Slide 1

CS540
Computer Networks II

Sandy Wang
chwang_98@yahoo.com

"Data and Computer Communications", 10/e, by William Stallings, Chapter 11 "Local Area Network Overview".

In this chapter, we look at the underlying technology and protocol architecture of LANs. Chapters 12 and 13 are devoted to a discussion of specific LAN systems.

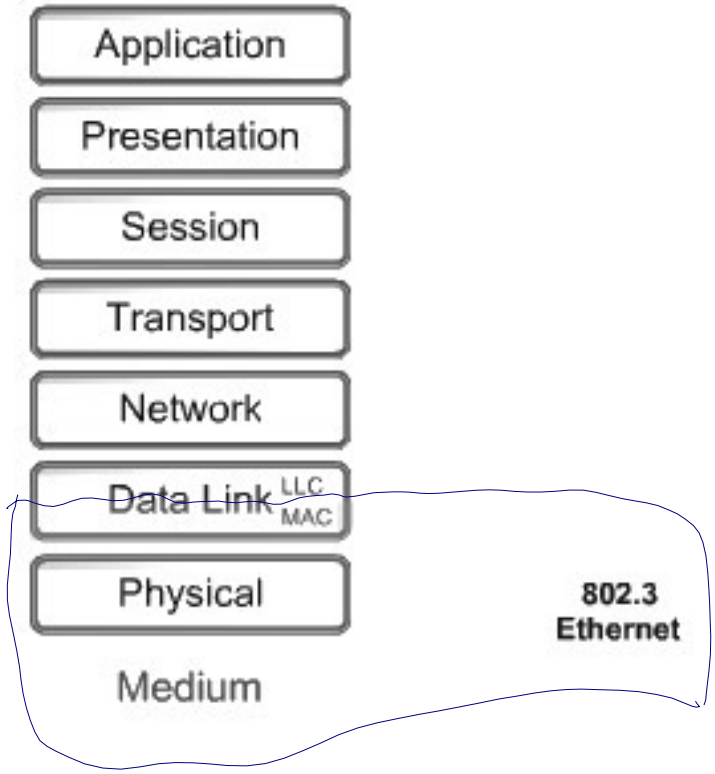
Slide 3

Topics

1. Overview
2. LAN Switching
3. IPv4
4. IPv6
5. Routing Protocols -- RIP, RIPng, OSPF
6. Routing Protocols -- ISIS, BGP
7. MPLS
8. Midterm Exam
9. Transport Layer -- TCP/UDP
10. Congestion Control & Quality of Service (QoS)
11. Access Control List (ACL)
12. Application Layer Protocols
13. Application Layer Protocols continue
14. Others – Multicast, SDN
15. Final Exam

Slide 4

Ethernet and the OSI Model



fyi

Ethernet and the OSI Model

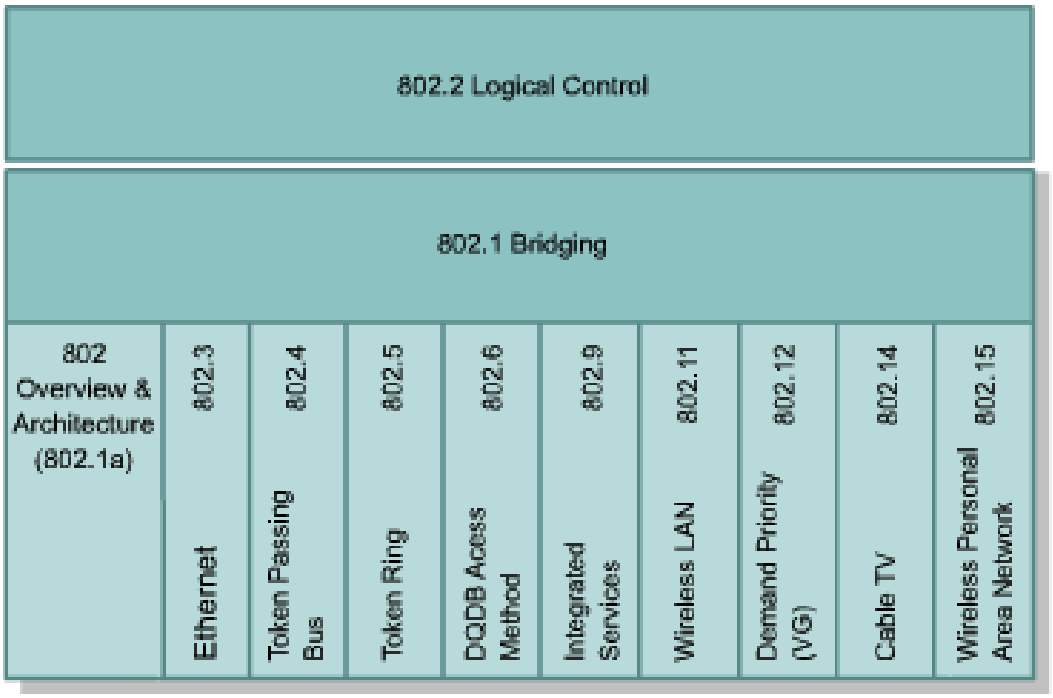
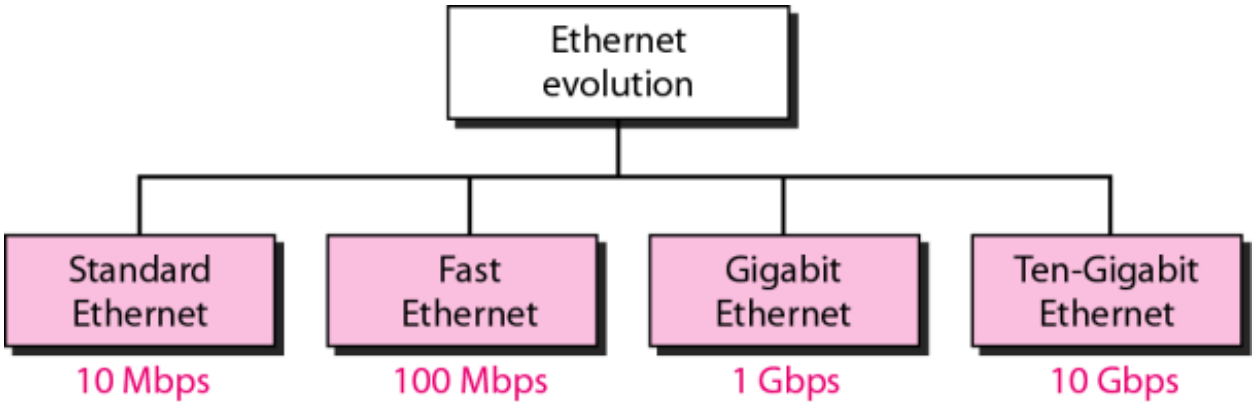
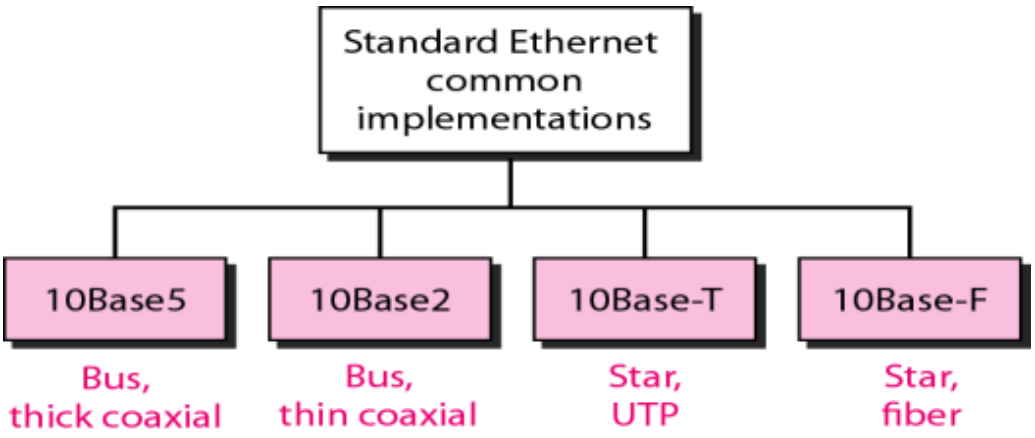


Figure 13.3 *Ethernet evolution through four generations*

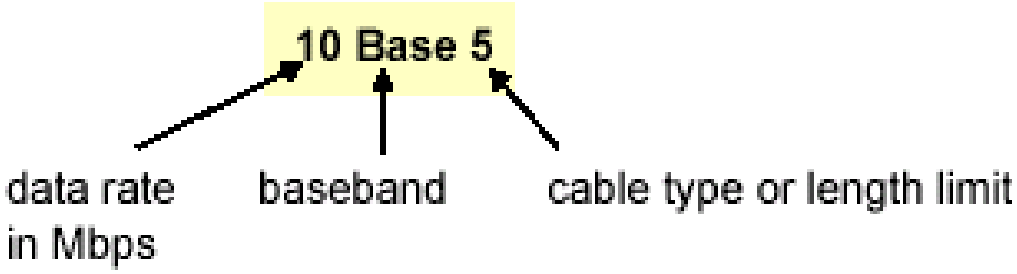


fyi

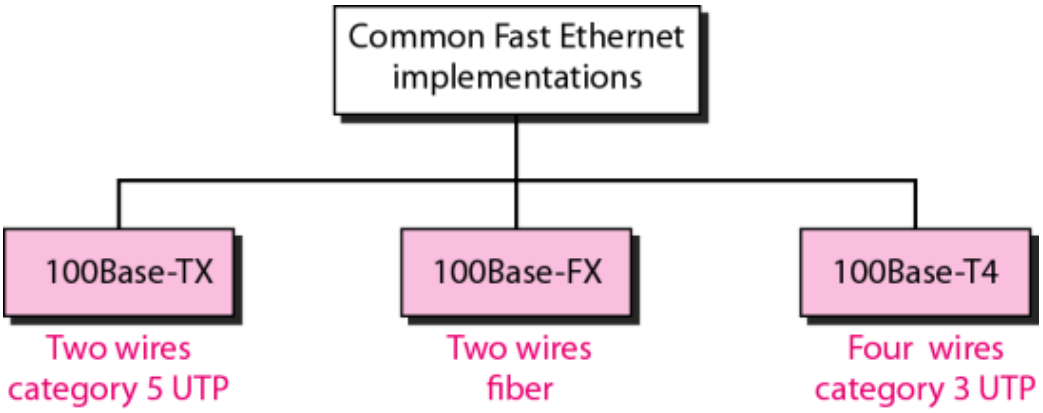
Categories of traditional Ethernet



- <data rate><Signaling method><Max segment length or cable type>

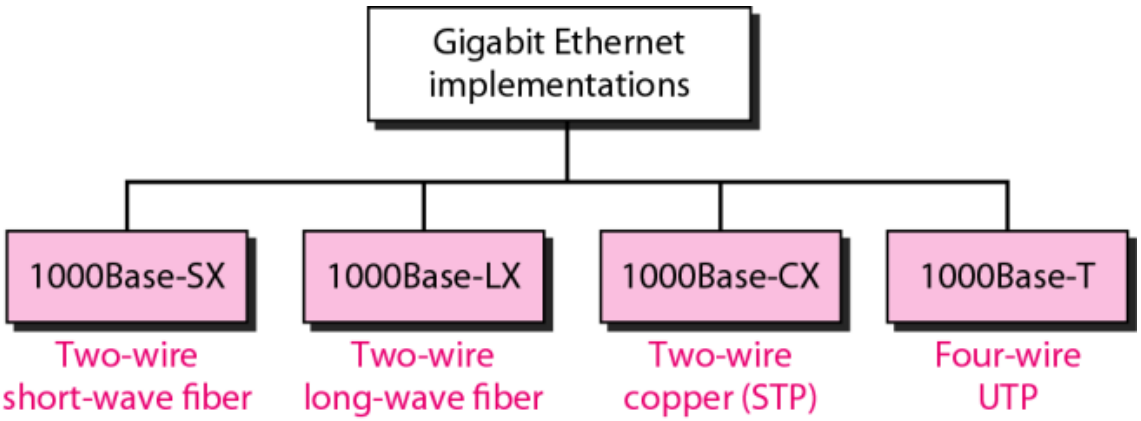


Fast Ethernet implementations



fyi

Gigabit Ethernet implementations



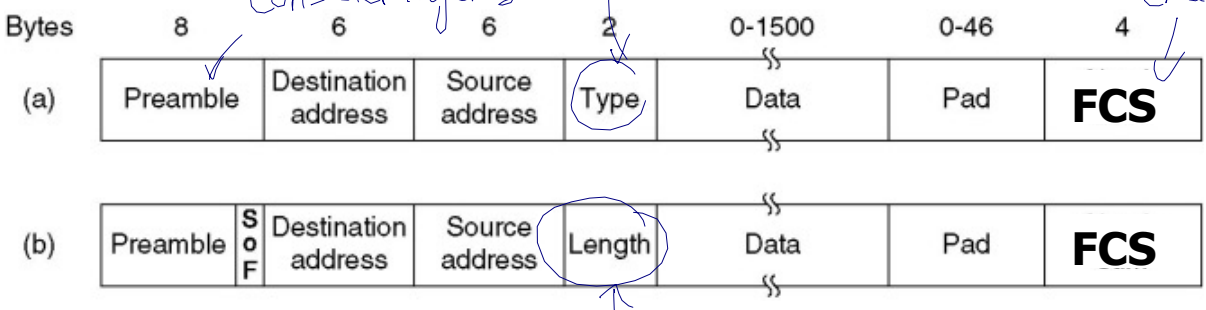
lack full duplex switch mode

Ethernet Frame Format

what is inside → IPv6 / IPv4 ...

consider layer 1 not consider in L2

checksum.



Frame formats. (a) DIX Ethernet , (b) IEEE 802.3.

how many bytes

L2 frame.

14 bytes header

4 bytes trailer

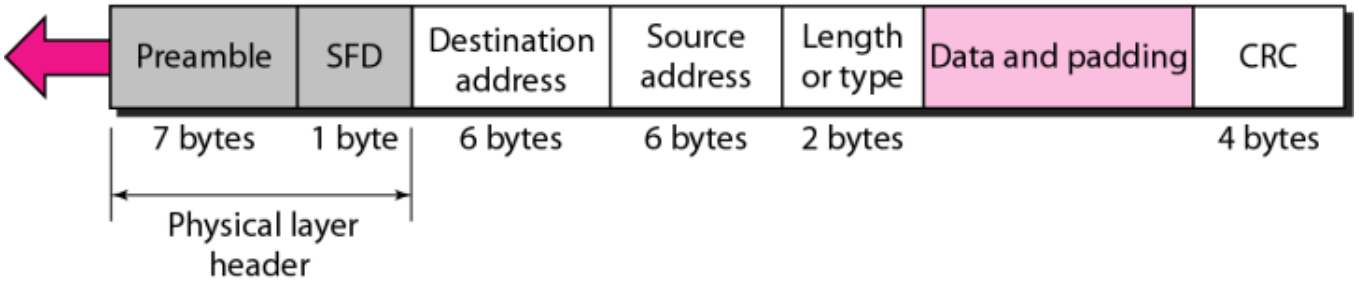
switch mode → store & fwd

→ cut thru

802.3 MAC frame

Preamble: 56 bits of alternating 1s and 0s.

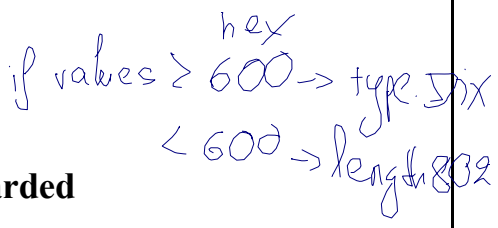
SFD: Start frame delimiter, flag (10101011)



- Length/Type – Length if less than 0x0600, otherwise protocol type
- If less than 46 bytes data, padding is required

Ethernet Frame

- Preamble:
  - 8 bytes with pattern 10101010 used to synchronize receiver, sender clock rates.
  - In IEEE 802.3, eighth byte is start of frame (10101011)
- Addresses: 6 bytes (explained latter)
- Type (DIX)
  - Indicates the type of the **Network layer protocol** being carried in the **payload (data)** field, **mostly IP** but others may be supported such as IP (**0800**), Novell IPX (**8137**) and AppleTalk (**809B**), ARP (**0806**))
  - Allow **multiple network layer** protocols to be supported on a single machine (multiplexing)
  - Its value starts at **0600h (=1536 in decimal)**
- Length (IEEE 802.3): number of bytes in the **data field**.
  - Maximum 1500 bytes **(= 05DCh)**
- CRC: checked at receiver, if error is detected, the frame is **discarded**
  - CRC-32
- Data: carries data encapsulated from the upper-layer protocols
- Pad: Zeros are added to the data field to make the **minimum data length = 46 bytes**



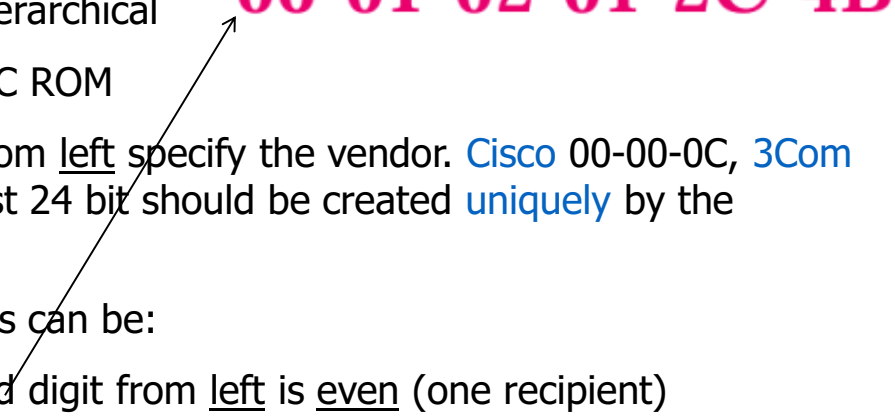
Ethernet Provides Unreliable, connectionless Service

does not know relationship between frames.

- **Ethernet data** link layer protocol provides **connectionless service** to the network layer
  - No handshaking between sending and receiving adapter.
- **Ethernet protocol provides *Unreliable* service to the network layer :**
  - Receiving adapter doesn't send ACK or NAK to sending adapter
  - This means stream of datagrams passed to network layer can have gaps (missing data)
    - Gaps will be filled if application is using reliable transport layer protocol
      - Otherwise, application will see the gaps

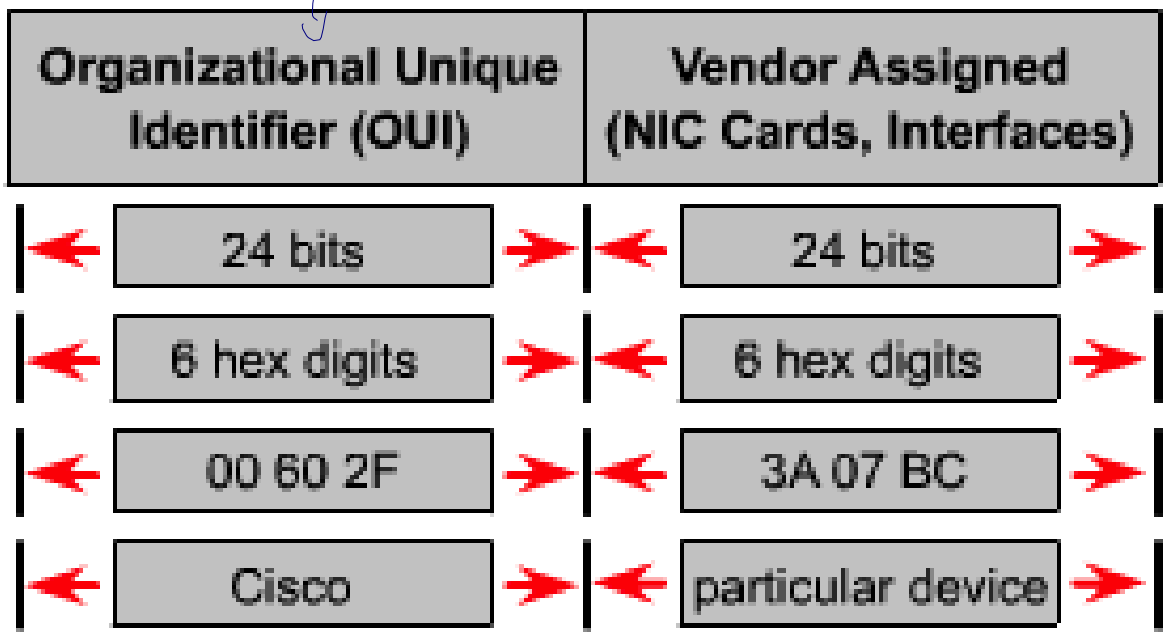
Ethernet address

- Six bytes = 48 bits
- Flat address not hierarchical
- Burned into the NIC ROM
- First three bytes from left specify the vendor. Cisco 00-00-0C, 3Com 02-60-8C and the last 24 bit should be created uniquely by the company
- Destination Address can be:
  - Unicast: second digit from left is even (one recipient)
  - Multicast: Second digit from left is odd (group of stations to receive the frame – conferencing applications)
  - Broadcast (ALL ones) (all stations receive the frame)
- Source address is always Unicast



Naming

vendor id



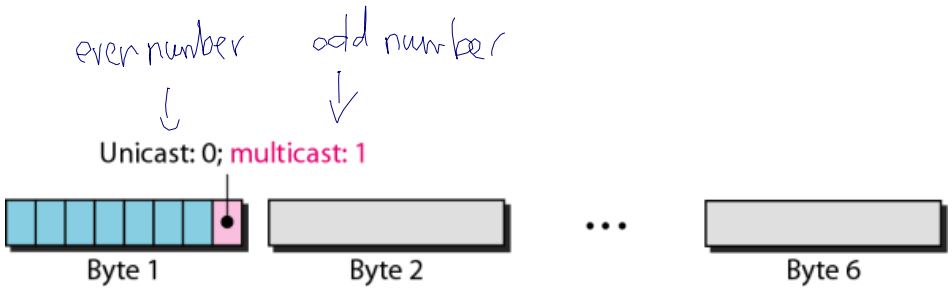
Note

The least significant bit of the first byte defines the type of address. If the bit is 0, the address is unicast; otherwise, it is multicast.

Note

The broadcast destination address is a special case of the multicast address in which all bits are 1s.

Figure 13.7 *Unicast and multicast addresses*



Example 13.1

Define the type of the following destination addresses:

a. 4A:30:10:21:10:1A ***b. 47:20:1B:2E:08:EE***

c. FF:FF:FF:FF:FF:FF

Solution

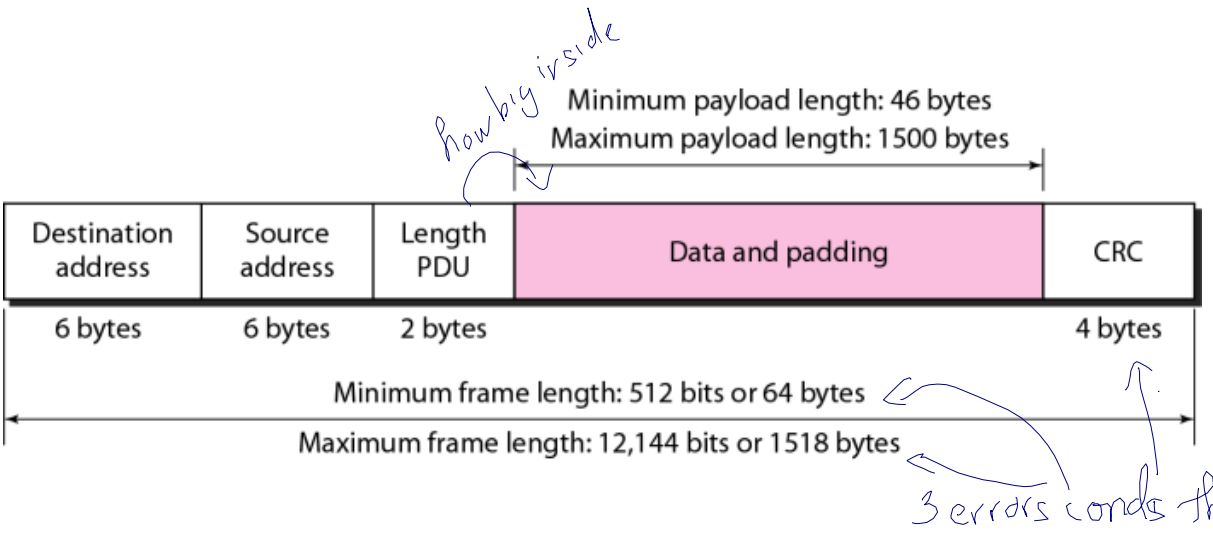
To find the type of the address, we need to look at the second hexadecimal digit from the left. If it is even, the address is unicast. If it is odd, the address is multicast. If all digits are F's, the address is broadcast. Therefore, we have the following:

a. This is a unicast address because A in binary is 1010.

b. This is a multicast address because 7 in binary is 0111.

c. This is a broadcast address because all digits are F's.

Figure 13.5 *Minimum and maximum lengths*



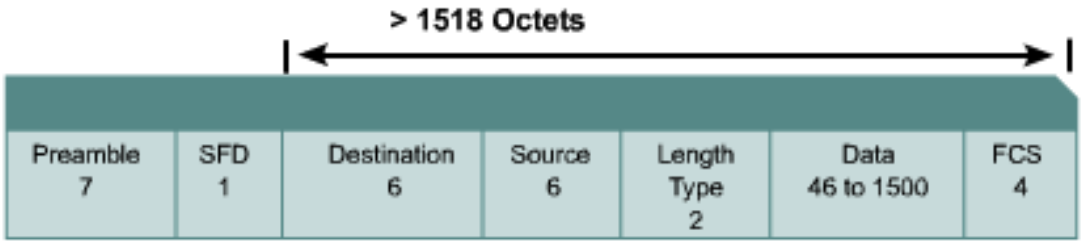
Note

Frame length:

Minimum: 64 bytes (512 bits)

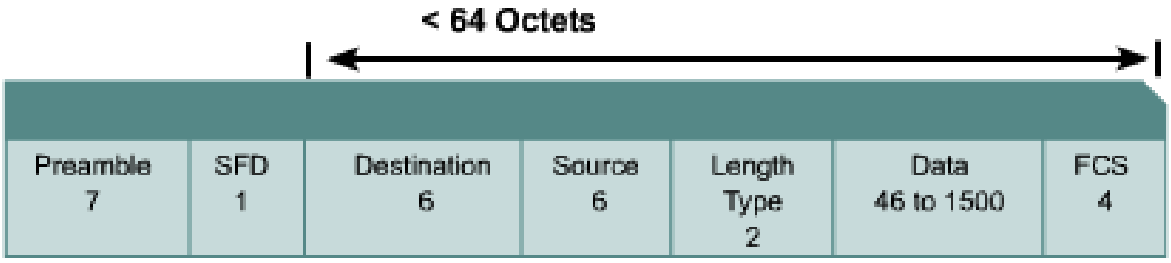
Maximum: 1518 bytes (12,144 bits)

Ethernet Errors



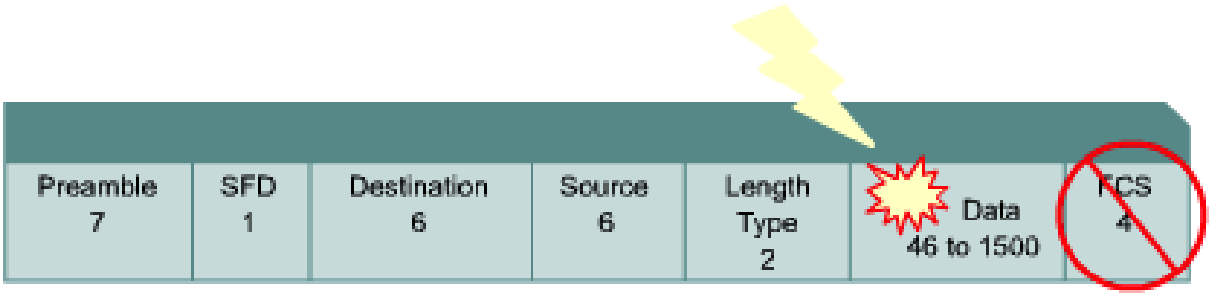
Jabber and Long Frames are both in excess of the maximum frame size. Jabber is significantly larger.

Ethernet Errors

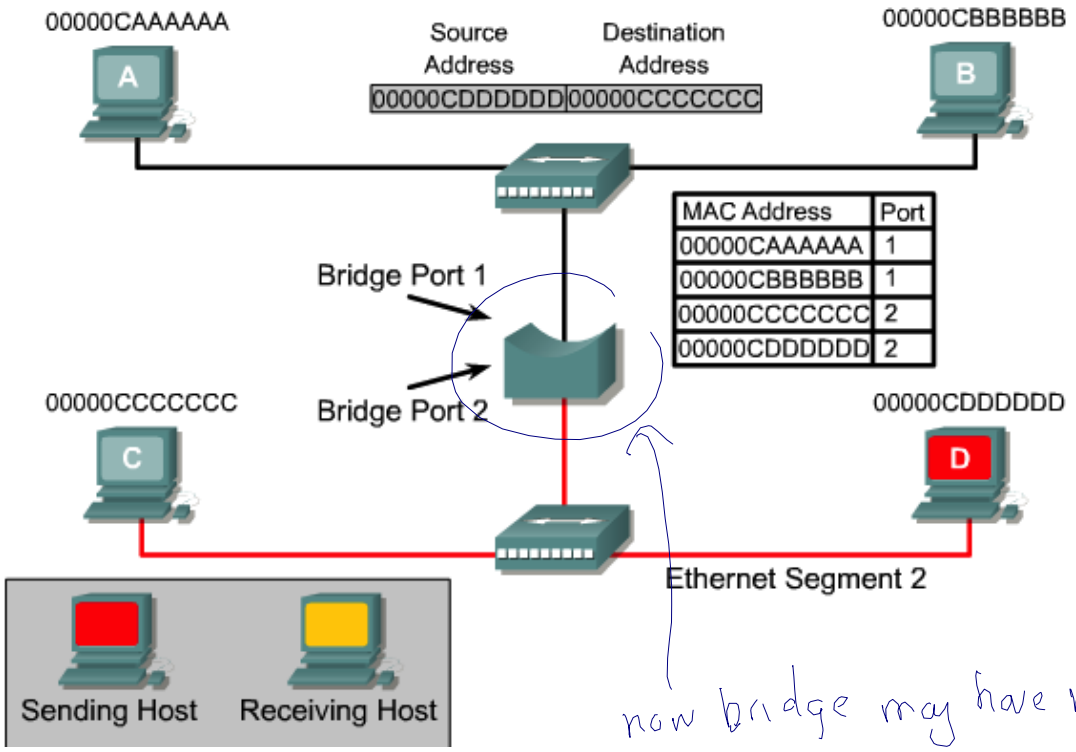


Short frames are properly formed in all but one aspect and have valid FCS checksums, but are less than the minimum frame size (64 octets).

FCS Errors


Preamble
7
SFD
1
Destination
6
Source
6
Length
Type
2
Data
46 to 1500
FCS
4

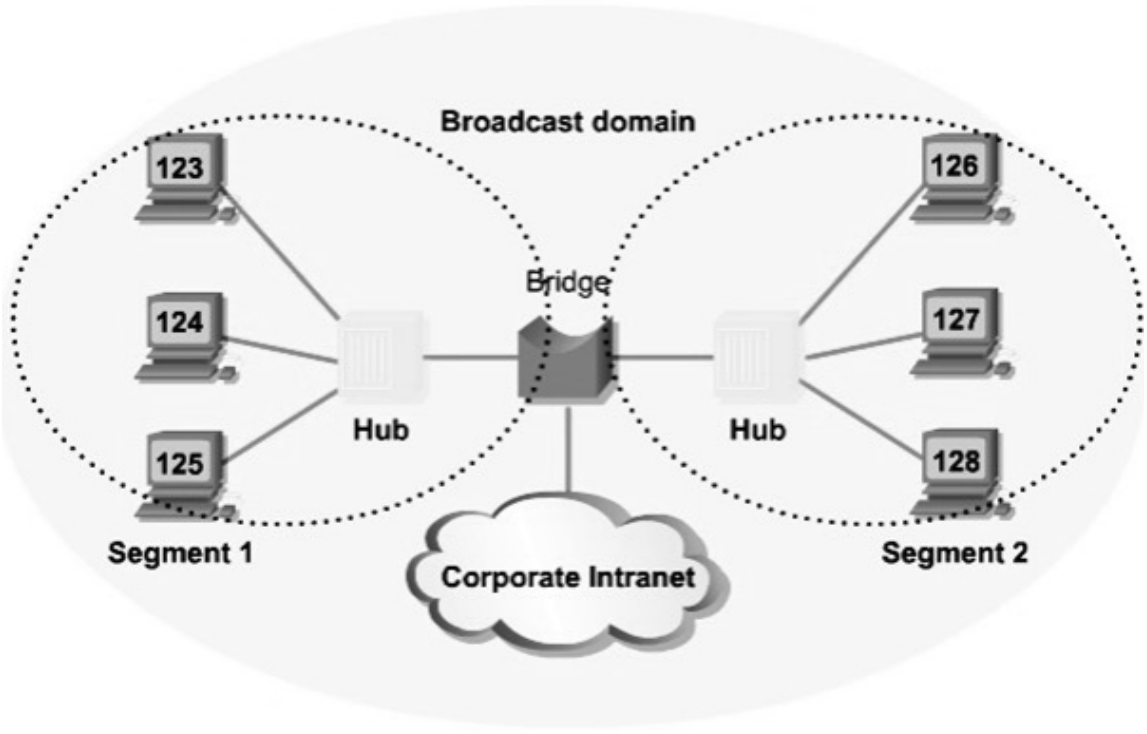
Layer 2 Bridging



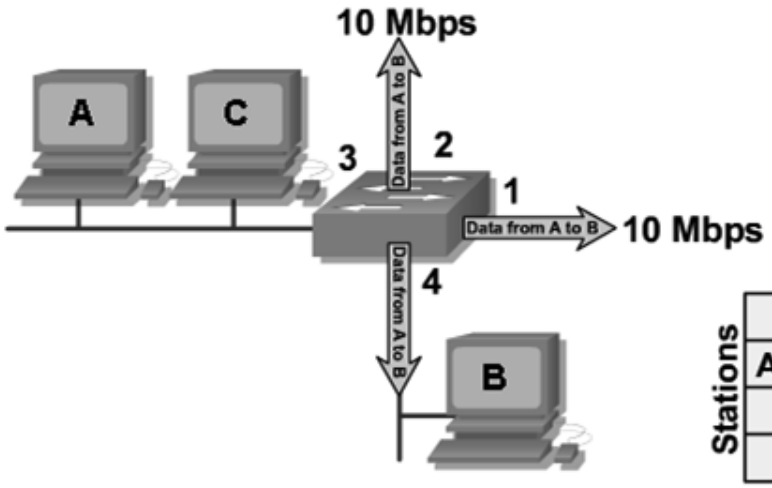
now bridge may have more than 2 ports

Bridges

bridge separates devices into different domains

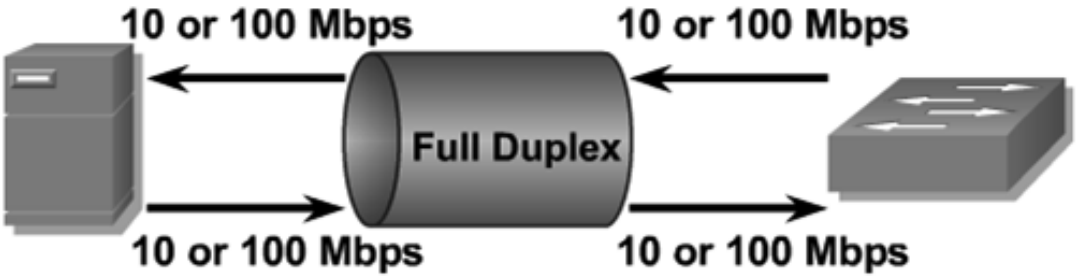


Switch Operation ← "multiport" bridge



- Forward packets based on MAC address in forwarding table
- Operates at OSI Layer 2
- Learns a station's location by examining source address

Full Duplex

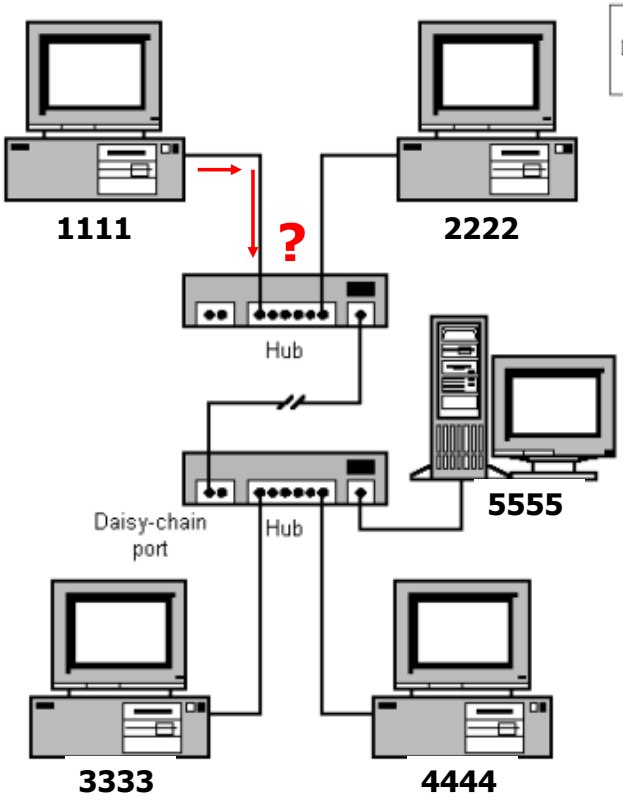


- Doubles bandwidth between nodes
- Collision-free transmission
- Two 10- or 100- Mbps data paths

Switch Modes

- Store and Forward - A switch receives the entire frame before sending it out the destination port.
- Cut-Through - A switch starts to transfer the frame as soon as the destination MAC address is received.

Sending and receiving Ethernet frames via a hub



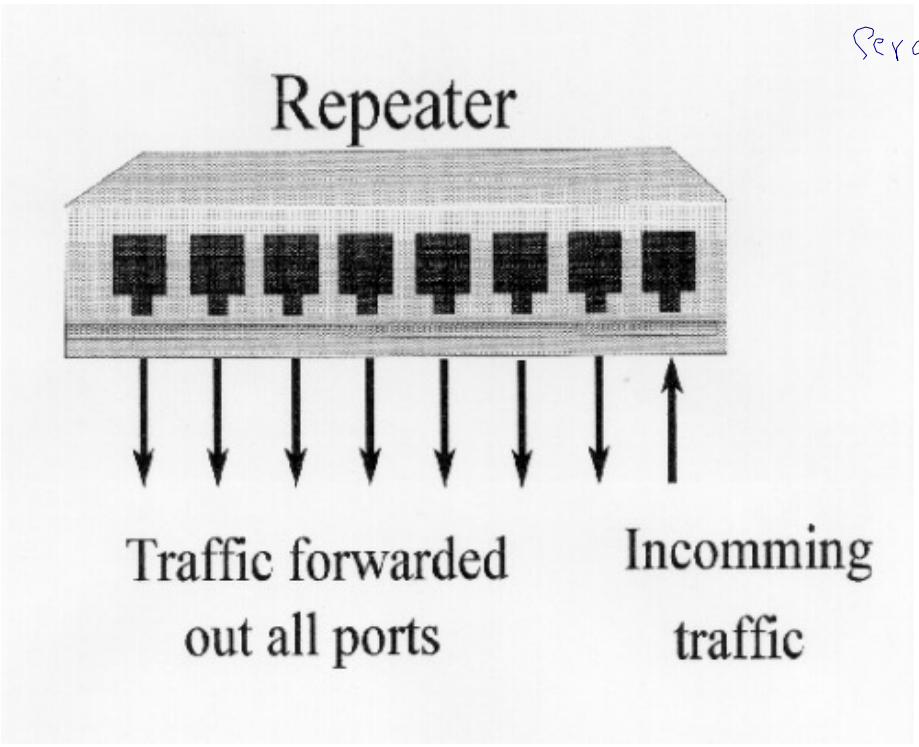
Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

3333 1111

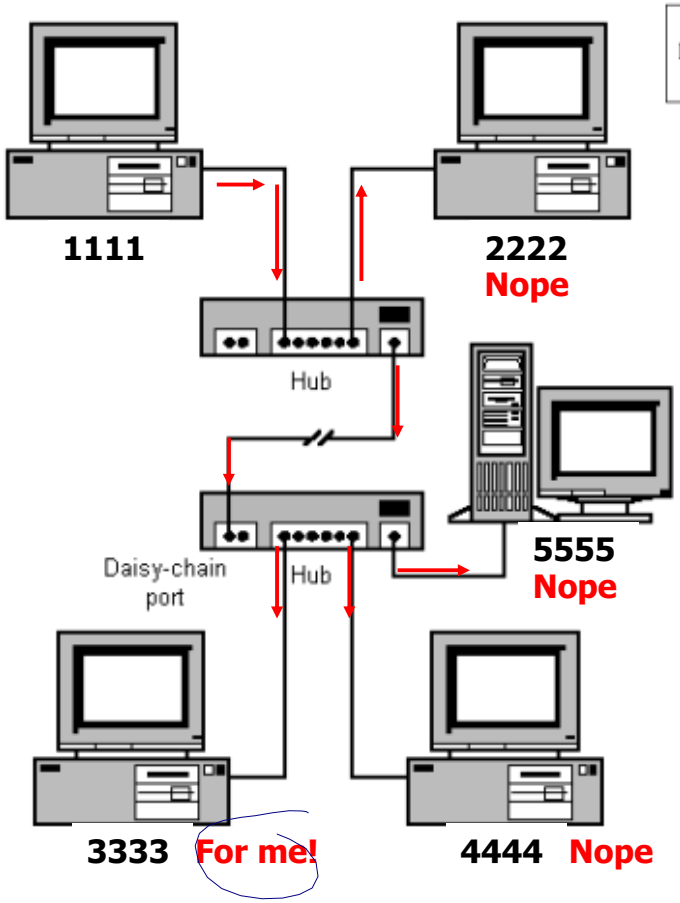
- So, what does a hub do when it receives information?
- Remember, a hub is nothing more than a multiport repeater.

- Sending and receiving Ethernet frames via a hub

incoming traffic
Send out to all ports



Sending and receiving Ethernet frames via a hub

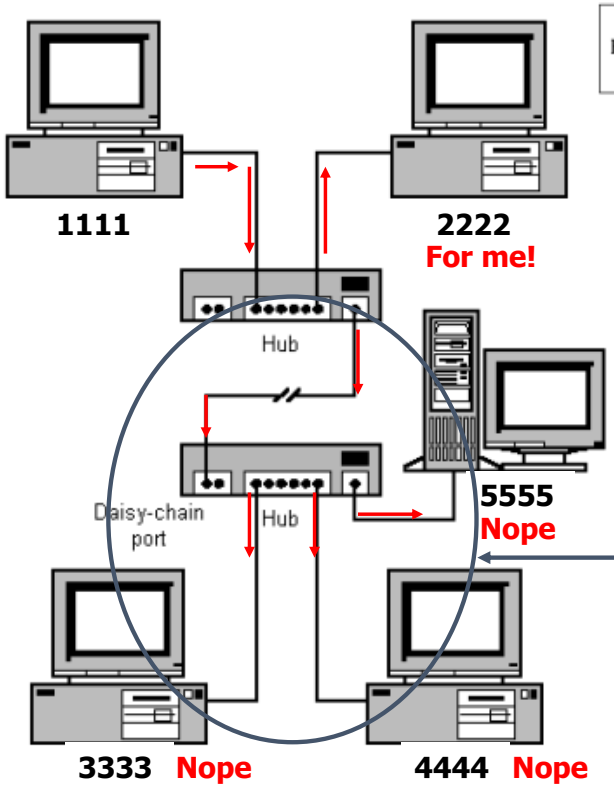


Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

3333 1111

- The hub will **flood** it out all ports except for the incoming port.
- Hub is a layer 1 device.
- A hub does NOT look at layer 2 addresses, so it is fast in transmitting data.
- Disadvantage with hubs: A hub or series of hubs is a single **collision domain**.
- A collision will occur if any two or more devices transmit at the same time within the collision domain.
- More on this later.

Sending and receiving Ethernet frames via a hub



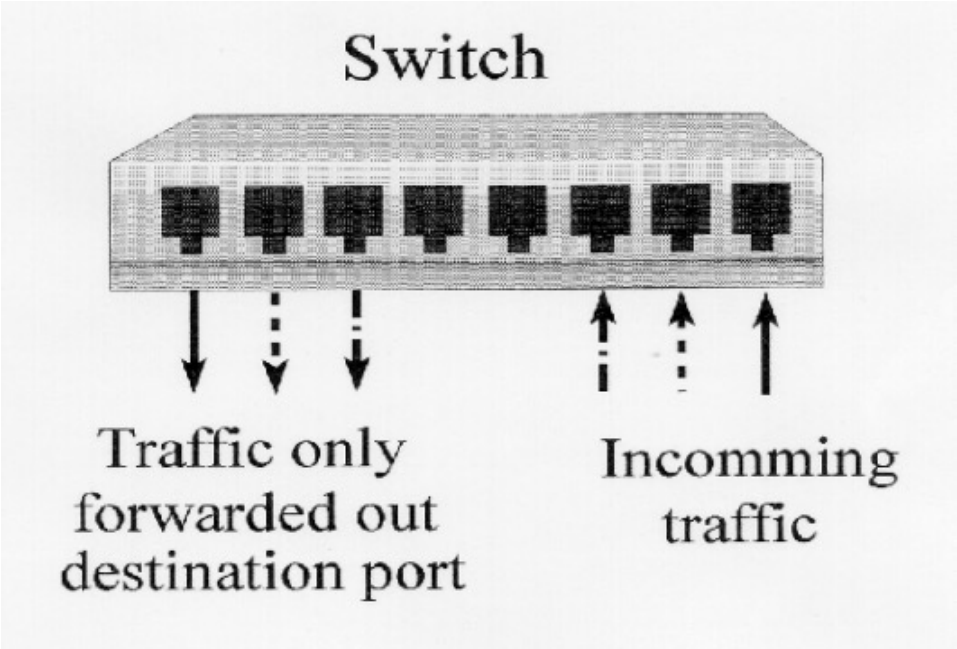
Preamble	Destination Address	Source Address	Type	Data	Pad	CRC
	2222	1111				

- Another disadvantage with hubs is that is take up unnecessary bandwidth on other links.

Wasted bandwidth

go back 25

Sending and receiving Ethernet frames via a switch



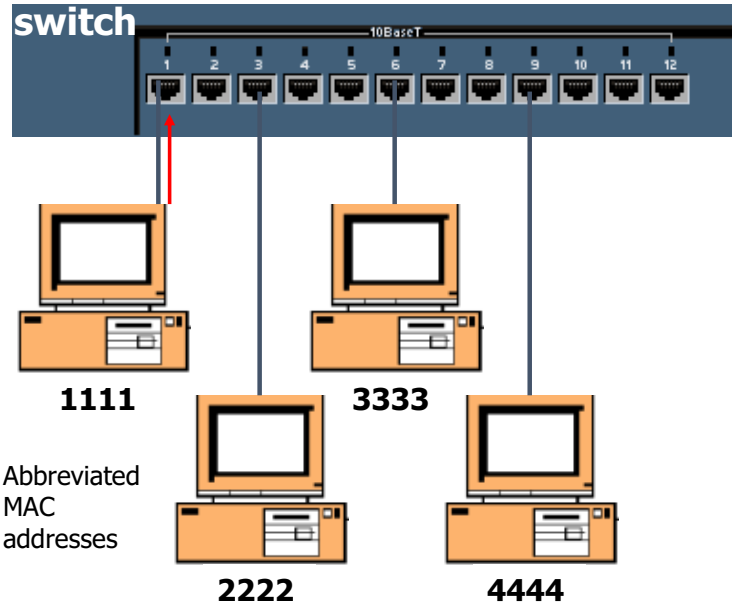
Sending and receiving Ethernet frames via a switch

Source Address Table

Port	Source MAC Add.	Port	Source MAC Add.

Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

3333 1111



- Switches are also known as **learning bridges** or **learning switches**.
- A switch has a source address table in cache (RAM) where it stores source MAC address after it learns about them.
- A switch receives an Ethernet frame it searches the source address table for the Destination MAC address.
- If it finds a match, it **filters** the frame by only sending it out that port.
- If there is not a match if **floods** it out all ports.

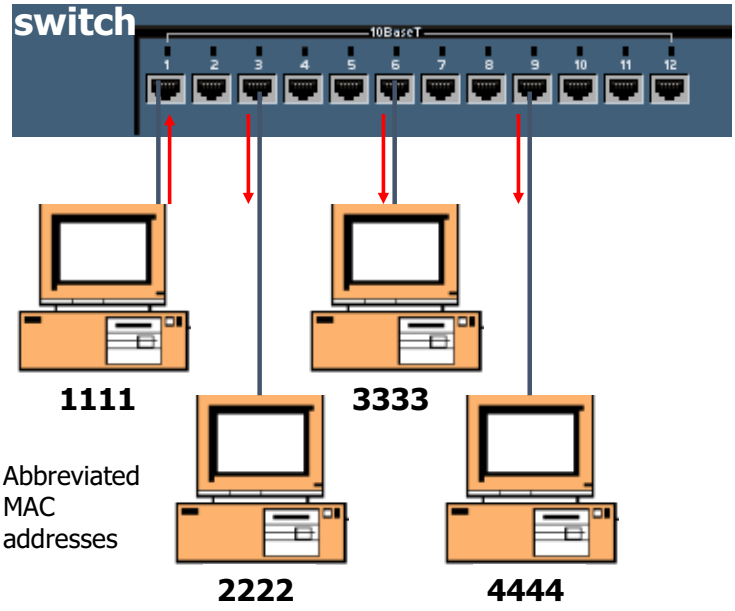
Every switch will do these steps:

1. Source learning :
2. Dest. lookup.
   hit → send out.
3. Not hit → flood. (not back to the src).

No Destination Address in table, Flood

Source Address Table

Port	Source MAC Add.	Port	Source MAC Add.
1	1111		



Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

3333 1111

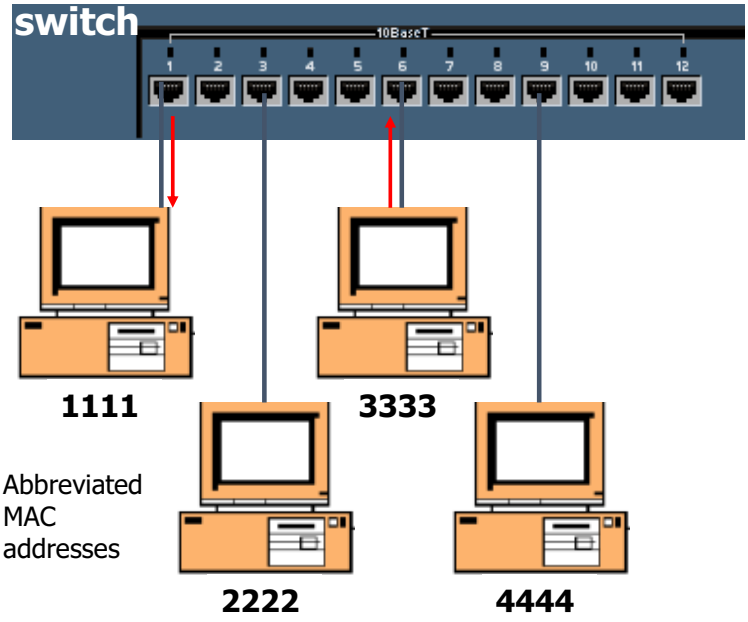
- How does it learn source MAC addresses?
- First, the switch will see if the SA (1111) is in it's table.
- If it is, it resets the timer (more in a moment).
- If it is NOT in the table it adds it, with the port number.
- Next, in our scenario, the switch will **flood** the frame out all other ports, because the DA is not in the source address table.

Destination Address in table, Filter

3 sends to 1

Source Address Table

Port	Source MAC Add.	Port	Source MAC Add.
1	1111	6	3333



Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

1111 3333

- Most communications involve some sort of client-server relationship or exchange of information. (You will understand this more as you learn about TCP/IP.)
- Now 3333 sends data back to 1111.
- The switch sees if it has the SA stored.
- It does NOT so it adds it. (This will help next time 1111 sends to 3333.) ✓
- Next, it checks the DA and in our case it can **filter** the frame, by sending it only out port 1.

Destination Address in table, Filter

each entry has a time out (usually 4 hrs)

Source Address Table

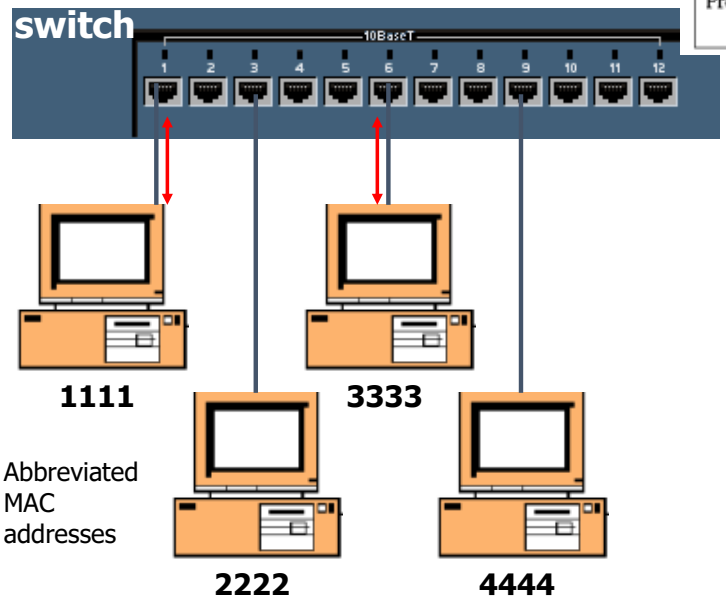
Port	Source MAC Add.	Port	Source MAC Add.
1	1111	6	3333

Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

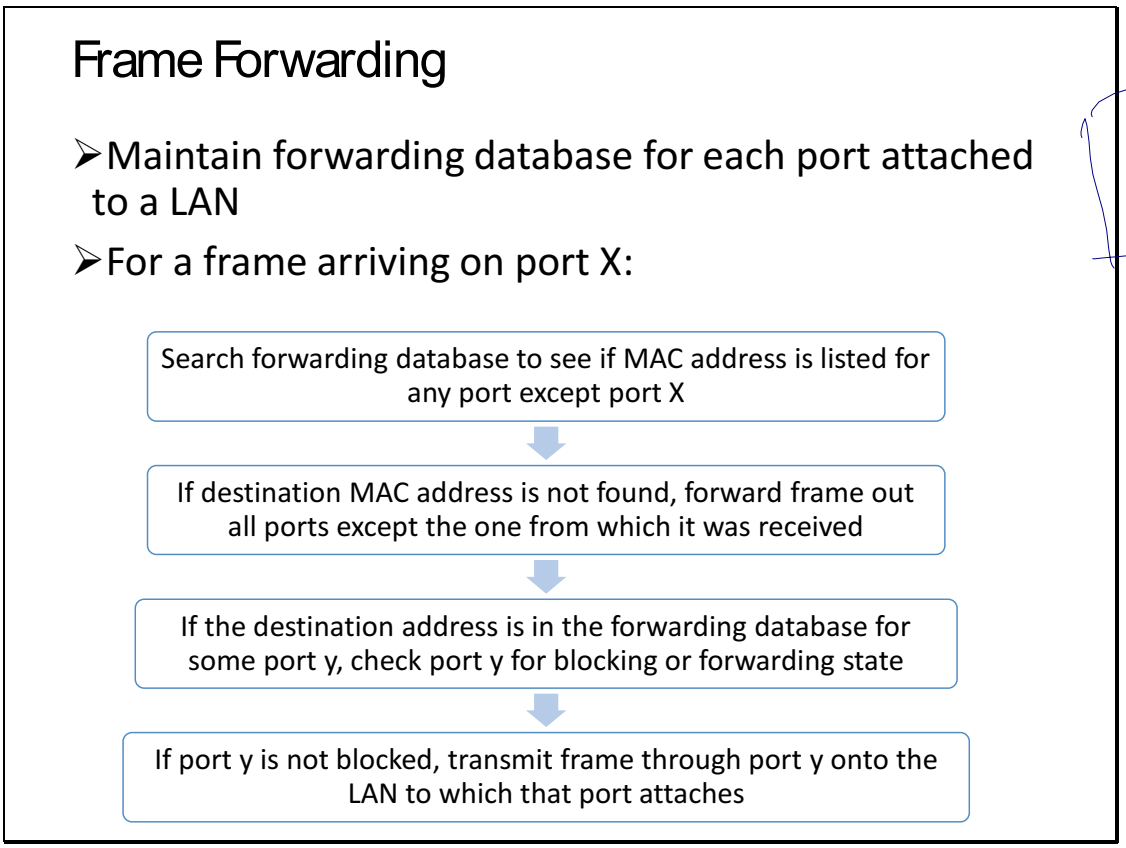
3333 1111

Preamble	Destination Address	Source Address	Type	Data	Pad	CRC

1111 3333



- Now, because both MAC addresses are in the switch's table, any information exchanged between 1111 and 3333 can be sent (filtered) out the appropriate port.



remember 3 steps

In this scheme, a bridge maintains a forwarding database for each port attached to a LAN. The database indicates the station addresses for which frames should be forwarded through that port. We can interpret this in the following fashion. For each port, a list of stations is maintained. A station is on the list if it is on the "same side" of the bridge as the port. For example, for bridge 102 of Figure 11.8, stations on LANs C, F, and G are on the same side of the bridge as the LAN C port, and stations on LANs A, B, D, and E are on the same side of the bridge as the LAN A port. When a frame is received on any port, the bridge must decide whether that frame is to be forwarded through the bridge and out through one of the bridge's other ports. Suppose that a bridge receives a MAC frame on port *x* **.**

The following rules are applied:

1. Search the forwarding database to determine if the MAC address is listed for any port except port **x .**

2. If the destination MAC address is not found, forward frame out all ports except the one from which it was received. This is part of the learning process described subsequently.

3. If the destination address is in the forwarding database for some port y, then determine whether port y is in a blocking or forwarding state. For reasons explained later, a port may sometimes be blocked, which prevents it from receiving or transmitting frames.

4. If port y is not blocked, transmit the frame through port y onto the LAN to which that port attaches.

Address Learning

- Can preload forwarding database
- When frame arrives at port X, it has come from the LAN attached to port X
- Use source address to update forwarding database for port X to include that address
- Have a timer on each entry in database
- If timer expires, entry is removed
- Each time frame arrives, source address checked against forwarding database
  - If present timer is reset and direction recorded
  - If not present entry is created and timer set

The preceding scheme assumes that the bridge is already equipped with a forwarding database that indicates the direction, from the bridge, of each destination station. This information can be preloaded into the bridge, as in fixed routing. However, an effective automatic mechanism for learning the direction of each station is desirable. A simple scheme for acquiring this information is based on the use of the source address field in each MAC frame.

The strategy is this. When a frame arrives on a particular port, it clearly has come from the direction of the incoming LAN. The source address field of the frame indicates the source station. Thus, a bridge can update its forwarding database for that port on the basis of the source address field of each incoming frame. To allow for changes in topology, each element in the database is equipped with a timer. When a new element is added to the database, its timer is set. If the timer expires, then the element is eliminated from the database, since the corresponding direction information may no longer be valid. Each time a frame is received, its source address is checked against the database. If the element is already in the database, the entry is updated (the direction may have changed) and the timer is reset. If the element is not in the database, a new entry is created, with its own timer.

Ch. 7 – Spanning Tree Protocol

CCNA 3 version 3.0

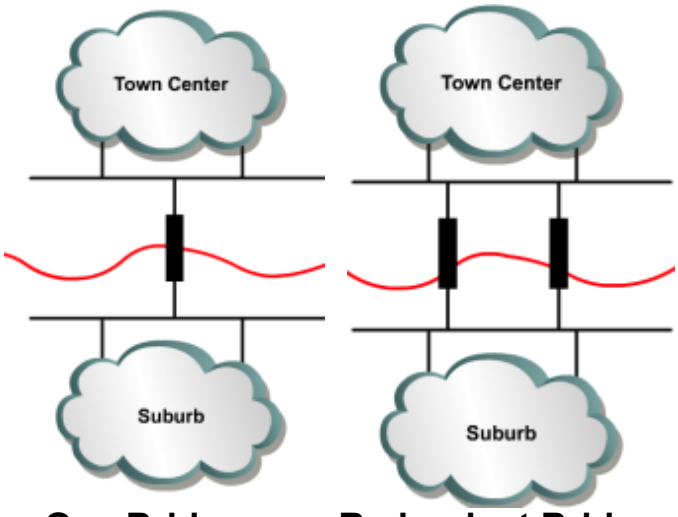
Overview

- Define redundancy and its importance in networking
- Describe the key elements of a redundant networking topology
- Define broadcast storms and describe their impact on switched networks
- Define multiple frame transmissions and describe their impact on switched networks
- Identify causes and results of MAC address database instability
- Identify the benefits and risks of a redundant topology
- Describe the role of spanning tree in a redundant-path switched network
- Identify the key elements of spanning tree operation
- Describe the process for root bridge election
- List the spanning-tree states in order
- Compare Spanning-Tree Protocol and Rapid Spanning-Tree Protocol

Redundancy

- Achieving such a goal requires extremely reliable networks.
- Reliability in networks is achieved by reliable equipment and by designing networks that are tolerant to failures and faults.
- The network is designed to reconverge rapidly so that the fault is bypassed.
- Fault tolerance is achieved by redundancy.
- Redundancy means to be in excess or exceeding what is usual and natural.

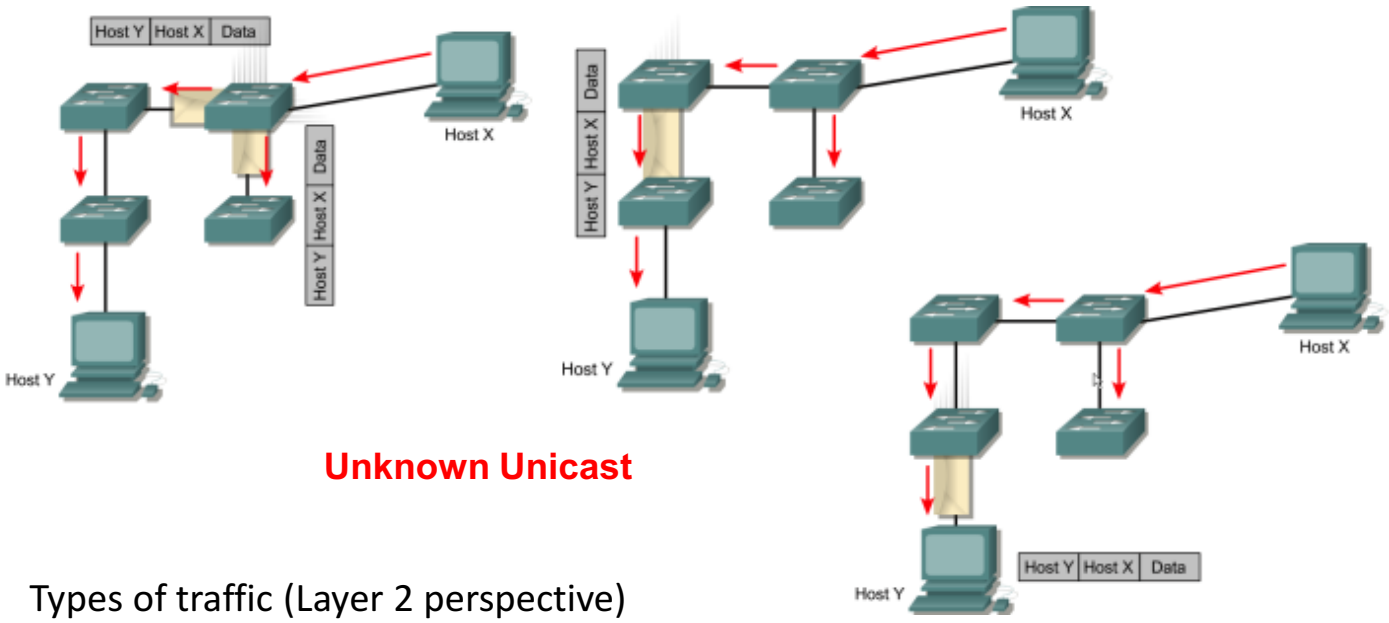
Redundant topologies



One Bridge **Redundant Bridges**

- A network of roads is a global example of a redundant topology.
- If one road is closed for repair there is likely an alternate route to the destination

Types of Traffic

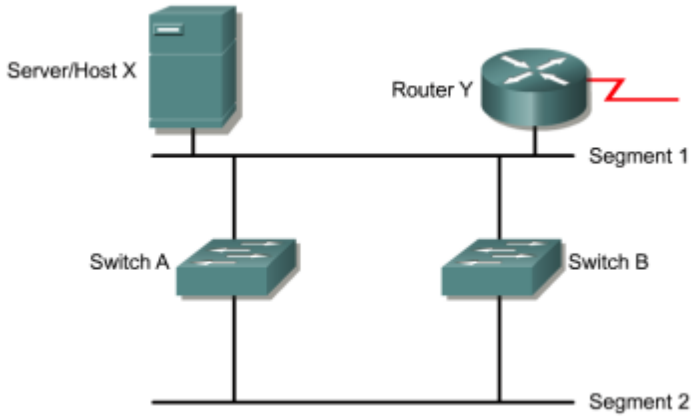


Unknown Unicast

Types of traffic (Layer 2 perspective)

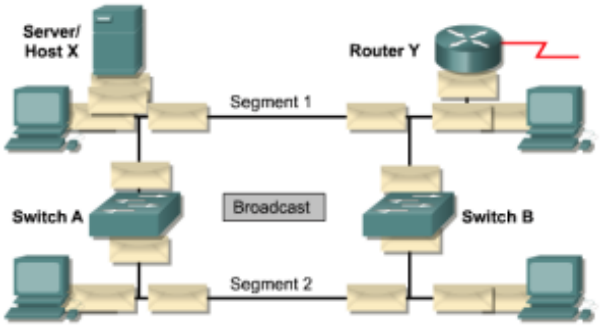
- Known Unicast: Destination addresses are in Switch Tables
- Unknown Unicast: Destination addresses are <u>not</u> in Switch Tables
- Multicast: Traffic sent to a group of addresses
- Broadcast: Traffic forwarded out all interfaces except incoming interface.

Redundant switched topologies



- Switches learn the MAC addresses of devices on their ports so that data can be properly forwarded to the destination.
- Switches will flood frames for unknown destinations until they learn the MAC addresses of the devices.
- Broadcasts and multicasts are also flooded. (Unless switch is doing Multicast Snooping or IGMP)
- A redundant switched topology ***may*** (STP disabled) cause broadcast storms, multiple frame copies, and MAC address table instability problems.

Broadcast Storm



A state in which a message that has been broadcast across a networkresults in even more responses, and each response results in still more responses in a snowball effect.
www.webopedia.com

- Broadcasts and multicasts can cause problems in a switched network.
- If Host X sends a broadcast, like an ARP request for the Layer 2 address of the router, then Switch A will forward the broadcast out all ports.
- Switch B, being on the same segment, also forwards all broadcasts.
- Switch B sees all the broadcasts that Switch A forwarded and Switch A sees all the broadcasts that Switch B forwarded.
- Switch A sees the broadcasts and forwards them.
- Switch B sees the broadcasts and forwards them.
- The switches continue to propagate broadcast traffic over and over.
- This is called a broadcast storm.

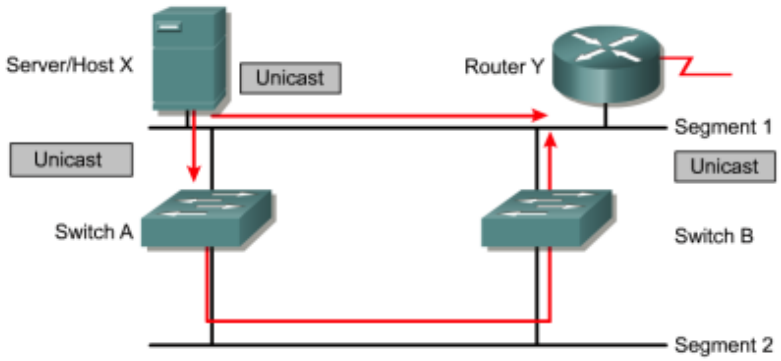
Spanning tree → not loop.
→ no device left out.

Spanning tree algo. missing.

BPDU ← meant for spanning tree calculation

01:80:C2:00:00:00 ← well known multicast address for spanning tree.

Multiple frame transmissions



- In a redundant switched network it is possible for an end device to receive multiple frames.
- Assume that the MAC address of Router Y has been timed out by both switches.
- Also assume that Host X still has the MAC address of Router Y in its ARP cache and sends a unicast frame to Router Y.

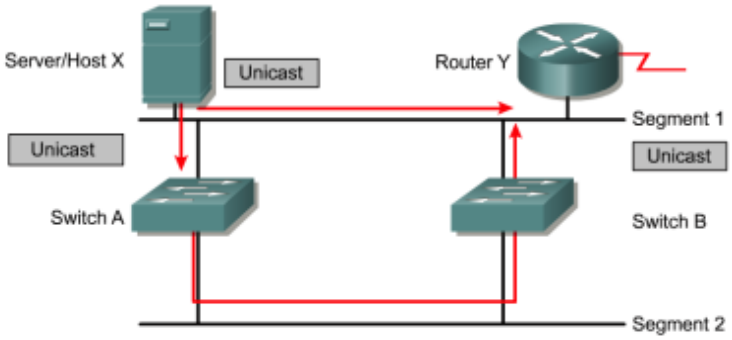
BPDU fields ← no memorization

closedown.

↓

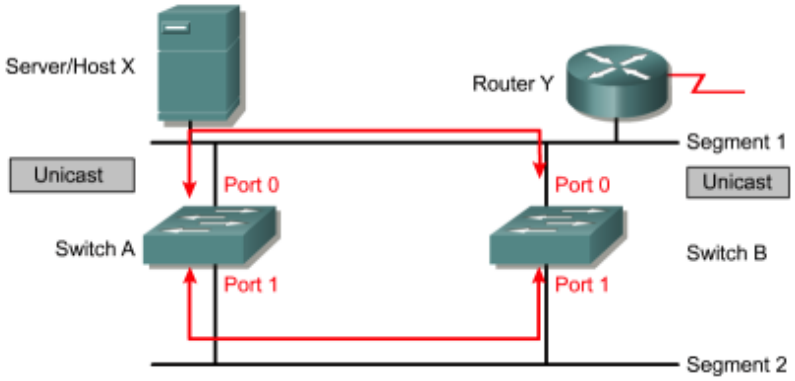
+ port roles, port states (after stabilize, state in. block state or fwd state), STP - Rapid Spanning tree protocol (RSTP) to shorten the 50s wait

Multiple frame transmissions



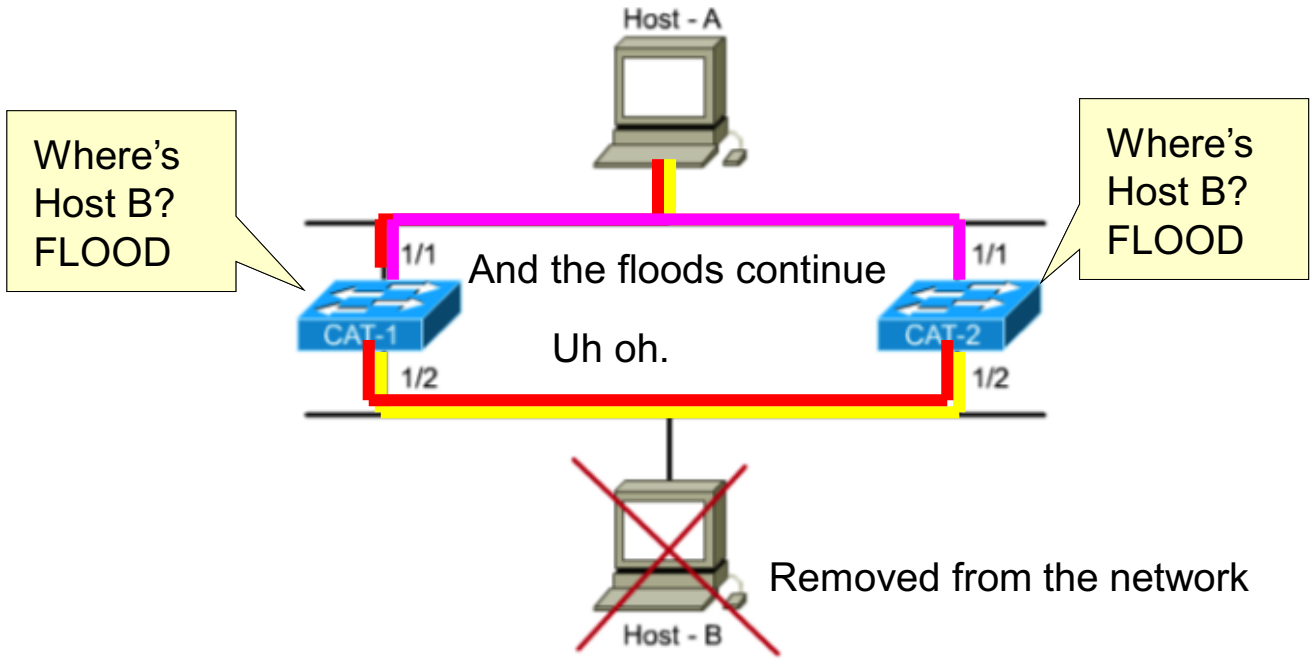
- The router receives the frame because it is on the same segment as Host X.
- Switch A does not have the MAC address of the Router Y and will therefore flood the frame out its ports. (Segment 2)
- Switch B also does not know which port Router Y is on.
- Note: Switch B will forward the the unicast onto Segment 2, creating multiple frames on that segment.
- After Switch B receives the frame from Switch A , it then floods the frame it received causing Router Y to receive multiple copies of the same frame.
- This is a causes of unnecessary processing in all devices.

Media access control database instability

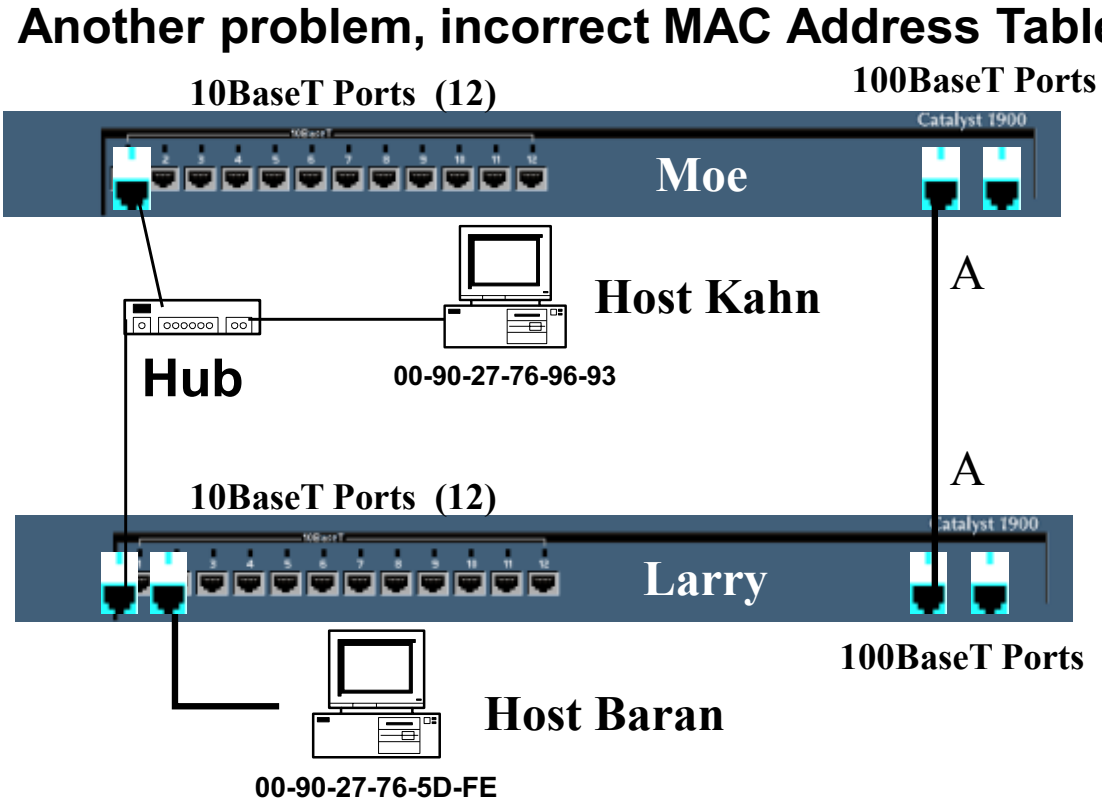


- In a redundant switched network it is possible for switches to learn the wrong information.
- A switch can incorrectly learn that a MAC address is on one port, when it is actually on a different port.
- Host X sends a frame directed to Router Y.
- Switches A and B learn the MAC address of Host X on port 0.
- The frame to Router Y is flooded on port 1 of both switches.
- Switches A and B see this information on port 1 and incorrectly learn the MAC address of Host X on port 1.

Layer 2 Loops - Flooded unicast frames

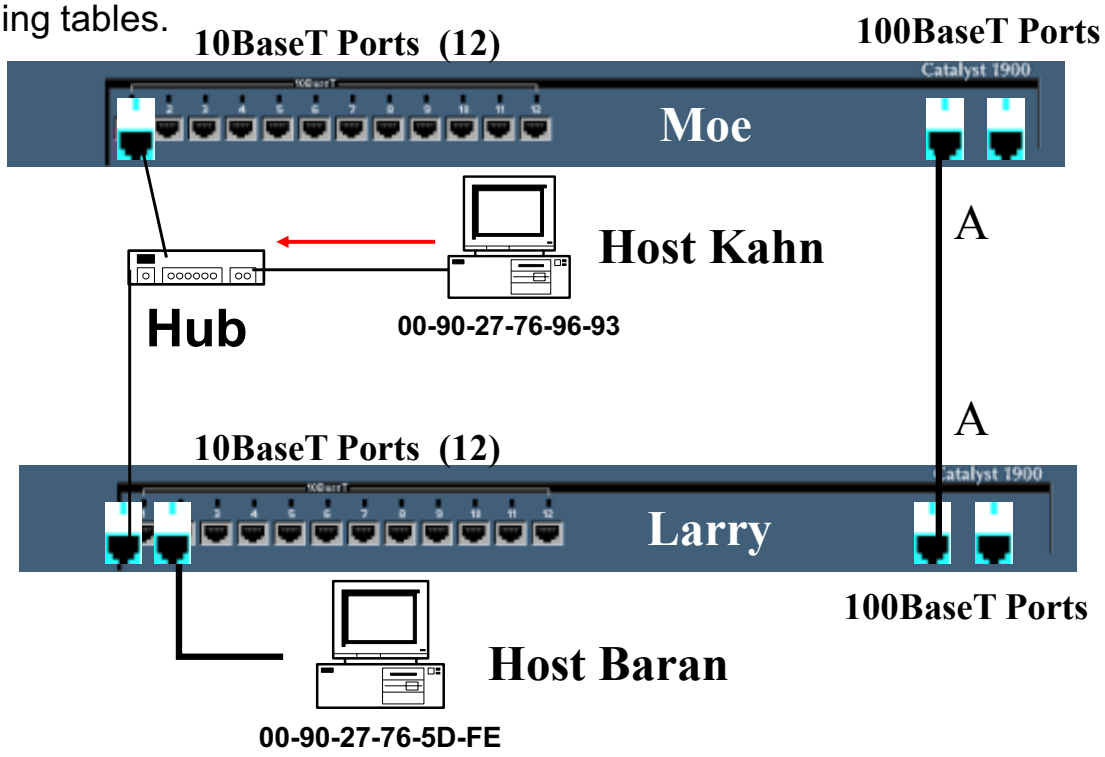


Redundant Paths and No Spanning Tree



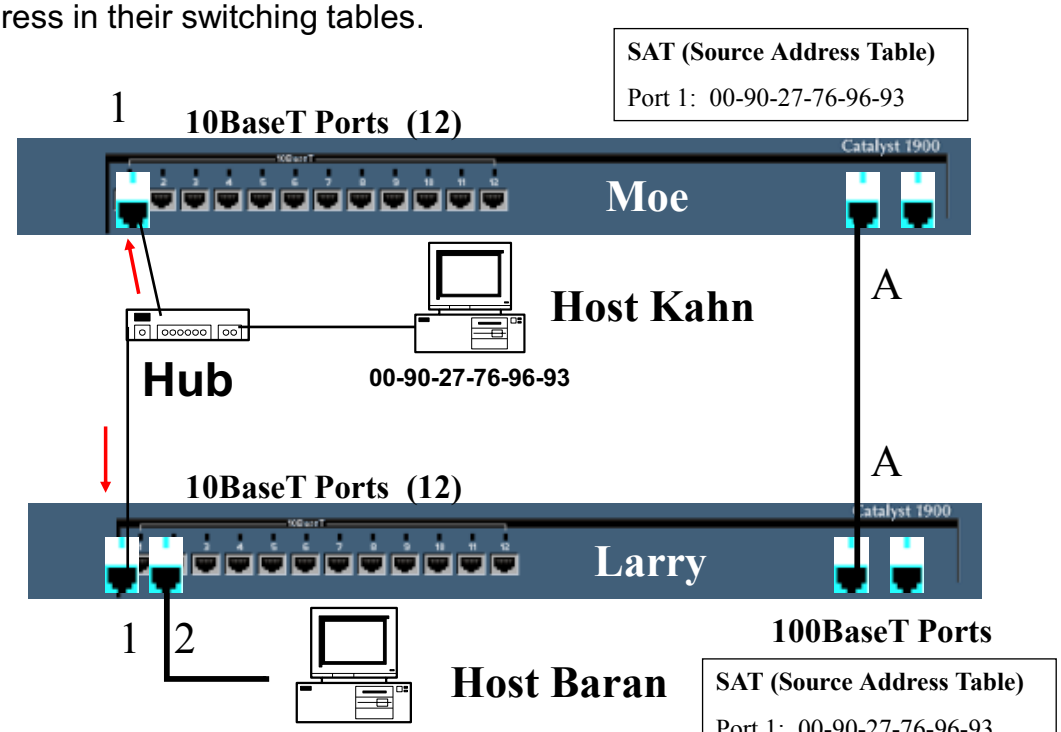
Redundant Paths and No Spanning Tree

Host Kahn sends an Ethernet frame to Host Baran. Both Switch Moe and Switch Larry see the frame and record Host Kahn's Mac Address in their switching tables.



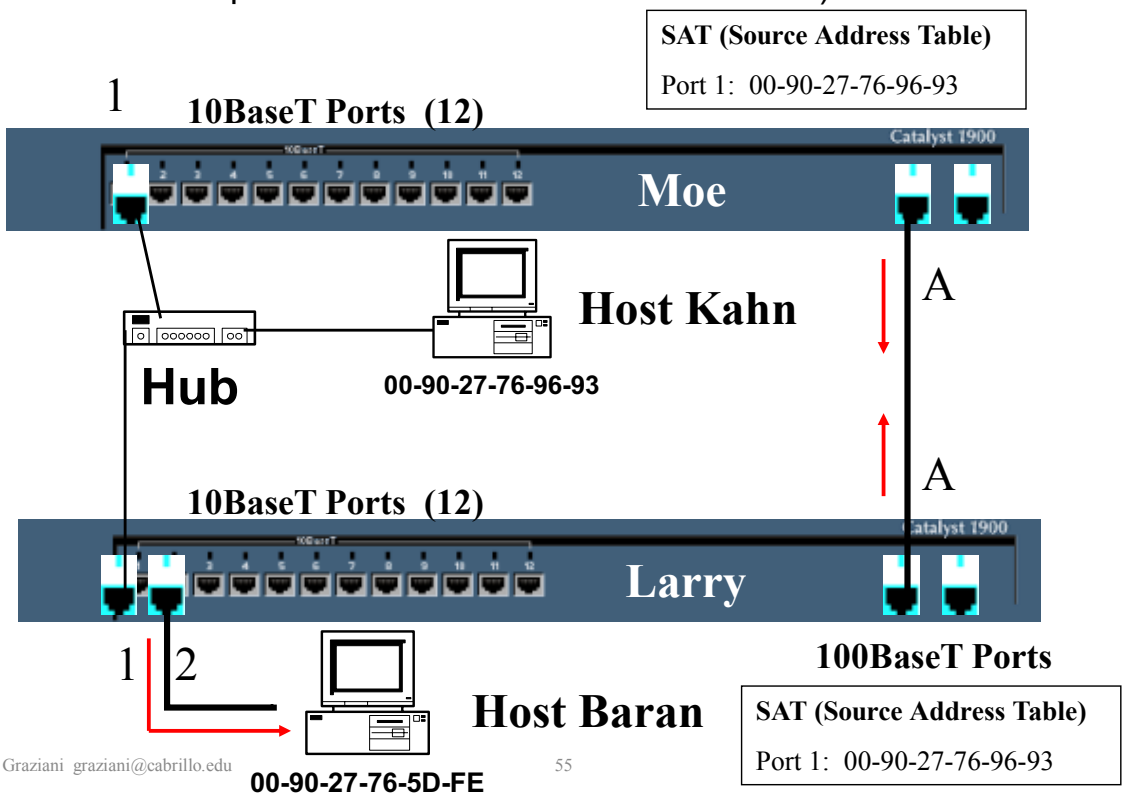
Redundant Paths and No Spanning Tree

Both Switch Moe and Switch Larry see the frame and record Host Kahn's Mac Address in their switching tables.



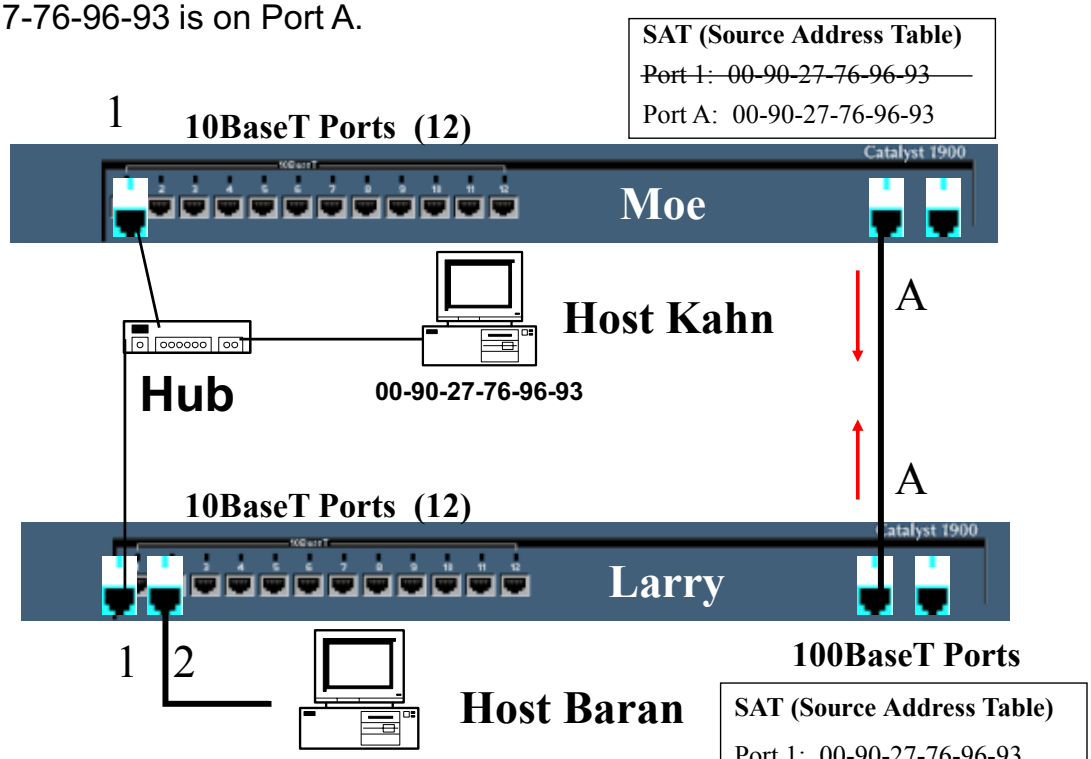
Redundant Paths and No Spanning Tree

Both Switches do not have the **destination MAC address** in their table so they **both flood** it out all ports. Host Baran receives the frame.)



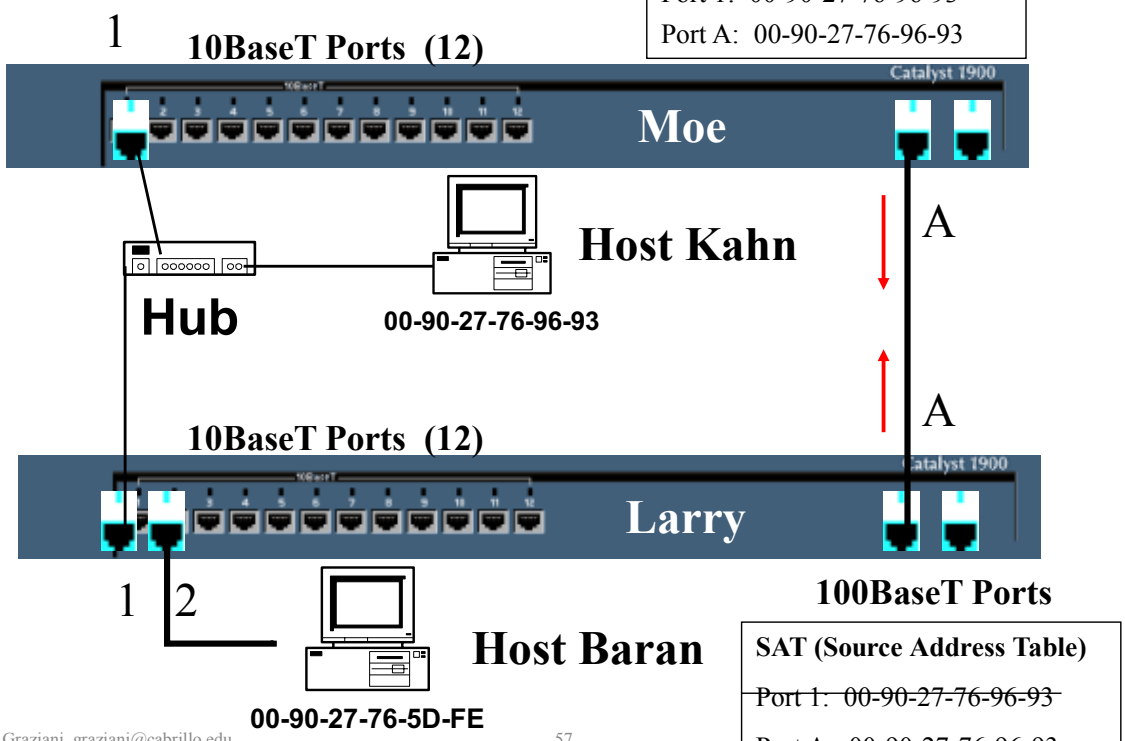
Redundant Paths and No Spanning Tree

Switch Moe now learns, **incorrectly**, that the Source Address 00-90-27-76-96-93 is on Port A.



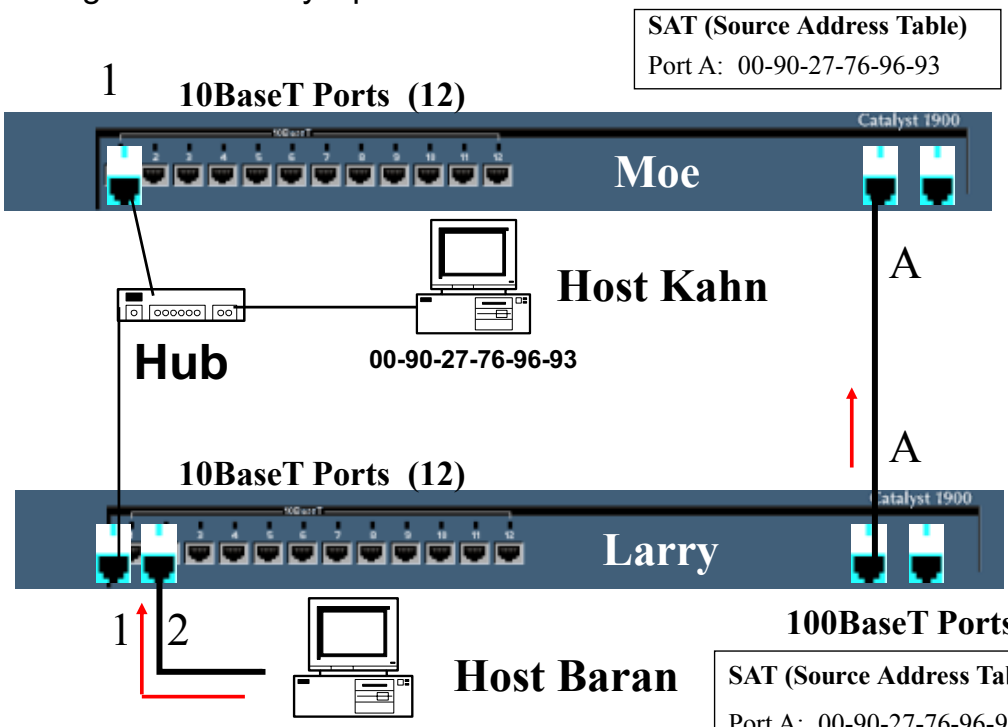
Redundant Paths and No Spanning Tree

Switch Larry also learns, **incorrectly**, that the Source Address 00-90-27-76-96-93 is on Port A.



Redundant Paths and No Spanning Tree

Now, when Host Baran sends a frame to Host Kahn, it will be sent the longer way, through Switch Larry's port A.



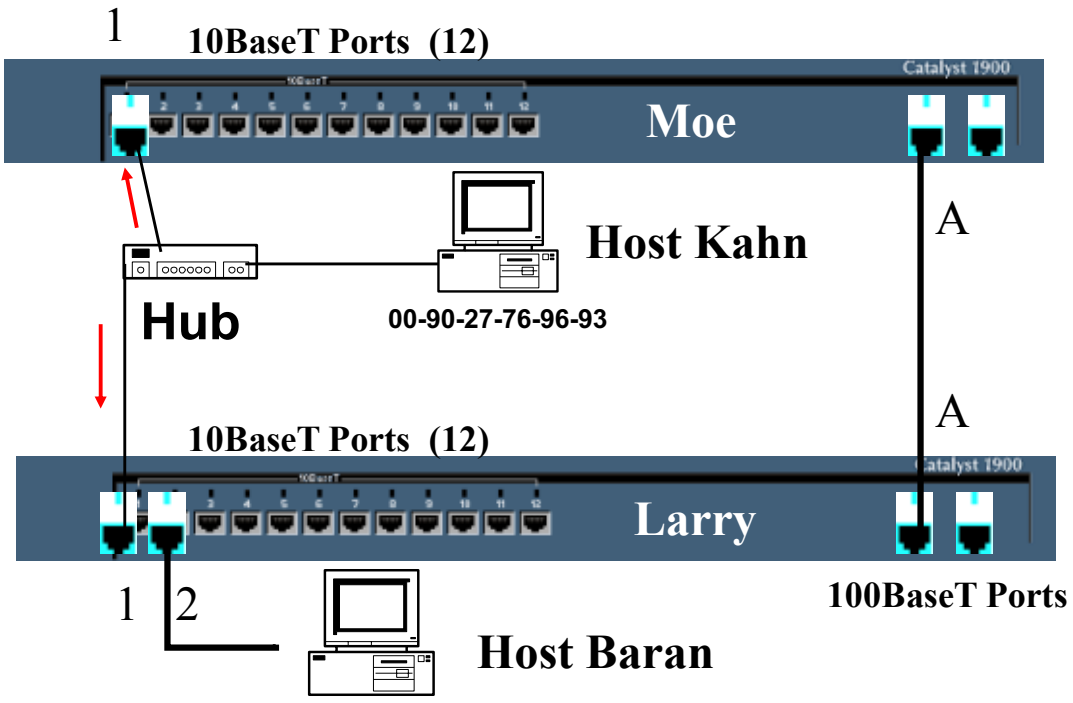
Redundant Paths and No Spanning Tree

- Then, the same confusion happens, but this time with Host Baran.
- Okay, maybe not the end of the world.
- Frames will just take a longer path and you may also see other "unexpected results."

- But what about broadcast frames, like ARP Requests?

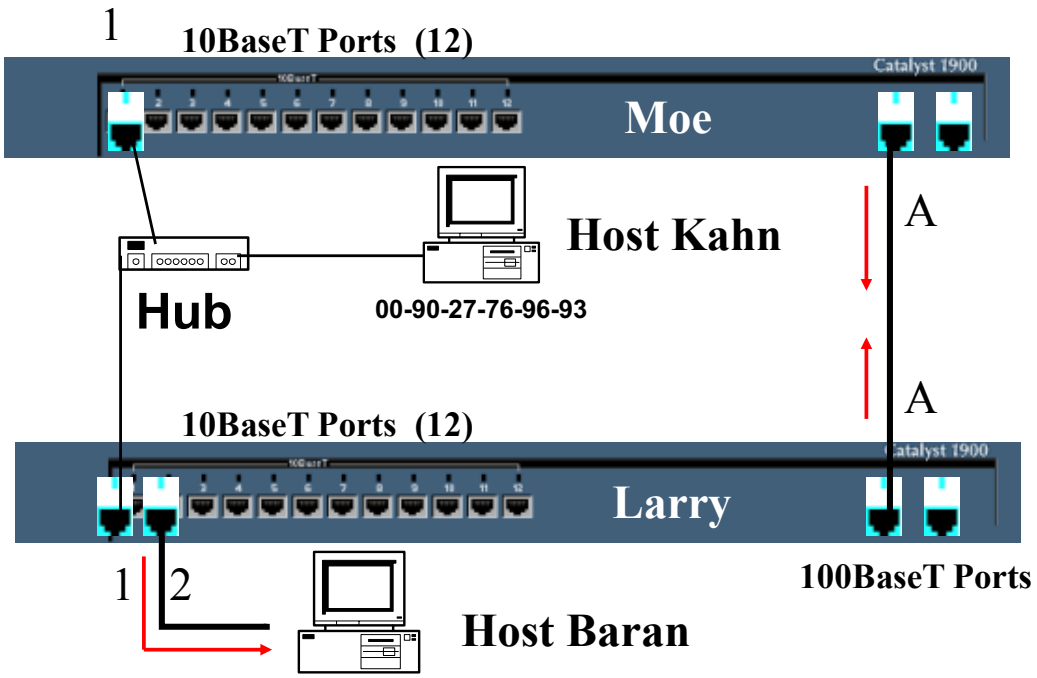
Broadcasts and No Spanning Tree

Lets, leave the switching tables alone and just look at what happens with the frames. Host Kahn sends out a layer 2 broadcast frame, like an ARP Request.



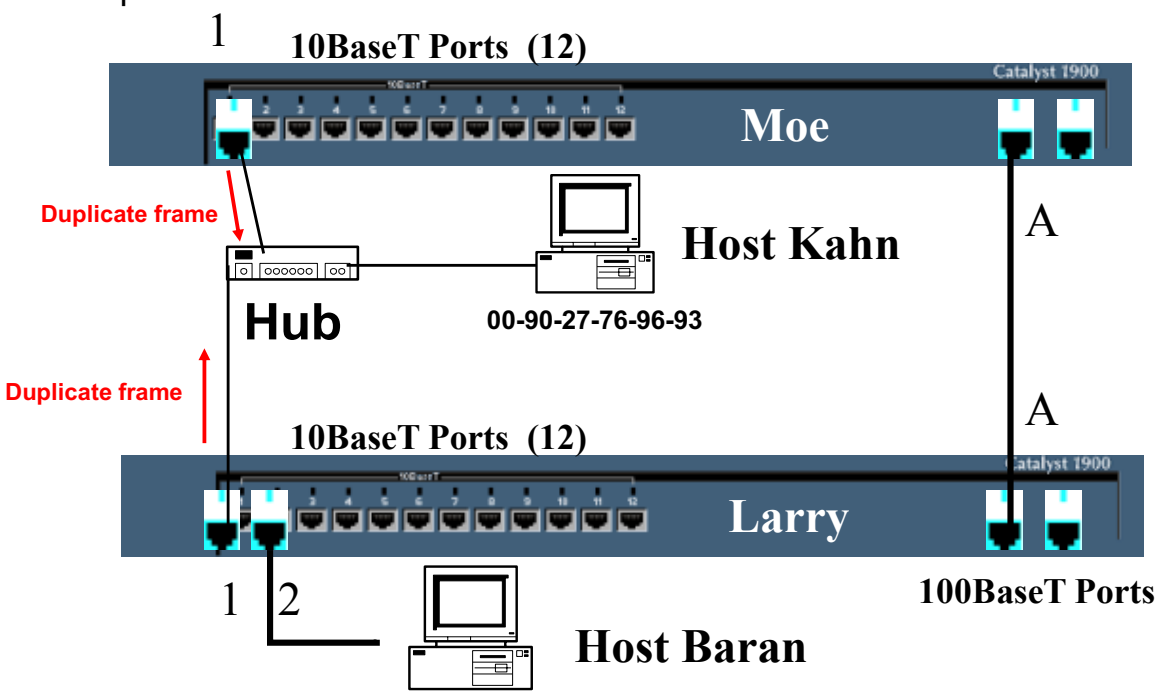
Broadcasts and No Spanning Tree

Because it is a layer 2 broadcast frame, both switches, Moe and Larry, **flood the frame out all ports**, including their port A's.



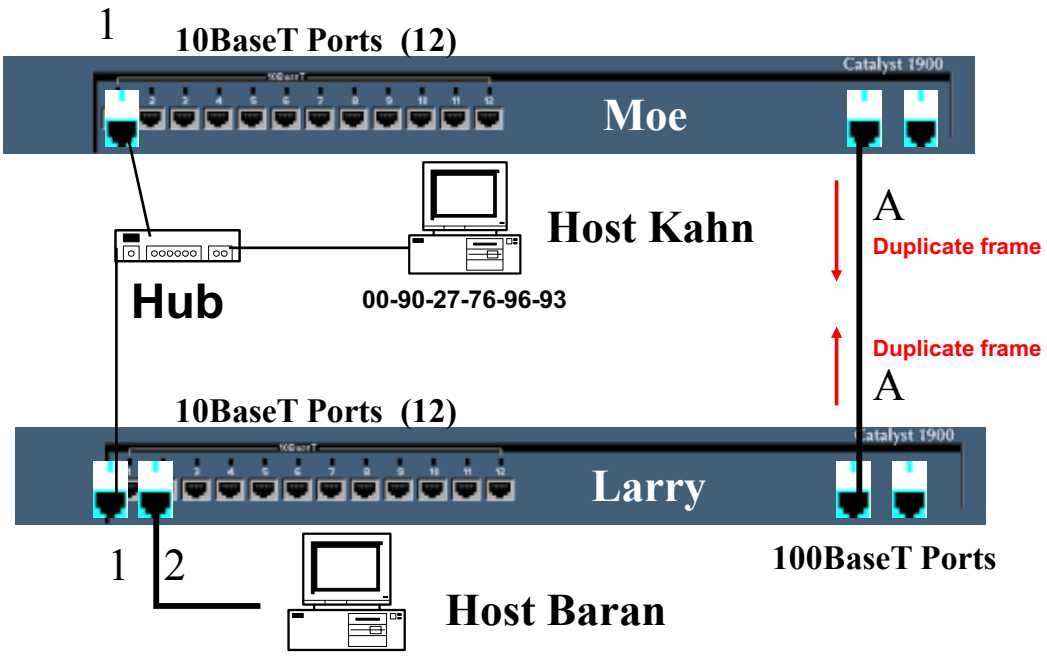
Broadcasts and No Spanning Tree

Both switches receive the same broadcast, but on a different port. Doing what switches do, **both switches flood the duplicate broadcast frame** out their other ports.



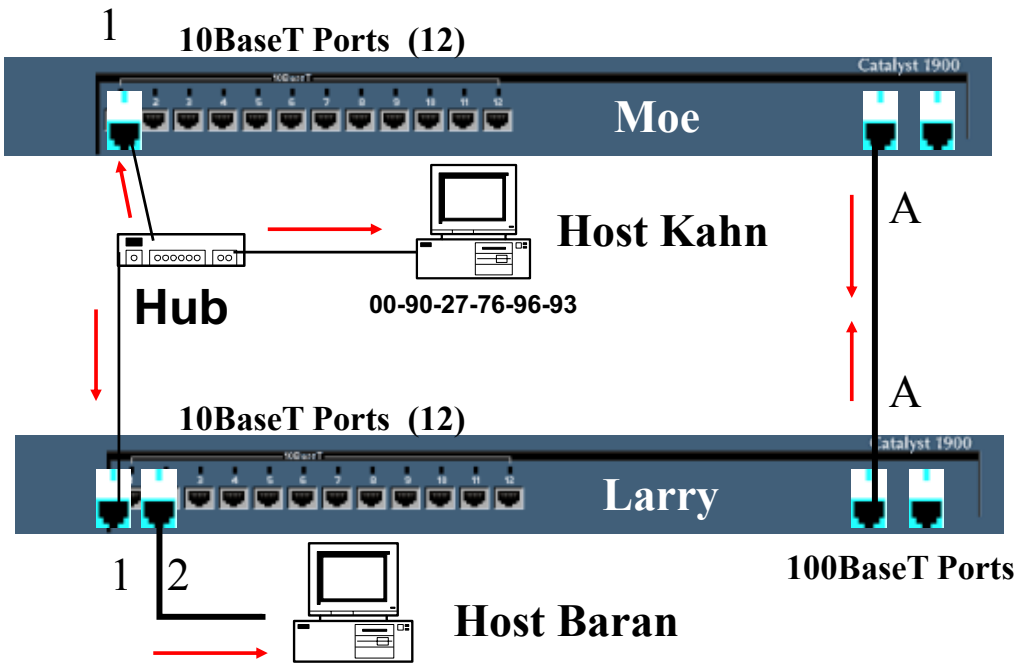
Broadcasts and No Spanning Tree

Here we go again, with the switches flooding the same broadcast again out its other ports. This results in duplicate frames, known as a ***broadcast storm***!

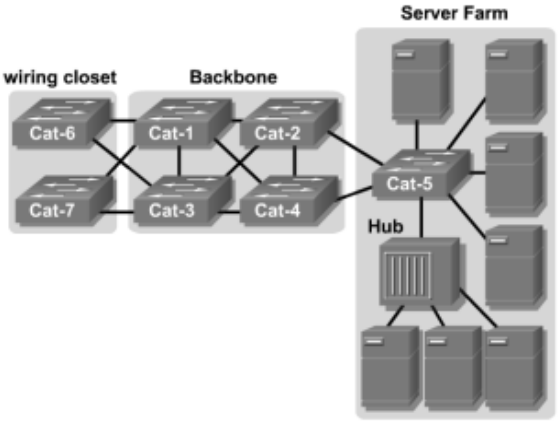


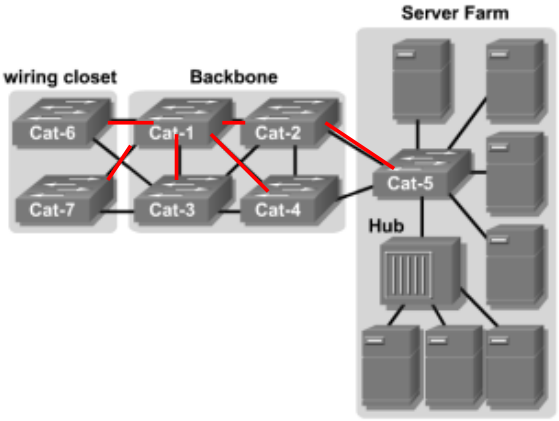
Broadcasts and No Spanning Tree

Remember, that layer 2 broadcasts not only take up network bandwidth, but must be processed by each host. This can severely impact a network, to the point of making it unusable.

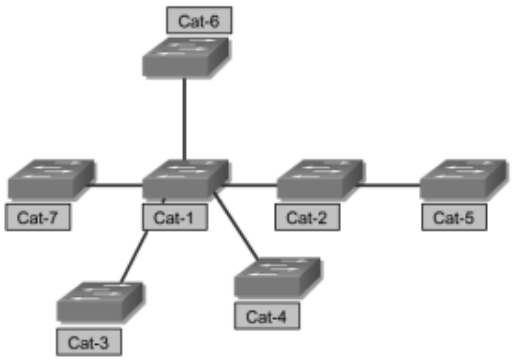


Redundant topology and spanning tree

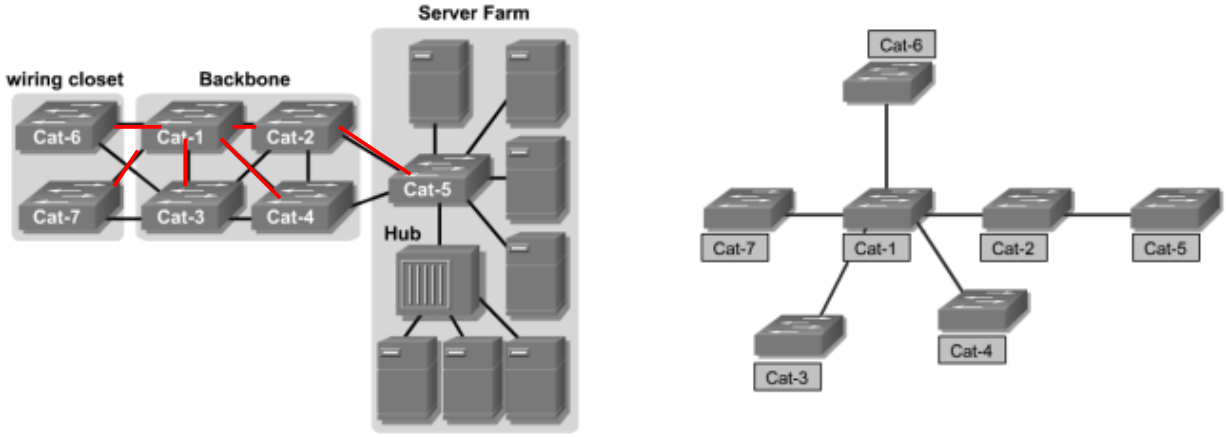




- Unlike IP, in the Layer 2 header there is no Time to Live (TTL).
- The solution is to allow physical loops, but create a loop free logical topology.
- The loop free logical topology created is called a tree.
- This topology is a star or extended star logical topology, the spanning tree of the network.

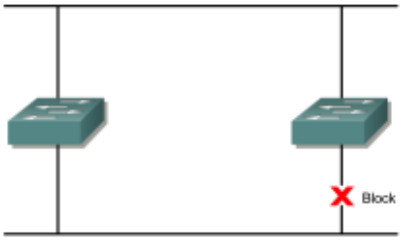


Redundant topology and spanning tree

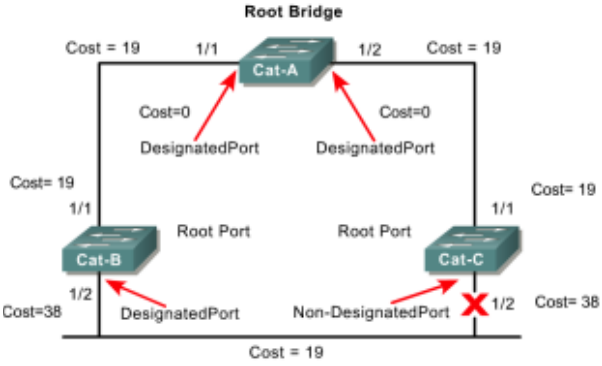


- It is a spanning tree because all devices in the network are reachable or spanned.
- The algorithm used to create this loop free logical topology is the **spanning-tree algorithm**.
- This algorithm can take a relatively long time to converge.
- A new algorithm called the **rapid spanning-tree algorithm** is being introduced to reduce the time for a network to compute a loop free logical topology. (later)

Spanning-Tree Protocol (STP)



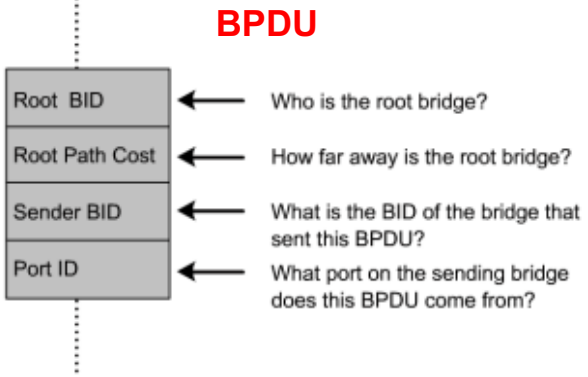
Link Speed	Cost(Revised IEEE Spec)	Cost (Previous IEEE Spec)
10 Gbps	2	1
1 Gbps	4	1
100 Mbps	19	10
10 Mbps	100	100

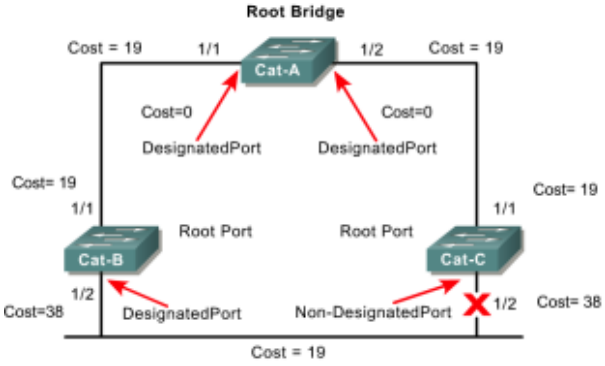


We will see how this works in a moment.

- Shortest path is based on cumulative link costs.
- Link costs are based on the speed of the link.
- The Spanning-Tree Protocol establishes a root node, called the root bridge.
- The Spanning-Tree Protocol constructs a topology that has one path for reaching every network node.
- The resulting tree originates from the **root bridge**.
- **Redundant links** that are not part of the shortest path tree are **blocked**.

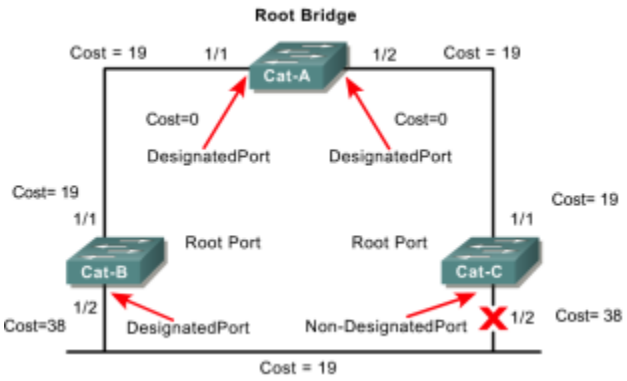
Spanning-Tree Protocol (STP)





- It is because certain paths are blocked that a loop free topology is possible.
- Data frames received on blocked links are dropped.
- The Spanning-Tree Protocol requires network devices to exchange messages to detect bridging loops.
- Links that will cause a loop are put into a blocking state.
- topology, is called a **Bridge Protocol Data Unit (BPDU).**
- BPDUs continue to be received on blocked ports.
- This ensures that if an active path or device fails, a new spanning tree can be calculated.

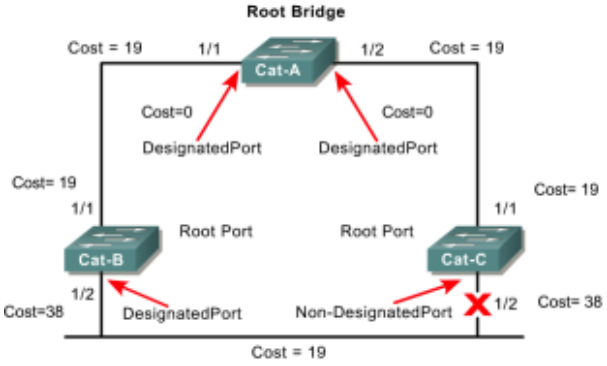
Spanning-Tree Protocol (STP)



BPDUs contain enough information so that all switches can do the following:

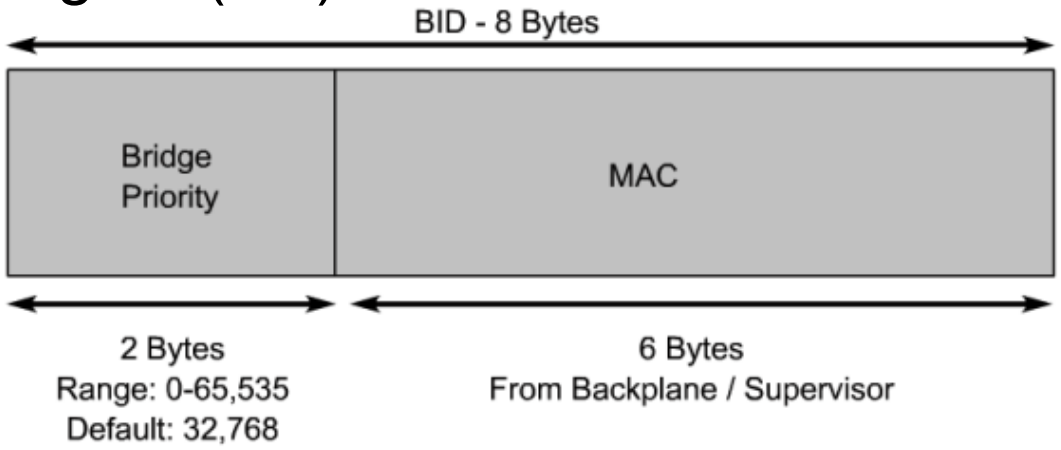
- Select a **single switch that will act as the root** of the spanning tree
- Calculate the **shortest path from itself to the root switch**
- **Designate one of the switches as the closest one to the root**, for each LAN segment. This bridge is called the "**designated switch**".
  - The designated switch handles all communication from that LAN towards the root bridge.
- Choose one of its ports as its root port, for each non-root switch.
  - This is the interface that gives the best path to the root switch.
- Select ports that are part of the spanning tree, the designated ports. Non-designated ports are blocked.

Two Key Concepts: BID and Path Cost



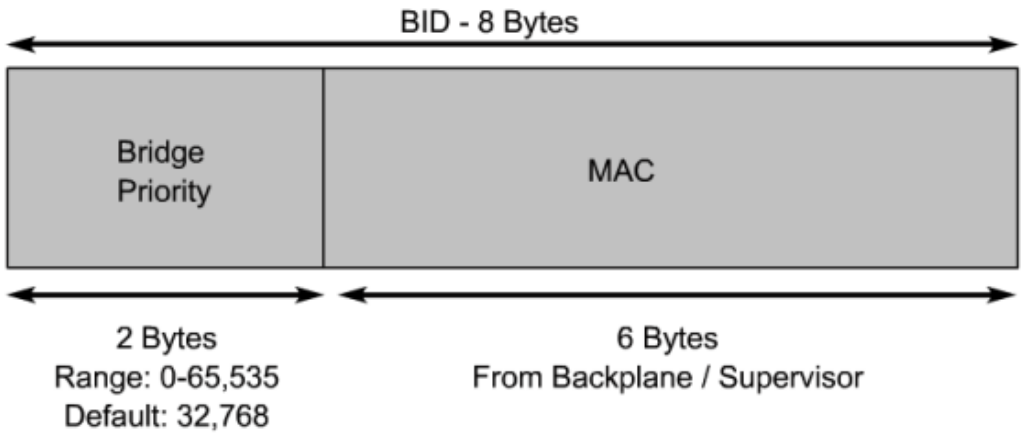
- STP executes an algorithm called Spanning Tree Algorithm (STA).
- STA chooses a reference point, called a root bridge, and then determines the available paths to that reference point.
  - If more than two paths exists, STA picks the best path and blocks the rest
- STP calculations make extensive use of two key concepts in creating a loop-free topology:
  - **Bridge ID**
  - **Path Cost**

Bridge ID (BID)



- **Bridge ID (BID)** is used to identify each bridge/switch.
- The BID is used in determining the center of the network, in respect to STP, known as the root bridge.
- Consists of two components:
  - **A 2-byte Bridge Priority**: defaults to **32,768** or 0x8000.
  - **A 6-byte MAC address**

Bridge ID (BID)



- **Bridge Priority** is usually expressed in **decimal format** and the **MAC address** in the BID is usually expressed in **hexadecimal format**.
- BID is used to elect a root bridge (coming)
- **Lowest Bridge ID is the root.**
- If all devices have the same priority, the bridge with the lowest MAC address becomes the root bridge. (Yikes!)

Path Cost

Link Speed	Cost(Revised IEEE Spec)	Cost (Previous IEEE Spec)
10 Gbps	2	1
1 Gbps	4	1
100 Mbps	19	10
10 Mbps	100	100

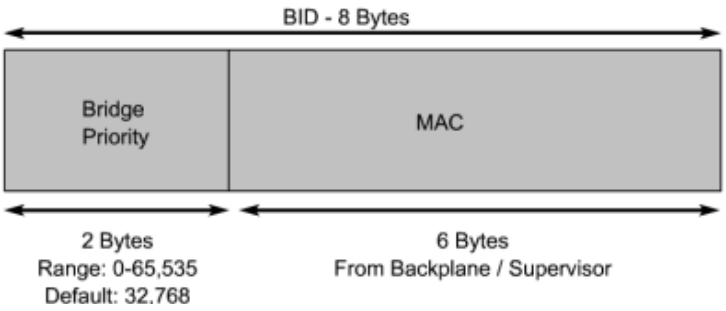
- Bridges use the concept of cost to evaluate how close they are to other bridges.
- This will be used in the STP development of a loop-free topology .
- **Originally, 802.1d** defined cost as 1000/bandwidth of the link in Mbps.
  - Cost of 10Mbps link = 100 or 1000/10
  - Cost of 100Mbps link = 10 or 1000/100
  - Cost of 1Gbps link = 1 or 1000/1000
- Running out of room for faster switches including 10 Gbps Ethernet.

Path Cost

Link Speed	Cost(Revised IEEE Spec)	Cost (Previous IEEE Spec)
10 Gbps	2	1
1 Gbps	4	1
100 Mbps	19	10
10 Mbps	100	100

- IEEE modified the most to use a non-linear scale with the new values of:
  - 4 Mbps 250 (cost)
  - 10 Mbps 100 (cost)
  - 16 Mbps 62 (cost)
  - 45 Mbps 39 (cost)
  - 100 Mbps 19 (cost)
  - 155 Mbps 14 (cost)
  - 622 Mbps 6 (cost)
  - 1 Gbps 4 (cost)
  - 10 Gbps 2 (cost)

Path Cost



Link Speed	Cost(Revised IEEE Spec)	Cost (Previous IEEE Spec)
10 Gbps	2	1
1 Gbps	4	1
100 Mbps	19	10
10 Mbps	100	100

- You can modify the path cost by modifying the cost of a port.
  - Exercise caution when you do this!
- BID and Path Cost are used to develop a loop-free topology .
- Coming very soon!
- But first the **Four-Step STP Decision Sequence**

Four-Step STP Decision Sequence

- When creating a loop-free topology, STP always uses the same four-step decision sequence:

  Four-Step decision Sequence

  Step 1 - Lowest BID

  Step 2 - Lowest Path Cost to Root Bridge

  Step 3 - Lowest Sender BID

  Step 4 - Lowest Port ID

- Bridges use Configuration BPDUs during this four-step process.
  - There is another type of BPDU known as Topology Change Notification (TCN) BPDU.

Four-Step STP Decision Sequence

BPDU key concepts:

- Bridges save a copy of only the best BPDU seen on every port.
- When making this evaluation, it considers all of the BPDUs received on the port, as well as the BPDU that would be sent on that port.
- As every BPDU arrives, it is checked against this four-step sequence to see if it is more attractive (lower in value) than the existing BPDU saved for that port.
- Only the lowest value BPDU is saved.
- Bridges send configuration BPDUs until a more attractive BPDU is received.
- Okay, lets see how this is used...

Three Steps of Initial STP Convergence

- The STP algorithm uses three simple steps to converge on a loop-free topology.
- Switches go through three steps for their initial convergence:

STP Convergence

Step 1 Elect one Root Bridge
Step 2 Elect Root Ports
Step 3 Elect Designated Ports

- All STP decisions are based on a the following predetermined sequence:

Four-Step decision Sequence

Step 1 - Lowest BID

Step 2 - Lowest Path Cost to Root Bridge

Step 3 - Lowest Sender BID

Step 4 - Lowest Port ID

Three Steps of Initial STP Convergence

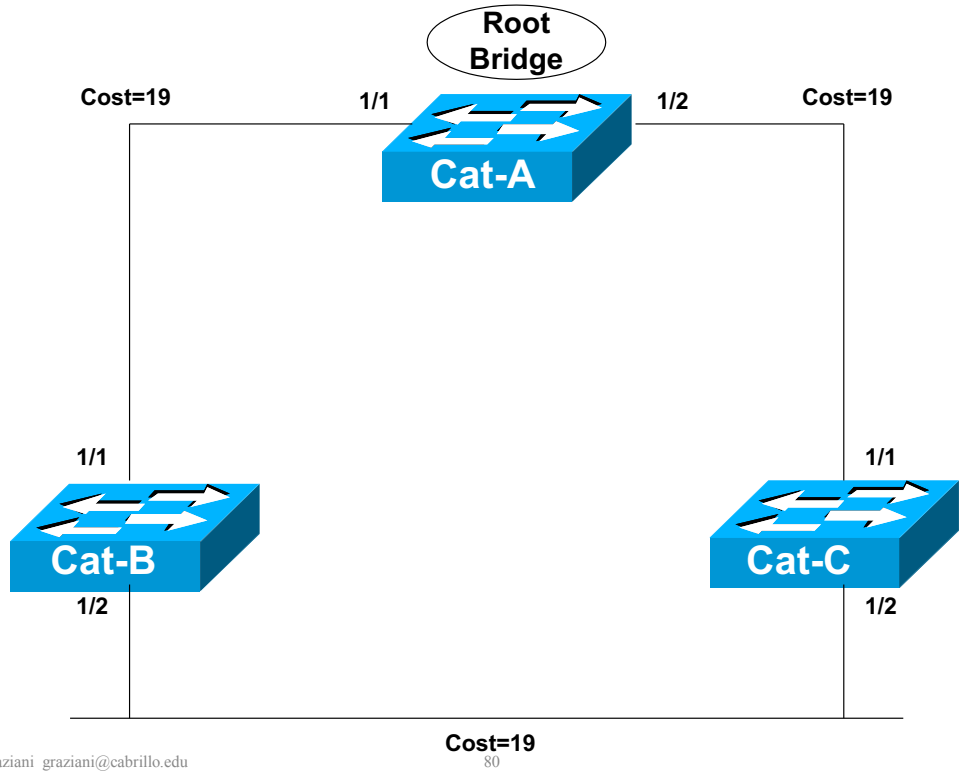
STP Convergence

Step 1 Elect one Root Bridge

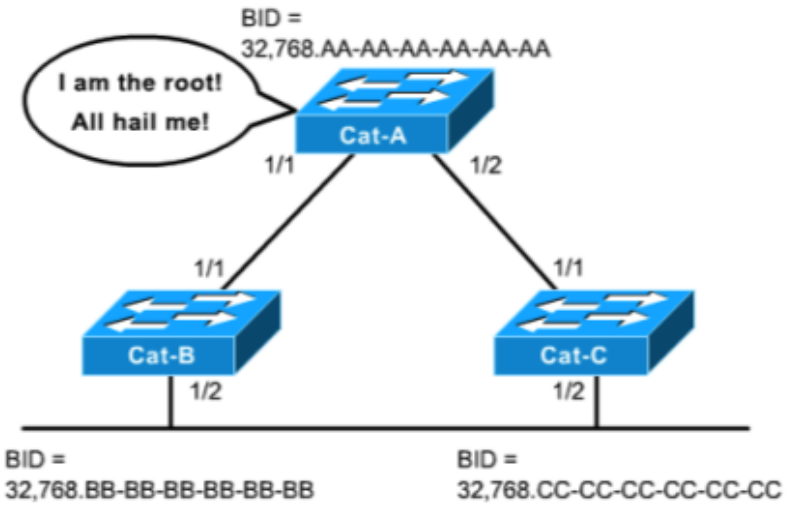
Step 2 Elect Root Ports

Step 3 Elect Designated Ports

Step 1 Elect one Root Bridge



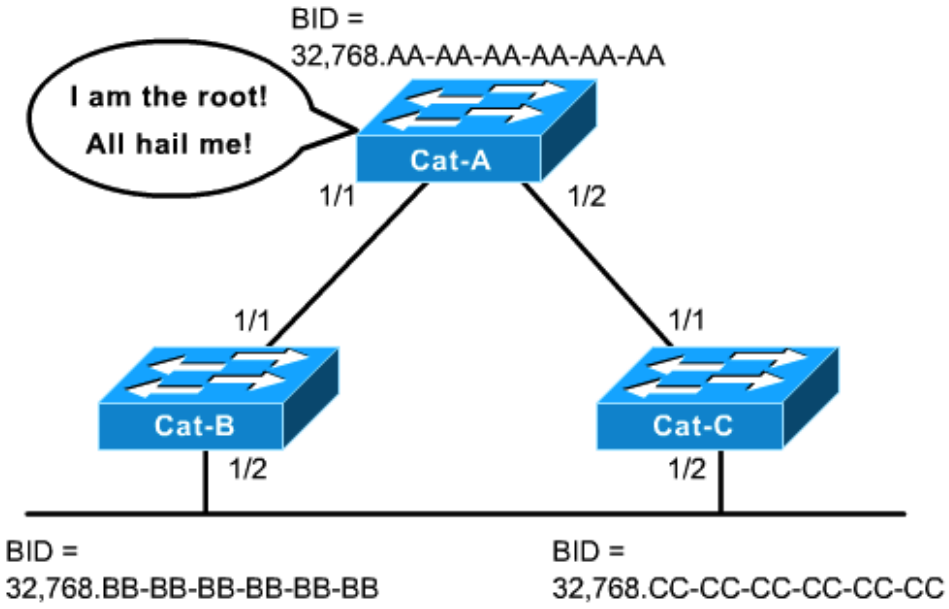
Step 1 Elect one Root Bridge



- When the network first starts, all bridges are announcing a chaotic mix of BPDUs.
- All bridges immediately begin applying the four-step sequence decision process.
- Switches need to elect a single Root Bridge.
- Switch with the **lowest BID** wins!
- Note: Many texts refer to the term “highest priority” which is the “lowest” BID value.
- This is known as the “Root War.”

Step 1 Elect one Root Bridge

Cat-A has the lowest Bridge MAC Address, so it wins the Root War!



All 3 switches have the same default Bridge Priority value of 32,768

Step 1 Elect one Root Bridge

BPDU

802.3 Header

```
  Destination:   01:80:C2:00:00:00  Mcast 802.1d Bridge group
  Source:        00:D0:C0:F5:18:D1
  LLC Length:    38
```

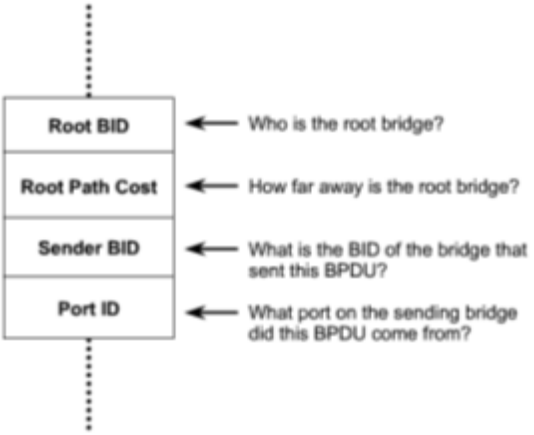
802.2 Logical Link Control (LLC) Header

```
  Dest. SAP:     0x42  802.1 Bridge Spanning Tree
  Source SAP:    0x42  802.1 Bridge Spanning Tree
  Command:       0x03  Unnumbered Information
```

802.1 - Bridge Spanning Tree

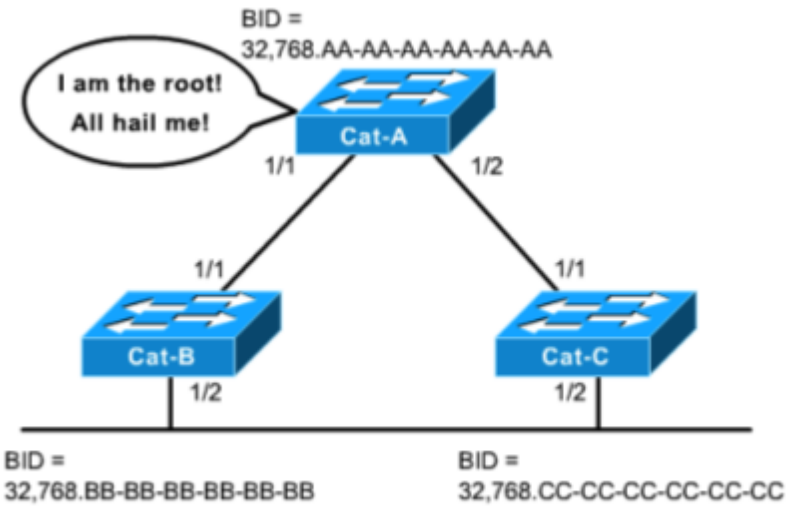
```
  Protocol Identifier:  0
  Protocol Version ID:  0
  Message Type:         0  Configuration Message
  Flags:                %00000000
  Root Priority/ID:     0x8000/ 00:D0:C0:F5:18:C0
  Cost Of Path To Root: 0x00000000  (0)
  Bridge Priority/ID:   0x8000/ 00:D0:C0:F5:18:C0
  Port Priority/ID:     0x80/ 0x1D
  Message Age:          0/256 seconds (exactly 0  seconds)
  Maximum Age:          5120/256 seconds (exactly 20  seconds)
  Hello Time:           512/256 seconds (exactly 2  seconds)
  Forward Delay:        3840/256 seconds (exactly 15  seconds)
```

Its all done with BPDUs!



Configuration BPDUs are sent every 2 seconds by default.

Step 1 Elect one Root Bridge



- At the beginning, all bridges assume they are the center of the universe and declare themselves as the Root Bridge, by placing its own BID in the Root BID field of the BPDU.
- Once all of the switches see that Cat-A has the lowest BID, they are all in agreement that Cat-A is the Root Bridge.

Three Steps of Initial STP Convergence

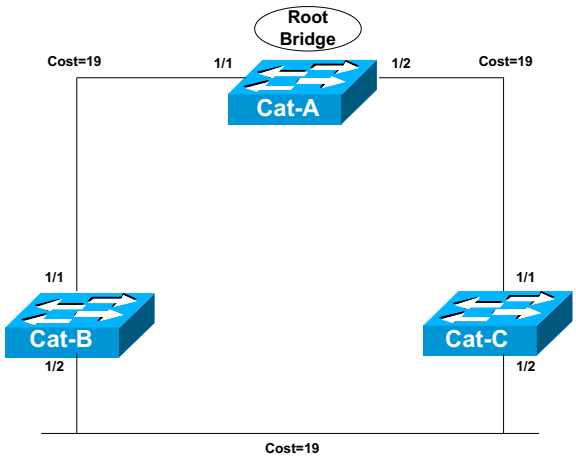
STP Convergence

Step 1 Elect one Root Bridge

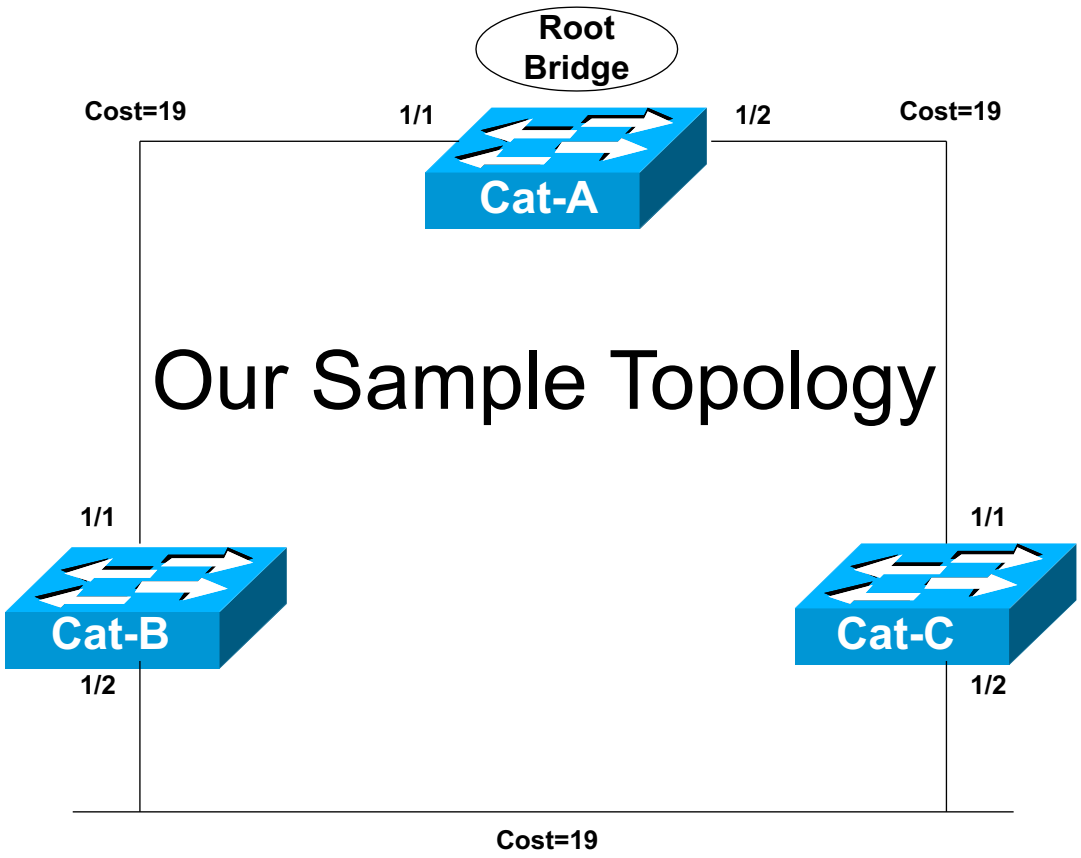
Step 2 Elect Root Ports

Step 3 Elect Designated Ports

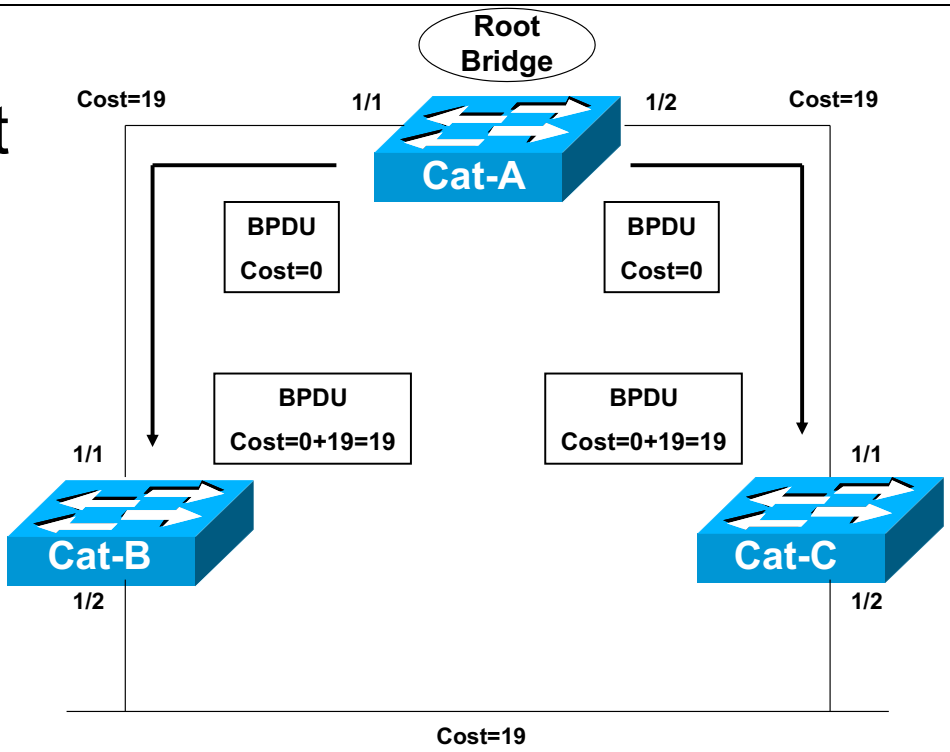
Step 2 Elect Root Ports



- Now that the Root War has been won, switches move on to selecting **Root Ports**.
- A bridge's **Root Port** is the ***port closest to the Root Bridge***.
- Bridges use the **cost** to determine closeness.
- **Every non-Root Bridge will select one Root Port!**
- Specifically, bridges track the **Root Path Cost**, the cumulative cost of all links to the Root Bridge.


Root
Bridge
Cost=19
1/1
1/2
Cost=19
Cat-A
Our Sample Topology
1/1
1/1
Cat-B
Cat-C
1/2
1/2
Cost=19

Step 2 Elect Root Ports



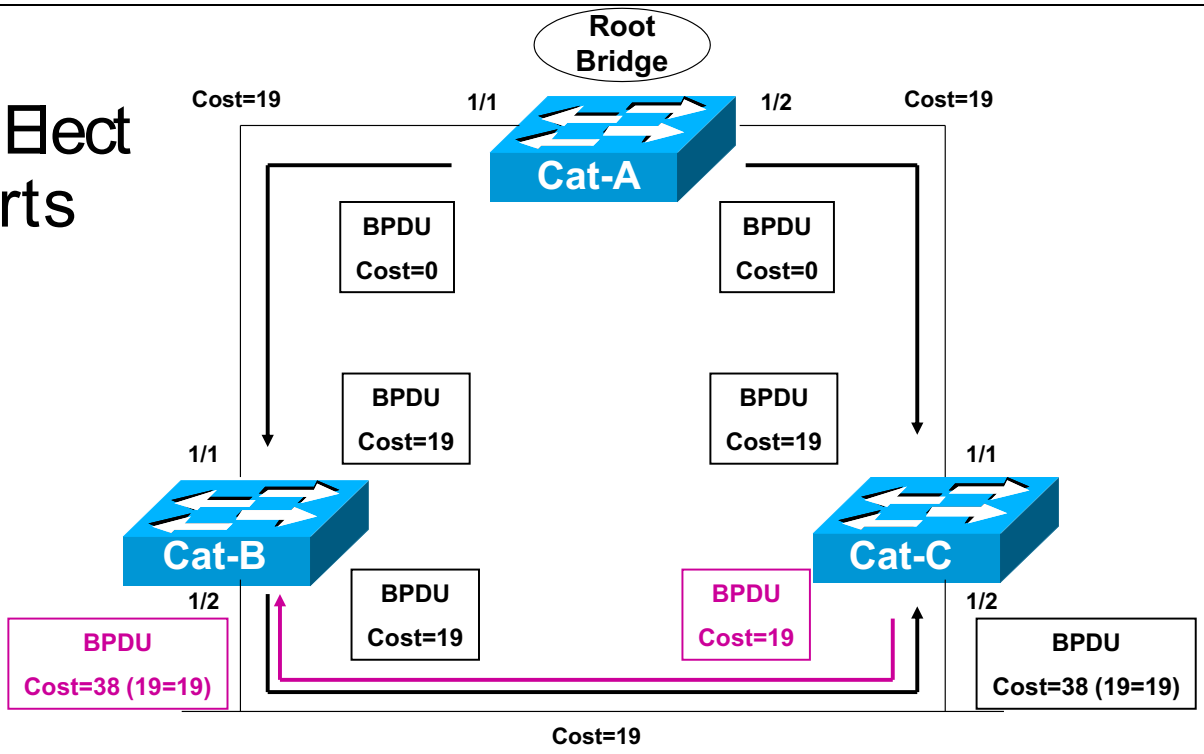
Step 1

- Cat-A sends out BPDUs, containing a Root Path Cost of 0.
- Cat-B receives these BPDUs and adds the Path Cost of Port 1/1 to the Root Path Cost contained in the BPDU.

Step 2

- Cat-B adds Root Path Cost 0 PLUS its Port 1/1 cost of 19 = 19

Step 2 Elect Root Ports



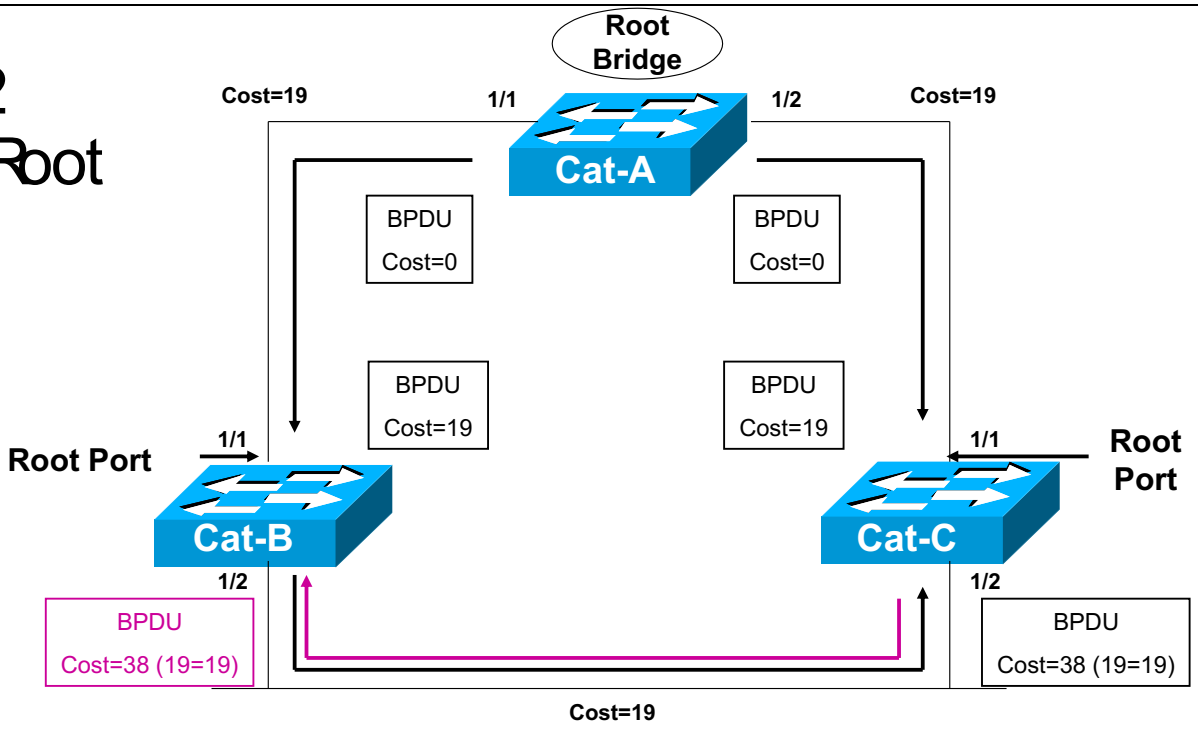
Step 3

- Cat-B uses this value of 19 internally and sends BPDUs with a Root Path Cost of 19 out Port 1/2.

Step 4

- Cat-C receives the BPDU from Cat-B, and increased the Root Path Cost to 38 (19+19). (Same with Cat-C sending to Cat-B.)

Step 2 Elect Root Ports



Step 5

- Cat-B calculates that it can reach the Root Bridge at a cost of 19 via Port 1/1 as opposed to a cost of 38 via Port 1/2.
- Port 1/1 becomes the Root Port for Cat-B, the port closest to the Root Bridge.
- Cat-C goes through a similar calculation. Note: Both Cat-B:1/2 and Cat-C:1/2 save the best BPDU of 19 (its own).

Three Steps of Initial STP Convergence

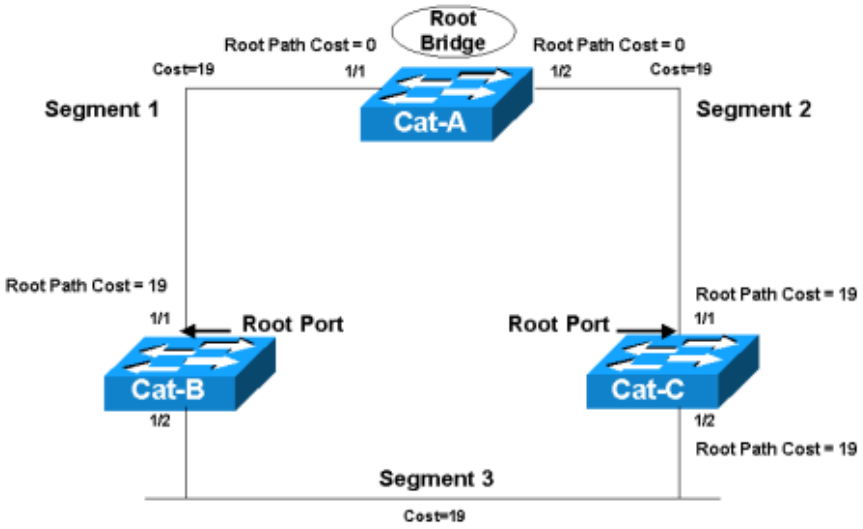
STP Convergence

Step 1 Elect one Root Bridge

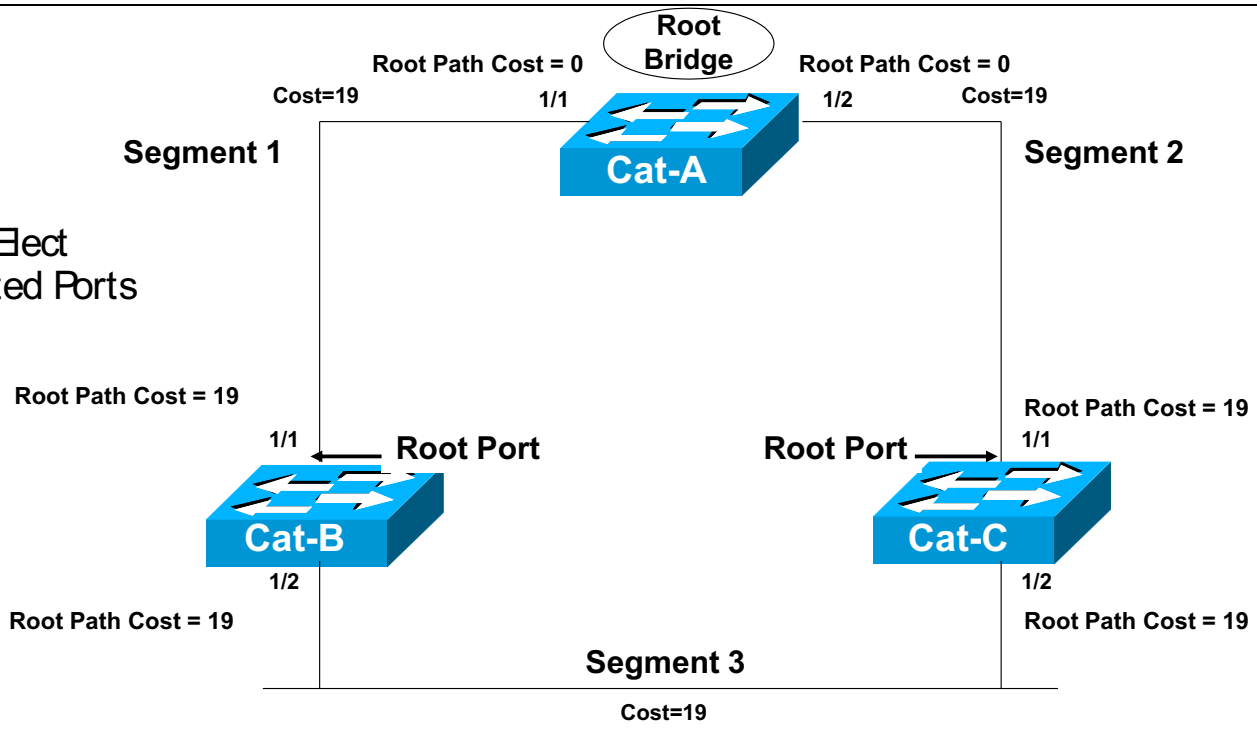
Step 2 Elect Root Ports

Step 3 Elect Designated Ports

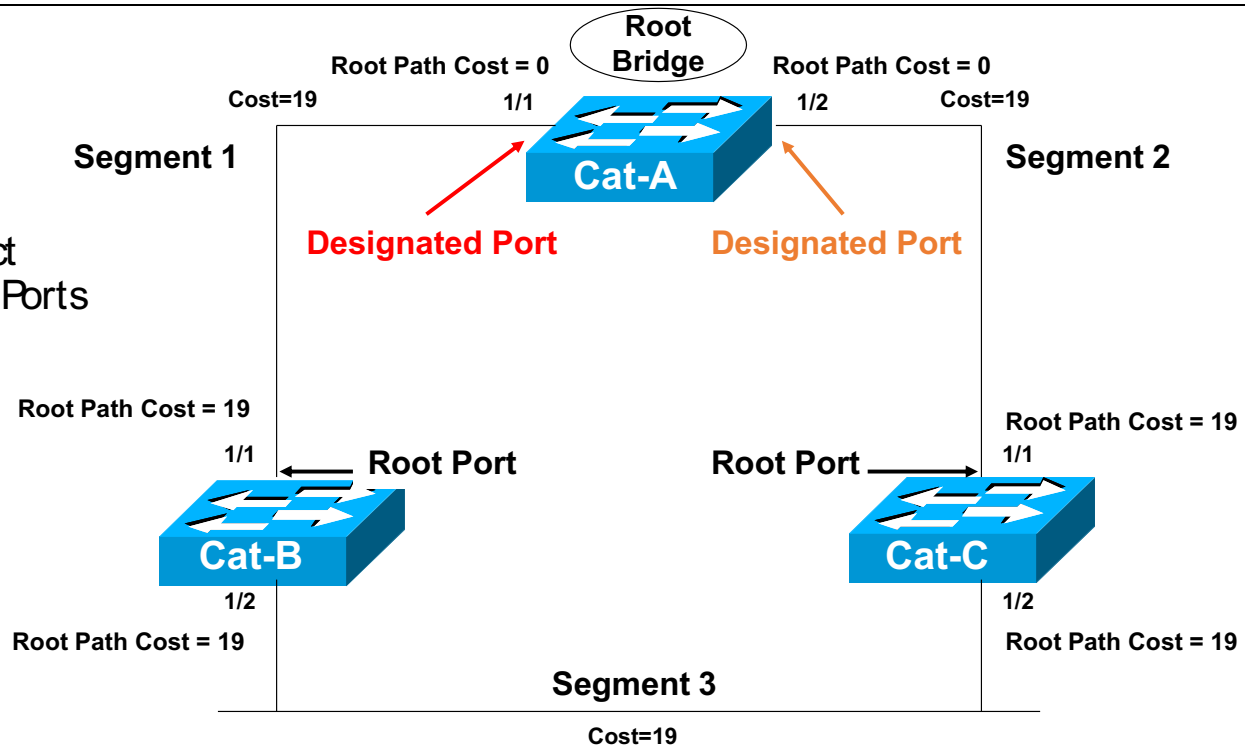
Step 3 Elect Designated Ports



- The loop prevention part of STP becomes evident during this step, electing designated ports.
- A **Designated Port** functions as ***the single bridge port that both sends and receives traffic to and from that segment and the Root Bridge.***
- **Each segment in a bridged network has one Designated Port, chosen based on cumulative Root Path Cost to the Root Bridge.**
- The switch containing the Designated Port is referred to as the **Designated Bridge** for that segment.
- To locate Designated Ports, lets take a look at each segment.
- **Root Path Cost**, the cumulative cost of all links to the Root Bridge.



- **Segment 1**: Cat-A:1/1 has a Root Path Cost = 0 (after all it has the Root Bridge) and Cat-B:1/1 has a Root Path Cost = 19.
- **Segment 2**: Cat-A:1/2 has a Root Path Cost = 0 (after all it has the Root Bridge) and Cat-C:1/1 has a Root Path Cost = 19.
- **Segment 3**: **Cat-B:1/2** has a **Root Path Cost = 19** and **Cat-C:1/2** has a **Root Path Cost = 19**. ***It's a tie!***



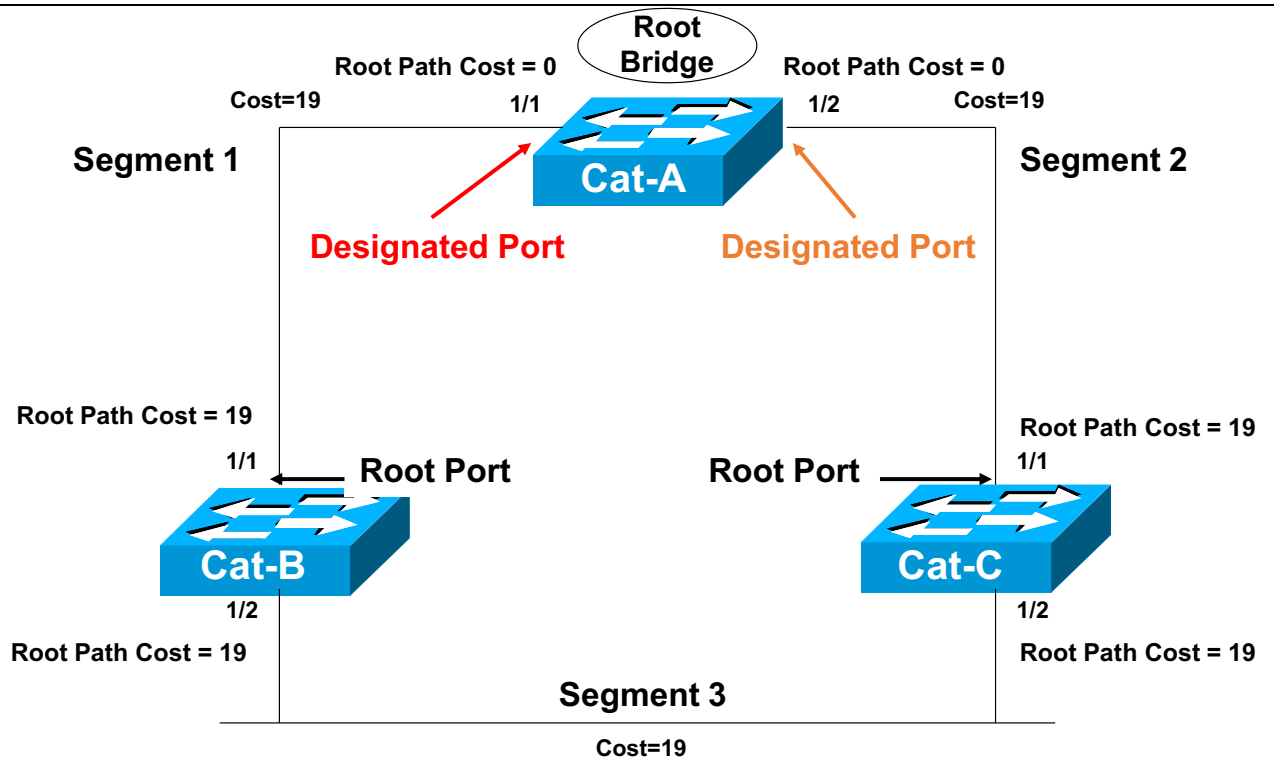
Step 3 Elect Designated Ports

Segment 1

- Because Cat-A:1/1 has the lower Root Path Cost it becomes the **Designate Port for Segment 1**.

Segment 2

- Because Cat-A:1/2 has the lower Root Path Cost it becomes the **Designate Port for Segment 2**.

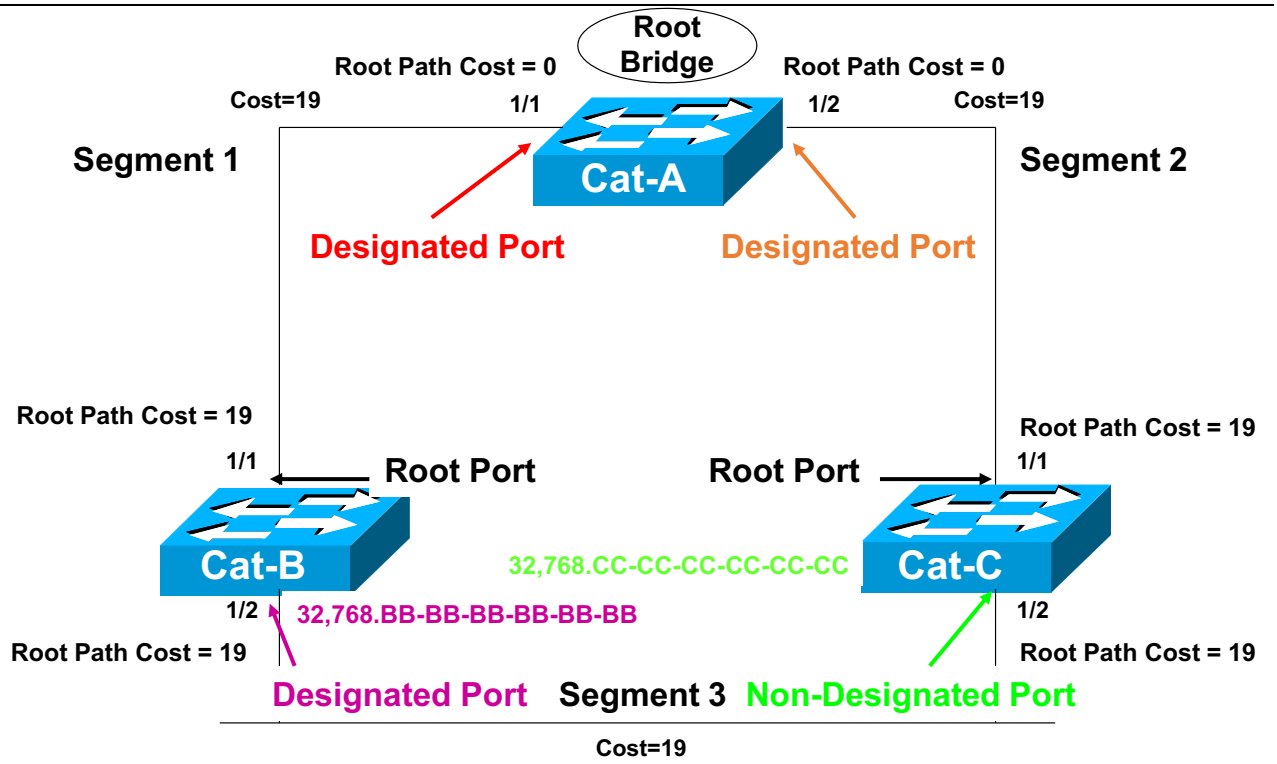


Segment 3

- **Both Cat-B and Cat-C** have a **Root Path Cost of 19**, a tie!
- When faced with a tie (or any other determination) STP always uses the four-step decision process:

  1. Lowest Root BID; 2. Lowest Path Cost to Root Bridge;

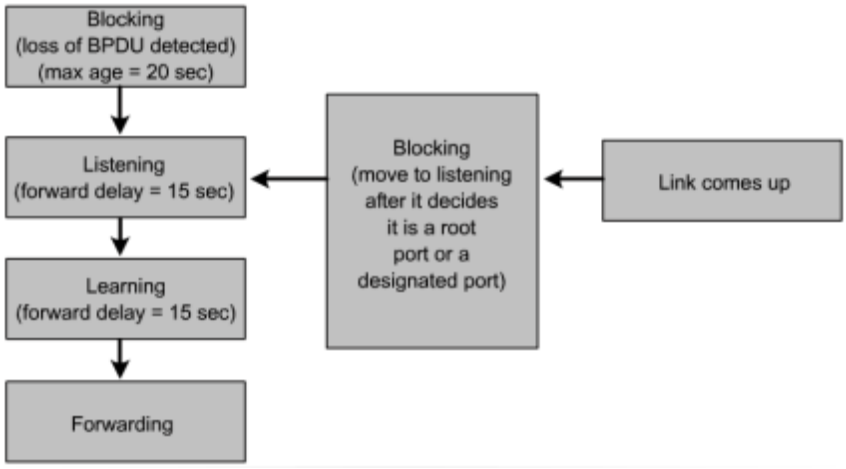
  3. Lowest Sender BID; 4. Lowest Port ID



Segment 3 (continued)

- 1) All three switches agree that Cat-A is the Root Bridge, so this is a tie.
- 2) Root Path Cost for both is 19, also a tie.
- 3) The sender's BID is lower on Cat-B, than Cat-C, so Cat-B:1/2 becomes the **Designated Port for Segment 3**.
- Cat-C:1/2 therefore becomes the **non-Designated Port for Segment 3**.

Stages of spanning-tree port states



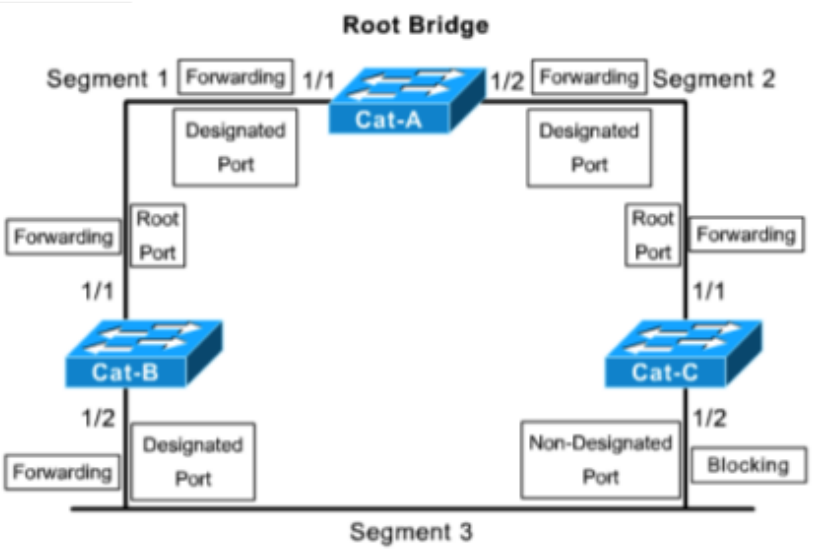
- Time is required for (BPDU) protocol information to propagate throughout a switched network.
- Topology changes in one part of a network are not instantly known in other parts of the network.
- There is propagation delay.
- A switch should not change a port state from inactive (Blocking) to active (Forwarding) immediately, as this may cause data loops.
- Each port on a switch that is using the Spanning-Tree Protocol has one of five states,

Slide 98

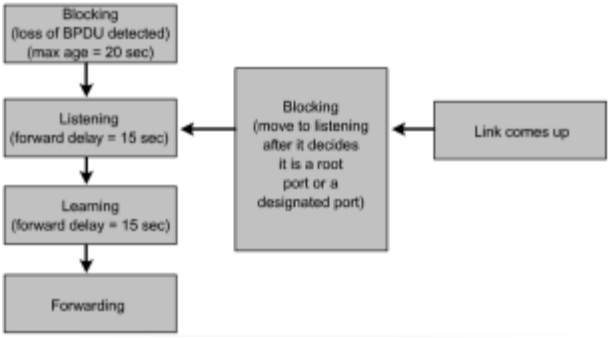
State	Purpose
Forwarding	Sending / receiving user data
Learning	Building bridging table
Listening	Building "active" topology
Blocking	Receives BPDUs only
Disabled	Administratively down

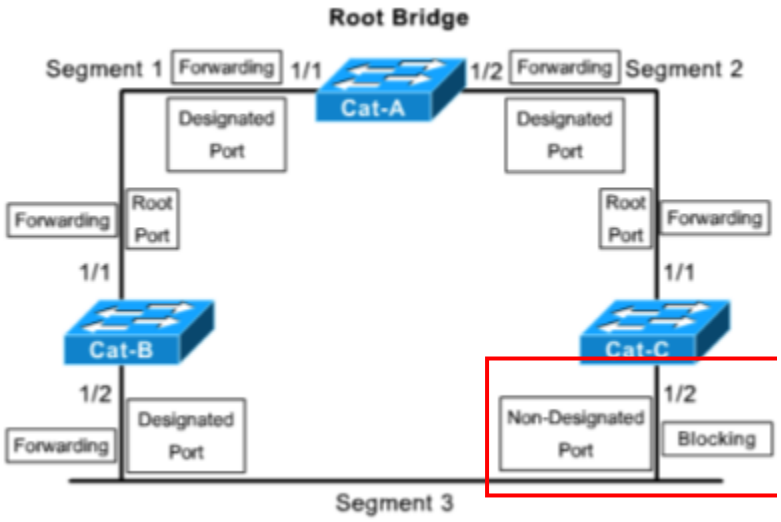
Designated Ports & Root Ports

Non-Designated Ports



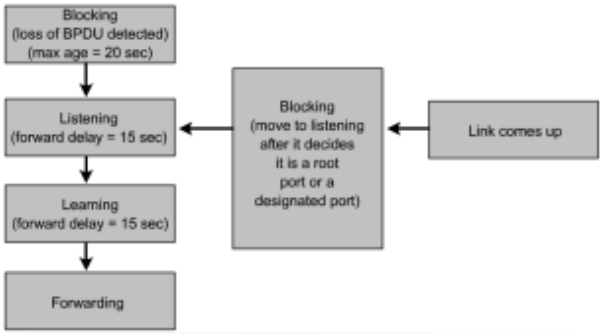
STP Port States

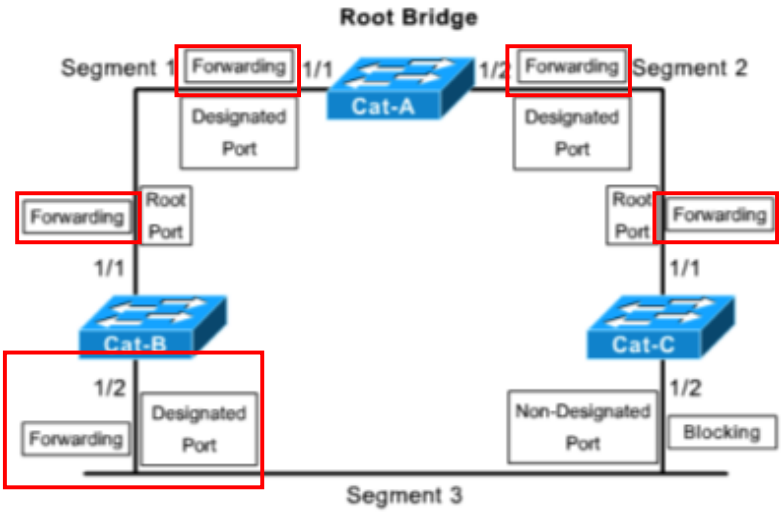




- In the **blocking state**, ports can only receive BPDUs.
  - Data frames are discarded and no addresses can be learned.
  - It may take up to 20 seconds to change from this state.
- Ports go from the blocked state to the **listening state**.
  - Switch **determines if there are any other paths to the root bridge**.
  - The **path that is not the least cost path to the root bridge goes back to the blocked state**.
  - The listening period is called the forward delay and lasts for 15 seconds.
  - In the listening state, user data is not being forwarded and MAC addresses are not being learned.
  - BPDUs are still processed.

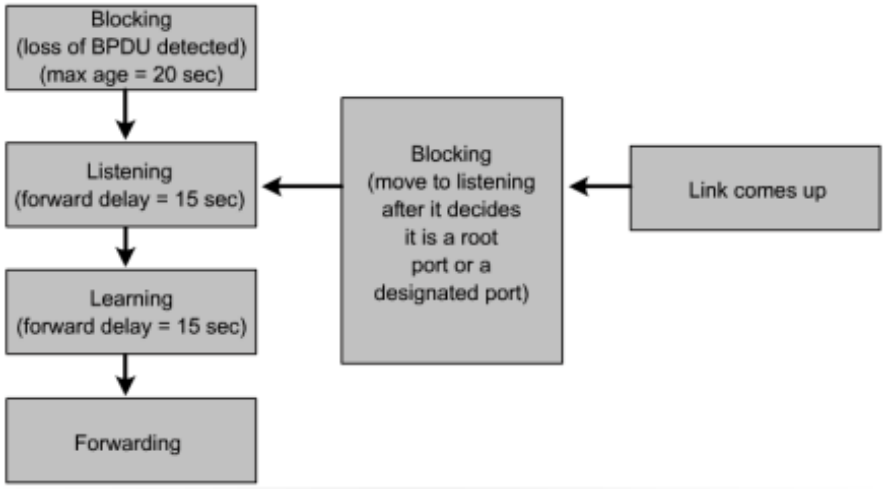
STP Port States



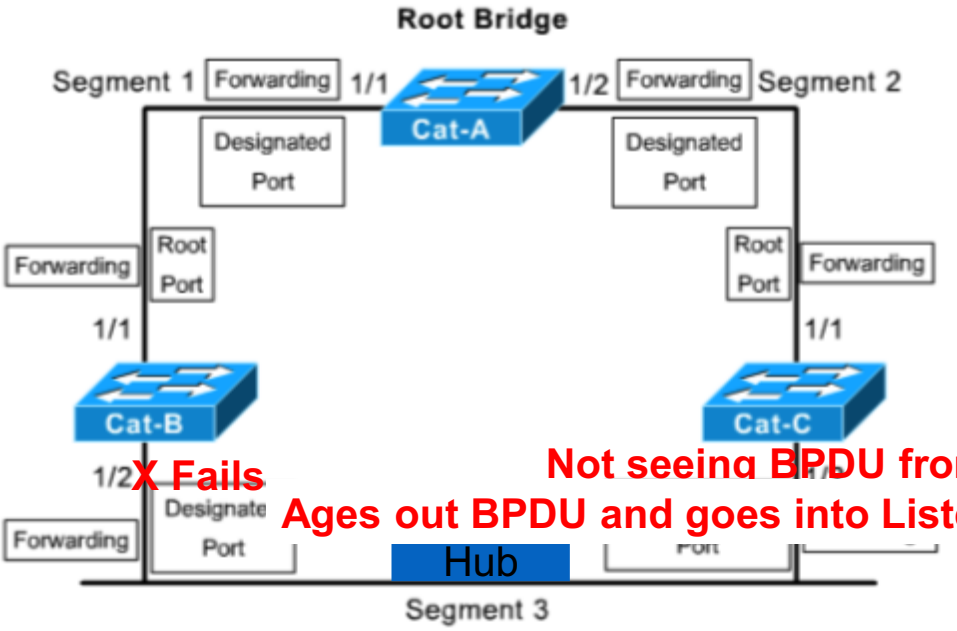


- Ports transition from the listening to the **learning state**.
  - In this state **user data is not forwarded, but MAC addresses are learned** from any traffic that is seen.
  - The learning state lasts for 15 seconds and is also called the forward delay.
  - BPDUs are still processed.
- A port goes from the learning state to the **forwarding state**.
  - In this state user data is forwarded and MAC addresses continue to be learned.
  - BPDUs are still processed.

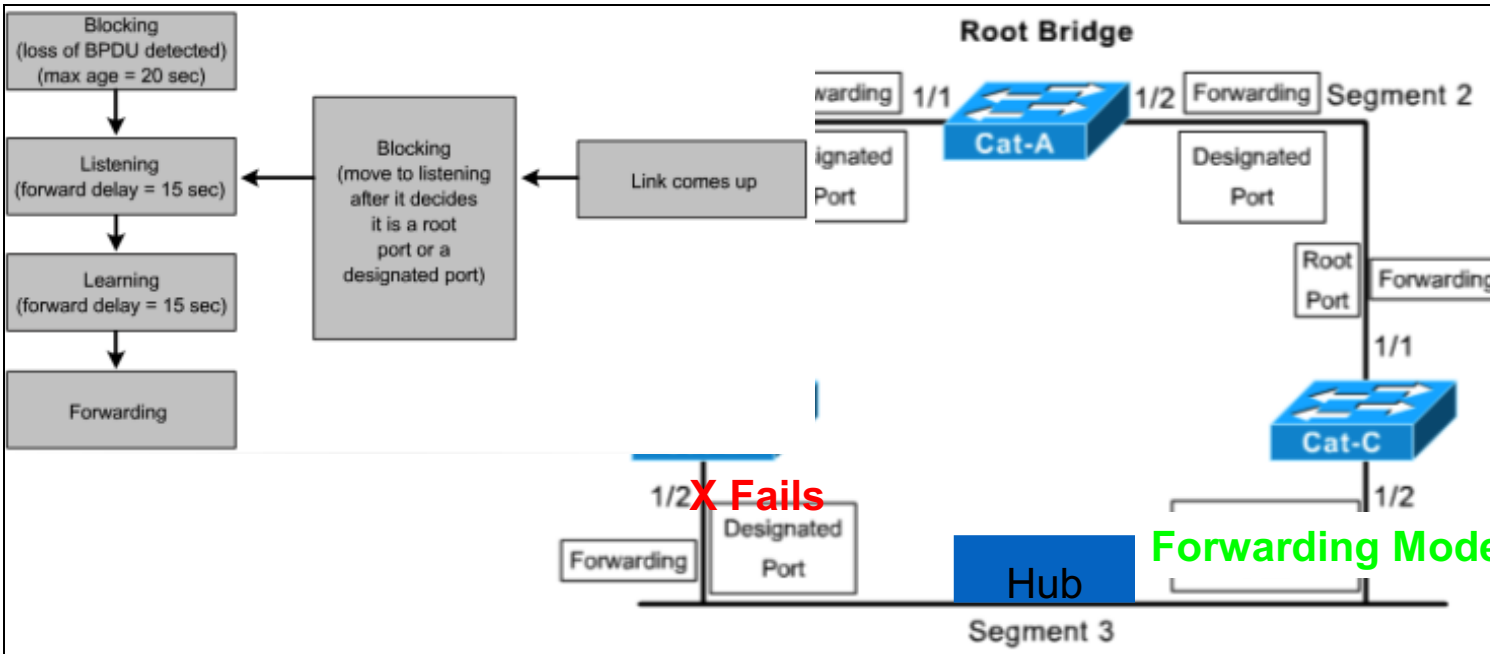
STP Timers



- Some details have been left out, such as timers, STP FSM, etc.
- The time values given for each state are the default values.
- These values have been calculated on an assumption that there will be a maximum of seven switches in any branch of the spanning tree from the root bridge.
- These are discussed in CCNP 3 Multilayer Switching.

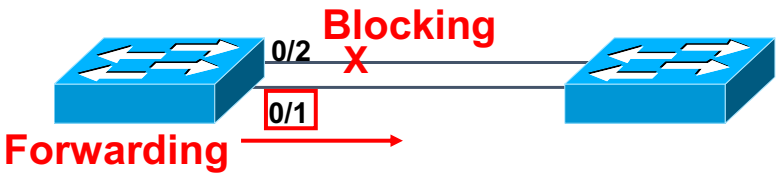


- Cat-B:1/2 fails.
- Cat-C has no immediate notification because it's still receiving a link from the hub.
- Cat-C notices it is not receiving BPDUs from Cat-B.
- **20 seconds (max age)** after the failure, Cat-C ages out the BPDU that lists Cat-B as having the DP for segment 3.
- This causes **Cat-C:1/2 to transition into the Listing state (15 seconds)** in an effort to become the DP.



- Because Cat-C:1/2 now offers the most attractive access from the Root Bridge to this link, it **eventually transitions to Learning State (15 seconds), then all the way into Forwarding mode.**
- In practice this will take **50 seconds (20 max age + 15 Listening + 15 Learning)** for Cat-C:1/2 to take over after the failure of Cat-B:1/2.

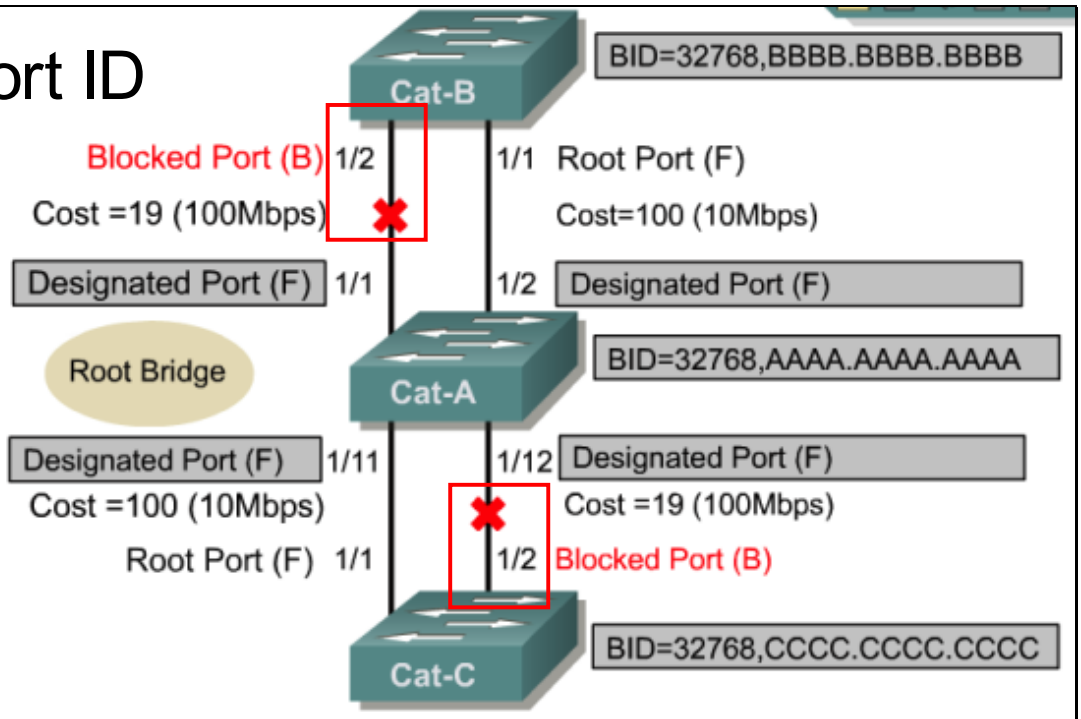
Port Cost/Port ID



Assume path cost and port priorities are default (32). Port ID used in this case. Port 0/1 would forward because it's the lower than Port 0/2.

- If the path cost and bridge IDs are equal (as in the case of parallel links), the switch goes to the port priority as a tiebreaker.
- Lowest port priority wins (all ports set to 32).
- You can set the priority from 0 – 63.
- If all ports have the same priority, the port with the lowest port number forwards frames.

Port Cost/Port ID



- If all ports have the same priority, the port with the lowest port number forwards frames.

STP Convergence Recap

- Recall that switches go through three steps for their initial convergence:

STP Convergence
Step 1 Elect one Root Bridge
Step 2 Elect Root Ports
Step 3 Elect Designated Ports

- Also, all STP decisions are based on a the following predetermined sequence:

Four-Step decision Sequence

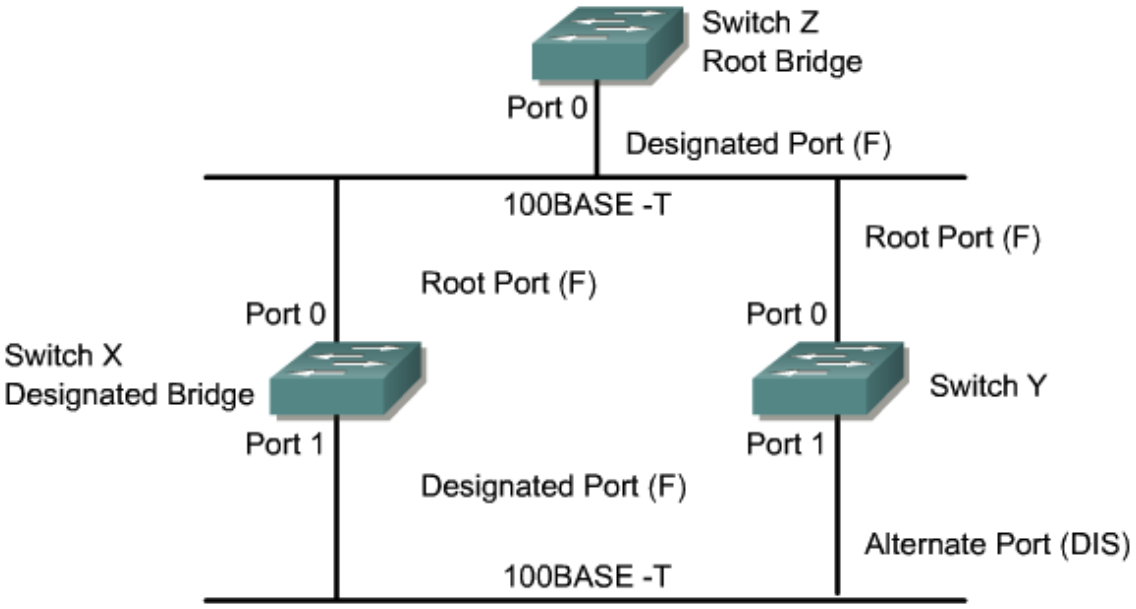
Step 1 - Lowest BID

Step 2 - Lowest Path Cost to Root Bridge

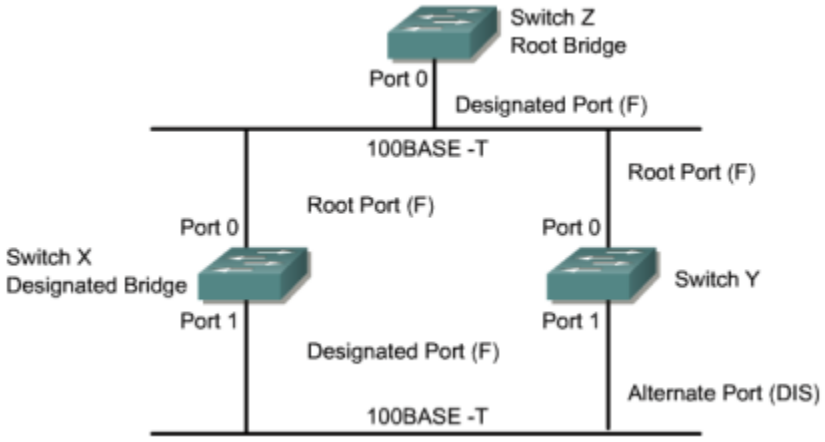
Step 3 - Lowest Sender BID

Step 4 - Lowest Port ID

Rapid Spanning Tree Protocol (RSTP)

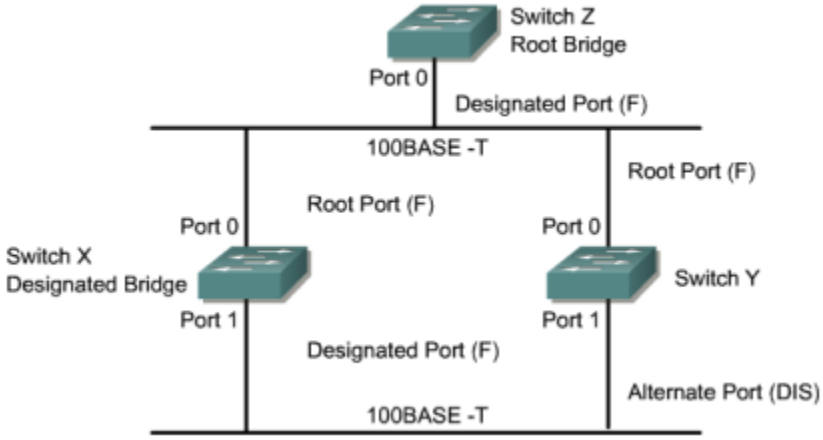


Rapid Spanning Tree Protocol (RSTP)



- The Rapid Spanning-Tree Protocol is defined in the IEEE 802.1w LAN standard. The standard and protocol introduce the following:
  - Clarification of port states and roles
  - Definition of a set of link types that can go to forwarding state rapidly
  - Concept of allowing switches, in a converged network, to generate their own BPDUs rather than relaying root bridge BPDUs
- The “blocked” state of a port has been renamed as the “discarding” state.

RSTP Link Types

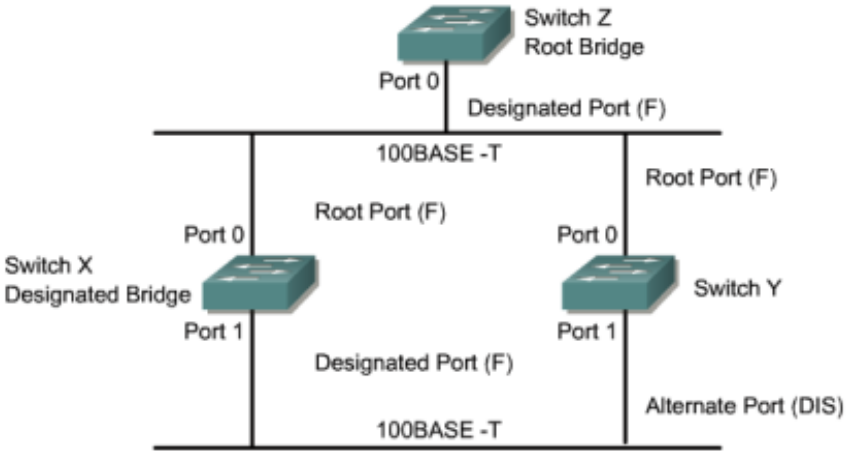


- Link types have been defined as point-to-point, edge-type, and shared.
- These changes allow failure of links in switched network to be learned rapidly.
- Point-to-point links and edge-type links can go to the forwarding state immediately.
- Network convergence does not need to be any longer than 15 seconds with these changes.
- The Rapid Spanning-Tree Protocol, IEEE 802.1w, will eventually replace the Spanning-Tree Protocol, IEEE 802.1D

RSTP Port States

STP (802.1D) Port State	RSTP (802.1w) Port State	Is Port Included in Active Topology?	Is Port Learning Mac Addresses?
Disabled	Discarding	No	No
Blocking	Discarding	No	No
Listening	Discarding	No	No
Learning	Learning	No	Yes
Forwarding	Forwarding	Yes	Yes

RSTP Port Roles



- The role is now a variable assigned to a given port.
- The root port and designated port roles remain.
- The blocking port role is now split into the **backup** and **alternate** port roles.
- The Spanning Tree Algorithm (STA) determines the role of a port based on Bridge Protocol Data Units (BPDUs).
- To keep things simple, the thing to remember about a BPDU is that there is always a way of comparing any two of them and deciding whether one is more useful than the other.
- This is based on the value stored in the BPDU and occasionally on the port on which they are received.

Rapid Spanning Tree Protocol (RSTP)

Evolution of STP

Cisco's Implementation	Spanning Tree Protocol Process	IEEE Standard
Spanning Tree Protocol (STP): • 802.1D • Common Spanning Tree (CST) • Mono Spanning Tree (MST)	ST = Spanning Tree 	
Cisco Enhancements (First evolution): • Portfast • Uplinkfast • Backbonefast		RSTP: • 802.1w • Edge Fast (Cisco Port Fast) • Uplink Fast RSTP (Cisco Uplink Fast) • Backbone Fast Engine (Cisco Backbone Fast)
Cisco Enhancements (Second Evolution): • PVST: ISL • PVST+: ISL & 802.1Q • Includes previous enhancements • Additional enhancements: o BPDU Guard o Root Guard		
Cisco MISTP: • Uses PVST+ • Includes previous enhancements • Catalyst 4000/6000		MST (Multiple Spanning Tree): • 802.1s • Uses RSTP

- RSTP adds features to the standard similar to vendor proprietary features including Cisco's Port Fast, Uplink Fast and Backbone Fast.
- Cisco recommends that administrators upgrade to the IEEE 802.1w standard when possible.

Cisco's Port Fast and RSTP's Edge Fast

- A common problem is with DHCP and STP Port States.
- The workstation will power up and start looking for a DHCP servers before its port has transitioned to Forwarding State.
- The workstation will not be able to get a valid IP address, and may default to an IP address such as 169.x.x.x.
- Spanning-tree PortFast causes a port to enter the spanning-tree forwarding state immediately, bypassing the listening and learning states.
- You can use PortFast on switch ports connected to a single workstation or server to allow those devices to connect to the network immediately, instead of waiting for the port to transition from the listening and learning states to the forwarding state.
- **Caution** PortFast should be used *only* when connecting a single end station to a switch port.
  - If you enable PortFast on a port connected to another networking device, such as a switch, you can create network loops.

Defining VLANs

- Broadcast domain consisting of a group of end stations not limited by physical location and communicate as if they were on a common LAN
- Membership by:
  - Port group
  - MAC address
  - Protocol information

A VLAN is a broadcast domain consisting of a group of end stations, perhaps on multiple physical LAN segments, that are not constrained by their physical location and can communicate as if they were on a common LAN. Some means is therefore needed for defining VLAN membership. A number of different approaches have been used for defining membership, including the following:

Membership by port group: Each switch in the LAN configuration contains two types of ports: a trunk port, which connects two switches, and an end port, which connects the switch to an end system. A VLAN can be defined by assigning each end port to a specific VLAN. This approach has the advantage that it is relatively easy to configure. The principle disadvantage is that the network manager must reconfigure VLAN membership when an end system moves from one port to another.

Membership by MAC address: Since MAC-layer addresses are hard-wired into the workstation's network interface card (NIC), VLANs based on MAC addresses enable network managers to move a workstation to a different physical location on the network and have that workstation automatically retain its VLAN membership. The main problem with this method is that VLAN membership must be assigned initially. In networks with thousands of users, this is no easy task. Also, in environments where notebook PCs are used, the MAC address is associated with the docking station and not with the notebook PC. Consequently, when a notebook PC is moved to a different docking station, its VLAN membership must be reconfigured.

Membership based on protocol information: VLAN membership can be assigned based on IP address, transport protocol information, or even higher-layer protocol information. This is a quite flexible approach, but it does require switches to examine portions of the MAC frame above the MAC layer, which may have a performance impact.

Communicating VLAN Membership

Switches need to know VLAN membership

- Configure information manually
- Network management signaling protocol
- Frame tagging (IEEE802.1Q)

Switches must have a way of understanding VLAN membership (that is, which stations belong to which VLAN) when network traffic arrives from other switches; otherwise, VLANs would be limited to a single switch. One possibility is to configure the information manually or with some type of network management signaling protocol, so that switches can associate incoming frames with the appropriate VLAN.

A more common approach is frame tagging, in which a header is typically inserted into each frame on interswitch trunks to uniquely identify to which VLAN a particular MAC-layer frame belongs. The IEEE 802 committee has developed a standard for frame tagging, IEEE 802.1Q, which we examine in the next chapter.

Switching Concepts

STP States

States	Purpose
Blocking	Receives BPDUs only
Listening	Building "active" topology
Learning	Building bridging table
Forwarding	Sending and receiving user data
Disabled	Administratively down

Slide 118

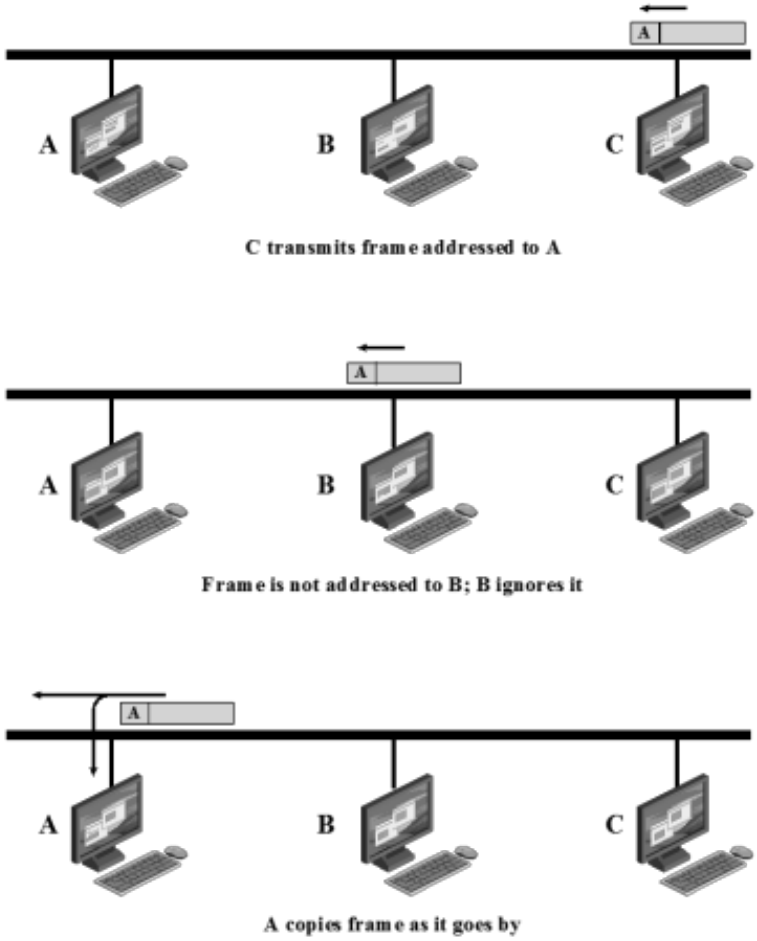


Figure 11.1 Frame Transmission on a Bus LAN

Figure 11.1 illustrates the bus scheme. In this example, station C wishes to transmit a frame of data to A. The frame header includes A's address. As the frame propagates along the bus, it passes B. B observes the address and ignores the frame. A, on the other hand, sees that the frame is addressed to itself and therefore copies the data from the frame as it goes by.

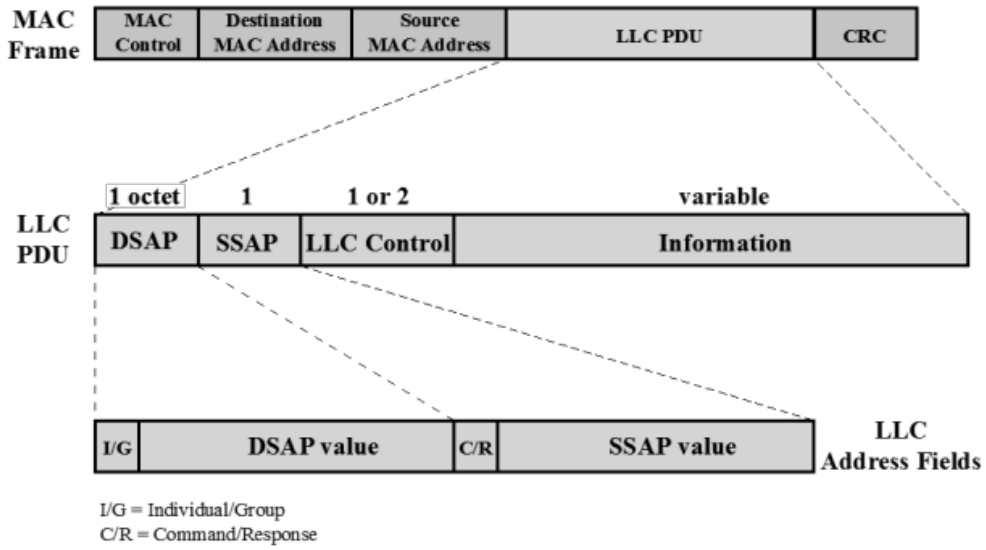


Figure 11.5 LLC PDU in a Generic MAC Frame Format

All three LLC protocols employ the same PDU format (Figure 11.5), which consists of four fields. The DSAP (Destination Service Access Point) and SSAP (Source Service Access Point) fields each contain a 7-bit address, which specifies the destination and source users of LLC. One bit of the DSAP indicates whether the DSAP is an individual or group address. One bit of the SSAP indicates whether the PDU is a command or response PDU. The format of the LLC control field is identical to that of HDLC (Figure 7.7), using extended (7-bit) sequence numbers.

For type 1 operation , which supports the unacknowledged connectionless service, the unnumbered information (UI) PDU is used to transfer user data. There is no acknowledgment, flow control, or error control. However, there is error detection and discard at the MAC level.

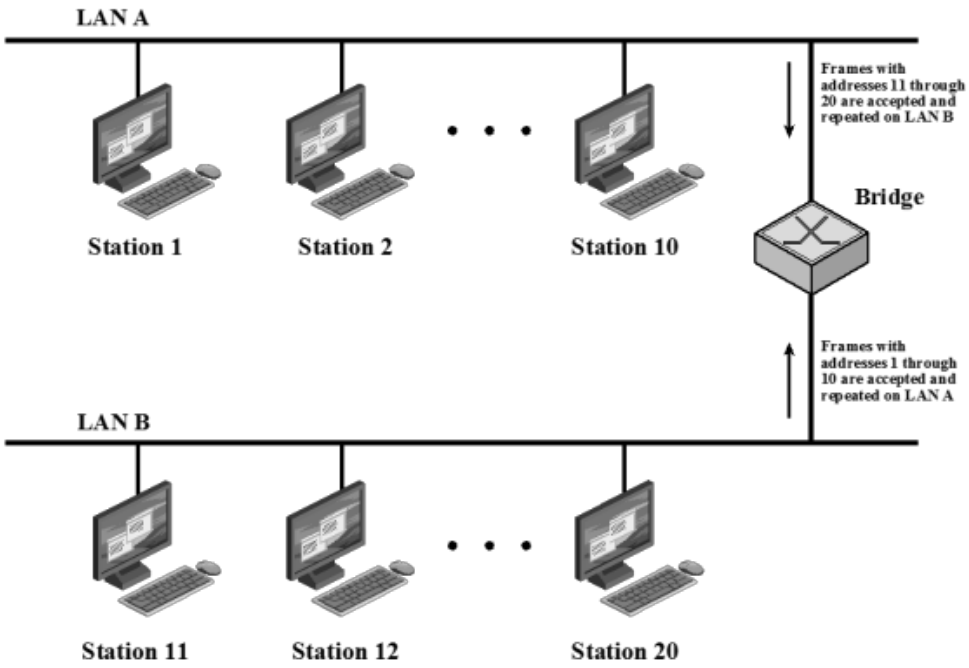


Figure 11.6 Bridge Operation

Figure 11.6 illustrates the action of a bridge connecting two LANs, A and B, using the same MAC protocol. In this example, a single bridge attaches to both LANs; frequently, the bridge function is performed by two "halfbridges," one on each LAN. The functions of the bridge are few and simple:

- Read all frames transmitted on A and accept those addressed to any station on B.
- Using the medium access control protocol for B, retransmit each frame on B.
- Do the same for B-to-A traffic.

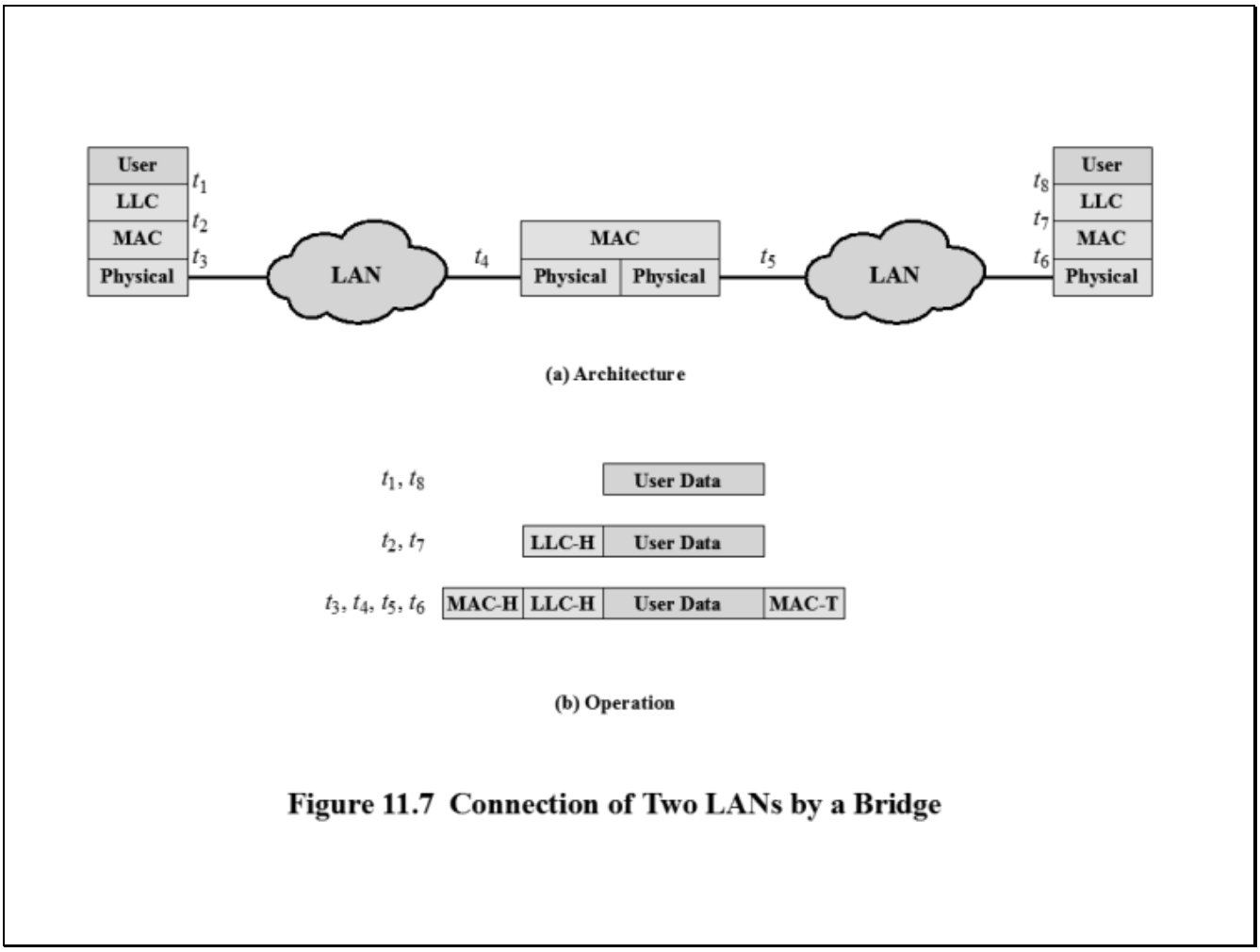


Figure 11.7 Connection of Two LANs by a Bridge

The IEEE 802.1D specification defines the protocol architecture for MAC bridges. Within the 802 architecture, the endpoint or station address is designated at the MAC level. Thus, it is at the MAC level that a bridge can function. Figure 11.7 shows the simplest case, which consists of two LANs connected by a single bridge. The LANs employ the same MAC and LLC protocols. The bridge operates as previously described. A MAC frame whose destination is not on the immediate LAN is captured by the bridge, buffered briefly, and then transmitted on the other LAN. As far as the LLC layer is concerned, there is a dialogue between peer LLC entities in the two endpoint stations. The bridge need not contain an LLC layer because it is merely serving to relay the MAC frames.

Figure 11.7b indicates the way in which data are encapsulated using a bridge. Data are provided by some user to LLC. The LLC entity appends a header and passes the resulting data unit to the MAC entity, which appends a header and a trailer to form a MAC frame. On the basis of the destination MAC address in the frame, it is captured by the bridge. The bridge does not strip off the MAC fields; its function is to relay the MAC frame intact to the destination LAN. Thus, the frame

is deposited on the destination LAN and captured by the destination station.

The concept of a MAC relay bridge is not limited to the use of a single bridge to connect two nearby LANs. If the LANs are some distance apart, then they can be connected by two bridges that are in turn connected by a communications facility. The intervening communications facility can be a network, such as a wide area packet-switching network, or a point-to-point link. In such cases, when a bridge captures a MAC frame, it must encapsulate the frame in the appropriate packaging and transmit it over the communications facility to a target bridge. The target bridge strips off these extra fields and transmits the original, unmodified MAC frame to the destination station.