Spanning Tree Protocols

One Picture Summary

Basic Operational Parameters

STP multicast address:	01-80-C2-00-00-00
Bridge ID:	2 byte priority + 6 byte MAC
Port ID	1 byte priority (128 default) + 1 byte interface ID. (0x 8005,)
Root Bridge	Switch with **lowest** bridge ID
Max Age	20 secs default, then last valid BPDU will be dropped

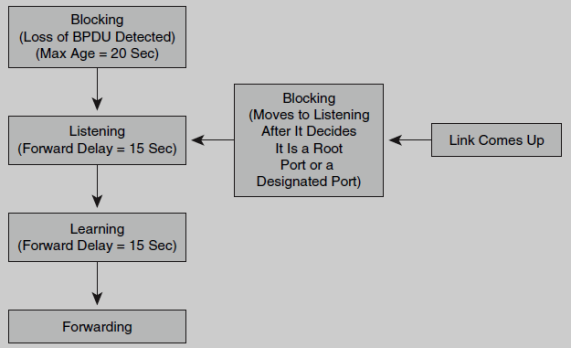
Root Port: only one for each switch, Port who receives inbound BPDU with the lowest root path cost

Designated Port: Only one in each collision domain. The port residing on the switch who has the lowest root path cost to root bridge wills wins. If tied, then lower sender bridge ID, lower sender port ID.

Table 7-3 *STP Path Cost*

Link Bandwidth	Old STP Cost	New STP Cost
4 Mbps	250	250
10 Mbps	100	100
16 Mbps	63	62
45 Mbps	22	39
100 Mbps	10	19
155 Mbps	6	14
622 Mbps	2	6
1 Gbps	1	4
10 Gbps	0	2

2-24 *Spanning-Tree Port States*



BPDU Types

UplinkFast Configuration

//designed for fast recover of direct topology change, the candidate root ports are named as Alternate port.

- Uplinkfast tracks all possible paths to root bridge, so failover can take place without delay.
- Uplinkfast only available on non-root bridges, it might modify local switch to prevent it becoming root –bridge. It ensure the local switch is the farthest to the root bridge, namely leaf-node switch, not a transit node to root bridge. (often used in access switch)
- During failover, switch updates its local CAM table for uplink and multicast (0100-0ccd-cdcd) to upstream switch to notify the new downstream port link.
- Multicast packet is sent out at a rate specified by max-update-rate parameters in unit of packets per second. (0~65535, 150 packets per second)

Switch(config)# **spanning-tree uplinkfast [xx (pps)]**
Swtich# **show spanning-tree uplinkfast**

Backbone Fast Configuration

Designed for fast recovery of indirect topology change, and should be enabled on all switches for the sake of RLQ

When non-root bridge receives inferior BPDUs due to indirect topology change, it will immediate confirm the root-bridge reachability by sending out Root Link Query (RLQ). Thus non-root bridge can instantly transition its blocked port to listening state rather than waiting for Max-age timer to response to the inferior BPDUs.

- If inferior BPDUs on Blocking port, then alternative paths to root bridge includes root port and all other blocking ports.
- If inferior BPDUs on root port, then all blocking ports are alternative paths to root bridge.
- If inferior BPDUs on root port and no ports are blocked, then loss connectivity to root bridge is assumed.

Switch(config)# **spanning-tree backbonefast**
Swtich# **show spanning-tree backbonefast**

Securing STP

STP Root Guard

- Prevent a port to become a root port
- As long as superior BPDUs are being received, the port will be kept in the ***root-inconsistent STP state***, in which no data can be received or sent except listening BPDUs.

Switch(config-if)# **spanning-tree guard root**
Swtich# **show spanning-tree inconsistentports**

STP BPDU Guard

- If any BPDU detected, port will be put into ***errdisable state*** immediately.

Switch(config)# **spanning-tree portfast bpduguard default**
Switch(config-if)# **spanning-tree bpduguard enable**
Swtich# **show spanning-tree bpduguard**

Configuration BPDU:
Used for spanning-tree computation, sent out by root bridge **every 2 seconds** by default.
Topology ChangeNotification (TCN) BPDU:
Used to announce changes in the network topology, not periodical packet by triggered by network change.

Table 7-2 *Configuration BPDU Message Content*

Field Description	Number of Byt
Protocol ID (always 0)	2
Version (always 0)	1
Message Type (Configuration or TCN BPDU)	1
Flags	1
Root Bridge ID	8
Root Path Cost	4
Sender Bridge ID	8
Port ID	2
Message Age (in 256ths of a second)	2
Maximum Age (in 256ths of a second)	2
Hello Time (in 256ths of a second)	2
Forward Delay (in 256ths of a second)	2

Topology Change Procedure

Direct Topology Change:
- Physical link on root port failure
- Non-root bridge immediately remove the best BPDU, and wait for new BPDU from root bridge
- Root-bridge send out configuration BPDU with TCN set
- .non-root bridge receives new BPDU from root bridge and store it as Best BPDU, port transition from block to forward state (30 secs)

Indirect Topology Change:
- Traffic filter between root bridge and non-root bridge
- No BPDU is removed, and No TCN BPDU is sent as no physical interface failure.
- After Max Age (20 secs) expires, best BPDU is flushed, non-root bridge waits for new root BPDU from any of its ports.
- When root BPDU received on other ports, it is stored as bet BPDU.
- The port which receives this best BPDU transition to Forward state. (30 secs + Max Age 20 secs)

insignificant Topology Change:
- Non-PortFast switch access port goes up or down
- Will not change STP change, but trigger TCN BUPD send out to root bridge, which further age out MAC table
- In large network it could cause ***unknown unicast flood***

Protect from unexpected BPDU loss

Loop Guard on blocked port
- If a blocked port receiving a steady flow of incoming BPDU, experiencing a sudden stop of BPDU flow, Loop guard temporarily moves the port into ***loop-inconsistent state*** until the BPDU flow recovers.
- Loop Guard actions on per VLAN basis.

Switch(config)# **spanning-tree loopguard default**
Switch(config-if)# **spanning-tree guard loop**

UDLD – Unidirectional Link Detection
- Monitoring link by periodical UDLD message on both direction (15 secs by default) (interval should be long enough to detect problem before port transition from block to forward which takes 50 secs at most)
- **UDLD Normal Mode**: when unidirectional link detected, UDLD mark the port as having an ***undetermined state*** and generate a syslog message, but allow port to continue its operation.
- **UDLD Aggressive Mode**: when unidirectional link condition is detected, switch takes action to re-establish the link by sending out UDLD message once a second. After 8 UDLD messages sent with no echo reply, port is placed in ***the Errdisable State***.
- UDLD Operation is independent in each physical link, even in EtherChannel interface. In one physical link, UDLD operation is independent on two directions. **UDLD must be enabled on both switches**.

Switch(config)# **udld {enable | aggressive | message time xxxx}**
Switch(config-if)# **udld {enable | aggressive | disable}**

Completely disable STP on Portfast port by BPDU filtering
- BPDU filtering feature only take effect on PortFast ports.
- BPDU filtering prevents port from sending or processing any BPDU packets.

Switch(config)# **spanning-tree portfast bpdufilter default**
Switch(config-if)# **spanning-tree bpdufilter {enable | diable}**

Advanced STP

Step 1 Change Notice: Non-Root switch continually sends TCN BPDU out its root port every Hello Time Interval until it gets an ACK from upstream neighbors, only informing the change and no details about the change.
Step 2 Change broadcast: Upon receiving the TCN BPDU, Root Bridge sends out Configuration BPDU with Topology Change Flag set, which are relayed to all other Non-Root Bridges.
Step 3 MAC table flush: the only reaction of non-root bridge receiving configuration BPDU is to shorten the aging out of MAC table from default 300 secs to Forward Delay Value 15 secs.

Types of STP

Name	Trunk	BPDU scope	Standard
CST	802.1Q based	Native VLAN BPDU version 0	IEEE 802.1D
PVST	Cisco ISL based	Per VLAN	Provide load balance.
PVST+	802.1Q and ISL	Per VLAN	Interoperate with PVST & CST
RSTP		BPDU version 2 Aims at fast convergence.	IEEE 802.1w, can co-exist with MST & PVST+
RPVST+			RSTP + PVST+
MST		Aims at reducing CPU consuming.	IEEE 802.1s, compatible with all above STP types.

By default, a switch operates in Per-VLAN STP (PVST+) mode using traditional 802.1D STP. RSTP cannot be used until a different spanning-tree mode (MST or RPVST+) is enabled.
STP architecture is primarily driven by Traffic Pattern & Network Topology.

Configuration Commands

Bridge ID manipulation

Switch(config)# **spanning-tree vlan xx priority xx**
Switch(config)# **spanning-tree vlan xx root {primary | secondary }**

Port ID manipulation

Switch(config-if)# **spanning-tree [vlan xx] cost xx**
Switch(config-if)# **spanning-tree [vlan xx] port-priority xx**

STP Timer manipulation

//only need to modify timers on root bridge as it propagate them to all non-root bridges.
Switch(config)# **spanning-tree [vlan xx] hello-time xx**
Switch(config)# **spanning-tree [vlan xx] forward-time xx**
Switch(config)# **spanning-tree [vlan xx] max-age xx**

PortFast Configuration

- Port go up or down will not trigger TCN BUPD
- STP is still running as normal on portfast port just in case
- 2 benefits, PC not wait , avoid unknown unicast flood.

Switch(config)# **spanning-tree portfast default**
Switch(config-if)# **spanning-tree portfast**
Swtich# **show spanning-tree interface ga0/1 portfast**

RSTP

- RSTP BPDU uses different version (2) than the 802.1D STP (version 0), thus they can co-exist in a same switch, RSTP can revert to 802.1D on a per-port basis.
- Proposal Bit in BPDU is used in RSTP

Port Role	Port Transition	Port State	Link Type
Root	Not rely on link type Transit immediately when receiving superior BPDU.	Discarding Learning Forwarding	Point to point Shared Derived from duplex mode, can also be overridden by manual setting
Designated	Rapid Transit only on P2P link type.		
Alternate	Transit not depends on link type.		
Backup			
Disabled			

Tips to bear in mind:
- In all Port stats, port will send out, accept and process BPDU frames.
- In RSTP, port states and roles can transition independently of each other.
- Unlike PortFast, Edge port becomes normal STP port when receiving BPDU.
- **When a designated port is in a discarding or learning state (and only in this case), it sets the proposal bit in the BPDUs and sends out, then wait for agreement. If no agreement received, it reverts to 802.1D to slowly transition to forwarding state.**
- Root port rapidly transition to forwarding state as soon as superior BPDU received from the root and it puts the non-edge designated ports in discarding state. **This operation is called sync** (between neighbor switches).

How to achieve fast converge?
- Every switch port sends out BPDU at Hello Time Interval regardless of port state and role, 3 BPDUs are missed in a row, that neighbor is presumed to be down (**fast failure detection, 6 secs**), infos related to that port ages out immediately.
- Select Alternate and backup port in advance.
- All port states accept and process BPDU, Root & Designated Port can rapidly transition through Proposal & Agreement mechanism (**neighbor switch interaction for transition to forwarding fastly, 4 secs**), and all non-edge designated ports be in discarding sate, Proposal & Agreement handshake will gradually spread to other leaves of the STP Tree (**synchronization**).

MST

- Number of MST instances required depends on number of logical topologies existed.
- Instead of each instance for each VLAN or one STP instance for all VLANs, In each MST Region, MST allow you to configure up to 16 STP instance (MSTI) with each hosts any number of VLANs.
- MST is designed to interoperate with all other forms of STP.
- MST Regions: same MST configuration name, MST configuration revision number, MST instance-to-VLAN mapping table.
- CST regards MST region as a single “Switch”, and maintain the loop-free topology. While inside MST region, Internal Spanning tree (IST- a single CST instance) is rinning to work out a loop-free topology.

Switch(config)# **spanning-tree mode mst**
Switch(config)# **spanning-tree mst configuration**
Switch(config-mst)# **name xxx**
Switch(config-mst)# **revision xxx**
Switch(config-mst)# **instance {instance-id, 0~15} vlan {vlan list}**