Ethernet

Moldován István

MŰEGYETEM 1782

Budapest University of Technology and Economics

Department of
Telecommunications and Media Informatics

The origins of Ethernet

Aloha

Origin. Send, then wait for ACK. If no ACK, resend after random time

Slotted Aloha

News: Send only in time slots

CSMA

CSMA = Carrier Sense Multiple Access

New: first sense for carrier, only send if no carrier detected

CSMA/CD

CD = Collision Detection

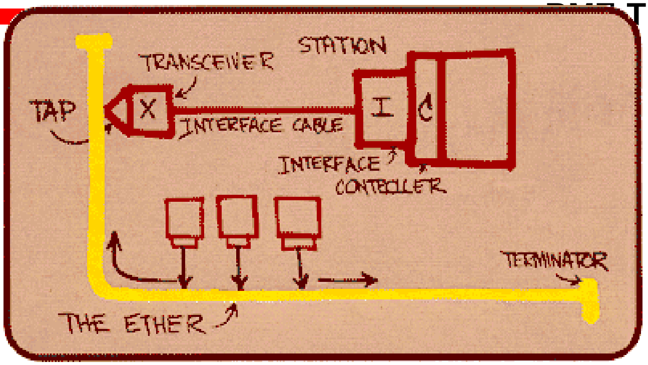
New: Stop sending when collision detected

Start of Ethernet

- 1972 Dr Robert Metcalfe

1976 first mention of Ethernet name

- The original DIX Ethernet V2 standard
  - 1982 (DEC-Intel-Xerox)
- Az IEEE 802.3
  - 10Base-5 - 1983
  - 10Base-2 - 1988
  - 10Base-T - 1990



The first Ethernet picture

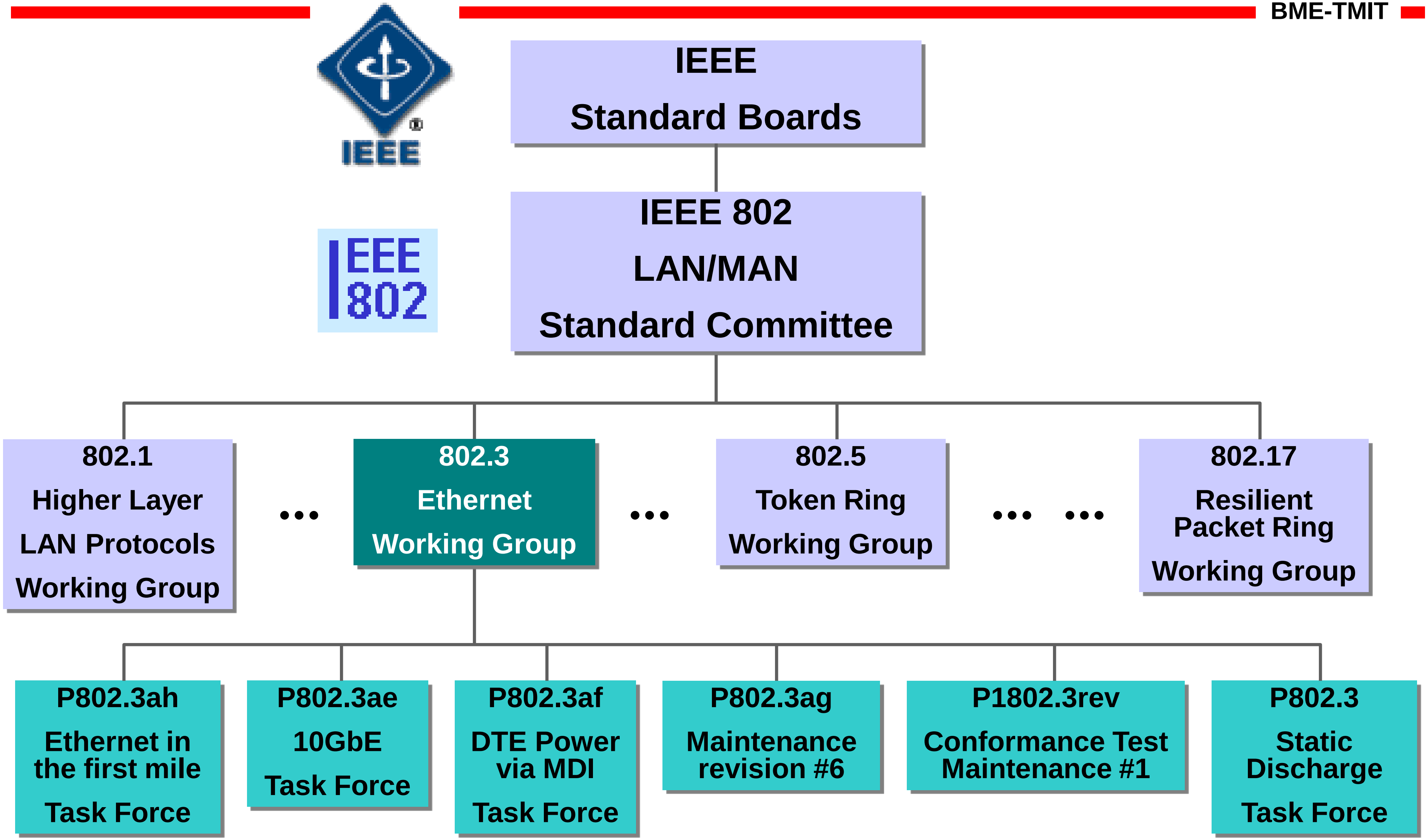
- Ethernet evolution – after 10M
  - 100BASE-TX (Fast Ethernet)
    - IEEE 802.3u: 1995
  - 1000BASE-X (Gigabit Ethernet)
    - IEEE 802.3z: July 1998
  - 1000BASE-T (Gigabit on Copper)
    - IEEE 802.3ab June 1999
  - 10 Gigabit Ethernet (IEEE 802.3ae)
    - IEEE 802.3ae 2002
  - 40GBE, 100GBE available now

Ethernet and OSI model

	OSI MODEL	TCP / IP		Exchange Unit
7	**Application Layer** Communication Type: E-mail, FTP, client/server…	**FTP, HTTP, SMTP, DNS, Telnet**	**Application Protocol**	**APDU**
6	**Presentation Layer** Encryption, data conversion: ASCII to EBCDIC, BCD to binary…		**Presentation Protocol**	**PPDU**
5	**Session Layer** Starts, stops sessions. Maintains orders.		**Session Protocol**	**SPDU**
4	**Transport Layer** Ensures delivery of entire file or message.	**TCP, UDP**	**Transport Protocol**	**Segments**
3	**Network Layer** Routes data to different LANs, WANs based on Network address.	**IP** **(ICMP, ARP, RARP)**		**Packet**
2	**Data Link (MAC) Layer** Transmit packets from node to node based on station address.	**Ethernet IEEE 802.3**		**Frame**
1	**Physical Layer** Electrical signals and cabling.			**Bits**

OSI = **O**pen **S**ystem **I**nterconnection

IEEE 802 Groups



IEEE

Standard Boards

IEEE 802

LAN/MAN

Standard Committee

802.1

Higher Layer

LAN Protocols

Working Group

…

802.3

Ethernet

Working Group

…

802.5

Token Ring

Working Group

… …

802.17

Resilient Packet Ring

Working Group

P802.3ah

Ethernet in the first mile

Task Force

P802.3ae

10GbE

Task Force

P802.3af

DTE Power via MDI

Task Force

P802.3ag

Maintenance revision #6

P1802.3rev

Conformance Test Maintenance #1

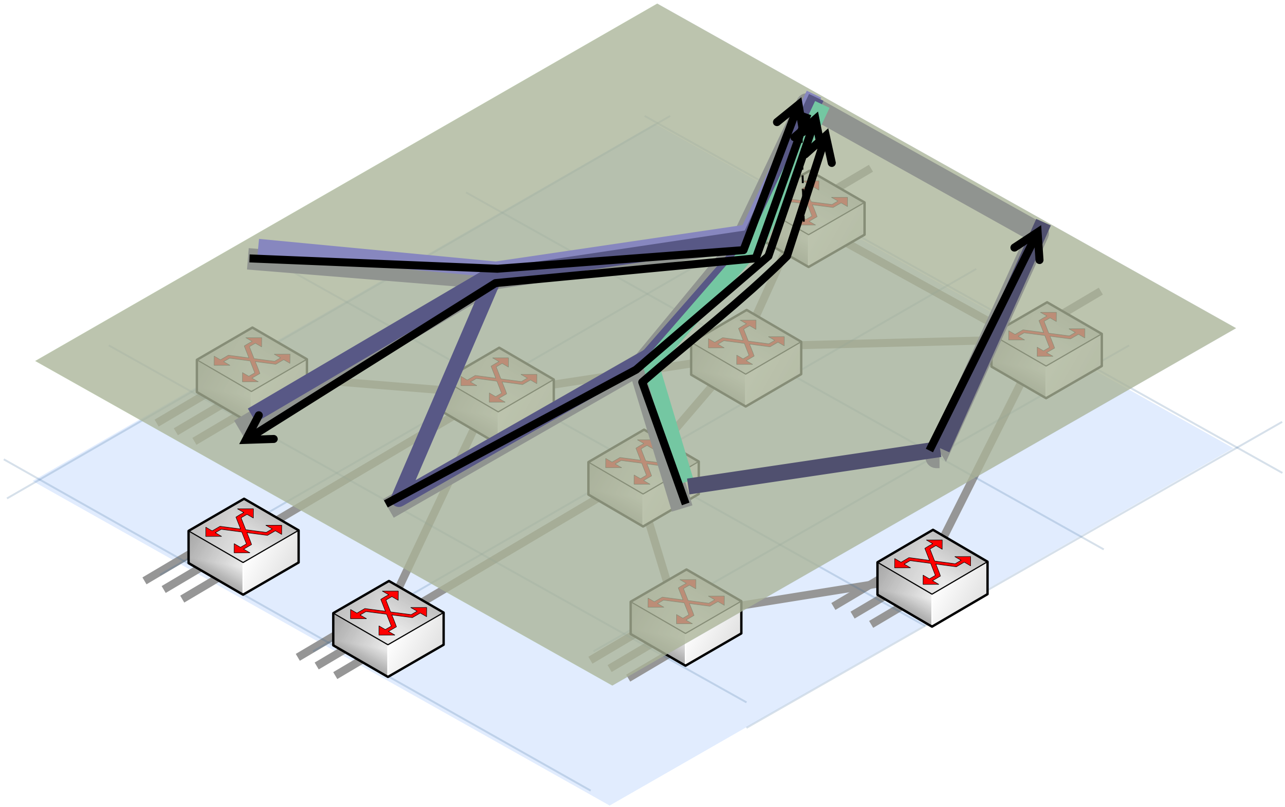
P802.3

Static Discharge

Task Force

ETHERNET OPERATION

Ethernet Forwarding



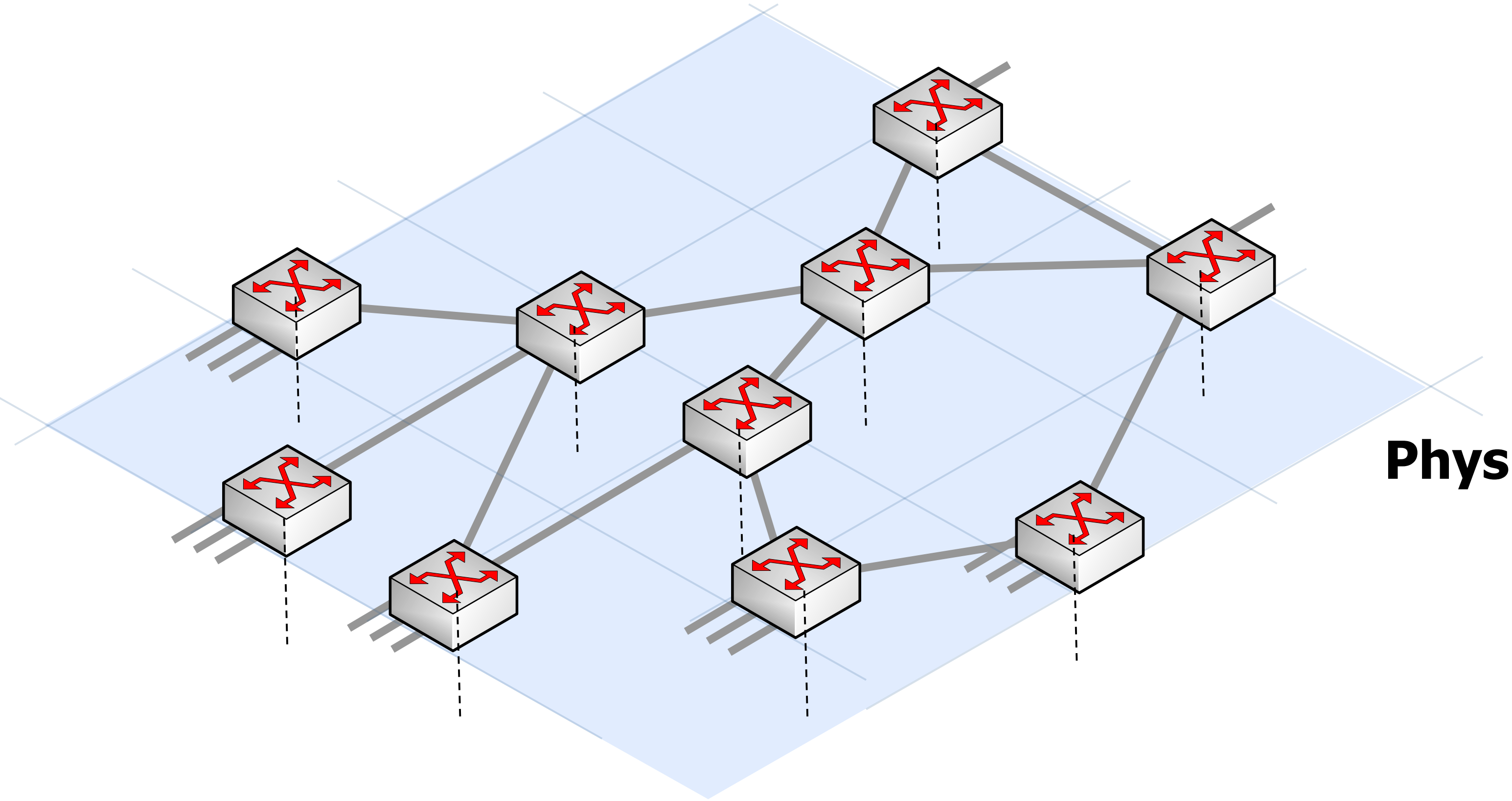
MAC Forwarding Topology

VLAN Forwarding Topology

Active (Spanning Tree) Topology

Physical Topology

Physical topology



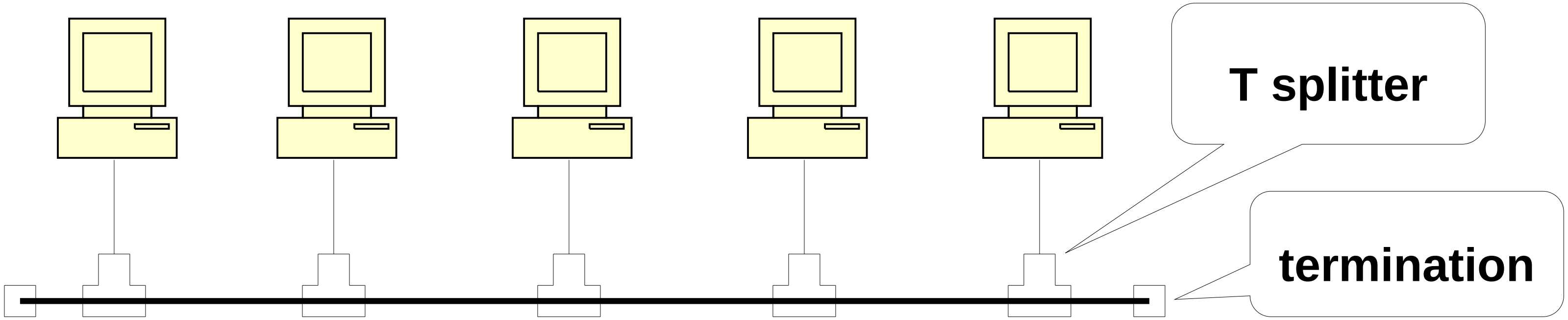
Physical topology

Physical Topology

- Ethernet Layer 2 topology
  - Determined by physical connections between switches
- It still can be an overlay topology
  - Eg. when optical overlay is used
- Properties
  - Links
  - Link speeds
  - Aggregated links (Etherchannel, 802.3ad)

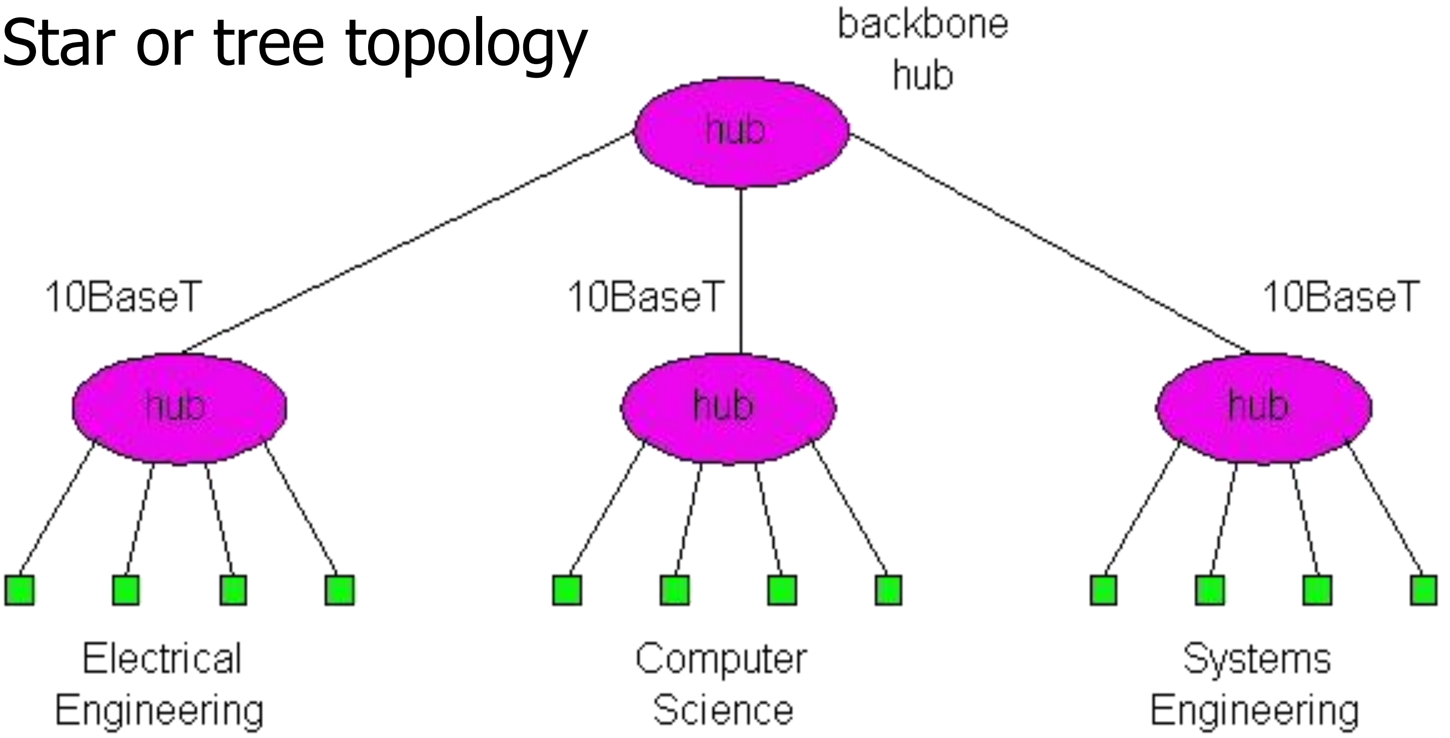
Physical connections - 1

- Coax, 10base2
  - 10: 10Mbps; 2: 200 m cable max.
  - Thin coaxial cable
- Longer distance:
  - Repeater needed



Physical connections - 2

- 10BaseT and 100BaseT up to 10GBE
  - 10, 100, 1000, 10000 MBps
  - T: Twisted Pair
  - Star or tree topology

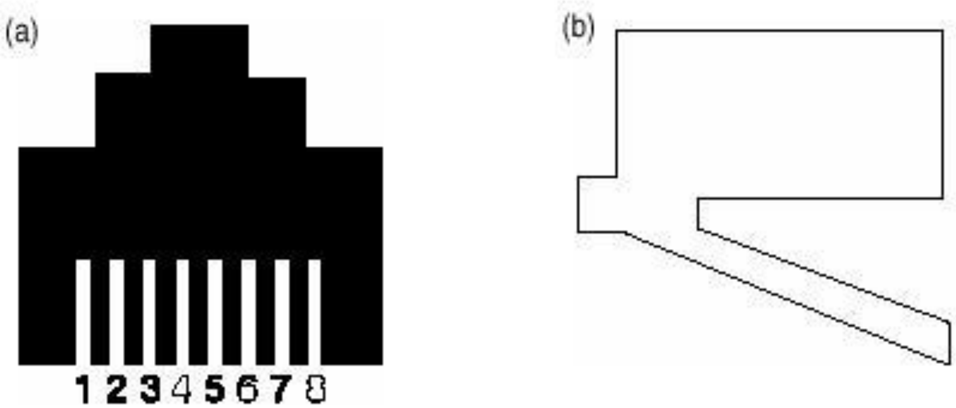


Physical connections - 3

- GE: Gigabit Ethernet
  - TX – twisted pair
  - SX/LX/FX – Optical connetion
- 10GE
  - Optical or twisted pair
- Higher speeds: 25, 40, 100Gbps
  - Usually optical, but TP also available
- 802.11: WLAN
  - It’s also Ethernet!

UTP – Category 5

- RJ-45



- **Pinout (10/100)**

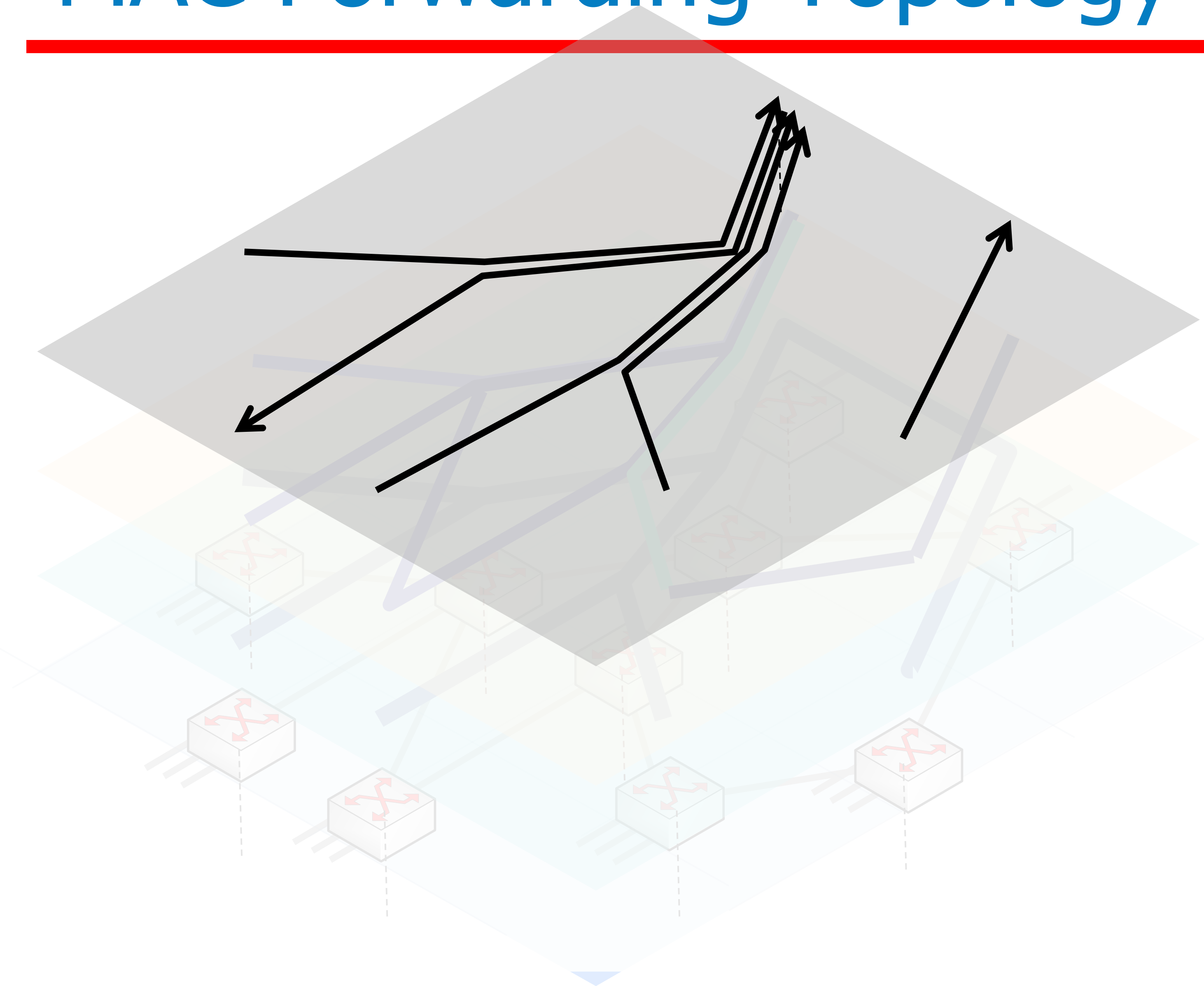
1 TD+ (Transmit Data)
2 TD- (Transmit Data)
3 RD+ (Receive Data)
4 Not used
5 Not used
6 RD- (Receive Data)
7 Not used
8 Not used

GB Ethernet uses all pairs!

Optical Ethernet

- Longer distances
  - Extends the reach up to kilometers
- Point-to-Point connection
- Usually reached with SFP modules
  - Different SFP types
    - Different distances
    - Different „colors” - WDM

MAC Forwarding Topology



MAC Forwarding topology

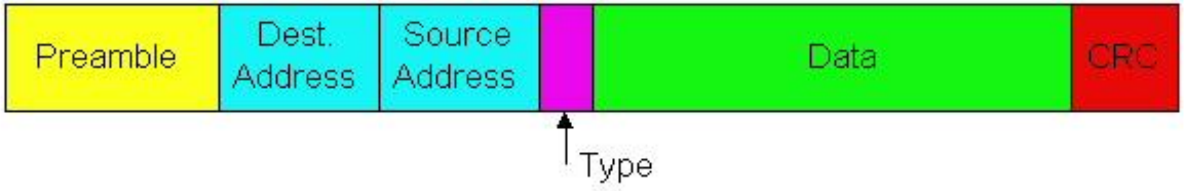
VLAN Forwarding topology

Active (Spanning Tree) topology

Physical topology

Frame format - 1

- Ethernet frame
  - I, II, 802.3 (802.2 SNAP needed for Ethernet II compatibility)
- IEEE 802.3 Data Link Control (DLC)

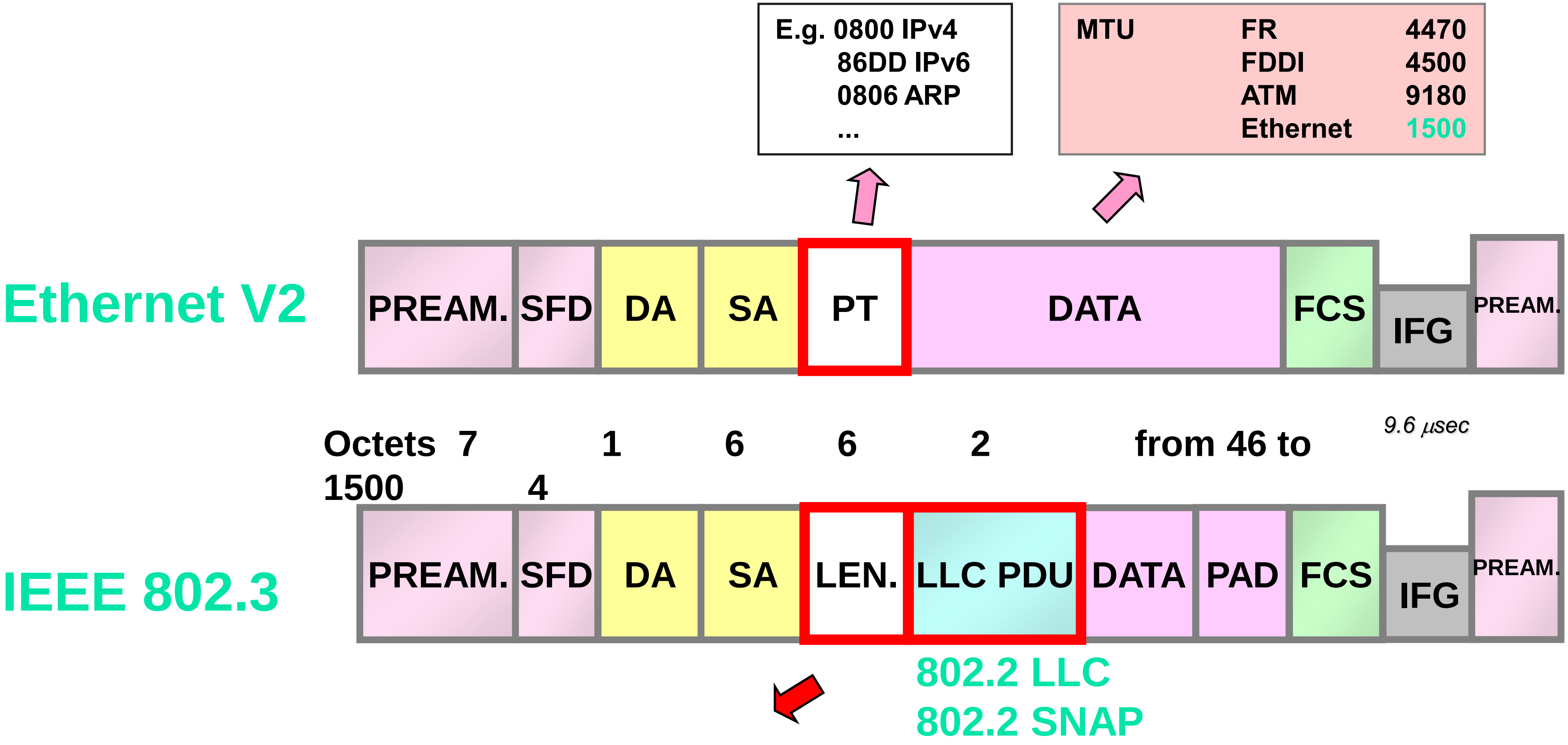


- Preamble and CRC are handled by the hardware:
  - 7 byte 10101010 followed by 10101011, needed for receiver synchronization
- IEEE 802.3 requires LLC header after the DLC

Frame format - 3

- Address: 6 byte
  - All stations receive the frame, but all drop except the one which is the destination
  - Special address: Broadcast – FF:FF:FF:FF:FF:FF
- Type field: 2 bytes
- CRC: 4 bytes, the receiver drops the frame with CRC error
- Data: maximum 1500 bytes, minimum 46 bytes
  - Maximum 9000 byte – GE Jumbo frame

Frame format - 3



E.g. 0800 IPv4
86DD IPv6
0806 ARP
...

MTU		
	FR	4470
	FDDI	4500
	ATM	9180
	Ethernet	1500

Octets 7 1 6 6 2 from 46 to
1500 4

802.2 LLC
802.2 SNAP

Difference between Ethernet V2 and 802.3

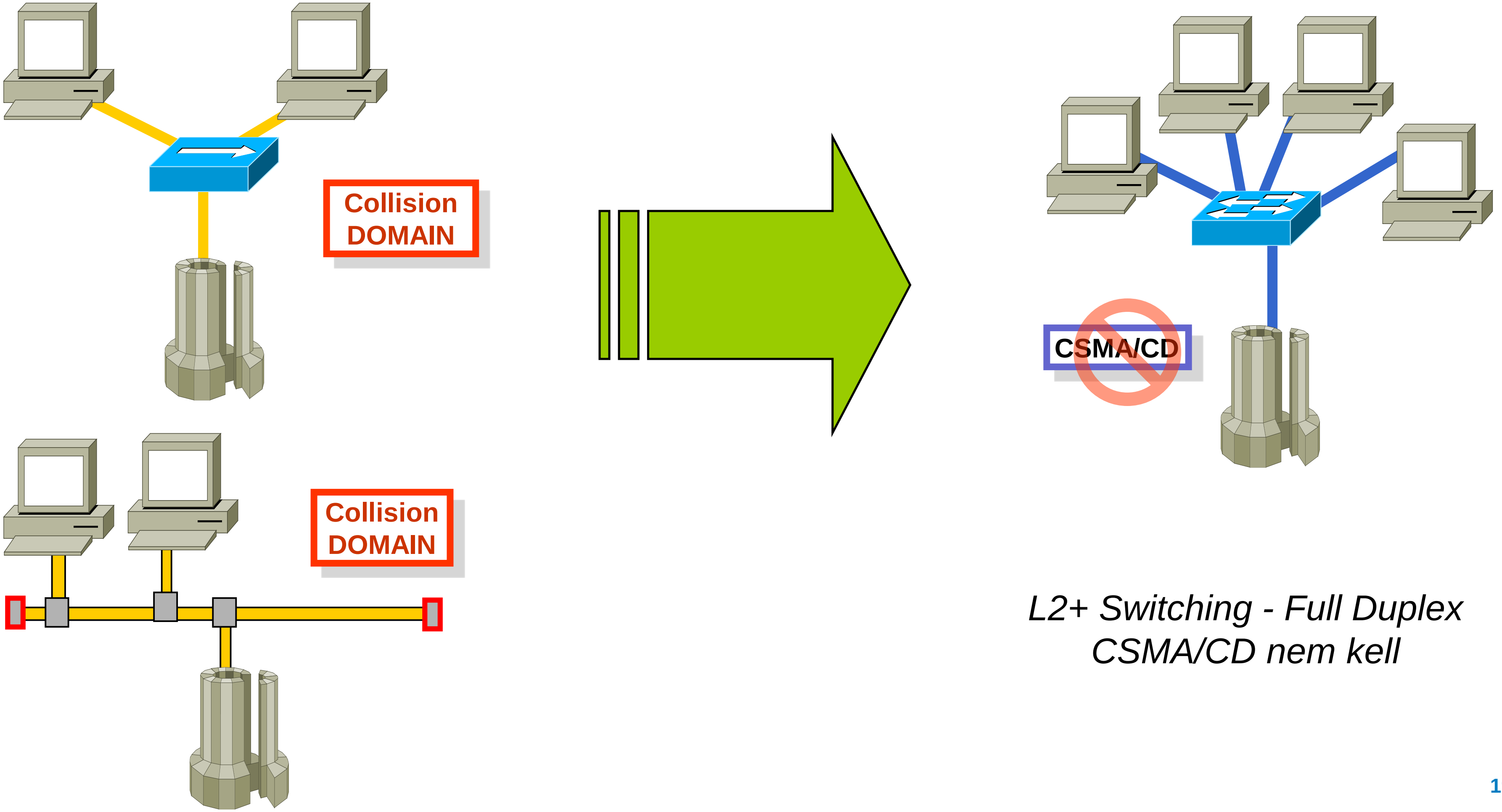
Maximum Frame Size is max.1518 (decimal), or 0x05EE Hex

EthernetV2 Ethertype is always greater than 0x05EF

http://www.iana.org/assignments/ethernet-numbers

No more collisions!

FDX & Microsegmentation
No collision



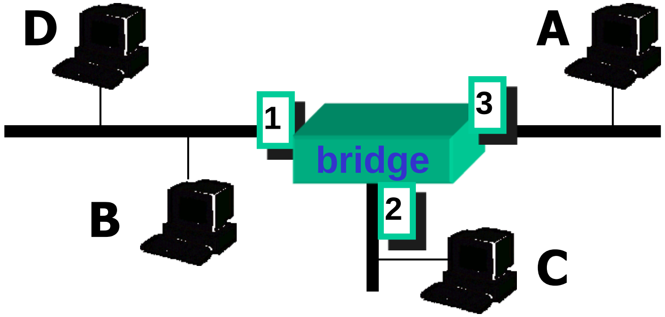
L2+ Switching - Full Duplex
CSMA/CD nem kell

Bridging - operation

- Target: transparent operation
  - Automatic plug-n-play operation
  - Automatic config
  - Cooperation with existing LAN technologies
- 3 main functionalities:
  1. forwarding
  2. MAC learning
  3. Loop avoidance: Spanning Tree

Ethernet Bridge Operation

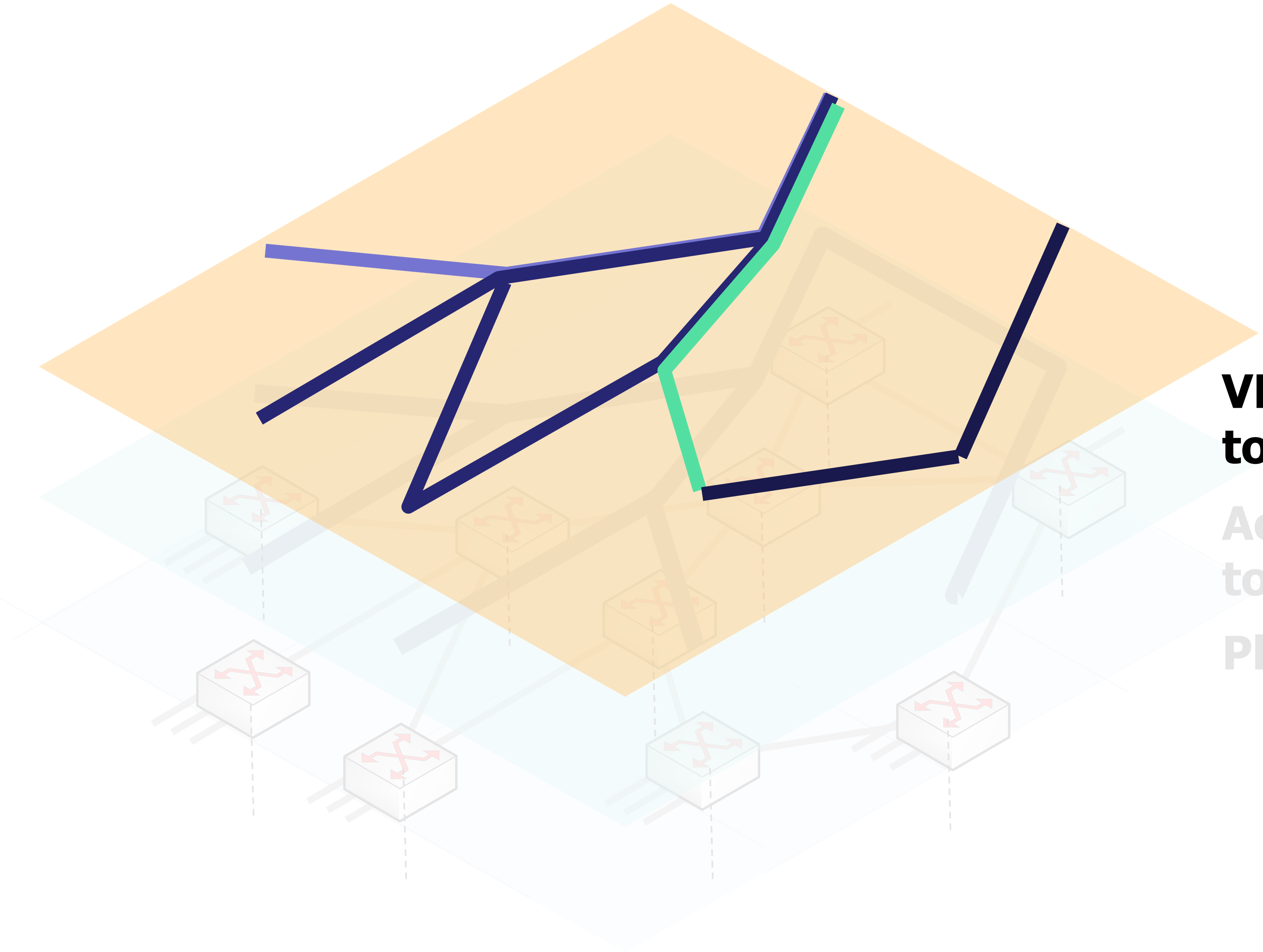
- Frame forwarding based on destination MAC address
  - MAC addresses supposed to be unique
- If destination not known: flooding
  - and learn the source MAC
- If destination MAC is already learned, forward only to that port



- Example:
  - A->D: broadcast
  - D->A: port 3
    - learn D's MAC
    - C->D: port 1

MAC addr.	Port
A	3
B	1
C	2

VLAN topology



VLAN Forwarding topology

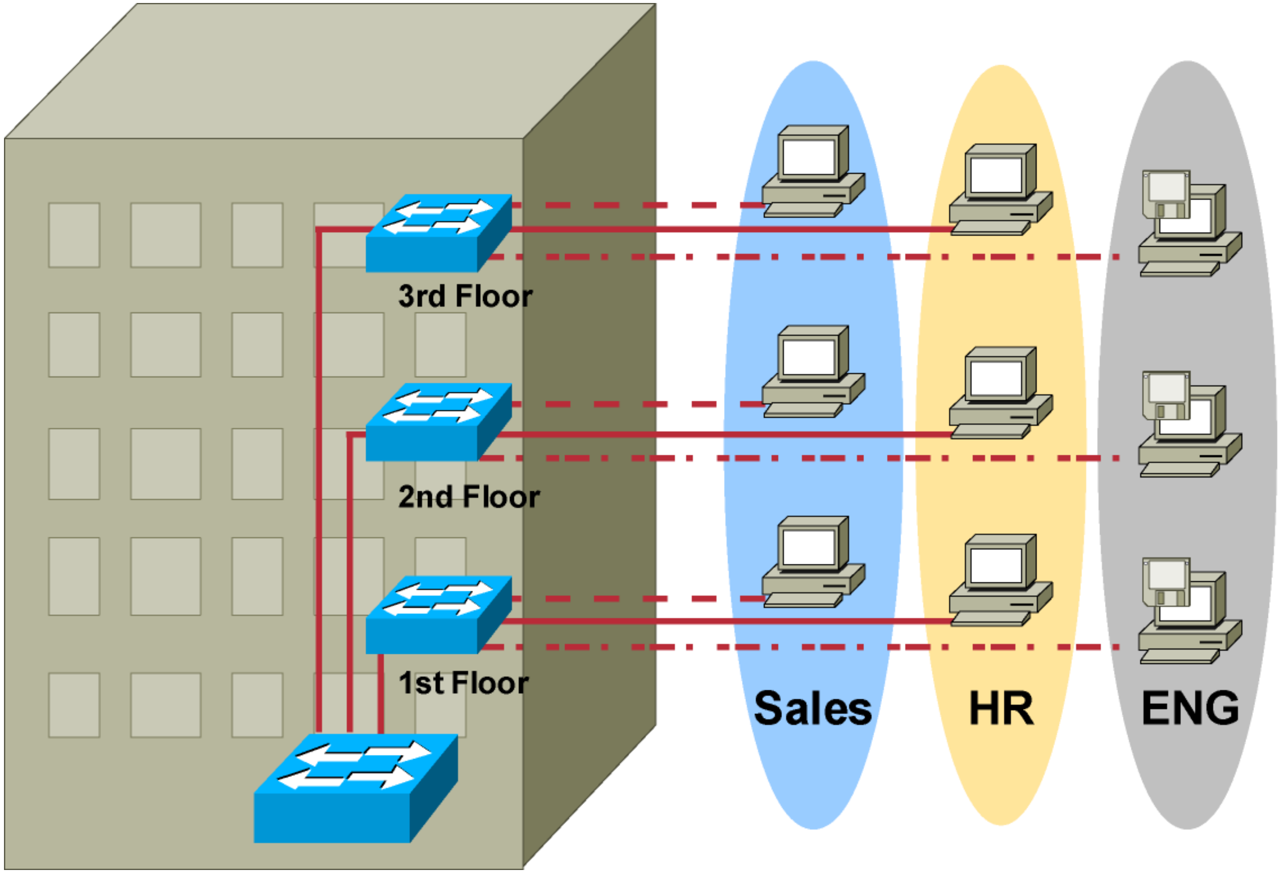
Active (Spanning T topology

Physical topology

VLAN

- LAN (Local Area Network): domain
  - Within the LAN everybody receives a broadcast
  - Limited by L3 devices (usually gateways/routers)
  - The limits are determined by cabling
  - To communicate out, router/GW is needed
  - To find an other device, adress resolution is needed (ARP)
- VLAN (Virtual LAN): administratively created broadcast domain
  - The admins determine who is in
  - Limits are virtual, not physical
  - Different VLANs do not see each other’s traffic

VLANs

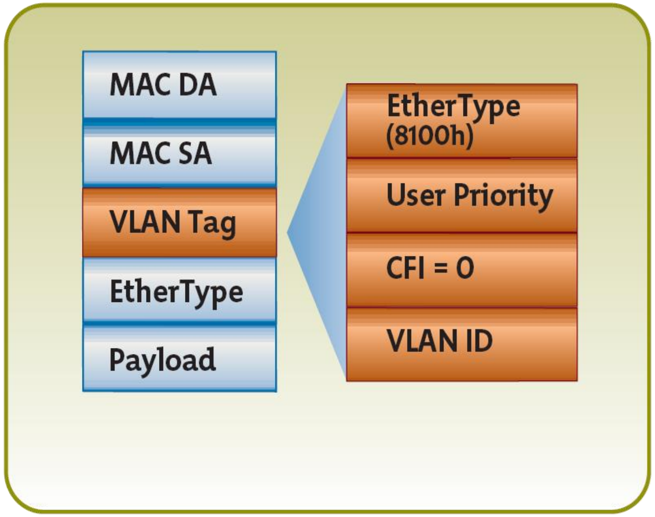


- Layer 2 connectivity
- Logical setup
- Single broadcast domain
- Management
- Security

1 VLAN = 1 Broadcast Domain = 1 Logical Subnet

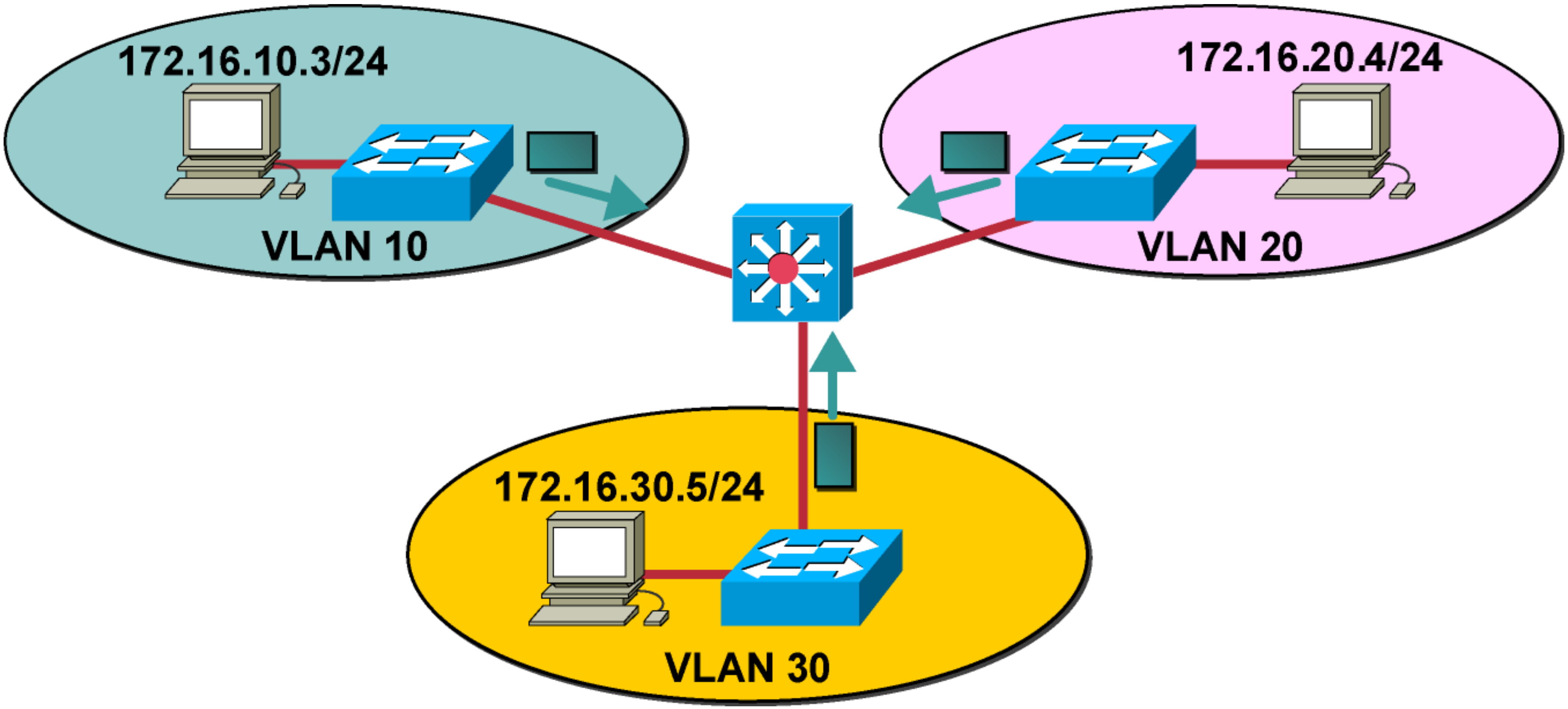
VLANs

- Virtual LANs introduced by IEEE 802.1Q
  - VLAN tag, 4096 VLANs possible
- Traffic separation by filtering
  - Filtering at ingress port
  - Filtering at egress ports
  - Does not interact with path selection!
    - It follows the Spanning Tree



- Q-in-Q, Provider Bridges (IEEE 802.1ad)
  - 4096 VLANs not enough in a provider network
  - Stacked VLANs
- Mac-in-Mac, Provider Backbone Bridges (IEEE802.1ah)
  - Solves MAC address scalability by MAC encapsulation

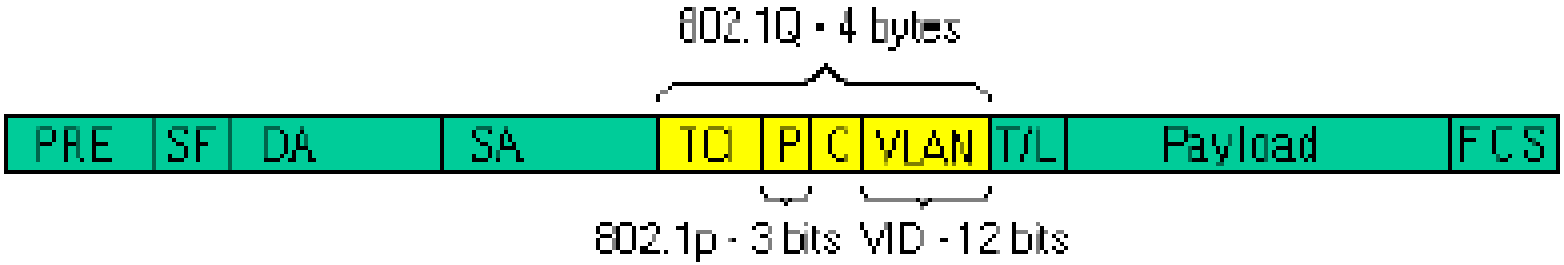
Traffic between VLANs



- No level 2 connection
- Only through an IP level router/gateway

Tagged Frame

- TCI (Tag Control Info): 8100 shows 802.1p/Q VLAN
- P: priority(0..7)
- C (Canonical Indicator): used for Token Ring
- VLAN: VID (0..4095)



VLAN operation - Filters

- Ingress filtering
  - Filtering if packets are tagged
  - Tagging if required
- Switching
  - As usual, based on learning bridge operation
  - Flooding if needed
- Egress filtering
  - Filter outgoing
  - Remove tag if needed

VLAN tagging

- Port-based VLANs: physical inteface based
- MAC-based VLANs: preconfigured MAC table
- Protocol-based VLANs: VLANs for each protocol: UDP, TCP, or even higher
- IP subnet based(not used)

VLAN trunk

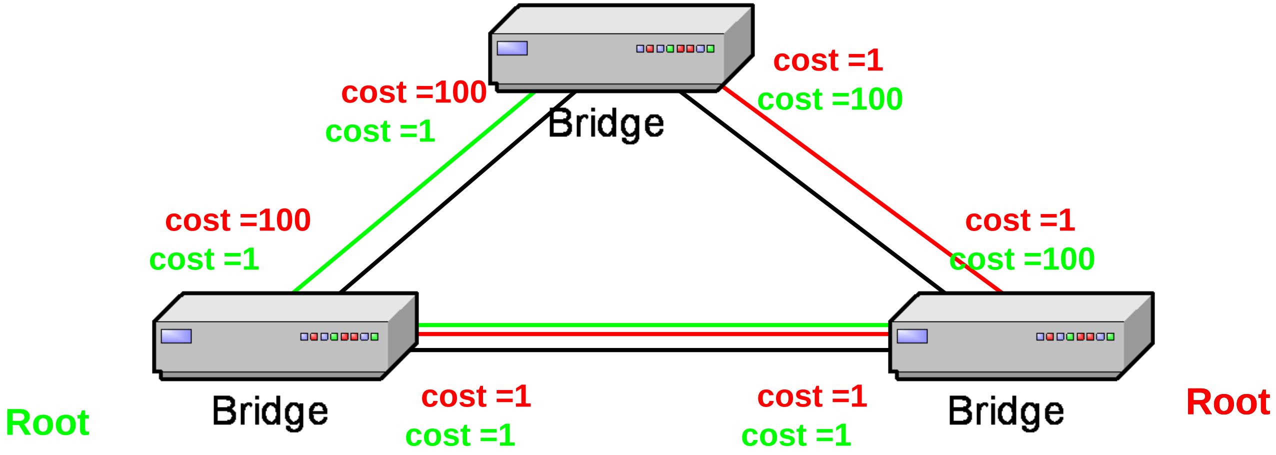
- On the uplink
  - „trunk port”
  - Tagged packets only
  - Filtering
- The trunk may also be „untagged”
  - Remove tag after filtering at egress

With VLANs and MSTP we can do

- Protection
  - Multiple disjoint trees
  - VLAN 1 assigned to primary tree, VLAN 2 to backup tree
  - On failure, traffic is switched to VLAN 2, using the backup tree
  - (requires IP level switching/failover logic)
- Traffic Engineering
  - Load balancing
  - paths can be “engineered”
  - traffic mapping to different engineered paths
- Of course, for a simple tree physical topology it is useless

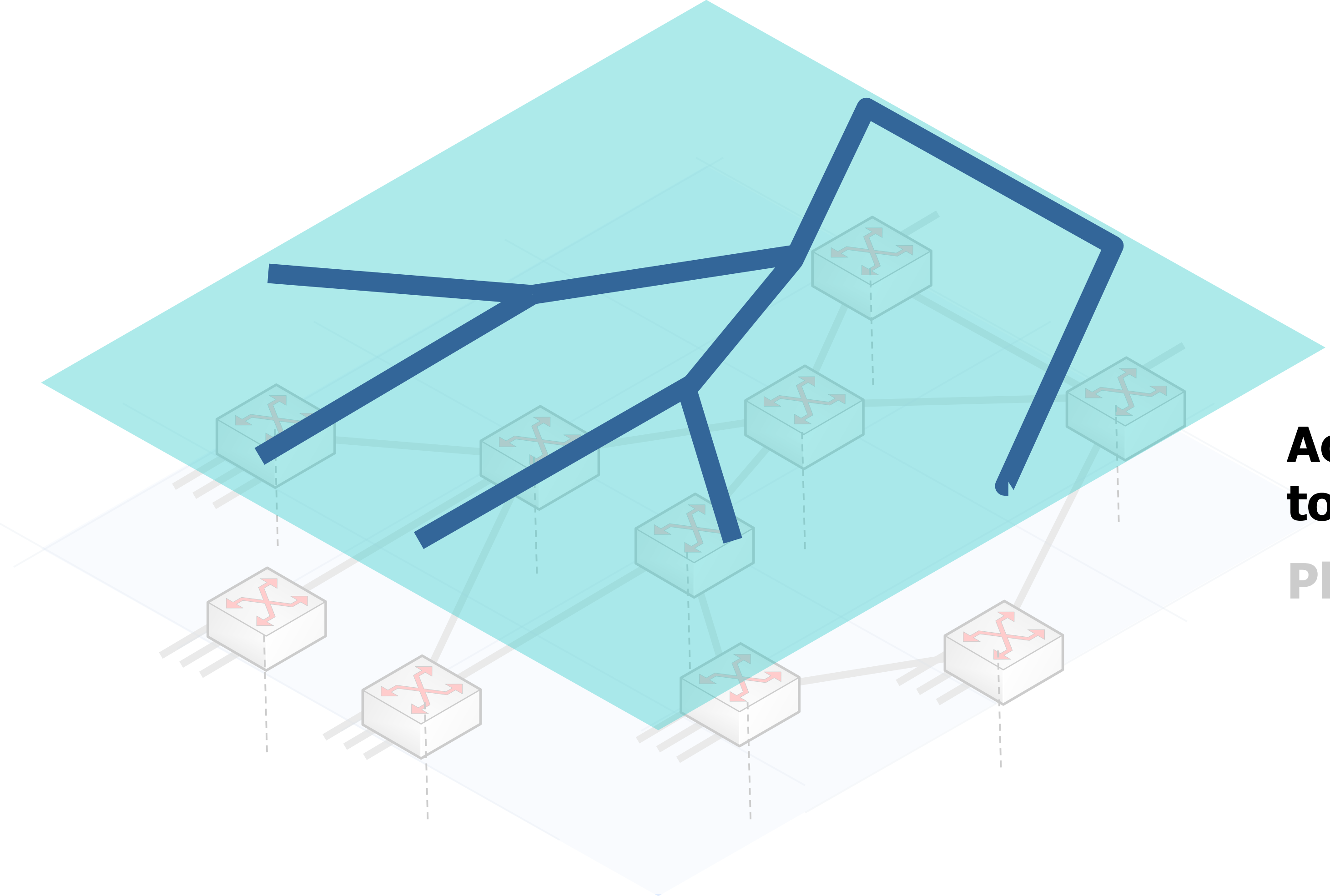
MSTP optimization

- MSTP requires configuration
- Trees are set up by setting different port costs



- Port cost assignment:
  - 1 for forwarding, (#of bridges+1) for blocking

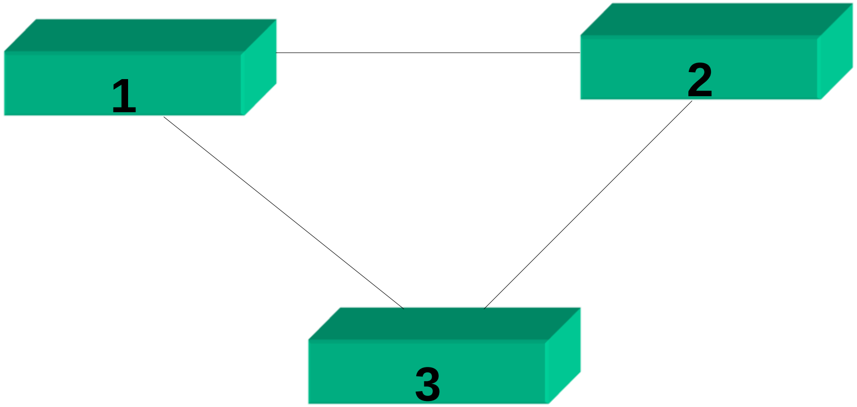
Active Topology



Active (Spanning T
topology

Physical topology

Redundancy - loop



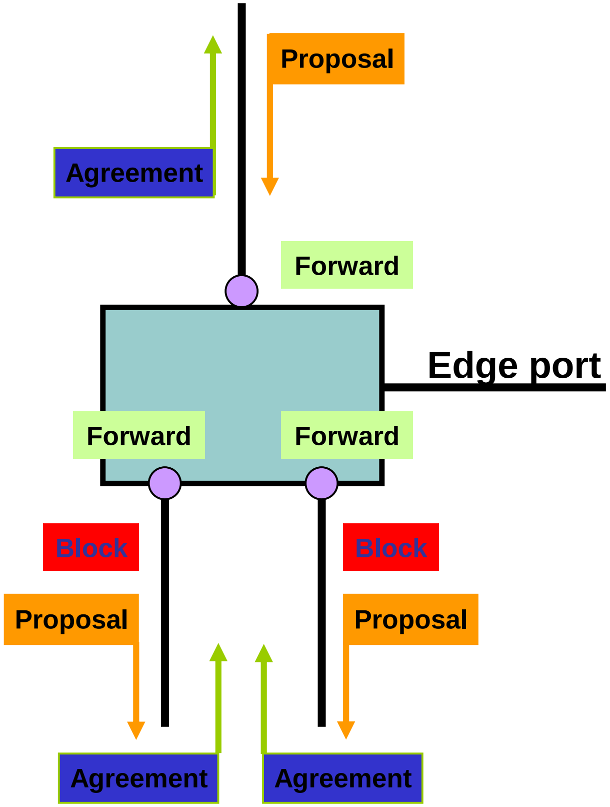
1. Broadcast packet arrives at 1. It is forwarded to 2 and 3
2. 2 sends to 3
3. 3 sends to 2
4. 2 and 3 both send it back to 1
   - Loop!

STP Bridge

- Avoid loops
  - Reduces topology to a tree
- Learning bridge based
- Packets travel along the tree only
  - In the direction of the root
- 802.1d

IEEE 802.1w sequence of events

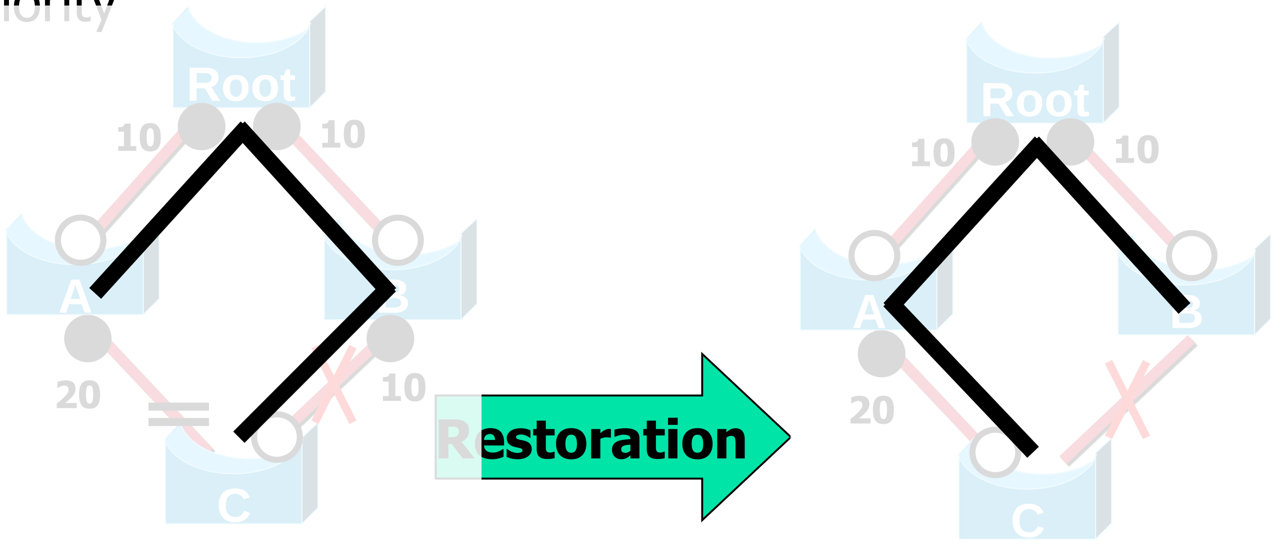
- Receive a proposal
  - Block all other non-edge ports
- Send an agreement back
  - Put the new root port to forwarding
- Send out proposals on other ports
- Receive agreement from others
  - Put ports into forwarding



RSTP operation

- Distributed operation
  - Uses BPDUs to communicate

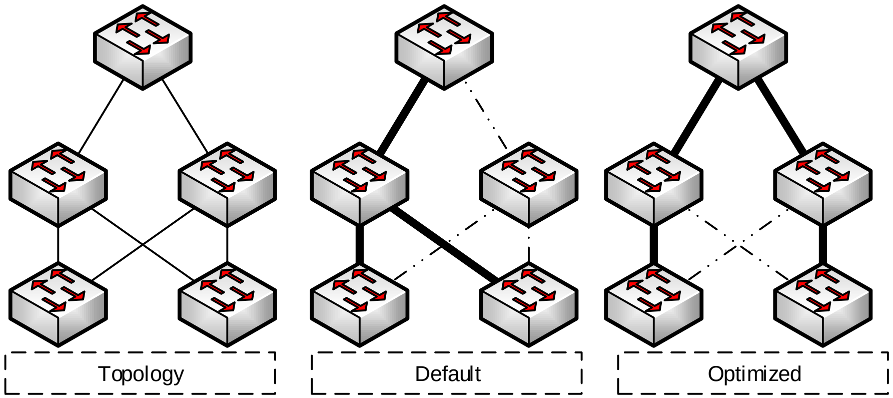
- Parameters affecting the active topology
  - Bridge ID (priority)
  - Port cost, priority

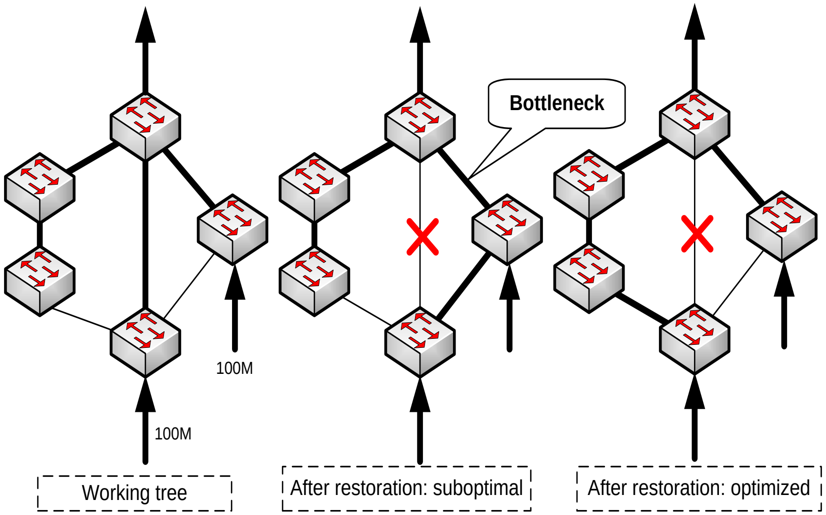


- The resulting topology is unambigously determined

RSTP optimization

- RSTP constructs the loop-free forwarding topology based on link cost and bridge ID
  - May not be optimal
- In case of failure
  - With default cost set we don’t have bandwidth guarantees
    - – The restored topology may also be suboptimal
  - With optimization we give bandwidth bounds even after restoration (if possible)





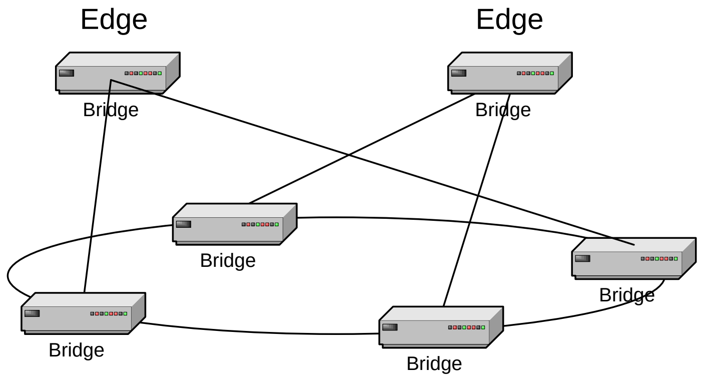
MSTP

- RSTP disadvantage: bad resource utilization
- Cisco: PVST (Per-VLAN Spanning tree)
  - Each VLAN: an RSTP
  - Many VLANs – not scalable, unnecessary
- IEEE: MSTP
  - Multiple spanning trees
  - VLANs assigned to trees

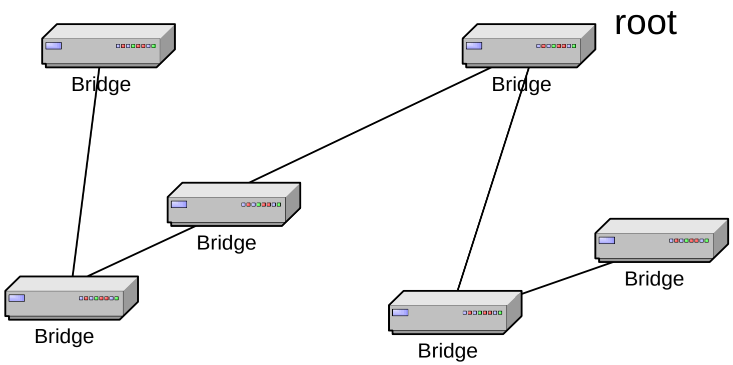
MSTP operation

- RSTP based, technology upgrade
- Max. 64 tree(MST instance)
- For each tree we can set
  - root
  - Link cost/priority
  - VLAN assignment
- 1 VLAN to 1 tree only!

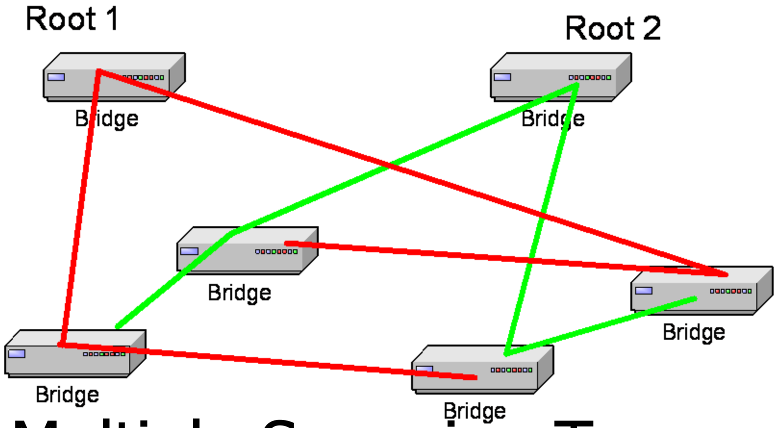
MSTP Advantages



- Network Topology: 2 exits
- Ring - redundancy
  - Higher reliability



- STP: one tree

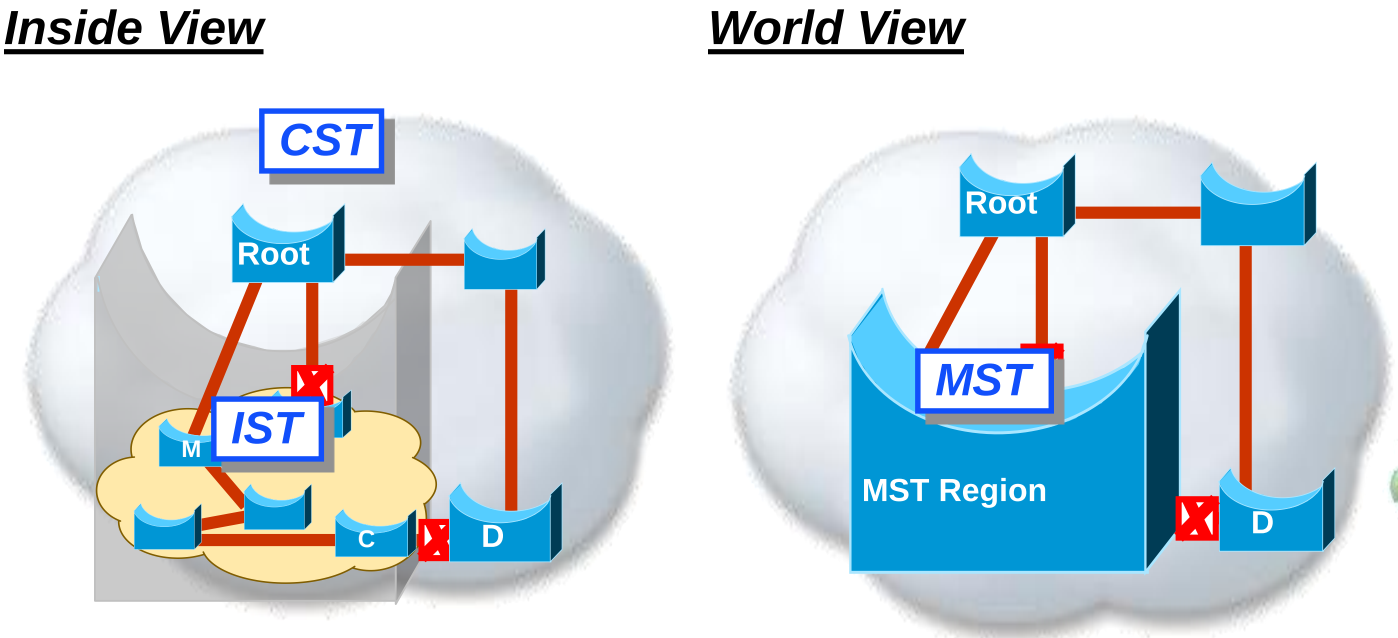


- Multiple Spanning Tree
  - 2 trees

Evolution to multiple trees & regions

- Why regions?
  - Different administrative control over different parts of the L2 network
  - Not all switches in the network might run/support MST - different kinds of STP divide network into STP regions
  - All benefits of MST are available INSIDE the region, outside it is single instance (topology) for all VLANs
- MST region is a linked group of MST switches with same MST configuration
  - Inside region: many instances
    - IST – Internal Spanning Tree (instance 0), always exists on ALL ports
    - MSTI - Multiple Spanning Tree Instance
  - Outside of region: one instance

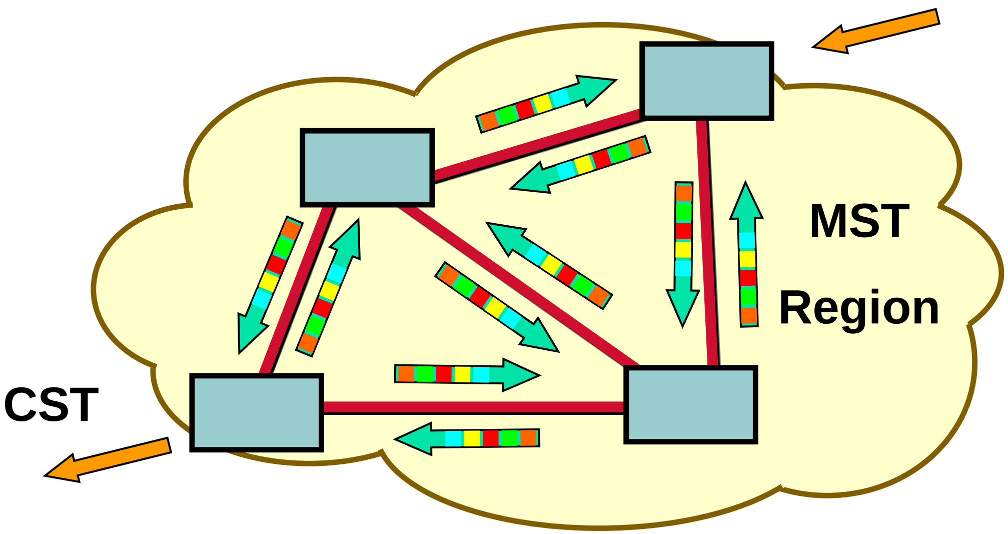
802.1s: CST, IST, MST - Lots of Trees ...



- **CST 802.1Q Common SPT** => Single Instance only
- **IST 802.1s Internal SPT** => receives and sends BPDUs to the CST represents the MST to the Outside World as CST Bridge
- **MST 802.1s Multiple SPT** => represent several VLANs mapped to a single MST Instance

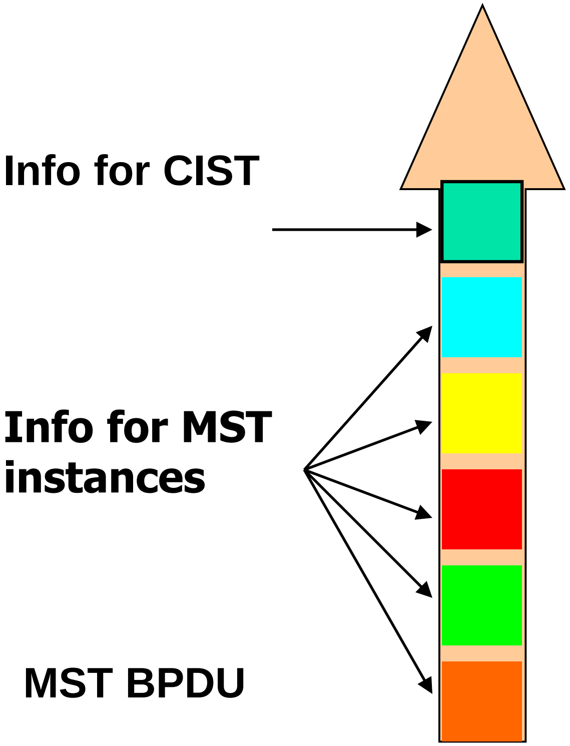
MST instances

- MSTIs are STP instances, defined only in a region
- MSTIs are not connected to the outer world
- One BPDU is sent with info for all trees
- Only one has timer related parameters (IST instance)
- The MST BPDUs are sent on all ports
- BPDUs are sent in all directons unlike in 802.1D where designated bridge sends only



MST Region

CST



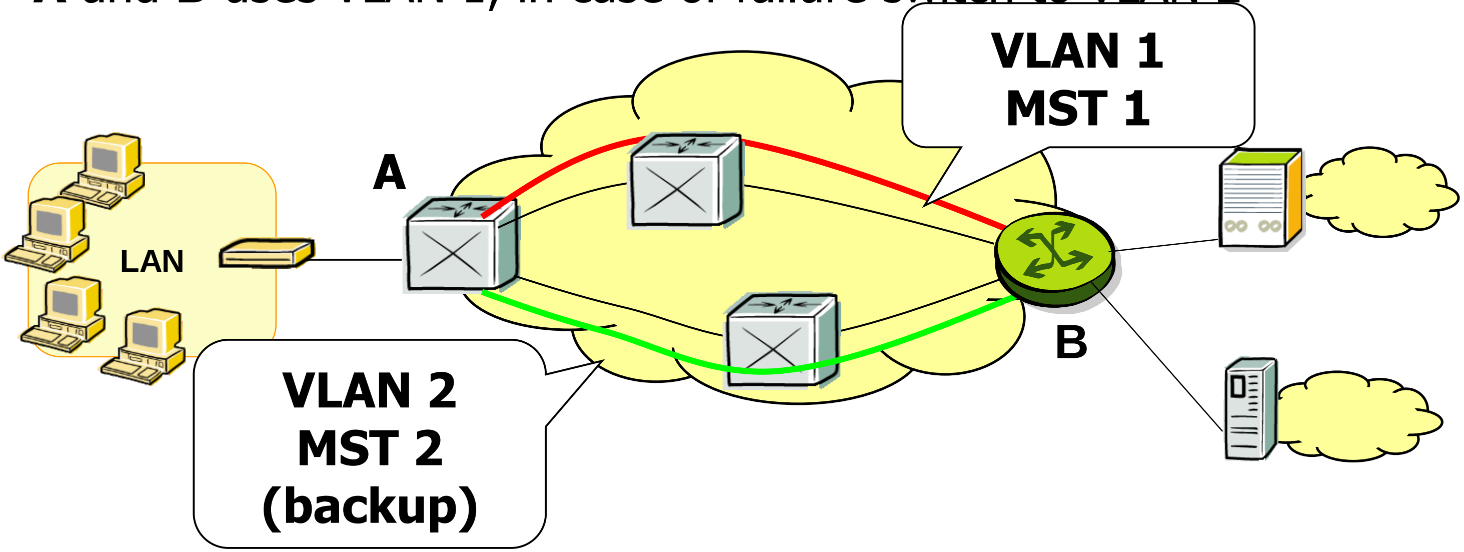
Info for CIST

Info for MST instances

MST BPDU

Protection switching

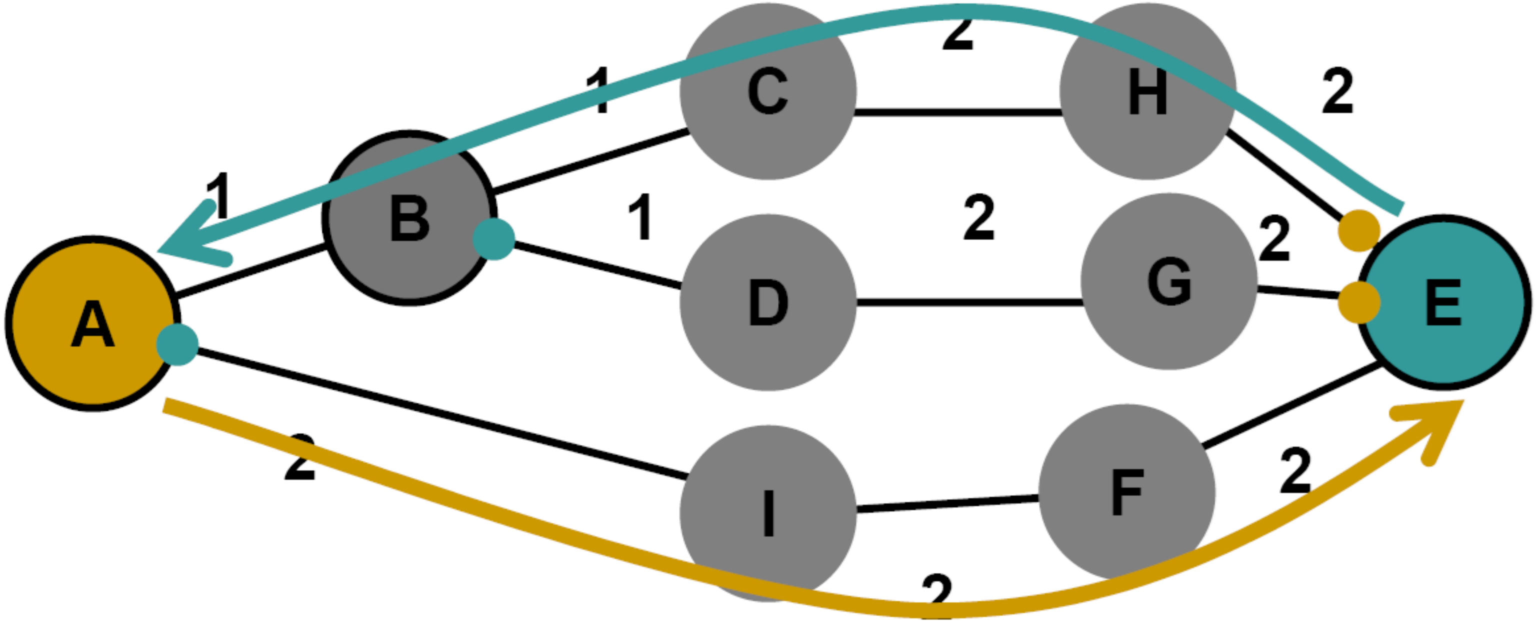
- Using MSTP
  - 2 MSTI trees, two paths: red and green
  - VLAN 1 -> MST 1, VLAN 2 -> MST 2
  - **A** and **B** uses VLAN 1, in case of failure switch to VLAN 2



- **Alternatives: 802.3ad Link Aggregation**
  - **uses redundant links for load balancing and protection**

Shortest Path Bridging

- IEEE 802.1aq
- Multiple trees rooted at each bridge
  - Each using shortest path
- Problem
  - MAC learning requires symmetrical paths


A
B
C
D
E
F
G
H
I
1
1
1
2
2
2
2
2
2
2
2