC switches, the port roles are as shown in Figure 4-24.

Figure 4-24 Port roles in an NPV network

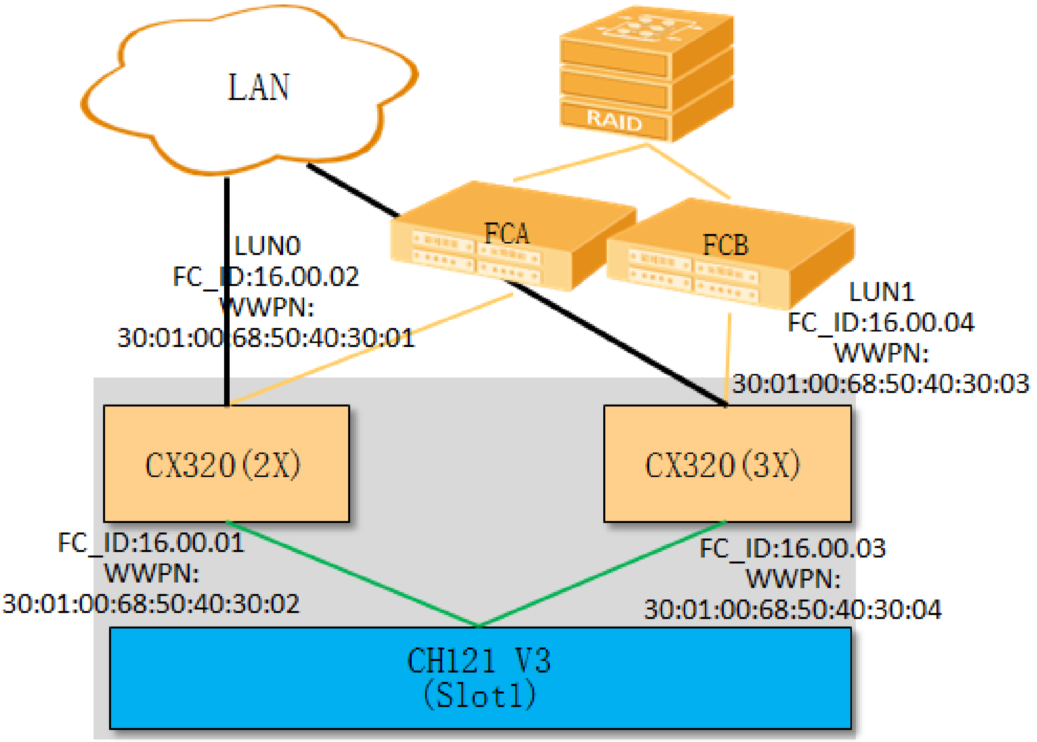


In the figure, the NPV instances of the CX320 switch modules use VF_Ports to connect to CNAs of compute nodes and use NP_Ports to connect to FC switches.

- NPV network configuration example

As shown in Figure 4-25, CX320 switch modules are installed in slots 2X and 3X; an MX517 is installed in flexible card slot 1 of each CX320; the compute node is installed in slot 1; an MZ510 is installed on the compute node; each CX320 uses one 10GE link to connect to the LAN and one FC link to connect to an FC switch so that network reliability is ensured.

Figure 4-25 NPV network configuration example



Step 1 Plan the configuration process.

1. Configure the MZ510 working mode to NIC + FCoE.
2. Configure the ports of the flexible card as FC ports.
3. Create an NPV instance, specify an FCoE VLAN, and add FC and FCoE ports to the NPV instance.

Step 2 Prepare data.

Prepare the following data for this example.

Table 4-11 Device information

Device	Port No.	WWPN/VLAN	FC_ID	Connected To
CX320 (2X)	FC 2/21/1	-	-	FCA
	10GE 2/1/1	2094, 2093	-	CH121 V3
	10GE 2/17/1	2093	-	LAN

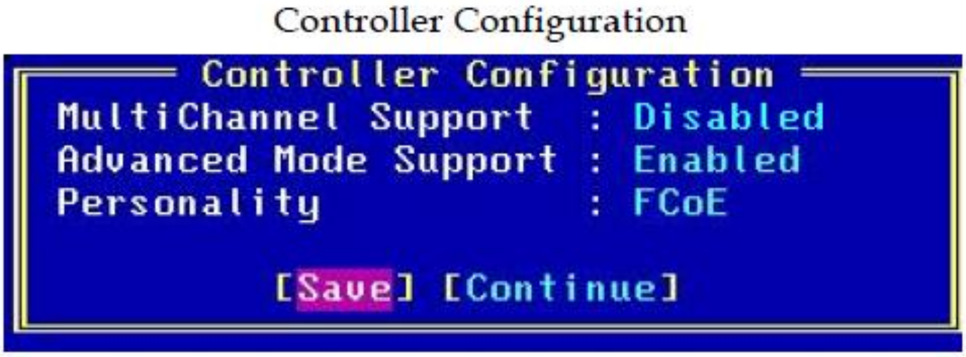
Device	Port No.	WWPN/VLAN	FC_ID	Connected To
CX320 (3X)	FC 3/21/1	-	-	FCB
	10GE 3/1/1	2094, 2093	-	CH121 V3
	10GE 3/17/1	2093		LAN
CH121 V3	-	30:00:00:68:50:40:30:02	16.00.01	CX320 (2X)
	-	30:00:00:68:50:40:30:04	16.00.03	CX320 (3X)
FCA		30:00:00:68:50:40:30:01	16.00.02	CX320 (2X)
FCB		30:00:00:68:50:40:30:03	16.00.04	CX320 (3X)

Step 3 Perform the configuration.

Configure the MZ510.

1. During the BIOS startup, press **Ctrl+P** when "Press <Ctrl><P> for PXESelect (TM) Utility" is displayed. The **Controller Configuration** screen shown in Figure 4-26 is displayed.

Figure 4-26 Configuring the MZ510

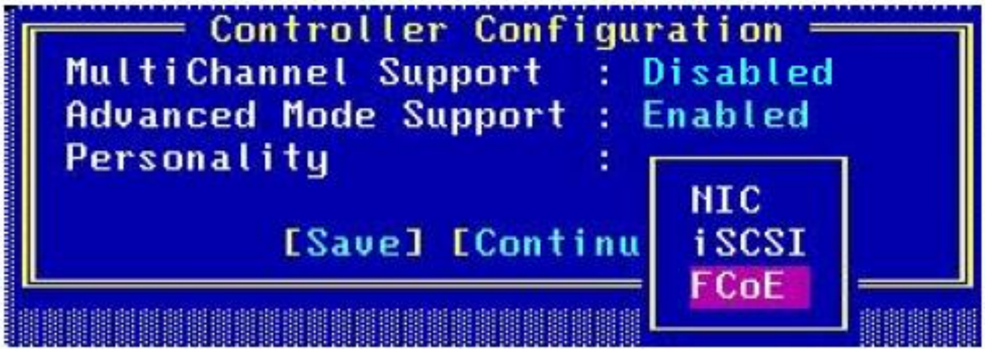


Controller Configuration

```
Controller Configuration
MultiChannel Support   : Disabled
Advanced Mode Support  : Enabled
Personality            : FCoE

        [Save] [Continue]
```

2. Select **Personality** by pressing arrow keys and press **Enter**. The **Personality** options shown in Figure 4-27 are displayed.

Figure 4-27 Setting the NIC working mode



Controller Configuration

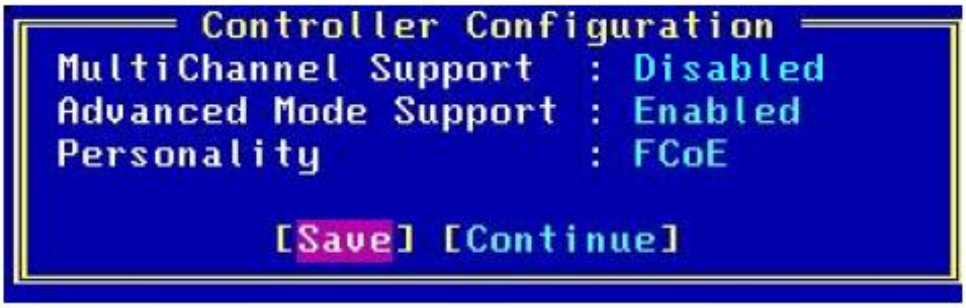
MultiChannel Support : Disabled

Advanced Mode Support : Enabled

Personality :

3. Select **FCoE** by pressing arrow keys and press **Enter**. Then select **Save** by pressing arrow keys and press **Enter**, as shown in Figure 4-28.

Figure 4-28 Saving the configuration



Controller Configuration

MultiChannel Support : Disabled

Advanced Mode Support : Enabled

Personality : FCoE

Configure FC ports.

# Before configuring FC ports, insert FC optical modules into the ports. If the Link indicator is steady green, the port is up.

# Configure port 10GE 2/21/1 of the CX320 in slot 2X as an FC port.

```
<HUAWEI> system-view

[~HUAWEI] sysname CX320_2X

[*HUAWEI] commit

[~CX320_2X] port mode fc 10GE 2/21/1

Warning: This operation will cause all the other configurations on the port to be lost.
Continue?[Y/N]:Y

[*CX320_2X] commit
```

# Check the FC port status.

```
[~CX320_2X] display interface fc 2/21/1

FC2/21/1 current state : UP (ifindex:144)

Line protocol current state : UP
```

......

Create an NPV instance.

```
# Create VLAN 2093 for Ethernet services.
[~CX320_2X] vlan 2093
# Create FCoE port 1.
[*CX320_2X] interface fcoe-port 1
[*CX320_2X-FCoE-Port1] quit
# For the CX320 in slot 2X, create an NPV instance npv1 and specify VLAN 2094 as the FCoE VLAN and VLAN 2095 as the NPV VLAN.
[*CX320_2X] fcoe npv1 npv
[*CX320_2X-fcoe-npv-npv1] vlan 2094
[*CX320_2X-fcoe-npv-npv1] npv-vlan 2095
[*CX320_2X] quit
# Add port 10GE 2/1/1 of the CX320 in slot 2X to FCoE port 1.
[*CX320_2X-10GE2/1/1] fcoe-port 1
[*CX320_2X-10GE2/1/1] quit
# Set the ports connected to CH121 compute nodes as hybrid ports and add the FCoE and Ethernet VLANs.
[*CX320_2X] interface 10GE 2/1/1
[*CX320_2X-10GE2/1/1] port link-type hybrid
[*CX320_2X-10GE2/1/1] port hybrid tagged vlan 2094
[*CX320_2X-10GE2/1/1] port hybrid tagged vlan 2093
[*CX320_2X-10GE2/1/1] fcoe-port 1
[*CX320_2X-10GE2/1/1] quit
# Set ports connected to the LAN as access ports and add the Ethernet VLAN.
[*CX320_2X] interface 10GE 2/17/1
[*CX320_2X-10GE2/17/1] port link-type access
[*CX320_2X-10GE2/17/1] port default vlan 2093
# Add port FC 2/21/1 of the CX320 in slot 2X to the npv1 instance.
[*CX320_2X] fcoe npv1
[*CX320_2X-fcoe-fcf-npv1] member interface fc 2/21/1
[*CX320_2X-fcoe-fcf-npv1] member interface fcoe-port 1
[*CX320_2X-fcoe-fcf-npv1] fcoe role np-port interface fc 2/21/1
[*CX320_2X] commit
```

Verify the configuration.

# View registration information of servers and storage.

```
[~CX320_2X] display fcoe instance npv

FCoE instance with NPV type:

------------------------------------------------------------------------------

Instance name                                  : npv1

VLAN                                                : 2094

Instance MAC                                    : 20:0b:c7:23:42:03

FKA-ADV-Period(ms)                         : 8000

Number of FCoE Port(VF & F)            : 1

Number of FCoE Port(VNP & NP)          : 1

Number of online VF(F)-Port               : 1

Number of online ENode VN(N)-Port     : 1

------------------------------------------------------------------------------

Total: 1
```

Configure CX320_3X.

The configuration procedure is the same as that of the CX320 in slot 2X, except that the FC port numbers and the numbers of the ports connected to compute nodes.

----End

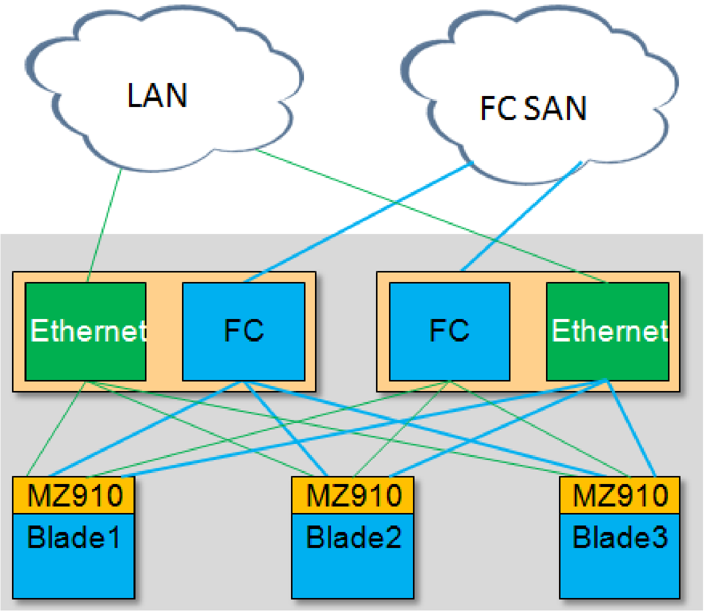
4.5 FC Networking

4.5.1 Multi-Plane Switch Module

4.5.1.1 Overview

The CX9xx series multi-plane switch modules include the switch module providing 10GE+ 8G FC ports(for example, CX911 and CX912) and the switch module providing GE+ 8G FC ports (for example, CX915). Figure 4-29 shows the internal structure of multi-plane switch modules.

Figure 4-29 Internal structure of multi-plane switch modules



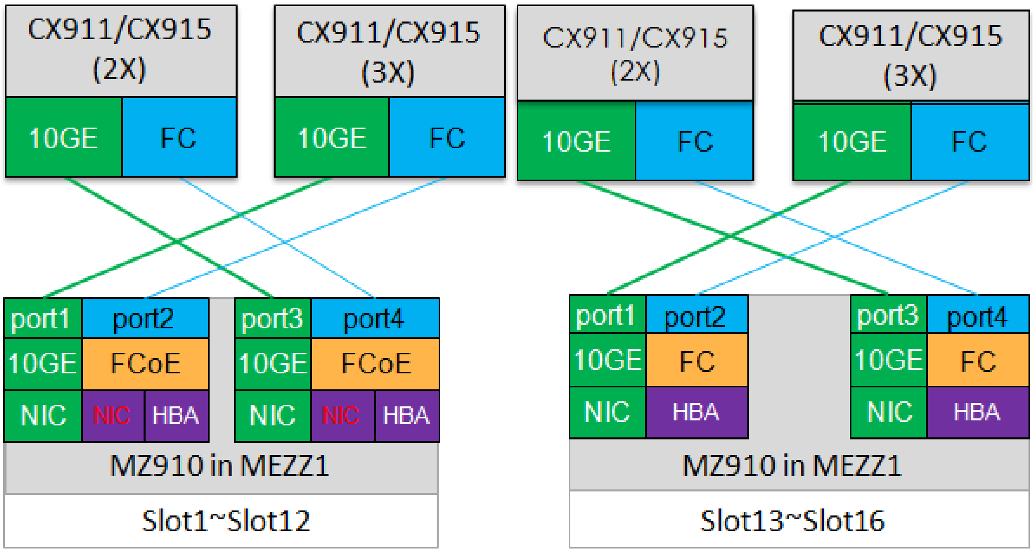
Ethernet and FC planes are independent in management, control, and data channels, similar to two independent switches combined into one switch module. The two planes use the same hardware management and monitoring systems, improving system integration. The CX210 provides only the FC switching plane. Like the CX912, CX210 integrates the MX210 module.

4.5.1.2 MZ910 Port Types

The FC switching planes of the CX911 and CX915 integrate the Qlogic's FCoE switches (MX510s), and the FC switching plane of the CX912 integrates the Brocade 300 FC switches (MX210s). The MX510 switch module provides 12 FCoE ports and 12 FC ports, and the E9000 supports a maximum of 16 blades. Therefore, if a multi-plane switch module is used with the MZ910, port types of different slots are different.

The MZ910 provides four ports. Ports 1 and 3 are 10GE Ethernet ports, and ports 2 and 4 are FC or FCoE ports. (The MM910 configures the MZ910 port types based on the switch module type.) Figure 4-30 shows port types of the MZ910s connected to the CX911s or CX915s.

Figure 4-30 MZ910 port types



From the OS perspective, each FCoE port has a NIC and an HBA, and each 10GE port has a NIC. Therefore, if the CX911 or CX915 is used with the MZ910, each MZ910 in slots 1 to 12 provides four NICs and two HBAs and each MZ910 in slots 13 to 16 provides two NICs and two HBAs. See Table 4-12.

Table 4-12 Number of NICs/HBAs

Switch Module	Slot Number	MZ910 Port Type	Number of NICs	Number of HBAs
CX911/CX915	01-12	FCoE	4	2
	13-16	FC	2	2
CX912	01-16	FC	2	2

4.5.1.3 MX210/MX510 Working Modes

Table 4-13 MX210/MX510 working mode

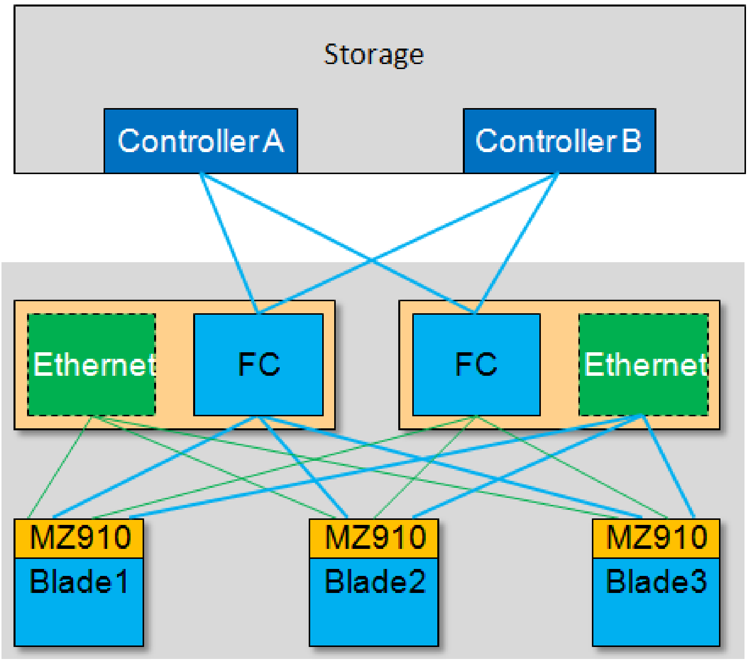
Connected Device	Vender/Model	MX510 Working Mode	MX210 Working Mode
FC storage device (direct connection)	NA	Full fabric	Native (switch)
FC switch	MDS series: MDS9100/MDS9500 Nexus series: Nexus 5000/Nexus7000	Transparent (NPV)	Access Gateway (NPV)

Connected Device	Vender/Model	MX510 Working Mode	MX210 Working Mode
	Qlogic 3000/QLogic 5000/QLogic 9000	Full fabric	Access Gateway (NPV)
	FC switch: Brocade 300/Brocade 5000/Brocade 7500	Transparent (NPV)	Access Gateway or Native (switch, requiring the full fabric license)

When MX210s (Native) are connected to Brocade switches through E-Ports, the Full Fabric license needs to be configured.

4.5.2 Connecting MX210s/MX510s to Storage Devices

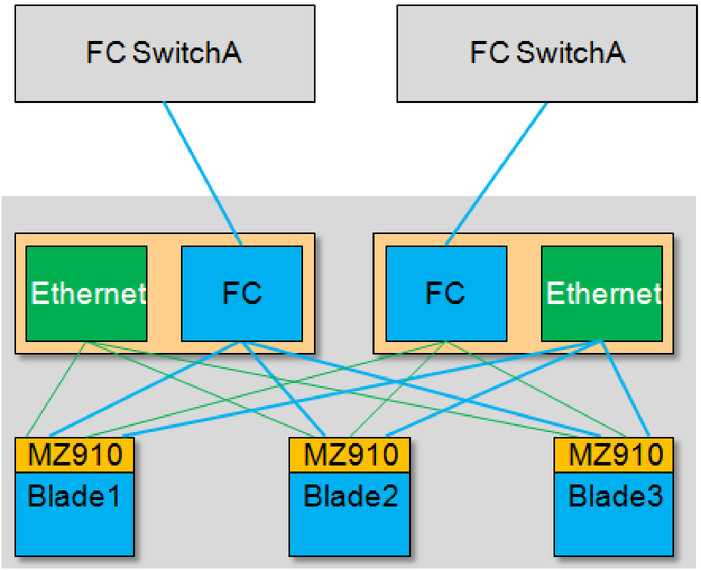
Figure 4-31 Connecting MX210s/MX510s to storage devices



Connect the MX510s of the CX911/CX915, MX210s of the CX912, and CX210s to storage devices in crossover mode to ensure system reliability and storage performance. The MX210s and MX510s in default configuration can directly connect to storage devices.

4.5.3 Connecting MX210s/MX510s to FC Switches

Figure 4-32 Connecting MX210s/MX510s to FC switches



Connect MX210s/MX510s to FC switches in parallel connection mode. (The crossover connection mode is not recommended. Cross-connections may cause multiple paths to connect to one switch.)

4.5.3.1 MX210 Load Balancing and Failover

Same as the CX311, the CX911 and CX915 use the MX510 as the FC plane module. The link load balancing and failover of the CX911 and CX915 are also the same as that of the CX311. When registering with an external switch, the MX510 randomly selects an FC port from the available FC ports on the panel to transmit data. If the port is faulty, the HBA registers again and selects a new FC port.

The CX912 and CX210 integrate the MX210. When connecting a CX912 or CX210 to an external Brocade FC switch, bind the links using ISL Trunking. The link load balancing and failover of the CX912 and CX210 are as follows:

1. **Access Gateway**: Use N_Port trunking to implement link aggregation and load balancing. If a physical link is faulty, the CX912 and CX210 automatically switch traffic to another link. The HBA does not need to register again.
2. **Native (Switch)**: Use E_Port trunking to implement link aggregation and load balancing. If an FC link is faulty, the CX912 and CX210 automatically switch traffic to another link. The HBA does not need to register again.

To implement the ISL Trunking feature, the MX210s and external FC switches must support ISL Trunking, and the MX210 must be configured with an independent license. Compared with the MX510, the MX210 offers a shorter failover duration when an FC link is faulty.

A Acronyms and Abbreviations

Acronyms and Abbreviations	Full Name
BPDU	Bridge Protocol Data Unit
CNA	Converged Network Adapter
DAD	Dual-Active Detect
DCB	Data Center Bridging
DCBX	Data Center Bridging Exchange Protocol
DRR	Deficit Round Robin
ETS	Enhanced transmission selection
FC	Fibre Channel
FCoE	Fibre Channel over Ethernet
FCF	Fibre Channel Forwarder
FDR	Fourteen Data Rate
FIP	FCoE Initialization Protocol
FSB	FCoE Initialization Protocol Snooping Bridge
HBA	Host Bus Adapter
LACP	Link Aggregation Control Protocol
LAG	Link Aggregation Group
LACPDU	Link Aggregation Control Protocol Data Unit
LLDP	Link Layer Discovery Protocol
MSTP	Multiple Spanning Tree Protocol
NPV	N_Port Virtualizer
NPIV	N_Port_ID virtualization

Acronyms and Abbreviations	**Full Name**
PFC	Priority-based Flow control
PG	Port Group
PVST	Per-VLAN Spanning Tree