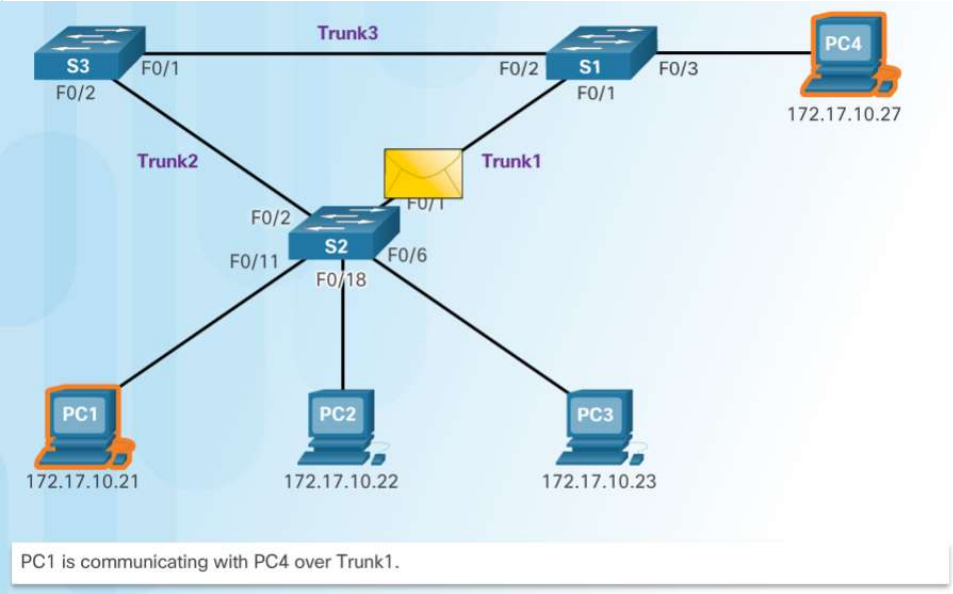
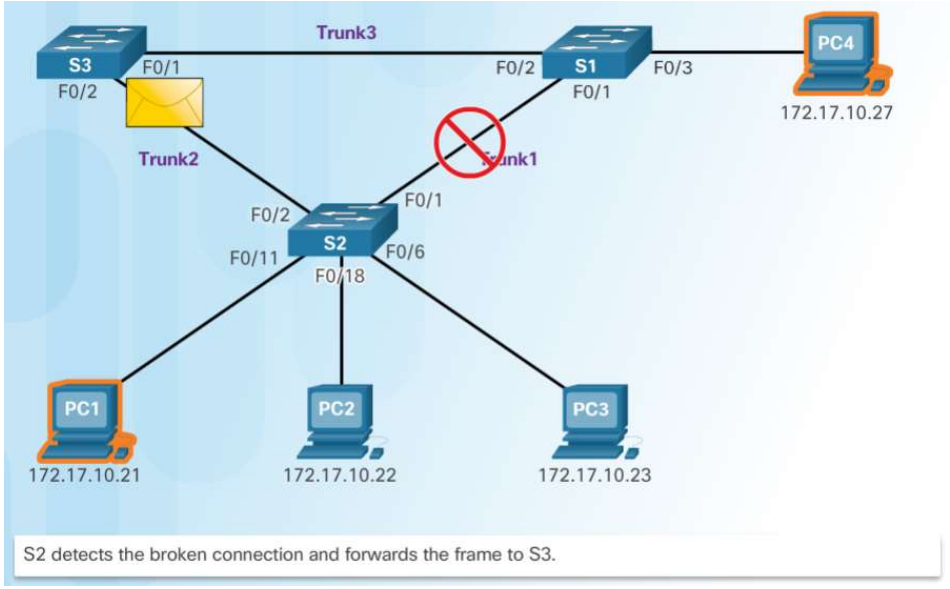
Spanning Tree Protocol (STP)

Chien-Chao Tseng

Redundancy at OSI Layers 1 and 2

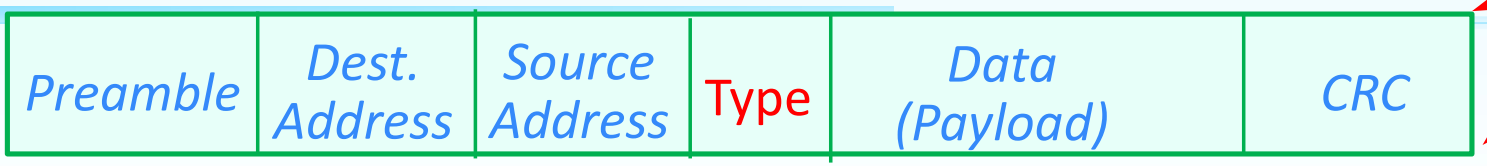
- Multiple cabled paths between switches provide **physical redundancy**
  - ➢Improves reliability and availability of networks
  - – **Alternate physical paths** for data to traverse the network
    - ➢Possible to access network resources, despite path disruption

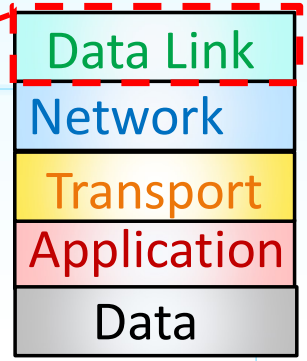




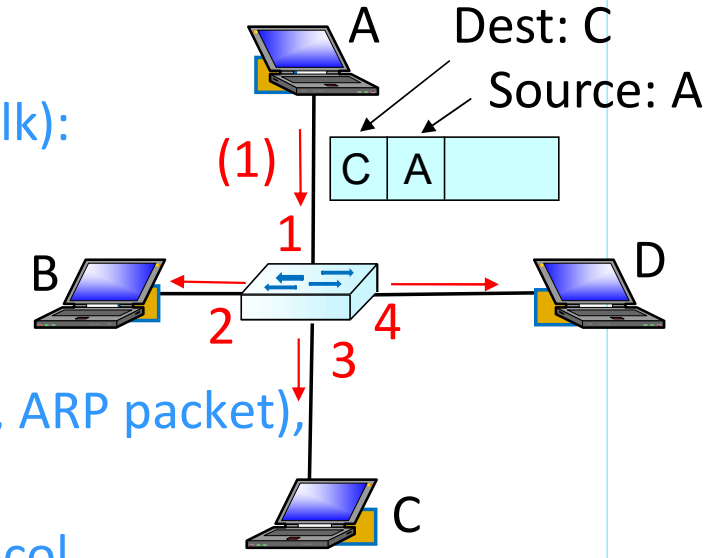
Source: Cisco

Ethernet Frame Structure





- Preamble: 7 bytes, to synchronize receiver, sender clock rates, 10101010
  - Start Frame Delimiter (SFD): One byte, 10101011
- Mac Addresses: 6 bytes, Flat
- Type: indicates higher layer protocol
  - Mostly IP, but others possible (e.g., Novell IPX, AppleTalk):
    - 0x0800: IPv4, 0x0806: ARP
- CRC: checked at receiver, frame dropped if error detected
- Behavior of an adapter on receiving a frame with
  - Matching Destination Address, Broadcast Address (e.g., ARP packet), or Participating Multicast Address
    - passes data in the frame to the Network Layer protocol
  - Otherwise, discards frame

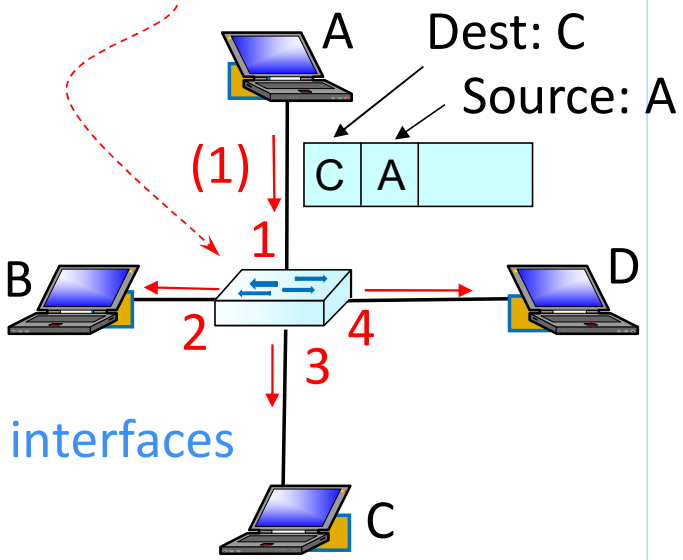


Switch Table and Self-Learning

- Each switch has a switch table
- **Table entry:**
  - MAC Address of Host, Interface to Reach Host, Time Stamp
  - Looks like a routing table

MAC Addr	Interface	TTL

- How does switch create and maintain table entries?



- **MAC Learning**

  Self-learning which hosts can be reached through which interfaces
  - When switch receives a frame, it
    - Learns location (port) of sender (incoming LAN segment)
    - Records Sender-MAC/Port pair in switch table

Flooding, Forwarding and Self-learning (MAC Learning)

- Assume switch table is initially empty

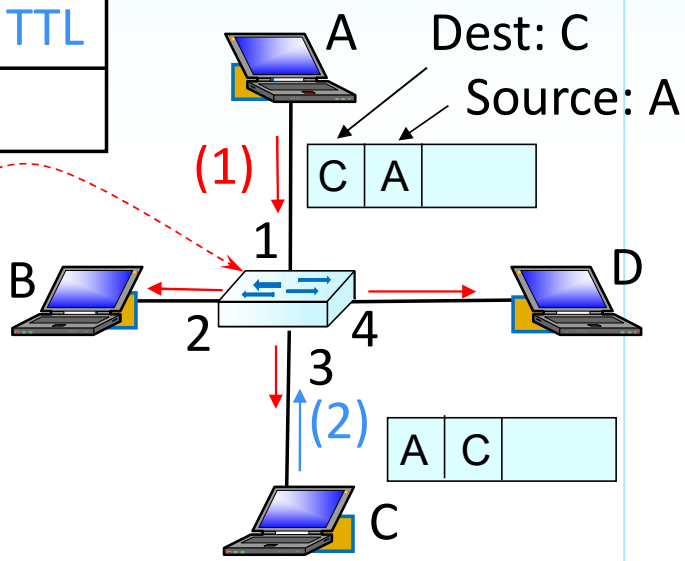
MAC Addr	Interface	TTL

1) A→C:
Destination C location unknown
   - Flood
   - MAC Learning
     - Update switch table

MAC Addr	Interface	TTL
A	*1*	*60*

2) C → A:
Destination A location known
   - Unicast (Selectively send on just one link)
   - MAC Learning
     - Update switch table

MAC Addr	Interface	TTL
A	*1*	*60*
C	*3*	*60*



Frame Filtering/Forwarding Algorithm

- Algorithm: when a switch receives frame

Skip

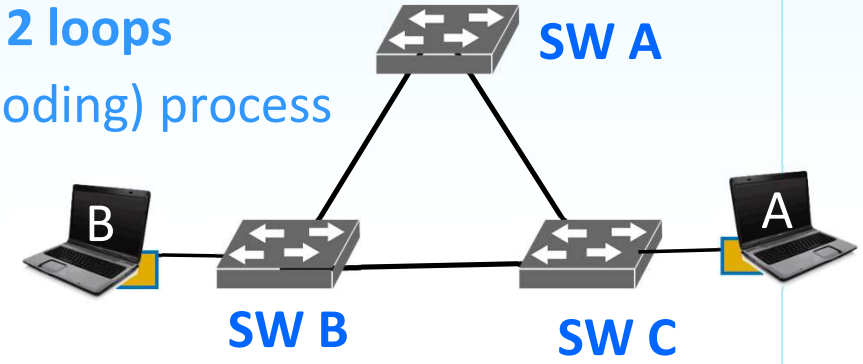
1. records incoming link, MAC address of sending host
2. indexes switch table using MAC destination address
3. if entry found for destination

```
    then {
            if destination on segment from which frame arrived
                then drop frame
                 else forward frame on interface indicated by entry
     }
    else flood  /* forward on all interfaces except arriving
                   interface */
```

- Note: Spanning Tree Protocol (STP, later in class)
  - https://www.youtube.com/watch?v=japdEY1UKe4, Step by Step, by CertBros

Logical Layer 2 Loops

- **Physical redundant paths** may cause **logical Layer 2 loops**
  - Due to switch **MAC learning** and **forwarding** (flooding) process
- Recall: How switches react to
  - Broadcast messages or
  - Unicast messages with **unknown** addresses
  - ➢ Forward frame out of every port, except the one that receives it

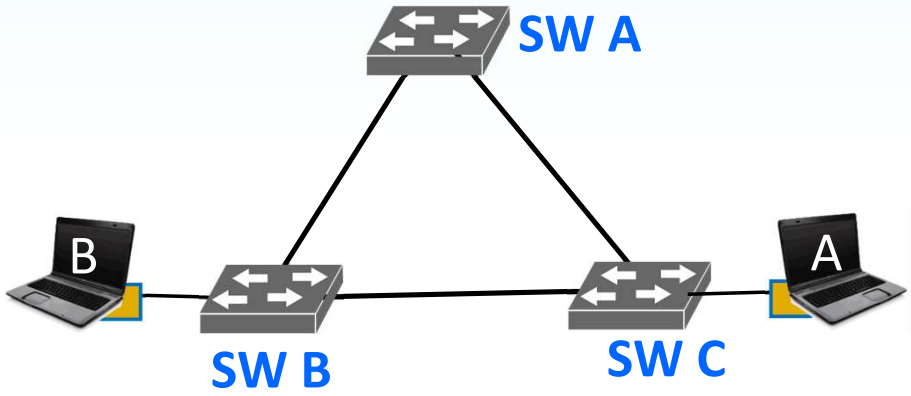


- **Three primary issues**:
  1. **Broadcast Storm**
  2. **MAC Database (address table) Instability**
  3. **Multiple-frame transmission**

  ➢**Solution: Spanning Tree Protocol (STP)**
- Reference: https://www.youtube.com/watch?v=japdEY1UKe4
  - Spanning Tree Protocol Explained | Step by Step, by CertBros

Primary Issues of Layer 2 Loops

1. **Broadcast storm—**
   Each switch may flood broadcasts endlessly,
   (if without some loop-avoidance process.)
2. **MAC Database (Address Table) Instability—**
   Copies of the same frame being received
   on different ports of the switch.
   (MAC learning)



3. **Multiple-frame Transmission—**
   Multiple copies of unicast frames may be delivered to destination stations.
   – Can cause unrecoverable errors
     - Many protocols expect to receive only a single copy of each transmission

➢ Uses **STP** to prevent loops when using redundant switches

Spanning Tree Protocol – Overview

- **Objective**:

  to ensure "**only one logical path** between all destinations" on the network

- **Idea:**

  Intentionally blocking ports of redundant paths that could cause a loop

  – User data is prevented from entering or leaving that port

  – But Bridge Protocol Data Unit (BPDU) frames are allowed to pass through.

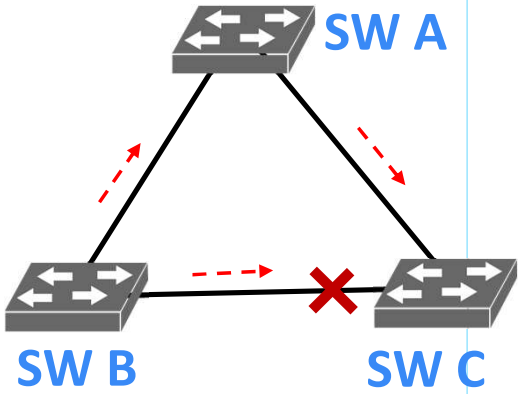
    - BPDU: control message used by STP
      - Encapsulated in an Ethernet frame

- **Multiple** physical paths still exist,

  – But disabled to prevent the loops from occurring

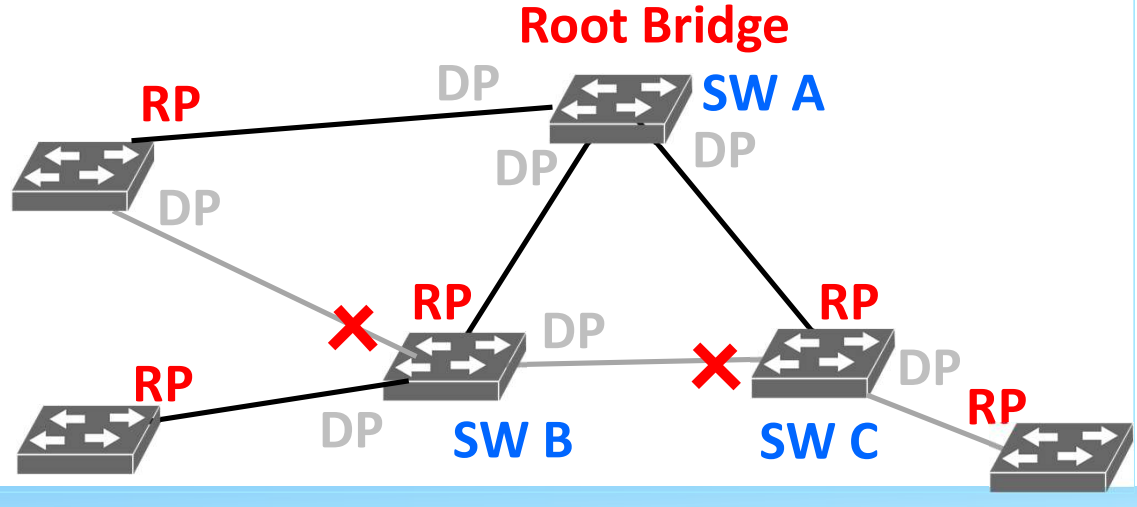
  – On a link or switch failure

    - STP recalculates the paths and unblocks the necessary ports
      - ➢ Allow the redundant path to become active



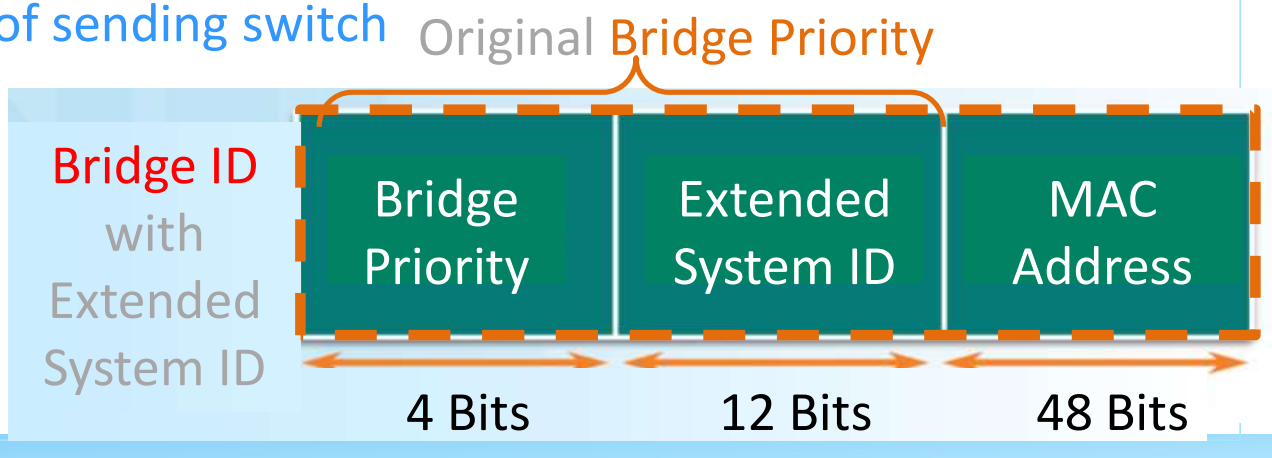
Spanning Tree Algorithm (STA)

- IEEE 802.1D STP and Rapid STP (RSTP)
  use **Spanning Tree Algorithm** (STA) to determine **which switch ports** on a network must be put in **Blocking State** (to prevent loops)
- **STA** designates **a single switch** as the **Root Bridge**
  - Used as the **reference point** for all **path calculations**
- **Root Ports (RPs)**: per-switch
  Ports on a switch with the **best route** to the **Root Bridge** (in terms of path cost.)
  - Selected on a **per-switch basis**
  - **Root Bridge** is the only bridge that does not have a Root Port.
- **Designated Port (DP)**: per-link
  Port on **a link (segment)** with the **best route** to Root Bridge
  - Selected on a **per-link basis**



Bridge ID (BID)

- **Bridge ID (BID)**: switch, 8 Bytes, divided into two parts
  - **Bridge Priority**: 2 Bytes originally,
    - automatically assigned, but configurable
    - value between 0 and 65,535, default is 32,768
    - **Extended System ID** (optional): borrowing 12 bits from original Bridge Priority
      - a VLAN ID or
      - a multiple spanning tree protocol (MSTP) instance ID
  - **MAC Address**: MAC address of sending switch

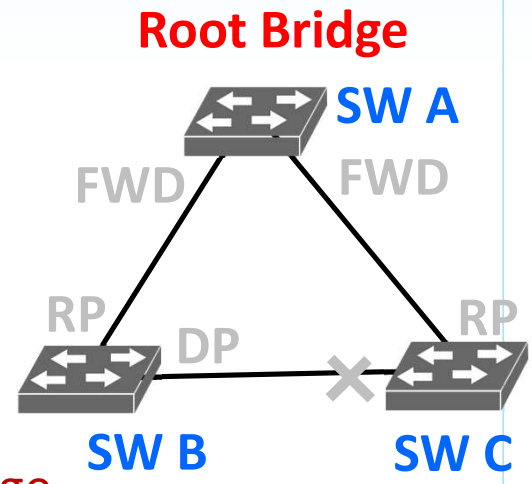


Root Bridge Election

- BPDU: Bridge Protocol Data Unit

  Contains information necessary to configure and maintain spanning tree topology.

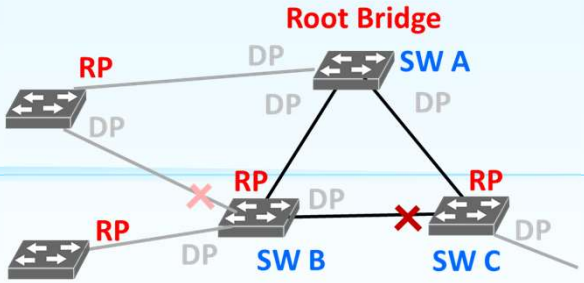
- Root Bridge Election
  - All switches participating in STP exchange BPDU frames,
    - BPDU carrying (sender) Bridge ID (BID), containing
      - Bridge Priority
      - Extended System ID (Optional)
      - MAC Address (of sending switch)
  - Switch with the lowest BID automatically becomes Root Bridge



- After Root Bridge elected

  BPDU originated from Root Bridge contains Root Bridge ID and Root Path Cost

Root Path Cost



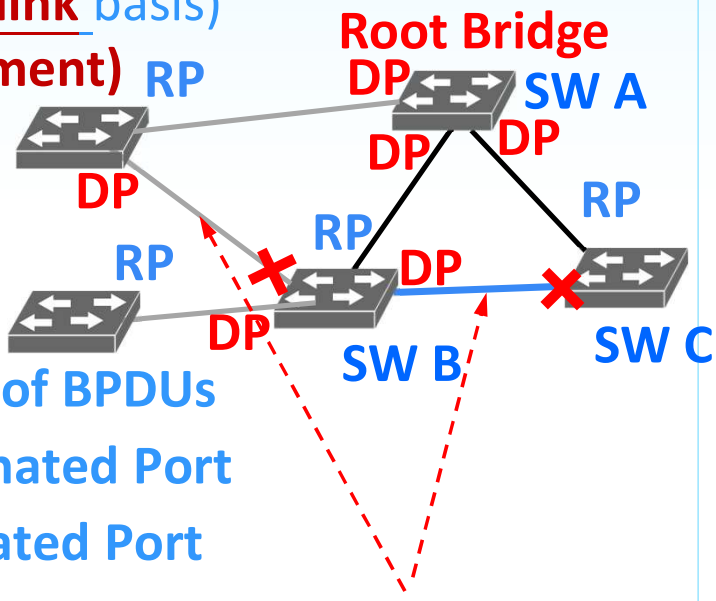
- ✓ BPDU contains Root Path Cost
- When a switch receives a BPDU,
  it calculates the **Internal Root Path Cost** of the ingress port
  - **Internal Root Path Cost = Root Path Cost** (received) **+ port cost** (ingress)
    - ➢ Sum of individual port costs **along the best path** from **the port** to **Root Bridge**
- **Default port cost** defined by the port speed

Link Speed	Cost (Revised IEEE Specification)	Cost (Previous IEEE Specification)
10 Gb/s	2	1
1 Gb/s	4	1
100 Mb/s	19	10
10 Mb/s	100	100

Designated Port and Non-Designated Port

- Recall: **Designated Ports** (Selected on a **per-segment/link** basis) port with the **best route** to Root Bridge on a **link (segment)**

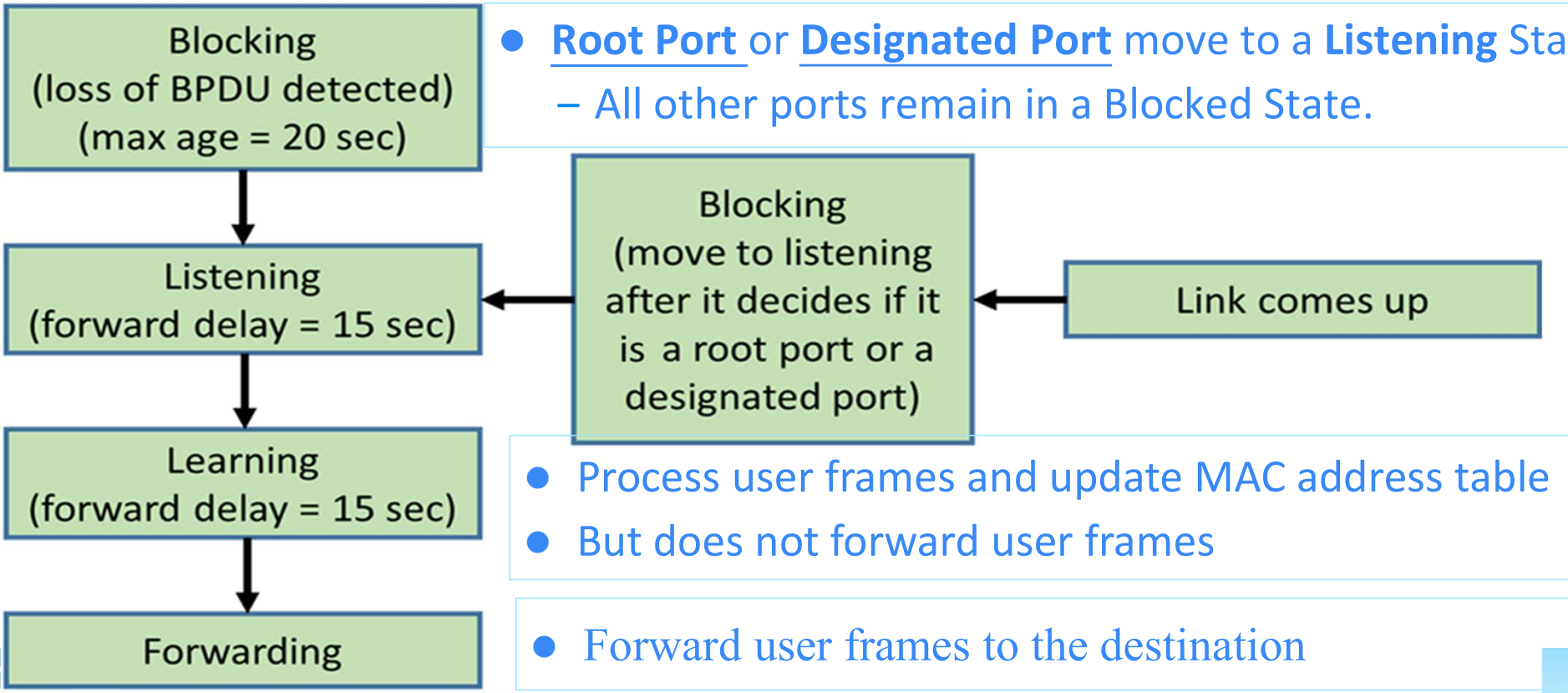
1) **All ports** on **Root Bridge** are **Designated Ports**
2) The **other end** of the **Root Port** is **Designated Port**.
3) For any other **segment** not including a Root Port,
   - Selected on a **Per-segment Basis** based on the **Cost of BPDUs**
     - Port with **lower Internal Root Path Cost** is **Designated Port**
     - The other end of **Designated Port** is **Non-Designated Port**



- **Root Port** and **Designated Port** in **Forwarding State**
- **Non-Designated Port** in **Blocking State**
  - Does not forward frames and discards frames received
  - Only listen to and process BPDUs (does not send BPDUs)
- Disabled Port: administratively shut down, (not in STP and not forwarding frames)

State Transition in STP

- When STP enabled, every port is in a **Blocking** State
  - Only listening to and processing BPDUs on its interfaces.
- Switch Ports in a blocking state at the election time



- **Root Port** or **Designated Port** move to a **Listening** State,
  - All other ports remain in a Blocked State.

- Process user frames and update MAC address table
- But does not forward user frames

- Forward user frames to the destination

STP Modes

- Several STP modes may exist in a switch [1] Cisco IOS use PVST+ as default mode

Protocol	Standard	Resources Needed	Convergence	Tree Calculation
STP	802.1D	Low	Slow	All VLANs
PVST+ [1]	Cisco	High	Slow	Per VLAN
RSTP	802.1w	Medium	Fast	All VLANs
Rapid PVST+	Cisco	Very high	Fast	Per VLAN
MSTP	802.1s	Medium or high	Fast	Per Instance

- PVSP: Per-VLAN Spanning Tree
- RSTP: Rapid STP
- Multiple Spanning Tree Protocol (MSTP)