1-4(B). STP Concepts

Note: We're dealing with the original 802.1D Spanning Tree definition here. It's good for academic discussion, but in real life, you'll use something faster that works per-VLAN.

STP PURPOSE

L3 packets have a Time-To-Live (TTL) counter that is reduced at every router they pass through. When the packet's TTL hits zero, the packet dies. At L2, there is no such counter, so frames within a LAN can loop endlessly..

STP (Spanning Tree Protocol)—Blocks carefully selected links between switches so that every switch is still reachable, but frames can't loop back to switches they've already visited. Effectively, it turns a web of switches into a tree of switches.

Looping Frame Effects

- Broadcast Storms—by definition, broadcasts are flooded out all ports except the one they came in, just like multicasts and frames for MAC addresses that have never been learned. Thus, if the physical topology allows broadcast frames to loop, they will loop, and perhaps multiply. By definition, each host must process a broadcast frame, since it's addressed to "everyone." The result of a broadcast storm is that the bandwidth gets used up and hosts get overloaded. Windows machines will crash and networking careers will burn.
- MAC Table Instability—each switch is constantly learning new destinations for the same MAC address as literally the same frame loops around and re-enters the switch from a new direction.
- Multiple Frames Arrive at Host—Frames don't have a serial number, so if the frame loops, it'll re-arrive endlessly and the host will have to process it as if it was new.

STP doesn't change the connect/notconnect status of a port. It just adds logic to the forwarding decision: "act normally unless the port is blocked."

Convergence—The process of all switches agreeing on which ports should be blocked, either initially or after a physical topology change (link up or down).

STP DATA

Bridge ID (BID)—Concatenation of Priority & MAC.

802.1D	Bits	Description
Priority	16	In Per-VLAN Spanning Tree (Chapter 3), this will be further divided
MAC	48	

Hello BPDU (Bridge Protocol Data Unit)

- Root Bridge ID (Sender's current belief)
- Sender's BID (Bridge Priority, MAC)
- Sender's Root Cost
- Root Switch's Timer Values—hello, max age, forward delay

STP ALGORITHM (INITIALIZATION)

STP identifies every port that needs to forward frames. When it's done, the leftovers are blocked. The result is a tree topology that flows toward one "root" switch. A question that asks for blocked ports forces you to go through the entire process of identifying every single port serving any other role. The leftovers are blocked. There are no shortcuts.

- Root Switch—Elected. All of its interfaces forward. (They're all "designated ports.")
- RPs (Root Ports)—Every other switch (non-root) chooses one "root" port with the best path to the root switch (lowest "root cost") to forward frames.
- DPs (Designated Ports)—For each link (network segment), whichever switch has the lowest root cost is the designated switch for that segment and its interface is chosen to forward traffic to and from that segment.
- All other ports are blocking

PICK A ROOT SWITCH

Switch with lowest BID (Priority, MAC) wins—Since it's a concatenation, setting a lower priority overrides everything. This is easy to remember because the lowest MAC could easily belong to some antique in a poorly-connected corner of your net, badly suited to such a key role. There are no ties because the MAC address is unique.

The election starts with each switch putting its own BID as the root switch in the BPDUs (hellos) it sends out. Each time a switch receives a BPDU with a lower root BID, it starts naming that switch as the root. Eventually, they all agree (converge).

The root switch choice can be pre-empted by a priority change on a non-root switch because its periodic "hellos" will have a different BID—see the section "Topology Changes."

CHOOSE ROOT PORTS ON NONROOT SWITCHES

Each non-root switch has exactly one root port, the one with the lowest cost to the root switch. The switch listens to the root cost in incoming BPDUs and adds the cost of its own interface where the hello was received to get that port's "root cost." This happens to equal the sum of the costs of all ports the frame would exit when traveling to the root switch in that direction (because the root switch's cost is 0).

- Port with lowest root cost becomes the RP
- Ties broken with lowest BID (Priority, MAC) on the *upstream* switch (the BID is in the hello)

And for situations where the same two switches have multiple links between them

- Ties broken with lowest port priority (Range 0-255, default 128) on the upstream switch
- Ties broken with lowest port number on the upstream switch

Default Port Costs—Costs have been updated over time to cope with faster hardware. By default, Cisco uses 1998 values, but the 2004 values can be used instead:

```
S2(config)# spanning-tree pathcost method long
```

Must be entered on each switch; doesn't propagate from root.

Speed	10 Mbps	100 Mbps	1 Gbps	10 Gbps	100 Gbps	1 Tbps
Cost 1998	100	19	4	2		
Cost 2004	2,000,000	200,000	20,000	2,000	200	20

CHOOSE DESIGNATED PORT FOR EACH LAN SEGMENT

All switches on a segment compete to feed that segment—Remember this is link-centric, not switch-centric.

- Switch offering the lowest cost to the root switch wins
  - All ports on the root bridge have cost 0 and are designated ports—Always. Any blocking will happen at the other end.
  - Switchports with only hosts or routers always win (become designated & forwarding) because hosts and routers don't send BPDUs.
- Ties between lowest cost switches broken by lowest BID (Priority, MAC)
- Ties within the winning switch (parallel links to a hub or a cable looping back) broken by lowest interface priority, then interface number (on the winning upstream switch)

BLOCKED PORTS

Everything Else. Always happens to be on the downstream side because that switch has a higher root cost, placing it "downstream" and preventing those ports from being designated for their LAN segment.

STABLE-STATE BEHAVIOR

Root switch sends BPDUs out all active interfaces every 2 seconds (default) with root-cost = 0.

Non-root switches receive them on their root ports and forward them out all designated ports after changing sender BID and root cost.

TOPOLOGY CHANGES

If hellos stop or list different details, a switch recalculates the topology.

Timers—all timers are dictated by the root switch

- Hello—default 2 seconds
- MaxAge Default 10x hello = 20 secs.)—How long to wait after hellos stop before re-topo.
- Forward Delay (default 15 secs.)—How long a port stays in each of the intermediate states (listening & learning) while transitioning to forwarding.

After MaxAge passes without a hello (default 20 secs.), or the switch notices that its own root port has failed (immediately) the switch reruns STP choices.

- Chooses new root switch based on BPDUs (maybe self)—Why? It's possible that the L2 network has just bisected, with the old root switch no longer part of our switch's network.
- If not root itself, chooses new Root Port (RP) and decides whether any of its ports are DP on their segments.

STP separates the concepts of port roles (DP, RP) from port states (forwarding, blocking, etc.) Ports can move immediately from forwarding to blocking, but when moving from blocking to forwarding, must move through two extra states. Each lasts "forward delay" seconds, default 15 secs.

- Listening—No forwarding. The MAC address table is flushed of all addresses for the port
- Learning—No forwarding. Switch learns MACs received on interface.

Convergence following an event therefore takes 30 seconds (+ 20 for MaxAge, if the failed port wasn't its own).

Note: Because ports moving from forwarding to blocking take 0 seconds and blocking to forwarding takes 2xFowardDelay seconds. The network may be hosed for 30-50 seconds (default).

Port States

State	Forwards Data	Learns Macs	Stable / Transitory
Blocking	N	N	Stable
Listening	N	N	Transitory
Learning	N	Y	Transitory
Forwarding	Y	Y	Stable
Disabled	N	N	Stable

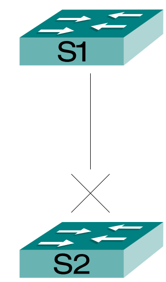
PER-VLAN STP

Cisco has added extensions which allow different tree topologies on different VLANs, spreading traffic across links that would be completely unused if all VLANs used the same tree. To accomplish this, the bridge ID is slightly modified. Bridge priority is split into priority and VLAN subfields.

802.1D	Cisco PVSTP+	Bits	Description
Priority	Priority	4	Priority is still the high 4 bits of a 16-bit numbers—values 0-65,535, granularity multiples of 4096. Default 32,768.
	VLAN ID	12	Often the priority is still displayed as the 16 bits, adding VLAN ID to the priority. For example, vlan 11 with priority of 32768 would show as a priority of 32779.
MAC	MAC	48	

Show commands still display the priority field as all 16 bits. The result is like adding the VLAN ID to the priority. For example, in VLAN 1, the default priority of 32,768 would display as 32,769 and in VLAN 2, it would be 32,770.

Again, STP doesn't affect the general status of ports; they remain up/up and connected. It only adds an extra step to the switching logic to prevent frames being forwarded out ports that aren't in the STP forwarding state. This is easier to remember thanks to Cisco's STP working per-VLAN—one VLAN isn't stopped from using a port simply because it has been blocked by another VLAN's STP.



Since MAC addresses aren't learned on ports in the blocking state, the MAC address table won't contain entries suggesting that the switch forward frames out a port that STP would prevent. For example, S2 won't have any entries pointing to S1.

At the other end, S1 obviously won't hear any frames coming from S2 and won't generate any MAC address table entries pointing out that port.

All of this can be seen with the command

```
S# show mac address-table dynamic
```