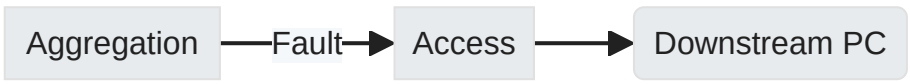
STP Principles and Configuration

1 STP Principles and Configuration

1.1 STP Overview

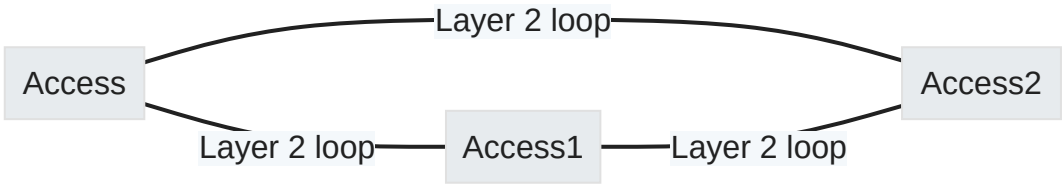
1.1.1 Network Without Redundancy

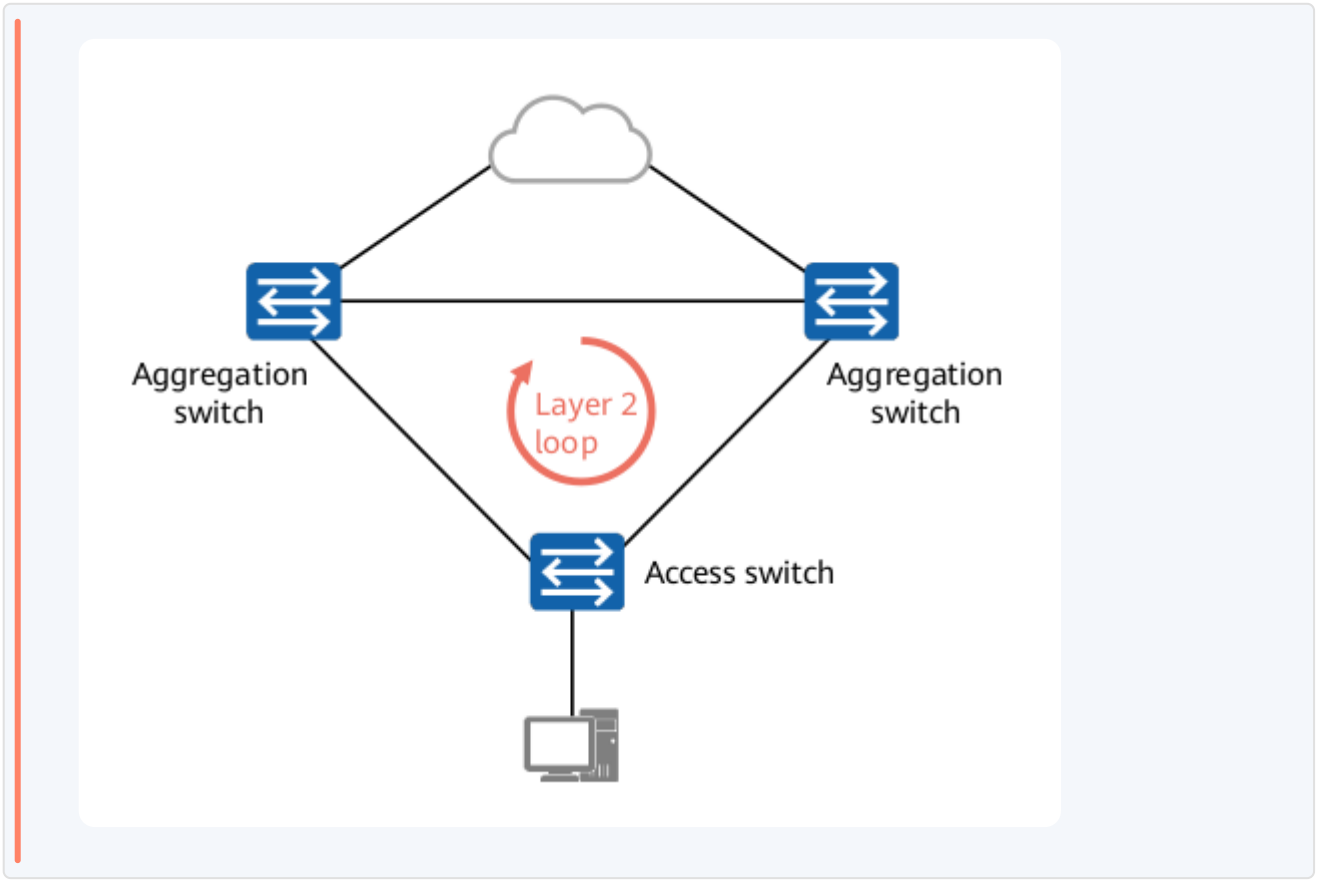
- If a fault occurs, downstream hosts are disconnected.
- Single Point of Failure (SPOF) issue.



1.1.2 Redundancy and Loops

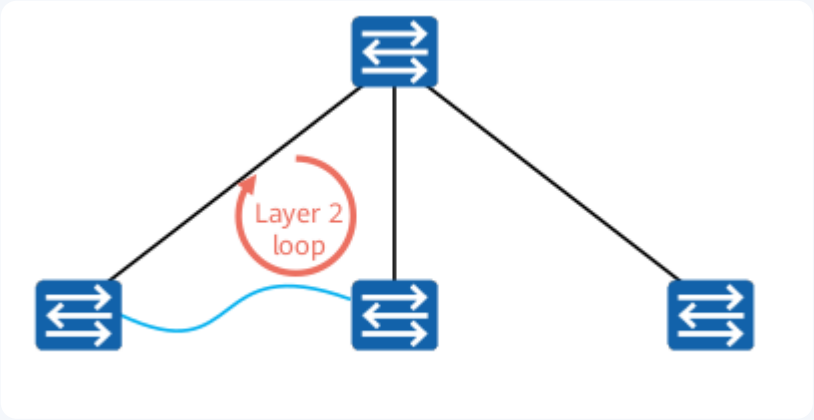
- Redundant links improve network reliability.
- However, they introduce Layer 2 loops.



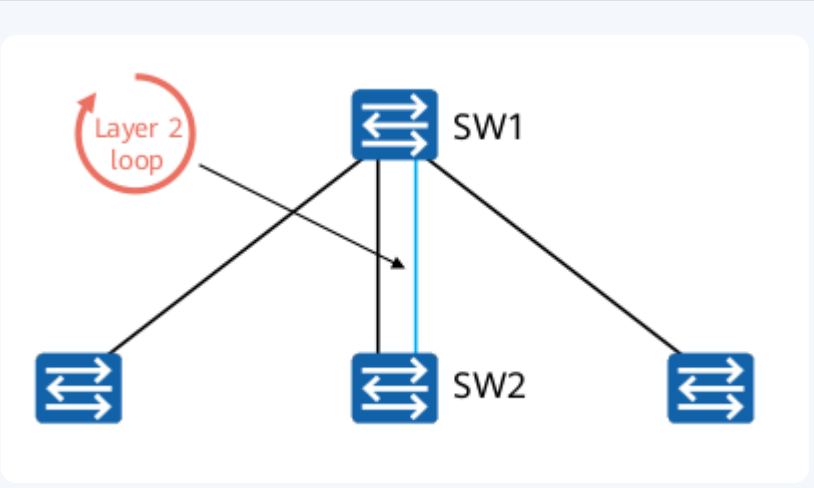


1.1.3 Human Errors Causing Loops

1.1.3.1 Case 1: Incorrect Cable Connections



1.1.3.2 Case 2: Incorrect Manual Configurations

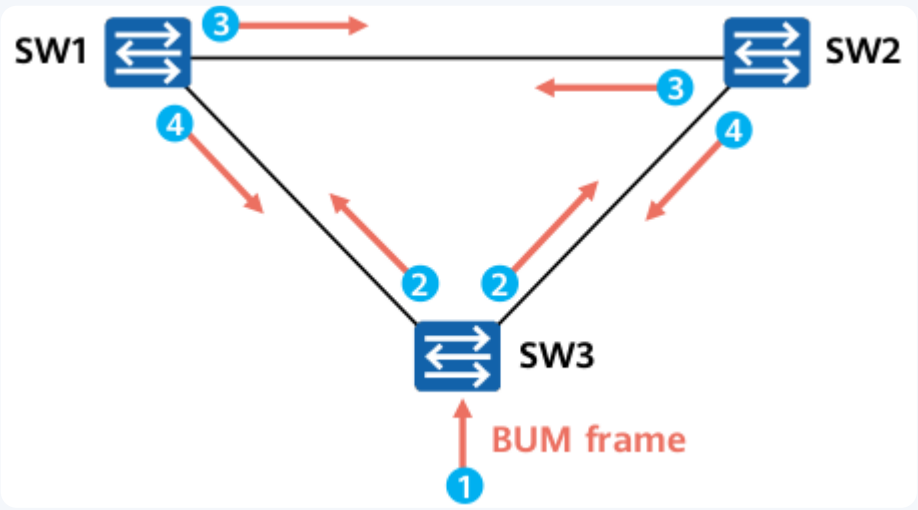


binding the link between SW1 and SW2 to a logical link (aggregation link)

1.1.4 Issues Caused by Layer 2 Loops

1.1.4.1 Issue 1: Broadcast Storm

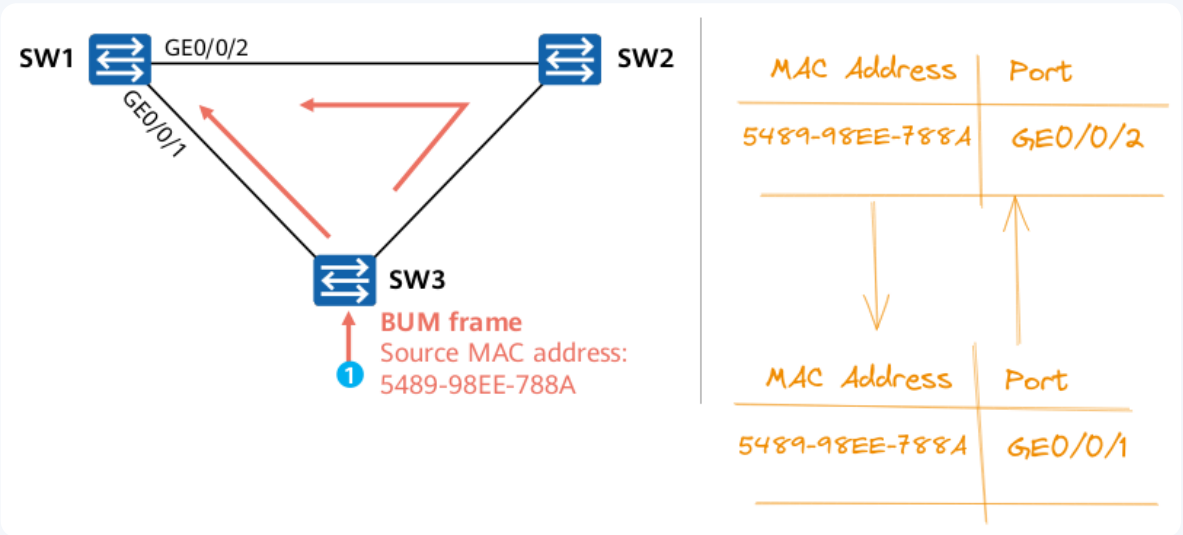
- BUM frames (broadcast, unknown unicast, multicast) are flooded repeatedly.
- Causes network resource exhaustion and service interruption.



When SW3 receives the BUM frames, it floods the frames. After SW1 and SW2 receive the BUM frames, they flood the frames again.

1.1.4.2 Issue 2: MAC Address Flapping

- Frequent switching of MAC address between ports.
- Results in unstable MAC address table entries.

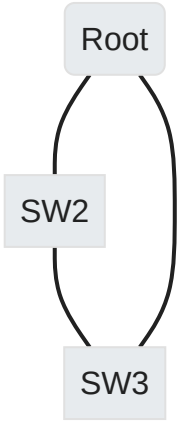


SW1 is used as an example. The MAC address of 5489-98EE-788A is frequently switched between GE0/0/1 and GE0/0/2, causing MAC

address flapping

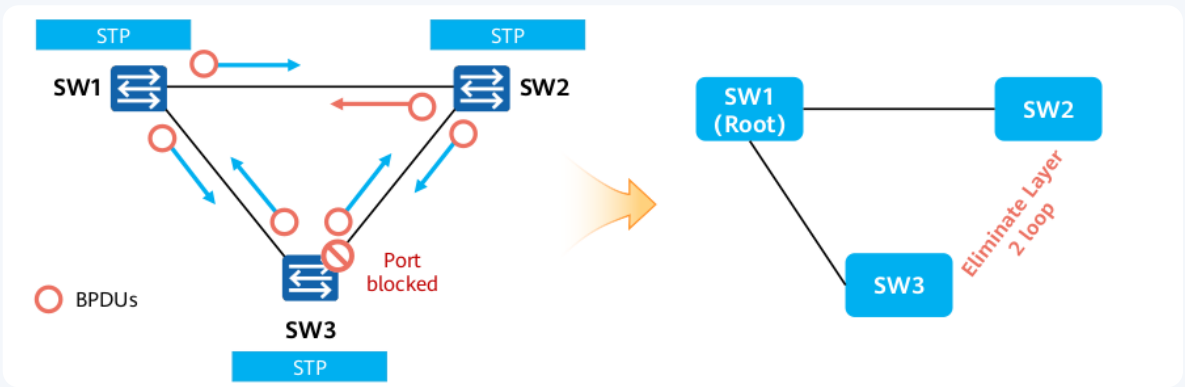
1.1.5 Spanning Tree Protocol (STP)

1.1.5.1 STP Introduction



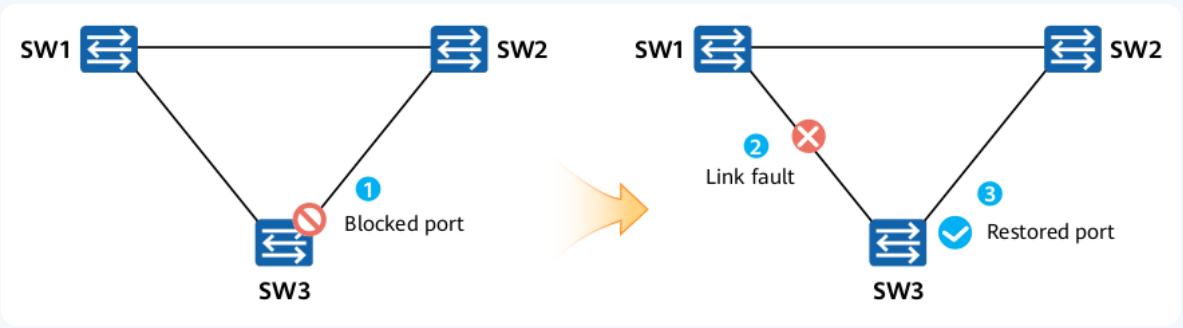
STP prevents loops by:

- Blocking one or more ports on the network.
- Calculating a loop-free tree topology.



1.1.5.2 Dynamic Response to Topology Changes

STP monitors topology changes and adjusts accordingly:



1.1.6 Comparison of Layer 2 vs. Layer 3 Loops

Factor	Layer 2 Loop	Layer 3 Loop
Root Cause	Redundancy or wrong cabling.	Routing loop.
Prevention	Protocols like STP.	TTL field in IP headers.
Impact	Data frame forwarding issues.	Infinite packet forwarding.

1.2 Basic Concepts and Working Mechanism of STP

1.2.1 Bridge ID (BID)

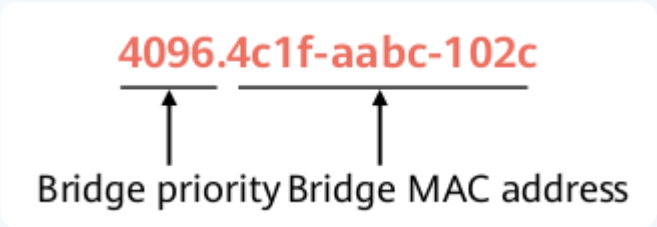
- BID is a unique identifier for each switch running STP.
- It consists of a `16-bit bridge priority` and a `48-bit MAC address`.
- The device with the smallest BID becomes the `root bridge`.

> As defined in IEEE 802.1D,

> The default bridge priority is 32768.

> The bridge priority can be changed but must be a multiple of 4096.

> If the priorities are the same, devices compare MAC addresses. A smaller MAC address indicates a `Root Bridge`.

> 
>

1.2.2 Root Bridge

- The root bridge serves as the root of an STP network tree.
- Elected based on lowest BID.

> Once the network is set up,root bridge regularly sends out configuration `BPDUs` to other devices at specific intervals so they can pass on info about any changes, keeping the network running smoothly.

1.2.3 STP Cost

- Each port has a cost that contributes to the Root Path Cost (RPC).
- Higher port bandwidth → Lower cost.

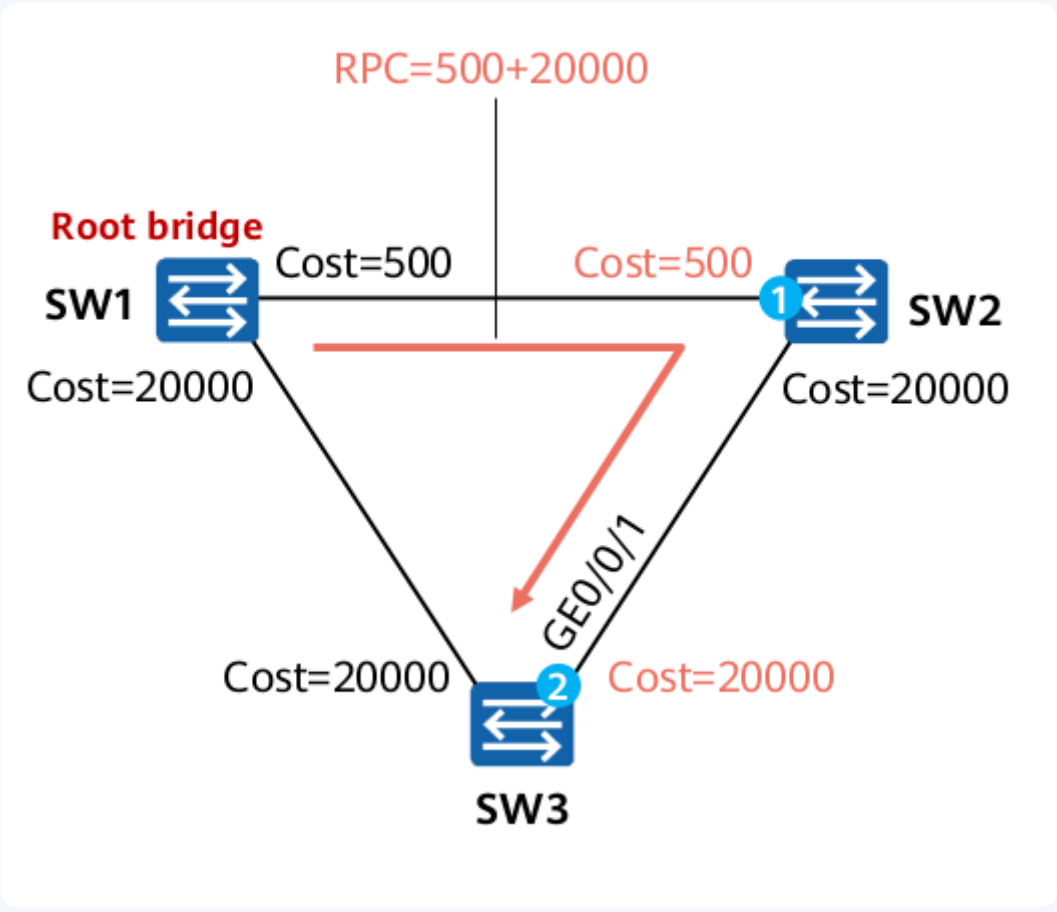
> Huawei switches use IEEE 802.1t to calculate the path cost.

1.2.3.1 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
...				

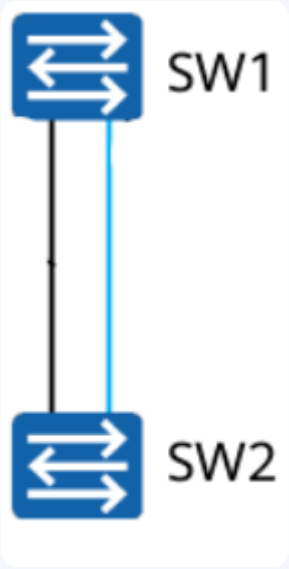
1.2.3.2 Root Path Cost (RPC)

```
RPC = Sum of port costs along the path to root
```



1.2.4 Port ID (PID)

`PID = Port Priority + Port Number (12 bits)`

> PIDs uses to identify ports and to elect a designated port in a specific scenario.
>
> 
>

> Default port priority is `128` . Ranges from 0 to 240 Must be a multiple of `16` .

1.2.5 Bridge Protocol Data Units (BPDUs)

> STP selects the root bridge by transmitting configuration BPDUs between switches and determines the role and status of each switch port.

> Once the network is stable, only the Root Bridge sends out configuration messages by itself, while other bridges send these messages only after getting them from devices upstream devices.

1.2.5.1 Format of Configuration BPDUs

Byte Field	Parameter	Description
2	PID	Protocol Identifier, always set to 0 for STP (Spanning Tree Protocol).
1	BPDU Type	Identifies the type of BPDU: 0x00 for configuration, 0x80 for TCN (Topology Change Notification).
1	Flags	Flags indicating status such as Topology Change or Acknowledgment. Only the first two and last two bits are used in STP.
8	Root ID	Identifier of the root bridge in the network.
4	RPC	Root Path Cost, represents the total cost of the path from the current bridge to the root bridge.
8	Bridge ID	Identifier of the bridge sending this BPDU.
2	Port ID	Identifier of the port sending this BPDU, includes priority and number.
2	Message Age	Time since BPDU left root bridge, increases by one with each hop and indicates distance to root. Default is every second.
2	Max Age	Maximum age before considering a link faulty if no BPDU received; default is typically set to 20 seconds.
2	Hello Time	Frequency at which root bridge sends out BPDUs; default is typically set to every 2 seconds.
2	Forward Delay	Time spent in Listening and Learning states; default is typically set to about 15 seconds.

1.2.5.2 Types of BPDUs

BPDU Type	Description
Configuration BPDU	Key to topology calculation
Topology Change Notification (TCN)	Notifies switches about topology changes

1.2.5.3 BPDU Comparison Rules

Use this order when comparing fields within a configuration BPDU:

1. **Smallest Root Identifier (`BID` of root):** to elect root bridge
2. **Smallest Root Path Cost (`RPC`):** to change port Role to Root port
3. **Smallest Bridge Identifier (`BID` of sender):** If there's a tie in path costs, STP looks at who sent the BPDU. The switch with the smallest BID wins and its path is used.
4. **Smallest Port Identifier (`PID`):** If there's still a tie after looking at BIDs, STP looks at port IDs. Each port on a switch has its own identifier, and once again, smaller is better - so the port with the smallest ID is chosen for forwarding traffic.

1.2.6 Port Roles in STP

Role	Description
Designated Port	Forwards data frames; one per network segment
Root Port	Optimal path to root; one per non-root switch
Alternate Port	Blocked to prevent loops

1.2.7 STP Calculation

1.2.7.1 Root Bridge Election

1.2.7.1.1 Steps:

1. Each switch considers itself as the root bridge initially.

2. Switches compare received BPDUs and elect the switch with the smallest BID as the root bridge.

> Plan STP network in advance by setting the bridge priority to 0 for the planned root switch to ensure Continuity.

1.2.7.2 Port Roles

1.2.7.2.1 Root Port Selection

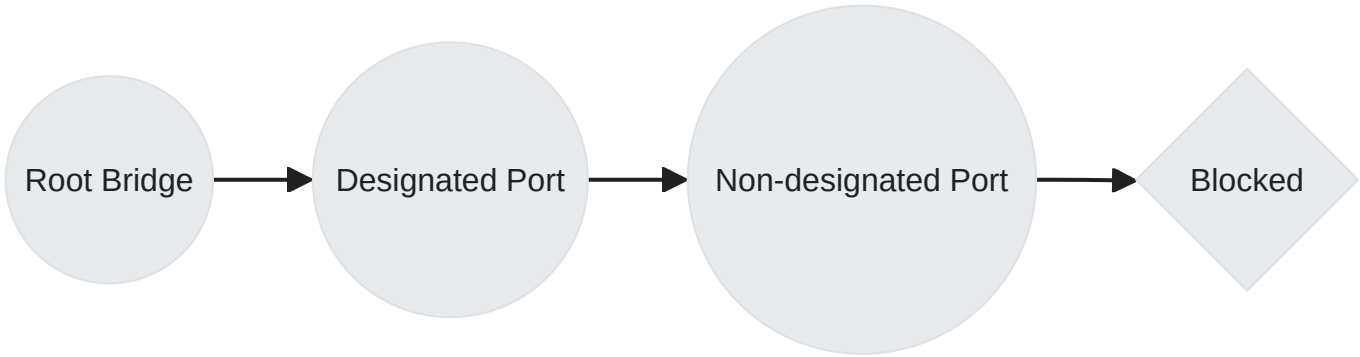
- Each non-root bridge selects one root port with **optimal path** to the root bridge based on:
  - Smallest Root Path Cost (RPC)
  - Smallest Bridge ID (BID)
  - Smallest Port ID (PID)

1.2.7.2.2 Designated Port Selection

- A designated port is elected on each link to ensure unique and optimal paths.
- Uses same criteria as root port selection: RPC $->$ BID $->$ PID (Smallest)

1.2.7.2.3 Non-designated Port Blocking

- Any port that is neither a root nor designated port becomes a non-designated or alternate port.
- Non-designated ports are blocked by STP to prevent loops.

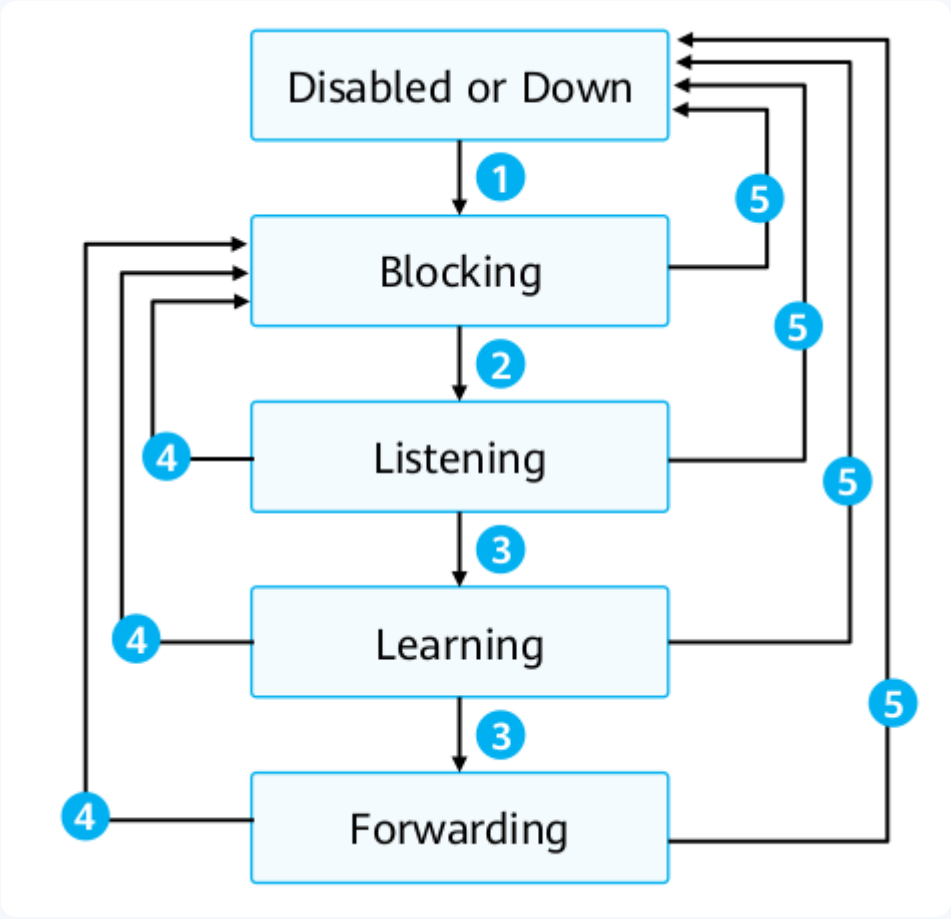


1.2.8 STP Port States

State	Description
Disabled	No BPDU/service data frame processing.
Blocking	Listens to BPDUs without forwarding data frames or learning MAC addresses.
Listening	Considers port as candidate for root/designated but still in calculation.The port can send and receive BPDUs but cannot send or receive service data frames or learn MAC addresses.
Learning	Creates MAC entries without forwarding traffic.
Forwarding	Sends/receives service data frames and processes BPDUs; final state for active ports.

> **Service data frames:** are the actual network traffic carrying user data, such as emails or website content

1.2.8.1 State Transitions



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.

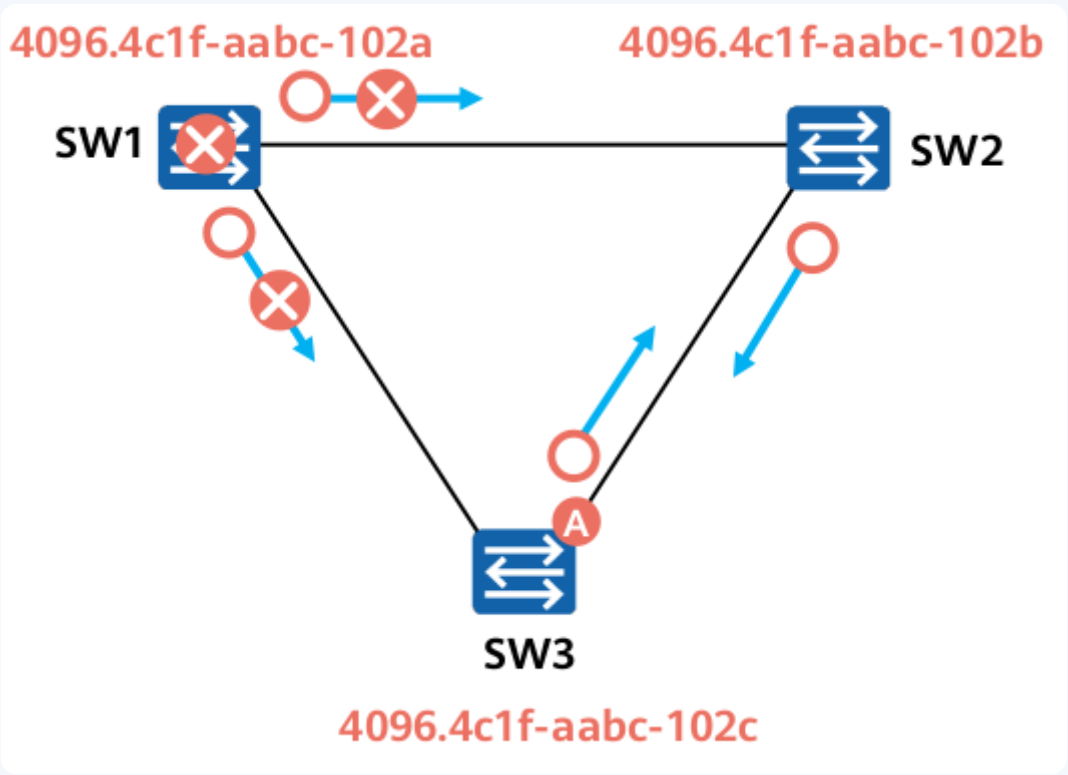
> Upon Forward Delay timer expiration, if a port remains the root port or designated port, it means no better path was discovered and STP received no information prompting a change in the port's status.

4. The port is no longer the root port or designated port.
5. The port is disabled or the link fails.

> **STP transitions** require 15 seconds from listening to learning state and another 15 seconds from learning to forwarding state.

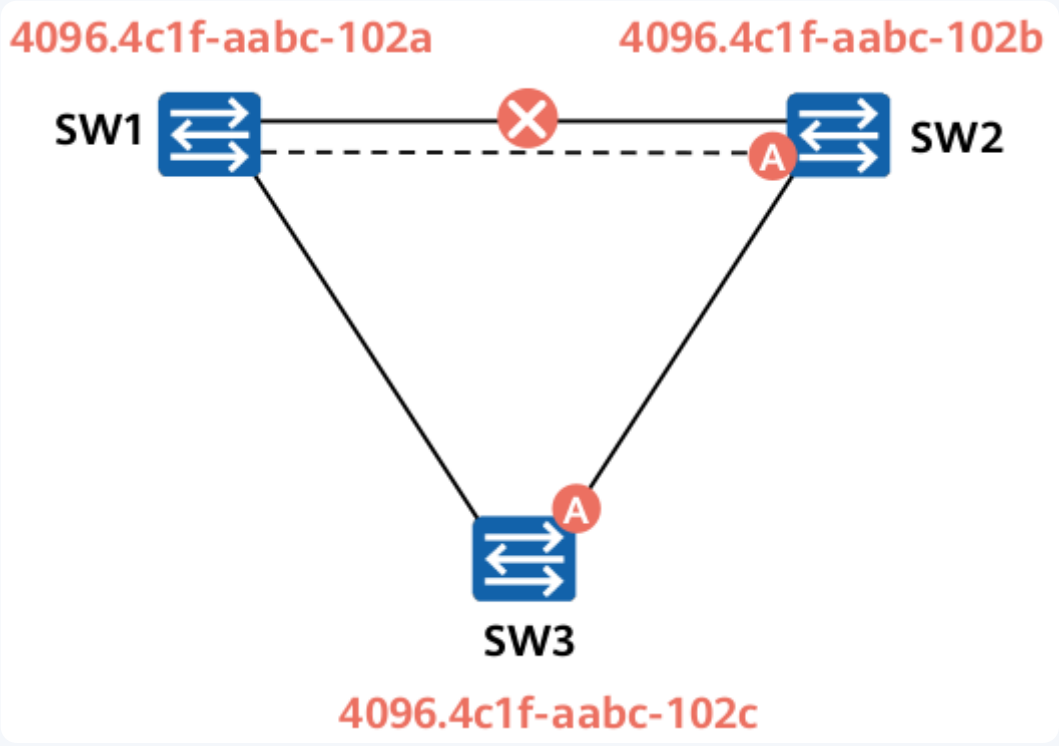
1.2.9 Topology Changes Handling

1.2.9.1 Root Bridge Fault



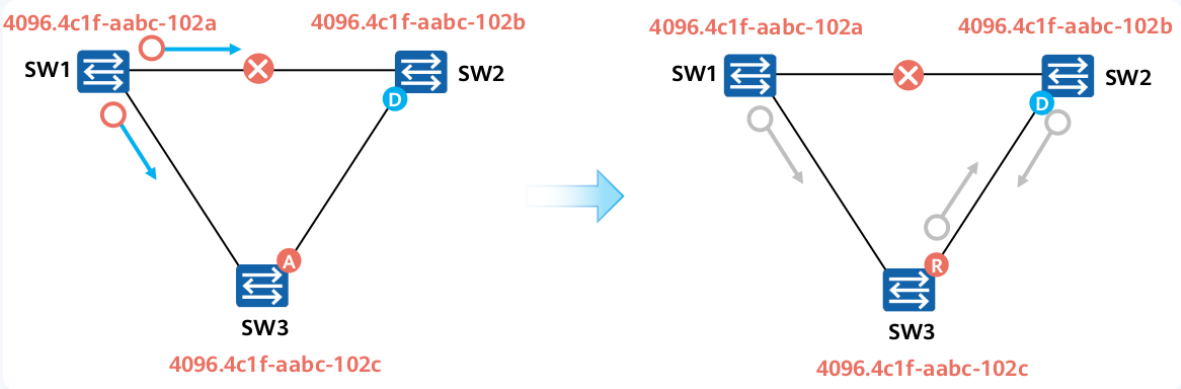
1. The root bridge SW1 fails and stops communicating with other switches.
2. SW2 detects the failure after waiting 20 seconds, prompting a re-election among the remaining switches to choose a new root bridge.
3. SW3 becomes the new root bridge, and its port A starts forwarding traffic after 30 seconds (two Forward Delay intervals from listening to forwarding 30 sec).
4. Network recovery from root bridge failure takes approximately 50 seconds total.

1.2.9.2 Direct Link Failure Recovery



1. When SW2's root port link fails, the alternate port becomes the new root port and starts forwarding traffic after 30 seconds(two Forward Delay intervals from listening to forwarding 30 sec)..
2. The alternate port moves from blocking to listening, then learning, and finally to forwarding state during this transition.

1.2.9.3 Indirect Link Failure Recovery

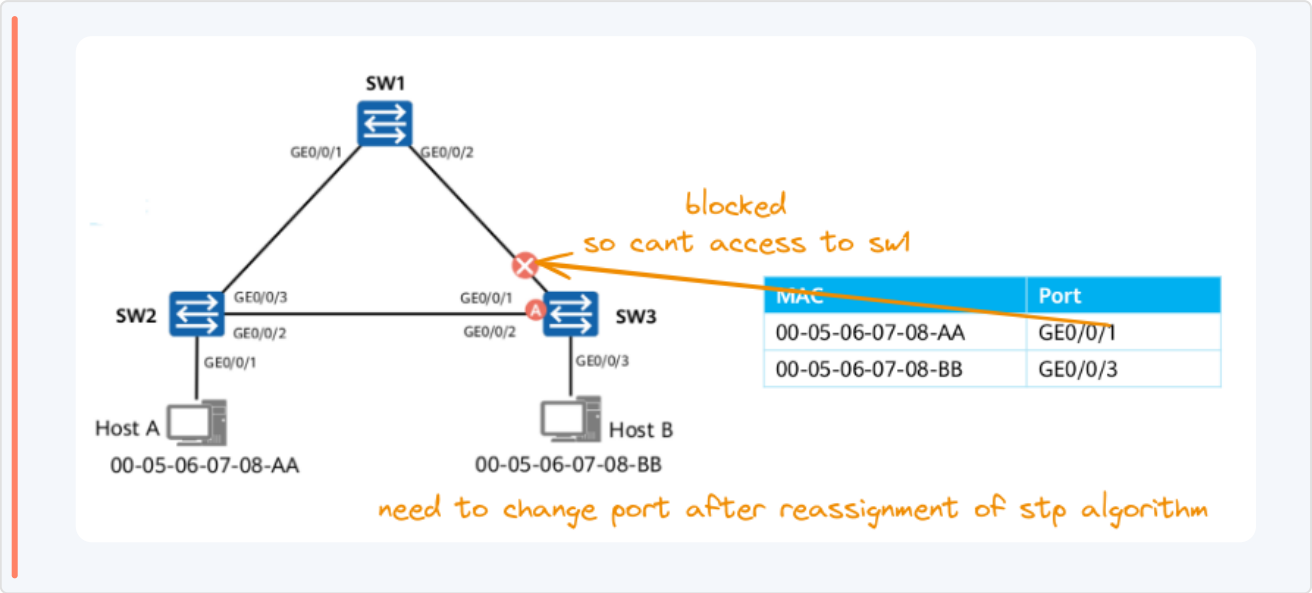


1. An indirect link failure between SW1 and SW2 prevents SW2 from receiving STP BPDUs from SW1.

2. After 20 seconds, the Max Age timer on SW2 expires, indicating outdated topology information.
3. SW2 may assume it's the root bridge and starts issuing its own BPDUs.
4. When SW3 receives these BPDUs from both switches, it compares them and realizes the BPDU from SW1 is better, so it doesn't switch to using SW2 as the root bridge.
5. The alternate port on SW3 transitions from Blocking to Listening for 15 seconds, then Learning for another 15 seconds, before moving to Forwarding state.
6. It takes about 50s to recover from an Unphysical link failure, which is equal to the value of the Max Age timer plus twice the value of the Forward Delay timer.

1.2.9.4 Topology Change Notification (TCN)

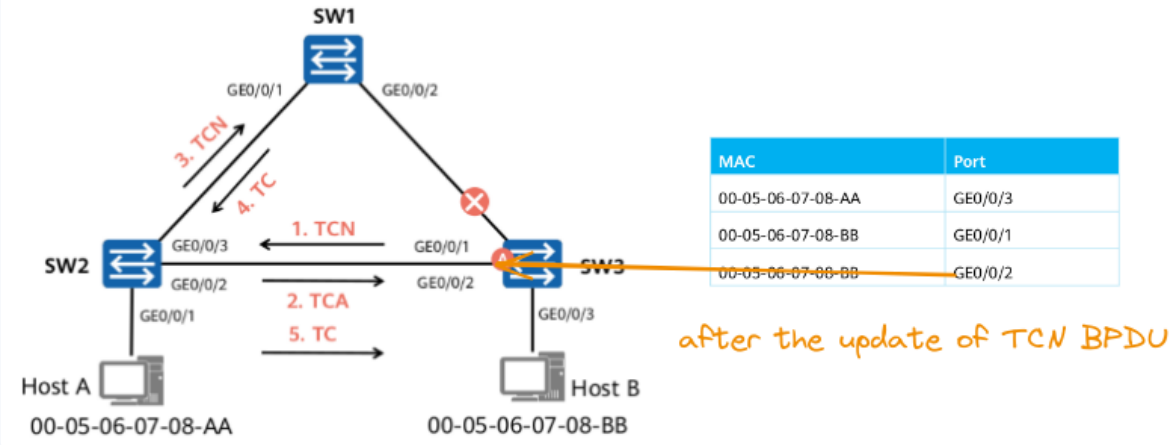
Problem:



- SW3's root port is broken, causing the network's path layout to change.
- Host B can't get frames from Host A because of outdated MAC address information in the switches.
- Switches use a MAC address table to decide where to send data, with entries aging out after 300 seconds by default.

- After a network change, switches must quickly update their MAC address tables to prevent sending data the wrong way.

When topology changes occur:



1. Switches send out TCN BPDUs.
2. Root bridge sets TC flag in BPDU To update Mac Table address, Quickly others to shorten MAC entry aging time from default (300s) to Forward Delay value (~15s).

1.3 Basic STP Configurations

1.3.1 Configuration Commands

1.3.1.1 Set Working Mode

```
[Huawei] stp mode  <stp> || <rstp> || <mstp>
```

- `stp` : Classic Spanning Tree Protocol.
- `rstp` : Rapid Spanning Tree Protocol.

- `mstp` : Multiple Spanning Tree Protocol.

> Default mode is MSTP.

1.3.1.2 Root Bridge Configuration

```
[Huawei] stp root primary   // Sets switch as root bridge with priority 0.
[Huawei] stp root secondary // Sets switch as secondary root with priority 4096.
```

1.3.1.3 Set STP Priority

```
[Huawei] stp priority <priority>
```

> Default priority is 32768.

1.3.1.4 Configure Path Cost Method

```
[Huawei] stp pathcost-standard <dot1d-1998> || <dot1t> || <legacy>
```

> All switches must use the same path cost calculation method.

> Default path cost is dot1t on Huawei.

1.3.1.5 Set Port Path Cost

```
[Huawei-GigabitEthernet0/0/1] stp cost <cost>
```

1.3.1.6 Set Port Priority

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

> Default priority is 128.

1.3.1.7 Enable STP/RSTP/MSTP Globally

```
[Huawei] stp enable
```

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

1.3.1.8 Checking Port States

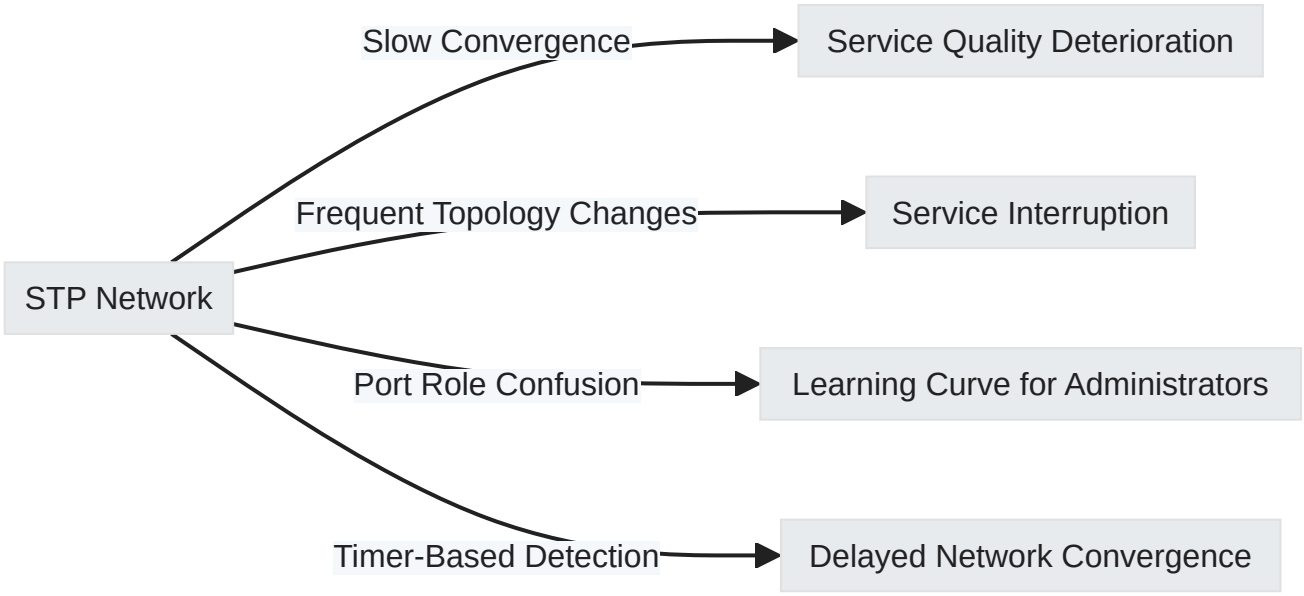
```
<Huawei> display stp brief
```

MSTID	Port	Role	State
0	GigabitEthernet0/0/21	ROOT	FORWARDING
0	GigabitEthernet0/0/22	ALTE	DISCARDING

1.4 Improvements Made in RSTP

1.4.1 Disadvantages of STP

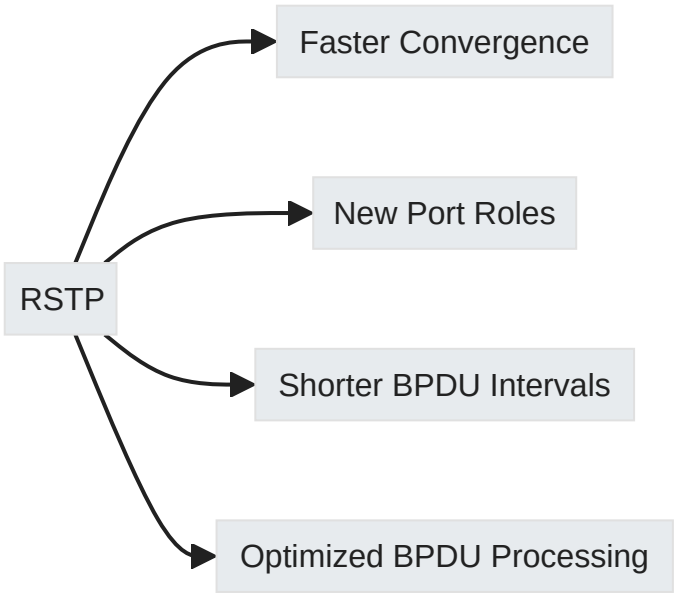
- Frequent network changes can cause slow protocol convergence, resulting in potential Interference or degraded performance of network services.
- The complexity and Unexpected nature of port roles and states in some protocols can be challenging for inexperienced network administrators to understand and manage.
- Protocols that rely on timers to detect changes in the network topology may not recognize these changes promptly, leading to a delayed response.
- Protocols that send Bridge Protocol Data Units (BPDUs) across the whole network Hurt from slow convergence times, as the propagation of these units takes time and can slow down the overall response to changes.



1.4.2 Advantages of RSTP over STP

- Faster convergence compared to STP.
- Defines new port roles for better understanding and deployment simplicity.
- Shorter timeout intervals for BPDUs.
- Optimized BPDU processing methods.

> RSTP defined in IEEE 802.1w.



1.4.3 Port Roles in RSTP

1.4.3.1 Traditional Port Roles

- Root Port (R)
- Designated Port (D)

1.4.3.2 New Port Roles in RSTP

- **Alternate Port (A):** Provides an alternative path to the root bridge when the root port fails.
- **Backup Port (B):** Serves as a backup path from the root bridge.

1.4.4 Edge Ports in RSTP

1.4.4.1 Characteristics of Edge Ports:

> An edge port connects directly to user terminals, not switches. It transitions immediately to Forwarding state upon initialization so no delay occur.

> If an edge port receives a configuration BPDU, it reverts to a common STP port and triggers recalculation of the spanning tree, potentially causing network flapping.

1.4.5 Port States in RSTP vs. STP

1.4.5.1 Reduced Number of States in RSTP:

1. **Discarding:** Neither forwards traffic nor learns MAC addresses.
2. **Learning:** Learns MAC addresses but does not forward traffic.
3. **Forwarding:** Forwards user traffic and learns MAC addresses.

1.4.5.2 Corresponding States Between Protocols:

STP State	RSTP State	Port Role
Disabled	Discarding	Disabled
Blocking	Discarding	Alternate or Backup
Listening	Discarding	Root or Designated
Learning	Learning	Root or Designated
Forwarding	Forwarding	Root or Designated

> Discarding: This combines Disabled, Blocking, and Listening from STP

> into one state where the port isn't forwarding frames.

1.5 STP Advancement

1.5.1 Overview

- STP and RSTP ensure a loop-free topology for Ethernet networks.
- Both protocols have limitations when dealing with multiple VLANs.
- All VLANs share one spanning tree, leading to inefficient link utilization.

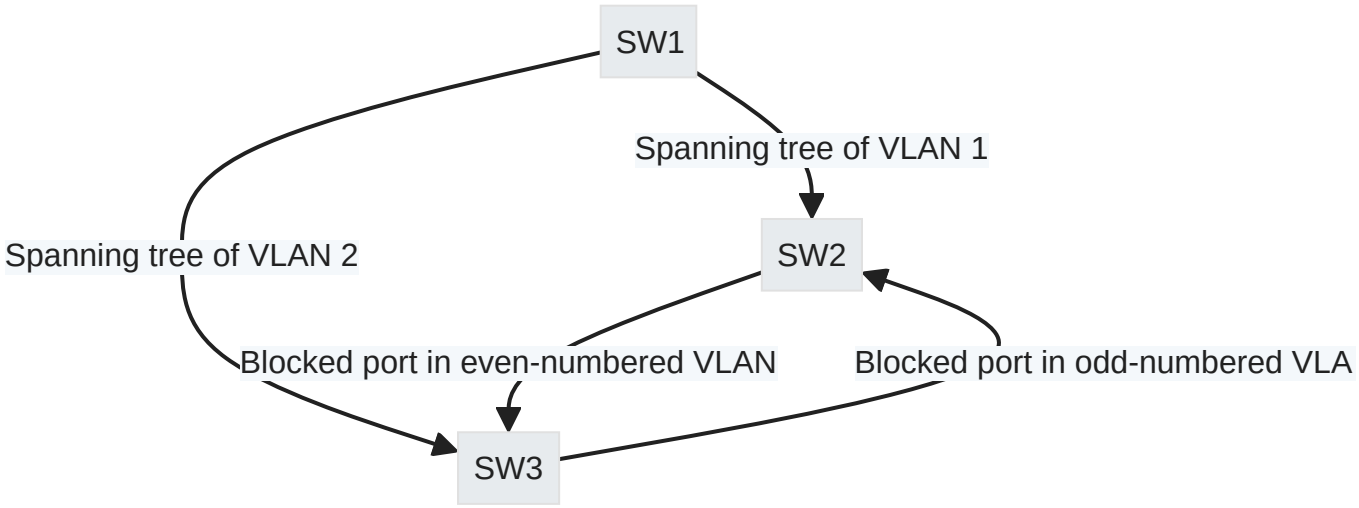
1.5.2 Defects of STP/RSTP

> **Note**
>
> A single spanning tree across all VLANs can result in Insufficient load balancing and underutilization of network links.

1.5.2.1 Issues

- **Inter-VLAN Load Balancing:** Cannot be performed effectively.
- **Blocked Links:** Do not transmit any traffic, causing potential packet transmission failures within VLANs.

1.5.3 Huawei's VBST (VLAN-Based Spanning Tree)



1.5.3.1 Benefits

---

1. Eliminates loops per VLAN.
2. Implements link multiplexing for improved efficiency.
3. Reduces configuration and maintenance costs.

1.5.4 MSTP (Multiple Spanning Tree Protocol)

- Defined by IEEE 802.1s standard.
- Compatible with STP/RSTP; provides rapid convergence and load balancing.
- MSTP divides a switching network into multiple regions, each of which has multiple spanning trees that are independent of each other.

> **region:** is a subsection of a network that operates its own MSTP, functioning as a singular switch externally.

> **spanning tree:** is a loop-free network path architecture ensuring a single active route between devices to prevent network issues.

1.5.4.1 Key Concepts

- **MSTIs:** Multiple spanning trees that are independent within different regions.
- **VLAN Mapping:** Multiple VLANs can share a single spanning tree within an MST instance.

1.5.5 Network Optimization Techniques

1.5.5.1 iStack Networking

`iStack` technology enables stacking switches to function as a single logical device, simplifying network management and enhancing performance.

> **Example**
>
> Merging individual light switches into a single master switch for ease of control.

1.5.5.2 Smart Link

`Smart Link` is designed for dual-uplink networks, providing fast failover without the need for STP by having one active and one standby link.

> Smart Link does not involve **protocol packet exchange**