IK2204

Layer 2 Issues

Bridges, Hubs, Switches, VLANs,
Spanning Tree...

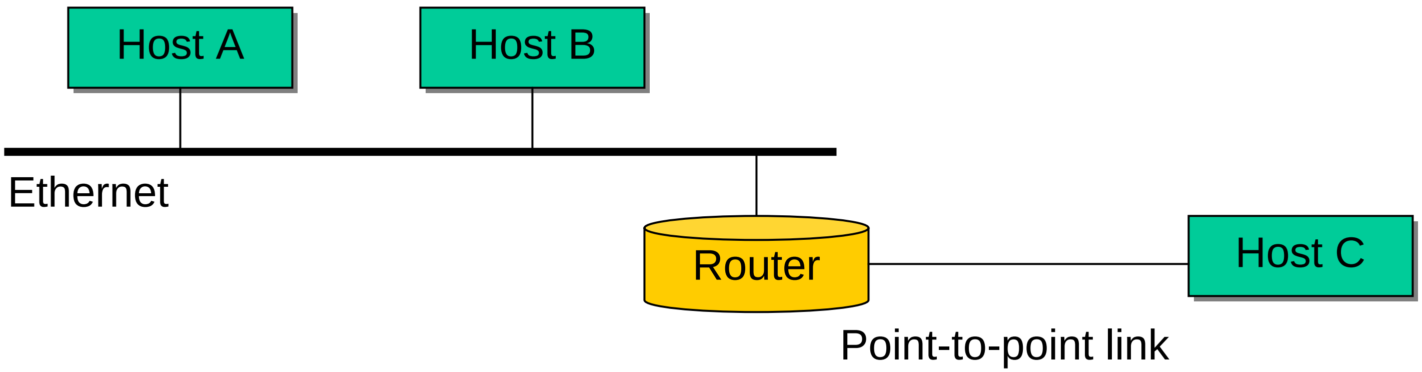
Contents

- Perlman, chapter 3 and 5
- Video lectures no. 2-11
- And then some.....

Different Link Types

- Point-to-point
  - A link that connects exactly two interfaces.
  - Dial-up, V.35, etc.
- Multicast (broadcast)
  - a link that supports a native mechanism at the link layer for sending packets to all (i.e., broadcast) or a subset of all neighbors.
  - Ethernet, FDDI, Wireless, etc.
- Non-Broadcast Multiple Access
  - More than two interfaces can attach, but no native form of multicast or broadcast.
  - ATM, X.25, Frame Relay, etc.

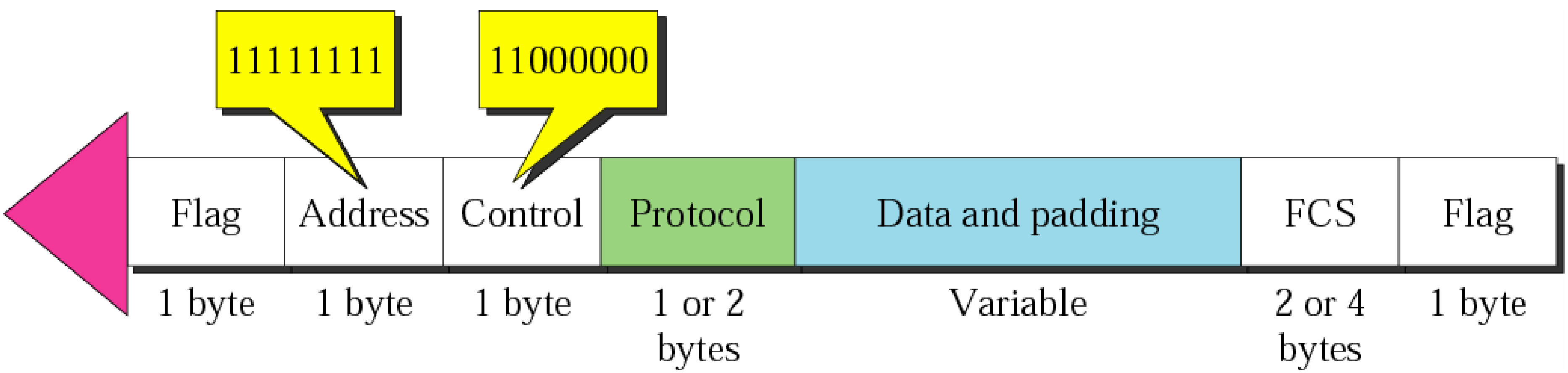
Point-to-Point Links



- Point-to-point link connects exactly 2 interfaces
- Widespread point-to-point protocols:
  - SLIP (CSLIP) for transmission of IP datagrams over serial lines
  - PPP (Point-to-Point Protocol)

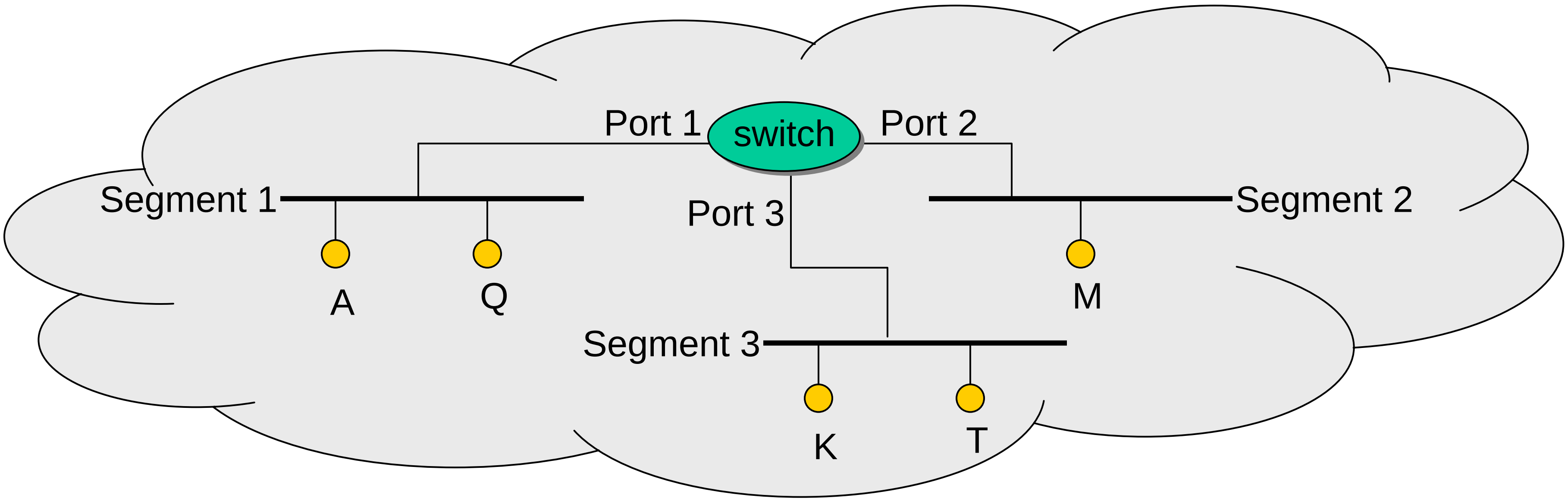
PPP (RFC 1331, 1332)

- Link Control Protocol (LCP) to establish, configure, and test data-links (includes option negotiation)
- Network Control Protocol (NCP) to carry data, e.g., IP (RFC 1332)
- CRC check on every frame
- Dynamic negotiation of IP address of each end



©The McGraw-Hill Companies, Inc., 2000

Ethernet LAN



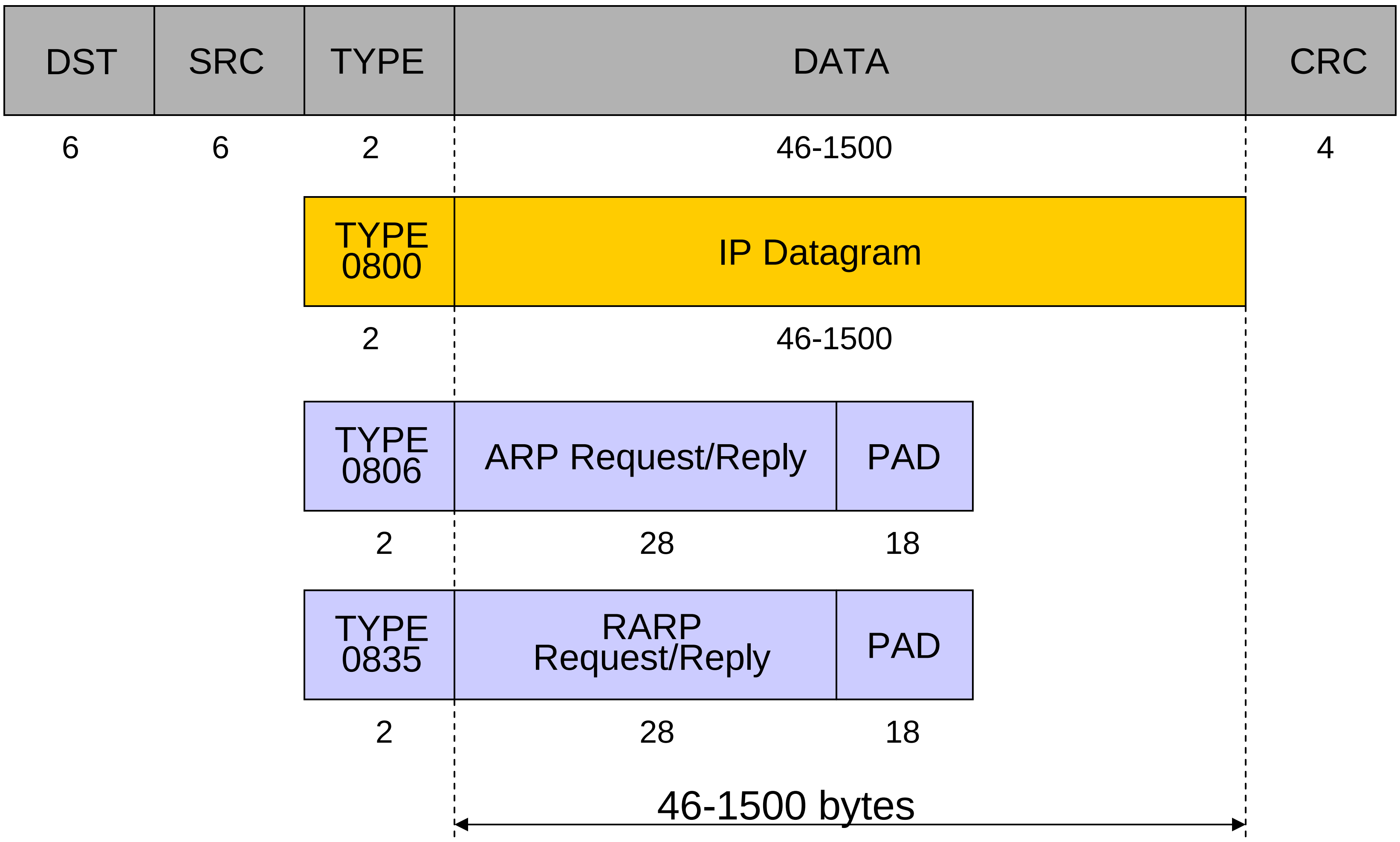
- Ethernet LAN normally built with *switches* and *hubs*
- Nodes attached to a hub are on the same *segment* (L1)
  - No hubs shown in figure
- Ethernet is a broadcast link technology
- A LAN is a *broadcast domain*

Ethernet Frame (DIX)

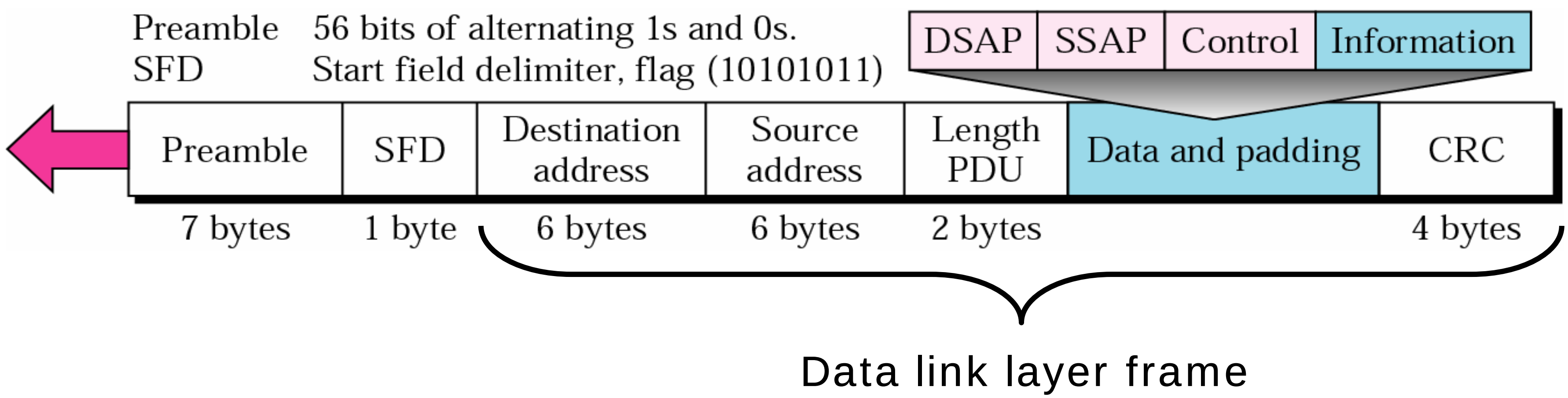
DST	SRC	TYPE	DATA	CRC
6	6	2	46-1500	4

- DIX: Digital, Intel, Xerox
- DST: Destination Address, 48-bit MAC address
- SRC: Source Address, 48-bit MAC address
- Type: Type of data carried, e.g., IP packet
- Data: min size is 46 bytes, max size is 1500 bytes
- CRC: Cyclic Redundancy Check (or FCS – Frame Check Sum)
- 48-bit MAC address is normally written 12:34:56:78:9a:bc
  - Multicast (8th bit is 1), unicast, or broadcast (all bits are 1)

Ethernet Encapsulation (RFC 894)



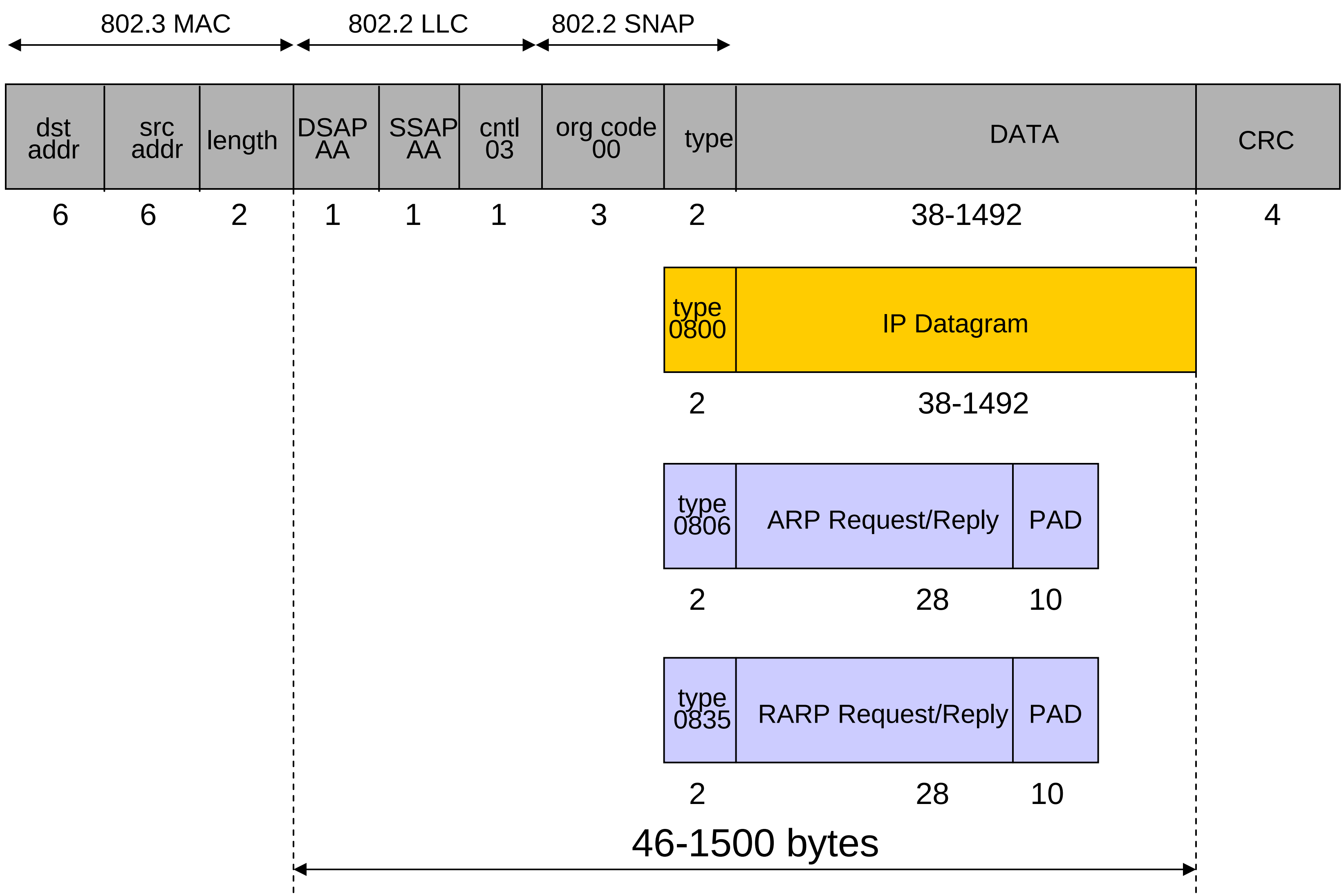
IEEE 802.3 MAC Frame Format



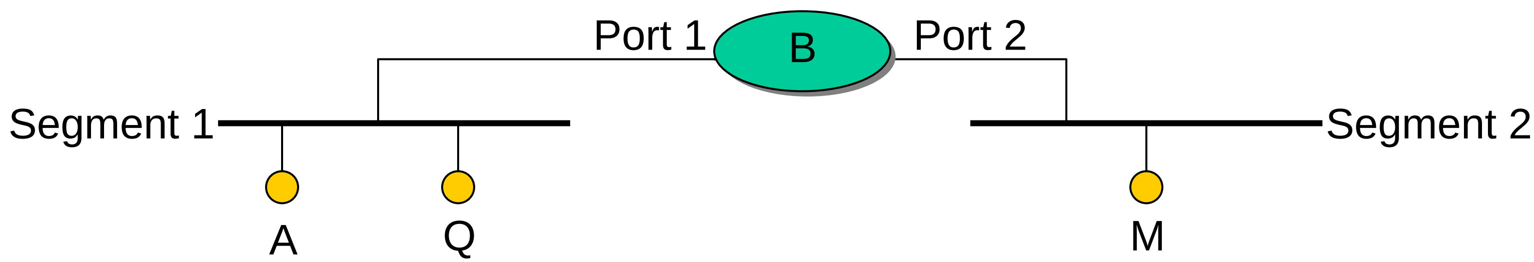
©The McGraw-Hill Companies, Inc., 2000

- 18 bytes of overhead at link layer (26 bytes on the wire)
- Length/PDU
  - Length if less than 1518
    - IEEE 802.3 format
  - Otherwise PDU type
    - RFC 894

IEEE 802.3 Encapsulation (RFC 1042)



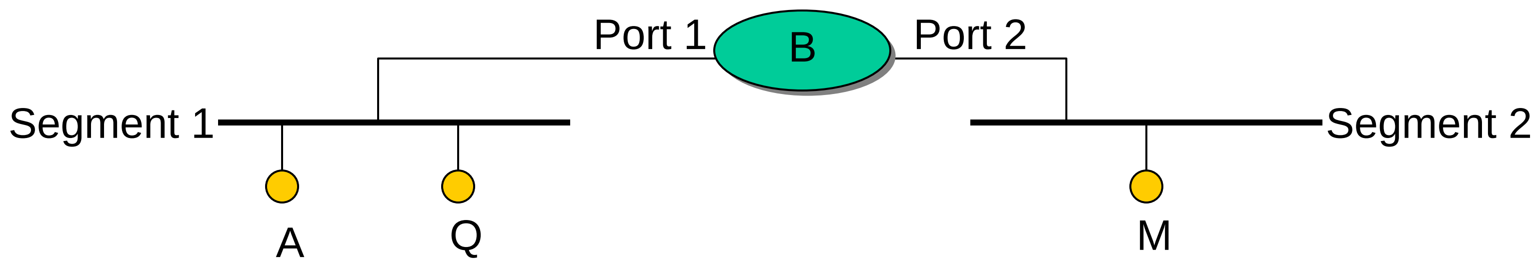
Bridges



- A bridge operates on the link layer—buffers frames
- A bridge does not modify the structure or content of a frame
- A bridge is transparent, i.e., invisible to the clients
- A *multiport* bridge can connect many segments—a *switch*
- A bridge has a *forwarding table*—to map addresses to ports

Address	**Port**
A	1
Q	1
M	2

Standard Bridges



Standard bridge compared to a ”no-frills” bridge:

- Learn which stations that belong to each port
  - Automatically
- Avoid loops
  - Automatically through the Spanning Tree Protocol

To conform to the 802.1D standard, a bridge must implement

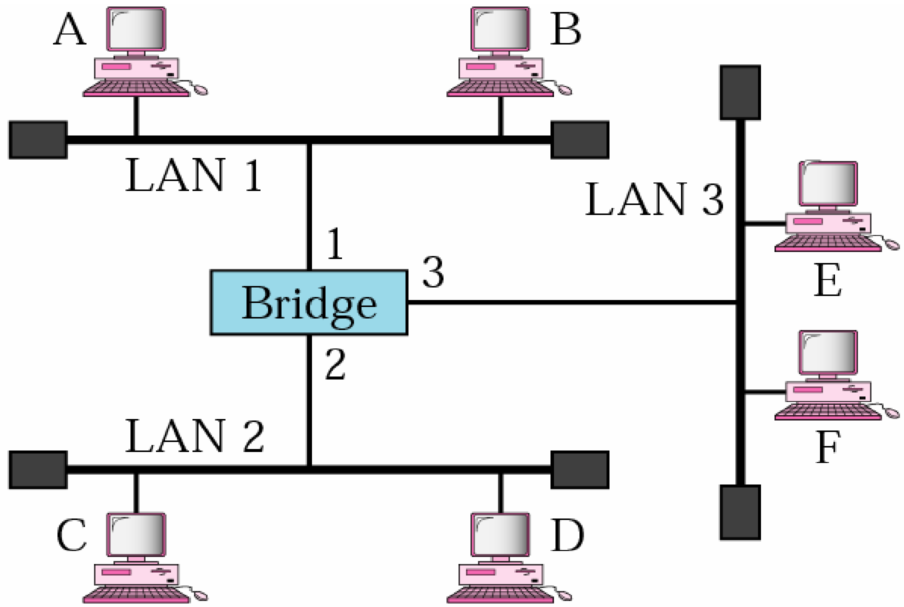
- Learning
- Spanning tree

Learning Bridges

Strategy of a learning bridge:

1. Listen promiscously, receiving every packet
2. Store the source address of every packet in a forwarding table (or *learning table)*
3. For each packet, check destination address with learning table
   - If not found, *flood*: forward on all (except receiving) ports
   - If addr found, forward only on port specified in learning table.
     If specified port == receiving port, drop packet
4. Bridge ages each entry in the learning table.
   - Deletes entry if no traffic is received from that source after a certain period of time

Learning Bridges Example



Address	Port

a. Original

Address	Port
A	1

b. After A sends a frame to D

Address	Port
A	1
E	3

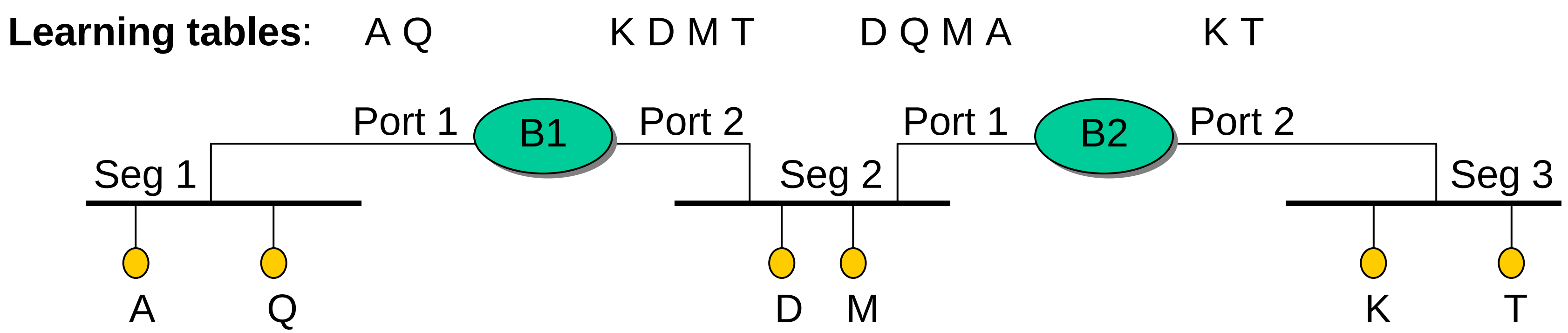
c. After E sends a frame to A

Address	Port
A	1
E	3
B	1

d. After B sends a frame to C

©The McGraw-Hill Companies, Inc., 2000

Learning Bridges—Observations



B1 cannot distinguish between Segment 2 and Segment 3

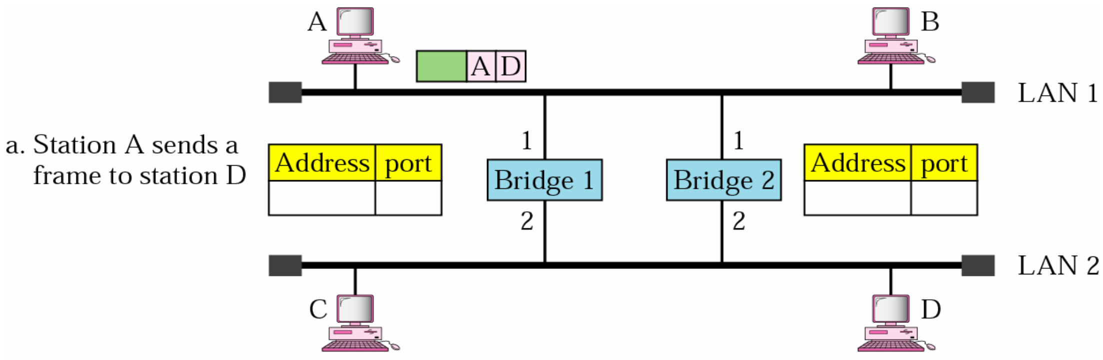
B2 cannot distinguish between Segment 1 and Segment 2

Bridges are not aware of the existence of other bridges

However,

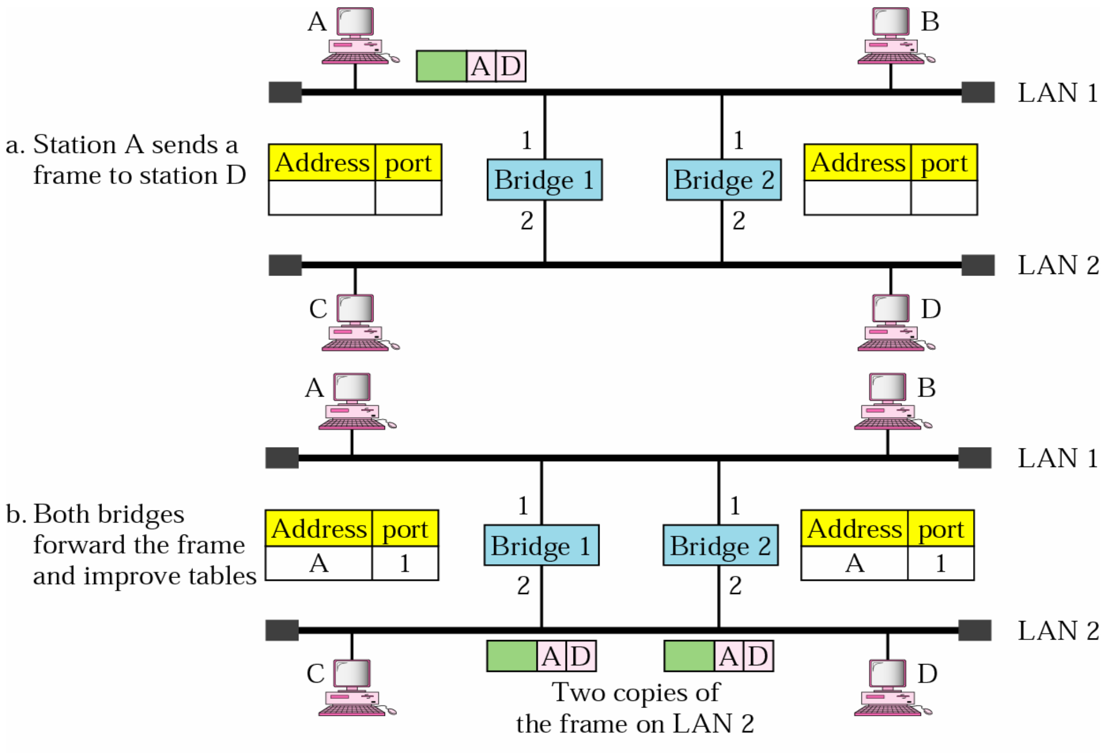
- Bridge concept works for any number of ports
- Bridge concept works for any *tree* (loop-free) topology

Learning Bridges—Loop Problem

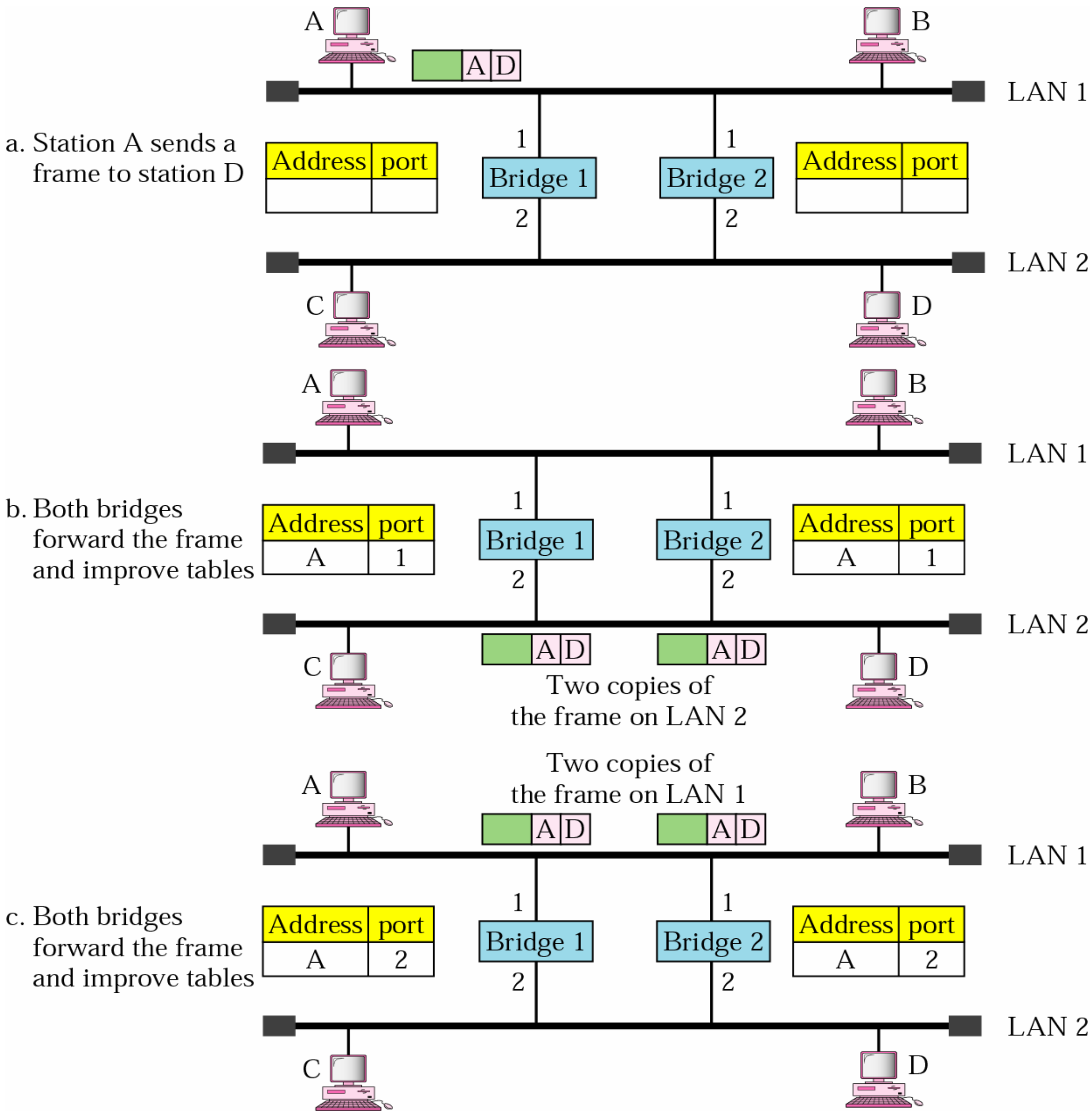


©The McGraw-Hill Companies, Inc., 2000

Learning Bridges—Loop Problem

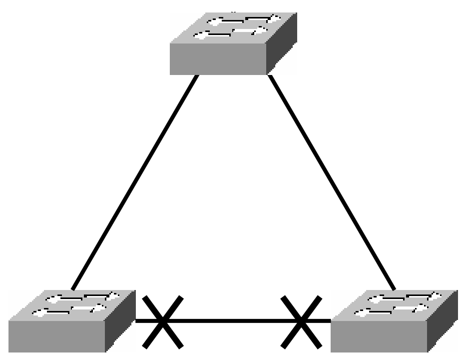


Learning Bridges—Loop Problem



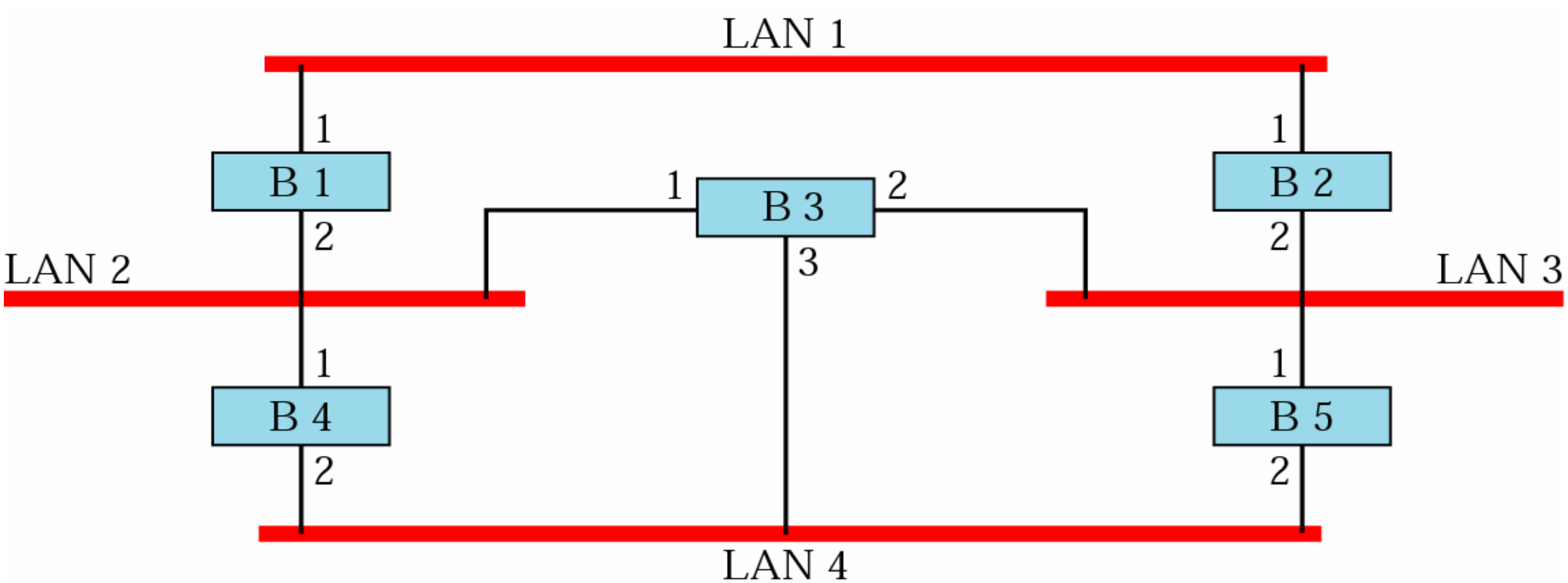
Spanning Tree—Basic Idea

- Discover a subset of the topology that is loop-free (a tree)
- Just enough connectivity so that there is a path between every pair of segments where physically possible
  - The tree is spanning
- Disable (block) all other ports
- All bridges exchange configuration messages, called *bridge protocol data units* (BPDUs), allowing them to calculate a spanning tree



Spanning Tree Starting Point

- Each bridge has a unique ID
- Each port has a unique ID within the bridge
- A *cost* can be calculated for each path between two bridges
  - A default cost value is used for each port (can be configured)

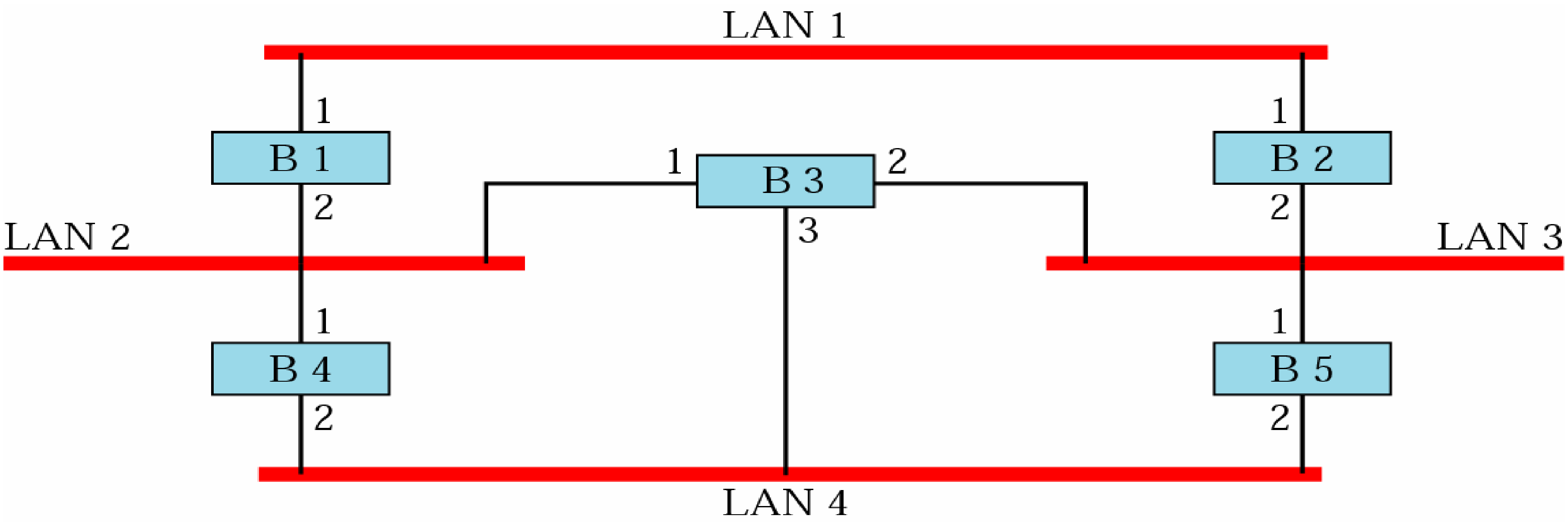


©The McGraw-Hill Companies, Inc., 2000

Spanning Tree Process

1. The node with the smallest ID is selected the root bridge
2. On each bridge, select a root port
   - Port with the least cost path to the root bridge
3. On each LAN segment, select a designated bridge
   - Bridge with least cost path to root bridge
     - If two bridges have same cost, select the bridge with smallest ID
   - Mark the corresponding port as the designated port
4. Forward frames only on marked ports
   - Designated ports and root ports
   - Block on the others

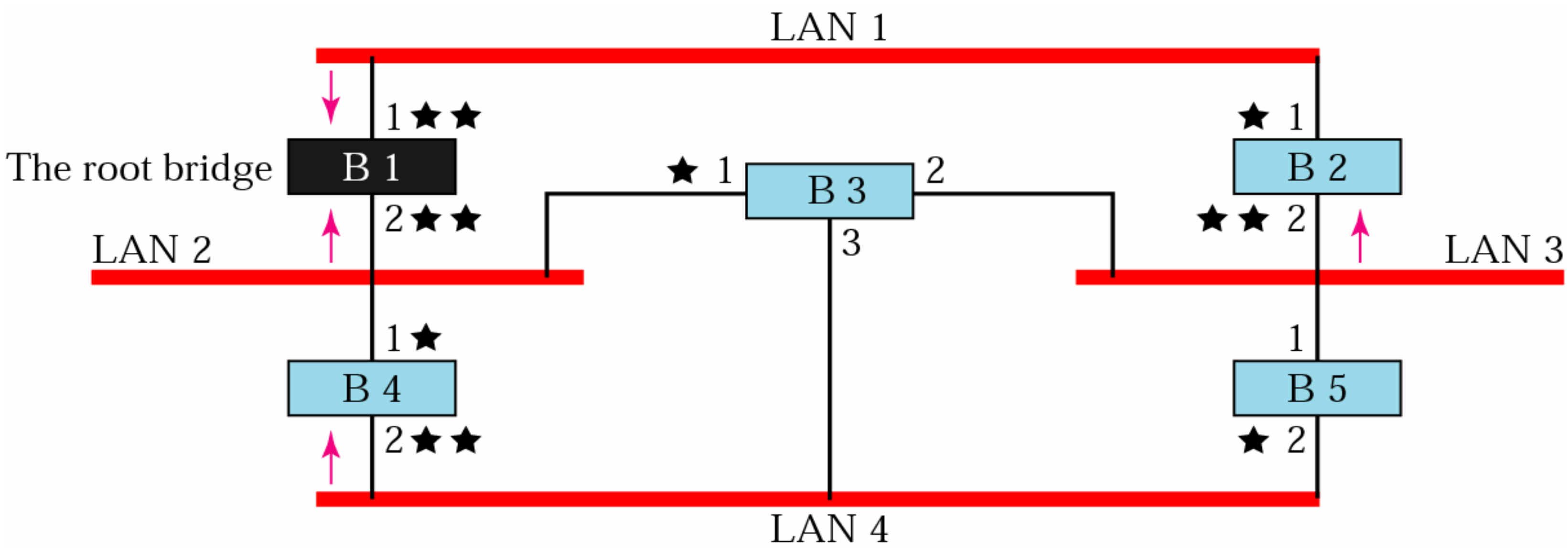
Before Spanning Tree



©The McGraw-Hill Companies, Inc., 2000

- B1 has the smallest ID
  - then assume B4 < B2 < B3 < B5

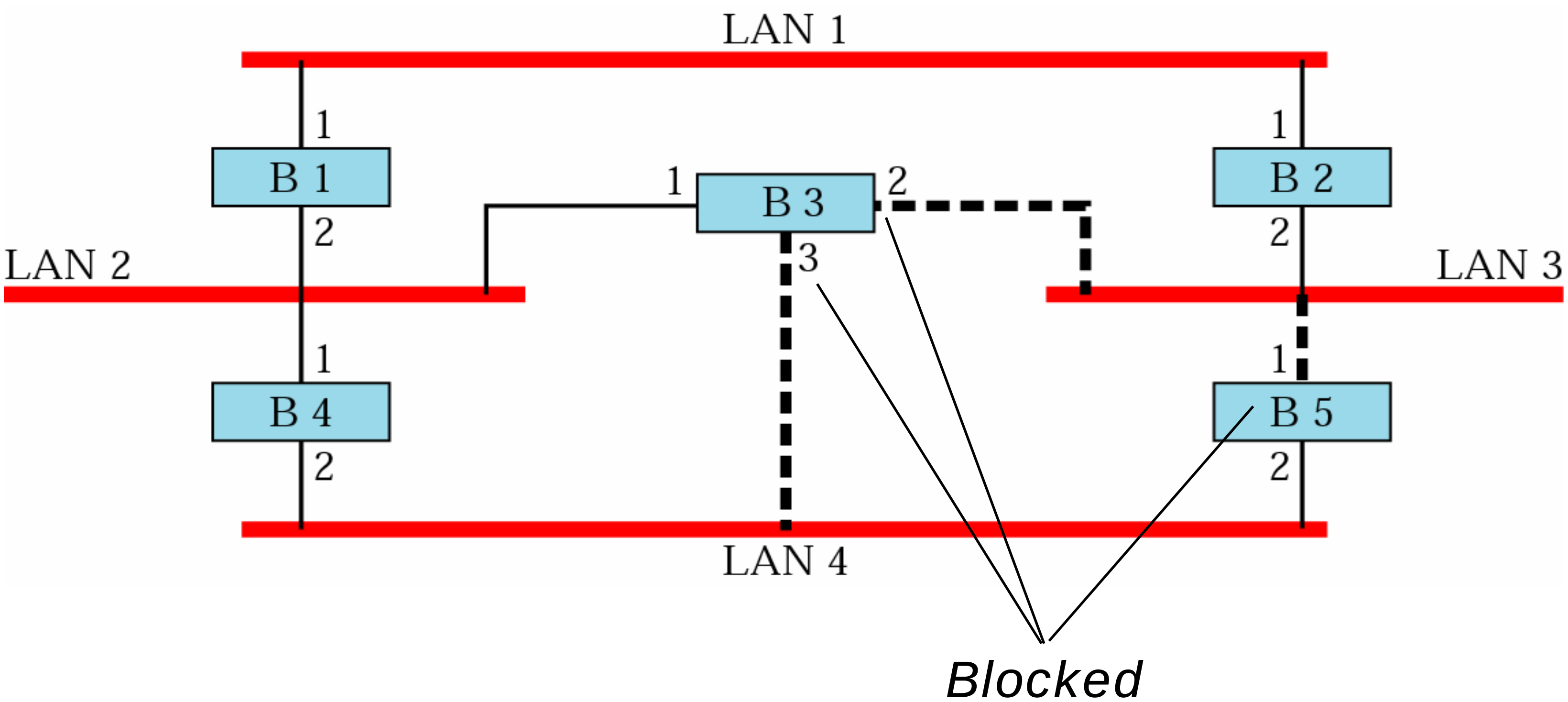
Applying Spanning Tree



©The McGraw-Hill Companies, Inc., 2000

- B1 has the smallest ID (then B4, B2, B3, B5)
- * * means designated bridge for a segment
- * means root port

Forwarding Ports and Blocking Ports



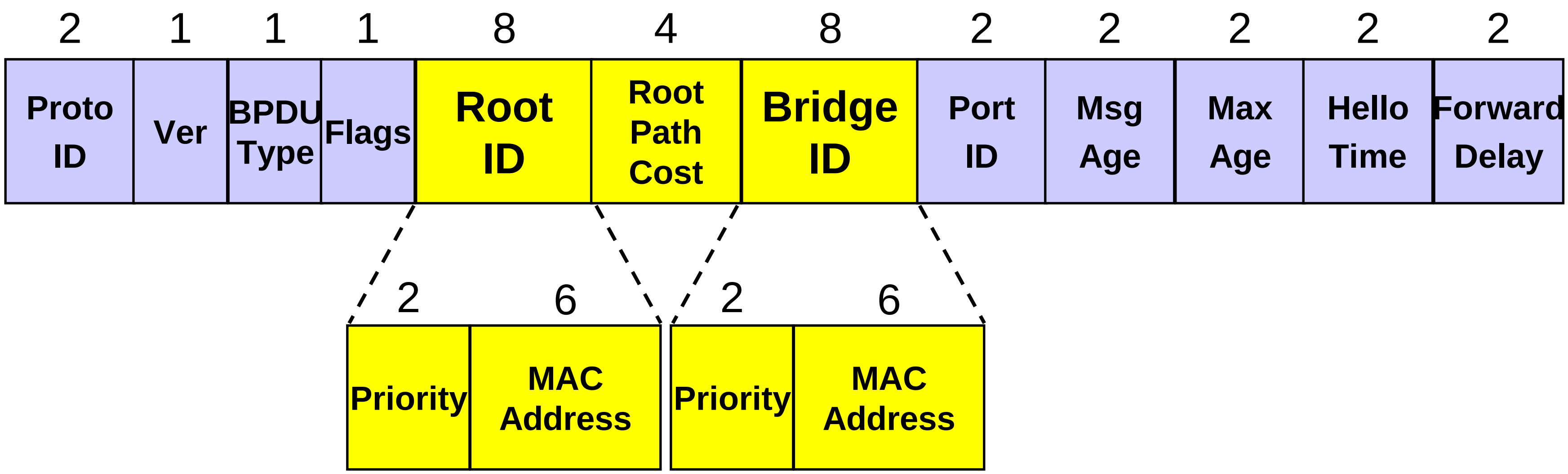
©The McGraw-Hill Companies, Inc., 2000

- Note that STP is not a routing protocol
  - It does not optimize routing
  - Traffic conectration towards the root

Spanning Tree Protocol

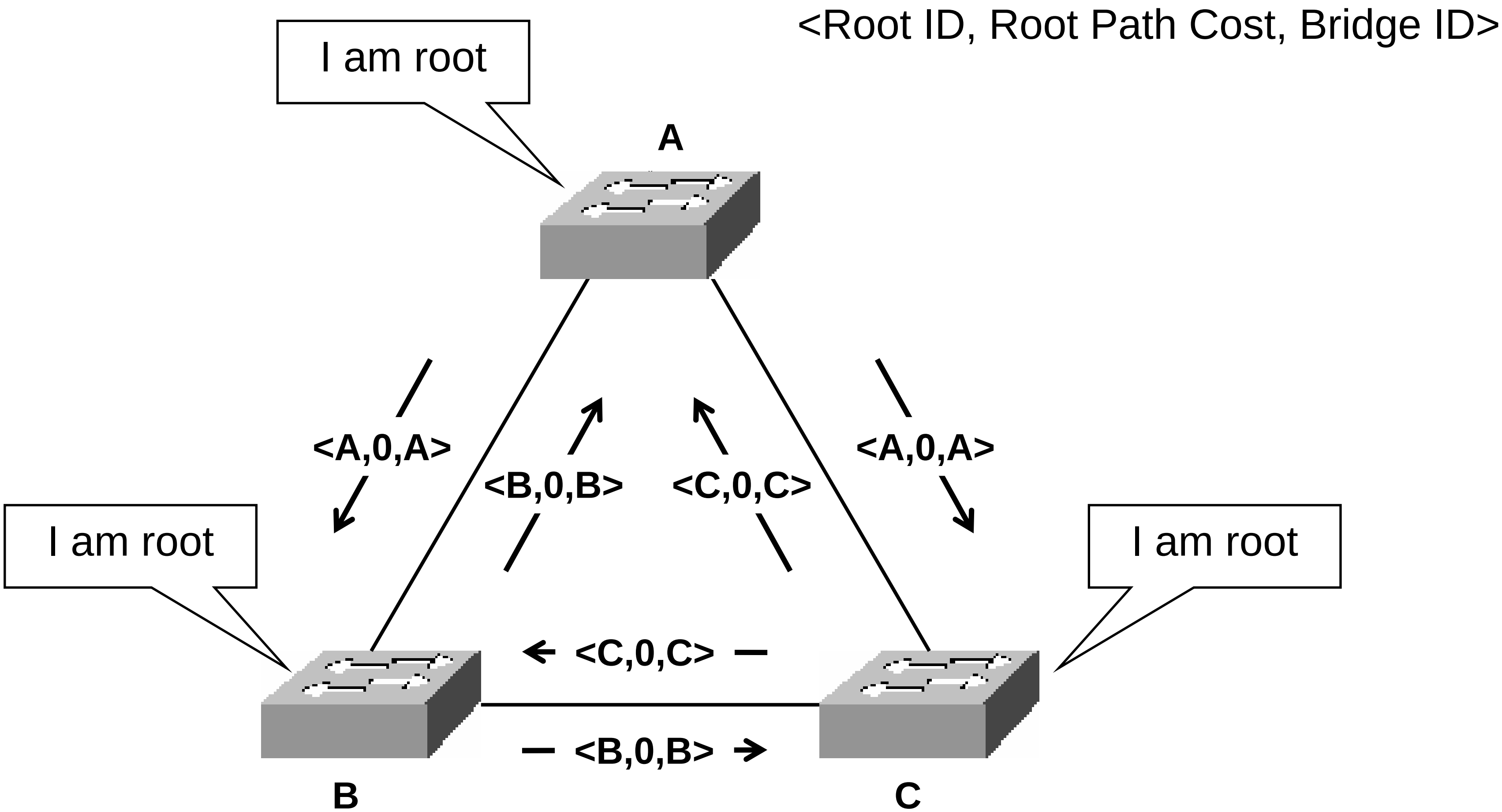
- Defined in IEEE 802.1D MAC Bridges standard
- Protocol to calculate a spanning tree
- Convergence
  - All bridges should reach a unified view of the spanning tree
- Special frames sent between neighbor switches
  - Bridge Protocol Data Units, BPDUs
  - Not forwarded!
- Based on bridge identifiers
  - Strict ordering
  - Based on MAC addresses

Bridge Protocol Data Unit (BPDU)



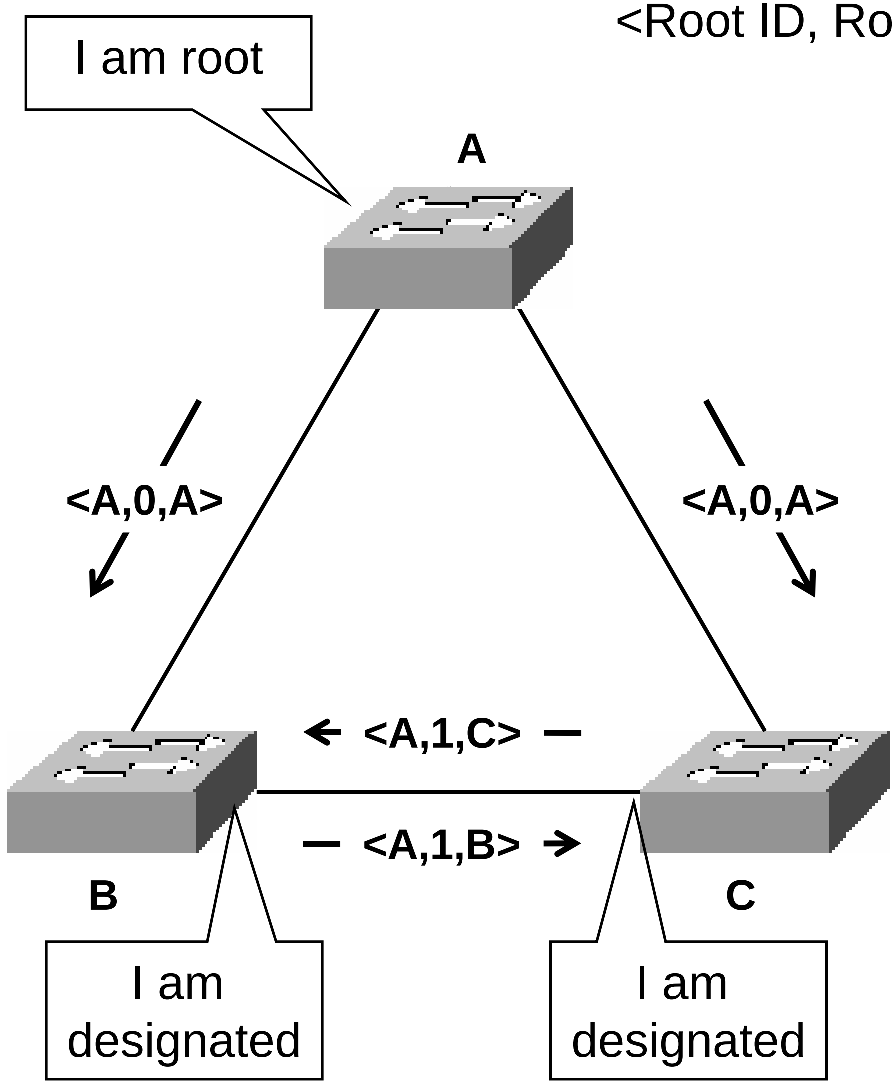
- Sent as an 802.1 frame
  - Destination MAC address 01-80-C2-00-00-00 (multicast)
  - BPDU Type 0
- Sent periodically (Hello Time) by root bridges
  - Triggers sending of BPDUs in designated bridges

Initial State

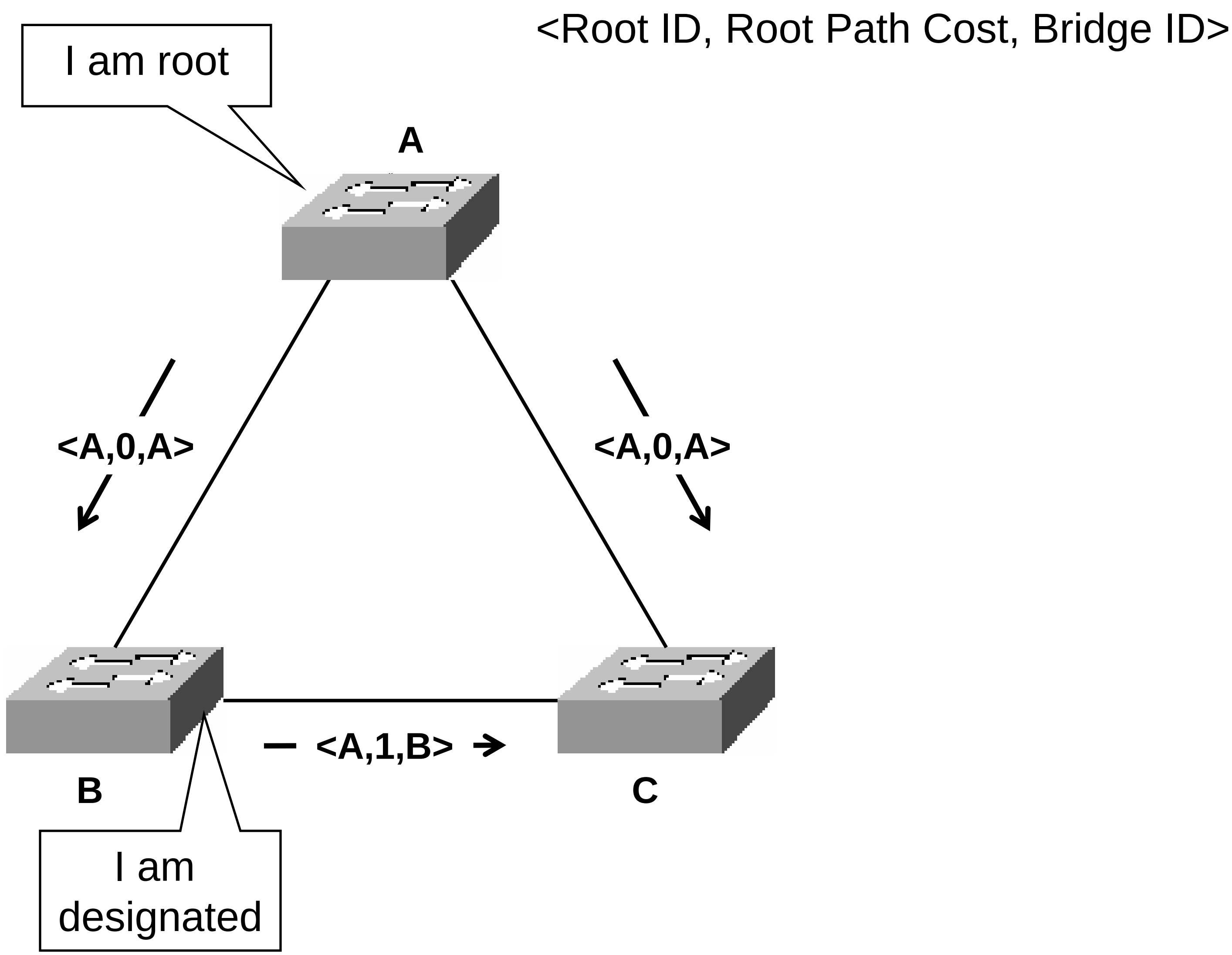


Root Bridge Recognized

<Root ID, Root Path Cost, Bridge ID>

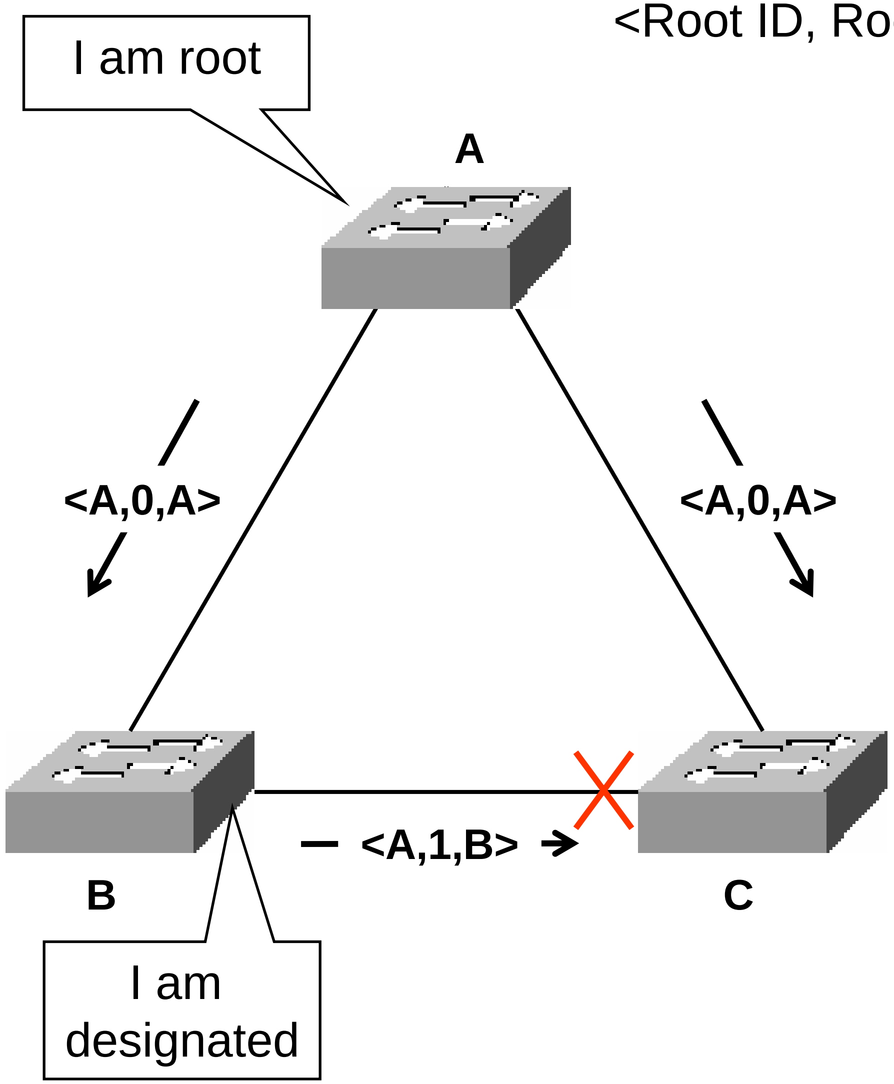


Designated Bridge Recognized



Ports Disabled

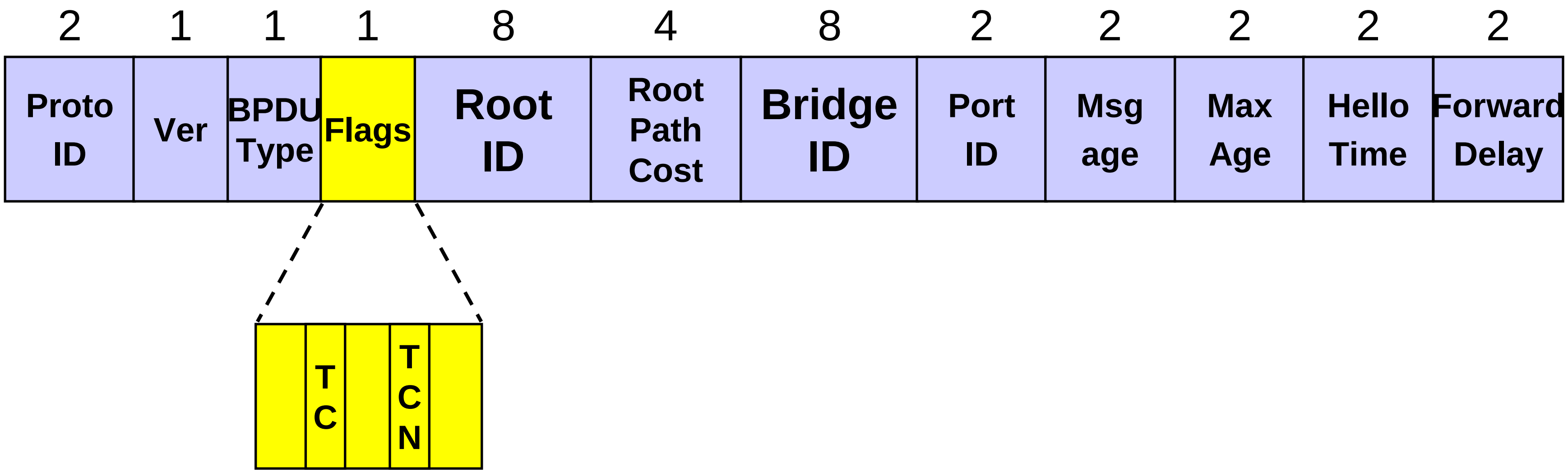
<Root ID, Root Path Cost, Bridge ID>



Topology Changes and Learning Table

- Entries in learning table expires
  - Normally after 5 minutes
  - Incorrect forwarding
    - “Black hole”
- Solution:
  - Faster expiration time when network configuration has changed
  - Topology Change Notification PDU

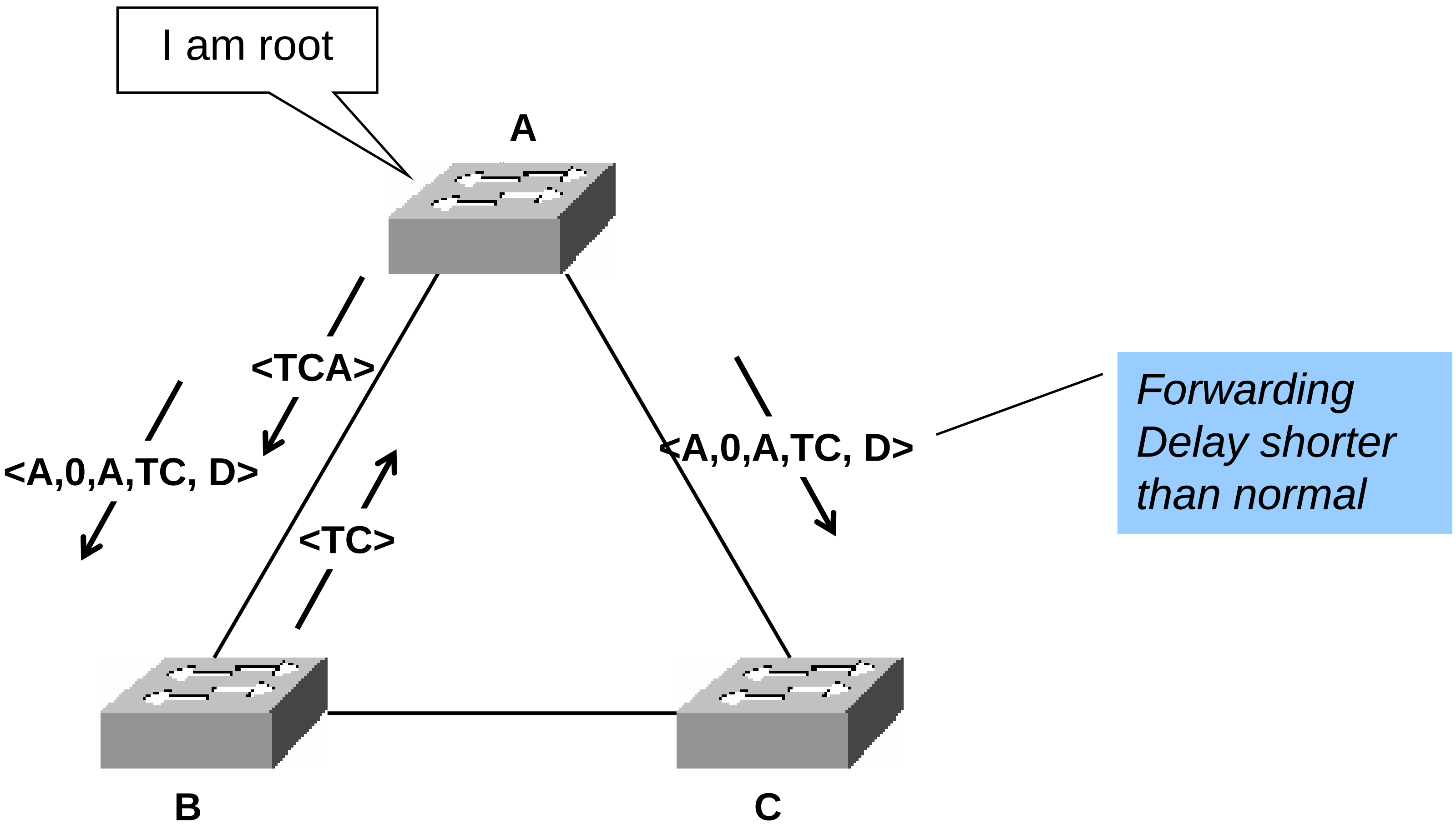
Topology Change Notification



- Topology change at a bridge
  - Port failure
  - No periodic configuration BPDUs
  - Port status change
- Bridge sends spontaneous BPDU
  - Topology Change Notification BPDU
  - BPDU type 0x80

Topology Change

TCN BPDU: <Flags>
Configuration BPDU: <Root ID, Root Path Cost, Bridge ID, Flags, Forward Delay>



Spanning Tree Protocol Timing

2	1	1	1	8	4	8	2	2	2	2	2
Proto ID	**Ver**	**BPDU Type**	**Flags**	**Root ID**	**Root Path Cost**	**Bridge ID**	**Port ID**	**Msg age**	**Max Age**	**Hello Time**	**Forward Delay**

- Protocol is timer driven
- Too short timers can give loops and instabilities
- Too long timers can give long convergence times
  - Until network reaches a stable spanning tree configuration

Rapid Spanning Tree Protocol

- Ordinary STP takes 30 – 50 seconds to converge, with default settings
- Rapid Spanning Tree Protocol (RSTP)
  - First defined in IEEE 802.1w
  - Replaces STP after 2004 in 802.1D
  - Full-duplex mode
    - No shared links
  - Backwards compatible with STP
  - Converges in around one second

RSTP vs STP

- RSTP has two more port designations
  - Alternate Port—backup for Root Port
  - Backup port—backup for Designated Port on the segment
- In RSTP, all bridges send BPDUs automatically
  - While in STP, the root triggers BPDUs
- In RSTP, bridges act to bring the network to convergence
  - While in STP, bridges passively wait for time-outs before changing port states

Bridging Limitations

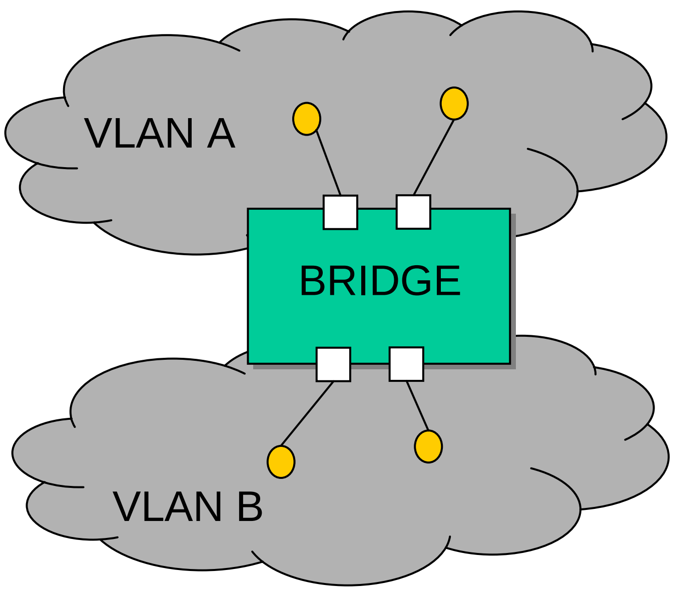
Can a bridged network be of an arbitrary size?

No, it is not scaling because of:

- Addressing
  - No structure in the MAC addresses → no aggregation can be used
- Table sizes (size of the learning table)
- Traffic congestion
  - Broadcast and multicast spans the whole broadcast domain
  - Broadcast domains must be kept to a reasonable size

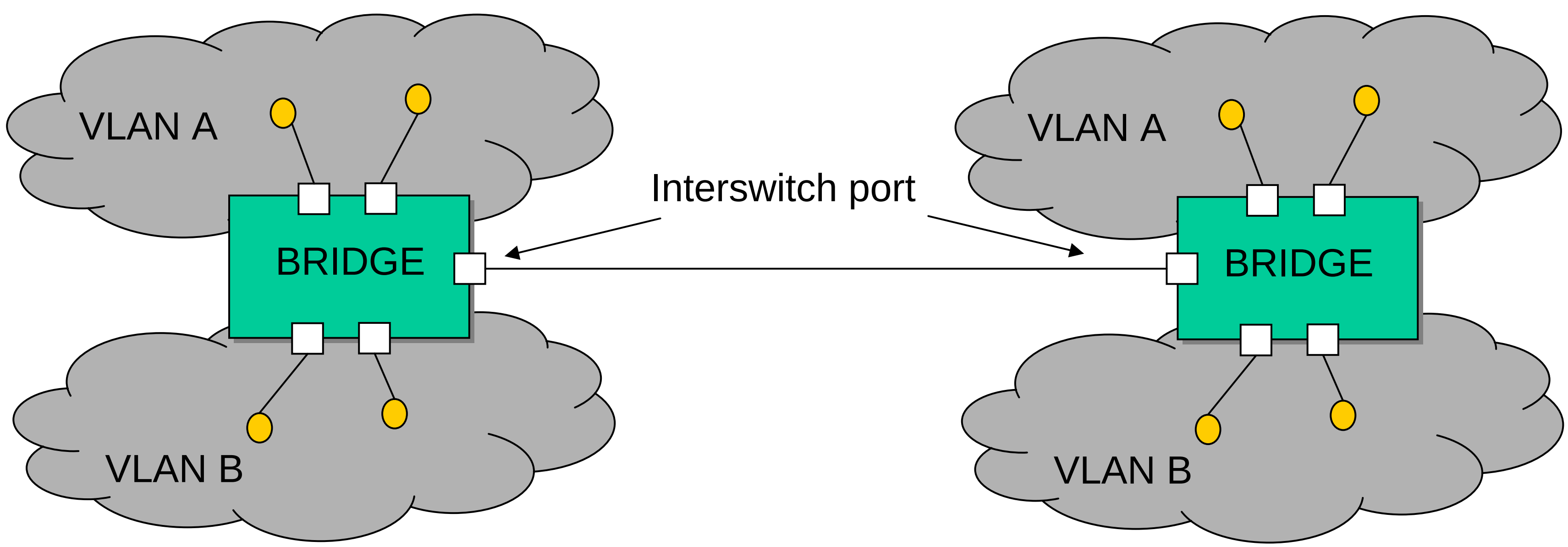
Virtual LAN (VLAN)

- A single bridge can act as multiple bridges by grouping the ports of the bridge into separate VLANs.
- Like a LAN, each VLAN is a layer 2 realm separated from other VLANs (separate broadcast domains) by separating ports on a bridge.
- Routing between VLANs requires a router (just like LANs)

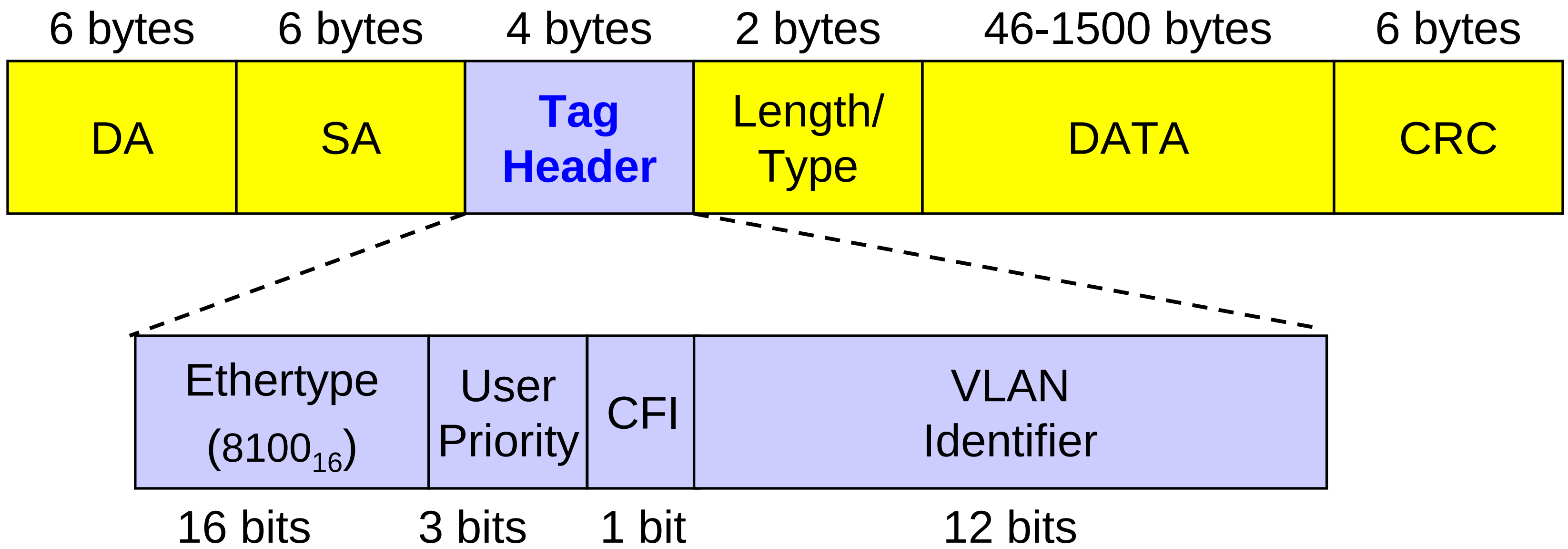


VLAN Grouping and Frame Tagging

- VLAN membership determined through port number
  - Ports 1, 2: VLAN 1; Ports 3, 4: VLAN 2, etc
- Frames belonging to different VLANs can be sent on the same link by VLAN encapsulation using frame tagging
  - Interswitch ports belong to several VLANs



IEEE 802.1Q Frame Tagging



- Tag header added to Ethernet header
  - Ethertype 810016
- 12-bit VLAN ID allows for 4096 VLANs

VLAN Extensions

- Multiple spanning tree protocol
  - Original STP computes one spanning tree for all VLANs
  - Multiple STP computes one spanning tree per VLAN
- VLAN Stacking (Q-in-Q)
  - IEEE 802.1ad ”Provider Bridge” standard
  - Extends tag space
  - Allows for providers to transport customer frames with tags
    - Outer tag: S-TAG or ”provider tag” (IEEE 802.1ad)
    - Inner tag: C-TAG or ”customer tag” (IEEE 802.1Q)

6 bytes	6 bytes	4 bytes	4 bytes	2 bytes
DA	SA	S-TAG	C-TAG	Length/ Type

Layer 2 Summary

- Point-to-point and Ethernet
- Bridging
  - Learning bridges
- Spanning Tree Protocol
- VLANs