Advanced Networks

UNIVERSITÀ DEGLI STUDI DI TRENTO

http://disi.unitn.it/locigno/index.php/teaching-duties/

A primer on modern LANs

Renato Lo Cigno

Copyright

Overview

- Modern LAN are all based on the IEEE 802 standard family
  - They are “switched”
  - Mostly rely on “fast ethernet” (& beyond) and WiFi (802.11)
- A switched LAN is a complex network
  - Can be hierarchical and support “virtual LANs”
  - Can have Routers “embedded” that provide subnetting of the IP addressing space
  - Can mix public and private addresses
  - Normally has a “frontear” protected by firewalls, where NAT (Network Address Translation) functions are also performed
- A switched LAN requires routing
  - Spanning Tree
  - Fast Spanning Tree and Beyond

Understanding LANS

- To grasp the world of modern LANs the key point is the 802.1 standard suite that defines the “interworking” environment
  - http://www.ieee802.org/1/
- Unfortunately the readings are huge and many recent documents are not public
- LANs are Ethernet ...
  - ... Ethernet is CSMA/CD ...
  - ... All I need to know is CSMA/CD 1-persistent with binary backoff
  - !?!?!
- Actually today “Ethernet” is only legacy and framing

Understanding LANS

- Switches come in many shapes, forms, size, performance and ... price!
- Not all switches are equal
  - Store and Forward
  - Cut through
  - Buffering
  - Backpressure to sources

- Switches solve the problem of collisions
  - but they do not solve the problem of sustained congestion

Switches

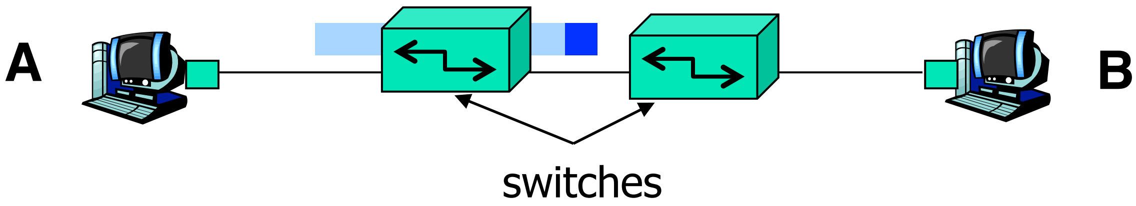
- Switches use the MAC address (like standard bridges) to take forwarding decisions
- The decision is taken via “backward learning”
  - Destinations on a port are learned reading the source address of packets incoming into the port
- Low end switches are little more than a cable concentrator
- Rarely they offer a throughput higher than 1-2 times the line speed
- Store & Forward switching
  - Similar to routing
  - High Switching time Ts (Ts > 1-2 transmission times Tx)
  - Tx = Packet size / Transmission speed

Switches

- High end switches often use the: cut through technology
  - The frame is not stored
  - Forwarded "on the fly" reading the destination MAC as the frame is decoded
  - Ts is just a few bytes transmission times, 2 or more orders of magnitude less than a S&F switch, as low as hundreds of ns
  - Cannot check the integrity of frames
- Modern LANs include the networks in Data Centers where performance issues are exasperated
- Cut through becomes fundamental to reduce latency in data access
- Data intensive, distributed computation makes the network one of the major bottlenecks

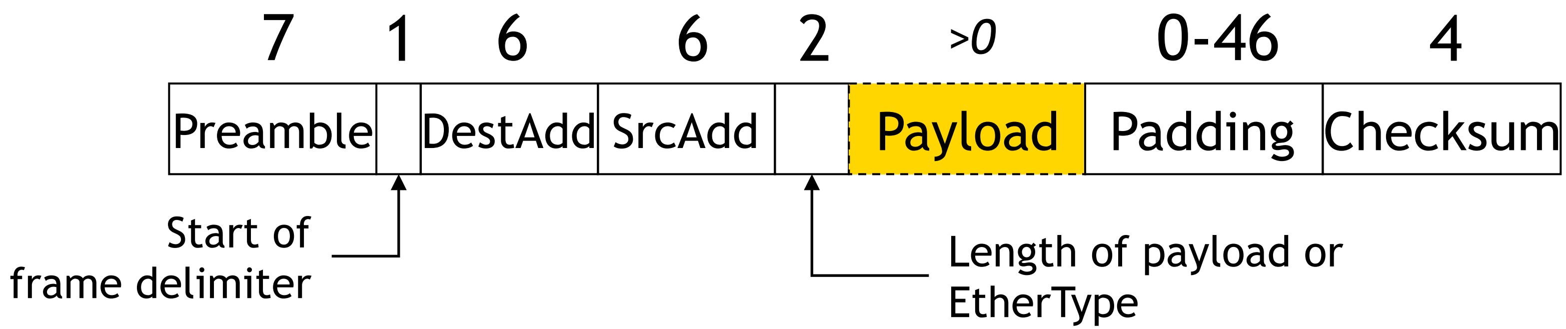
Cut Trough

- Start forward transmission as soon as possible
  - Do Look-up while inspecting header
  - If outgoing link is idle, start forwarding the frame
  - If frame is corrupted is forwarded all the same
- Transmission spans multiple links
  - Transmit the head of the frame via the outgoing link while still receiving the tail via the incoming link



locigno@disi.unitn.it

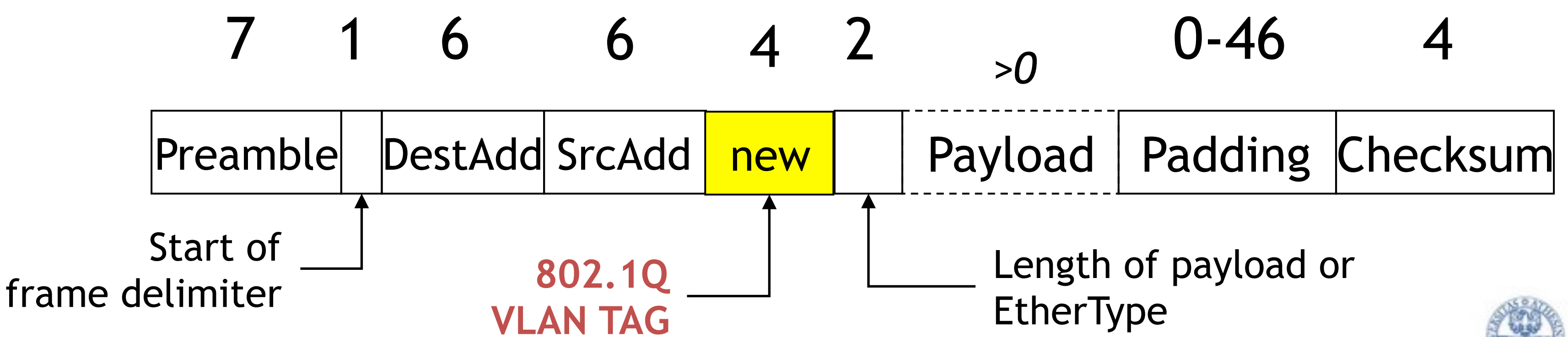
Ethernet & IEEE 802.3 frame format (legacy)



- Preamble (7 byte)
  - synchronizing sequence "10101010
  - Start of frame (1 byte) "10101011"
- Addresses (6 byte)
  - Desitnation and source address of the frame
- Length or type (2 byte)
  - lenght of the frame in bytes (0-1500)
  - if > 1536 means Protocol Type
- Payload
- Padding
  - guarantees the minimum frame length
- Checksum

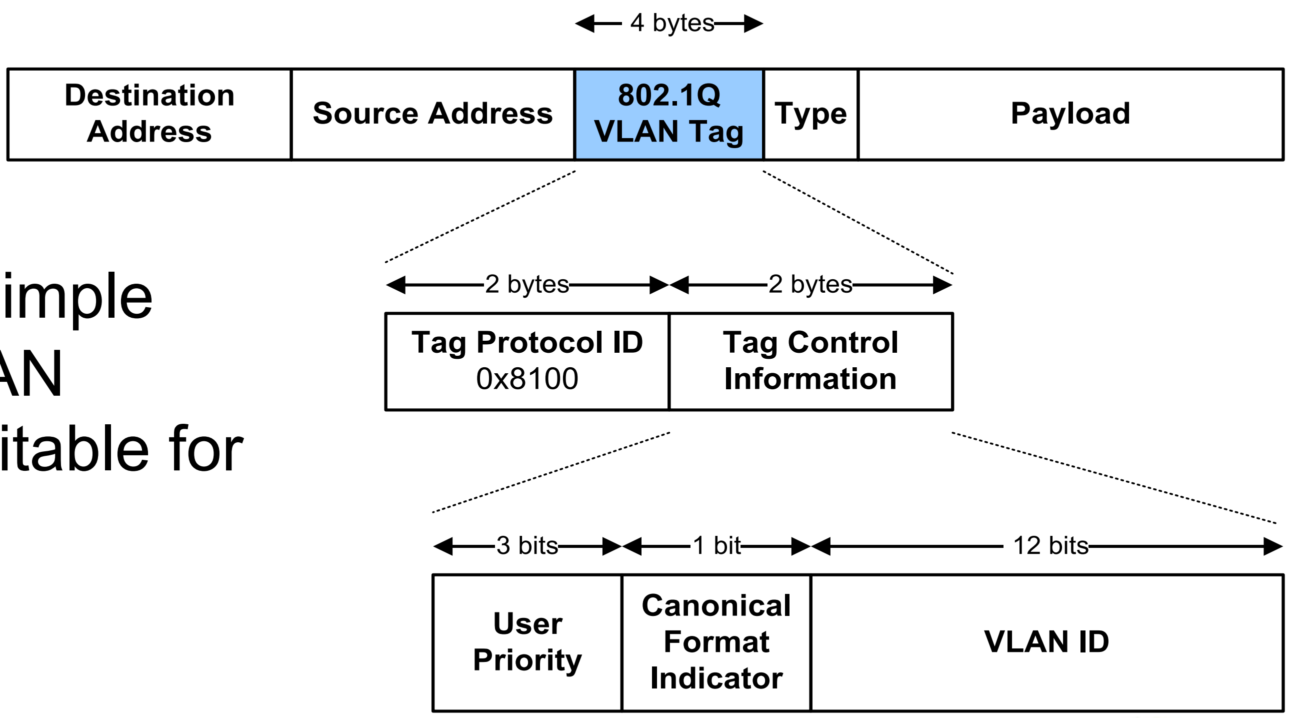
VLAN frame format

- VLANs can be defined at L1/2/3, but we’re concerned only with L2 MAC-based dynamically configurable VLANs
- The orignal frame format has been extended to support many new features and protocols
  - 802.1ad (Q-in-Q)
  - MPLS (Unicast and Multicast)
  - 802.1ae (MAC security)
  - 802.1x → authentication for LANs (EAP, EAPOL, ...)
  - ...
- 4 bytes added before the EtherType field



IEEE 802.1Q: VLAN Tagging

- Tag are normally transparent to endsystems
- VLAN tags are added/stripped by switches

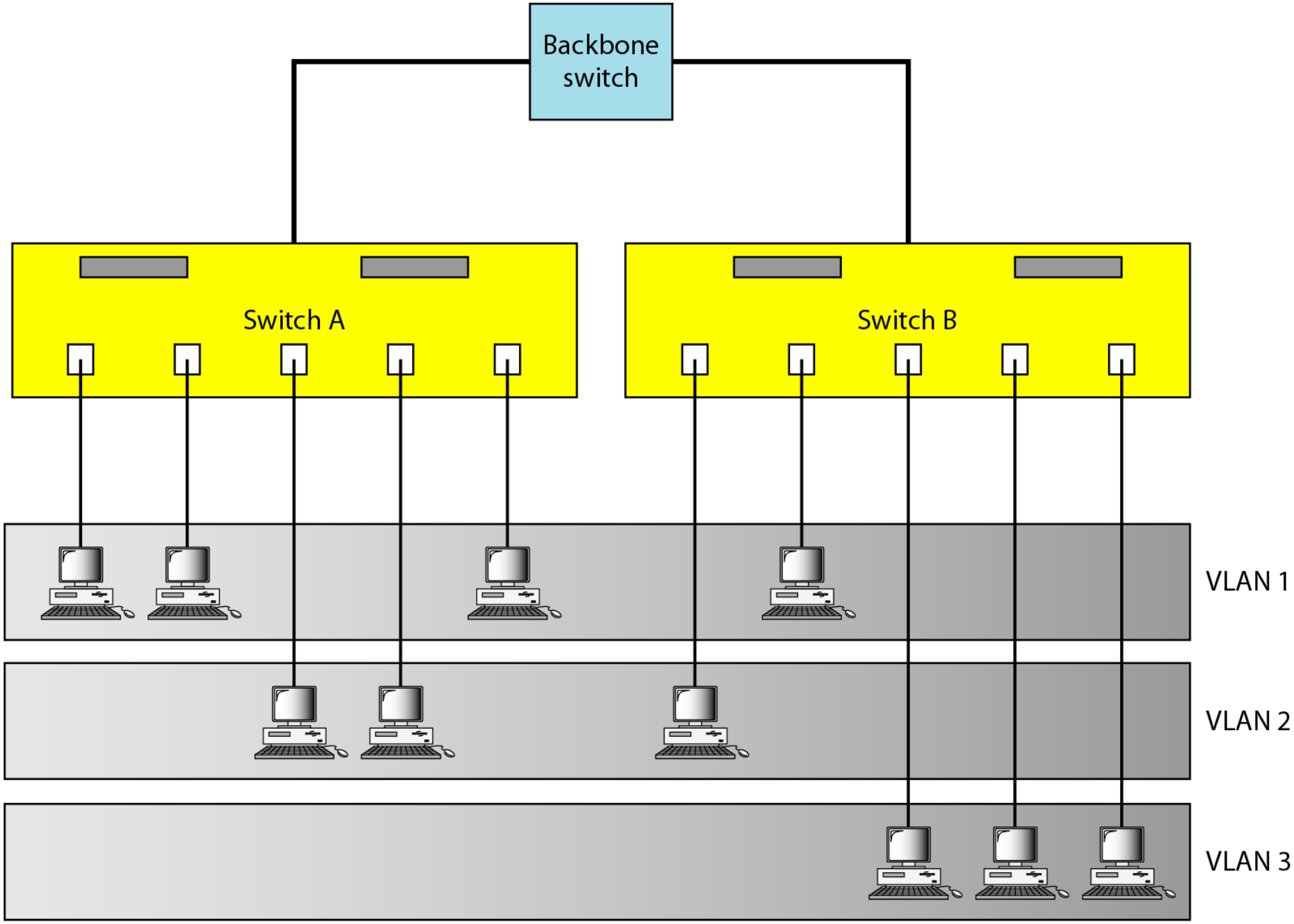


- 1Q provides a simple single-layer VLAN Environment suitable for simple LANS

802.1Q Tag Fields

- **Tag Protocol Identifier:**
  - Value 0x8100 identifies 802.1Q tag
- **User Priority:**
  - Can be used by sender to prioritize different types of traffic (e.g., voice, data)
  - 0 is lowest priority
- **Canonical Format Indicator:**
  - Used for compatibility between different types of MAC protocols
- **VLAN Identifier (VID):**
  - Specifies the VLAN (1 – 4094)
  - 0x000 indicates frame does not belong to a VLAN
  - 0xfff is reserved

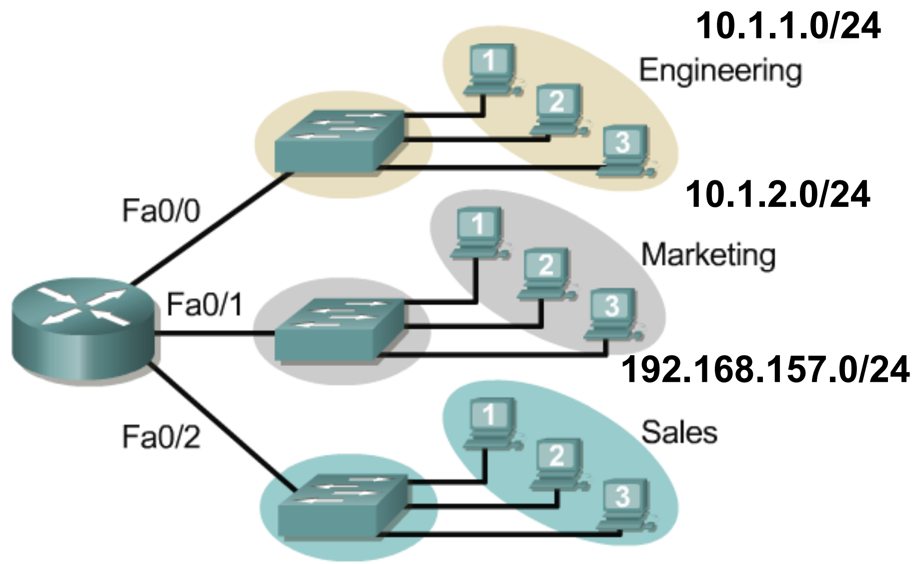
VLANs define logical broadcast domains



Broadcast domains with VLANs

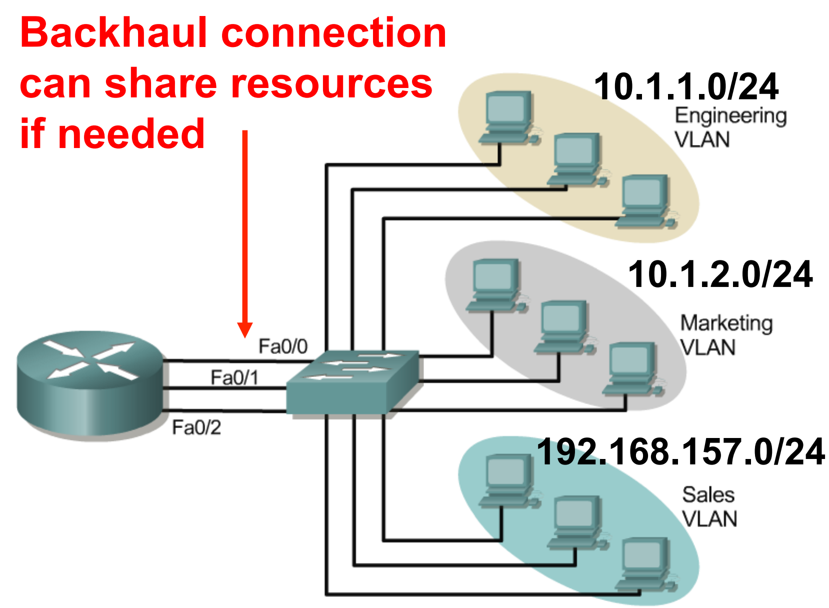
1) Without VLANs

User groups can be divided by subnets, but must be also on different switches to enforce separation



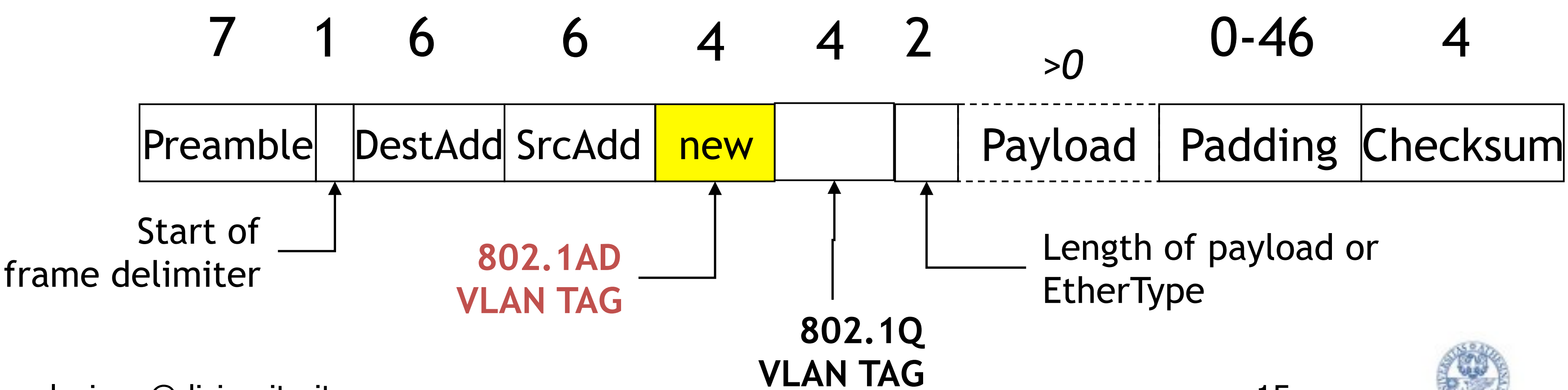
2) With VLANs

User groups can be divided by subnets, and be connected to the same switch, or be spread on different switches (see slides before)

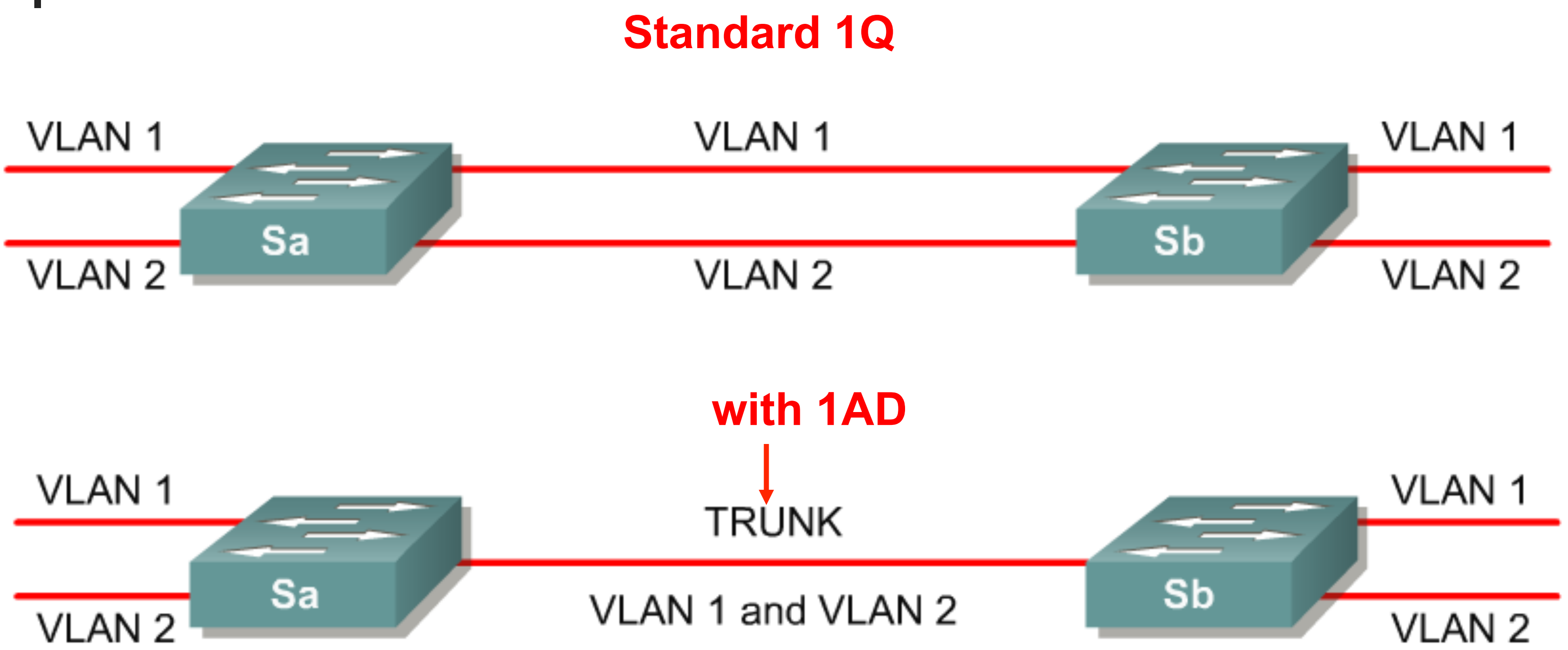


IEEE 802.1AD (Q-in-Q)

- Add another 4 bytes and enables up 16 millions VLANs, compared to the 4096 of 1Q
  - In principle the standard allows recursive nesting of tags, but more than 2 are never used (TTBOMK)
- The tags and fields have the same meaning of the 1Q
- Used and fundamental for Metro and Carrier grade Ethernets, not for simple LANs



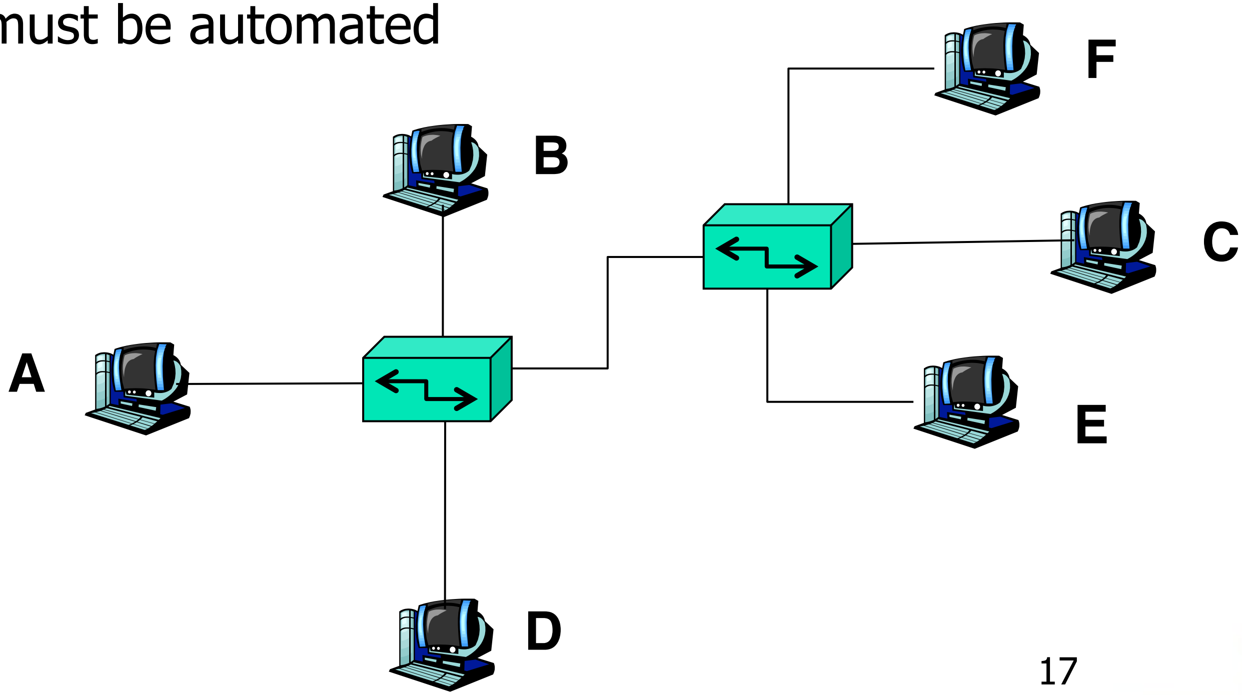
1AD is fundamental for trunking



- 1AD also known as VLAN Tagging
  - Allows operators to carry multiple VLANs across geographic links

Addresses Backward Learning

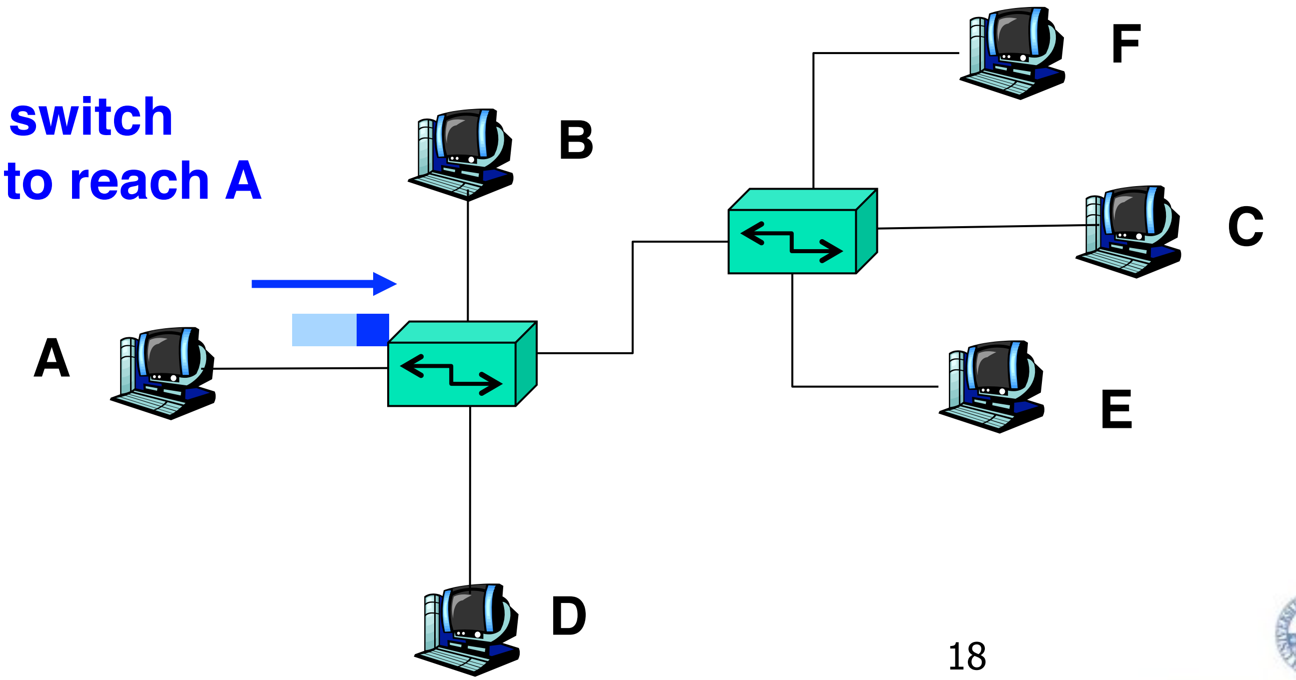
- Switches forward frames based on dest. addresses
  - Only on links that need them
- Switch table
  - Maps destination MAC address to outgoing interface
  - No algorithm to build the switch
  - Building must be automated



Backward Learning: Building the Table

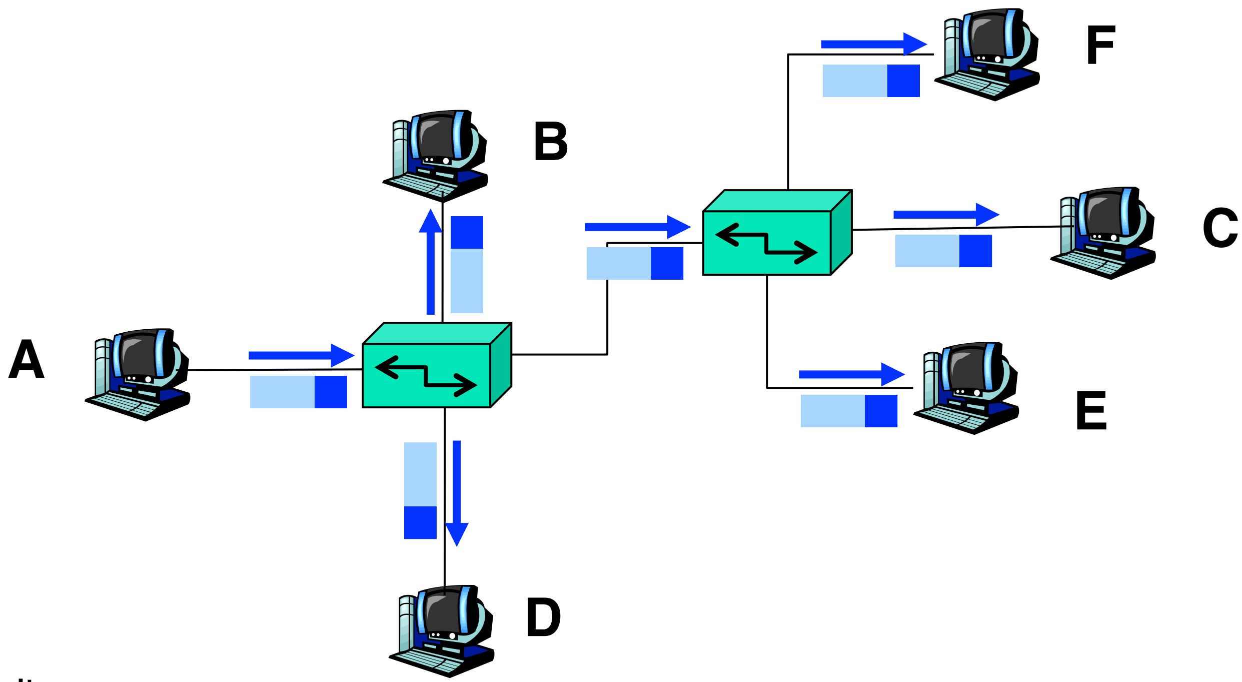
- When a frame arrives
  - Inspect the *source* MAC address
  - Associate the address with the *incoming* interface
  - Store the mapping in the switch table
  - Use a time-to-live field to eventually forget the mapping

The first switch learns how to reach A



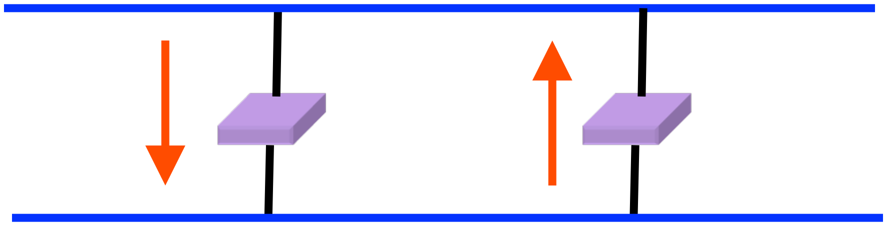
Backward Learning: Broadcast and Misses

- Miss: output port to destination is not in switch table
- Broadcast must go to everybody in any case
- When frame arrives with unfamiliar destination
  - Forward the frame out all of the interfaces except for the one where the frame arrived



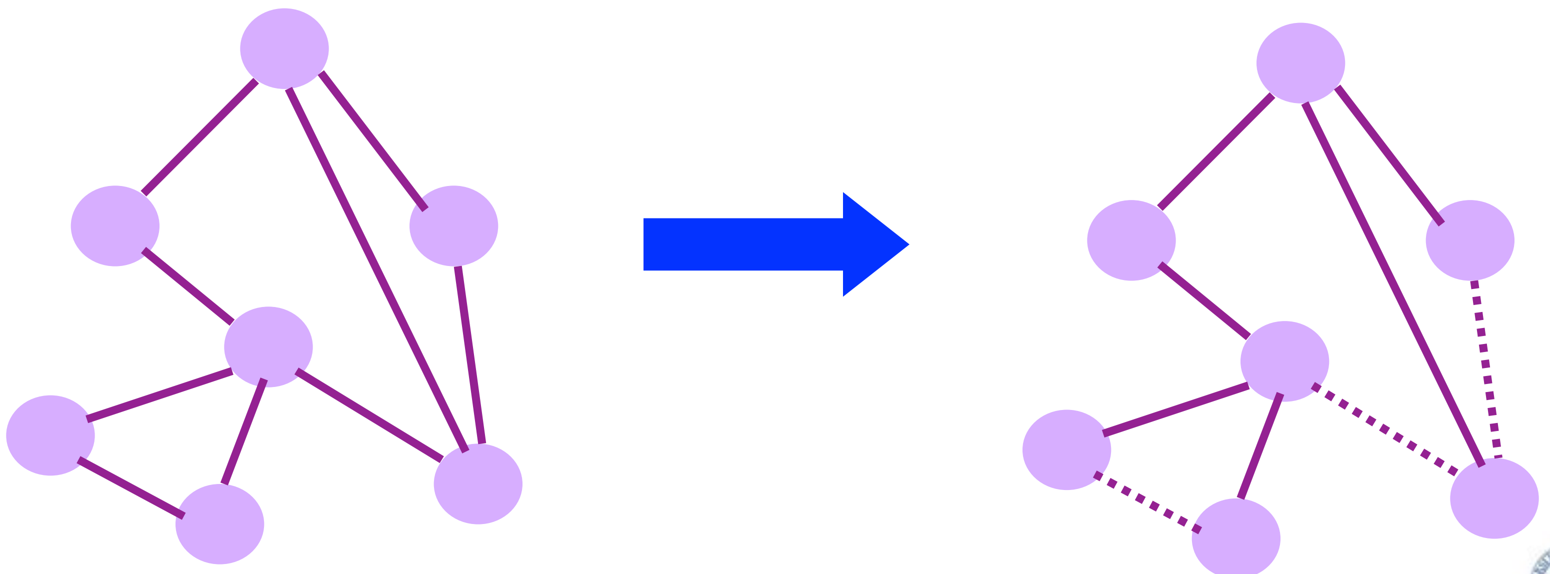
Broadcast Lead to Loops

- Switches need to broadcast frames
  - Upon receiving a frame with an *unfamiliar destination*
  - Upon receiving a frame sent to the *broadcast address*
- Broadcasting is implemented by flooding
- Flooding can lead to forwarding loops
  - E.g., if the network contains a cycle of switches
  - Either accidentally, or by design for higher reliability



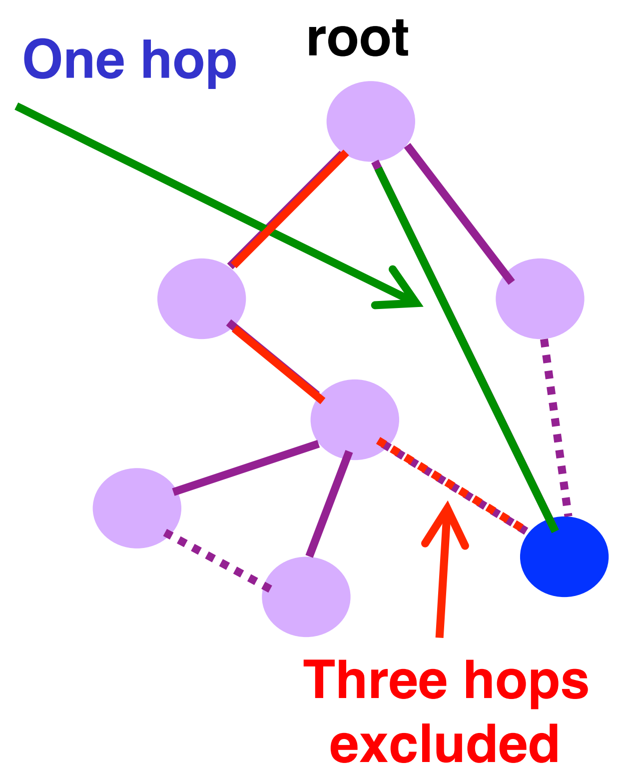
Solution: Spanning Trees

- Ensure the topology has no loops
  - Avoid using some of the links to avoid forming a loop
- Spanning tree
  - Sub-graph that covers all vertices but contains no cycles
  - MAC addresses are not structured, thus “routing” is not possible
  - The standard does not guarantee that the ST is minimum



Constructing the Spanning Tree

- Distributed algorithm
  - Switches cooperate to build the spanning tree
  - Reconfigure automatically when failures occur
- Key points of the algorithm
  - A “root” must be elected
    - The switch with the smallest (random) identifier
  - For each of its interfaces, a switch
    - identifies if the interface is on the shortest path from the root
    - excludes an interface from the tree if it is not on the SP to the root

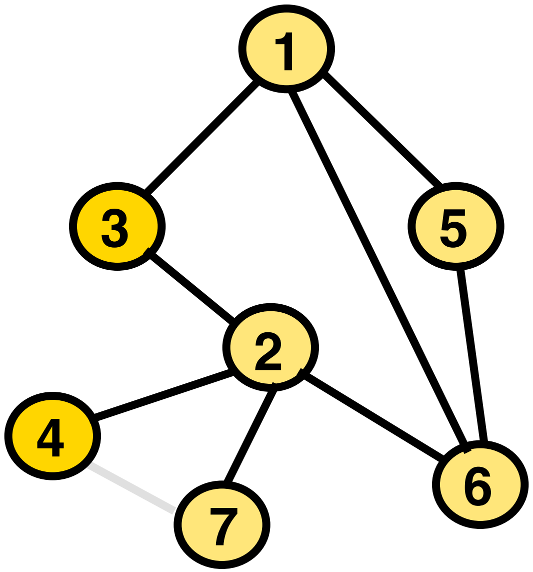


Steps in Spanning Tree Algorithm

- Use broadcast messages: (Y, d, X)
  - sent by node X, thinking Y is the root, the distance Y-X to root is d
- Initially, each switch sends a message out every interface identifying itself as the root
  - Switch A announces (A, 0, A)
- Switches update their view of the root
  - Upon receiving a message, check the root id
  - If the new id is smaller, start viewing that switch as root
- Switches compute their distance from the root
  - Add 1 to the distance received from a neighbor
  - Identify interfaces not on a shortest path to the root and exclude them from the spanning tree

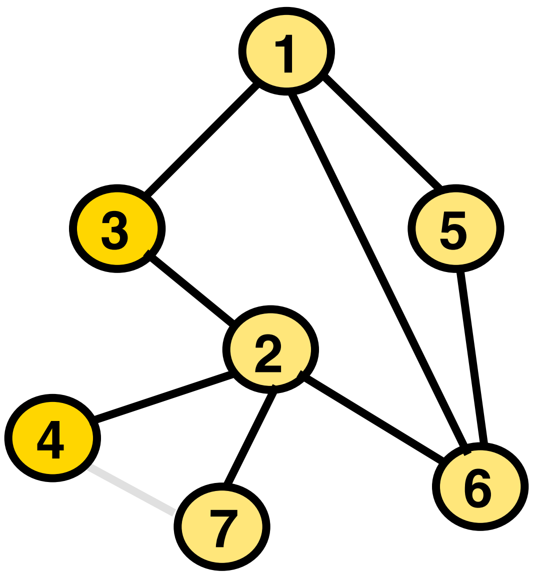
Example From Switch #4' s Viewpoint

- Switch #4 thinks it is the root
  - Sends (4, 0, 4) message to 2 and 7
- Switch #4 hears from #2
  - receives (2, 0, 2) message from 2
  - thinks that #2 is the root
  - realizes it is just one hop away
- Switch #4 hears from #7
  - receives (2, 1, 7) from 7
  - realizes this is a longer path
  - prefers its own one-hop path
  - removes 4-7 link from the tree



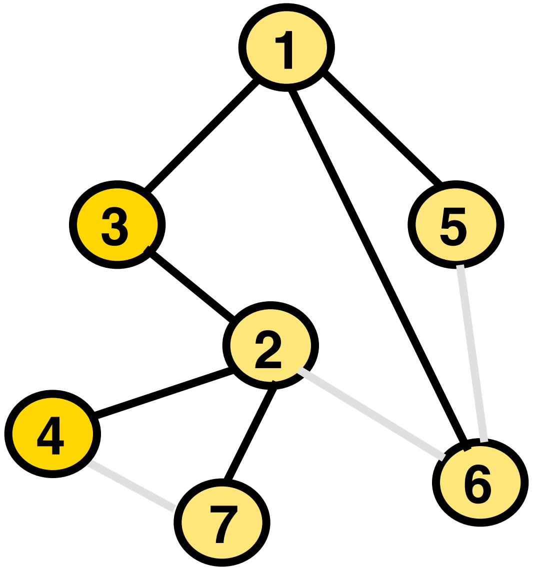
Example From Switch #4's Viewpoint

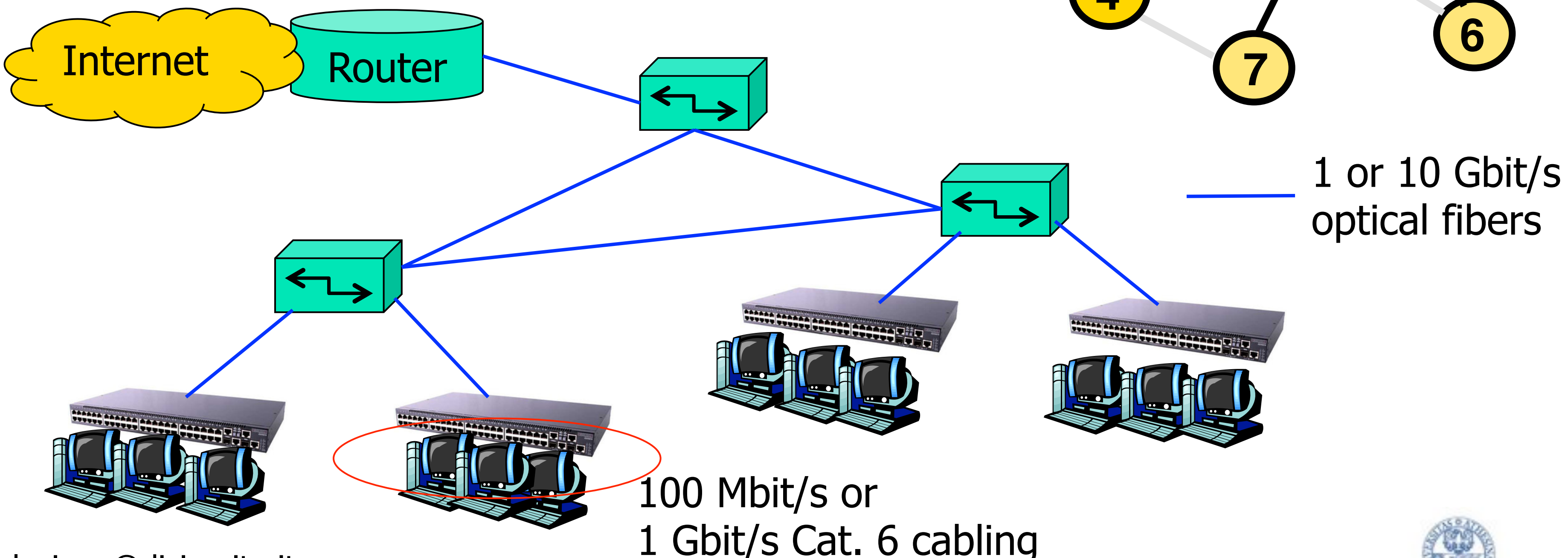
- Switch #2 hears about switch #1
  - Switch 2 hears (1, 1, 3) from 3
  - Switch 2 starts treating 1 as root
  - And sends (1, 2, 2) to neighbors
- Switch #4 hears from switch #2
  - Switch 4 starts treating 1 as root
  - And sends (1, 3, 4) to neighbors
- Switch #4 hears from switch #7
  - Switch 4 receives (1, 3, 7) from 7
  - And realizes this is a longer path
  - So, prefers its own three-hop path
  - And removes 4-7 link from the tree



Other switches

- Behave the same and the SP rooted in 1 remains the only active
- 1 becomes a bottleneck
  - Good network design is needed
  - Hierarchical switches with fast backbones





Robust Spanning Tree Algorithm

- Algorithm must react to failures
  - Failure of the root node
    - Need to elect a new root, with the next lowest identifier
  - Failure of other switches and links
    - Need to recompute the spanning tree
- Root switch continues sending messages
  - Periodically reannouncing itself as the root (1, 0, 1)
  - Other switches continue forwarding messages
- Detecting failures through timeout
  - Switch waits to hear from others
  - Eventually times out and claims to be the root
- **Very slow to reconfigure and converge**

802.1aq: Routing in LAN/VLAN

- Standard SP performs very poorly
- In 2012 a new amendment to the standard provides for real routing with link state, shortest path routing among switches
- Link costs are related to TX speed of the link:
  - $C = 2*10^{13} / LS$
  - LS = links speed from 100kbit/s to 10Tbit/s
- Works only for all switched LAN
  - No hubs
  - No 10bT (Coax cable)

802.1aq: Routing in LAN/VLAN

- Stations can belong to multiple VLAN
  - Just as a host can have multiple IP addresses
- If the topology permits it a switch may route based on the VLAN only, either internal or external (Q-in-Q)
- Routing based on VLANs partially solves the problem of address backward learning:
  - If the frame can be forwarded based on the VLAN, the address need not be known
  - Smaller forwarding tables
  - Less broadcast and flooding