Fondamenti di Internet

STP Principles and Configuration

Prof. Gianluca Reali

Foreword

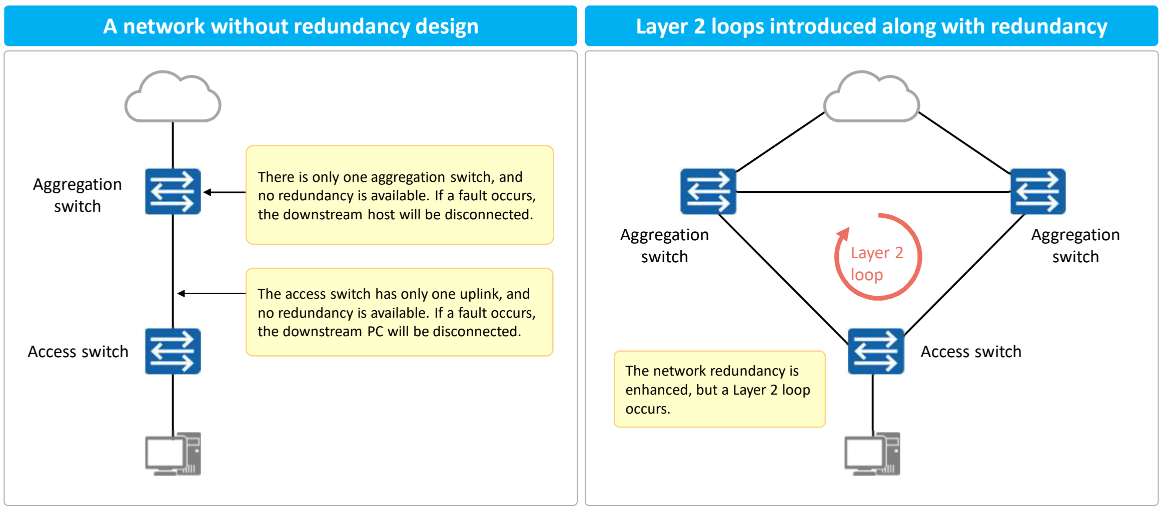
- On an Ethernet switching network, redundant links are used to implement link backup and enhance network reliability. However, the use of redundant links may produce loops, leading to broadcast storms and an unstable MAC address table. As a result, communication on the network may deteriorate or even be interrupted. To prevent loops, IEEE introduced the Spanning Tree Protocol (STP).
- Devices running STP exchange STP Bridge Protocol Data Units (BPDUs) to discover loops on the network and block appropriate ports. This enables a ring topology to be trimmed into a loop-free tree topology, preventing infinite looping of packets and ensuring packet processing capabilities of devices.
- IEEE introduced the Rapid Spanning Tree Protocol (RSTP) to improve the network convergence speed.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the causes and problems of Layer 2 loops on a campus switching network.
  - Describe basic concepts and working mechanism of STP.
  - Distinguish STP from RSTP and describe the improvement of RSTP on STP.
  - Complete basic STP configurations.
  - Understand other methods to eliminate Layer 2 loops on the switching network except STP.

Contents

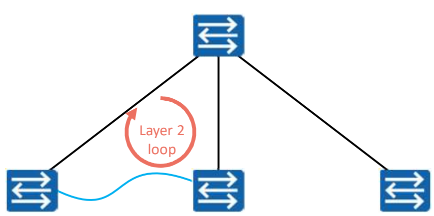
Technical Background: Redundancy and Loops on a Layer 2 Switching Network



- As LANs increase, more and more switches are used to implement interconnection between hosts. As shown in the figure, the access switch is connected to the upstream device through a single link. If the uplink fails, the host connected to the access switch is disconnected from the network. Another problem is the single point of failure (SPOF). That is, if the switch breaks down, the host connected to the access switch is also disconnected.
- To solve this problem, switches use redundant links to implement backup. Although redundant links improve network reliability, loops may occur. Loops cause many problems, such as communication quality deterioration and communication service interruption.

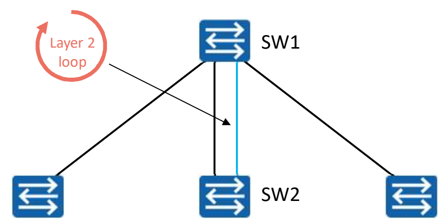
Technical Background: Layer 2 Loops Caused by Human Errors

Case 1



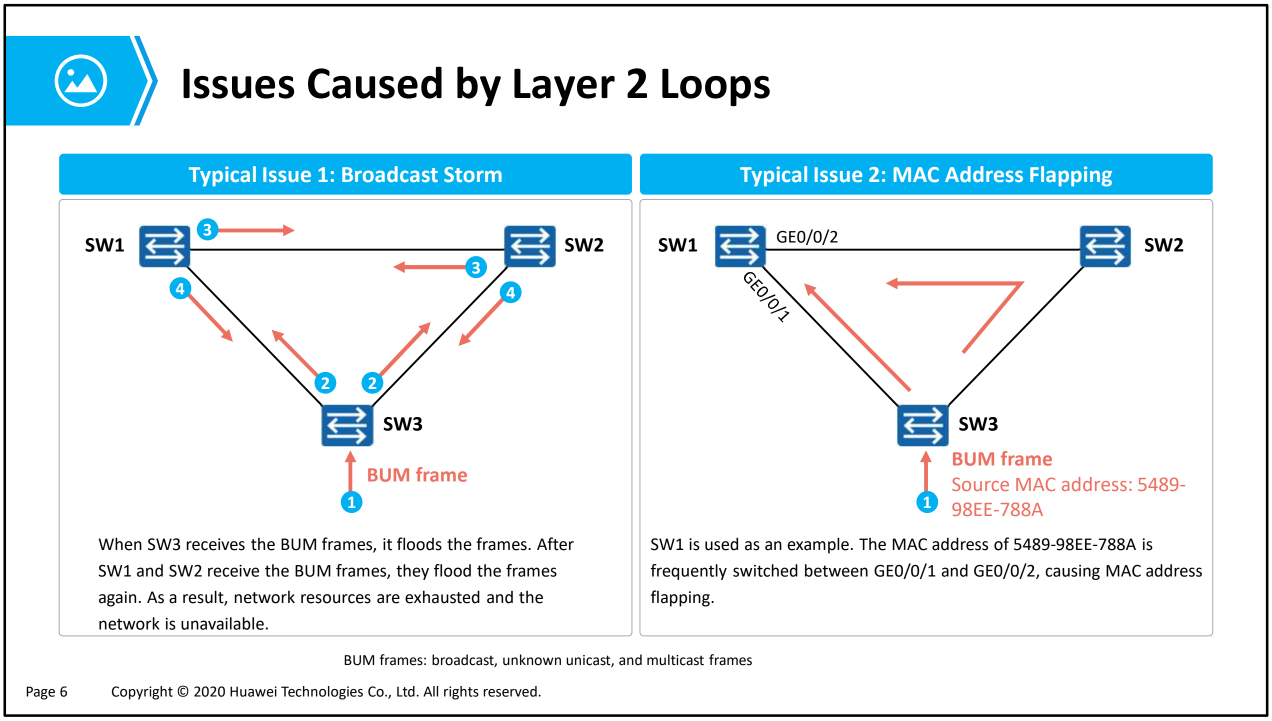
Incorrect operations: For example, connections of cables between devices are incorrect.

Case 2



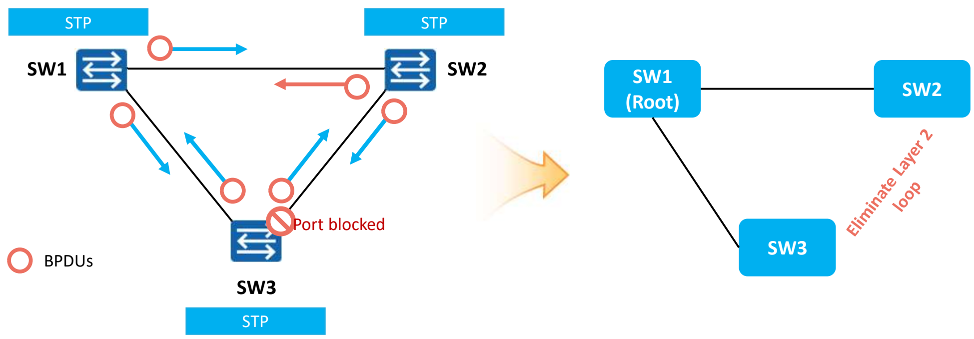
Incorrect manual configurations: For example, the network administrator does not bind the link between SW1 and SW2 to a logical link (aggregation link), causing Layer 2 loops.

- In practice, redundant links may cause loops, and some loops may be caused by human errors.



- Issue 1: Broadcast storm
  - According to the forwarding principle of switches, if a switch receives a broadcast frame or a unicast frame with an unknown destination MAC address from an interface, the switch forwards the frame to all other interfaces except the source interface. If a loop exists on the switching network, the frame is forwarded infinitely. In this case, a broadcast storm occurs and repeated data frames are flooded on the network.
  - In this example, SW3 receives a broadcast frame and floods it. SW1 and SW2 also forward the frame to all interfaces except the interface that receives the frame. As a result, the frame is forwarded to SW3 again. This process continues, causing a broadcast storm. The switch performance deteriorates rapidly and services are interrupted.
- Issue 2: MAC address flapping
  - A switch generates a MAC address table based on source addresses of received data frames and receive interfaces.
  - In this example, SW1 learns and floods the broadcast frame after receiving it from GE0/0/1, forming the mapping between the MAC address 5489-98EE-788A and GE0/0/1. SW2 learns and floods the received broadcast frame. SW1 receives the broadcast frame with the source MAC address 5489-98EE-788A from GE0/0/2 and learns the MAC address again. Then, the MAC address 5489-98EE-788A is switched between GE0/0/1 and GE0/0/2 repeatedly, causing MAC address flapping.

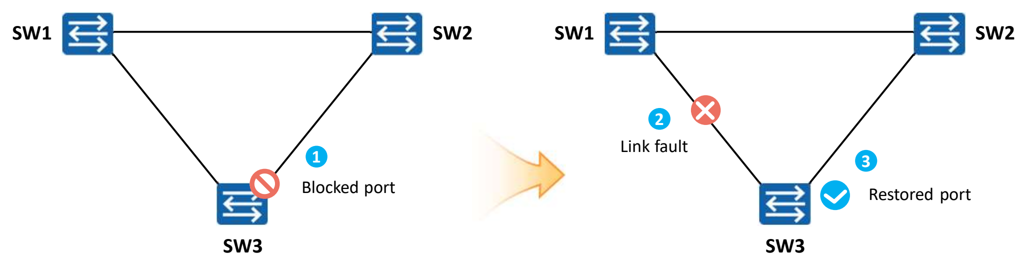
Introduction to STP



When STP is deployed on a network, switches exchange STP BPDUs and calculate a loop-free topology. Finally, one or more ports on the network are blocked to eliminate loops.

- On an Ethernet network, loops on a Layer 2 network may cause broadcast storms, MAC address flapping, and duplicate data frames. STP is used to prevent loops on a switching network.
- STP constructs a tree to eliminate loops on the switching network.
- The STP algorithm is used to detect loops on the network, block redundant links, and prune the loop network into a loop-free tree network. In this way, proliferation and infinite loops of data frames are avoided on the loop network.

STP Can Dynamically Respond to Network Topology Changes and Adjust Blocked Ports



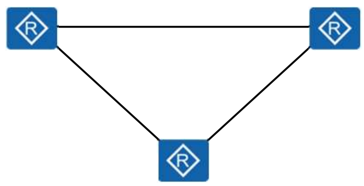
STP running on a switch continuously monitors the network topology. When the network topology changes, STP can detect the changes and automatically adjust the network topology.

Therefore, STP can solve the Layer 2 loop problem and provide a solution for network redundancy.

- As shown in the preceding figure, switches run STP and exchange STP BPDUs to monitor the network topology. Normally, a port on SW3 is blocked to prevent the loop. When the link between SW1 and SW3 is faulty, the blocked port is unblocked and enters the forwarding state.

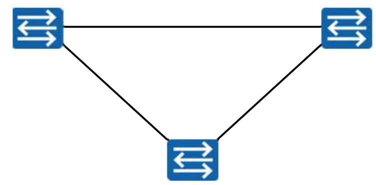
Q&A: Layer 2 and Layer 3 loops

Layer 3 loop



- Common root cause: routing loop
- Dynamic routing protocols have certain loop prevention capabilities.
- The TTL field in the IP packet header can be used to prevent infinite packet forwarding.

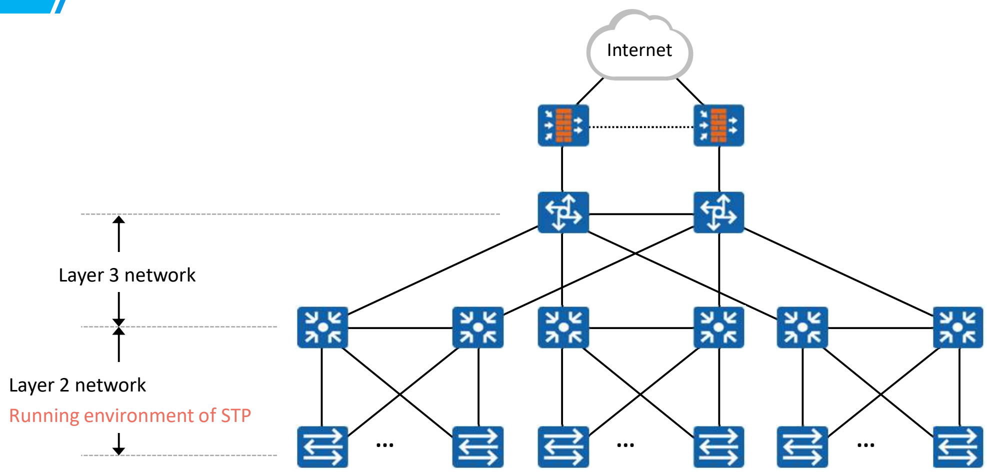
Layer 2 loop



- Common root cause: Layer 2 redundancy is deployed on the network, or cables are incorrectly connected.
- Specific protocols or mechanisms are required to implement Layer 2 loop prevention.
- The Layer 2 frame header does not contain any information to prevent data frames from being forwarded infinitely.

- Common loops are classified into Layer 2 and Layer 3 loops.
- Layer 2 loops are caused by Layer 2 redundancy or incorrect cable connections. You can use a specific protocol or mechanism to prevent Layer 2 loops.
- Layer 3 loops are mainly caused by routing loops. Dynamic routing protocols can be used to prevent loops and the TTL field in the IP packet header can be used to prevent packets from being forwarded infinitely.

Application of STP on a Campus Network



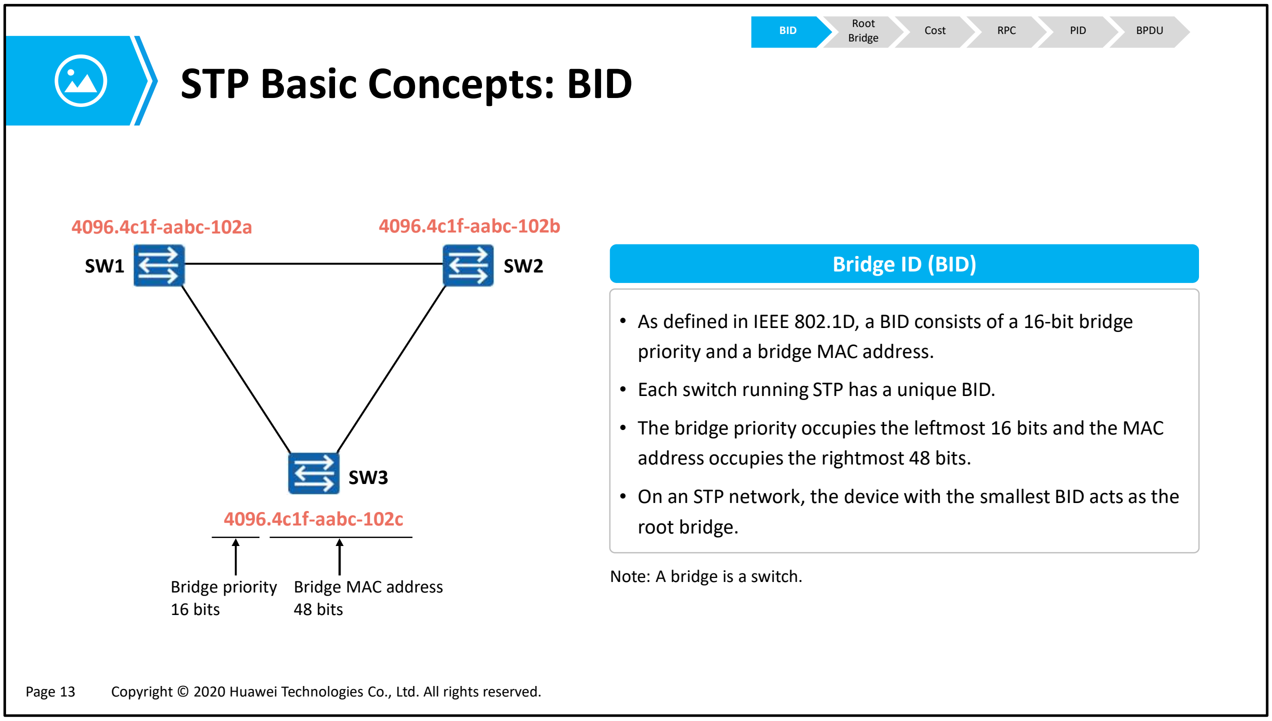
- STP is used on Layer 2 networks of campus networks to implement link backup and eliminate loops.

STP Overview

- STP is used on a LAN to prevent loops.
- Devices running STP exchange information with one another to discover loops on the network, and block certain ports to eliminate loops.
- After running on a network, STP continuously monitors the network status. When the network topology changes, STP can detect the change and automatically respond to the change. In this way, the network status can adapt to the new topology, ensuring network reliability.
- With the growth in scale of LANs, STP has become an important protocol for a LAN.

Contents

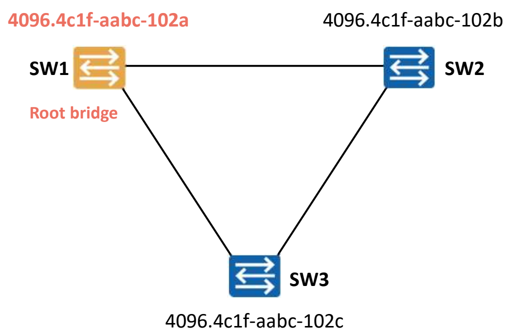
https://support.huawei.com/enterprise/en/doc/EDOC1100096325/99f4e7eb/understanding-stp-rstp



- In STP, each switch has a bridge ID (BID), which consists of a 16-bit bridge priority and a 48-bit MAC address. On an STP network, the bridge priority is configurable and ranges from 0 to 65535. The default bridge priority is 32768. The bridge priority can be changed but must be a multiple of 4096 (N.B. 2^12=4096, so bits 13 to 16 are used for changing priority values). The device with the highest priority (a smaller value indicates a higher priority) is selected as the root bridge. If the priorities are the same, devices compare MAC addresses. A smaller MAC address indicates a higher priority.
- As shown in the figure, the root bridge needs to be selected on the network. The three switches first compare bridge priorities. The bridge priorities of the three switches are 4096. Then the three switches compare MAC addresses. The switch with the smallest MAC address is selected as the root bridge.

- The mac address of a switch is usually the preceding number of the MAC-Address on the first port .
- Eg: The mac-address on the first port GE0/0/1 is **0019.e728.8101**. The mac address of the switch would be **0019.e728.8100.**

STP Basic Concepts: Root Bridge



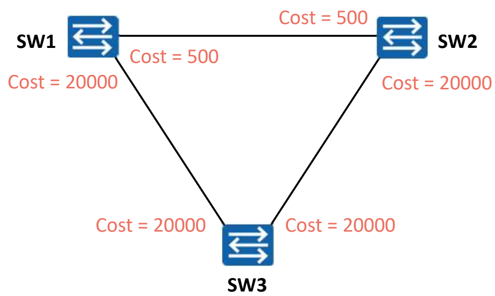
Root Bridge

- One of the main functions of STP is to calculate a loop-free STP tree on the entire switching network.
- The root bridge is the root of an STP network.
- After STP starts to work, it elects a root bridge on the switching network. The root bridge is the key for topology calculation of the spanning tree and is the root of the loop-free topology calculated by STP.
- On an STP network, the device with the smallest BID acts as the root bridge.

  During BID comparison, devices first compare bridge priorities. A smaller priority value indicates a higher priority of a device. The switch with the smallest priority value becomes the root bridge. If priority values are the same, the switch with the smallest MAC address becomes the root bridge.

- The root bridge functions as the root of a tree network.
- It is the logical center, but not necessarily the physical center, of the network. The root bridge changes dynamically with the network topology.
- After network convergence is completed, the root bridge generates and sends configuration BPDUs to other devices at specific intervals. Other devices process and forward the configuration BPDUs to notify downstream devices of topology changes, ensuring that the network topology is stable.

STP Basic Concepts: Cost



Cost

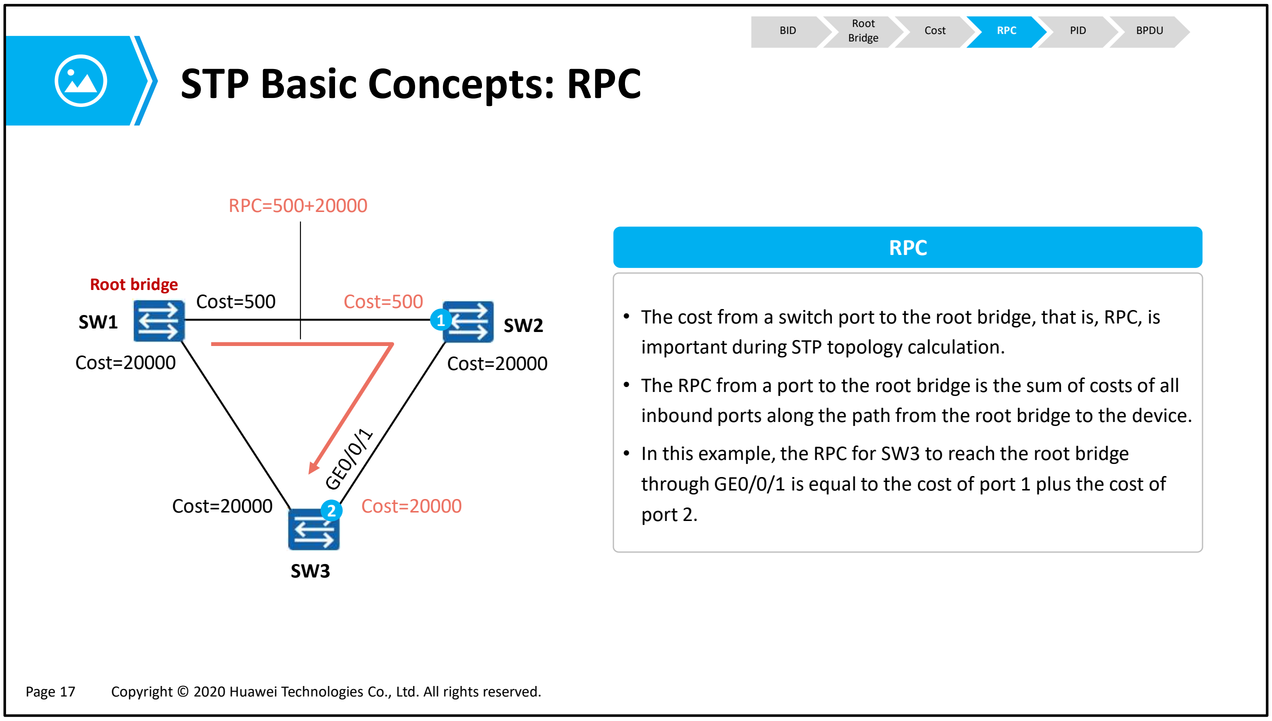
- Each STP-enabled port maintains a cost. The cost of a port is used to calculate the root path cost (RPC), that is, the cost of the path to the root.
- The default cost of a port is related to the rate, working mode, and STP cost calculation method used by a switch.
- A higher port bandwidth indicates a smaller cost.
- You can also run commands to adjust the cost of a port as required.

- Each port on a switch has a cost in STP. By default, a higher port bandwidth indicates a smaller port cost.
- Huawei switches support multiple STP path cost calculation standards to provide better compatibility in scenarios where devices from multiple vendors are deployed. By default, Huawei switches use IEEE 802.1t to calculate the path cost.

STP Basic Concepts: Cost Calculation Methods

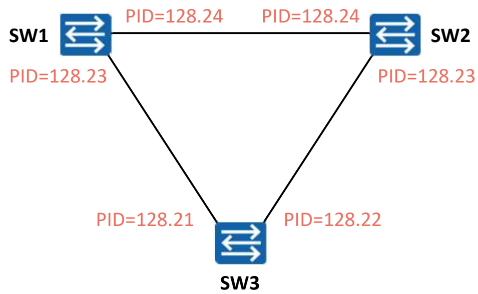
Port Rate	Port Mode	Recommended STP Cost		
		IEEE 802.1d-1998	IEEE 802.1t	Huawei Legacy Standard
100 Mbit/s	Half-duplex	19	200,000	200
	Full-duplex	18	199,999	199
	Aggregated link: two ports	15	100,000	180
1000 Mbit/s	Full-duplex	4	20,000	20
	Aggregated link: two ports	3	10,000	18
10 Gbit/s	Full-duplex	2	2000	2
	Aggregated link: two ports	1	1000	1
40 Gbit/s	Full-duplex	1	500	1
	Aggregated link: two ports	1	250	1
100 Gbit/s	Full-duplex	1	200	1
	Aggregated link: two ports	1	100	1
...				

The cost has a default value and is associated with the port rate. When the device uses different algorithms, the same port rate corresponds to different cost values.



- There may be multiple paths from a non-root bridge to the root bridge. Each path has a total cost, which is the sum of all port costs on this path. A non-root bridge compares the costs of multiple paths to select the shortest path to the root bridge. The path cost of the shortest path is called the root path cost (RPC), and a loop-free tree network is generated. The RPC of the root bridge is 0.

STP Basic Concepts: Port ID

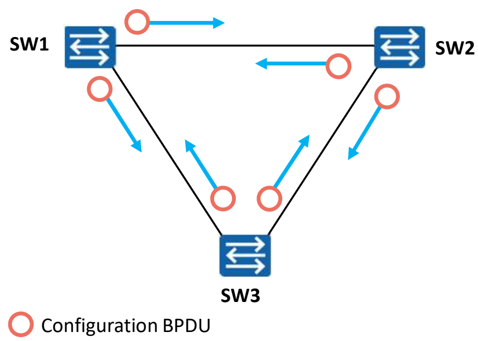


Port ID (PID)

- An STP-enabled switch uses PIDs to identify ports. A PID is used to elect a designated port in a specific scenario.
- A PID consists of the leftmost four bits (port priority) and the rightmost 12 bits (port number).
- The value of the port priority ranges from 0 to 240, with an increment of 16.
- An STP-enabled port maintains a default port priority, which is 128 on Huawei switches. You can run a command to change the priority as required.

- Each port on an STP-enabled switch has a port ID, which consists of the port priority and port number. The value of the port priority ranges from 0 to 240, with an increment of 16. That is, the value must be an integer multiple of 16. By default, the port priority is 128. The PID is used to determine the port role.

STP Basic Concepts: BPDU



Bridge Protocol Data Unit (BPDU)

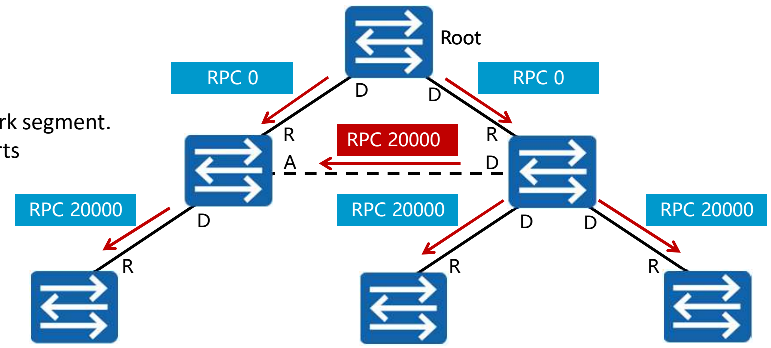
- BPDU is the basis for STP to work normally.
- STP-enabled switches exchange BPDUs that carry important information.
- There are two types of BPDUs:
  - Configuration BPDU
  - Topology Change Notification (TCN) BPDU
- Configuration BPDUs are the key to STP topology calculation. TCN BPDUs are triggered only when the network topology changes.

- Switches exchange BPDUs where information and parameters are encapsulated to calculate spanning trees.
- BPDUs are classified into configuration BPDUs and TCN BPDUs.
- A configuration BPDU contains parameters such as the BID, path cost, and PID. STP selects the root bridge by transmitting configuration BPDUs between switches and determines the role and status of each switch port. Each bridge proactively sends configuration BPDUs during initialization. After the network topology becomes stable, only the root bridge proactively sends configuration BPDUs. Other bridges send configuration BPDUs only after receiving configuration BPDUs from upstream devices.
- A TCN BPDU is sent by a downstream switch to an upstream switch when the downstream switch detects a topology change.

Spanning Tree Port Roles

STP operations:

1. Selects a root bridge.
2. Each non-root switch elects a root port.
3. Select a designated port for each network segment.
4. Blocks non-root and non-designated ports



- Spanning tree supports designated, root and alternate port roles.
- The root path cost enables port roles to be determined.

Page 20

- A converged spanning tree network defines that each interface be assigned a specific port role. Port roles are used to define the behavior of port interfaces that participate within an active spanning tree topology. For the spanning tree protocol, three port roles of designated, root and alternate are defined.
- The designated port is associated with a root bridge or a designated bridge of a LAN segment and defines the downstream path via which Configuration BPDU are forwarded. The root bridge is responsible for the generation of configuration BPDU to all downstream switches, and thus root bridge port interfaces always adopt the designated port role.
- The root port identifies the port that offers the lowest cost path to the root, based on the root path cost. The example demonstrates the case where two possible paths exist back to the root, however only the port that offers the lowest root path cost is assigned as the root port. Where two or more ports offer equal root path costs, the decision of which port interface will be the root port is determined by comparing the bridge ID in the configuration BPDU that is received on each port.
- Any port that is not assigned a designated or root port role is considered an alternate port, and is able to receive BPDU from the designated switch for the LAN segment for the purpose of monitoring the status of the redundant link, but will not process the received BPDU. The IEEE 802.1D-1990 standard for STP originally defined this port role as backup, however this was amended to become the alternate port role within the IEEE 802.1D-1998 standards revision.

Format of Configuration BPDUs

PID	PVI	BPDU Type	Flags	Root ID	RPC	Bridge ID	Port ID	Message Age	Max Age	Hello Time	Forward Delay

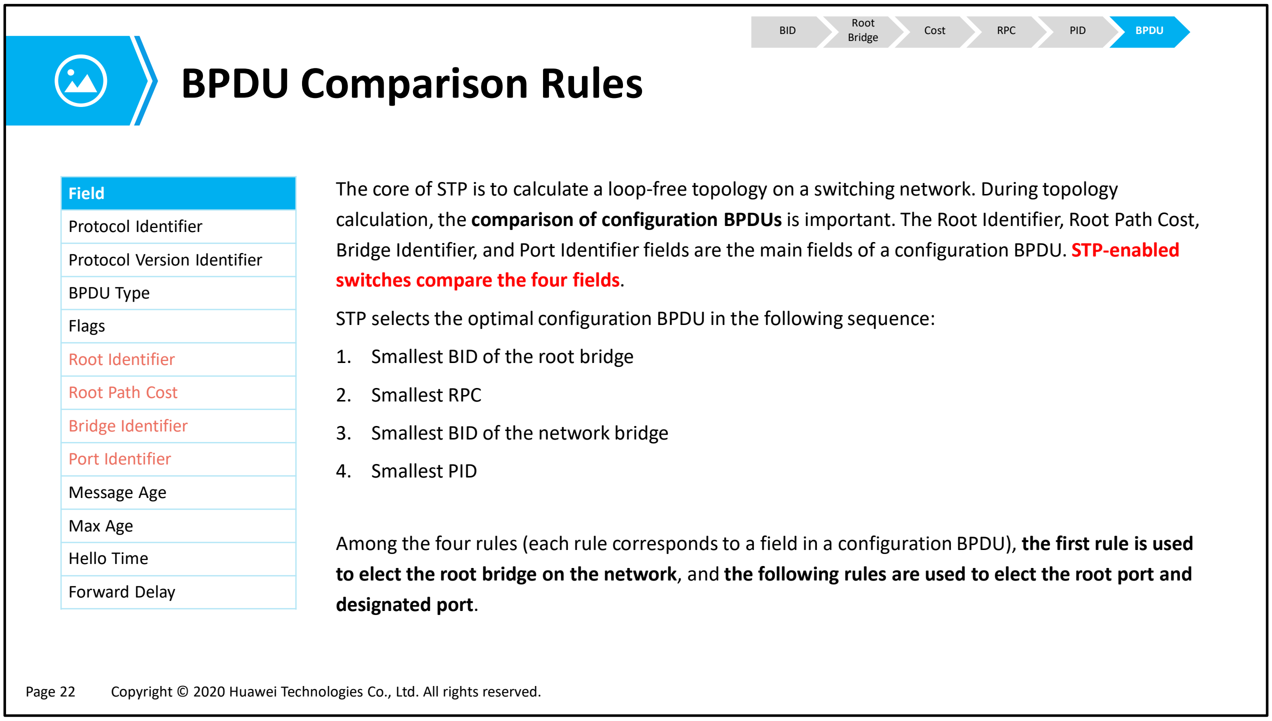
Byte	Field	Description
2	PID	Protocol identifier. For STP, the value of this field is always 0.
1	PVI	Protocol version identifier: For STP, the value of this field is always 0.
1	BPDU Type	Type of BPDUs. The value 0x00 indicates a configuration BPDU and the value 0x80 indicates a TCN BPDU.
1	Flags	Indicates whether the network topology has changed. The rightmost bit is the Topology Change (TC) flag. The leftmost bit is the Topology Change Acknowledgment (TCA) flag.
8	Root D	BID of the root bridge.
4	RPC	STP cost of the path from the current port to the root bridge.
8	Bridge ID	BID of the sender.
2	Port ID	ID of the port that sends this BPDU, which consists of the port priority and port number.
2	Message Age	Number of seconds after a BPDU is sent from the root bridge. The value increases by 1 each time the BPDU passes through a network bridge. It refers to the number of hops to the root bridge.
2	Max Age	If the bridge does not receive any BPDU for a period of time and the lifetime of the BPDU reaches the maximum, the network bridge considers that the link connected to the port is faulty. The default value is 20s.
2	Hello Time	Interval at which the root bridge sends configuration BPDUs. The default value is 2s.
2	Forward Delay	Time that is spent in Listening or Learning state. The default value is 15s.

- Message Age: Records the time that has elapsed since the original BPDU was generated on the root bridge.

  If the configuration BPDU is sent from the root bridge, the Message Age value is 0. Otherwise, the Message Age value is the total time spent to transmit the BPDU from the root bridge to the local bridge, including the transmission delay. The Message Age value of a configuration BPDU increases by 1 each time the configuration BPDU passes through a bridge. Max AgeThe Max Age timer specifies the aging time of BPDUs. This parameter is configurable on the root bridge.

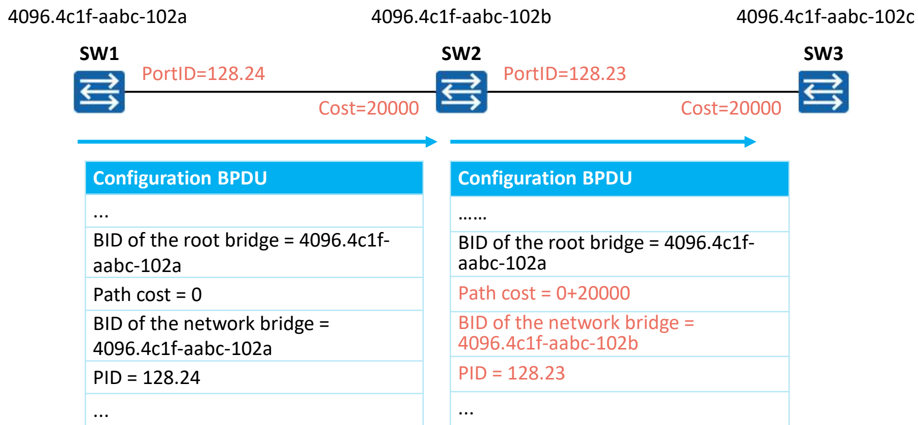
- The Max Age value is encapsulated in configuration BPDUs and transmitted on the entire network to ensure consistency. Upon receipt of a configuration BPDU, a non-root bridge compares the Message Age value with the Max Age value in the received configuration BPDU. If the Message Age value is smaller than or equal to the Max Age value, the non-root bridge forwards the configuration BPDU. If the Message Age value is greater than the Max Age value, the non-root bridge discards the configuration BPDU. In this case, the network is considered too large and the non-root bridge disconnects from the root bridge.

- When a BPDU is received, a counter in initialized at the Message Age value. If the value is better then the one stored, a timer is initialixed with that value. If no BPDUs are received by the stored Max Age, the information is aged-out.



- STP operations:
    1. Selects a root bridge.
    2. Each non-root switch elects a root port.
    3. Select a designated port for each network segment.
    4. Blocks non-root and non-designated ports.
- STP defines three port roles: designated port, root port, and alternate port.
- A designated port is used by a switch to forward configuration BPDUs to the connected network segment. Each network segment has only one designated port. In most cases, each port of the root bridge is a designated port.
- The root port is the port on the non-root bridge that has the optimal path to the root bridge. A switch running STP can have only one root port, but the root bridge does not have any root port.
- If a port is neither a designated port nor a root port, the port is an alternate port. The alternate port is blocked.

Configuration BPDU Forwarding Process

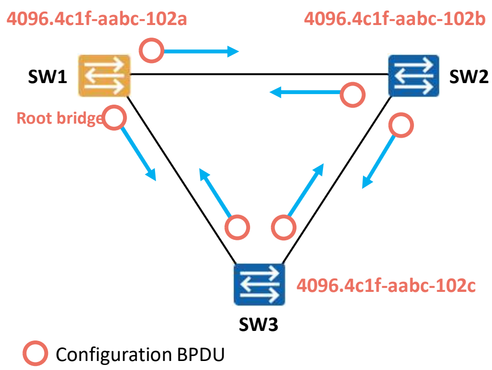


- When a switch starts, it considers itself as the root bridge and sends configuration BPDUs to each other for STP calculation.

Select Root Bridge | Select Root Port | Select Designated Port | Block Non-designated Port

STP Calculation (1)

When a switch starts, it considers itself as the root bridge and sends configuration BPDUs to each other for STP calculation.



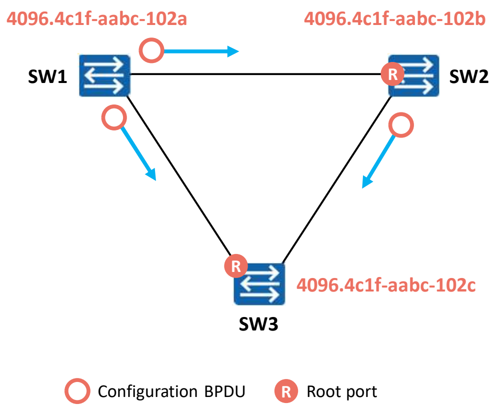
Select a Root Bridge on the Switching Network

- After STP starts to work on a switching network, each switch sends configuration BPDUs to the network. The configuration BPDU contains the BID of a switch.
- The switch with the smallest bridge ID becomes the root bridge.
- Only one root bridge exists on a contiguous STP switching network.
- The role of the root bridge can be preempted.
- To ensure the stability of the switching network, you are advised to plan the STP network in advance and set the bridge priority of the switch that is planned as the root bridge to the minimum value 0.

- What is a root bridge?
  - The root bridge is the root node of an STP tree.
  - To generate an STP tree, first determine a root bridge.
  - It is the logical center, but not necessarily the physical center, of the network.
  - When the network topology changes, the root bridge may also change. (The role of the root bridge can be preempted.)
- Election process:
  1. When an STP-enabled switch is started, it considers itself as the root bridge and declares itself as the root bridge in the BPDUs sent to other switches. In this case, the BID in the BPDU is the BID of each device.
  2. When a switch receives a BPDU from another device on the network, it compares the BID in the BPDU with its own BID.
  3. Switches exchange BPDUs continuously and compare BIDs. The switch with the smallest BID is selected as the root bridge, and other switches are non-root bridges.
  4. As shown in the figure, the priorities of SW1, SW2, and SW3 are compared first. If the priorities of SW1, SW2, and SW3 are the same, MAC addresses are compared. The BID of SW1 is the smallest, so SW1 is the root bridge, and SW2 and SW3 are non-root bridges.
- Note:
  - The role of the root bridge can be preempted. When a switch with a smaller BID joins the network, the network performs STP calculation again to select a new root bridge.

Select Root Bridge | **Select Root Port** | Select Designated Port | Block Non-designated Port

STP Calculation (2)



Select a Root Port on Each Non-root Bridge

- Each non-root bridge selects a root port from its ports.
- A non-root bridge has only one root port.
- When a non-root-bridge switch has multiple ports connected to the network, the root port receives the optimal configuration BPDU.
- The root port is located on each non-root bridge and has the shortest distance away from the root bridge.

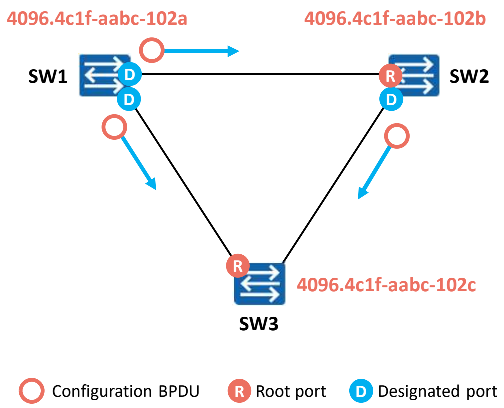
- What is a root port?
  - A non-root bridge may have multiple ports connected to a network. To ensure that a working path from a non-root bridge to a root bridge is optimal and unique, the root port needs to be determined among ports of the non-root bridge. The root port is used for packet exchange between the non-root bridge and the root bridge.
  - After the root bridge is elected, the root bridge still continuously sends BPDUs, and the non-root bridge continuously receives BPDUs from the root bridge. Therefore, the root port closest to the root bridge is selected on all non-root bridges. After network convergence, the root port continuously receives BPDUs from the root bridge.
  - That is, the root port ensures the unique and optimal working path between the non-root bridge and the root bridge.
- Note: A non-root bridge can have only one root port.
- Election process:
  1. A switch has multiple ports connected to a network. Each port receives a BPDU carrying main fields such as RootID, RPC, BID, and PID. The ports compare these fields.
  2. First, RPCs are compared.STP uses the RPC as an important basis to determine the root port. A smaller RPC indicates a higher priority of selecting the root port. Therefore, the switch selects the port with the smallest RPC as the root port.
  3. When the RPCs are the same, BIDs in the BPDUs received by ports of a switch are compared. A smaller BID indicates a higher priority of electing the root port, so the switch

selects the port with the smallest BID as the root port.

4. When the BIDs are the same, PIDs in the BPDUs received by ports of a switch are compared. A smaller PID indicates a higher priority of electing the root port, so the switch selects the port with the smallest PID as the root port.
5. When the PIDs are the same, PIDs of ports on the local switch are compared. A smaller PID indicates a higher priority of electing the root port, so the switch selects the port with the smallest PID as the root port.

Select Root Bridge | Select Root Port | Select Designated Port | Block Non-designated Port

STP Calculation (3)



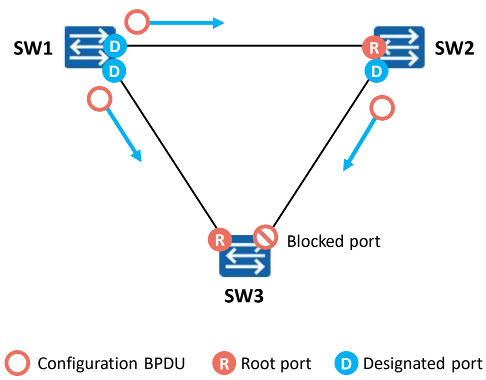
A designated port is elected on each link.

- After the root port is elected, the non-root bridge uses the optimal BPDU received on the port to calculate the configuration BPDU and compares the calculated configuration BPDU with the configuration BPDUs received by all ports except the root port.
  - If the former is better, the port is a designated port.
  - If the latter is better, the port is not a designated port.
- In most cases, all ports on the root bridge are designated ports.

- What is a designated port?
  - The working path between each link and the root bridge must be unique and optimal. When a link has two or more paths to the root bridge (the link is connected to different switches, or the link is connected to different ports of a switch), the switch (may be more than one) connected to the link must determine a unique designated port.
  - Therefore, a designated port is selected for each link to send BPDUs along the link.
- Note: Generally, the root bridge has only designated ports.
- Election process:
  1. The designated port is also determined by comparing RPCs. The port with the smallest RPC is selected as the designated port. If the RPCs are the same, the BID and PID are compared.
  2. First, RPCs are compared. A smaller value indicates a higher priority of electing the designated port, so the switch selects the port with the smallest RPC as the designated port.
  3. If the RPCs are the same, BIDs of switches at both ends of the link are compared. A smaller BID indicates a higher priority of electing the designated port, so the switch selects the port with the smallest BID as the designated port.
  4. If the BIDs are the same, PIDs of switches at both ends of the link are compared. A smaller PID indicates a higher priority of electing the designated port, so the switch selects the port with the smallest PID as the designated port.

Select Root Bridge → Select Root Port → Select Designated Port → **Block Non-designated Port**

STP Calculation (4)

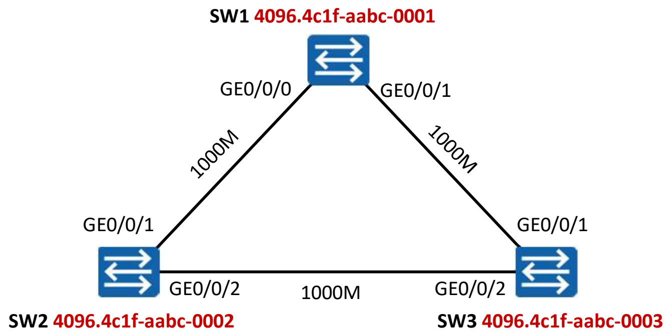


Block Non-designated Port

- On a switch, a port that is neither a root port nor a designated port is called a non-designated port.
- The last step of STP operations is to block the non-designated port on the network. After this step is complete, the Layer 2 loop on the network is eliminated.

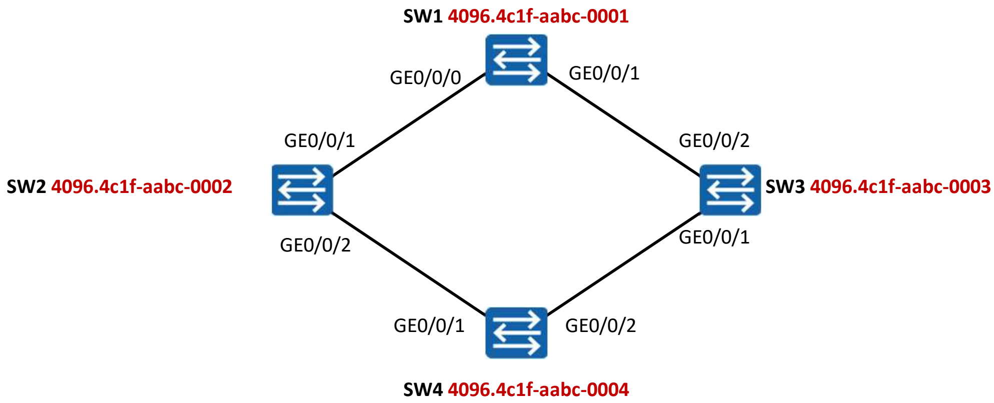
- What is a non-designated port (alternate port)?
  - After the root port and designated port are determined, all the remaining non-root ports and non-designated ports on the switch are called alternate ports.
- Blocking alternate ports
  - STP logically blocks the alternate ports. That is, the ports cannot forward the frames (user data frames) generated and sent by terminal computers.
  - Once the alternate port is logically blocked, the STP tree (loop-free topology) is generated.
- Note:
  - The blocked port can receive and process BPDUs.
  - The root port and designated port can receive and send BPDUs and forward user data frames.

Quiz 1: Identify the Root Bridge and Port Roles



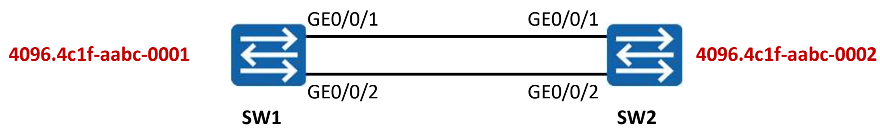
- As shown in the figure, the root bridge is selected first. If the three switches have the same bridge priority, the switch with the smallest MAC address is selected as the root bridge.
- GE0/0/1 on SW2 is closest to the root bridge and has the smallest RPC, so GE0/0/1 on SW2 is the root port. Similarly, GE0/0/1 on SW3 is also the root port.
- Then designated ports are selected. SW1 is elected as the root bridge, so GE0/0/0 and GE0/0/1 on SW1 are designated ports. GE0/0/2 on SW2 receives configuration BPDUs from SW3 and compares the BIDs of SW2 and SW3. SW2 has a higher BID than SW3, so GE0/0/2 on SW2 is the designated port.
- GE0/0/2 on SW3 is the alternate port.

Quiz 2: Identify the Root Bridge and Port Roles in the Following Topology



- As shown in the figure, the root bridge is selected first. If the four switches have the same bridge priority, the switch with the smallest MAC address is selected as the root bridge.
- GE0/0/1 on SW2 is closest to the root bridge and has the smallest RPC. Therefore, GE0/0/1 on SW2 is the root port. Similarly, GE0/0/2 on SW3 is the root port. The two ports on SW4 have the same RPC. The BID of SW2 connected to GE0/0/1 on SW4 and the BID of SW3 connected to GE0/0/2 on SW4 are compared. The smaller the BID, the higher the priority. Given this, GE0/0/1 on SW4 is selected as the root port.
- Then designated ports are selected. SW1 is elected as the root bridge, so GE0/0/0 and GE0/0/1 on SW1 are designated ports. GE0/0/2 on SW2 receives configuration BPDUs from SW4 and compares the BIDs of SW2 and SW4. SW2 has a higher BID than SW4, so GE0/0/2 on SW2 is the designated port, and GE0/0/1 on SW3 is the designated port.
- GE0/0/2 on SW4 is the alternate port.

Quiz 3: Identify the Root Bridge and Port Roles in the Following Topology

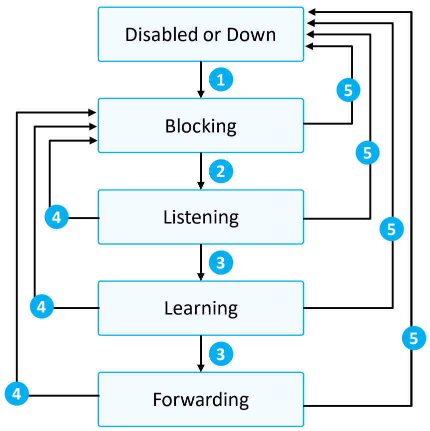


- As shown in the figure, the root bridge is selected first. If the two switches have the same bridge priority, the switch with a smaller MAC address is selected as the root bridge. SW1 is selected as the root bridge.
- Then the root port is selected. The two ports on SW2 have the same RPC and BID. The PIDs of the two ports are compared. The PID of G0/0/1 on SW2 is 128.1, and the PID of G0/0/2 on SW2 is 128.2. The smaller the PID, the higher the priority. Therefore, G0/0/1 of SW2 is the root port.
- SW1 is the root bridge, so GE0/0/1 and GE0/0/2 on SW1 are designated ports.
- GE0/0/2 on SW2 is the alternate port.

STP Port States

Port State	Description
Disabled	The port cannot send or receive BPDUs or service data frames. That is, the port is Down.
Blocking	The port is blocked by STP. A blocked port cannot send BPDUs but **listens to BPDUs**. In addition, the blocked port cannot send or receive service data frames or learn MAC addresses.
Listening	STP considers the port in Listening state as the root port or designated port, but the port is still in the STP calculation process. In this case, the port can **send and receive BPDUs** but cannot send or receive service data frames or learn MAC addresses.
Learning	A port in Learning state **listens to service data frames** but cannot forward them. After receiving service data frames, **the port learns MAC addresses**.
Forwarding	A port in Forwarding state **can send and receive service data frames and process BPDUs**. Only the root port or designated port can enter the Forwarding state.

STP Port State Transition



1. When a port is initialized or activated, it automatically enters the blocking state.
2. The port is elected as the root port or designated port and automatically enters the Listening state.
3. The **Forward Delay timer expires** and the port is still the root port or designated port.
4. The port is no longer the root port or designated port.
5. The port is disabled or the link fails.

- The figure shows the STP port state transition. The STP-enabled device has the following five port states:
- Forwarding: A port can forward user traffic and BPDUs. Only the root port or designated port can enter the Forwarding state.
- Learning: When a port is in Learning state, a device creates MAC address entries based on user traffic received on the port but does not forward user traffic through the port. The Learning state is added to prevent temporary loops.
- Listening: A port in Listening state can forward BPDUs, but cannot forward user traffic.
- Blocking: A port in Blocking state can only receive and process BPDUs, but cannot forward BPDUs or user traffic. The alternate port is in Blocking state.
- Disabled: A port in Disabled state does not forward BPDUs or user traffic.

Also each STP switch ports move through these five states in the time below:

- **Initialization to Blocking (0 seconds)**
- **Blocking to Listening (20 seconds)**
- **Listening to Learning (15 seconds)**
- **Learning to Forwarding (15 seconds)**

- **Disabled (A network administrator can disable the switch port at any time)**

When STP is enabled every switch in the network starts in the blocking state and transitions to listening and learning states. If properly configured, the ports then stabilize to the forwarding or blocking state until a change in the network is made. A port in the blocking state doesn't participate in frame forwarding, and after initialization, a BPDU is sent to each port in the switch. The switch assumes it's the root until it exchanges BPDU's with other switches in the network. This BPDU exchange establishes which switch in the network is the root switch. If only one switch resides in the network, no exchange occurs, and after the forward delay timer expires, the ports move to a listening state.

- A switch always enters the blocking state following switch initialization.

A port in the blocking state:

- Discards frames received from the attached network segment
- Discards frames switched from another port for forwarding
- Doesn't incorporate a host location into its address database; because there is no learning needed at this time.
- Receives BPDU's from the network segment and directs them to the switch system module.
- Unlike ports in the listening, learning, and forwarding state, a port in the blocking state doesn't process BPDU's received from the switch system module.
- It does receive and respond to network management messages, such as a network administrator disabling the port.

Always keep in mind that a port in a blocking state will not send or receive any data traffic across the network segment, but it will listen to STP BPDU messages. Also, when a switch running STP is powered on, all ports are in the blocking state.

- **Blocking:**

The port remains in the blocking state for *20 seconds* before transitioning to the listening state.

- **Listening:**

The port is listening to STP messages in the form of BPDU's and determining how the network topology is configured. While in the listening state, the port is not forwarding frames. The port is in listening state for *15 seconds* before transitioning to the learning state.

- **Learning**

The port is adding MAC addresses to the MAC address table. While in the learning state, the port is not forwarding frames.   The port is in the learning state for *15 seconds* before transitioning to the forwarding state.

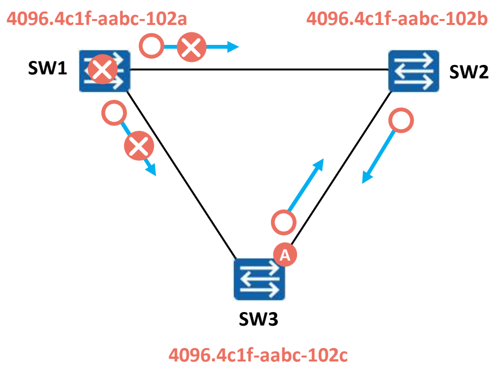
- **Forwarding**

The port is sending and receiving data across the network segment as normal, and after the network has converged on a topology, the port will be in either the forwarding or blocking state.

- **Disabled:**

The port can be administratively disabled at any time by the network administrator, as I mentioned to you before. At that point it will not receive BPDU's or forward traffic across any network segments. The switch cannot put a port into a disable state and therefore cannot take a port out of the disable state; only the network administrator can enable or disable a port.

Topology Change: Root Bridge Fault

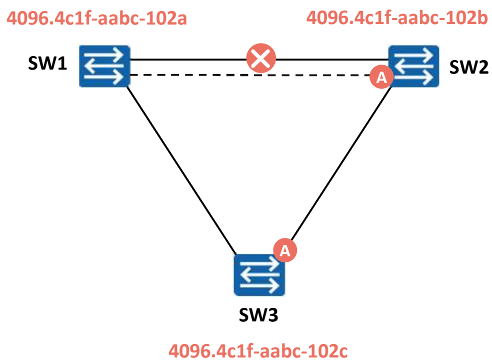


Root Bridge Fault Rectification Process

1. SW1 (root bridge) is faulty and stops sending BPDUs.
2. **SW2 waits for the Max Age timer (20s) to expire**. In this case, the record about the received BPDUs becomes invalid, and SW2 cannot receive new BPDUs from the root bridge. SW2 learns that the upstream device is faulty.
3. Non-root bridges send configuration BPDUs to each other to elect a new root bridge.
4. **After re-election**, port A of **SW3** transitions to the Forwarding state **after two intervals of the Forward Delay timer** (the default interval is 15s).
   - A non-root bridge starts root bridge re-election after BPDUs age.
   - Due to the root bridge failure, **it takes about 50s to recover** from a root bridge failure.

- Root bridge fault:
  - On a stable STP network, a non-root bridge periodically receives BPDUs from the root bridge.
  - If the root bridge fails, the downstream switch stops sending BPDUs. As a result, the downstream switch cannot receive BPDUs from the root bridge.
  - If the downstream switch does not receive BPDUs, the Max Age timer (the default value is 20s) expires. As a result, the record about the received BPDUs becomes invalid. In this case, the non-root bridges send configuration BPDUs to each other to elect a new root bridge.
- Port state:
  - The alternate port of SW3 enters the Listening state from the Blocking state after 20s and then enters the Learning state. Finally, the port enters the Forwarding state to forward user traffic.
- Convergence time:
  - It takes about 50s to recover from a root bridge failure, which is equal to the value of the Max Age timer plus twice the value of the Forward Delay timer.

Topology Change: Direct Link Fault



Direct Link Fault Rectification Process

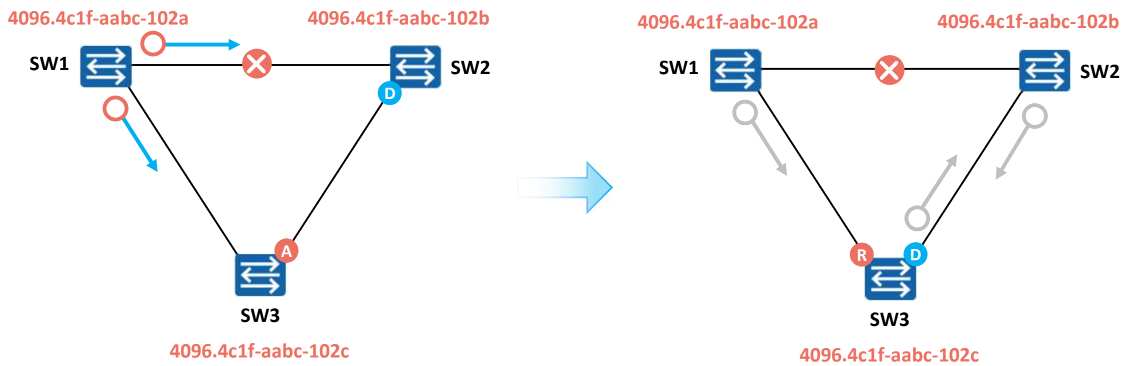
On a stable network, when SW2 detects that the link of the root port is faulty, the alternate port of SW2 enters the Forwarding state **after twice the value of the Forward Delay timer** (the default value is 15s).

- After SW2 detects a fault on the direct link, it switches the alternate port to the root port.
- If a direct link fails, **the alternate port restores to the Forwarding state after 30s**.

- Direct link fault:
  - When two switches are connected through two links, one is the active link and the other is the standby link.
  - When the network is stable, SW2 detects that the link of the root port is faulty, and the alternate port enters the Forwarding state.
- Port state:
  - The alternate port transitions from the Blocking state to the Listening, Learning, Forwarding states in sequence.
- Convergence speed:
  - If a direct link fails, the alternate port restores to the Forwarding state after 30s.

Topology Change: Indirect Link Fault

- When the indirect link fails, the alternate port on SW3 restores to the Forwarding state. It takes about **50s to recover from an indirect link failure**.



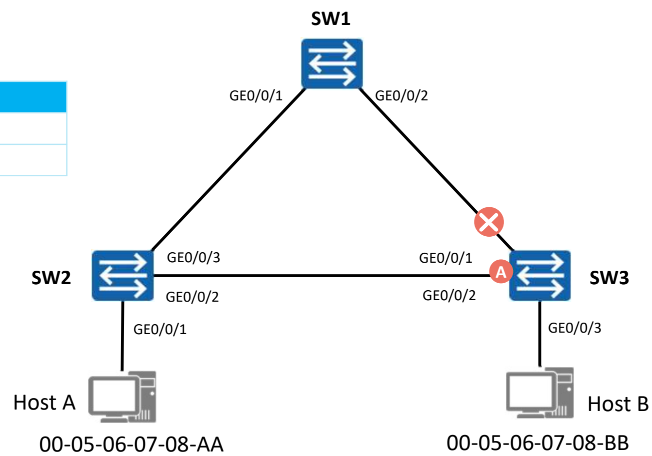
- Indirect link fault:
  - On a stable STP network, a non-root bridge periodically receives BPDUs from the root bridge.
  - If the link between SW1 and SW2 is faulty (not a physical fault), SW2 cannot receive BPDUs from SW1. The Max Age timer (the default value is 20s) expires. As a result, the record about the received BPDUs becomes invalid.
  - In this case, the non-root bridge SW2 considers that the root bridge fails and considers itself as the root bridge. Then SW2 sends its own configuration BPDU to SW3 to notify SW3 that it is the new root bridge.
  - During this period, the alternate port of SW3 does not receive any BPDU that contains the root bridge ID. After the Max Age timer expires, the port enters the Listening state and starts to forward the BPDU that contains the root bridge ID from the upstream device to SW2.
  - After the Max Age timer expires, SW2 and SW3 receive BPDUs from each other almost at the same time and perform STP recalculation. SW2 finds that the BPDU sent by SW3 is superior, so it does not declare itself as the root bridge and re-determines the port role.
- Port state:
  - The alternate port of SW3 enters the Listening state from the Blocking state after 20s and then enters the Learning state. Finally, the port enters the Forwarding state to forward user traffic.
- Convergence time:
  - It takes about 50s to recover from an indirect link failure, which is equal to the value of the Max Age timer plus twice the value of the Forward Delay timer.

The MAC Address Table Is Incorrect Because the Topology Changes

MAC address table SW2

MAC	Port
00-05-06-07-08-AA	GE0/0/1
00-05-06-07-08-BB	GE0/0/3

As shown in the figure, the root port of SW3 is faulty, causing the spanning tree topology to re-converge. After the spanning tree topology re-converges, Host B cannot receive frames sent by Host A. This is because switches forward data frames based on the MAC address table. By default, the aging time of MAC address entries is 300s. How is forwarding restored rapidly?



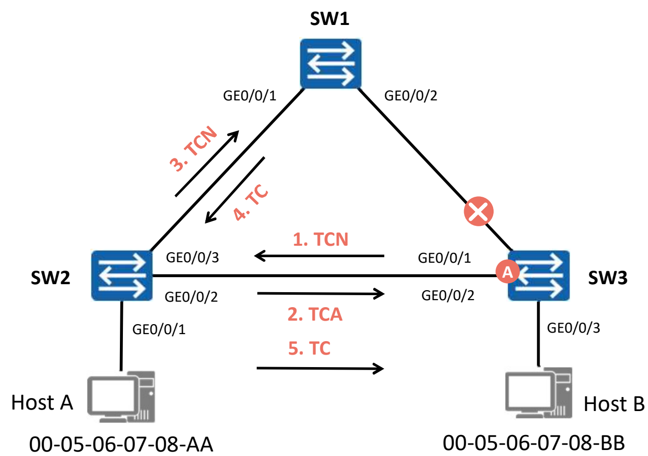
- On a switching network, a switch forwards data frames based on the MAC address table. By default, the aging time of MAC address entries is 300 seconds. If the spanning tree topology changes, the forwarding path of the switch also changes. In this case, the entries that are not aged in a timely manner in the MAC address table may cause data forwarding errors. Therefore, the switch needs to update the MAC address entries in a timely manner after the topology changes.
- In this example, the MAC address entry on SW2 defines that packets can reach Host A through GE0/0/1 and reach Host B through GE0/0/3. The root port of SW3 is faulty, causing the spanning tree topology to re-converge. After the spanning tree topology re-converges, Host B cannot receive frames sent by Host A. This is because the aging time of MAC address entries is 300s. After a frame sent from Host A to Host B reaches SW2, SW2 forwards the frame through GE0/0/3.

The MAC Address Table Is Incorrect Because the Topology Changes

MAC address table SW2

MAC	Port
00-05-06-07-08-AA	GE0/0/1
00-05-06-07-08-BB	GE0/0/3
00-05-06-07-08-BB	GE0/0/2

- TCN BPDUs are generated when the network topology changes.
- Packet format: protocol identifier, version number, and type
- Topology change: The TCA and TC bits in the Flags field of configuration BPDUs are used.



- When the network topology changes, the root bridge sends TCN BPDUs to notify other devices of the topology change. The root bridge generates TCs to instruct other switches to age existing MAC address entries.
- The process of topology change and MAC address entry update is as follows:
  1. After SW3 detects the network topology change, it continuously sends TCN BPDUs to SW2.
  2. After SW2 receives the TCN BPDUs from SW3, it sets the TCA bit in the Flags field of the BPDUs to 1 and sends the BPDUs to SW3, instructing SW3 to stop sending TCN BPDUs.
  3. SW2 forwards the TCN BPDUs to the root bridge.
  4. SW1 sets the TC bit in the Flags field of the configuration BPDU to 1 and sends the configuration BPDU to instruct the downstream device to change the aging time of MAC address entries from 300s to the value of the Forward Delay timer (15s by default).
  5. The incorrect MAC address entries on SW2 are automatically deleted after 15s at most. Then, SW2 starts to learn MAC address entries again and forwards packets based on the learned MAC address entries.

Contents

Basic STP Configuration Commands (1)

1. Configure a working mode.

   [Huawei] **stp mode** { **stp** | **rstp** | **mstp** }

   The switch supports three working modes: STP, RSTP, and Multiple Spanning Tree Protocol (MSTP). By default, a switch works in MSTP mode. On a ring network running only STP, the working mode of a switch is configured as STP; on a ring network running RSTP, the working mode of a switch is configured as RSTP.

2. (Optional) Configure the root bridge.

   [Huawei] **stp root primary**

   Configure the switch as the root bridge. By default, a switch does not function as the root bridge of any spanning tree. After you run this command, the priority value of the switch is set to 0 and cannot be changed.

3. (Optional) Configure the switch as the secondary root bridge.

   [Huawei] **stp root secondary**

   Configure the switch as the secondary root bridge. By default, a switch does not function as the secondary root bridge of any spanning tree. After you run this command, the priority value of the switch is set to 4096 and cannot be changed.

Basic STP Configuration Commands (2)

1. (Optional) Configure the STP priority of a switch.

   [Huawei] **stp priority** *priority*

   By default, the priority value of a switch is 32768.

2. (Optional) Configure a path cost for a port.

   [Huawei] **stp pathcost-standard** { **dot1d-1998** | **dot1t** | **legacy** }

   Configure a path cost calculation method. By default, the IEEE 802.1t standard (**dot1t**) is used to calculate path costs. All switches on a network must use the same path cost calculation method.

   [Huawei-GigabitEthernet0/0/1] **stp cost** *cost*

   Set the path cost of the port.

Basic STP Configuration Commands (3)

1. (Optional) Configure a priority for a port.

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

Configure a priority for a port. By default, the priority of a switch port is 128.

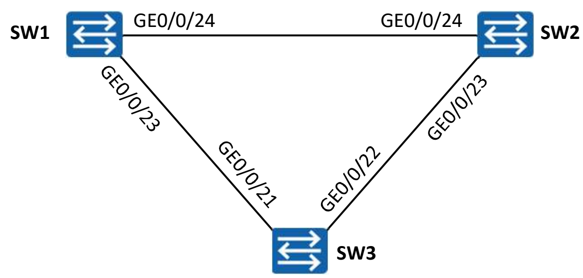
2. Enable STP, RSTP, or MSTP.

```
[Huawei] stp enable
```

Enable STP, RSTP, or MSTP on a switch. By default, STP, RSTP, or MSTP is enabled on a switch.

- Note that the effect of the command stp priority depends on the wiev from which it is issued. In sytem view it sets the switch priority. In a port view it sets the port priority.

Case 1: Basic STP Configurations



- Deploy STP on the three switches to eliminate Layer 2 loops on the network.
- Configure SW1 as the root bridge and block GE0/0/22 on SW3.

SW1 configuration:

```
[SW1] stp mode stp
[SW1] stp enable
[SW1] stp priority 0
```

SW2 configuration:

```
[SW2] stp mode stp
[SW2] stp enable
[SW2] stp priority 4096
```

SW3 configuration:

```
[SW3] stp mode stp
[SW3] stp enable
```


Case 1: Basic STP Configurations

Check brief information about STP states of ports on SW3.

```
<SW3> display stp brief
 MSTID   Port                     Role    STP State      Protection
   0     GigabitEthernet0/0/21    ROOT    FORWARDING     NONE
   0     GigabitEthernet0/0/22    ALTE    DISCARDING     NONE
```

Contents

Disadvantages of STP

- STP ensures a loop-free network but is slow to converge, leading to service quality deterioration. If the network topology changes frequently, connections on the STP network are frequently torn down, causing frequent service interruption.
- STP does not differentiate between port roles according to their states, making it difficult for less experienced administrators to learn about and deploy this protocol.
    - Ports in Listening, Learning, and Blocking states are the same for users because none of these ports forwards service traffic.
    - In terms of port use and configuration, the essential differences between ports lie in the port roles but not port states.
    - Both root and designated ports can be in Listening state or Forwarding state, so the port roles cannot be differentiated according to their states.
- The STP algorithm does not determine topology changes until the timer expires, delaying network convergence.
- The STP algorithm requires the root bridge to send configuration BPDUs after the network topology becomes stable, and other devices process and spread the configuration BPDUs through the entire network. This also delays convergence.

RSTP Overview

- RSTP defined in IEEE 802.1w is an enhancement to STP. RSTP optimizes STP in many aspects, provides **faster convergence**, and is **compatible with STP**.
- RSTP introduces **new port roles**. When the root port fails, the switch can enable the alternate port to obtain an alternate path from the designated bridge to the root bridge. RSTP defines **three states for a port** based on whether the port forwards user traffic and learns MAC addresses. In addition, RSTP introduces the **edge port**. The port connecting a switch to a terminal is configured as an edge port that enters the Forwarding state immediately after initialization, thus improving the working efficiency.

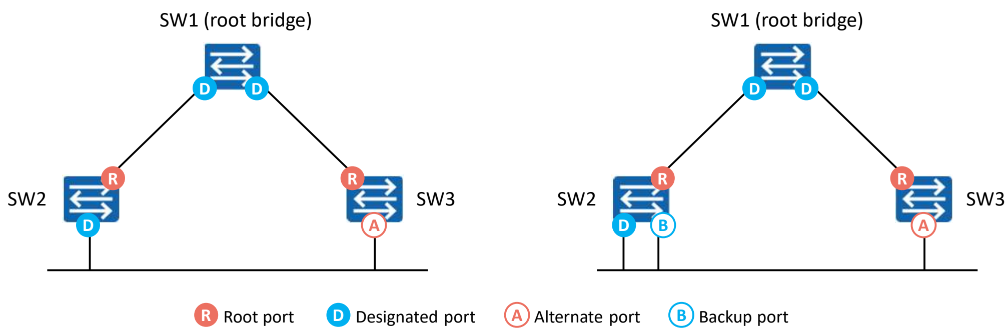
- The IEEE 802.1w standard released in 2001 defines RSTP. RSTP is an improvement on STP and implements fast network topology convergence.
- RSTP is evolved from STP and has the same working mechanism as STP. When the topology of a switching network changes, RSTP can use the Proposal/Agreement mechanism to quickly restore network connectivity.
- RSTP removes three port states, defines two new port roles, and distinguishes port attributes based on port states and roles. In addition, RSTP provides enhanced features and protection measures to ensure network stability and fast convergence.
- RSTP is backward compatible with STP, which is not recommended because STP slow convergence is exposed.
- Improvements made in RSTP:
    - RSTP processes configuration BPDUs differently from STP.
        - When the topology becomes stable, the mode of sending configuration BPDUs is optimized.
        - RSTP uses a shorter timeout interval of BPDUs.
        - RSTP optimizes the method of processing inferior BPDUs.
    - RSTP changes the configuration BPDU format and uses the Flags field to describe port roles.
    - RSTP topology change processing: Compared with STP, RSTP is optimized to accelerate the response to topology changes.

Improvements Made in RSTP

- RSTP processes configuration BPDUs differently from STP.
  - When the topology becomes stable, **the mode of sending configuration BPDUs is optimized.**
  - RSTP uses a **shorter timeout** interval of BPDUs.
  - RSTP optimizes the method of processing inferior BPDUs.
- RSTP changes the configuration BPDU format and **uses the Flags field to describe port roles**.
- RSTP topology change processing: Compared with STP, RSTP is optimized to **accelerate the response to topology changes**.

Port Roles in RSTP

- RSTP adds port roles to help understand RSTP and simplify RSTP deployment.



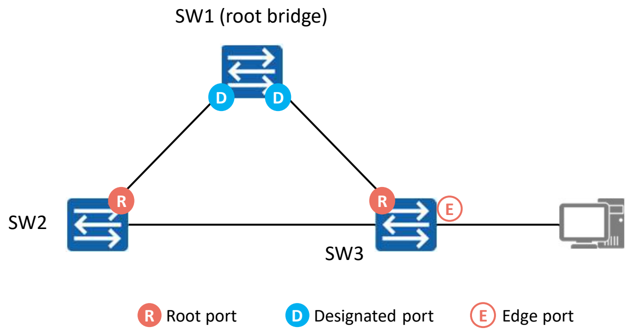
RSTP defines four port roles: root port, designated port, alternate port, and backup port.

A backup path to downstream nodes, where redundant links exist on the same LAN segment as the designated port.
An alternate path to the root bridge that differs from the path provided by the root port of the switch.

- From the perspective of configuration BPDU transmission:
  - An alternate port is blocked after learning a configuration BPDU sent from another network bridge.
  - A backup port is blocked after learning a configuration BPDU sent from itself.
- From the perspective of user traffic:
  - An alternate port acts as a backup of the root port and provides an alternate path from the designated bridge to the root bridge.
  - A backup port backs up a designated port and provides a backup path from the root bridge to the related network segment.

Edge Port

- An edge port is located at the edge of a region and does not connect to any switching device.



Generally, an edge port is directly connected to a user terminal. The edge port can transition from the Disabled state to the Forwarding state.

- In STP, it takes 15 seconds for the port of a switch connected to a user terminal to transition from Disabled to Forwarding. During this period, the user terminal cannot access the Internet. If the network changes frequently, the Internet access status of the user terminal is unstable.
- An edge port is directly connected to a user terminal and is not connected to any switching device. An edge port does not receive or process configuration BPDUs and does not participate in RSTP calculation. It can transition from Disabled to Forwarding without any delay. An edge port becomes a common STP port once it receives a configuration BPDU. The spanning tree needs to be recalculated, which leads to network flapping.

RST BPDU

PID	PVI	BPDU Type	Flags	Root ID	RPC	Bridge ID	Port ID	Message Age	Max Age	Hello Time	Fwd Delay

Bit7	Bit6	Bit5	Bit4	Bit3	Bit2	Bit1	Bit0
TCA	Agreement	Forwarding	Learning	Port Role		Proposal	TC

Port Role = 00 Unknown
01 Alternate/Backup Port
10 Root Port
11 Designated Port

- Unused fields of the STP BPDU are active within RSTP.
- New capabilities are introduced as part of RSTP.

- The BPDU format employed in STP is also applied to RSTP with variance in some of the general parameters. In order to distinguish STP configuration BPDU from Rapid Spanning Tree BPDU, thus known as RST BPDU, the BPDU type is defined. STP defines a configuration BPDU type of 0 (0x00) and a Topology Change Notification BPDU (TCN BPDU) of 128 (0x80), RST BPDU are identified by the BPDU type value 2 (0x02). Within the flags field of the RST BPDU, additional parameter designations are assigned to the BPDU fields.
- The flags field within STP implemented only the use of the Topology Change (TC) and Acknowledgement (TCA) parameters as part of the Topology Change process while other fields were reserved. The RST BPDU has adopted these fields to support new parameters. These include flags indicating the proposal and agreement process employed by RSTP for rapid convergence, the defining of the port role, and the port state.

Port States in RSTP

- RSTP deletes two port states defined in STP, reducing the number of port states to three.
  - If the port does not forward user traffic or learn MAC addresses, it is in Discarding state.
  - If the port does not forward user traffic but learns MAC addresses, it is in Learning state.
  - If the port forwards user traffic and learns MAC addresses, it is in Forwarding state.

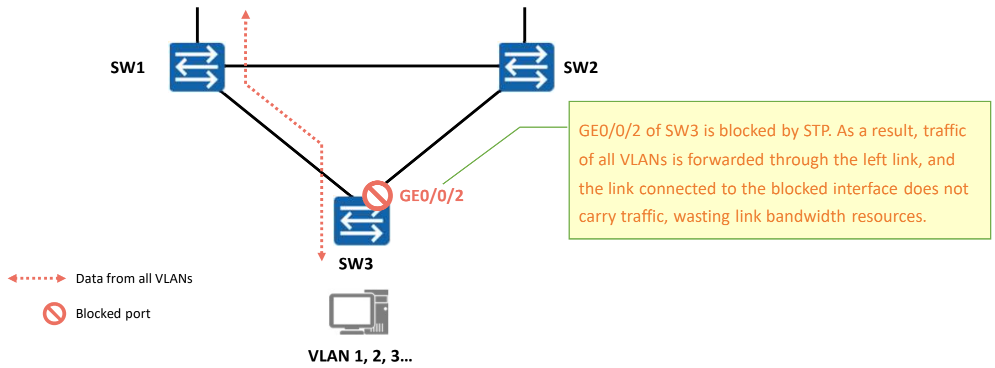
STP Port State	RSTP Port State	Port Role
Forwarding	Forwarding	Root port or designated port
Learning	Learning	Root port or designated port
Listening	Discarding	Root port or designated port
Blocking	Discarding	Alternate port or backup port
Disabled	Discarding	Disabled port

- RSTP deletes two port states defined in STP, reducing the number of port states to three.
  1. A port in Discarding state does not forward user traffic or learn MAC addresses.
  2. A port in Learning state does not forward user traffic but learns MAC addresses.
  3. A port in Forwarding state forwards user traffic and learns MAC addresses.

Contents

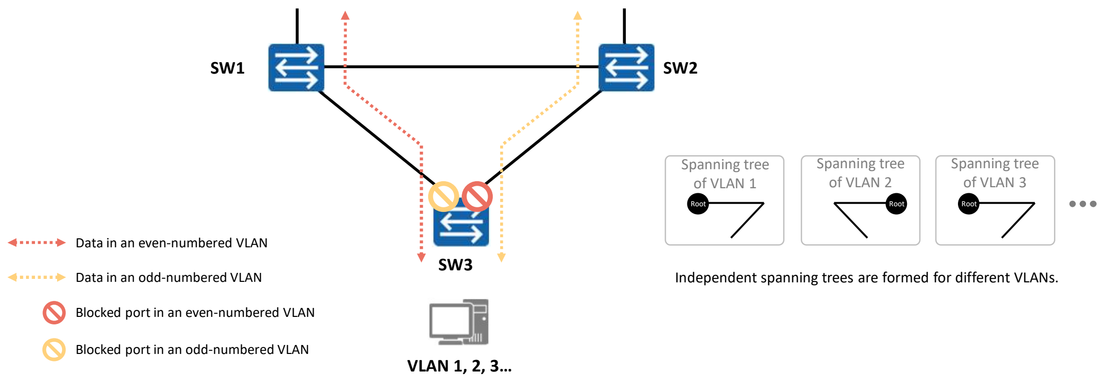
Defects of STP/RSTP: All VLANs Share One Spanning Tree

- RSTP, an enhancement to STP, allows for fast network topology convergence.
- STP and RSTP both have a defect: All VLANs on a LAN share one spanning tree. As a result, inter-VLAN load balancing cannot be performed, and blocked links cannot transmit any traffic, which may lead to VLAN packet transmission failures.



VBST

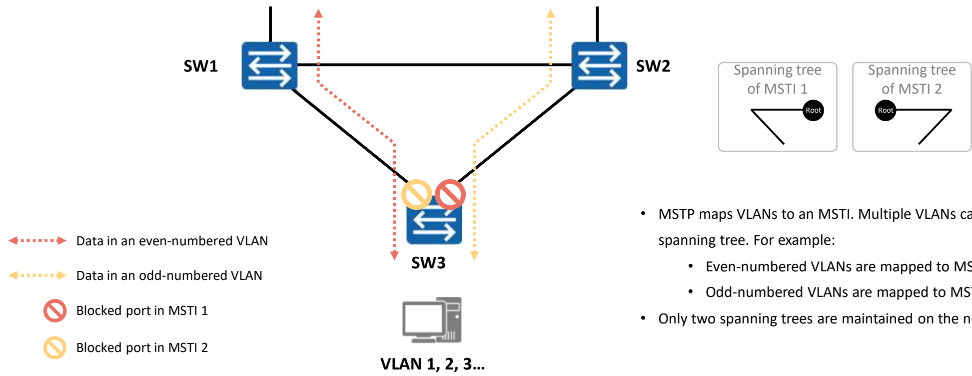
- Huawei provides the VLAN-based Spanning Tree (VBST). VBST constructs a spanning tree in each VLAN so that traffic from different VLANs is load balanced along different spanning trees.



- VBST brings in the following benefits:
  - Eliminates loops.
  - Implements link multiplexing and load balancing, and therefore improves link use efficiency.
  - Reduces configuration and maintenance costs.
- If a great number of VLANs exist on a network, spanning tree computation for each VPN consumes a huge number of switch processor resources.

MSTP

- To fix the defects, the IEEE released the 802.1s standard that defines the **Multiple Spanning Tree Protocol (MSTP)** in 2002.
- MSTP is compatible with STP and RSTP, and can rapidly converge traffic and provides multiple paths to load balance VLAN traffic.



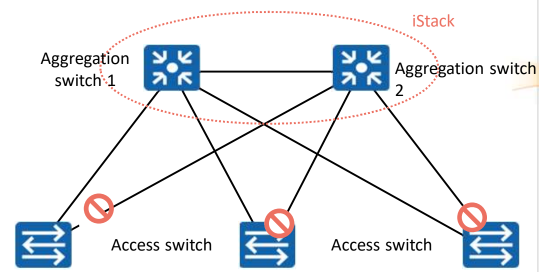
- MSTP maps VLANs to an MSTI. Multiple VLANs can share one spanning tree. For example:
  - Even-numbered VLANs are mapped to MSTI 1.
  - Odd-numbered VLANs are mapped to MSTI 2.
- Only two spanning trees are maintained on the network.

MSTP Overview

- MSTP divides a switching network into multiple regions, each of which has multiple spanning trees that are independent of each other.
- Each spanning tree is called a **multiple spanning tree instance** (**MSTI**).
- An MSTI is the spanning tree corresponding to a set of VLANs.
- Binding multiple VLANs to a single MSTI reduces communication costs and resource usage.
- The **topology of each MSTI is calculated independently**, and traffic can be balanced among MSTIs.
- Multiple VLANs with the same topology can be mapped to a single MSTI. The forwarding state of the VLANs for an interface is determined by the interface state in the MSTI.

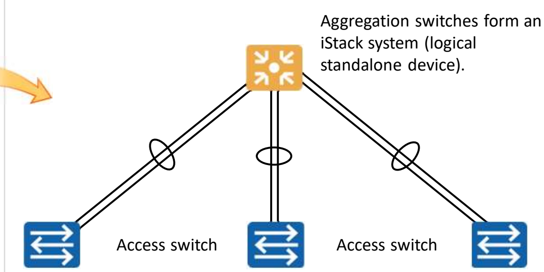
Stack and Tree Networking of Campus Networks

Traditional STP Networking



Two aggregation switches form a triangle Layer 2 loop with access switches, so STP must be deployed on the network. However, STP blocks ports on the network, causing a failure to fully utilize link bandwidth.

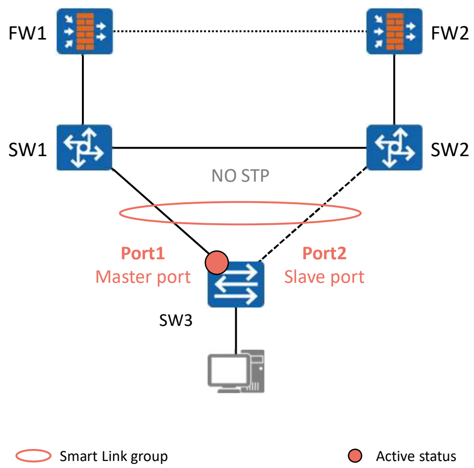
iStack Networking



Aggregation switches are stacked to form a single logical device, simplifying the network topology. In addition, link aggregation is deployed between aggregation switches and access switches to simplify the network topology to a tree topology, eliminating Layer 2 loops and improving link bandwidth utilization.

- Intelligent Stack (iStack) enables multiple iStack-capable switches to function as a logical device.
- Before an iStack system is set up, each switch is an independent entity and has its own IP address and MAC address. You need to manage the switches separately. After an iStack system is set up, switches in the iStack system form a logical entity and can be managed and maintained using a single IP address. iStack technology improves forwarding performance and network reliability, and simplifies network management.

Smart Link



Smart Link is tailored for dual-uplink networking.

- Smart Link is deployed on two switches where a host is dual-homed. When the network is normal, one of the two uplinks is active, and the other is in standby state (does not carry service traffic). In this way, a Layer 2 loop is eliminated.
- When the active link is faulty, traffic is switched to the standby link in milliseconds. This ensures proper data forwarding.
- Smart Link is easy to configure.
- Smart Link does not involve protocol packet exchange, therefore greatly improving speed and reliability.

- As shown in the figure, SW3 is connected to FW1 and FW2 through dual uplinks. In this way, Switch3 has two uplinks to the uplink device. Smart Link can be configured on SW3. In normal situations, the link on Port2 functions as a backup link. If the link on Port1 fails, Smart Link automatically switches data traffic to the link on Port2 to ensure service continuity.