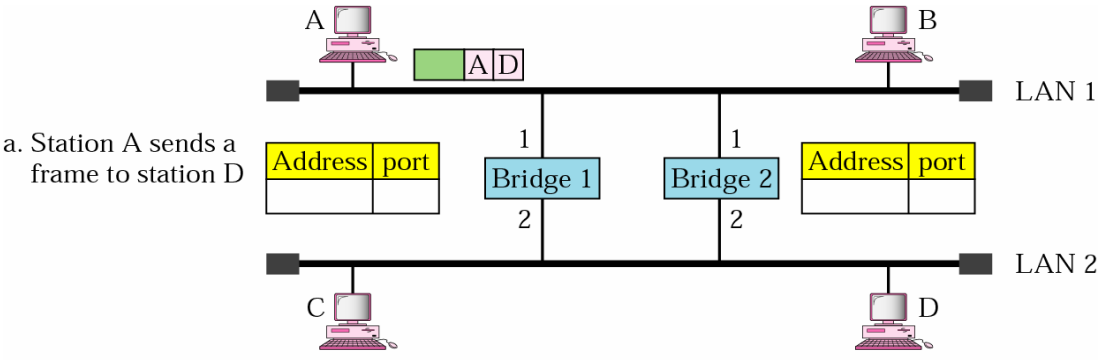
Spanning Tree Protocol

And Other Advanced Ethernet Topics

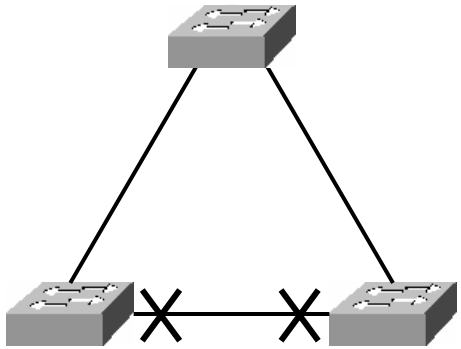
2E1623
Data Links and Local Area Networks

Learning Bridges—Loop Problem



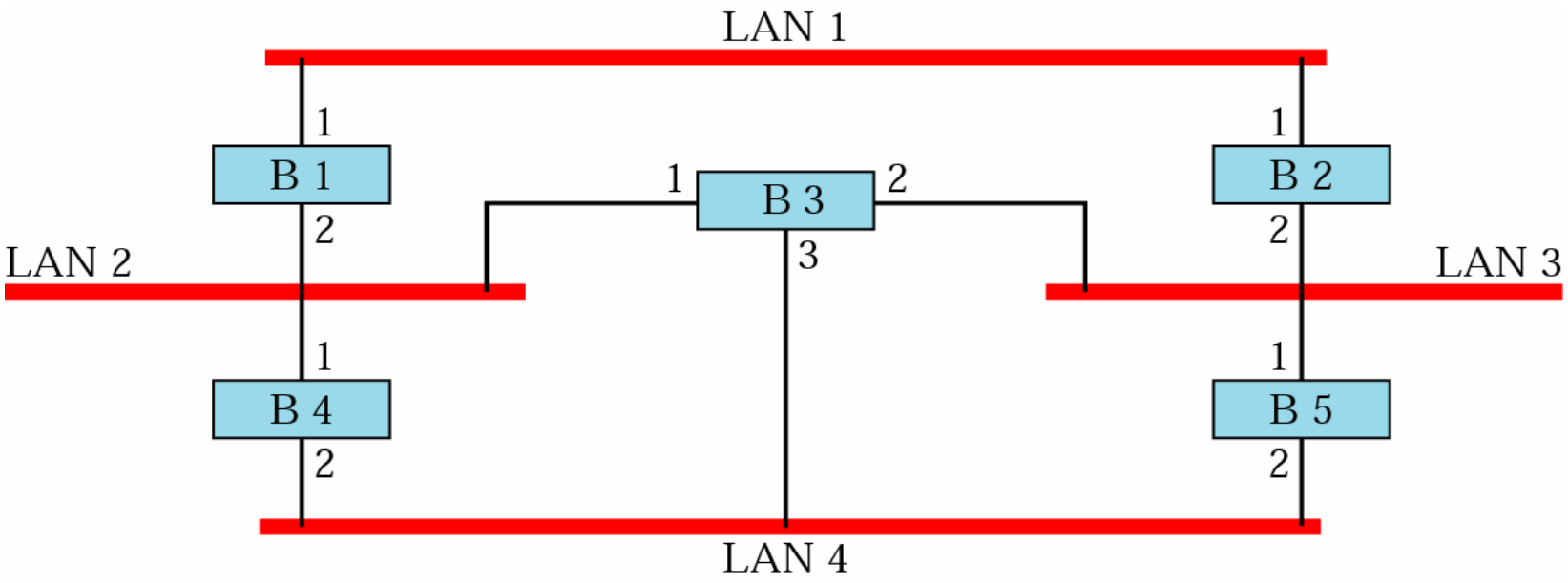
Spanning Tree—Basic Idea

- 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



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

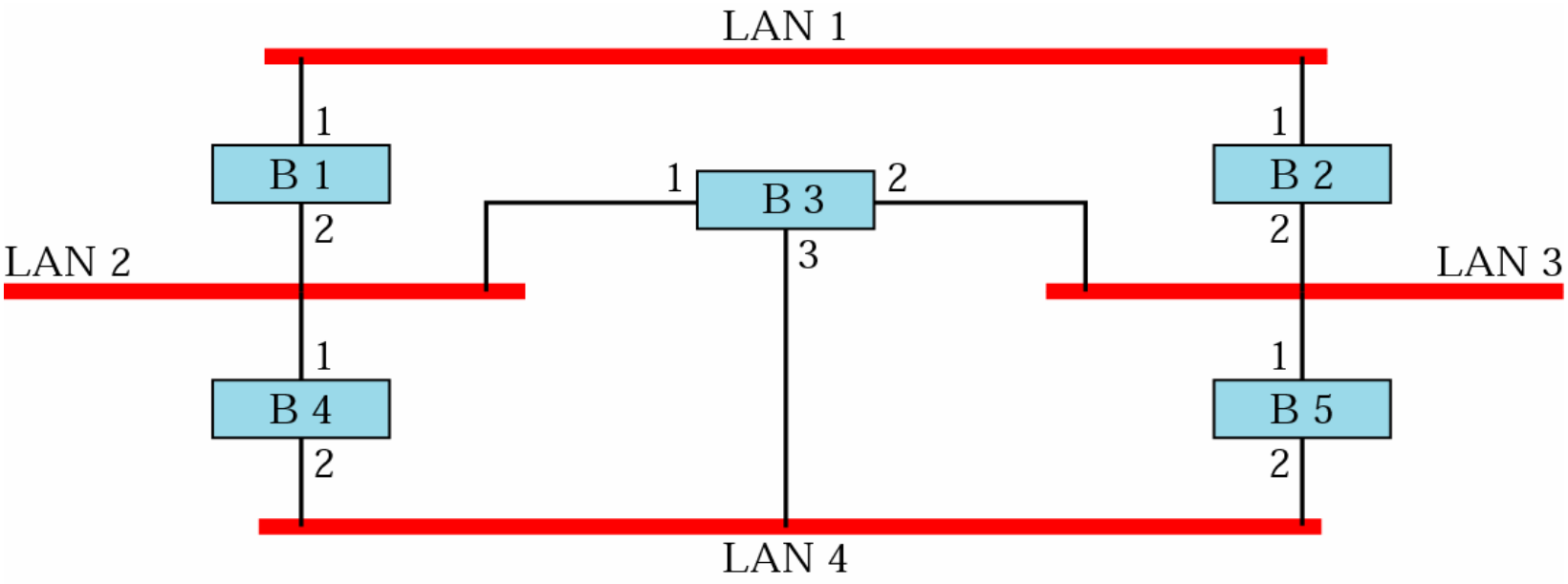


From B. A. Forouzan: Data Communications and Networking, 3rd ed, McGraw-Hill

Spanning Tree Process

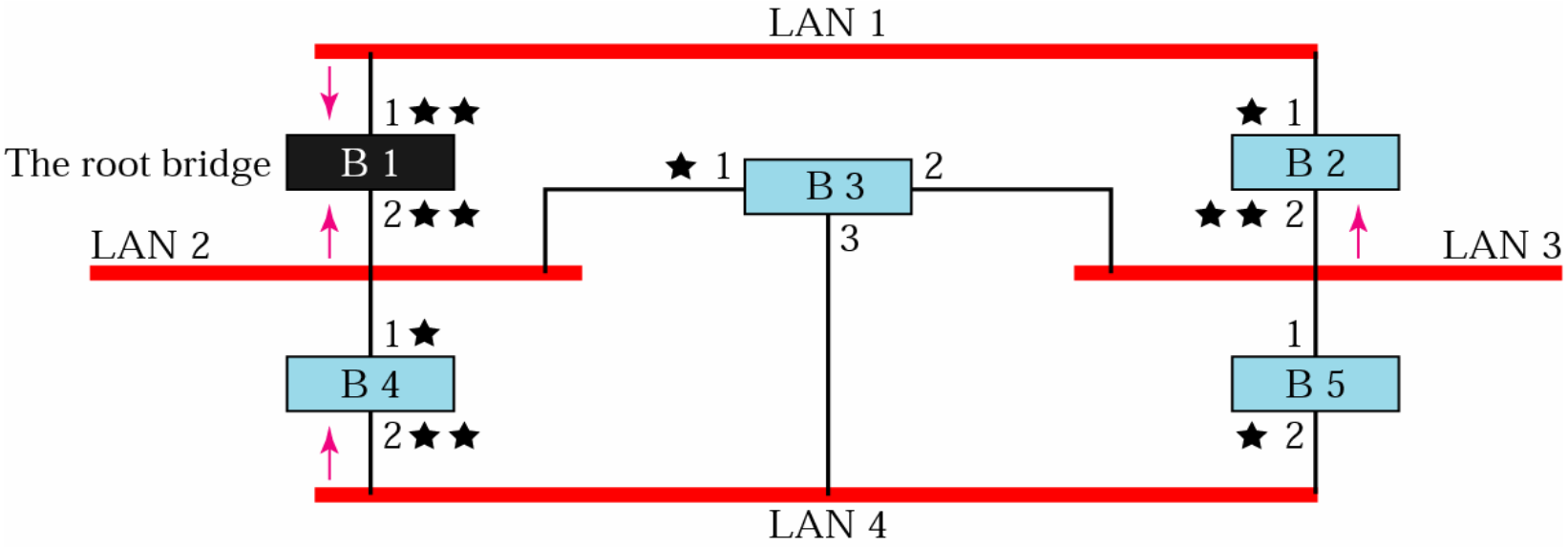
1. The node with the smallest ID is selected the *root bridge*
2. On each bridge, select a *root port*
   - Port with the least cost path to the root bridge
3. On each LAN segment, select a *designated bridge*
   - Bridge with least cost path to root bridge
     - If two bridges have same cost, select 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



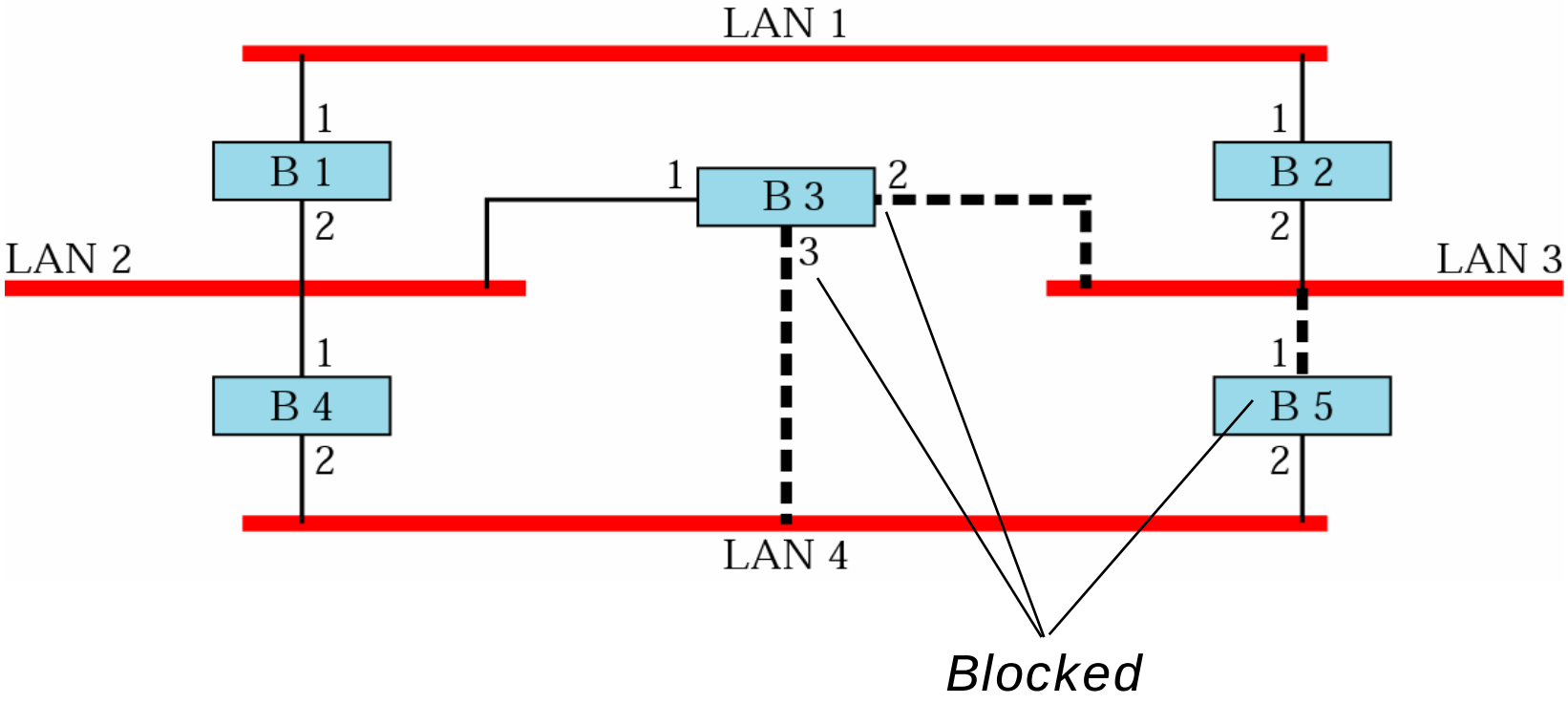
From B. A. Forouzan: Data Communications and Networking, 3rd ed, McGraw-Hill

Applying Spanning Tree



From B. A. Forouzan: Data Communications and Networking, 3rd ed, McGraw-Hill

Forwarding Ports and Blocking Ports

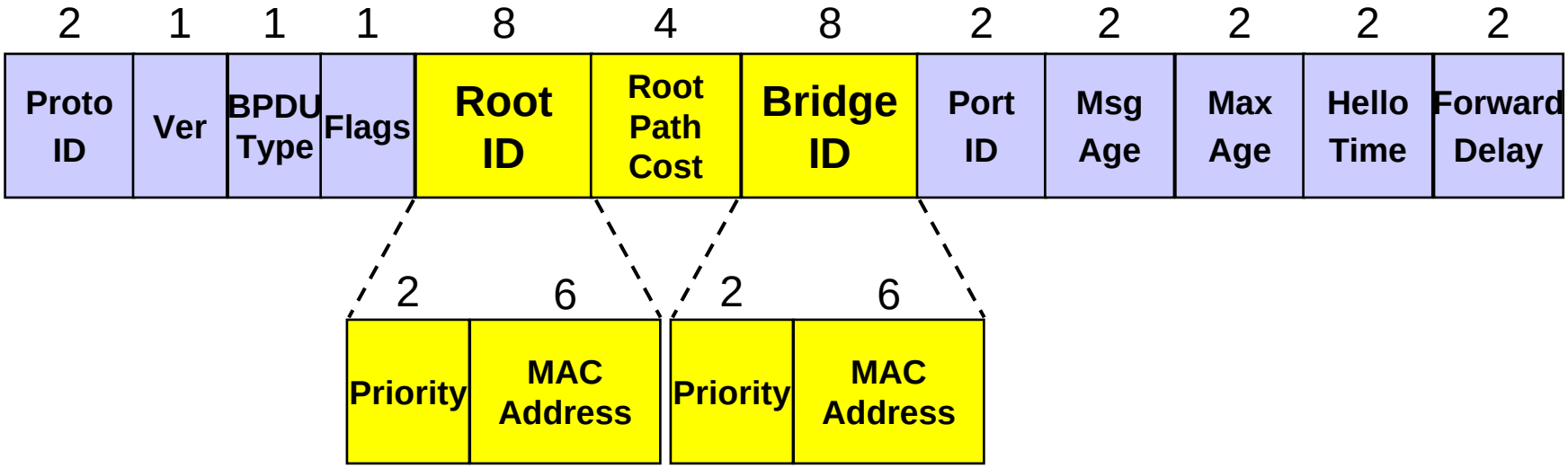


- Note that STP is not a routing protocol
  - In the sense that it does not optimize routing
  - Traffic concentration towards the root

Spanning Tree Protocol

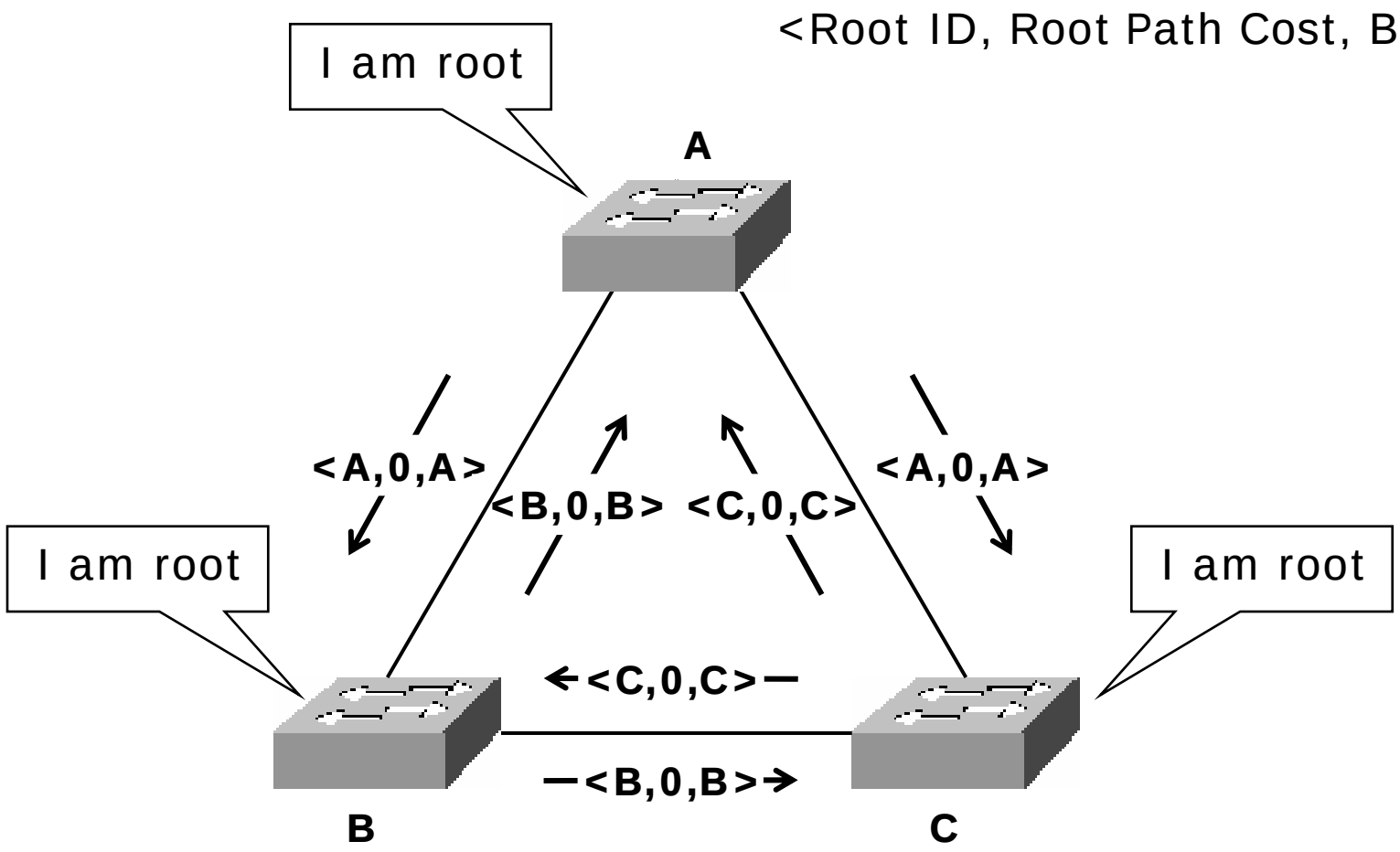
- Protocol to calculate a spanning tree
- Convergence
  - All bridges should reach a unified view of the spanning tree
- Special frames sent between neighbour switches
  - Bridge Protocol Data Units, BPDUs
  - Not forwarded!

Bridge Protocol Data Unit (BPDU)

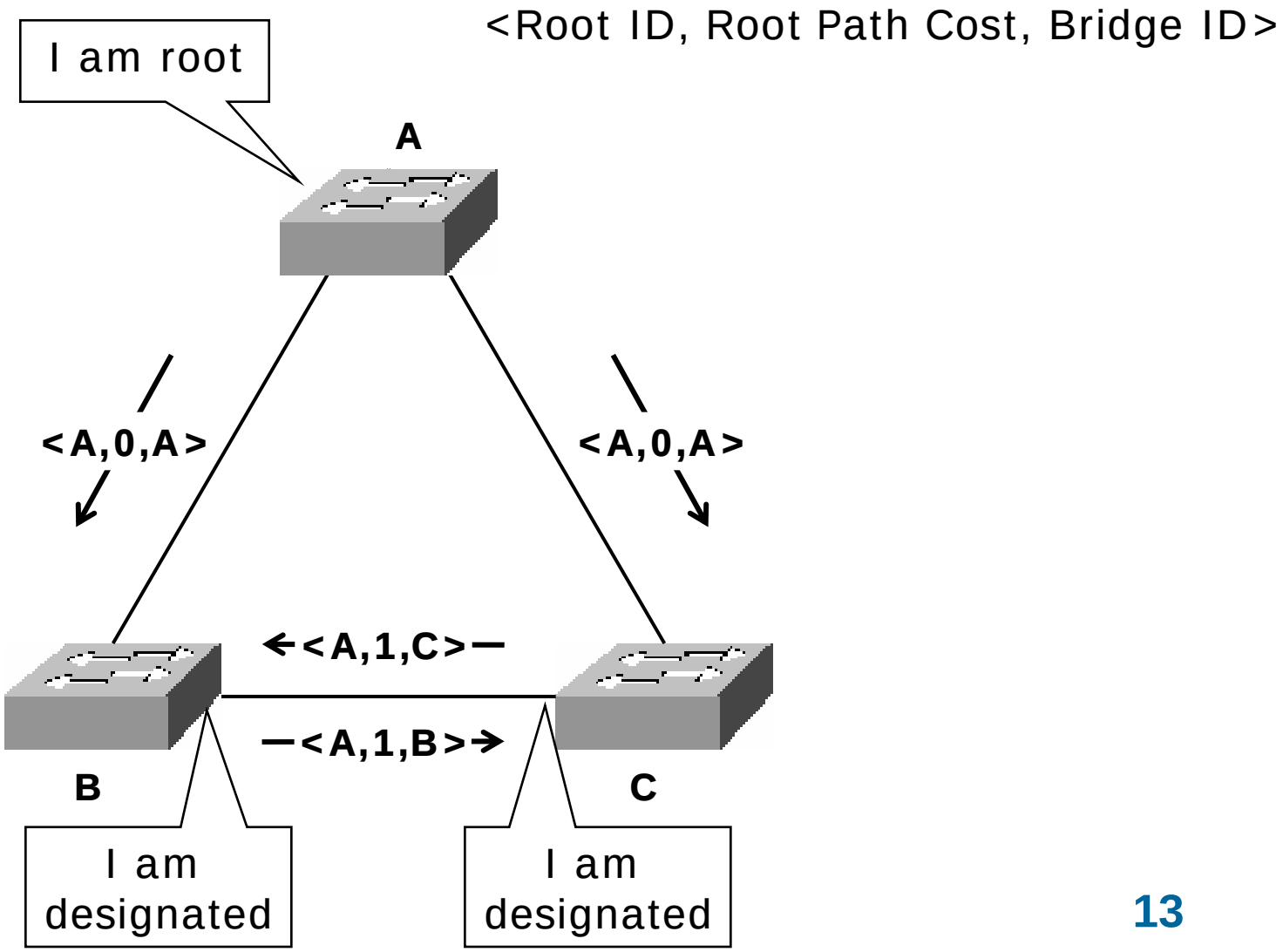


- Sent as an 802.1 frame
  - Destination MAC address 01-80-C2-00-00-00 (multicast)
  - BPDU Type 0
- Sent periodically (Hello Time) by root bridges
  - Triggers sending of BPDUs in designated bridges

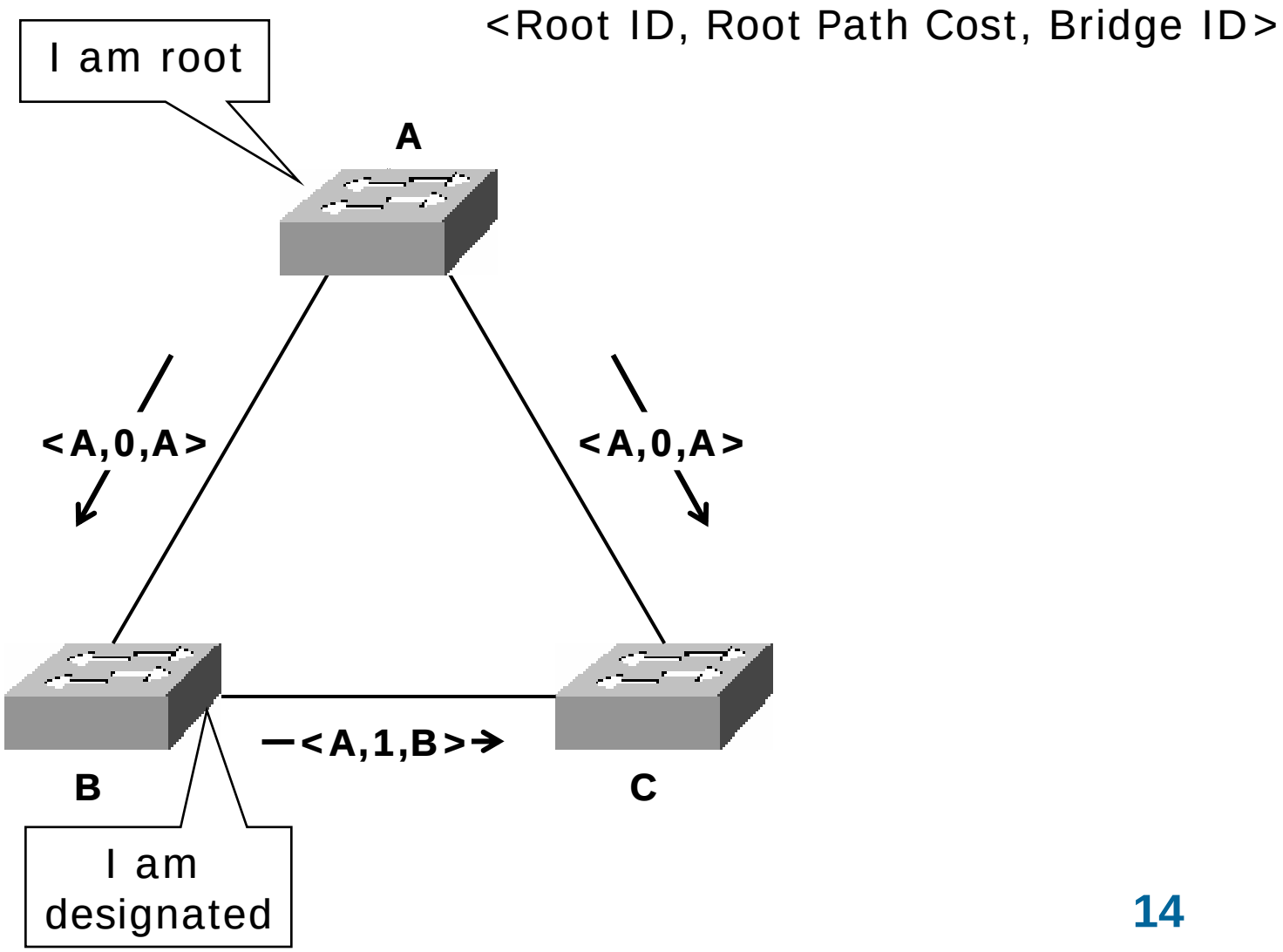
Initial State



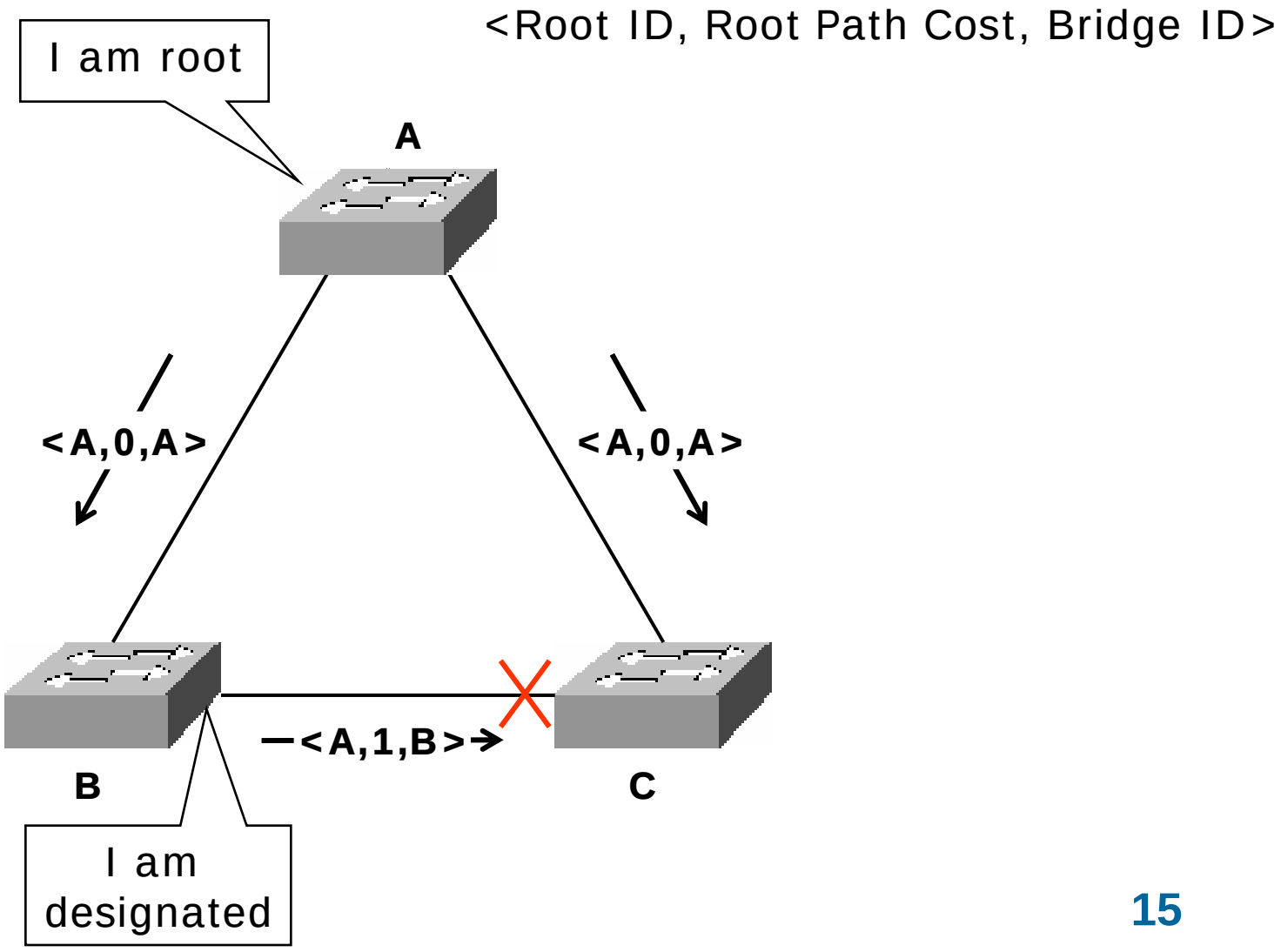
Root Bridge Recognized



Designated Bridge Recognized



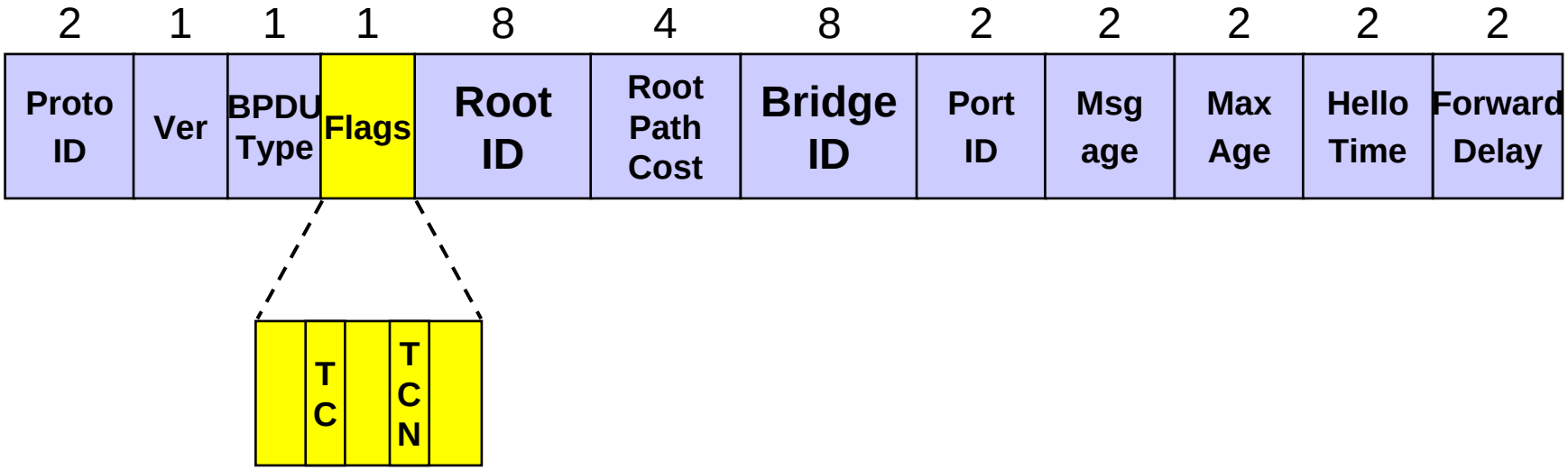
Ports Disabled



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

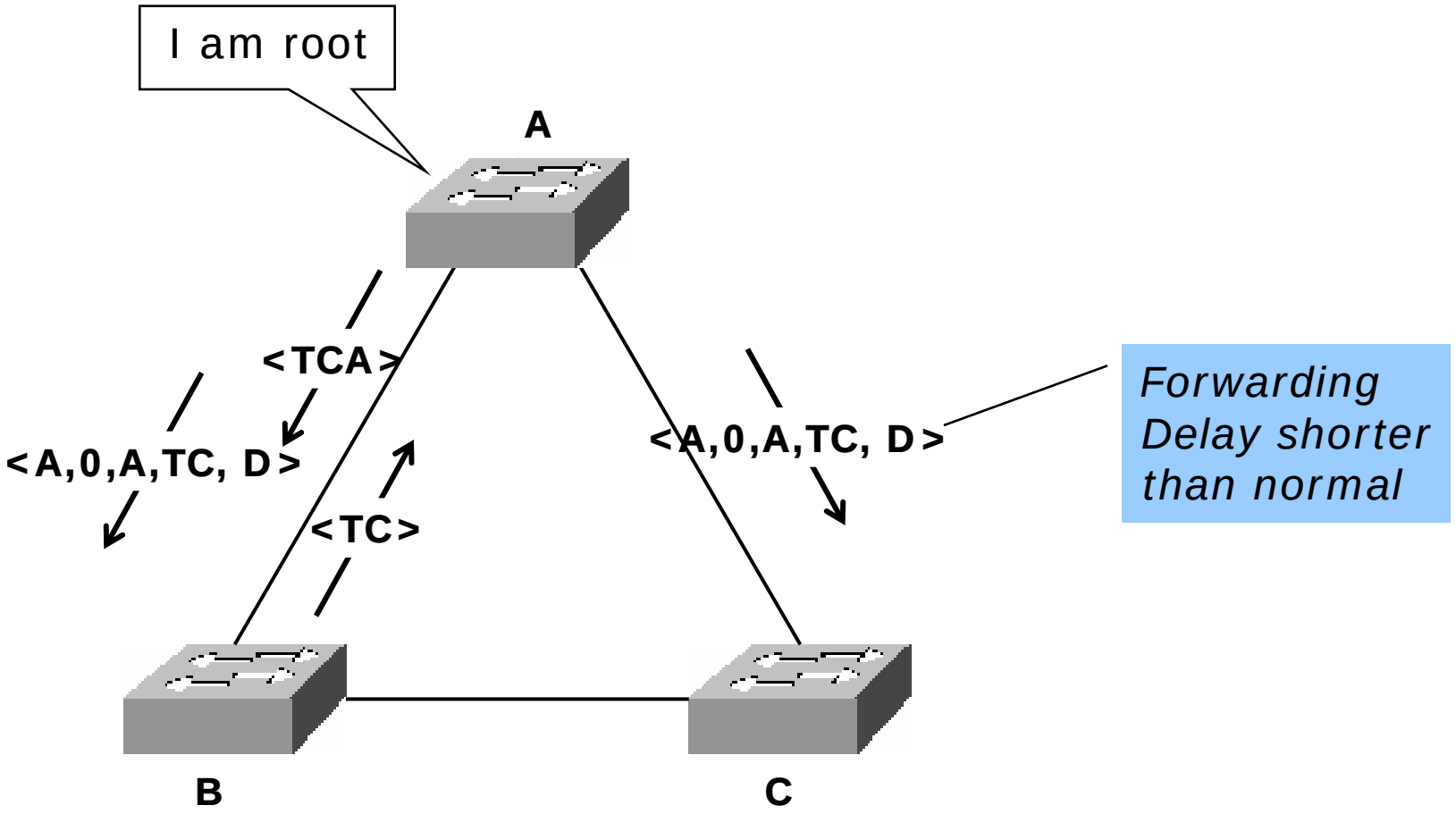
Topology Change Notification



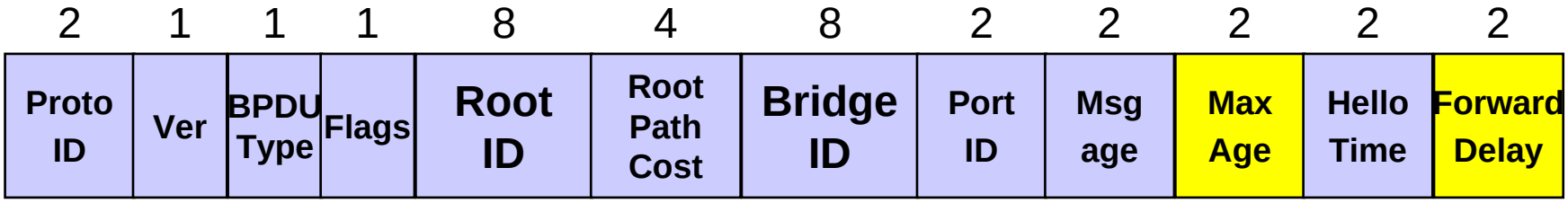
- Topology change at a bridge
  - Port failure
  - No periodic configuration BPDUs
  - Port status change
- Bridge sends spontaneous BPDU
  - Topology Change Notification BPDU
  - BPDU type 0x80

Topology Change

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

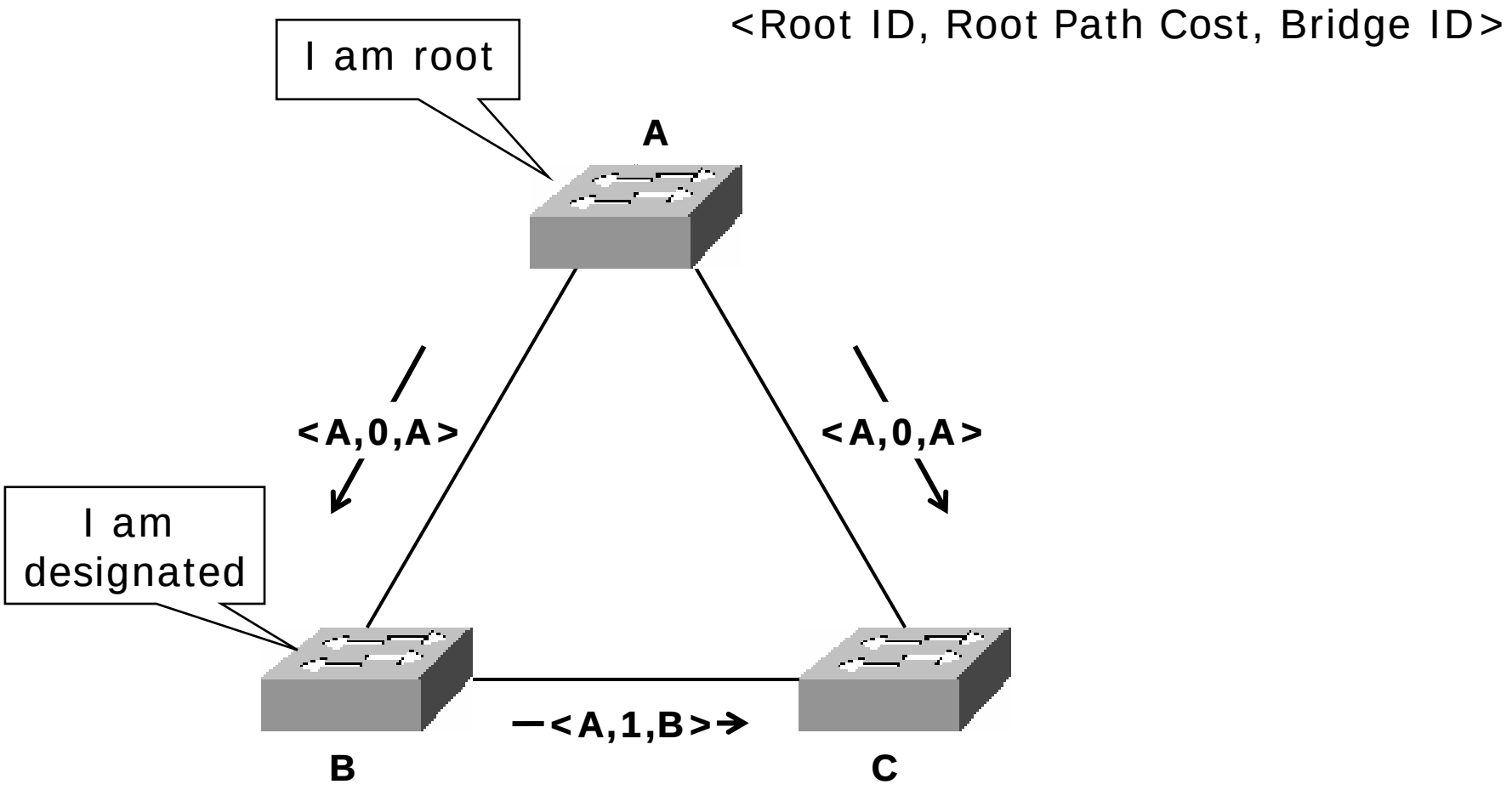


Spanning Tree Protocol Timing

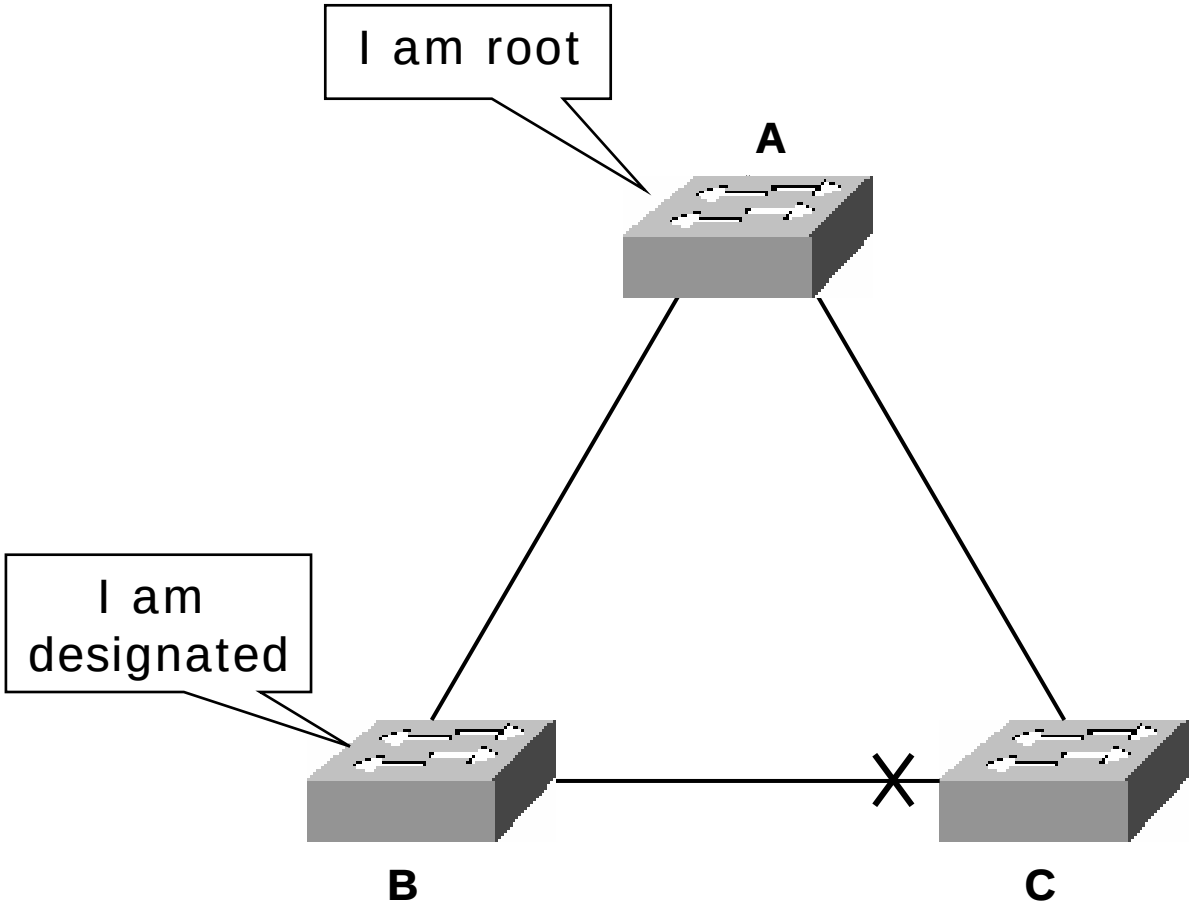


- Protocol is timer driven
- Too short timers can give loops and instabilities
- Too long timers can give long convergence times
  - Until network reaches a stable spanning tree configuration

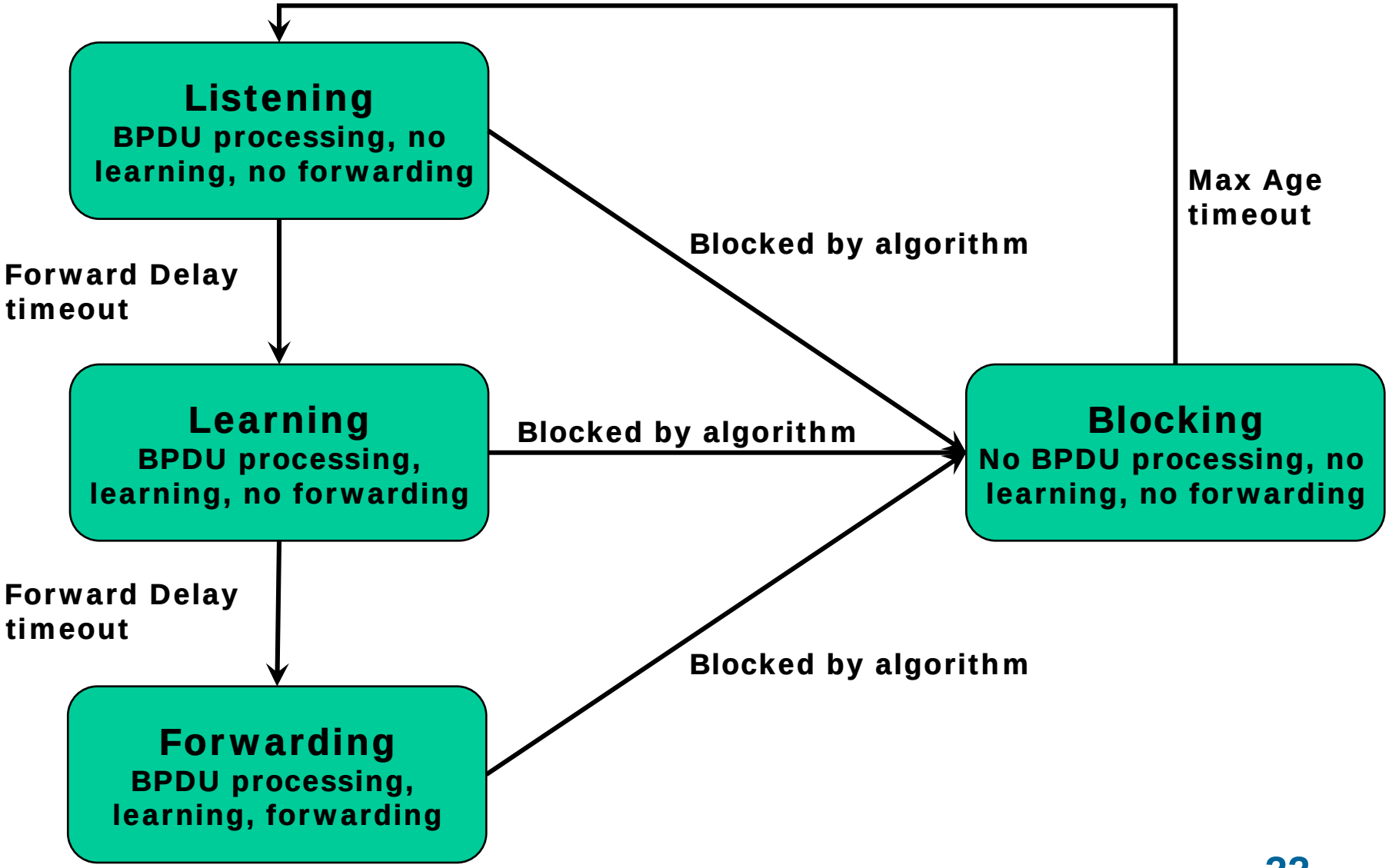
Designated Bridge Recognized



Ports Disabled



Port States


Listening
BPDU processing, no learning, no forwarding
Forward Delay timeout
Learning
BPDU processing, learning, no forwarding
Forward Delay timeout
Forwarding
BPDU processing, learning, forwarding
Blocked by algorithm
Blocked by algorithm
Blocked by algorithm
Blocking
No BPDU processing, no learning, no forwarding
Max Age timeout

Rapid Spanning Tree Protocol

- Ordinary STP takes 30 – 50 seconds to converge, with default settings
- Rapid Spanning Tree Protocol (RSTP)
  - IEEE 802.1w
  - Full-duplex mode
    - No shared links

RSTP vs STP

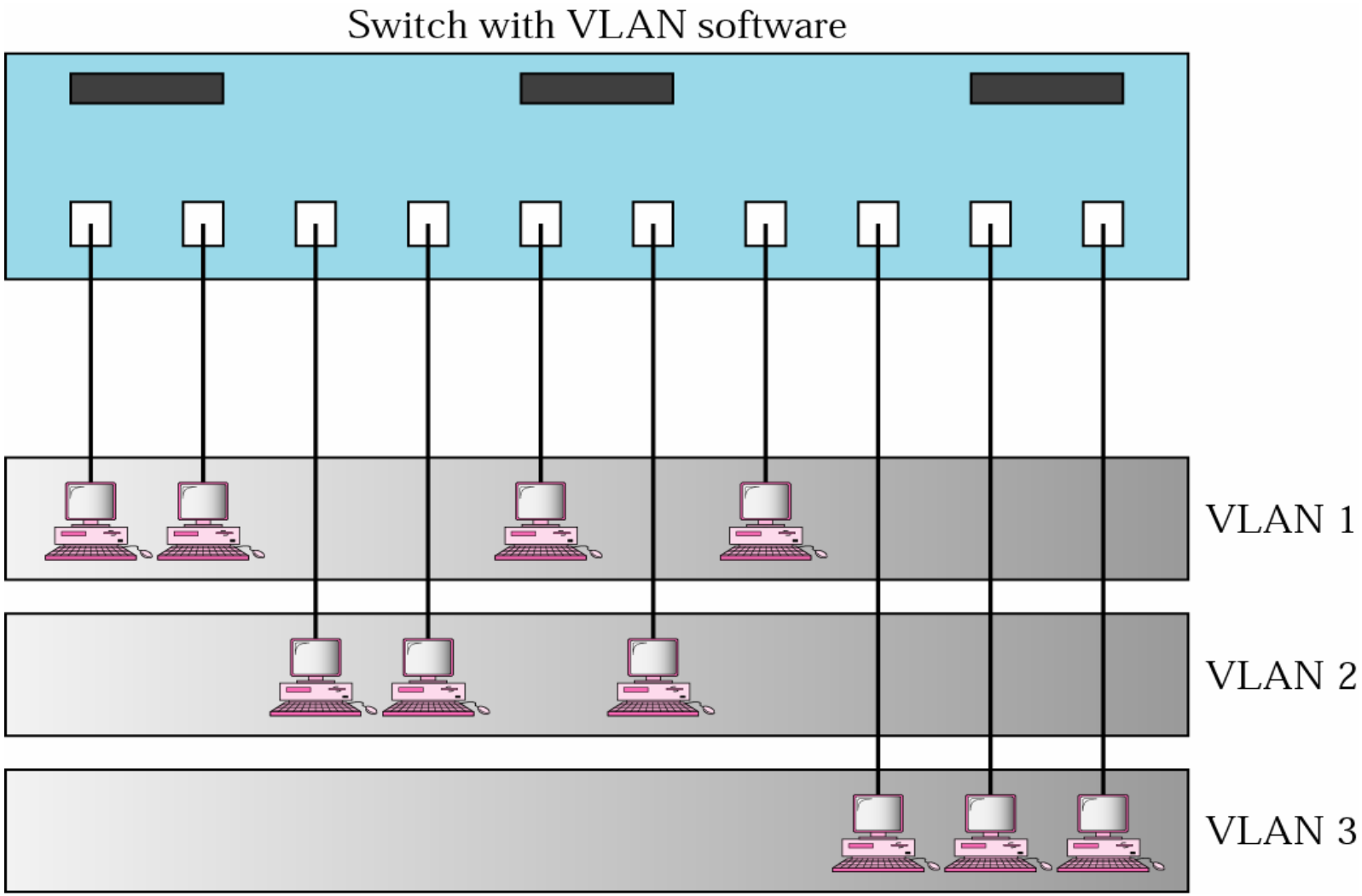
- RSTP has two more port designations
  - Alternate Port—backup for Root Port
  - Backup port—backup for Designated Port on the segment
- In RSTP, all bridges send BPDUs automatically
  - While in STP, the root triggers BPDUs
- In RSTP, bridges act to bring the network to convergence
  - While in STP,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

Virtual Local Area Networks

Virtual LANs (VLAN)

- Need a way to divide the LAN into different parts
  - Without physical reconfiguration
- Moving stations without reconfigurations
- Create virtual workgroups
- Keep broadcasts isolated
- Keep different protocols from each other

VLAN Divides LAN Into Logical Groups

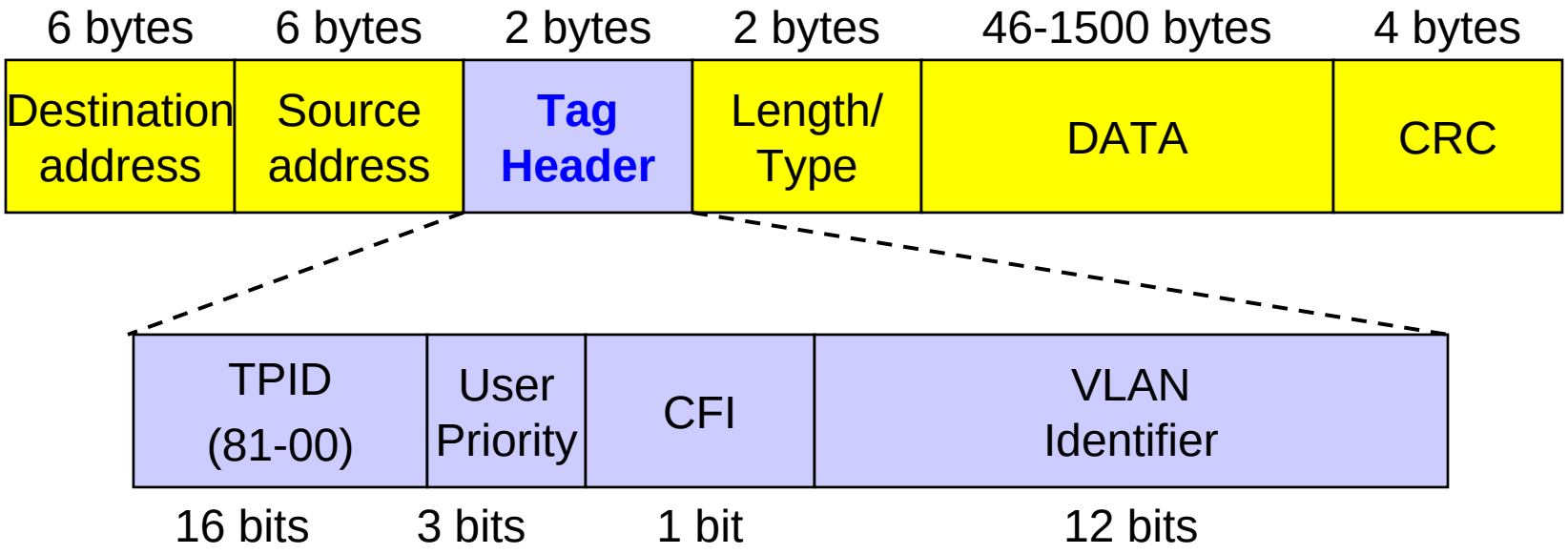


From B. A. Forouzan: Data Communications and Networking, 3rd ed, McGraw-Hill

VLAN Grouping

- How is VLAN membership determined?
  - Port number
    - Ports 1, 2, 7: VLAN 1
    - Ports 3, 4, 5, 6: VLAN 2
- MAC address
- Frame tagging
  - VLAN *trunking*
  - Many VLANs over the same link

Frame Tagging



- Tag header added to Ethernet header
  - IEEE 802.1Q
- 12-bit VLAN ID allows for 4096 VLANs

Spanning Trees and VLANs

- Per VLAN Spanning Tree
  - One spanning tree per VLAN
  - Many spanning tree instances to maintain
  - Different roots in different STs
    - load sharing
- Common Spanning Tree
  - One spanning tree for all VLANs
  - Simple, but all traffic goes the same way

Multiple Spanning Tree

- Multiple Spanning Tree Protocol (MSTP)
- Network organized in *regions*
- Regions have their own Multiple Spanning Tree Instances (spanning-tree topologies)
  - VLANs are associated to MSTIs
- One common spanning tree (CST) for the entire network
- MSTP based on RSTP (Rapid STP)

VLAN Signalling

- GVRP

Autonegotiation

Ethernet Autonegotiation

- Incompatible rx/tx modes
  - Full/half duplex
  - 10/100/1000 Mb/s
- Autonegotiation to allow two devices to agree on speed and duplex mode
- Based on 10BASE-T “heartbeat”
  - Normal Link Pulse
  - Sent every 16 ms on idle link

Link Code Word

- 100BASE-T “Fast Link Pulse”
- 16-bit code word, with a “Technology ability field” (8 bits)
  - 100BASE-T full duplex
  - 100BASE-T4
  - 100BASE-T
  - 10BASE-T full duplex
  - …
- Misconfiguration problems
  - Connectivity loss or performance degradation

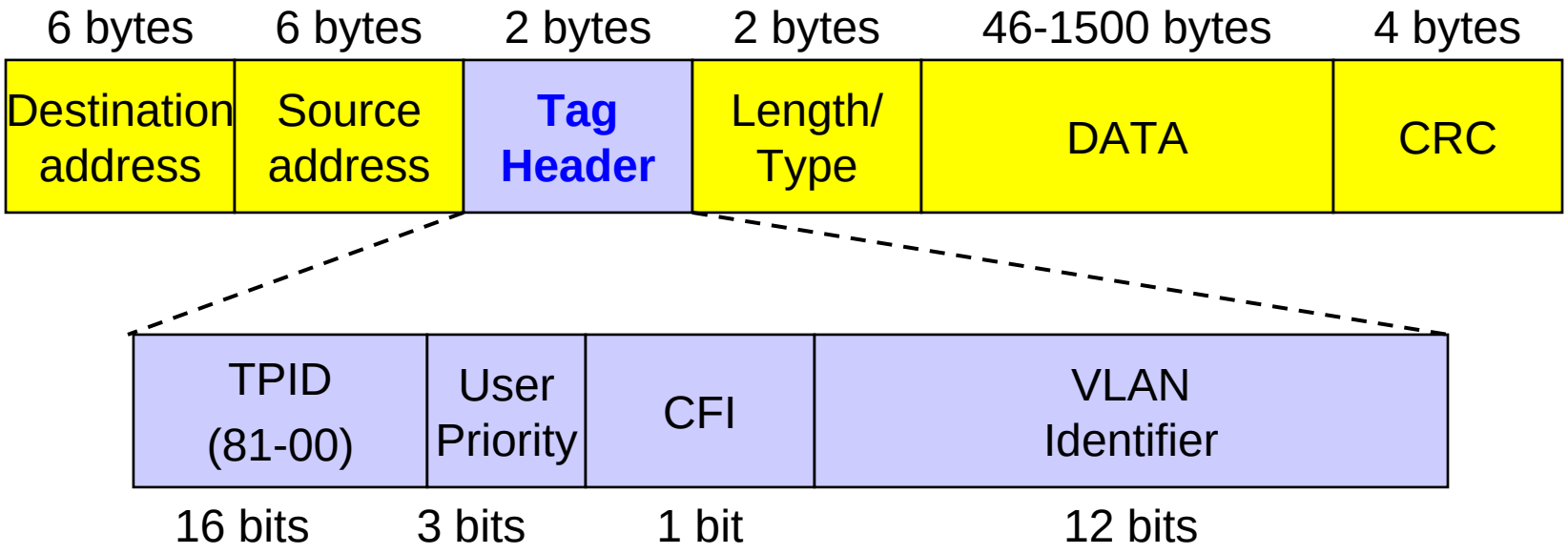
Ethernet PAUSE Frames

6 bytes	6 bytes	2 bytes	2 bytes	2 bytes	42 bytes	4 bytes
Destination address	Source address	Type (88-08)	Control operation (00)	Duration	Reserved	CRC

- Flow control
  - Full duplex mode only
  - Negotiated through auto-negotiation
- MAC Control Frame
- Destination address
  - 01:80:C2:00:00:01 (pre-defined multicast address) or address of remote bridge
- Duration is time in units of 512-bit times

Quality of Service and Multicast

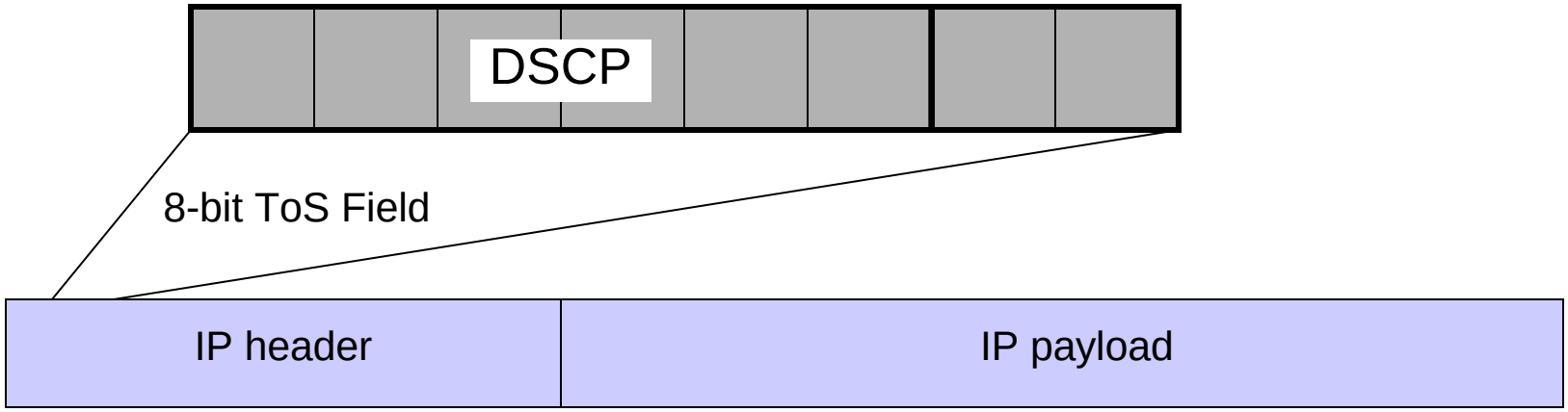
Quality of Service Switching



- Class of Service
  - User Priority field in IEEE 802.1Q header
  - 0 – 7, with 7 as highest priority value

DiffServ Packet Marking and Aggregation

- Each packet is marked with a DSCP (Differentiated Services Code Point) directly in the 8-bit IP ToS header field
  - 6 bits used → 64 possible code points (in practice much less is used)
  - Code points are unique within a domain – but may change at domain borders
- An ingress node aggregates packets into *behavior aggregates*, each marked by a unique code point (DSCP)



From QoS, RSVP and MPLS Lecture

Quality of Service Switching

- QoS processing depends on
  - QoS configuration of port
  - IP DSCP
    - Ignored, or set to predefined value
  - 802.1Q CoS
    - Ignored, set to predefined value, or computed from DSCP
  - ...

Multicast

- Multicast traffic is, by default, flooded
  - Increases traffic load
- Switches implement IGMP Snooping
  - Internet Group Management Protocol
    - Like ICMP
    - Monitor “Membership Reports” and “Leave Reports”
    - Only forward multicast frames to ports where there are receivers
  - MAC multicast group address is calculated from IP multicast address
    - 25 static bits (01:00:5e:0) plus last 23 bits from IP address
    - Hashing—multiple IP addresses map to the same MAC address

Summary

- Spanning Tree Protocol
  - Port disabling
  - Bridge PDUs
  - Rapid Spanning Tree Protocol (RSTP)
- VLANs
  - IEEE 802.1Q
  - Multiple Spanning Tree Protocol (MSTP)
- Autonegotiation
  - Half/full duplex
  - Speed
  - Flow control
  - …
- Quality of Service
- Multicast

Reading Instructions

- Behrouz A. Forouzan, ”Data Communications and Networking,” third edition
  - 14 Local Area Networks: Ethernet
    - 14.1 Traditional Ethernet
    - 14.2 Fast Ethernet
    - 14.3 Gigabit Ethernet
  - 16 Connecting LANs, Backbone Networks, and Virtual LANs
    - 16.1 Connecting Devices
    - 16.3 Virtual LANs
- Backes, F., "Transparent bridges for interconnection of IEEE 802 LANs," IEEE Network, Vol. 2, No. 1. 1988