Eth-Trunk, iStack, and CSS

Foreword

- As services develop and the campus network scale expands, users have increasingly demanding requirements on network bandwidth and reliability. Traditional solutions improve network bandwidth by upgrading devices and implement high reliability by deploying redundant links and using the Spanning Tree Protocol (STP), leading to low flexibility, time-consuming troubleshooting, and complex configuration.
- This chapter describes how to use Eth-Trunk, intelligent stack (iStack), and cluster switch system (CSS) technologies to improve network bandwidth and reliability.

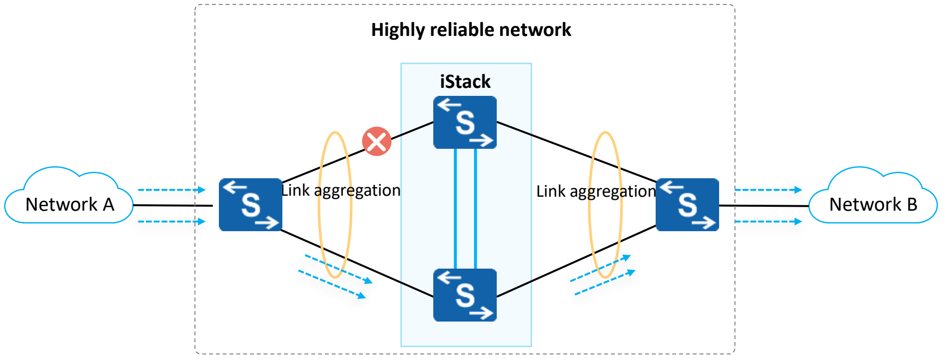
Objectives

- On completion of this course, you will be able to:
  - Understand the functions of link aggregation.
  - Understand the link aggregation types.
  - Understand the link aggregation negotiation process in Link Aggregation Control Protocol (LACP) mode.
  - Understand the advantages and principles of iStack and CSS.
  - Understand the common applications and networking of link aggregation and stacking technologies.

Contents

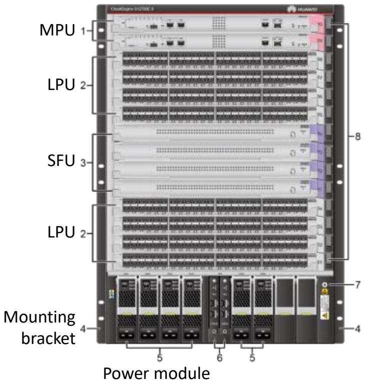
Network Reliability

- Network reliability refers to the capability of ensuring nonstop network services when a single point or multiple points of failure occur on a device or link.
- Network reliability can be implemented at the card, device, and link levels.



- As networks rapidly develop and applications become more and more diversified, various value-added services (VASs) are widely deployed. Network interruption may cause many service exceptions and huge economic losses. Therefore, the reliability of networks has become a focus.

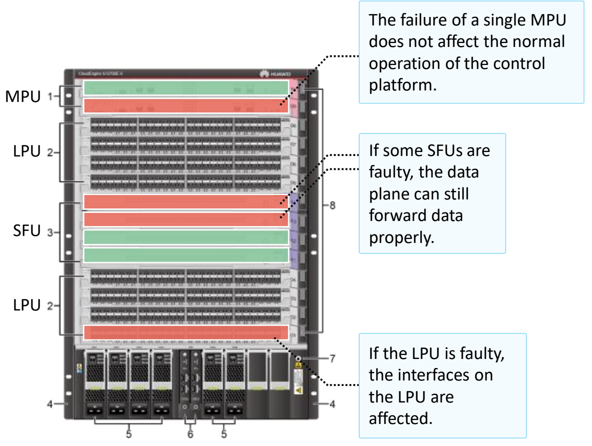
Card Reliability (1)



Front view of the S12700E-8 chassis

- A modular switch consists of a chassis, power modules, fan modules, main processing units (MPUs), switch fabric units (SFUs), and line processing units (LPUs).
- Chassis: provides slots for various cards and modules to implement inter-card communication.
- Power module: power supply system of the device
- Fan module: heat dissipation system
- MPU: responsible for the control plane and management plane of the entire system.
- SFU: responsible for the data plane of the entire system. The data plane provides high-speed non-blocking data channels for data switching between service modules.
- LPU: provides data forwarding functions on a physical device and provides optical and electrical interfaces of different rates.

Card Reliability (2)

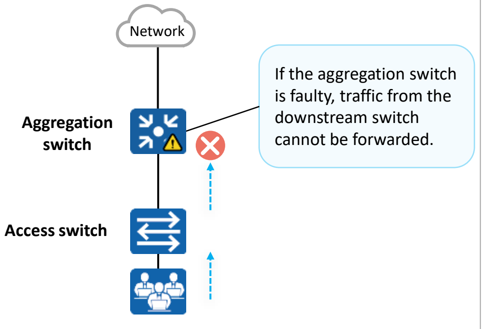


Front view of the S12700E-8 chassis

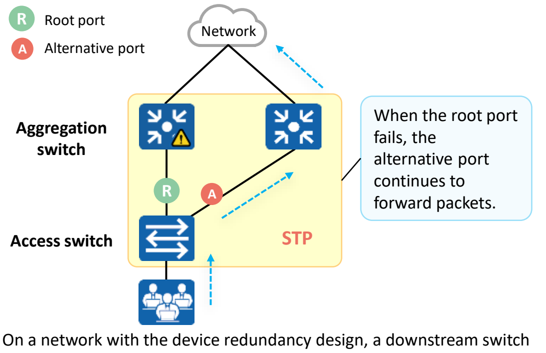
- For example, the S12700E-8 provides eight LPU slots, four SFU slots, two MPU slots, six power module slots, and four fan module slots.
- A modular switch can be configured with multiple MPUs and SFUs to ensure device reliability. If an SFU or MPU in a single slot is faulty, the switch can still run properly.
- After an LPU of a modular switch is damaged, interfaces on the LPU cannot forward data.

Device Reliability

No backup	Master/Backup mode



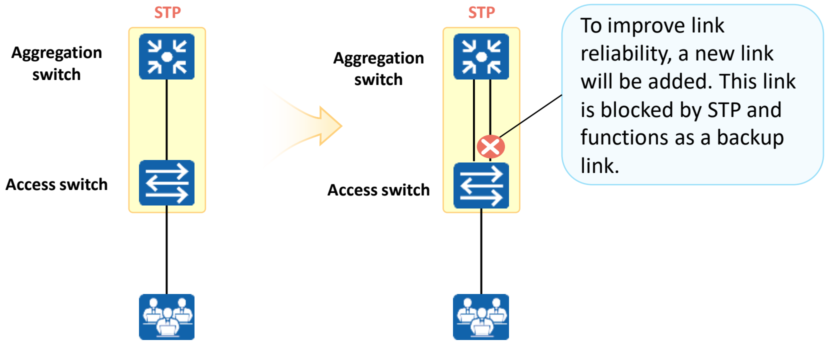
On a network without the device redundancy design, a downstream switch uses a single uplink. If the upstream switch or its interfaces fail, all downstream networks are interrupted.



On a network with the device redundancy design, a downstream switch is dual-homed to two upstream switches. The links work in active/backup mode. If the active link or upstream switch fails, traffic is switched to the backup link and forwarded through the backup device.

HUAWEI

Link Reliability

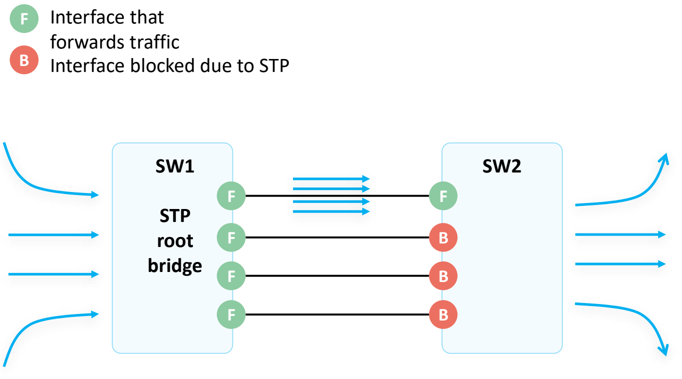


- To ensure link reliability, deploy multiple physical links between devices. To prevent loops, configure STP to ensure that traffic is forwarded on only one link, and other links function as backup links.

Contents

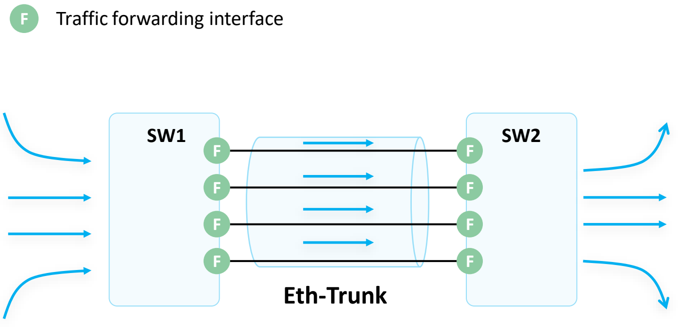
Increasing Link Bandwidth

- When multiple links exist between devices, traffic is forwarded on only one link due to STP. In this case, the inter-device link bandwidth remains unchanged.

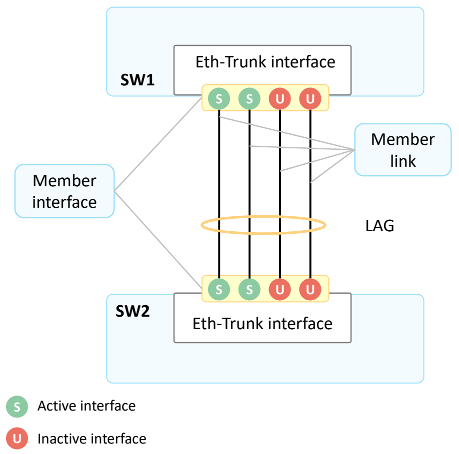


Eth-Trunk

- Ethernet link aggregation, also called Eth-Trunk, bundles multiple physical links into a logical link to increase link bandwidth, without having to upgrade hardware.



Basic Concepts of Eth-Trunk



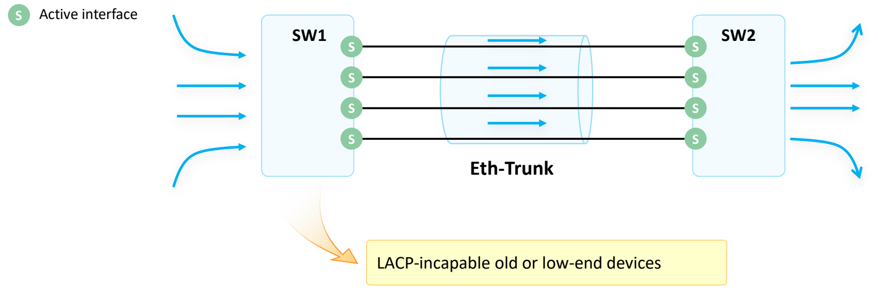
- A link aggregation group (LAG) is a logical link formed by bundling several links. Each LAG has one logical interface, known as an LAG interface or Eth-Trunk interface.
- Member interface and member link: Physical interfaces that constitute an Eth-Trunk interface are called member interfaces, and the link corresponding to a member interface is known as a member link.
- Active interface and active link: An active interface is also called a selected interface and is a member interface that participates in data forwarding. The link corresponding to an active interface is called an active link.
- Inactive interface and inactive link: An inactive interface is also called an unselected interface and is a member interface that does not participate in data forwarding. A link corresponding to an inactive interface is referred to as an inactive link.
- Link aggregation mode: Based on whether the Link Aggregation Control Protocol (LACP) is enabled, link aggregation can be classified into manual mode and LACP mode.
- Other concepts: upper and lower thresholds for the number of active interfaces

- An Eth-Trunk can be treated as a physical Ethernet interface. The only difference between the Eth-Trunk and physical Ethernet interface is that the Eth-Trunk needs to select one or more member interfaces to forward traffic.
- The following parameters must be the same for member interfaces in an Eth-Trunk:
  - Interface rate
  - Duplex mode
  - VLAN configurations: The interface type must be the same (access, trunk, or hybrid). For access interfaces, the default VLAN of the member interfaces must be the same. For trunk interfaces, the allowed VLANs and the default VLAN of the member interfaces must be the same.

Contents

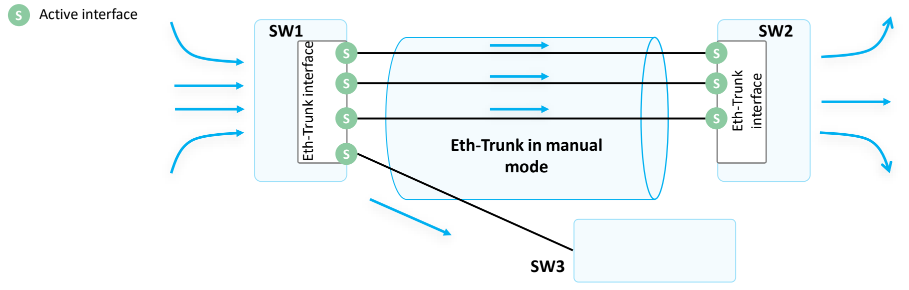
HUAWEI

Manual Mode



- Manual mode: An Eth-Trunk is manually created, and its member interfaces are manually configured. LACP is not used for negotiation between the two systems.
- In most cases, all links are active links. In this mode, all active links forward data and evenly share traffic. If an active link is faulty, the LAG automatically evenly shares traffic among the remaining active links.
- If one of the devices at both ends of an LAG does not support LACP, you can use the manual mode.

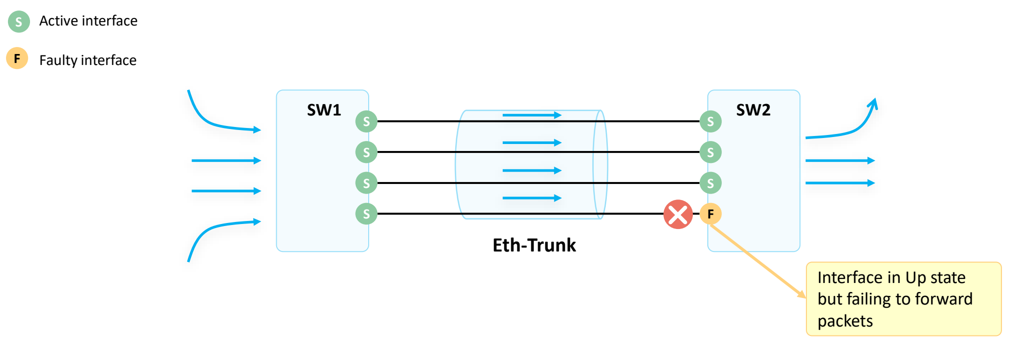
Defects of the Manual Mode (1)



- To ensure that the Eth-Trunk works properly, ensure that the peer interfaces of all member interfaces in the Eth-Trunk meet the following requirements:
  - The peer interfaces reside on the same device.
  - The peer interfaces are added to the same Eth-Trunk.
- In manual mode, devices do not exchange packets. Therefore, the configuration needs to be manually confirmed.

- As shown in the preceding figure, four interfaces of SW1 are added to an Eth-Trunk, but the peer end of one interface is SW3 instead of SW2. In this case, some traffic is load balanced to SW3, causing communication exceptions.

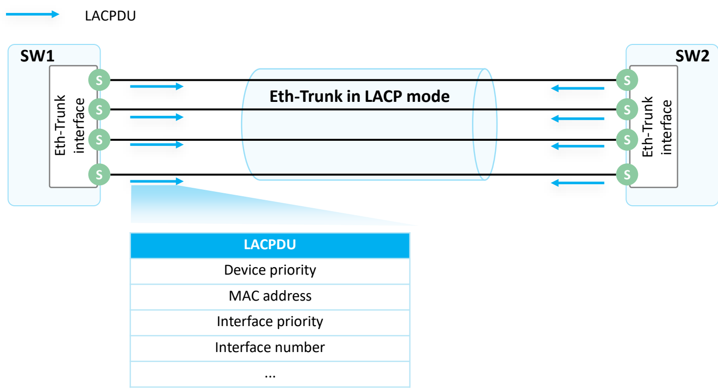
Defects of the Manual Mode (2)



- In manual mode, the device can determine whether the peer interface is working properly based only on the physical layer status.

Contents

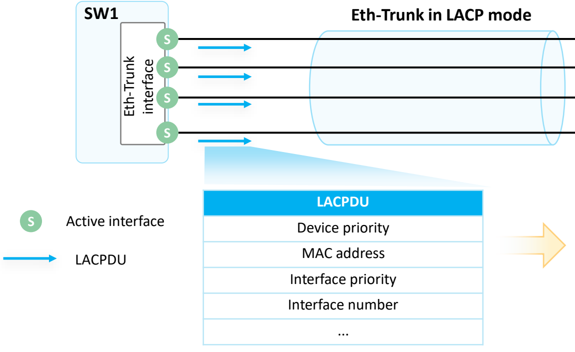
LACPDU



- LACP mode: A link aggregation mode that uses the LACP protocol. Devices exchange Link Aggregation Control Protocol Data Units (LACPDUs) to ensure that the peer interfaces are member interfaces that belong to the same Eth-Trunk and are on the same device.
- An LACPDU contains the device priority, MAC address, interface priority, and interface number.

System Priority

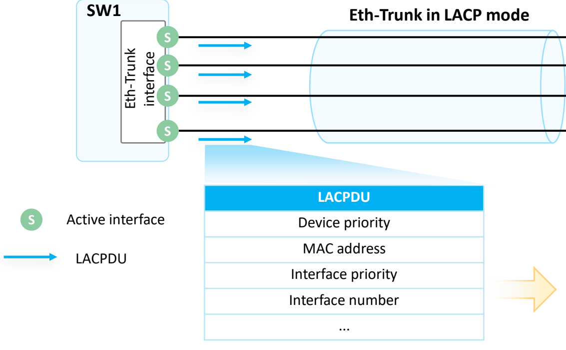
- In LACP mode, the number of active interfaces selected by devices at both ends must be consistent; otherwise, the Eth-Trunk cannot be set up. In this case, configure one end as the Actor. Then the other end selects active interfaces according to the Actor.
- The Actor is determined based on the LACP system priority. A smaller value indicates a higher priority.



By default, the LACP system priority is 32768. A smaller value indicates a higher priority. Generally, the default value is used. When the priorities are the same, LACP selects the Actor by comparing the MAC addresses. A smaller MAC address indicates a higher priority.

Interface Priority

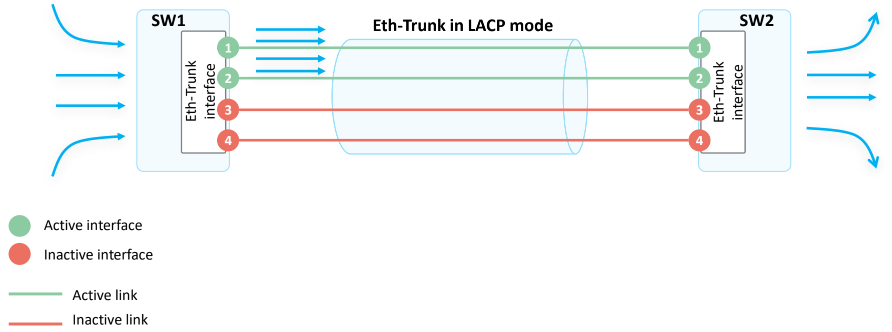
- After the Actor is selected, both devices select active interfaces based on the interface priorities of the Actor. A smaller LACP interface priority value indicates a higher priority.



By default, the LACP interface priority of an interface is 32768. A smaller value indicates a higher priority. Generally, the default value is used. When the priorities are the same, LACP selects active interfaces based on interface numbers. A smaller interface number indicates a higher priority.

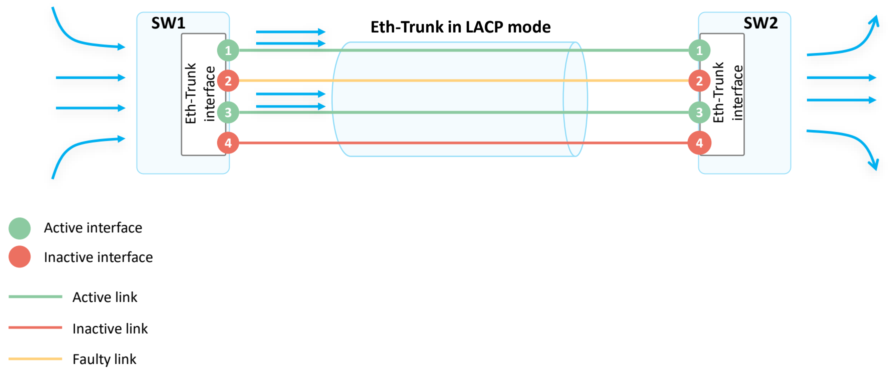
Maximum Number of Active Interfaces (1)

- In LACP mode, the maximum number of active interfaces can be configured. When the number of member interfaces exceeds the maximum number of active interfaces, the interfaces with higher priorities and smaller interface numbers are selected as active interfaces, and the other interfaces function as backup interfaces (inactive interfaces). In addition, the links corresponding to active interfaces become active links, and the links corresponding to inactive interfaces become inactive links. The switch sends and receives packets only through active interfaces.

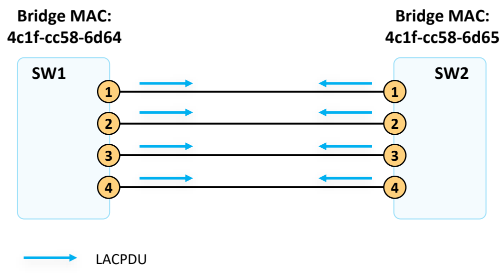


Maximum Number of Active Interfaces (2)

- If an active link fails, an inactive link with the highest priority (based on the interface priority and interface number) is selected to replace the faulty link. This ensures that the overall bandwidth does not change and services are not interrupted.



Active Link Election (1)

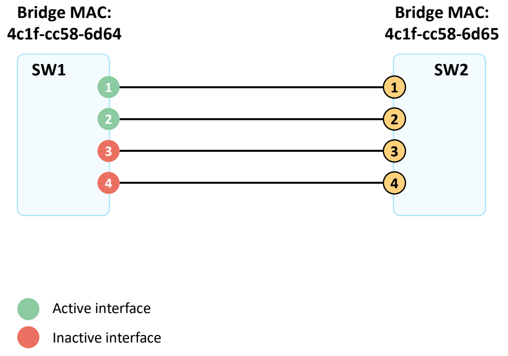


- An Eth-Trunk in LACP mode is set up between SW1 and SW2. The maximum number of active interfaces is set to 2 on SW1 and SW2.
- SW1 with a higher priority is elected as the Actor through LACPDUs.

HUAWEI

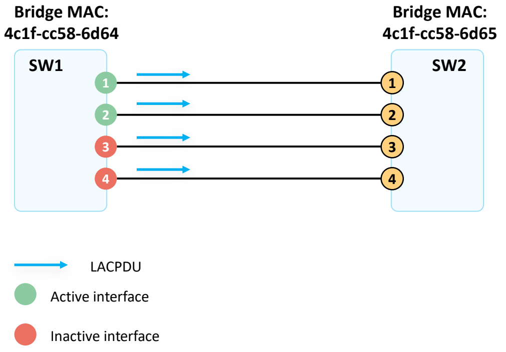
- Configure an Eth-Trunk in LACP mode between SW1 and SW2 and add four interfaces to an Eth-Trunk. The four interfaces are numbered 1, 2, 3, and 4. On SW1 and SW2, set the maximum number of active interfaces in the Eth-Trunk to 2 and retain the default settings for the other parameters (system priority and interface priority).
- SW1 and SW2 send LACPDUs through member interfaces 1, 2, 3, and 4.
- When receiving LACPDUs from the peer end, SW1 and SW2 compare the system priorities, which use the default value 32768 and are the same. Then they compare MAC addresses. The MAC address of SW1 is 4c1f-cc58-6d64, and the MAC address of SW2 is 4c1f-cc58-6d65. SW1 has a smaller MAC address and is preferentially elected as the Actor.

Active Link Election (2)



- SW1 compares the interface priorities and interface numbers to select active interfaces. Under the same interface priority, interfaces 1 and 2 have smaller interface numbers and are elected as active interfaces.

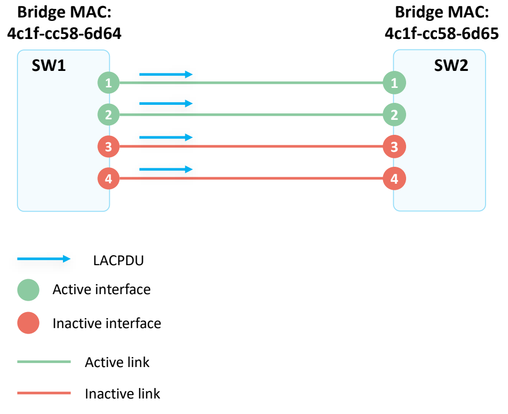
Active Link Election (3)



- SW1 notifies the peer end of the elected active interfaces through LACPDUs.

- LACP uses the following flags in an LACPDU to identify the interface status. If the three flags are set to 1, the interface is an active interface.
  - Synchronization
  - Collecting
  - Distributing
- If the three flags are set to 0, the interface is an inactive interface.

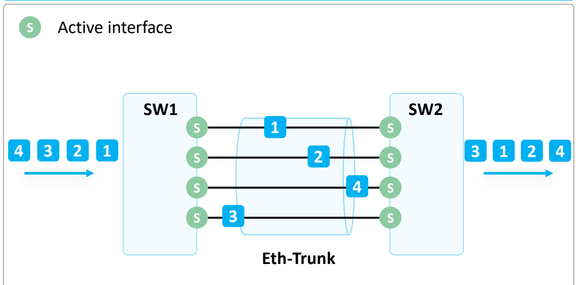
Active Link Election (4)



- SW2 determines the local active interfaces based on the election result of SW1 and the corresponding links become active links.
- In this way, the election of active links is complete.

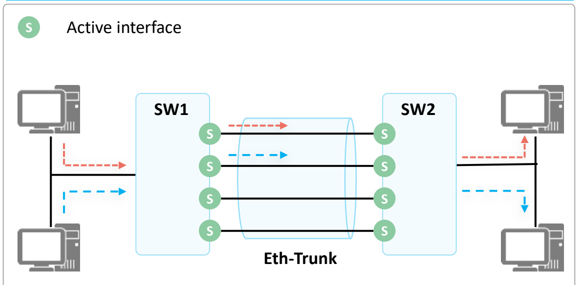
Load Balancing

Per-packet load balancing



When an Eth-Trunk is used to forward data, there are multiple physical links between devices at both ends of the Eth-Trunk. If data frames are forwarded on different links, data frames may arrive at the peer end in a different order in which they were transmitted, resulting in out-of-order packets.

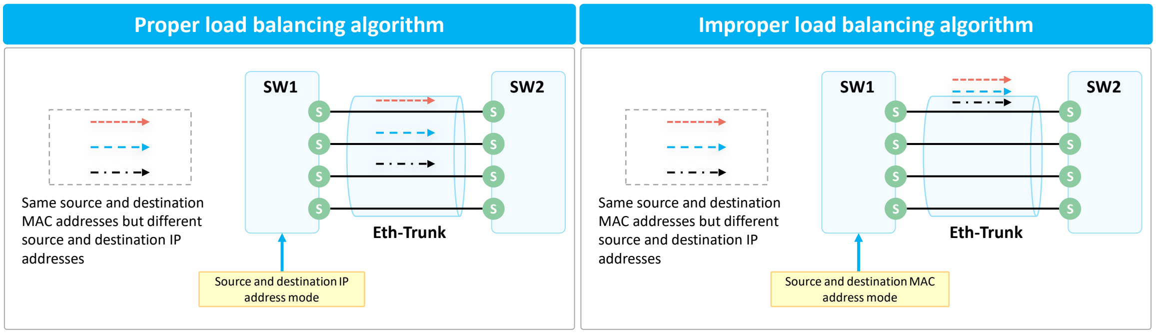
Per-flow load balancing



Load balancing based on flows is recommended for an Eth-Trunk. In this mode, a flow is load balanced to the same link. This ensures that frames of the same flow are transmitted over the same physical link and implements load balancing among physical links in an Eth-Trunk.

Load Balancing Mode

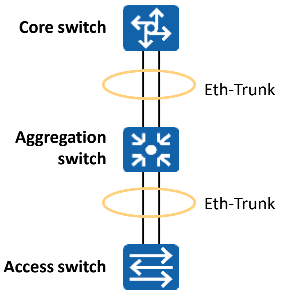
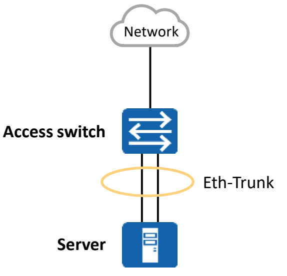
- An Eth-Trunk can load balance traffic based on IP addresses or MAC addresses of packets. You can configure different load balancing modes (valid locally only for outgoing packets) to distribute data flows to different member interfaces.
- Traffic can be load balanced based on: source IP address, source MAC address, destination IP address, destination MAC address, source and destination IP addresses, and source and destination MAC addresses.
- For actual services, you need to configure a proper load balancing mode based on traffic characteristics. If a service traffic parameter changes frequently, it is easier to load balance traffic if you use the load balancing mode based on this frequently-changing parameter.



- If the IP addresses of packets change frequently, load balancing based on the source IP address, destination IP address, or source and destination IP addresses is more suitable for load balancing among physical links.
- If MAC addresses of packets change frequently and IP addresses are fixed, load balancing based on the source MAC address, destination MAC address, or source and destination MAC addresses is more suitable for load balancing among physical links.
- If the selected load balancing mode is unsuitable for the actual service characteristics, traffic may be unevenly load balanced. Some member links have high load, but other member links are idle. For example, if the source and destination IP addresses of packets change frequently but the source and destination MAC addresses are fixed and traffic is load balanced based on the source and destination MAC addresses, all traffic is transmitted over one member link.

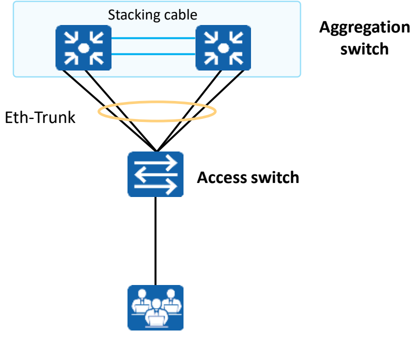
Contents

Typical Application Scenario (1)

Between switches	Between the switch and server
	
To ensure the bandwidth and reliability of links between switches, deploy multiple physical links between switches and add them to an Eth-Trunk.	To improve the access bandwidth and reliability of the server, bind two or more physical NICs into a NIC group and establish an Eth-Trunk with the switch.

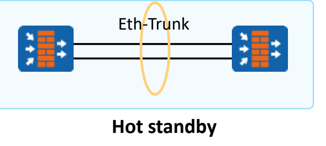
Typical Application Scenario (2)

Between a switch and stack



An iStack is a logical device consisting of two switches. A switch can be connected to the iStack through an Eth-Trunk to form a highly reliable loop-free network.

Heartbeat link of firewalls in hot standby mode



If two firewalls are deployed in hot standby mode, the heartbeat link is used to detect the status of the peer device. To prevent status detection errors caused by single-interface or single-link faults, you can create an Eth-Trunk and use it as the heartbeat link for status detection.

Contents

Configuration Commands (1)

1. Create an Eth-Trunk.

   ```
   [Huawei] interface eth-trunk trunk-id
   ```

   An Eth-Trunk interface is created, and the Eth-Trunk interface view is displayed.

2. Configure a link aggregation mode.

   ```
   [Huawei-Eth-Trunk1] mode {lacp | manual load-balance }
   ```

   To enable the LACP mode, run **mode lacp**. To enable the manual mode, run **mode manual load-balance**.

   Note: The link aggregation modes at both ends must be the same.

3. Add an interface to the Eth-Trunk (Ethernet interface view).

   ```
   [Huawei-GigabitEthernet0/0/1] eth-trunk trunk-id
   ```

   In the interface view, the interface is added to the Eth-Trunk.

Configuration Commands (2)

4. Add an interface to the Eth-Trunk (Eth-Trunk view).

   [Huawei-Eth-Trunk1] **trunkport** *interface-type { interface-number}*

   In the Eth-Trunk view, the interface is added to the Eth-Trunk. You can use either of the preceding commands to add an interface to an Eth-Trunk.

5. Enable interfaces at different rates to join the same Eth-Trunk interface.

   [Huawei-Eth-Trunk1] **mixed-rate link enable**

   By default, interfaces at different rates are not allowed to join the same Eth-Trunk, and only interfaces at the same rate can be added to the same Eth-Trunk.

6. Configure the LACP system priority.

   [Huawei] **lacp priority** *priority*

   A smaller priority value indicates a higher LACP system priority. By default, the LACP priority is 32768.

HUAWEI

Configuration Commands (3)

7. Configure the LACP interface priority.

```
[Huawei-GigabitEthernet0/0/1] lacp priority priority
```

The LACP interface priority is set in the interface view. By default, the LACP interface priority is 32768. A smaller priority value indicates a higher LACP interface priority.
You can run this command only after an interface is added to the Eth-Trunk.

8. Configure the maximum number of active interfaces.

```
[Huawei-Eth-Trunk1] max active-linknumber {number}
```

Ensure that the maximum number of active interfaces on the local end is the same as that on the peer end. The maximum number of active interfaces can be configured only in LACP mode.

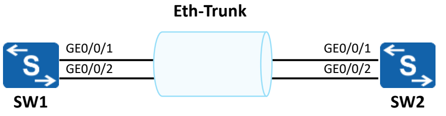
9. Configure the minimum number of active interfaces.

```
[Huawei-Eth-Trunk1] least active-linknumber {number}
```

The minimum number of active interfaces can be different on the local end and peer end and can be configured in both manual and LACP modes.
The minimum number of active interfaces is configured to ensure the minimum bandwidth. When the number of active links is smaller than the lower threshold, the Eth-Trunk interface goes down.

- The maximum number of active interfaces varies according to switch models. For example, the maximum number of active interfaces in an Eth-Trunk is 32 on the S6720HI, S6730H, S6730S, and S6730S-S, and is 16 on the S6720LI, S6720S-LI, S6720SI, and S6720S-SI. For details, see the product manual.
- The minimum number of active interfaces is configured to ensure the minimum bandwidth. If the bandwidth is too small, services that require high link bandwidth may be abnormal. In this case, you can disconnect the Eth-Trunk interface to switch services to other paths through the high reliability mechanism of the network, ensuring normal service running.

Example for Configuring an Eth-Trunk in Manual Mode



- Requirement description:
  - SW1 and SW2 are connected to the networks of VLAN 10 and VLAN 20.
  - SW1 and SW2 are connected through two Ethernet links. To provide link redundancy and enhance transmission reliability, configure an Eth-Trunk in manual mode between SW1 and SW2.

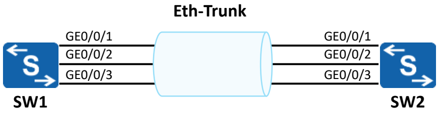
SW1 configuration:

```
[SW1] interface eth-trunk 1
[SW1-Eth-Trunk1] trunkport gigabitethernet 0/0/1 to 0/0/2
[SW1-Eth-Trunk1] port link-type trunk
[SW1-Eth-Trunk1] port trunk allow-pass vlan 10 20
```

SW2 configuration:

```
[SW2] interface eth-trunk 1
[SW2-Eth-Trunk1] trunkport gigabitethernet 0/0/1 to 0/0/2
[SW2-Eth-Trunk1] port link-type trunk
[SW2-Eth-Trunk1] port trunk allow-pass vlan 10 20
```


Example for Configuring an Eth-Trunk in LACP Mode (1)

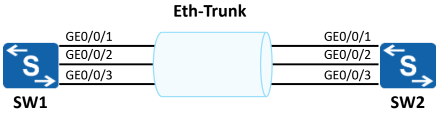


- Requirement description:
  - SW1 and SW2 are connected to the networks of VLAN 10 and VLAN 20.
  - SW1 and SW2 are connected through three Ethernet links. To provide link redundancy and enhance transmission reliability, configure an Eth-Trunk in LACP mode between SW1 and SW2, manually adjust the priority to configure SW1 as the Actor, and set the maximum number of active interfaces to 2. The other link functions as the backup link.

SW1 configuration:

```
[SW1] interface eth-trunk 1
[SW1-Eth-Trunk1] mode lacp
[SW1-Eth-Trunk1] max active-linknumber 2
[SW1-Eth-Trunk1] trunkport gigabitethernet 0/0/1 to 0/0/3
[SW1-Eth-Trunk1] port link-type trunk
[SW1-Eth-Trunk1] port trunk allow-pass vlan 10 20
[SW1-Eth-Trunk1] quit
[SW1] lacp priority 30000
```

Example for Configuring an Eth-Trunk in LACP Mode (2)



- Requirement description:
  - SW1 and SW2 are connected to the networks of VLAN 10 and VLAN 20.
  - SW1 and SW2 are connected through three Ethernet links. To provide link redundancy and enhance transmission reliability, configure an Eth-Trunk in LACP mode between SW1 and SW2, manually adjust the priority to configure SW1 as the Actor, and set the maximum number of active interfaces to 2. The other link functions as the backup link.

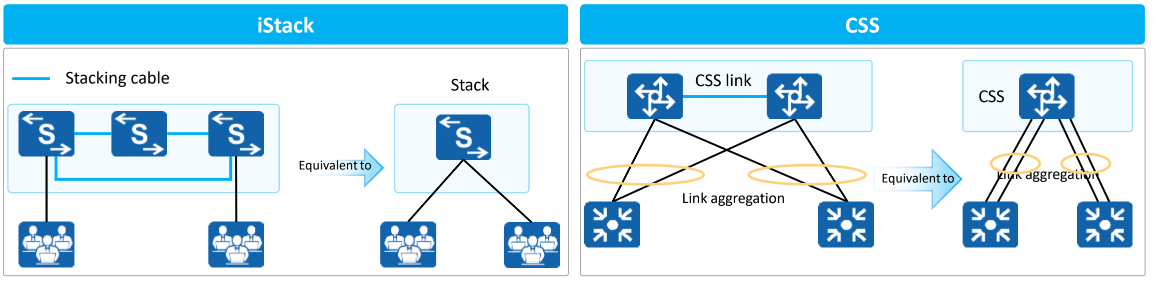
SW1 configuration:

[SW2] **interface eth-trunk 1**
[SW2-Eth-Trunk1] **mode lacp**
[SW2-Eth-Trunk1] **max active-linknumber 2**
[SW2-Eth-Trunk1] **trunkport gigabitethernet 0/0/1 to 0/0/3**
[SW2-Eth-Trunk1] **port link-type trunk**
[SW2-Eth-Trunk1] **port trunk allow-pass vlan 10 20**
[SW2-Eth-Trunk1] **quit**

Contents

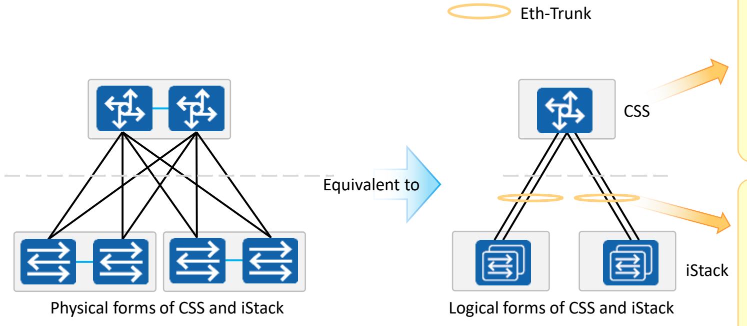
HUAWEI

Introduction to iStack and CSS



- iStack: Multiple iStack-capable switches are connected using stacking cables to form a logical switch that participates in data forwarding.
- Cluster switch system (CSS): Two CSS-capable switches are bundled into one logical switch.
- A CSS consists of only two switches. Generally, modular switches support CSS, and fixed switches support iStack.

Advantages of iStack and CSS



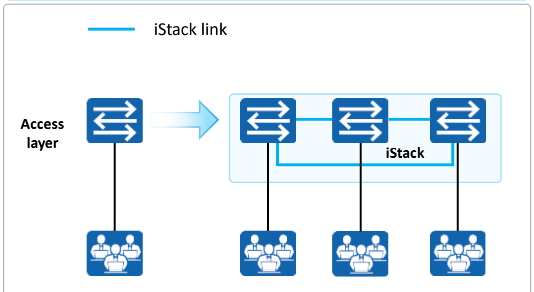
- One logical device simplifies O&M and facilitates management.
- If a physical device fails, the other device can take over the forwarding and control functions, preventing single points of failure.

- Inter-device link aggregation is implemented on a loop-free physical network, so STP does not need to be deployed.
- All links in the Eth-Trunk are used, and the link usage is 100%.

- Many-to-one virtualization: Switches can be virtualized into one logical switch (CSS) that has a unified control plane for unified management.
- Unified forwarding plane: Physical switches in a CSS use a unified forwarding plane, and share and synchronize forwarding information in real time.
- Inter-device link aggregation: Links between physical switches are aggregated into a single Eth-Trunk interface to interconnect with downstream devices.

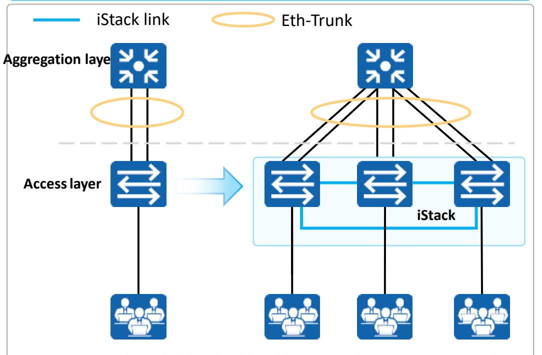
Application (1)

Extending the port quantity



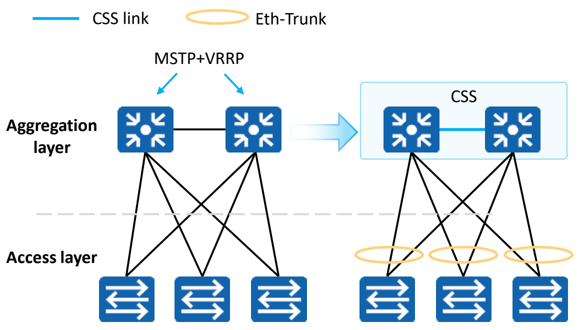
- When the port density of a switch cannot meet the access requirements, add new switches to set up an iStack to increase the number of ports.

Extending the bandwidth and implementing redundancy backup



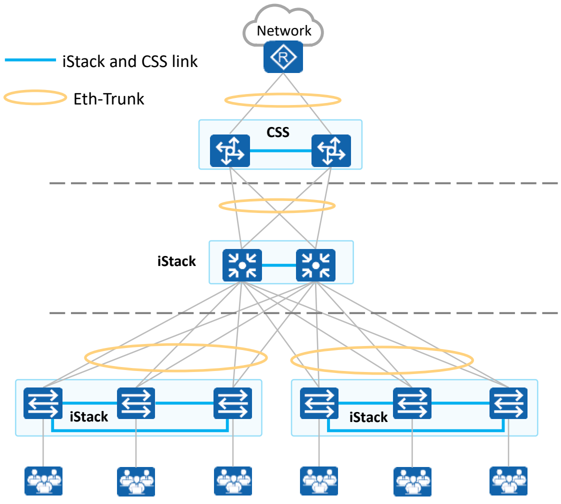
- To increase the uplink bandwidth, add new switches to set up an iStack and add multiple physical links of the member switches to an Eth-Trunk. This increases the uplink bandwidth, implements inter-device backup and inter-device link redundancy, and improves reliability.

Application (2)



- Two devices form a CSS and are virtualized into a single logical device. This simplified network does not require Multiple Spanning Tree Protocol (MSTP) or Virtual Router Redundancy Protocol (VRRP), so network configuration is much simpler. Additionally, inter-device link aggregation speeds up network convergence and improves network reliability.

Recommended Architecture



Core layer

- Core switches set up a CSS and use Eth-Trunks to connect to uplink and downlink devices, building a highly reliable and loop-free network.

Aggregation layer

- Aggregation switches set up an iStack and use Eth-Trunks to connect to uplink and downlink devices, building a highly reliable loop-free network.

Access layer

- Access devices that are geographically close to each other (such as access switches in a building) are virtualized into one logical device using iStack. This adds interfaces and simplifies management.
- An Eth-Trunk is used to connect to the aggregation layer. The logical network architecture is simple, and STP and VRRP are not required. This networking offers high reliability, high uplink bandwidth, and fast convergence.

Quiz

1. What are the differences between per-packet load balancing and per-flow load balancing?
2. How does an Actor be elected in LACP mode?
3. What are the advantages of CSS and iStack?

1. Packet disorder may occur if packets are load balanced to different links based on packets. If packets are load balanced to the same link based on flows, packet disorder will not occur. However, a single flow cannot make full use of the bandwidth of the entire Eth-Trunk.
2. Switches compare system priorities. A smaller value indicates a higher priority. If the system priorities are the same, the bridge MAC addresses are compared. A smaller bridge MAC address indicates a higher priority. The device with a higher priority becomes the Actor.
3. CSS and iStack simplify network management, improve network reliability, make full use of network link bandwidth, and use inter-device Eth-Trunk to construct a loop-free physical network.

Summary

- Link aggregation can be used to improve link reliability, utilization, and bandwidth. Link aggregation can be classified into static and LACP aggregation based on the aggregation mode.
- LACP uses packet negotiation to implement backup for active links. When a link fails, the backup link is elected as the active link to forward packets.
- To ensure the sequence in which packets arrive, link aggregation uses per-flow load balancing.
- iStack and CSS simplify network management and network structure, and improve network reliability.

HUAWEI


Thank You
www.huawei.com