NET 363
Introduction to LANs

Spanning Tree Protocol (STP)

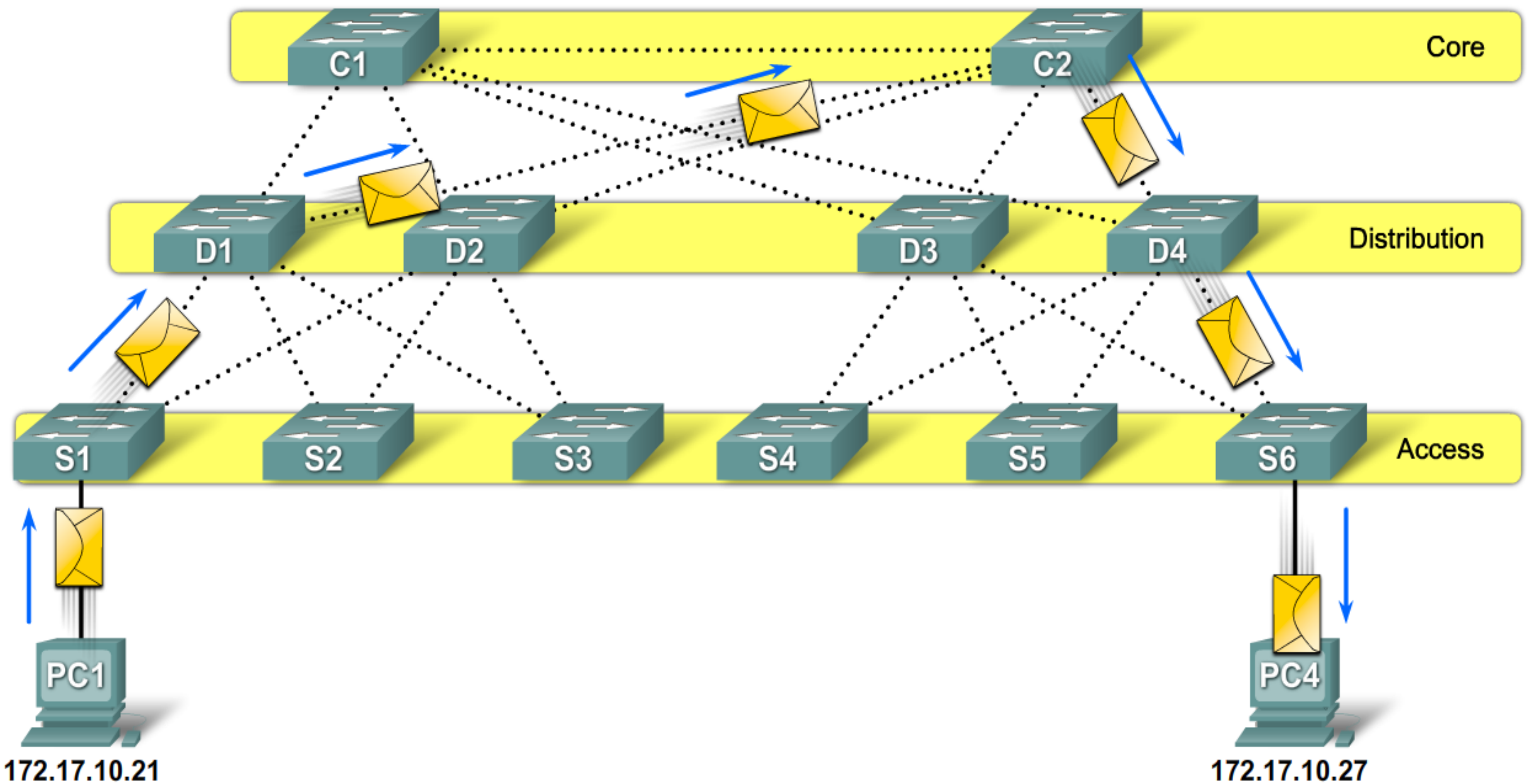
Greg Brewster
DePaul University

Providing Redundancy

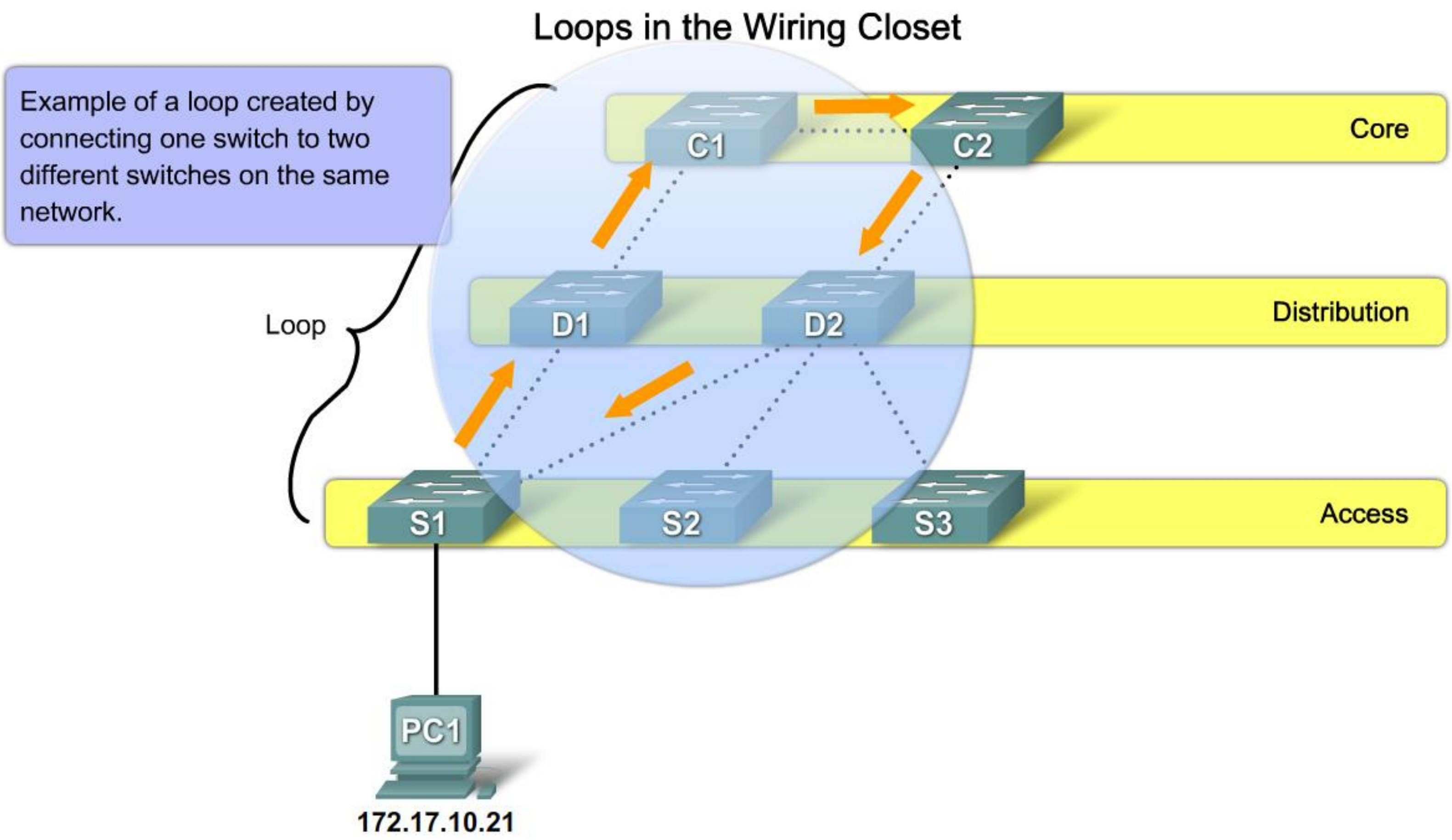
- In a LAN design, you want to ensure that traffic can always flow even if switches or links fail.
- This requires redundant paths
- This also creates the potential for loops
- A switched network <u>can't have active loops</u>
- A solution for basic parallel backup links is **Link Aggregation (<u>EtherChannel</u>)** (after midterm)
- A general solution for loops in switched environments is **<u>Spanning Tree Protocol</u>**
  - Provides loop-free network by blocking some ports.

The need for spanning trees:

Parallel switched paths for redundancy ➔ possible looping



A Loop



Loops in the Wiring Closet
Example of a loop created by connecting one switch to two different switches on the same network.
Loop
C1
C2
Core
D1
D2
Distribution
S1
S2
S3
Access
PC1
172.17.10.21

Purpose of Spanning Tree

Problems with Active Loops

Active Loops in a Switched LAN do provide redundancy and fault tolerance (which is good) but they may also cause these problems:

Considerations When Implementing Redundancy:

- **MAC database instability** - Instability in the content of the MAC address table results from copies of the same frame being received on different ports of the switch. Data forwarding can be impaired when the switch consumes the resources that are coping with instability in the MAC address table.
- **Broadcast storms** - Without some loop-avoidance process, each switch may flood broadcasts endlessly. This situation is commonly called a broadcast storm.
- **Multiple frame transmission** - Multiple copies of unicast frames may be delivered to destination stations. Many protocols expect to receive only a single copy of each transmission. Multiple copies of the same frame can cause unrecoverable errors.

Purpose of Spanning Tree

Possible Problem: Endless Packet Looping

- Without STP, Frames may continue to propagate between switches endlessly.
- There is no “Time-to-Live” in Ethernet to drop frames that are looping.
- It is possible for the MAC address table on a switch to constantly change with updates from looping broadcast frames, because same frame keeps arriving on different ports, resulting in MAC database instability.

Purpose of Spanning Tree

Possible Problem: Broadcast Storms

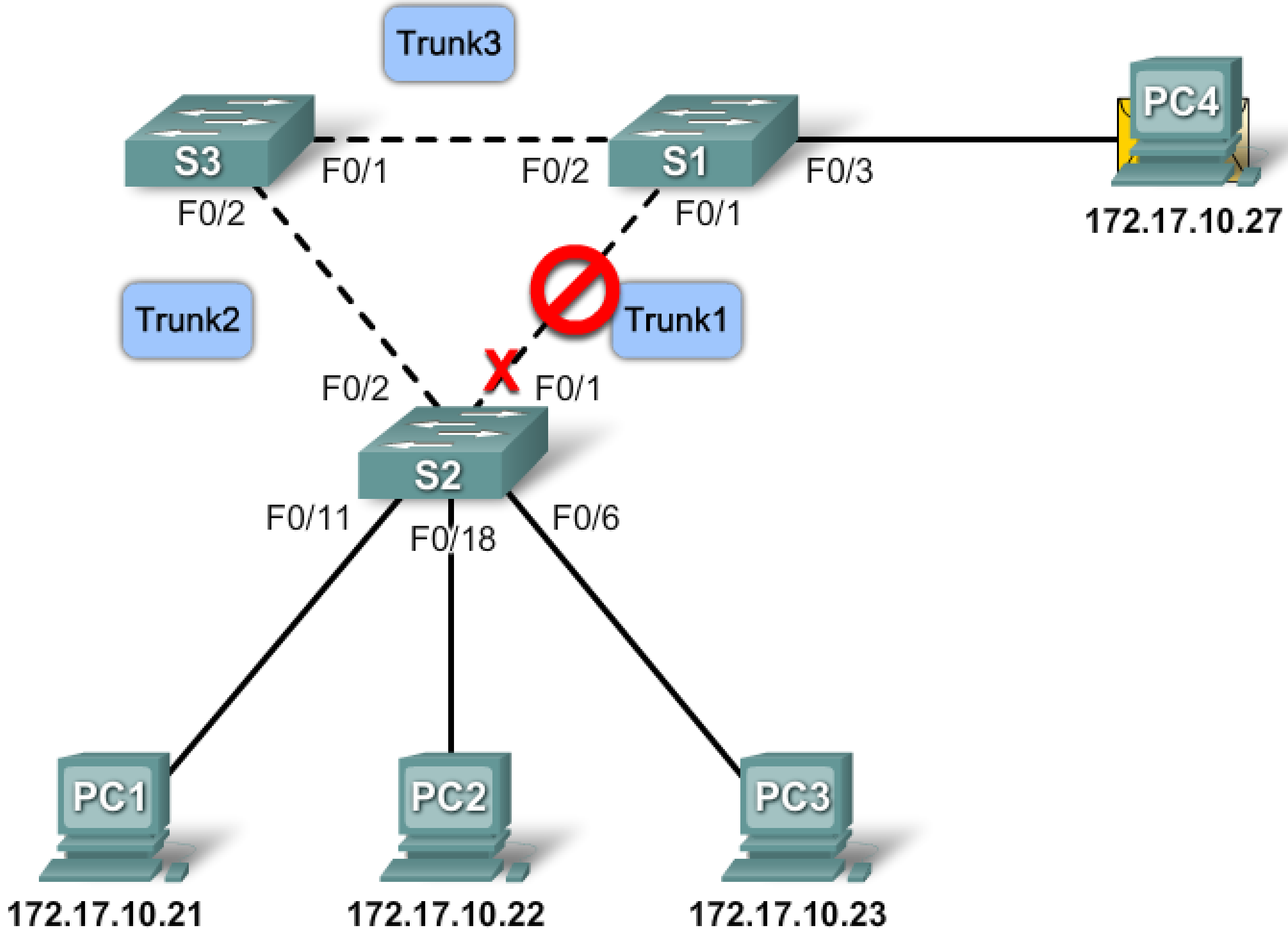
- A broadcast storm occurs when there are so many broadcast frames caught in a Layer 2 loop that all available bandwidth is consumed. This can cause a Denial of Service for users.
- A broadcast storm is inevitable on a looped network.
  - As more devices send broadcasts over the network, more traffic is caught within the loop; thus consuming more resources.
  - This eventually creates a broadcast storm that causes the network to fail.

STP Operation

Solution: Spanning Tree Protocol

- STP ensures that there is only one **active path** to each destination on the switched LAN by intentionally blocking redundant paths that could cause a loop.
- Switches implement STP by sending special Bridge Protocol Data Unit (**BPDU**) control packets out all ports periodically.
- STP causes switches to put each trunk port into either Forwarding State or Blocking State.
  - Data frames are sent in and out of Forwarding ports as usual.
  - No data frames are sent in or out of a Blocking ports.
  - BPDUs are sent in and out of all ports, regardless of state.
- If there are topology changes (a network cable or switch fails or comes back up), STP recalculates the paths and unblocks the necessary ports so there is always exactly one active path to each MAC destination.

Spanning Tree Protocol
S2 F0/1 is put into Blocking State



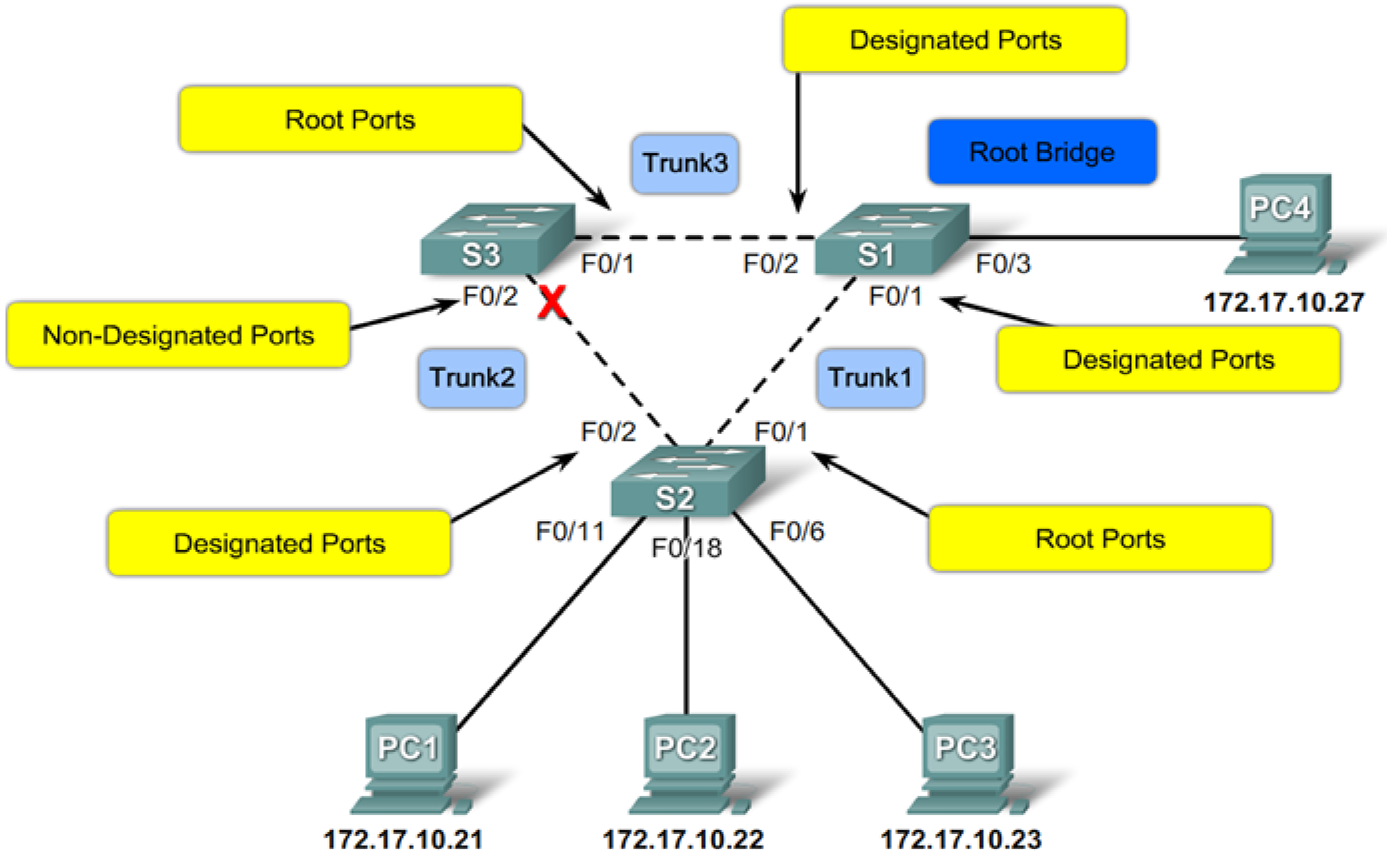
How does STP work?

- One switch is selected as the **root switch**
- All other switches determine the least-cost path from themselves to the root.
- Switch Ports on the least-cost path to the root switch are put in **Forwarding State**.
  - These ports actively send/receive data packets
  - They are called Root Ports or Designated Ports
- Switch Ports that are not on the least-cost path to the root are put into **Blocking State**.
  - They are called Alternate Ports or Blocked Ports
  - Switches do not forward packets out Alternate ports. Switches drop data packets arriving on Alternate ports.

STP Port Roles

- For every trunk port on every switch, STP assigns one of 3 possible Port Roles:
- Root Port
  - There is one Root Port on each non-root switch, which is the trunk port on the least-cost path towards the Root switch.
- Designated Port (DP)
  - For each inter-switch link, there is exactly one Designated Port, which provides the lowest-cost path to the Root off that link.
    - If both paths are equal cost, then highest Port Priority is DP
    - If both Port Priorities are equal, then lowest switch BID is DP
  - On a root switch, all ports are Designated Ports
- Backup Port or Alternate Port
  - Any port that is not Root or Designated is Alternate Port
  - These ports are put into Blocked State.

Spanning-Tree Algorithm Port Roles



STP Details (next few slides)

- How do you determine root switch?
  - Answer: Each switch has a Bridge ID (BID) value. Switch with the **lowest BID** will be the Root.
- How do switches determine the lowest-cost path to the root switch?
  - Answer: Each link between switches has a **Link Cost**. Adding up Cost of each link on a path to Root Switch gives the total **Path Cost** to the Root Switch.
- How do switches determine Port Roles?
  - Answer: Port Roles values assigned based on which Port connects to lowest-cost path to the Root Switch.

Choosing the Root Switch

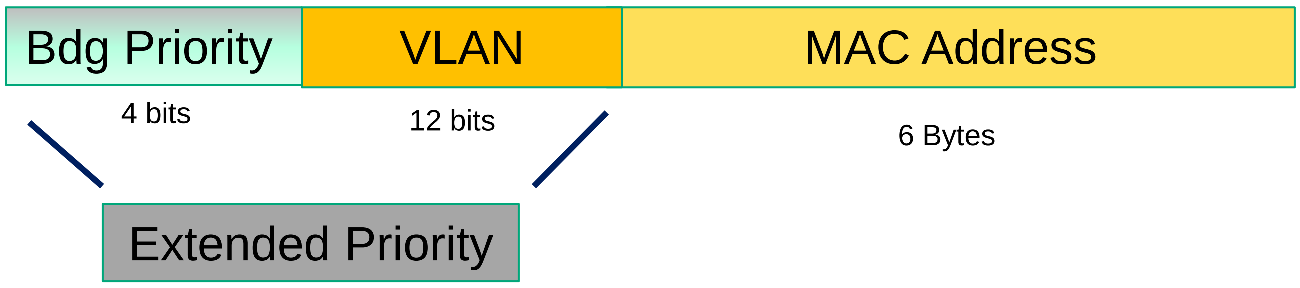
- How do you determine root switch?
  - Each switch has a Bridge ID (BID) value.
  - Initially, each switch assumes that it is the Root.
  - Each Switch regularly sends BPDUs out all ports containing
    - Its own BID value
    - The Root Switch BID (the lowest BID seen so far)
  - Eventually, every switch learns who has the lowest BID and all switches agree which switch is the Root.

The BPDU Message Format

Bytes	Field
2	Protocol ID
1	Version
1	Message type
1	Flags
8	Root ID
4	Cost of path
8	Bridge ID
2	Port ID
2	Message age
2	Max age
2	Hello time
2	Forward delay

BID (Bridge ID) value

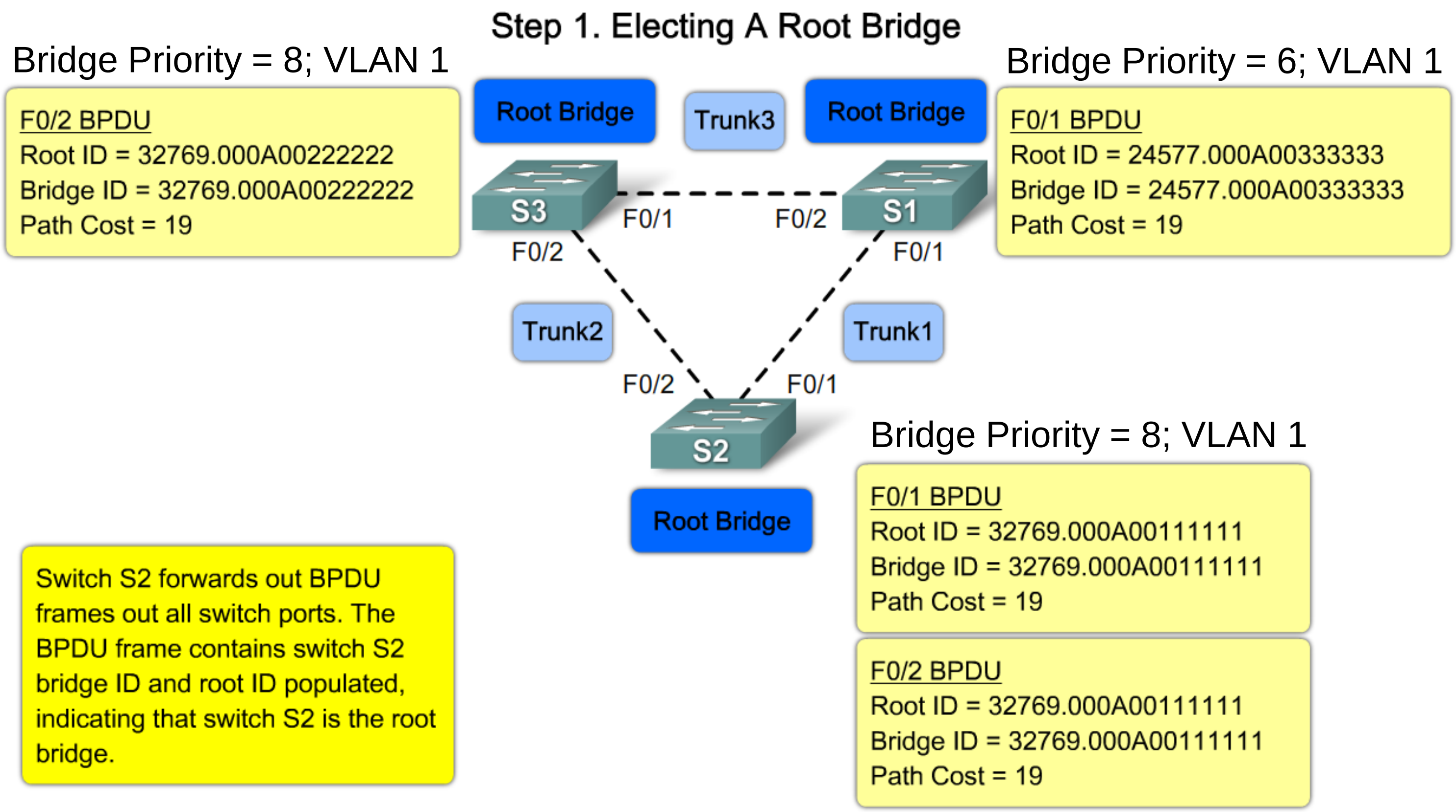
- BID value is 8 bytes
  - Extended Priority (2 bytes)
    - 4-bit Bridge Priority
    - 12-bit VLAN Number
  - MAC Address (6 bytes)



BID Priority

- Extended Priority (Extended System ID)
  - Bridge Priority is 4 bits (values 0-15)
  - Extended Priority = (Priority * 4096) + VLAN ID
  - Default Bridge Priority is 8
  - For VLAN 1, default Extended Priority = **32,769**
- When switches have different Priority values, Lowest Extended Priority determines who is Root.
- When all switches on a VLAN have same Priority values, the lowest MAC address determines the Root.

Choosing the Root Switch



S1, S2, S3 – each starts off assuming it is the root switch. Then, as they see BPDUs from neighbors, they learn that S1 has lowest BID, so eventually all agree that S1 is the Root switch.

Ways to set Priority
(and specify which Switch will be Root)

- Global config mode
  - **spanning-tree vlan *vlan-id* root primary**
    - This will set Extended Priority to (6 * 4096)+<vlan-id> or lower to ensure it is less than anything else seen on the same VLAN.
  - **spanning-tree vlan *vlan-id* root secondary**
    - This will set the Extended Priority to (7 * 4096)+<vlan-id>.
  - **Spanning-tree vlan *vlan-id* priority *value***
    - This will set the Extended Priority value directly to **value**

Details: Lowest-Cost Root Path

- How do switches determine the lowest-cost path to the root switch?
  - Answer: Each trunk link has a Link Cost.
  - By default, link cost is determined by link speed
    - Link Cost can also be set by admin in interface configuration mode using **spanning-tree cost <x>**
  - From each switch, Path Cost to the Root is calculated as the sum of Link Costs on path from this switch to Root.
  - Each Switch puts its Root Path Cost in BPDUs sent out to neighbor Switches.
  - Each switch calculates its own Root Path Cost through each neighbor as (Link Cost to Neighbor) + (Neighbor's reported Root Path Cost).

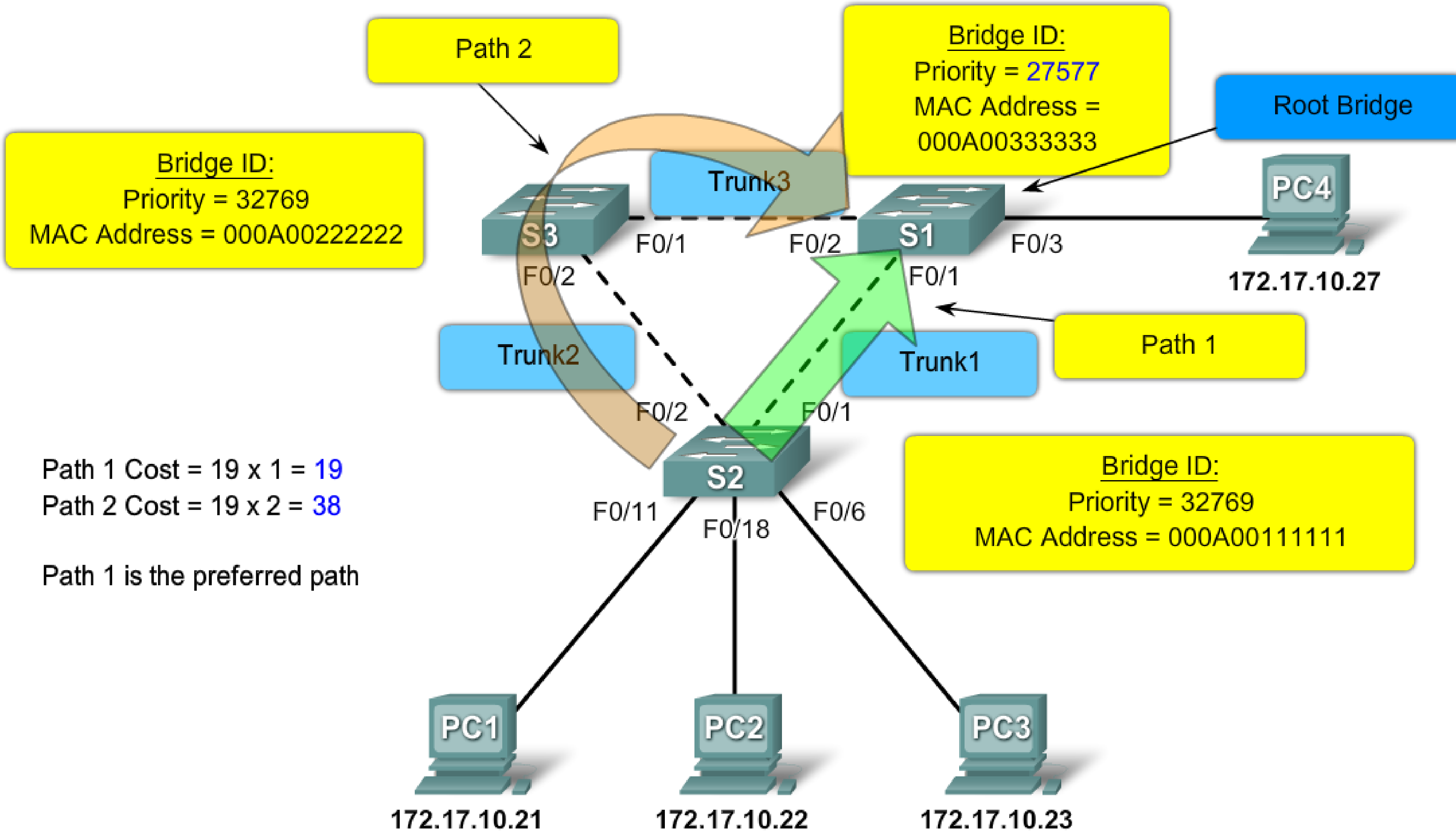
Default Link Cost Values

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

Admin can Manually Change the Link Cost (for example, to 25)

S1(config)# **int Fa0/1**
S1(config-if)# **spanning-tree cost 25**

Path Costs



Path 1 Cost = 19 x 1 = 19
Path 2 Cost = 19 x 2 = 38

Path 1 is the preferred path

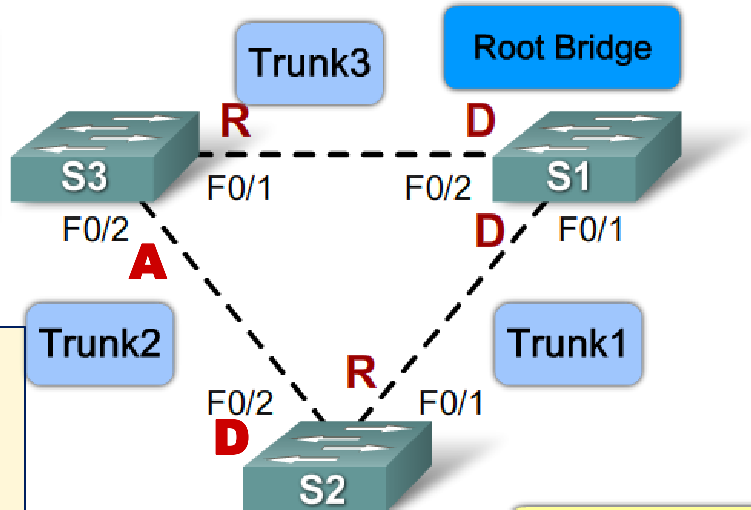
Determining Port Roles

- On Root Switch: all trunk ports are **Designated Ports**.
- For every other switch:
  - Determine Root Port: Trunk port on lowest-cost path to Root Switch is set to **Root Port** role.
    - Exactly one Root Port on each Switch (except Root Switch).
    - If 2 port path costs equal, lowest neighbor BID is Root port.
- For every inter-switch trunk (Port at each end), mark one end as Designated:
  - Port on lowest-cost path to Root is **Designated Port**
    - If Root Path Costs for both Ports are equal, then compare the 2 Switch BIDs – lower BID Port is **Designated Port**.
- Any remaining Ports with no assigned Role – mark as **Alternate Ports** (blocked)

Electing Designated Ports

Step 3. Electing Designated Ports and Non-Designated Ports

<u>F0/2 BPDU</u>
Root ID = 24577.000A00333333
Bridge ID = 32769.000A00222222
Path Cost = 38



For the <u>Trunk2 Ports</u>, the Root-Path Costs of both ends are <u>equal</u> (19). So S2 F0/2 is marked **Designated** because S2 BID < S3 BID. Then S3 F0/2 is **Alternate**.

<u>F0/2 BPDU</u>
Root ID = 24577.000A00333333
Bridge ID = 32769.000A00111111
Path Cost = 38

Switch S1 configures both of its switch ports in the designated role since it is the root bridge.

Port Roles on Point-to-Point Trunks

- For point-to-point trunks between switches, there are only two possibilities:
  - One end is Root Port, the other end is Designated Port
  - One end is Designated Port, the other end is Alternate Port
- Things get more complicated if there are Hubs connected between the Switches, but this is not required for NET 363. We will always assume the trunks between switches are Point-to-Point.

Topology Change

- Any switch sensing a topology change (switch or link up/down) will:
  - Set Topology Change (TC) Flag bit in outgoing BPDUs
  - Starts a TC While timer (2x hello timer)
  - Flush its MAC table (non-edge ports)
- Bridges receiving a BPDU with TC bit set will
  - Clear MAC Table addresses on all ports except one that received the BPDU with TC bit
  - Starts TC While timer and sends BPDUs with TC set on all Designated and Root Ports
- When TC While Timer expires:
  - All Switch ports proceed through 4 recovery states:
    - **Blocking -> Listening -> Learning -> Forwarding**

Port Processes in STP

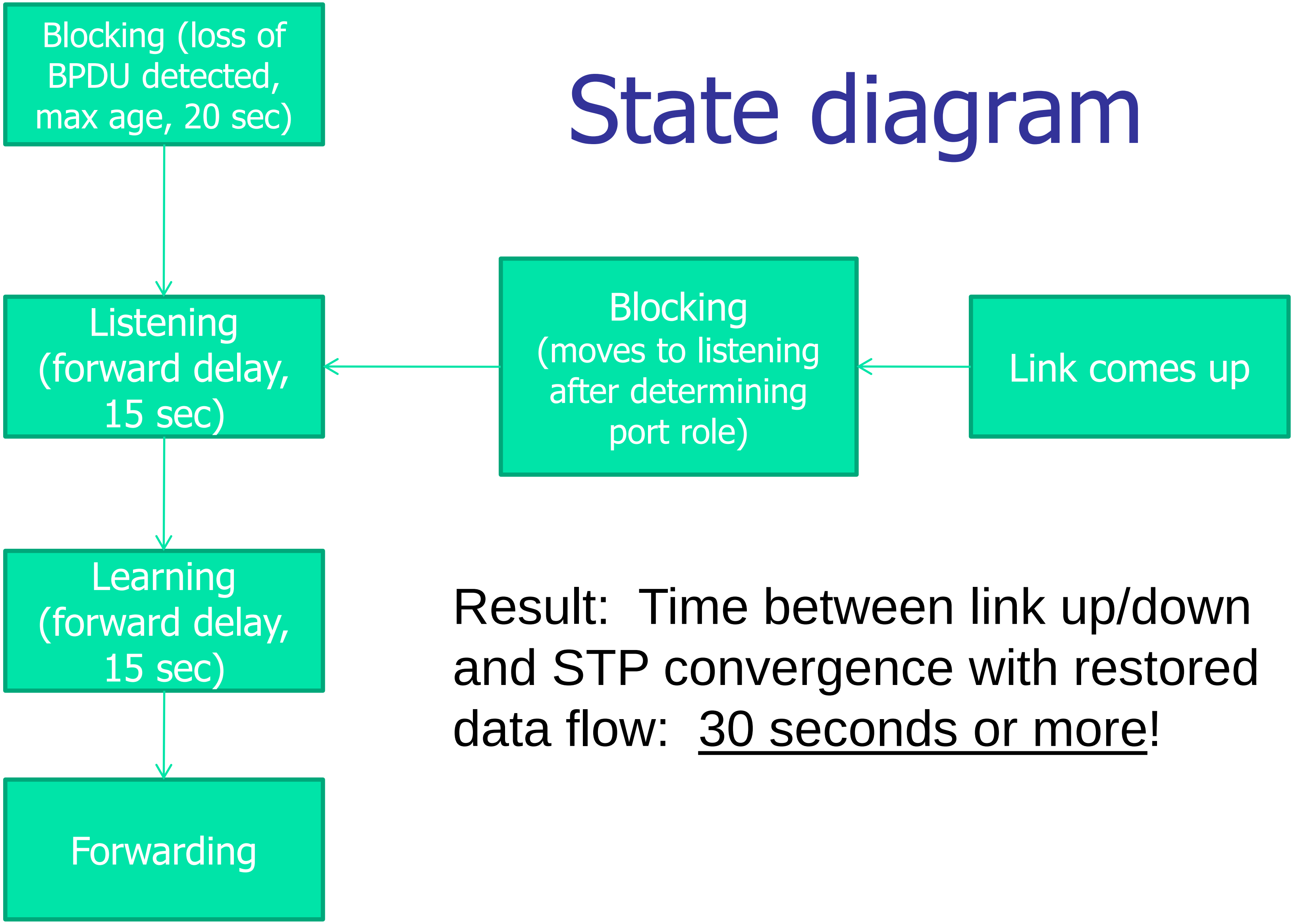
Processes	Blocking	Listening	Learning	Forwarding	Disable
Receives and process BPDUs	✓	✓ [1]	✓	✓	✗
Forward data frames received on interface	✗	✗	✗	✓	✗
Forward data frames switched from another interface	✗	✗	✗	✓	✗
Learn MAC addresses	✗	✗	✓	✓	✗

[1]Return to blocking if not lowest cost path to root bridge

Port States – STP

- **Blocking** – Discards data frames received on interface, does not learn addresses, receives / processes BPDUs, does not send BPDUs
- **Listening** – Discards data frames recvd on interface, does not learn addresses, receives / processes and transmits BPDUs
  - Return to Blocking state after Listening if not on least-cost path to Root
- **Learning** – Discards data frames received on interface or switching fabric, learns MAC addresses, recvs and transmits BPDUs
- **Forwarding** – Recvs/Forwards data frames recvd on interface, learns MAC addresses, recvs and transmits BPDUs

State diagram



Result: Time between link up/down and STP convergence with restored data flow: <u>30 seconds or more</u>!

STP Timers

- Hello time: time between BPDU sending
  - 1-10 sec, default 2 sec
- Forward delay: time spend in listening and learning states (ea)
  - 4-30 sec, default 15 sec
- Max age: controls max time bpdu config data is stored
  - 6-40 sec, default 20

Modifying Timers

- In general, Don't mess with timers!
- If necessary, you can adjust the network diameter setting (which adjusts timers accordingly).
- Example: To set the diameter on the root bridge to 5 switch hops:

S1(config)#spanning-tree vlan 1 root primary diameter 5

- Default diameter is 7