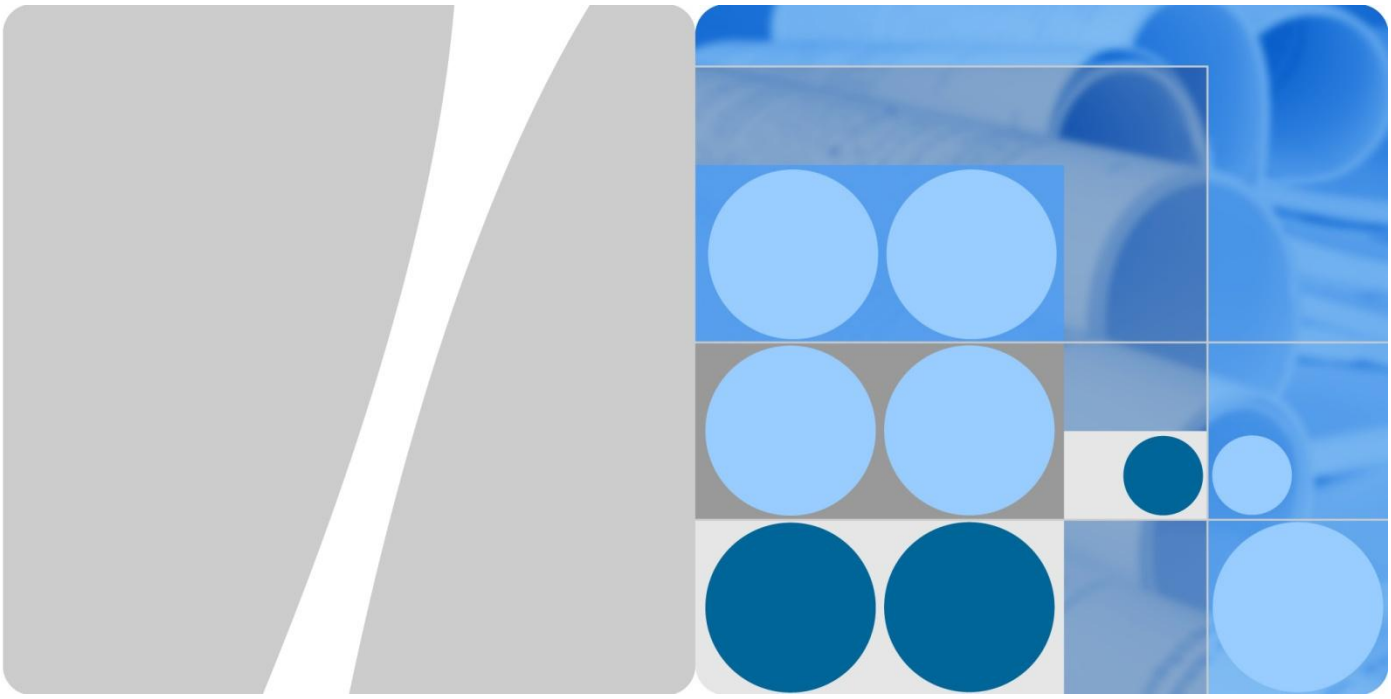


MSTP Technology White Paper

Issue 1.00

Date 2012-10-30

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions

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

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

Notice

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

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

Huawei Technologies Co., Ltd.

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

Website: http://enterprise.huawei.com

Contents

1 Overview

1.1 Review of STP

On a Layer 2 switching network, routers know how many hops a packet can pass, but switches do not. If a loop exists on the network, the packet is endlessly cycled in the loop and more response packets are generated. This results in a broadcast storm. When a broadcast storm occurs, all bandwidth is occupied. As a result, the network becomes unavailable.

The Spanning Tree Protocol (STP) aims to tackle this problem. STP, a Layer 2 management protocol, selectively blocks redundant links to remove any Layer 2 loop on a network. In addition, STP supports the link backup function.

Like other protocols, STP evolves with network development. Through the ongoing development of STP, defects are overcome and new features are developed. The IEEE 802.1D STP was the first widely used version. On the basis of IEEE 802.1D STP, IEEE 802.1w RSTP, PVST+, and IEEE 802.1s MSTP are subsequently generated. These derivative protocols are described later in this paper.

STP refer to the protocol defined in IEEE802.1D in the narrow sense, or IEEE802.1D STP and other STPs based on IEEE802.1D STP in the broad sense.

1.1.1 IEEE 802.1D STP

The basic idea of STP is simple. No loop exists in a tree that grows in nature. If a network grows like a tree, no loop can be generated on the network. STP defines the concepts of root bridge, root port, designated port, and path cost. These concepts are combined to cut redundant loops and implement link backup and path optimization on a tree network. The spanning tree algorithm is used to construct the tree.

To implement spanning tree functions, bridges need to exchange information. The exchanged information units are called bridge protocol data units (BPDUs). STP BPDUs are Layer 2 packets, in which the destination MAC address is multicast address 01-80-C2-00-00-00. STP-supporting bridges receive and handle STP BPDUs. The data field in an STP BPDU contains all the information used for STP calculation.

The STP working process is as follows:

The network root bridge is selected based on bridge IDs. A bridge ID consists of the priority level and MAC address of a bridge. The bridge with the smallest bridge ID is the root bridge on a network. All ports on the root bridge are connected to downstream bridges. Therefore, all ports on the root bridge are designated ports. Each downstream bridge selects the strongest branch as a path to the root bridge, and the corresponding ports are root ports. This process is repeated until the boundary of the network is reached.

After the designated ports and root ports are determined, a tree is established. When the spanning tree has stabilized after a period of time (30 seconds by default), the designated ports and root ports are in forwarding state and other ports are in blocked state. An STP BPDU is periodically sent from the designated ports of each bridge to maintain the link state. If the network topology is changed, the spanning tree is re-calculated and the port state is changed accordingly. This is the fundamental principle of the spanning tree. All other improved spanning tree protocols are implemented based on STP, and their basic ideas and concepts are similar.

With the wide application of STP and the development of network technologies, the defects of STP are unveiled. The major defect of STP is slow topology convergence.

If the network topology is changed, a new BPDU can be propagated on the whole network only after a delay. This delay is called the forwarding delay, and the default delay time in STP is 15 seconds. During the delay, if a port in forwarding state in the original topology does not find that it must stop forwarding in the new topology, a temporary loop may occur. To prevent temporary loops, STP uses a timer. That is, an intermediate state is inserted between the blocked state and the forwarding state of a port. In intermediate state, the port only learns the MAC address, but does not participate in packet forwarding. The time for both of the two stateful switchovers equals the forwarding delay. In this way, no temporary loop is generated when the network topology changes. This solution seems adequate but brings about at least twice the forwarding delay of convergence. Such a delay is unacceptable for certain real-time services, such as voice and video services.

1.1.2 IEEE 802.1w RSTP

To overcome the slow convergence of STP, the IEEE drafted the 802.1w standard in 2001 as a supplement to 802.1D. The IEEE 802.1w standard defines the Rapid Spanning Tree Protocol (RSTP). The following improvements were made in RSTP on the basis of STP so that the topology converges more rapidly (within 1 second), without waiting for the forwarding delay of STP:

1. Alternate port and backup port roles are set for fast transition of a root port and a designated port respectively. When a root port fails, the alternate port quickly replaces the root port and enters the forwarding state without delay. When a designated port fails, the backup port also quickly replaces the designated port and enters the forwarding state.
2. On a point-to-point link, a specified interface handshakes with the downstream bridge only once and then enters the forwarding state without delay. If the shared link is connected to more than three bridges, the downstream bridge does not respond to the handshake request from the designated port on the upstream bridge, but waits for twice the forwarding delay time to go into the forwarding state.
3. An edge port can directly enter the forwarding state without delay. An edge port is defined as a port that is connected to a terminal rather than through a bridge. A bridge cannot determine whether one of its ports is directly connected to a terminal. Therefore, the connection of an edge port to a terminal needs to be manually configured.

RSTP is backward compatible with STP and can be used for hybrid networking. Nevertheless, RSTP and STP are both single spanning trees (SSTs). An SST has its own defects, related mainly to the following three aspects:

1. A switching network has only one spanning tree. When the network is large, the convergence time is long and the probability of topology change is high.
2. The concept of VLAN is introduced in the IEEE 802.1Q standard. Since RSTP is an SST, all VLANs on a network share one spanning tree. To ensure normal communication in each VLAN on the network, the VLANs on the network must be distributed along the path of the spanning tree. Otherwise, the subscribers in a VLAN may be isolated and cannot communicate with each other because the internal links are blocked.
3. After being blocked, a link does not carry any traffic. Therefore, load balancing cannot be implemented, which causes a huge waste of bandwidth.

An SST cannot overcome these defects. With this background, the Multiple Spanning Tree Protocol (MSTP) supporting VLANs is put forward.

1.1.3 IEEE 802.1s MSTP

MSTP is a new spanning tree protocol defined in the IEEE 802.1s standard. MSTP introduces the concept of instance. In simple terms, STP and RSTP are based on ports, Per VLAN Spanning Tree (PVST)+ is based on VLANs, and MSTP is based on instances. An instance is a collection of VLANs. Binding multiple VLANs to an instance can save communication overhead and resource occupancy. The topology of each instance in MSTP is calculated independently. Therefore, load balancing can be implemented among these instances. Multiple VLANs with the same topological structure are mapped into an instance. The forwarding state of the ports in these VLANs depends on the state of the corresponding instance in MSTP. Instance 0 in MSTP has special functions and is called the common and internal spanning tree (CIST) instance. Other instances are called multiple spanning tree instances (MSTIs).

MSTP also introduces the concept of region. A region consists of the region name, revision level, and mappings between VLANs and instances. Only when these three items are the same, are the interconnected switches considered to be in a region. By default, the region name is the first MAC address of a switch, the revision level is 0, and all VLANs are mapped to instance 0. The switches in a region propagate and receive BPDUs of different MSTIs to ensure that all MSTIs are calculated in the whole region. The switches in different regions propagate and receive BPDUs only of the CIST instance. MSTP utilizes the CIST to ensure a loop-free network topology. MSTP also utilizes the CIST to maintain backward compatibility with STP and RSTP. Therefore, an MSTP region is externally equivalent to a switch, which is transparent to different regions, STP switches, and RSTP switches.

In summary, MSTP has significant advantages over the other STPs. MSTP has VLAN recognition capability and implements load balancing and fast port state transition similarly to RSTP. MSTP supports binding of multiple VLANs to one instance to reduce resource occupancy. In addition, MSTP is backward compatible with STP and RSTP.

1.2 STP Versions Supported by Huawei Switches

Huawei S-series switches support all STP versions in the IEEE standard, including IEEE 802.1D STP, IEEE 802.1W RSTP, and IEEE 802.1S MSTP.

2 Huawei STP Features and Related Technologies

This chapter describes technologies that Huawei uses to extend the standard STP versions, as well as application features, such as path route calculation, root bridge designation and backup, protection, and feature configuration. For the concepts and features of standard STP versions, see the protocol texts and the STP operation manuals of various products.

2.1 Software Modules Related to STP, RSTP, and MSTP

Two STP-related modules exist in the software platform of Huawei: RSTP module and MSTP module. The RSTP module implements the RSTP state machine in the IEEE 802.1w standard, and supports the STP-compatible mode and RSTP mode. The MSTP module implements the MSTP state machine in the IEEE 802.1s standard, and supports the STP-compatible mode, RSTP-compatible mode, and MSTP mode. The MSTP state machine in the IEEE 802.1s standard itself does not implement the RSTP-compatible mode. Support for the RSTP-compatible mode is an extension of IEEE 802.1s. By default, the RSTP module runs in RSTP mode, and the MSTP module runs in MSTP mode. The RSTP-compatible mode of the MSTP module is simulated on the basis of the MSTP state machine and the external behavior is the same as that of RSTP.

In RSTP-compatible mode, RSTP BPDUs are sent. STP BPDUs are sent after STP BPDUs are received. MSTP BPDUs can be processed properly and RSTP BPDUs are sent after MDTP BPDUs are received.

In RSTP-compatible or MSTP mode, if a port sends STP BPDUs, after the **mcheck** command is executed, RSTP/MSTP BPDUs are sent.

When the STP-compatible mode is switched to the RSTP-compatible or MSTP mode, the **mcheck** command is automatically executed and RSTP/MSTP BPDUs are sent.

In STP or RSTP mode, multiple instances can be configured and the state of each port of MSTIs is consistent with that of the CIST. For light CPU load, multiple instances are not recommended in STP or RSTP mode.

In practical device interconnection, STP, RSTP, and MSTP all comply with the backward compatible mode.

2.2 Calculation of Default Path Costs

Huawei STPs support three methods of calculating default path costs: method in the IEEE 802.1D standard, method in the IEEE 802.1t standard, and Huawei proprietary calculation method.

The default path costs in different cases are assigned as follows:

Port Speed	Link Type	Path Cost 802.1D-1998	Path Cost 802.1T	Path Cost Legacy
0		65,535	200,000,000	200,000
10 Mbit/s	Half-duplex	100	2,000,000	2,000
	Full-duplex	99	1,999,999	2,000
	Aggregated link 2 ports	95	1,000,000	1,800
	Aggregated link 3 ports	95	666,666	1,600
	Aggregated link 4 ports	95	500,000	1,400
100 Mbit/s	Half-duplex	19	200,000	200
	Full-duplex	18	199,999	200
	Aggregated link 2 ports	15	100,000	180
	Aggregated link 3 ports	15	66,666	160
	Aggregated link 4 ports	15	50,000	140
1000 Mbit/s	Full-duplex	4	20,000	20
	Aggregated link 2 ports	3	10,000	18
	Aggregated link 3 ports	3	6,666	16
	Aggregated link 4 ports	3	5,000	14
10 Gbit/s	Full-duplex	2	2,000	2
	Aggregated link 2 ports	1	1,000	1
	Aggregated link 3 ports	1	666	1
	Aggregated link 4 ports	1	500	1

For the basic calculation of default path costs in the IEEE 802.1D and IEEE 802.1t standards, see the protocol texts. The following describes the extension of the standard protocols.

The IEEE 802.1D and IEEE 802.1t standards do not stipulate that the path cost of a port in full-duplex mode must be different from that in half-duplex mode at the same link speed. The path cost of a port in full-duplex mode is usually slightly smaller than that in half-duplex mode, however.

For an aggregated link, the IEEE 802.1D standard does not specify that the priority level of an aggregated link must be different from that of a single port link. Therefore, the STP path cost of aggregated links in the IEEE 802.1D-1988 standard is not related to the number of aggregated links.

- The IEEE 802.1t standard recommended the following formula to calculate the default path cost:

  Path Cost = 20,000,000,000/link speed in kbit/s

  For an aggregated link, the link speed is the sum of the speeds of all unblocked ports in the aggregation group.
- Huawei proprietary algorithm for calculating the default path cost:

$$\text{Path Cost} = \begin{cases} 200{,}000 & (\text{Link Speed} = 0) \\ 2{,}200 - 20 * \text{LinkSpeed} & (0 < \text{Link Speed} \le 100) \\ 220 - 0.2 * \text{LinkSpeed} & (100 < \text{Link Speed} \le 1000) \\ 22 - 0.002 * \text{LinkSpeed} & (1000 < \text{Link Speed} \le 10{,}000) \\ 1 & (\text{Link Speed} > 10{,}000) \end{cases}, \text{Link Speed in Kbps}$$

For an aggregated link, the link speed is the sum of the speeds of all unblocked ports in the aggregation group.

The command line used to set a path cost calculation method is as follows:

[System view] [undo] stp pathcost-standard { dot1d-1998 | dot1t | legacy}

2.3 Root Bridge Designation and Root Bridge Backup

STP determines the root switch of a spanning tree through calculation. A user can also specify a switch as the root switch using a command provided by the switch.

After a switch is configured as the root switch or backup root switch, the priority of the switch cannot be modified.

A switch cannot serve as the root switch and backup root switch at the same time.

When the root switch fails or is powered off, the backup root switch replaces the original root switch to become the new root switch of the spanning tree. If a new root switch is specified, however, the backup root switch will not become the root switch. If multiple backup root switches are configured for the spanning tree, STP selects the one with the lowest MAC address to serve as the root switch when the original root switch fails.

Note the following principles when specifying a root switch:

1. A switch can be specified as the root of a spanning tree by setting the priority of the switch to 0 or by using a command.
2. No more than two root switches can be specified for a spanning tree. That is, no more than two switches can be specified as the root of a spanning tree.
3. Multiple backup root switches can be specified for a spanning tree. That is, two or more switches can be specified as the backup root of a spanning tree.
4. It is recommended that one root switch and multiple backup root switches be specified for a spanning tree.

By default, a switch does not function as the root switch or a backup root switch of a spanning tree.

- The RSTP module provides the following command for the user to configure the local switch as the root switch or backup root switch of a spanning tree:

  [System view] [undo] stp root {primary|secondary}

- The MSTP module provides the following command for the user to configure the local switch as the root switch or backup root switch of an instance of a spanning tree:

  [System view] [undo] stp [instance instance-id] root { primary | secondary }

  In the MSTP module, the current switch can be specified as the root switch or backup root switch of an instance (determined by the **instance** *instance-id* parameter) of a spanning tree. If the *instance-id* parameter is 0, the current switch is specified as the root switch or backup root switch of CIST.

The root types of a switch in different instances of a spanning tree are independent of each other. A switch can serve as the root switch or backup root switch in more than one instance. However, a switch cannot serve as the root switch and backup root switch at the same time in the same instance.

2.4 BPDU Protection

Commonly, the access ports of a device at the access layer are connected to a user terminal, such as a PC or file server, and are configured as edge ports to enable fast state transition on the ports. When these ports receive BPDUs, the system sets them as non-edge ports and recalculates the spanning tree. Therefore, flapping of network topology occurs. In normal situations, no BPDUs of spanning tree protocols are received on these ports. If pseudo BPDUs are sent to attack the switch, network flapping occurs.

The BPDU protection function prevents this type of network attacks.

When BPDU protection is enabled on a switch and an edge port receives a BPDU, the system disables the port and notifies the NMS that the port is disabled by MSTP. The disabled port can only be enabled by a network administrator. The BPDU protection function is recommended for the switch configured with an edge port.

By default, the BPDU protection function is disabled on a switch.

The command line used to configure the BPDU protection function is as follows:

[System view] [undo] stp bpdu-protection

2.5 Root Protection

If the root switch on a network is incorrectly configured or attacked, it may receive a BPDU with a higher priority. Then the root switch becomes a non-root switch, which changes the network topology. This also causes the traffic over high-speed links to be transmitted over low-speed links, and traffic becomes congested on the network. The root protection function prevents this situation.

After root protection is enabled on a port, the port keeps the role of the designated port in all instances. When the port receives a BPDU with a higher priority, the port transitions to the listening state and stops forwarding packets (that is, the links connected to the port are

disconnected). The port returns to the normal state if it receives no packet with a higher priority within a specified period.

Note the following points when configuring root protection:

1. Only one of loop protection, root protection, and edge port can be enabled on a port at one time.
2. In MSTP, root protection applies to all instances.

By default, the root protection function is disabled on a switch.

The command line used to configure the root protection function is as follows:

[Interface view] [undo] stp root-protection

2.6 Loop Protection

A switch maintains the state of the root ports and blocked ports using BPDUs from the upstream switch. These ports may fail to receive BPDUs from the upstream switch due to link congestion or unidirectional link failure. In this case, the switch reselects root ports. The original root ports become designated ports, and the blocked ports go into the forwarding state. As a result, a loop occurs on the switching network. The loop protection function prevents such a loop. With the loop protection function enabled, the ports whose roles are changed go into the discarding state, and the blocked ports remain in discarding state and do not forward packets.

Note the following points when configuring loop protection:

1. Only one of loop protection, root protection, and edge port can be enabled on a port at one time.
2. In MSTP, the loop protection function applies to instances whose port role is root, alternate, or backup.

By default, the loop protection function is disabled on a switch.

The command line used to configure the loop protection function is as follows:

[Interface view] [undo] stp loop-protection

2.7 TC Protection

According to the IEEE 802.1w and IEEE 802.1s standards, a switch purges the MAC table after detecting a topology change or receiving a transmission convergence (TC) packet. If the switch suffers a TC attack (receives TC packets continuously), the switch deletes MAC address entries repeatedly, which affects the forwarding service. With the TC protection enabled, the switch does not need to repeatedly delete MAC address entries, and normal operation of services is guaranteed.

By default, the TC protection function is enabled on a switch.

The command line used to configure the TC protection function is as follows:

[System view] stp tc-protection { undo}

2.8 Timeout Time Factor Configuration for a Switch

The roles of root ports, alternate ports, and backup ports remain unchanged if these ports receive a BPDU within each hello time. If one of these ports does not receive a BPDU for some reason (for example, the CPU is busy, STP BPDUs suffer interference from other protocol packets, or an aggregation sub-interface receives a BPDU and transparently transmits it to an aggregation interface) after the packet timeout time, the port is calculated as a designated port.

- The packet timeout time in STP is calculated as follows:

  MIN (MaxAge - EffectiveAge, 3 * HelloTime)
- In the implementation of Huawei STP, the packet timeout time is corrected as follows:

  MIN(MaxAge - EffectiveAge, 3 * HelloTime) * TimeoutFactor

The TimeoutFactor parameter in the commands can be set by a command line.

- STP module: [System view] stp timeout-factor number
- MSTP module: [System view] stp time-factor number

On a stable network, the value 5, 6, or 7 is recommended as the timeout time factor. By default, the timeout time factor for a switch is 3.

2.9 Configuration Digest Snooping

According to IEEE 802.1s, if interconnected switches implement interworking between MSTIs in the MSTP region, their region configurations (including region name, revision level, and mapping between VLANs and instances) must be consistent. When MSTP sends a BPDU, it puts the configuration ID into the BPDU. A configuration ID consists of region name, revision level, and configuration digest. The configuration digest is a 16-byte signature generated by encrypting the mapping between VLANs and instances through the HMAC-MD5 algorithm. The interconnected switches use the configuration ID to determine whether the switch that sends packets shares the same region with them.

After comparing section 13.7 in the IEEE 802.1Q – 2003 standard, Huawei finds that the calculation result of the configuration digest of a Cisco switch is inconsistent with the example given in the standard.

Because the configuration digest calculated by a Cisco MSTP switch does not comply with the IEEE 802.1s standard, the configuration digests calculated by an interconnected Huawei switch and Cisco switch are different, even if their region configurations are the same. Therefore, in practical application they are not considered to be in the same region. Only interworking in the CIST can be implemented, and interworking between MSTIs cannot.

Huawei MSTP switches provide a configuration digest snooping function to implement interworking with Cisco MSTP switches between MSTIs in a region.

Given that the region configuration of a Huawei switch is consistent with that of a Cisco switch, you can run a command to enable the configuration digest snooping function on any port connected to the Cisco switch. With configuration digest snooping enabled, the Huawei MSTP-enabled switch snoops the packet to replace the calculated configuration digest of the switch with the configuration digest in the packet. Therefore, the configuration ID of the Huawei switch is identical to that of the Cisco switch, and interworking between MSTIs can be implemented.

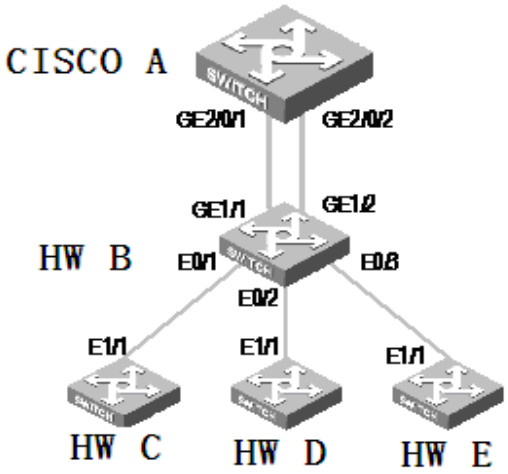
Note the following points when configuring configuration digest snooping:

1. Configuration digest snooping can be enabled on only one port. If this function is enabled on another port, it is automatically disabled on the previous port. In this way, configuration digests from different regions will not be received at the same time.
2. Configuration digest snooping can be enabled only when the region configurations of all switches in the region are identical. Otherwise, a broadcast storm may happen because the mappings between VLANs of the switches and instances are inconsistent.
3. When configuration digest snooping is enabled on some switches in a region, the region configurations of these switches can be modified when configuration digest snooping is disabled. Otherwise, inconsistent mappings between VLANs of all switches and instances may cause a broadcast storm during region configuration modification.
4. With configuration digest snooping enabled on a port, a switch always saves the configuration digest received last. Even if the port fails, the configuration digest received last is still valid.
5. If all the switches in the region are Huawei switches, it is unnecessary to enable configuration digest snooping.

   The command line used to configure the configuration digest snooping function is as follows:

   [Interface view] [undo] stp config-digest-snooping

In the following example, CISCO A is a switch supplied by Cisco, the other switches are all supplied by Huawei, the MSTP is enabled on all the switches, and the region configurations are all the same.



HW B is directly connected to CISCO A. Therefore, the configuration digest snooping function is enabled on GE1/1 or GE1/2. The configuration is as follows:

[HW B-GigabitEthernet1/2]stp config-digest-snooping

Although HW C, HW D, and HW E are not directly connected to CISCO A, they share the same region with CISCO A and they can obtain the configuration digest of CISCO A from HW B. Therefore, the configuration digest function needs to be enabled on their ports E1/1as follows:

[HW C-Ethernet1/1]stp config-digest-snooping

[HW D-Ethernet1/1]stp config-digest-snooping

[HW E-Ethernet1/1]stp config-digest-snooping

After the preceding configurations, all switches can implement interworking between MSTIs in the MSTP region.

2.10 No Agreement Check

Because the implementation mechanism of Cisco MSTP state machine is different from that of Huawei, the designated ports on Huawei switches connected to Cisco switches cannot quickly transition to the forwarding state.

Through analysis, Huawei found that the fast transition mechanism of Cisco MSTP is similar to that of RSTP. That is, MSTP packets sent from a designated port do not carry the agreement flag, and only root ports send the agreement flag. Therefore, the root ports of a Huawei switch are unable to receive packets with the agreement flag from the designated ports on the upstream switch and the syned flag cannot be set at all. As a result, the root ports on the Huawei switch cannot shake hands with the designated ports on the Cisco switch, and fast transition cannot be implemented.

The following command is configured on the port of the Huawei switch connected to the Cisco switch.

[Interface view] [undo] stp no-agreement-check

If the port connected to the Cisco switch is a root port, the syned flag of the root port is not checked when the allsyned flag is calculated, so fast transition is implemented.

2.11 Support for Standard 802.1s MSTP Packet Format

Standard 802.1s MSTP packet format Cisco MSTP packet format

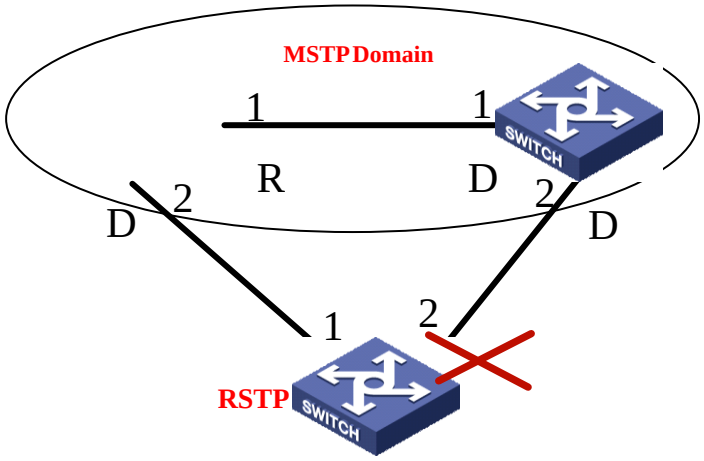
MST BPDU Standard Format		**MSTP BPDU Cisco Format**	
Field	**Octet**	**Field**	**Octet**
Protocol Identifier	2	Protocol Identifier	2
Protocol Version Identifier	1	Protocol Version Identifier	1
BPDU Type	1	BPDU Type	1
CIST Flags	1	CIST Flags	1
CIST Root Identifier	8	CIST Root Identifier	8
CIST External Path Cost	4	CIST External Path Cost	4

MST BPDU Standard Format		MSTP BPDU Cisco Format	
CIST Regional Root Identifier	8	CIST Bridge Identifier	8
CIST Port Identifier	2	CIST Port Identifier	2
Message Age	2	Message Age	2
Max Age	2	Max Age	2
Hello Time	2	Hello Time	2
Forward Delay	2	Forward Delay	2
Version 1 Length = 0	2	Version 1 Length = 0	2
Version 3 Length	2	Reserved 1	1
MST Configuration Identifier	50	Version 3 Length	2
CIST Internal Root Path Cost	4	MST Configuration Identifier	50
CIST Bridge Identifier	8	CIST Regional Root Identifier	8
CIST Remaining Hops	1	CIST Internal Root Path Cost	4
MSTI Configuration Messages (may be absent)		CIST Remaining Hops	1
		Reserved 2	1
		MSTI Configuration Messages (may be absent)	

MSTI Configuration Message Standard Format		MSTI Configuration Message Cisco Format	
Field	**Octet**	**Field**	**Octet**
MSTI Flags	1	MSTI Identifier	1
MSTI Regional Root Identifier	8	MSTI Flags	1
MSTI Internal Root Path Cost	4	MSTI Regional Root Identifier	8
MSTI Bridge Priority	1	MSTI Internal Root Path Cost	4
MSTI Port Priority	1	MSTI Bridge Identifier	8
MSTI Remaining Hops	1	MSTI Port Identifier	2
		MSTI Remaining Hops	1
		Reserved	1

A standard Ethernet frame contains 64 standard MSTI messages or 48 Cisco MSTI messages.

In the figure below, if the standard MSTP packet format is used, the packets received by the two ports on the RSTP switches are the same.



To implement interworking with Cisco MSTP switches, Huawei MSTP switches use the Cisco MSTP packet format.

Huawei switches can automatically identify the two packet formats.

The command lines used to configure the MSTP packet format are as follows:

- [Interface view] stp compliance { legacy | dot1s | auto }
- [Interface view] undo stp compliance

By default, Huawei switches send packets in the legacy format.

Cisco switches support MSTP packets in the standard format.

The command lines used to configure the MSTP packet format on Cisco switches are as follows:

- [Interface view] Spanning-tree pre-standard
- [Interface view] Undo Spanning-tree pre-standard

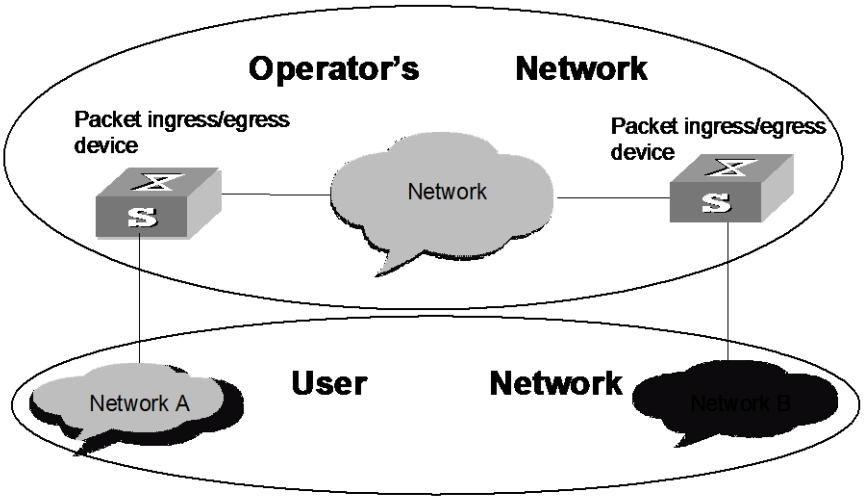
By default, Cisco switches send packets in the format automatically identified.

If Cisco switches support the standard packet format, as well as the standard configuration digest generation algorithm, Huawei switches can implement interworking with Cisco switches between MSTIs even if the configuration digest snooping function is disabled.

2.12 BPDU Tunnel

Through BPDU tunnel technology, a user's network located in different areas can transparently transmit BPDUs through specified VLAN VPNs within the operator's network. In addition, the spanning tree of the user's network and that of the operator's network are irrelevant to each other.

In the following figure, the top circle is an operator's network; the bottom circle is a user's network. The operator's network consists of packet receiving/transmitting devices. The user's network consists of user's networks A and B. On the operator's network, the destination MAC address of a BPDU is replaced with a special MAC address on a packet receiving/transmitting device on one side. On the device at the other side, the MAC address is retained. In this way, BPDUs can be transparently transmitted on the operator's network.



The command lines used to configure BPDU tunnels on a switch are as follows:

- [Interface view]bpdu-tunnel enable
- [Interface view]stp bpdu vlan *vlan-id*

Note the following precautions for BPDU tunnel configuration:

1. STP must be enabled on a switch with a BPDU tunnel established. Otherwise, after the BPDUs from the user's network arrive at the switch, these BPDUs are not sent to the CPU for processing. Therefore, the MAC addresses in the BPDUs are not replaced and BPDUs cannot be transmitted transparently.
2. The **vlan-vpn enable** command cannot be run on a port on which DOT1X, GVRP, GMRP, STP, and NTDP are enabled.

3 Interworking

3.1 Interworking Between STP, RSTP, and MSTP

The IEEE considers the interworking between STP, RSTP, and MSTP when drafting these standards. In hybrid networking, the standards guarantee that the network will not experience loop. Cooperation in fast transition between RSTP and MSTP, however, is subject to certain restrictions:

Due to the fast transition mechanism of designated ports in RSTP/MSTP (the state of a designated port can transition quickly only when the port receives a packet with the agreement flag from the downstream device), the following problem may occur:

RSTP is enabled on the upstream bridge while MSTP is enabled on the downstream bridge. In this case, the upstream bridge does not send any packet with the agreement flag to the downstream bridge, and the root port on the downstream bridge receives no packet with the agreement flag. This means that the MSTP-enabled port is not synchronized and the root port does not send any packet with the agreement flag to the designated port on the RSTP-enabled upstream bridge. Therefore, packets with the agreement flag in the MSTP region are suppressed, and the designated port on the RSTP-enabled upstream bridge can enter the forwarding state only after twice the forwarding delay time.

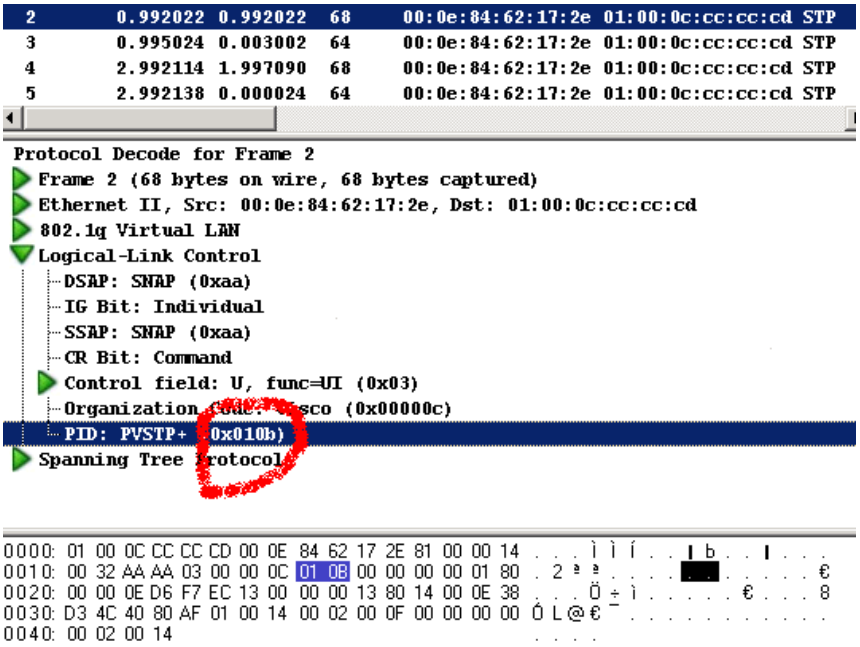
It is recommended that the MSTP-enabled bridge serve as the upstream bridge and the RSTP-enabled bridge serves as the downstream bridge. In this way, synchronization of the RSTP-enabled port does not require the root port to receive a packet with the agreement flag from the upstream bridge. The designated port on the MSTP-enabled upstream bridge can receive packets with the agreement flag from the root port on the RSTP-enabled downstream bridge, and the state of the designated port can transition quickly.

3.2 Interworking Between STP/RSTP/MSTP and PVST+

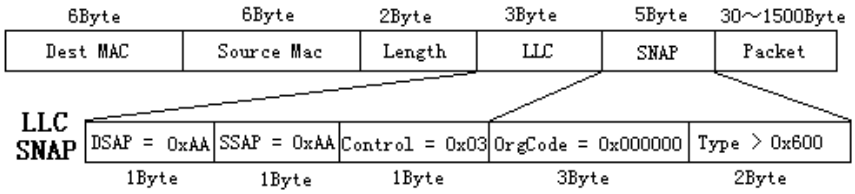
As explained in Chapter 1, STP, RSTP, and MSTP are all standard protocols defined by the IEEE. They can implement interworking to some extent with Cisco Per VLAN Spanning Tree (PVST)+.

Cisco PVST+ is based on VLANs. When a device that supports various IEEE STP standards is interconnected with a PVST+ device, no problem exists in STP interworking, if they are interconnected through access ports. The standard STP device considers the PVST+ device as one that supports the IEEE 802.1D standard. If they are interconnected through trunk ports, the standard STP device can implement interworking with VLAN 1 of the PVST+ device. In the other VLANs, the standard STP device is unable to recognize PVST+ packets. In this case, special processing is required.

The BPDU processing format for Cisco PVST+ is shown in the following figure:



The Ethernet encapsulation format of a PVST+ BPDU is SNAP (also known as Ethernet_SNA), which is shown in the following figure:



In the first figure, the value of the field circled in red is 0x010B. This is the Type field to be encapsulated in the SNAP format. In Ethernet encapsulation, the value of the Type field is required to be greater than 0x600 to distinguish between the Type and Length fields. However, the Type field in PVST+ BPDUs is less than 0x600. Due to the incorrect value of the Type field in PVST+ BPDUs, many devices may discard the PVST+ BPDUs, instead of forwarding them.

For a standard STP device that cannot transmit PVST+ BPDUs transparently, any physical loop must be blocked on the standard STP device. That is to say, the blocked port must be on the standard STP device, instead of the PVST+ device. Otherwise, a broadcast storm may occur in VLANs other than VLAN 1.

For a standard STP device that can transmit PVST+ BPDUs transparently, the standard device forwards PVST+ BPDUs as multicast packets in a VLAN. The PVST+ device can receive the desired PVST+ BPDUs correctly and then calculate and eliminate the loops in other VLANs. In this case, no special configuration is required on the standard STP device.

3.3 Interworking Between MSTP-Enabled Huawei and Cisco Switches in a Region

For information about interworking between an MSTP-enabled Huawei switch and an MSTP-enabled Cisco switch, see section 2.9 "Configuration Digest Snooping" and section 2.10 "No Agreement Check."

4 Appendix

4.1 Default Configurations of the RSTP Module

Item	Default Setting
Whether STP is enabled globally	By default, STP is enabled globally on the switches that support the intelligent resilient framework (IRF) and is disabled on other switches.
Whether a port is enabled	Enabled
STP operating mode	RSTP mode
Bridge priority	32768
Root bridge designation and root bridge backup	Disabled
Network diameter	7
Forward delay	15
Hello timer	2
Max age	20
BPDU timeout time factor	3
Maximum BPDU transmission rate	3
Edge interface	Disabled
Path cost standard	Legacy
Path cost	Automatically calculated according to the path cost standard
Interface priority	128
Point-to-point (P2P) link	Automatically calculated
BPDU protection	Disabled

Item	Default Setting
Root protection	Disabled
Loop protection	Disabled
TC protection	Enabled
Configuration digest snooping	Disabled

4.2 Default Configurations of the MSTP Module

Item	Default Setting
Whether STP is enabled globally	By default, STP is enabled globally on the switches that support the IRF and is disabled on other switches.
Whether a port is enabled	Enabled
Operating mode	MSTP mode
MST region name	Bridge MAC expressed in a hexadecimal character string
Revision level	0
Mapping between the VLAN and the MSTI	All VLANs are mapped to CIST.
Bridge priority	32768
Root bridge designation and root bridge backup	Disabled
Maximum number of hops in the MST region	20
Network diameter	7
Forward delay	15
Hello timer	2
Max Age	20
BPDU timeout time factor	3
Maximum BPDU transmission rate	3
Edge interface	Disabled
Path cost standard	Legacy
Path cost	Automatically calculated according to the path cost standard

Item	Default Setting
Interface priority	128
P2P link	Automatically calculated
BPDU protection	Disabled
Root protection	Disabled
Loop protection	Disabled
TC protection	Enabled
Configuration digest snooping	Disabled