CS-340 Introduction to Computer Networking

Lecture 14: Ethernet

Steve Tarzia

Many diagrams adapted from that by J.F Kurose and K.W. Ross

Last Lecture: Medium Access Control

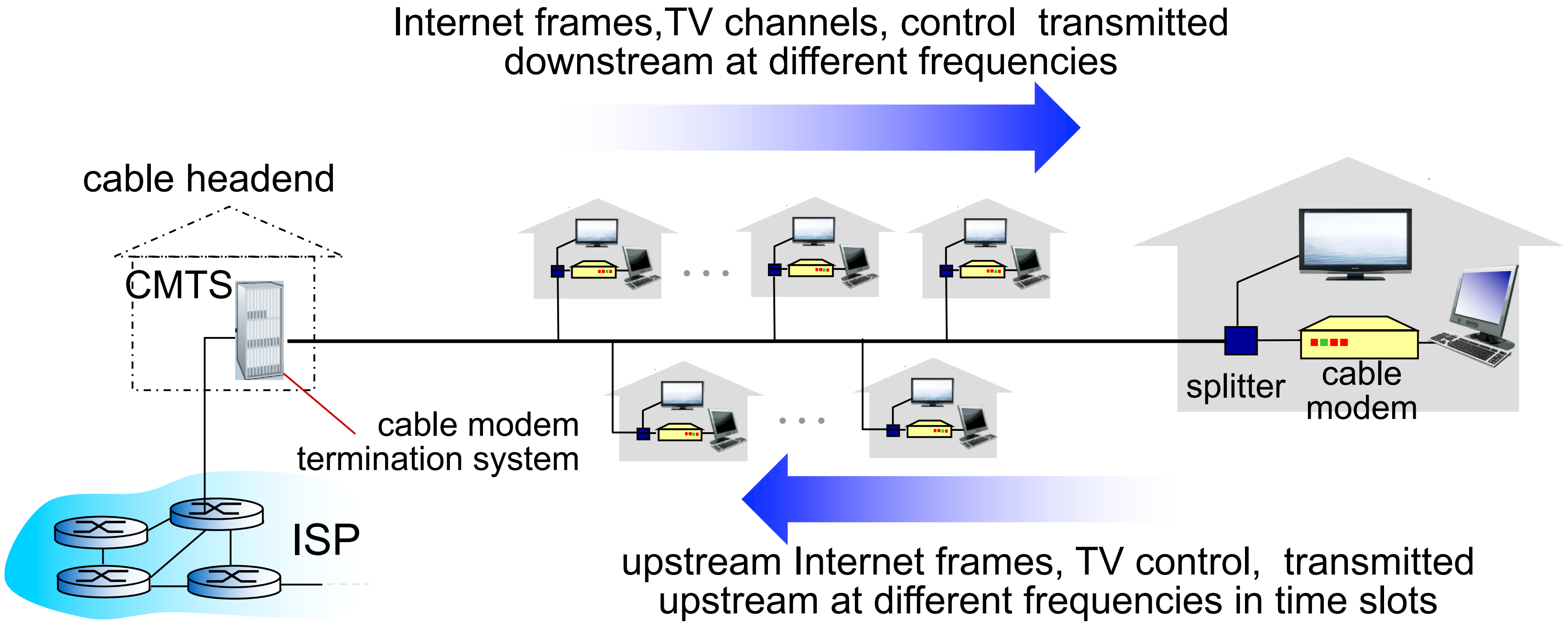
- Link-layer handles sharing a physical link/medium with multiple nodes.
- Medium Access Control / Multiple Access Protocol
  - Decide how to share the link.
  - Two nodes sending simultaneously is a **collision.** Packets are lost.

Three classes of sharing protocols:

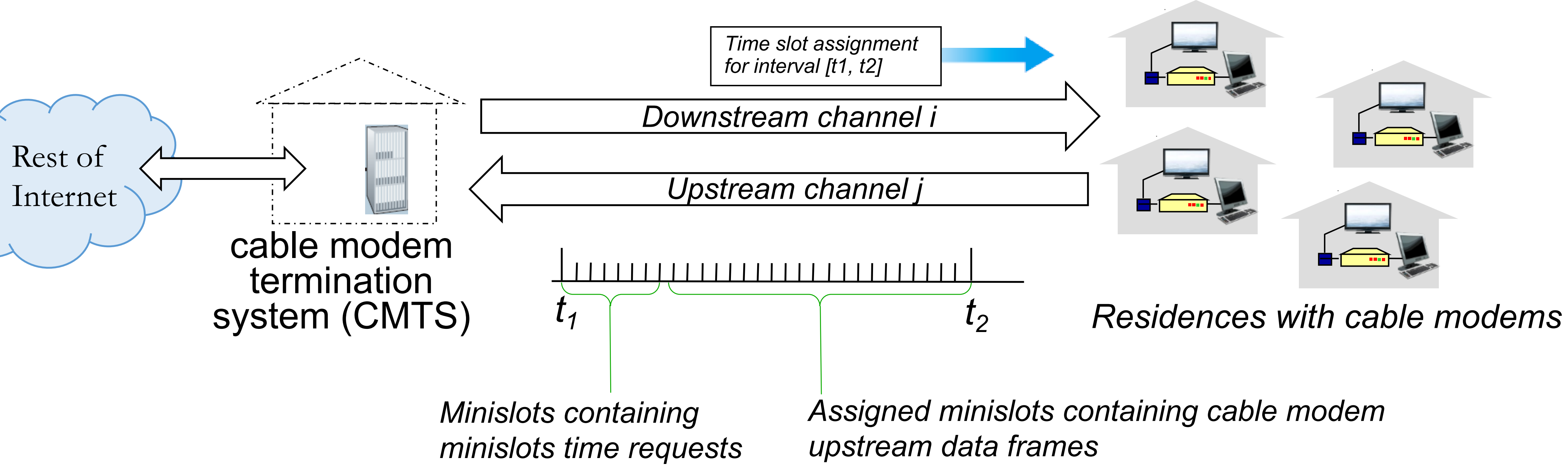
- **Channel Partitioning**:
  - Frequency Division Multiplexing - *WiFi*
  - Time Division Multiplexing
- **Turn-Taking**:
  - Polling - *Bluetooth*
  - Token-passing
- **Random Access**:
  - ALOHA (simple historical example)
  - CSMA/CD (Carrier Sense Multiple Access/Collision Detection) - *Ethernet, Wifi*


EG28 Communications Closet
CORNING

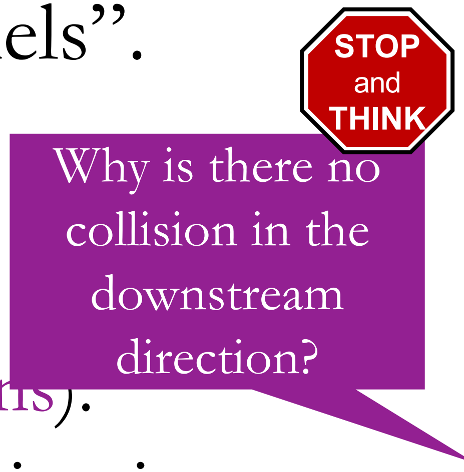
DOCSIS: Cable Internet Link-Layer Protocol



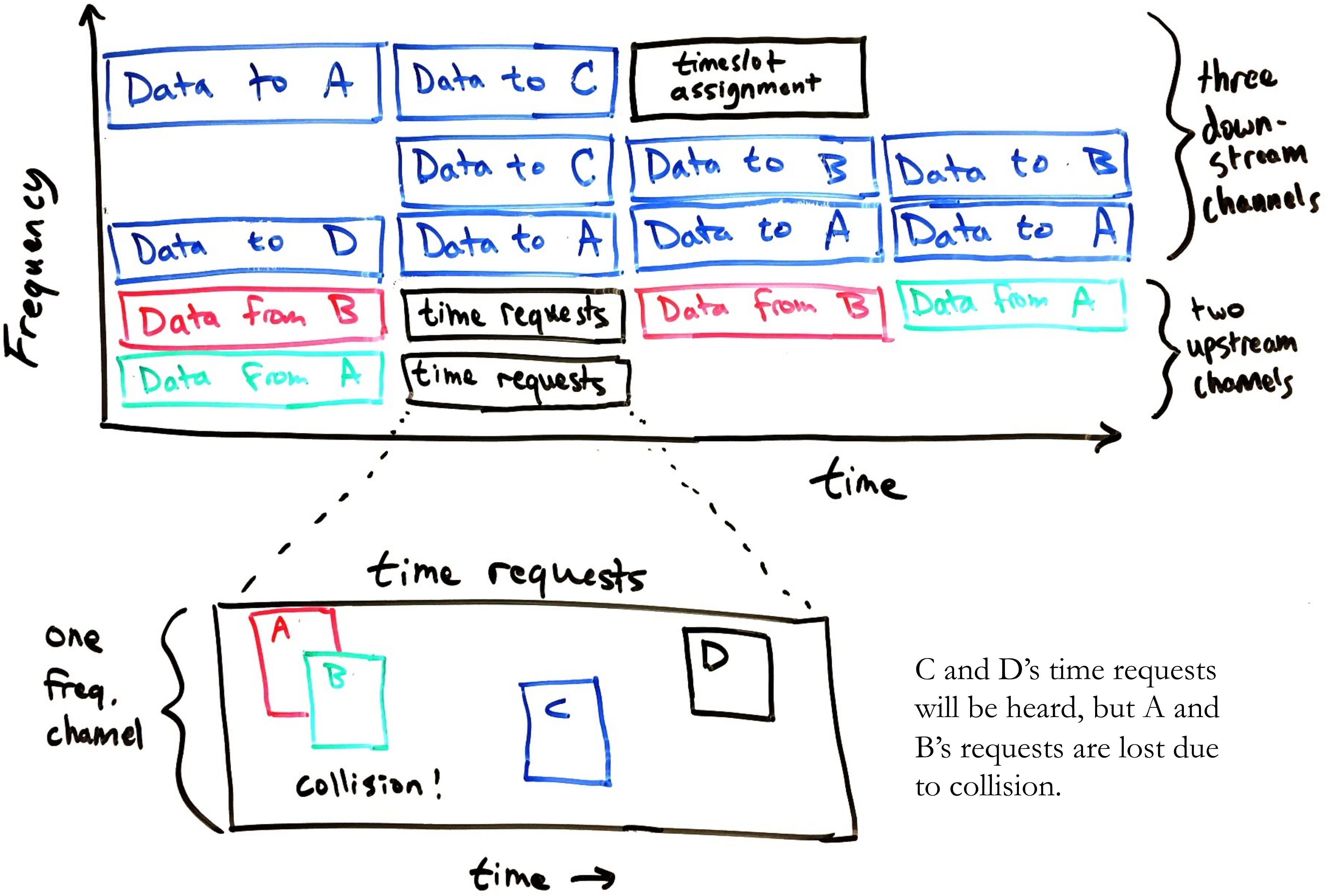
- Combines ideas from all three classes of multiple access protocols!
- All cable modems in a neighborhood share the same coaxial medium.
  - They are all connected to "one big wire." It's a shared, broadcast medium.
  - They can all "hear" each other's traffic, in both directions.



- FDM (channel partitioning) creates multiple up & down "channels".
- Upstream channels also use TDM:
  - Some time slots are for modems to send *time requests.* Time requests use a random-access protocol and may collide.
  - Remaining time slots are assigned to specific modems (taking turns).
- CMTS periodically broadcasts the time slot assignments, taking into account the time requests that were received.



A, B, C, and D are different DOCSIS cable modems connected to a single wire.



C and D's time requests will be heard, but A and B's requests are lost due to collision.

- Downstream channels are always reserved by the headend CMTS.
- Upstream channels are allocated exclusively to the different modems,
- Except the "time requests" slot is a free-for-all wherein anyone can request time using a random-access protocol

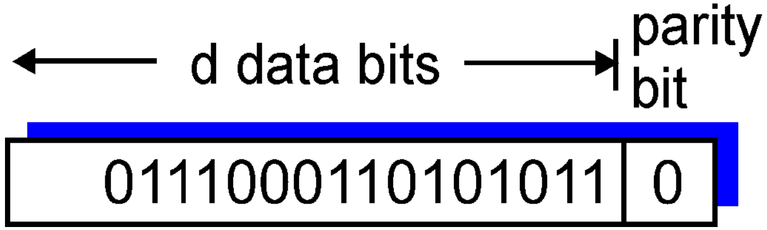
- FDM and TDM create many virtual channels.
- Taking-turns assigns channels to different modems
- Random access is used in one special time-request channel.

Error Detection and Correction

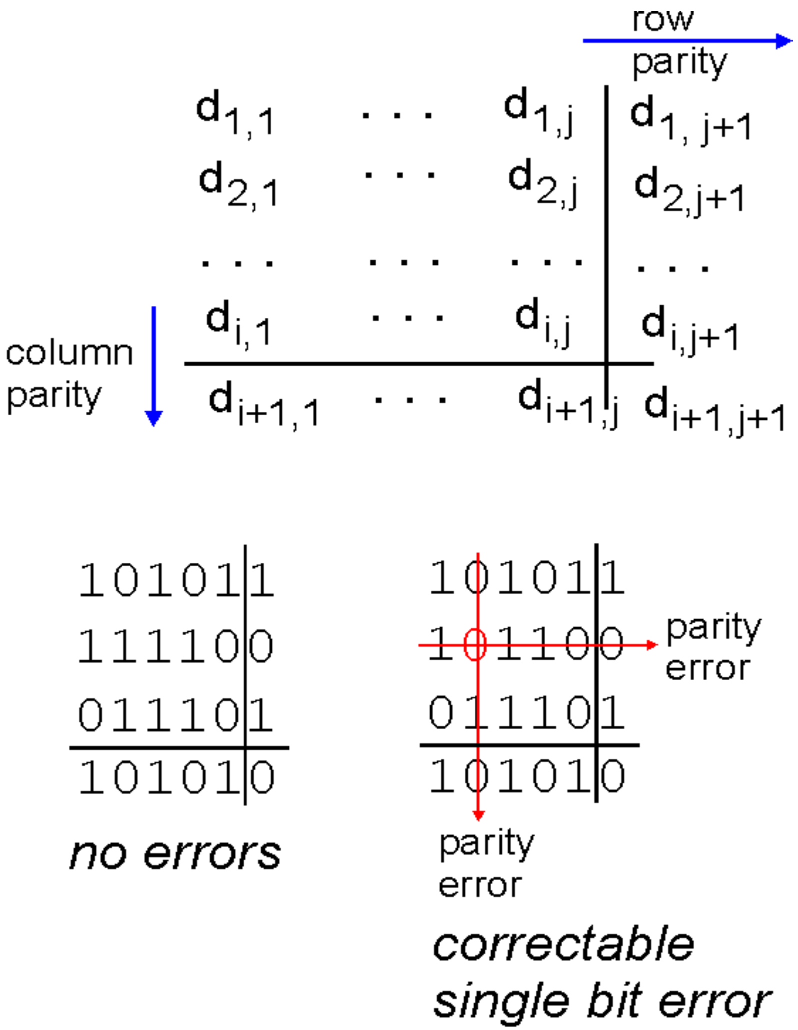
- The secondary purpose of the Link Layer (aside from MAC).
- Wireless media are especially prone to bit-flip errors, due to noise.
- **Error detection** – notice a bit error and discard the packet.
- **Error correction** – fix a bit error before delivering the packet.
- In both cases, additional bits of **redundant data** are added.

Parity

- Add a one or zero to make the total number of ones even.
- **Single-bit parity** detects a single bit error:



- **Two-dimensional bit parity** can correct errors:



Checksum (used in IPv4, UDP, and TCP headers)

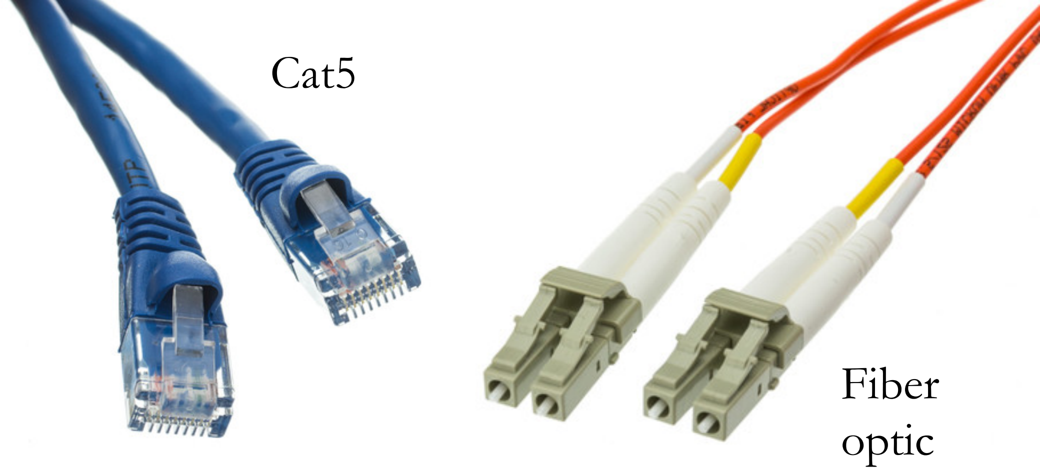
- Break the data into a sequence of 16-bit integers
- Add the integers
- *Wrap* the carry-out bits to the least-significant position.
- Finally, invert the result.

		1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
		1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
wraparound	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1
sum		1	0	1	1	1	0	1	1	1	0	1	1	1	1	0	0
checksum		0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1

Cyclic Redundancy Check (CRC)

- Error detection scheme that can be efficiently implemented using XOR circuitry.
- Very common in practice.
- Ethernet header uses CRC-32, using a specific 32-bit generator integer.

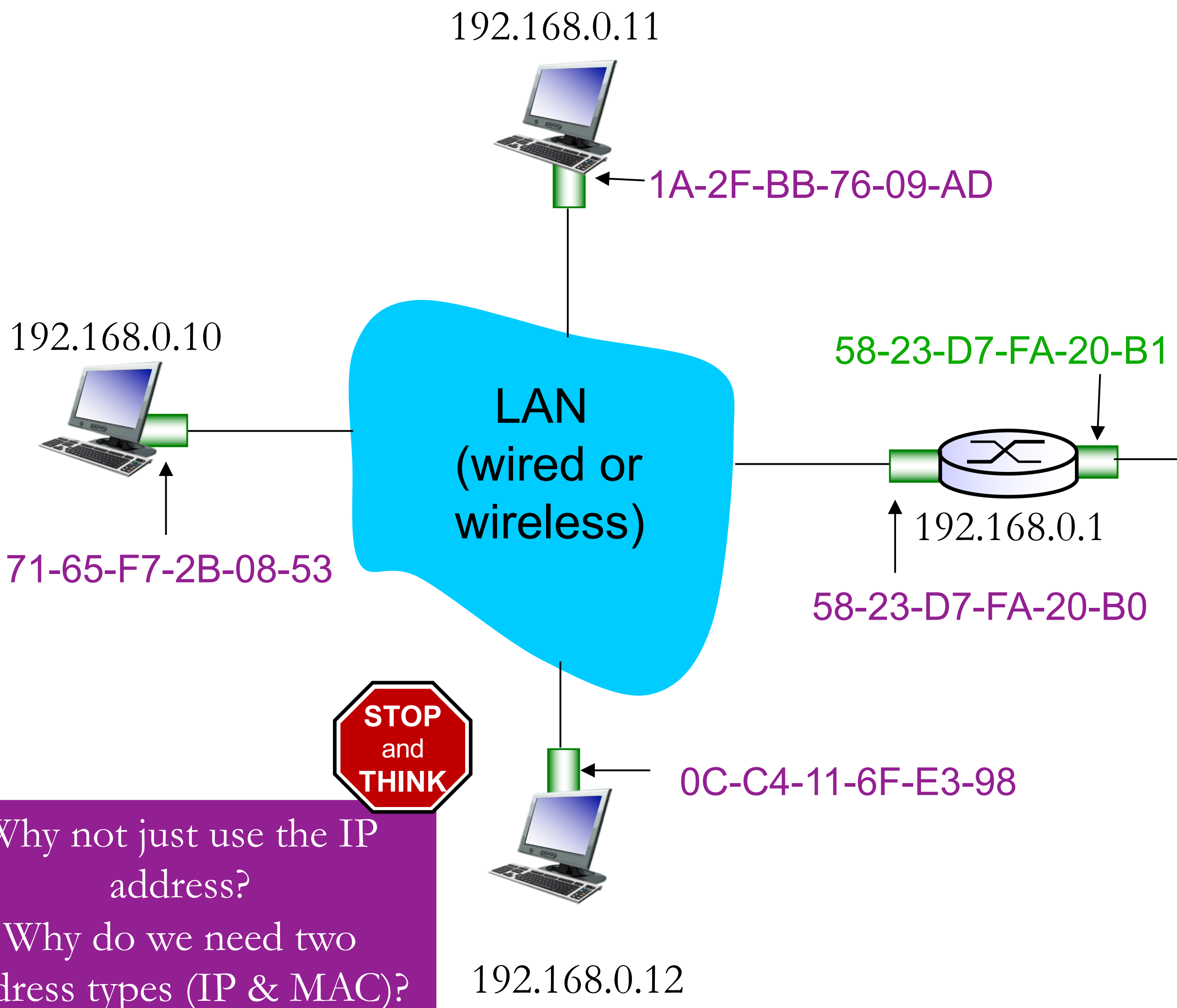
Ethernet



- A link-layer protocol for wired local area networks (LANs).
- Uses CSMA/CD: ***C**arrier-**S**ense **M**ultiple **A**ccess/ **C**ollision **D**etection*
  - Bandwidth utilization is good
  - Totally distributed, "plug and play"
- Uses Cat5 twisted pair or fiber-optic cabling.
- Same basic protocols also apply to WiFi networks (WiFi also uses FDM).
- Each **adapter** on the LAN has a unique **MAC address**: (***M**edia **A**ccess **C**ontrol address*)
  - 6-byte number expressed in hex, like 2B-39-0F-14-EE-A3.
  - MAC address is *permanent* and assigned at the factory.
  - Network device manufacturers buy blocks of MAC addresses from the IEEE.
    - Manufacturer is identified by MAC address prefix.
  - MAC address usually corresponds to a specific "plug" on a node.
- MAC address is used for communication **within a subnet**.

MAC addresses

- Previously, we assumed that links were point-to-point.
  - It was enough to set the packet's destination IP address and choose the correct link.
- If multiple nodes are listening on a link, we need the MAC address to identify the true destination.
- When sending a packet through the router, the destination IP address will be outside the network, eg., 2.0.0.1.
  - If the MAC destination is set to 58-23-D7-FA-20-B0, then the two other hosts will ignore the packet and the router will accept it and forward it.



Why not just use the IP address?
Why do we need two address types (IP & MAC)?

Ethernet frame

- Adds bytes before *and after* an IP datagram.

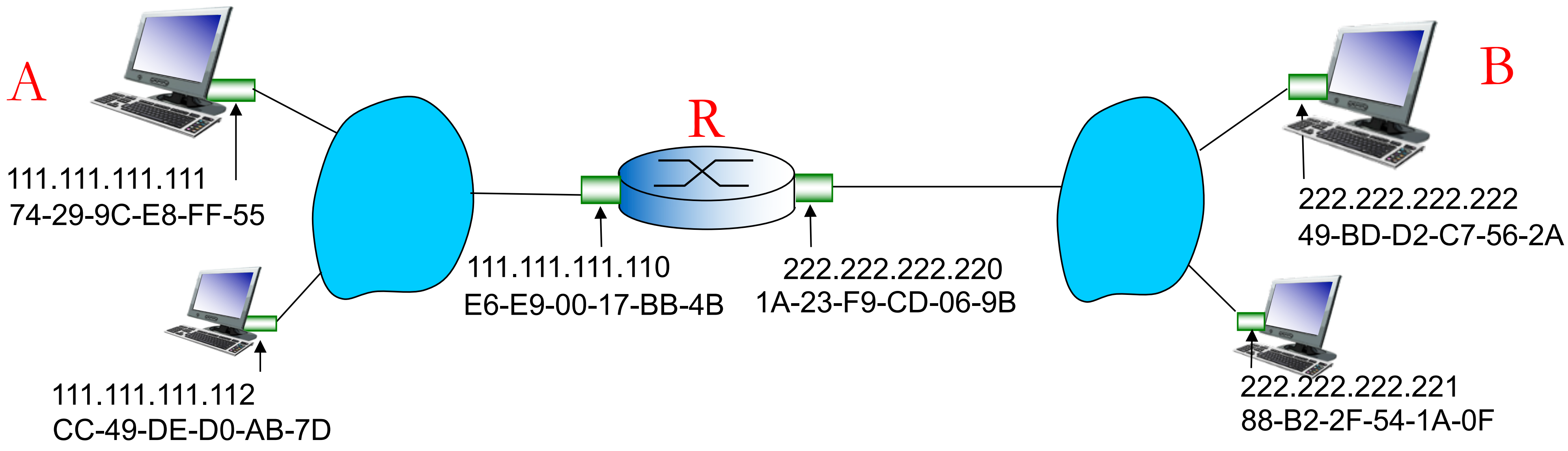
- Preamble is 8 bytes, 1010…1011
  - Used to synchronize bit-clock of the receiver to the sender.
- Type (2 bytes) usually indicates IP payload (ARP has a different type).
- CRC (4 bytes) is added after the payload, for bit-error detection.

Address Resolution Protocol (ARP)

- Allows nodes to **discover MAC address associated with an IP address.**
- Request lists a target IP address, and response gives the MAC address.
  - ARP requests are addressed to special *broadcast* MAC address, FF-FF-FF-FF-FF-FF.
  - All hosts on the subnet accept messages sent to the broadcast address.
- The host assigned that IP address replies with its MAC address:
  "You were looking for 192.168.0.10… Here I am, and here's my MAC address!"
  - Response can be addressed directly to the MAC address of the requester.
  - Responses are cached in adapter's ARP table.
  - Response has a TTL, typically 20 minutes.
- Like DNS, but responses come from all hosts on the subnet.
- Ethernet hosts are responsible for advertising their own existence.
  - Thus, Ethernet is *plug-and-play*.

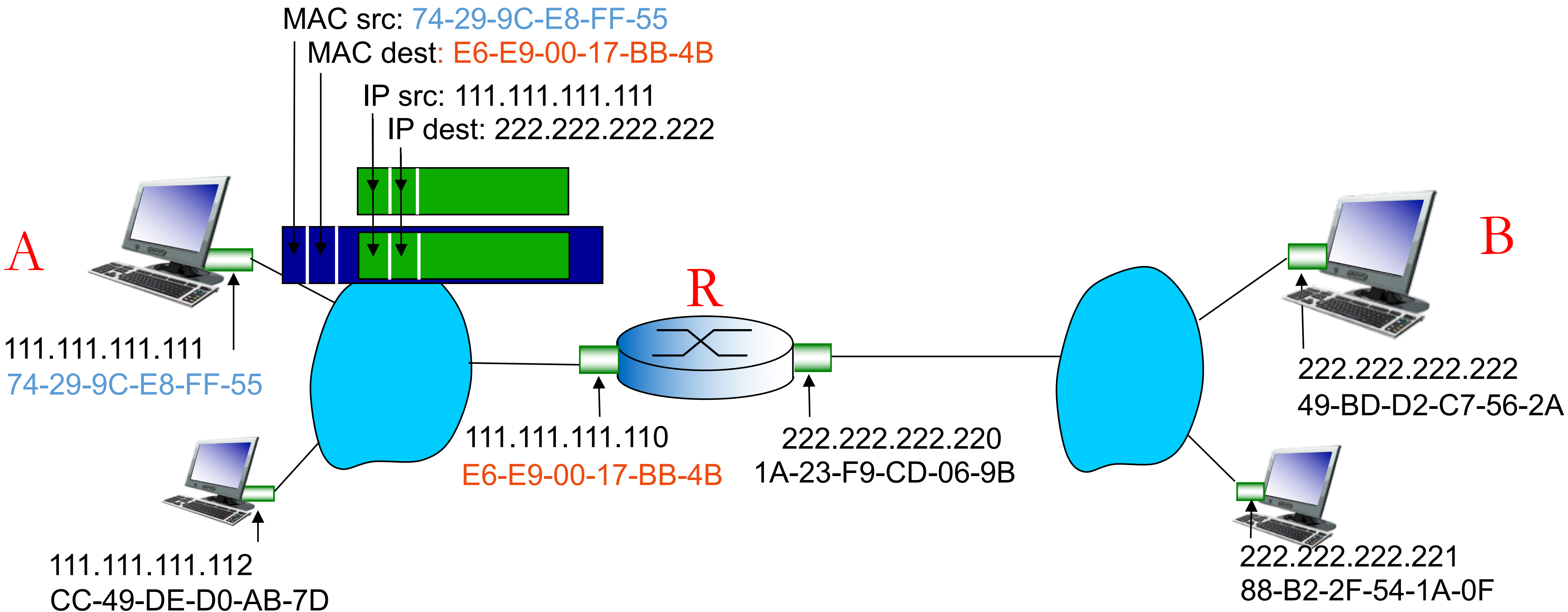
Routing between LANs/subnets

- Send datagram from A to B, via R
- Initially,
  - A knows B's IP address …from DNS response
  - A knows IP address of gateway router, R, …from IP configuration
  - A knows R's MAC address …from ARP response



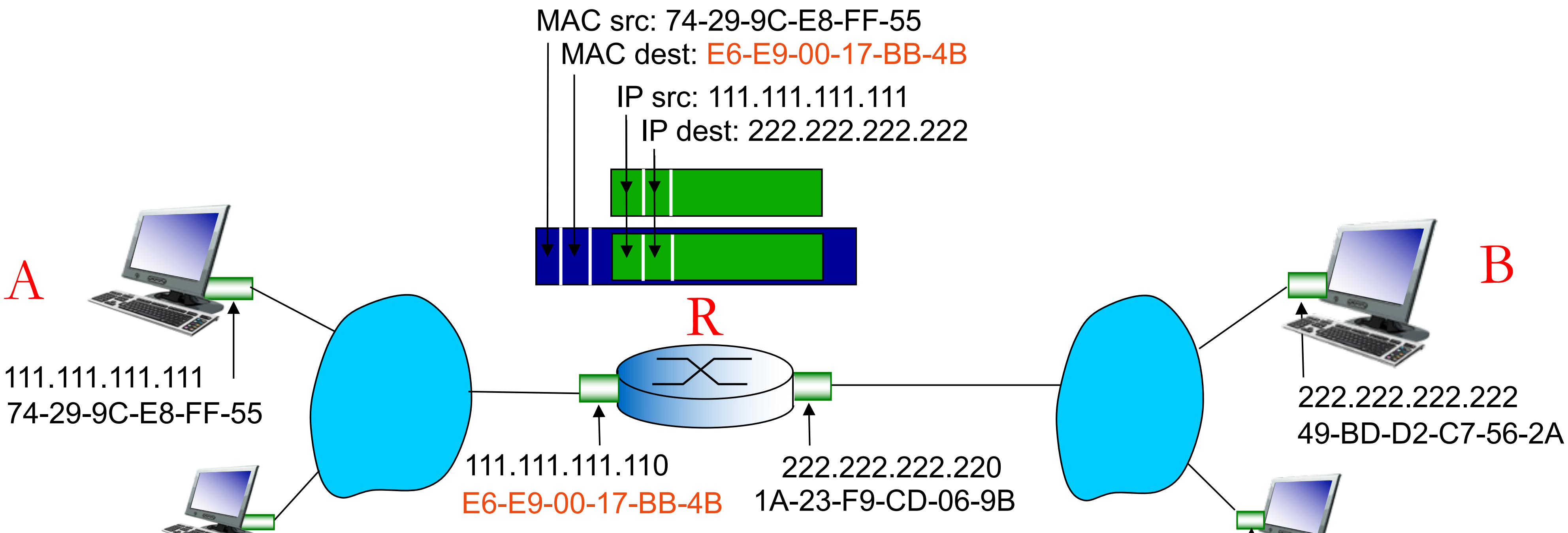
Routing: first hop, A to R

- A constructs IP datagram as expected, addressed end-to-end.
- Datagram is wrapped in an Ethernet Frame with MAC addresses relevant for the *first hop*.
  - *Source*: MAC of A, *Destination*: MAC of R.



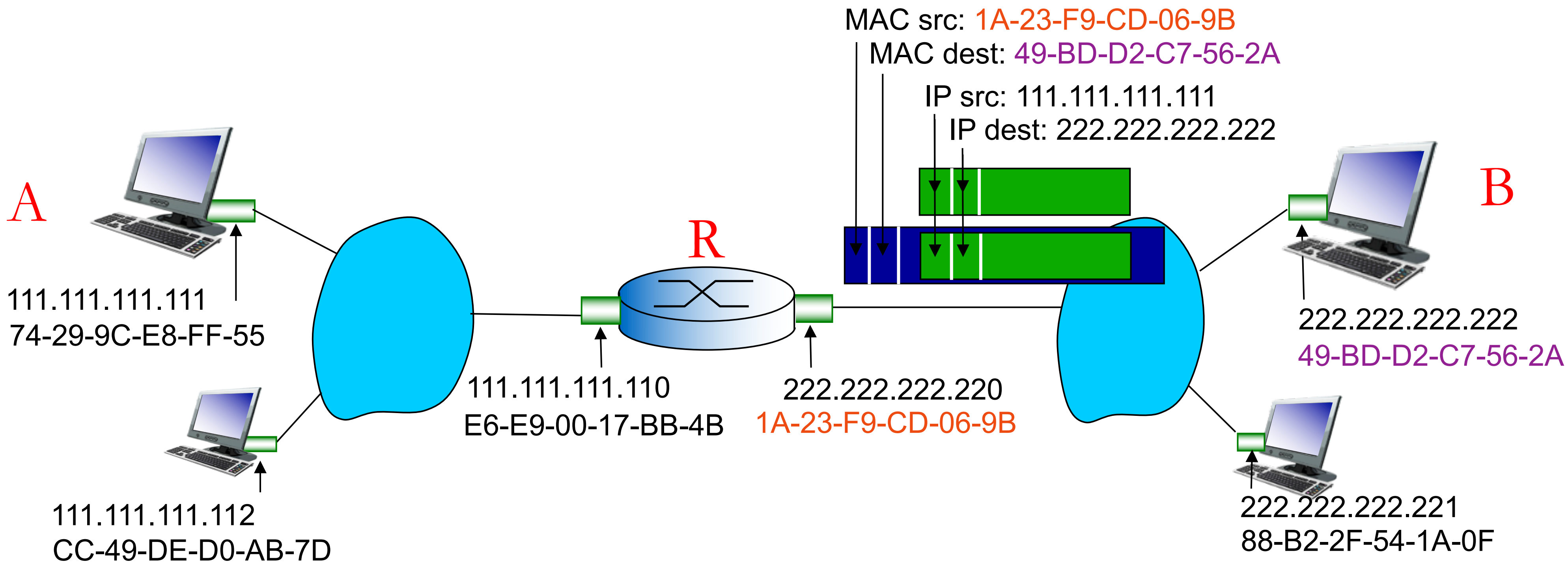
Routing: processing at R

- R accepts the packet because it sees its own MAC destination address.
- R looks at the IP payload, sees the destination IP address, checks its forwarding table, and decides to forward the packet to the right-side link.
  - The IP destination is within a subnet that R is on (at right).
  - The packet will next go to the MAC address of the destination, not a router.



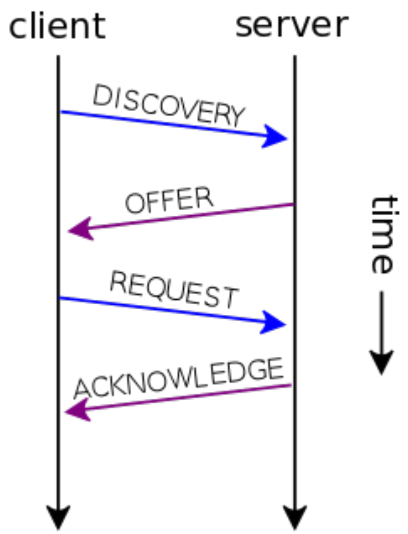
Routing: second hop, R to B

- R does an ARP query on the right-hand subnet looking for 222.222.222.222. B replies with its MAC address.
- R sends the packet with the MAC address of it's right-hand adapter as the source and B's MAC address as the destination.



Dynamic Host Configuration Protocol (DHCP)

- Recall that DHCP is how hosts automatically get IP configuration:
  - ***IP address*** • ***Subnet mask*** • ***Gateway address*** • ***DNS server address***
- When a host joins a network, it brings its permanent MAC address.
- Sends a DHCP request with broadcast destination, FF-FF-FF-FF-FF-FF.
- ***DHCP server*** (often the edge router) allocates IP addresses.
  - It listens for DHCP broadcasts & replies to the sender's MAC address (like ARP).
- DHCP involves four UDP packets:

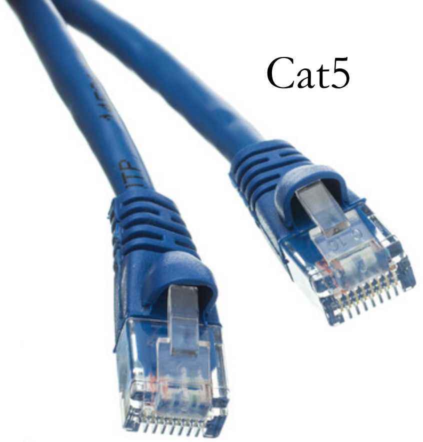


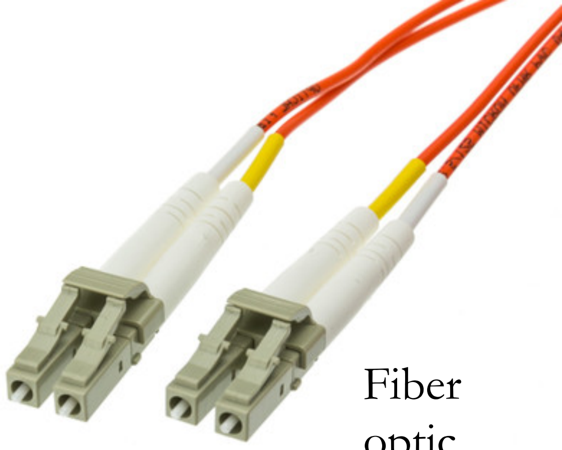
Advantage:
- Machine can move and join nearby network.

Disadvantage:
- IP address changes, so others cannot easily find you. Cannot run a server application.

Static IP configuration

- Hard-code machine's IP configuration, instead of using DHCP.
- Common for routers and servers, to make them permanently reachable at a certain IP address.
- ***Static DHCP*** combines both ideas:
  - A machine joining the network requests IP configuration with DHCP.
  - However, DHCP server is configured to always give certain machines specific IP addresses, based on MAC address observed in the request.
    - Eg., DHCP server is hard-coded to always return 222.222.222.222 in response to requests from 49-BD-D2-C7-56-2A
  - This gives the administrator one place to track IP address assignments.
  - Machines with unrecognized MAC addresses will get *dynamic* IP addresses from a pool of unused addresses.



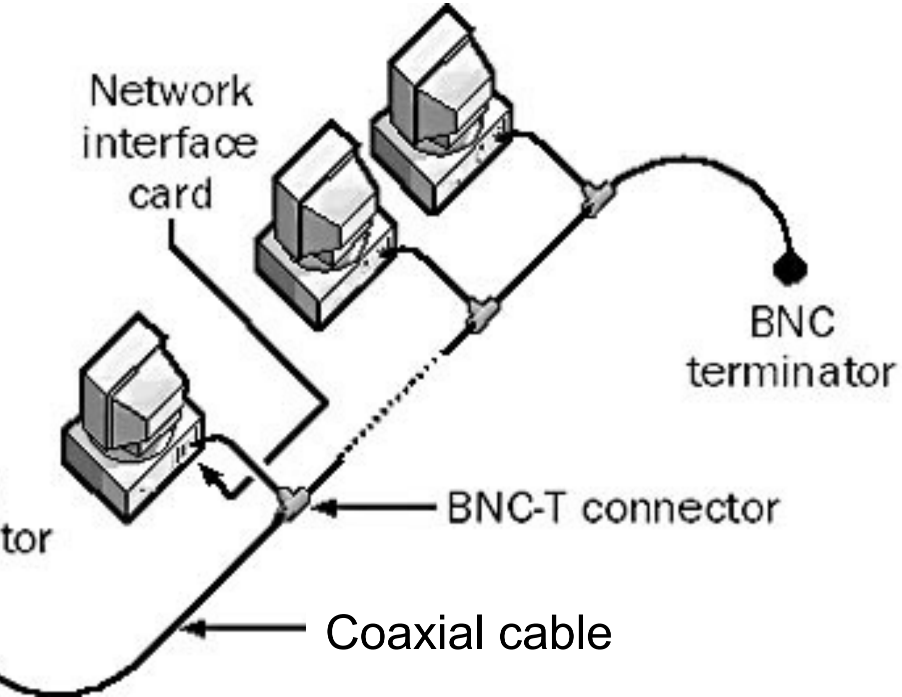


What makes wired Ethernet a shared medium?

Cat5 and Fiber Optic cables have dedicated channels in both directions!

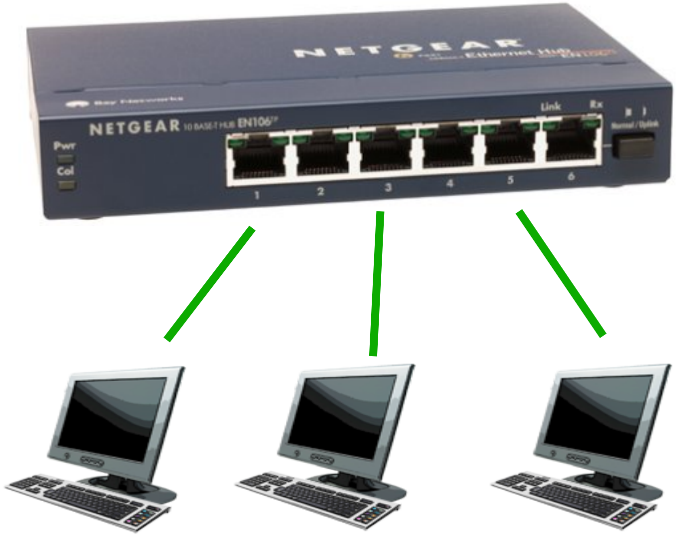
Ethernet began on shared electrical media

- Early Ethernet networks (1980s) connected many computers to a single coaxial cable wire.
- Used a **bus** topology.
- Single shared electrical channel



- Later networks (1990s) used cheaper twisted-pair cabling
- Used a **star** topology.
- Ethernet **hub** is at the center of the LAN and connects to each host.

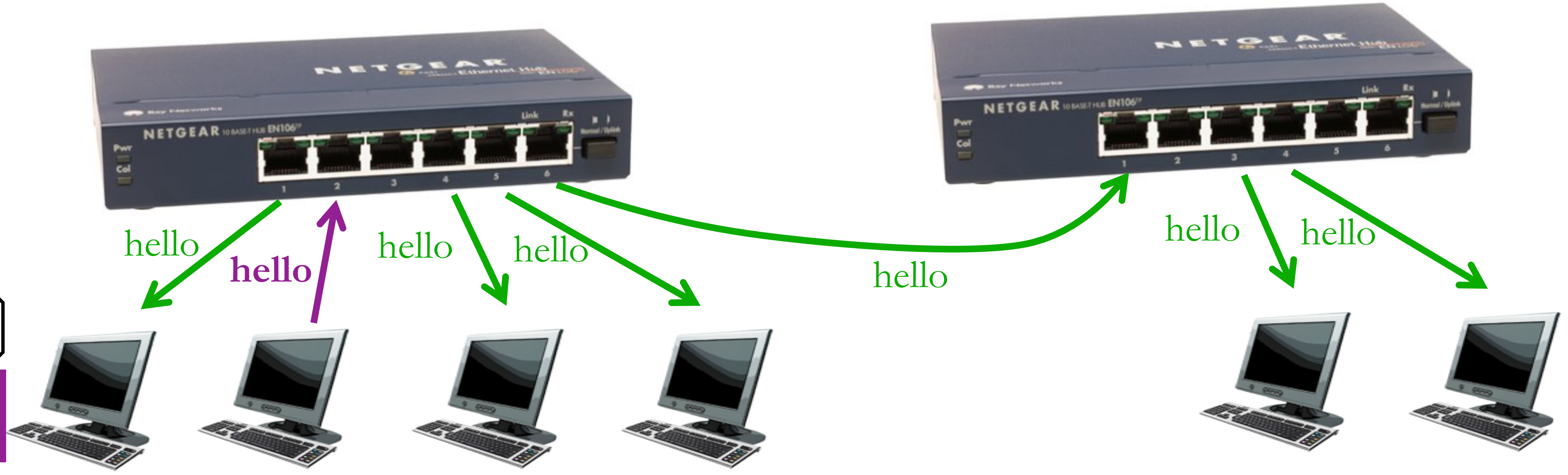
Ethernet Hub



Ethernet hubs *(early to mid 1990s)*

Note: "port" here means physical plug

- Data received by one port is broadcast to all the other <u>ports</u>.
- Machine should ignore packets addressed to different MAC address.
- Simple and easy to build (just read/send one bit at a time).
- Hub detects collisions and send special "jamming" signal to fail CRC check.



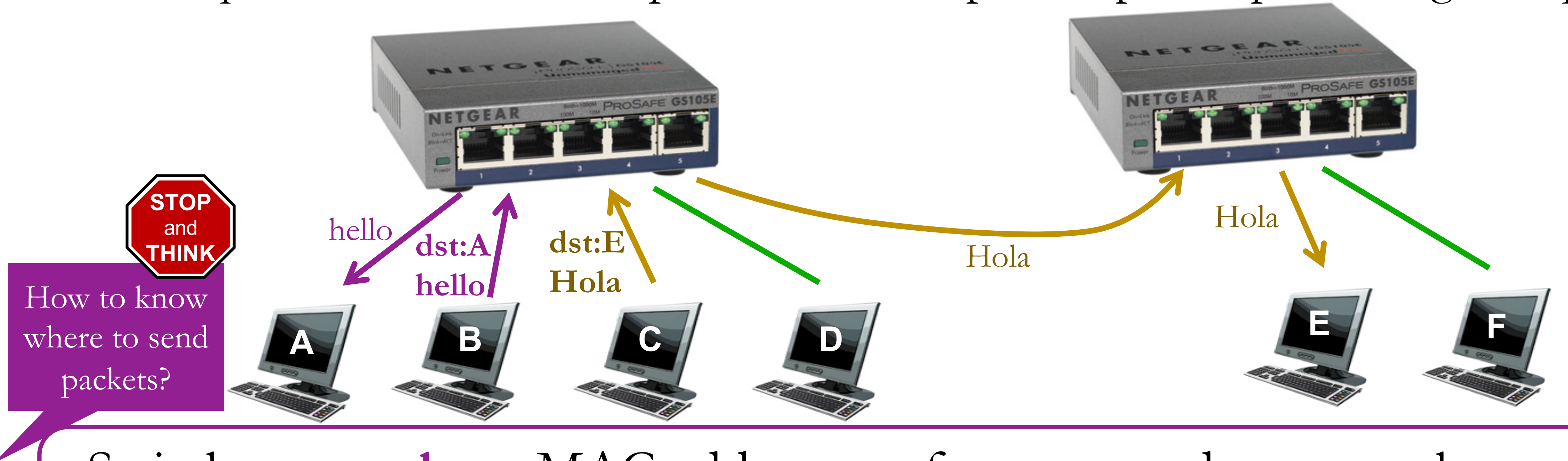
Shortcomings?

✗ Not scalable – a busy subnet will have many collisions.

✗ All machines must operate at same line speed (10 or 100Mbit/s)

Ethernet switches *(late 1990s–today)*

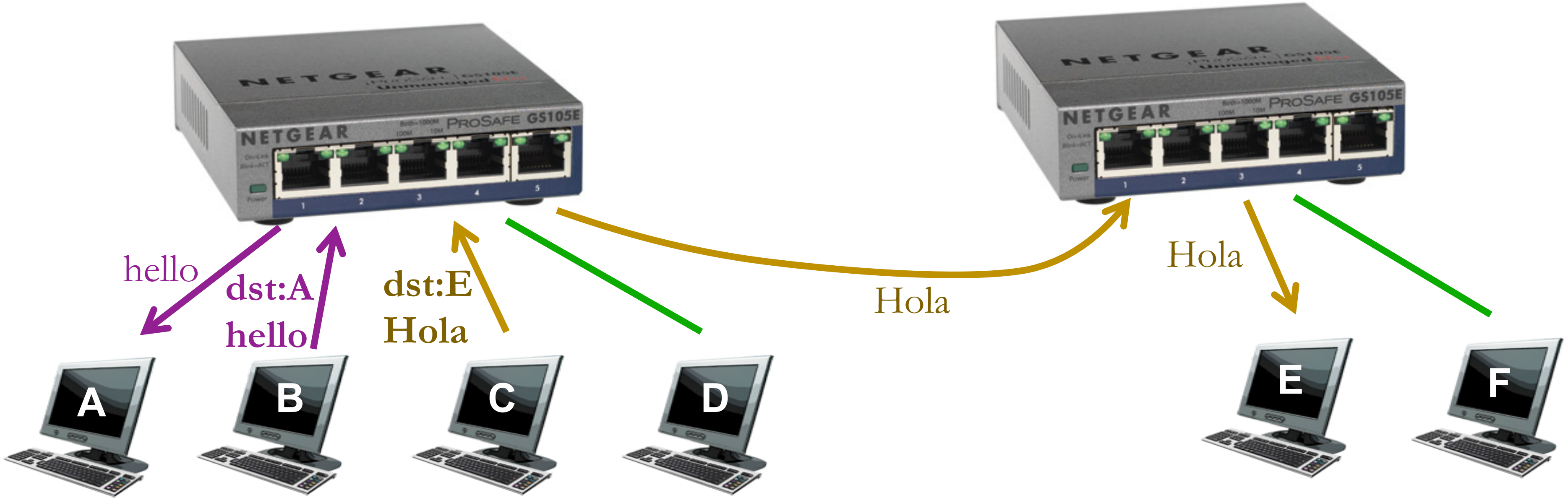
- *Avoid collision by avoiding broadcasts* – relay message to the **one** correct port.
  - Allow multiple pairs of ports to communicate simultaneously.
- Switches are **store and forward** devices, like routers.
  - Requires much more complex hardware – parallel packet processing and queueing.



- Switch **remembers** MAC addresses of recent senders on each port.
- If packet is addressed to an unknown address, broadcast to all ports.

Switches (continued)

- Switch allows a subnet to efficiently grow very large, because packet flows are isolated from each other and can happen in parallel.
- No special configuration is required on nodes or in switch.
  - It's entirely "plug and play," and compatible w/original ethernet design.
- If a port is connected to another switch, then it will relay traffic for many MAC addresses.



Switches avoid collision entirely

- Switch ports have output queues (like a router), so switch will wait to send a packet until the port is free.
  - Worst that can happen is that a packet is dropped due to a full queue.
- A message may arrive on a port at the same time we're sending because twisted pair and fiber cabling are **full duplex** media:
  - Have separate signal channels for sending and receiving.
  - Collision cannot be caused by incoming data.
- Coaxial cable and wireless are **half-duplex** media:
  - Share the same channel for sending and receiving.
  - Sending and receiving cannot occur simultaneously.
  - Collision can happen unless we carefully schedule transfers.
  - Can use FDM to support full-duplex operation.

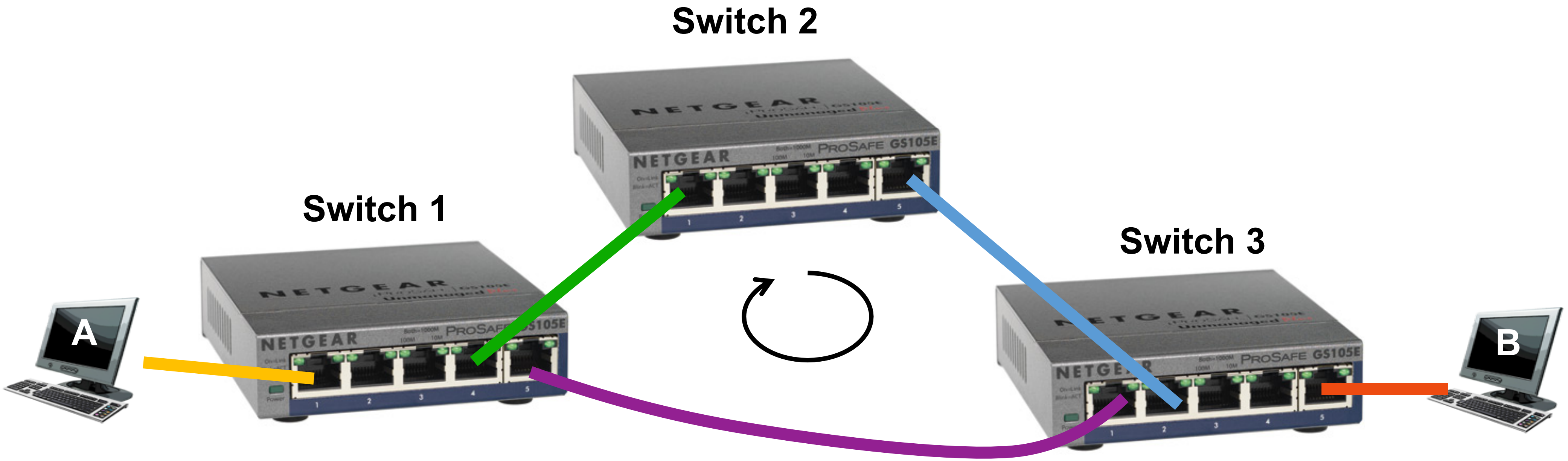
Cat5 twisted pair

LC fiber optic

Coaxial

Switching loops

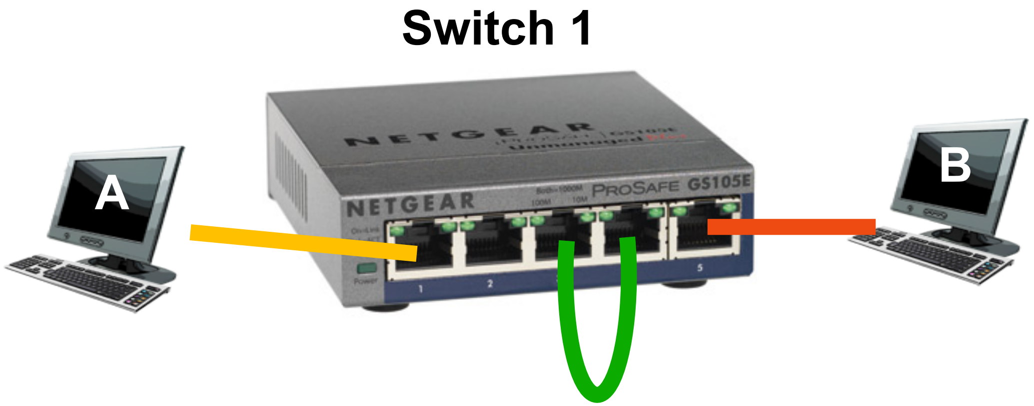
- What happens if you connect three Ethernet switches like this?



- Broadcast packets will travel in an infinite loop!

Switching loops *(cont.)*

- How about this?



- Same problem!

Ethernet Switch

- Chooses an outbound link using packet's destination *MAC address.*
- Forwarding rules are learned by inspecting traffic.

✗ Redundant links are not allowed. (Unless using *spanning tree protocol.*)

✓ No configuration is required, just "plug and play."

✗ ARP and DCHP (broadcast) traffic must be sent to all switch ports (on all connected switches).

  - Beyond ~1000 nodes, should break up the subnet with routers.

IP Router *(layer 3 switch)*

- Chooses an outbound link using packet's destination *IP address.*
- Forwarding rules are decided by IGP and BGP.

✓ Routing algorithm chooses *shortest* among multiple paths.

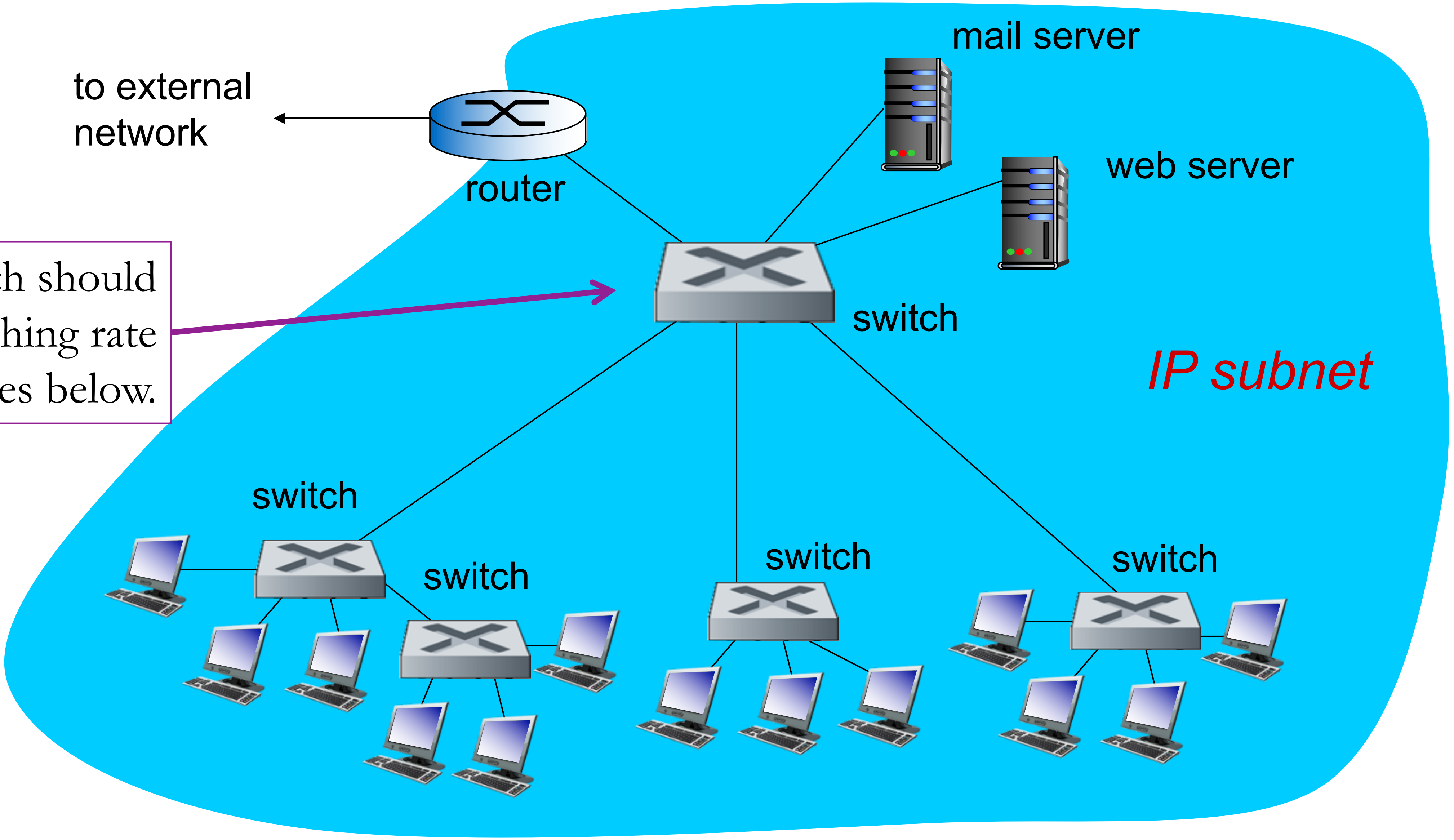
✗ Router must be configured to assign its IP addresses, IGP, etc.

✓ Isolates Ethernet broadcasts.

✓ Gives administrator greater control over where traffic is sent (*traffic engineering*).

A simple campus LAN

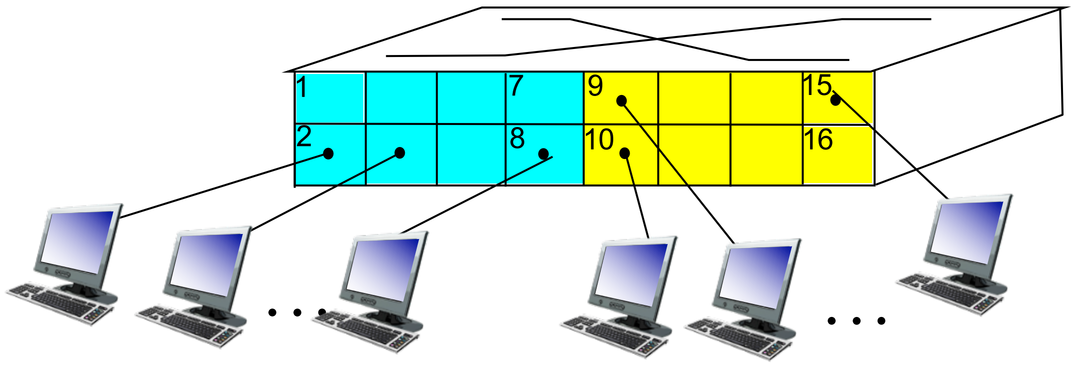
Central switch should have faster switching rate than edge switches below.



VLANs *(Virtual LANs)*

- VLANs divide a switch into multiple virtual switches
  - Allows large switches to be flexibly configured.
  - Often configured by port, but can also assign certain MAC addresses to certain VLANs.
- Often used to isolate private subnets for security.
- Traffic from one VLAN cannot flow into another VLAN (unless a router is connected).

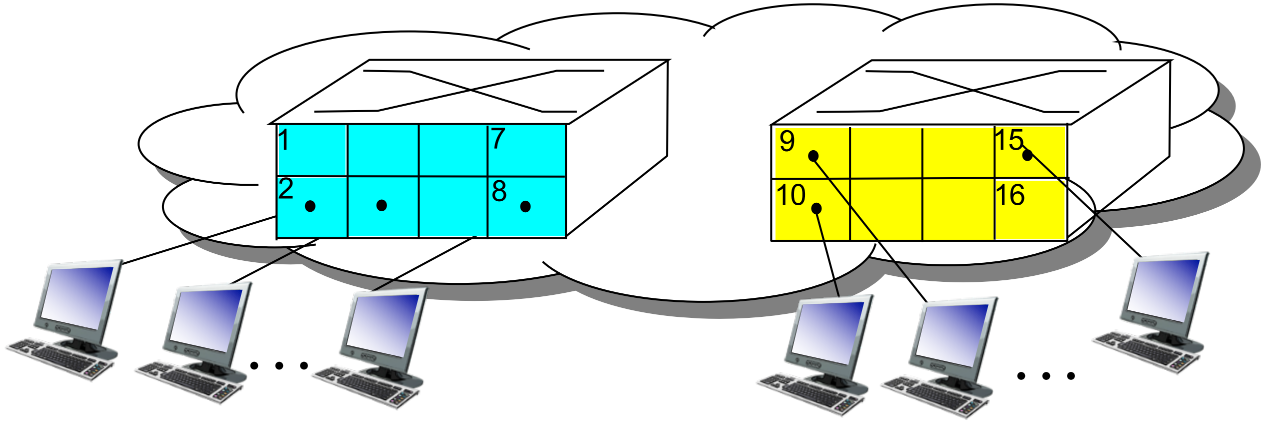
single physical switch ……



Electrical Engineering (VLAN ports 1-8)

Computer Science (VLAN ports 9-15)

… operates as *multiple* virtual switches

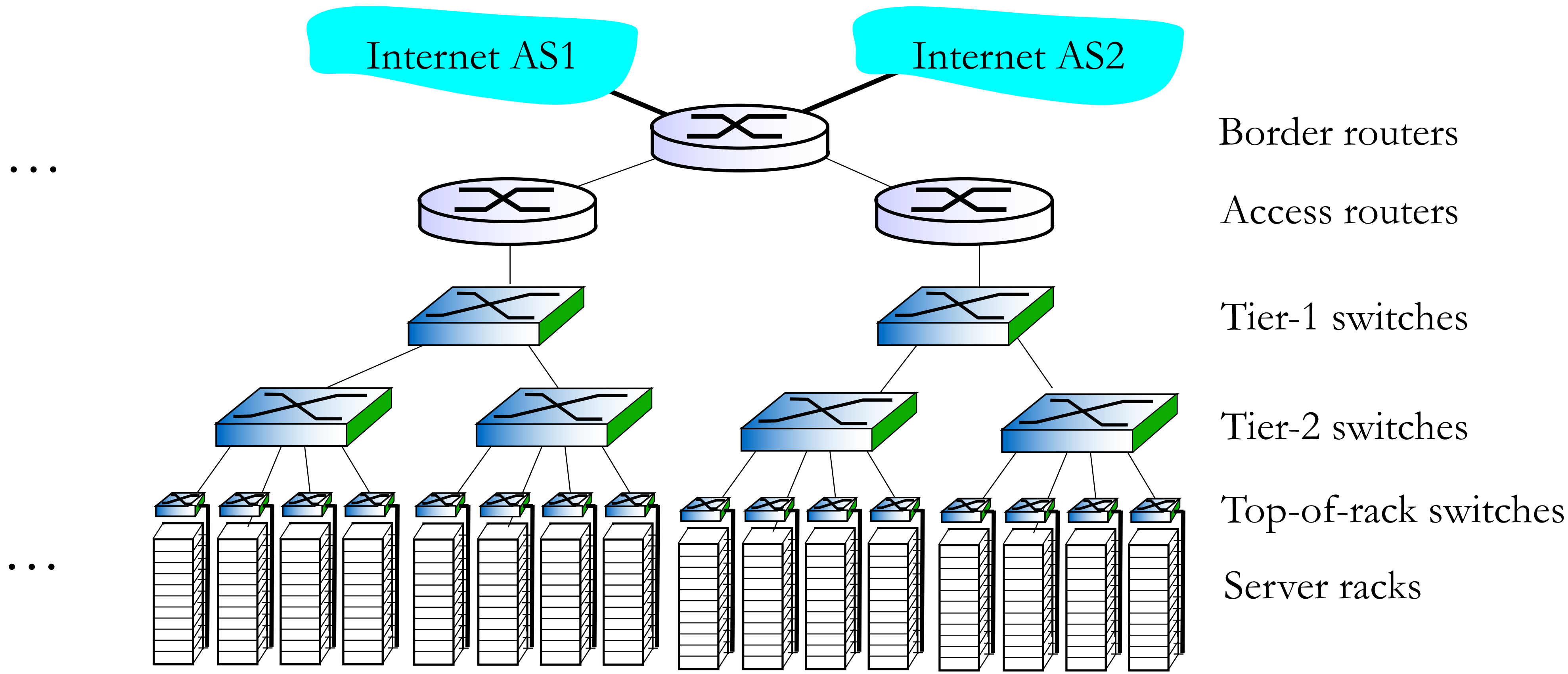


Electrical Engineering (VLAN ports 1-8)

Computer Science (VLAN ports 9-16)

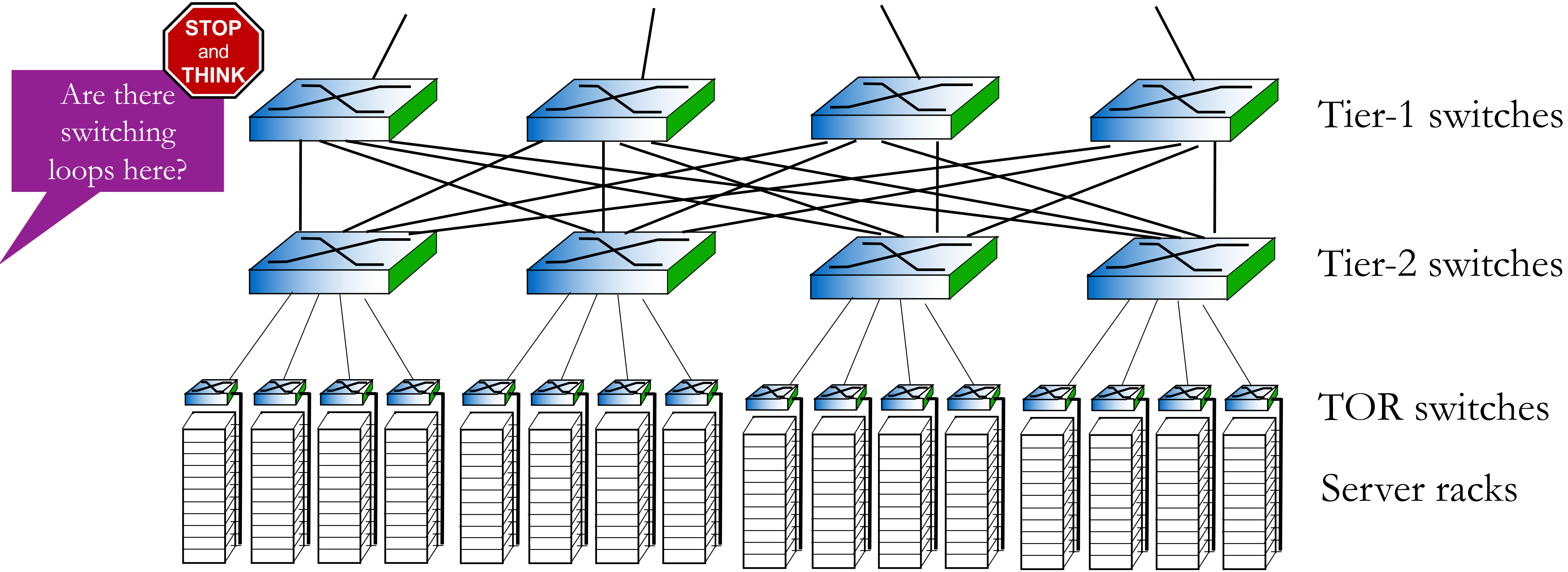
Data Center networks

- Data Centers house 10s or 100s of thousands of machines.
- Fast communication is essential
  - higher-level switches and routers can become bottlenecks.



Reducing bottlenecks in Data Center networks

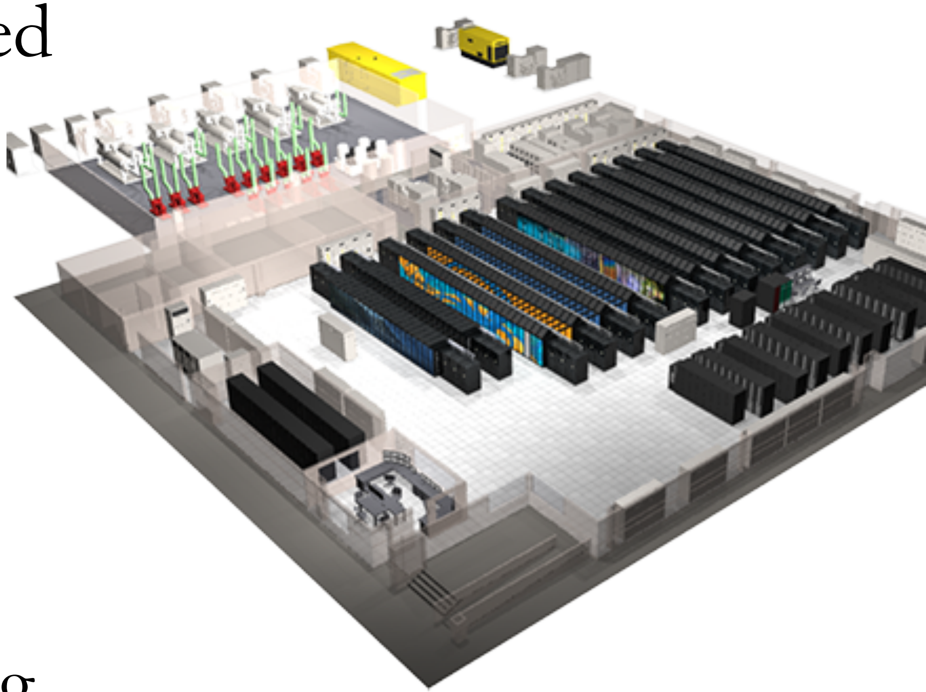
- Much data-center traffic goes between servers.
- Ideally, place related applications in same rack
- Use multiple interconnections to place less load on highest level switches.



Supercomputer (HPC) networks

- Supercomputers look superficially like data centers, but applications differ.
- The goal is to use entire cluster to do a single distributed computation.
- **Interconnect** speed is more important than speed of individual nodes.
- Typically uses **InifiniBand** links instead of Ethernet due to lower latency.
  - IB supports remote direct memory access (RDMA). OS need not run an interrupt handler when message arrives.
- COMP_ENG-358 Parallel Computing discusses various network topologies for parallel computing.

"Titan" Supercomputer at Oak Ridge National Lab

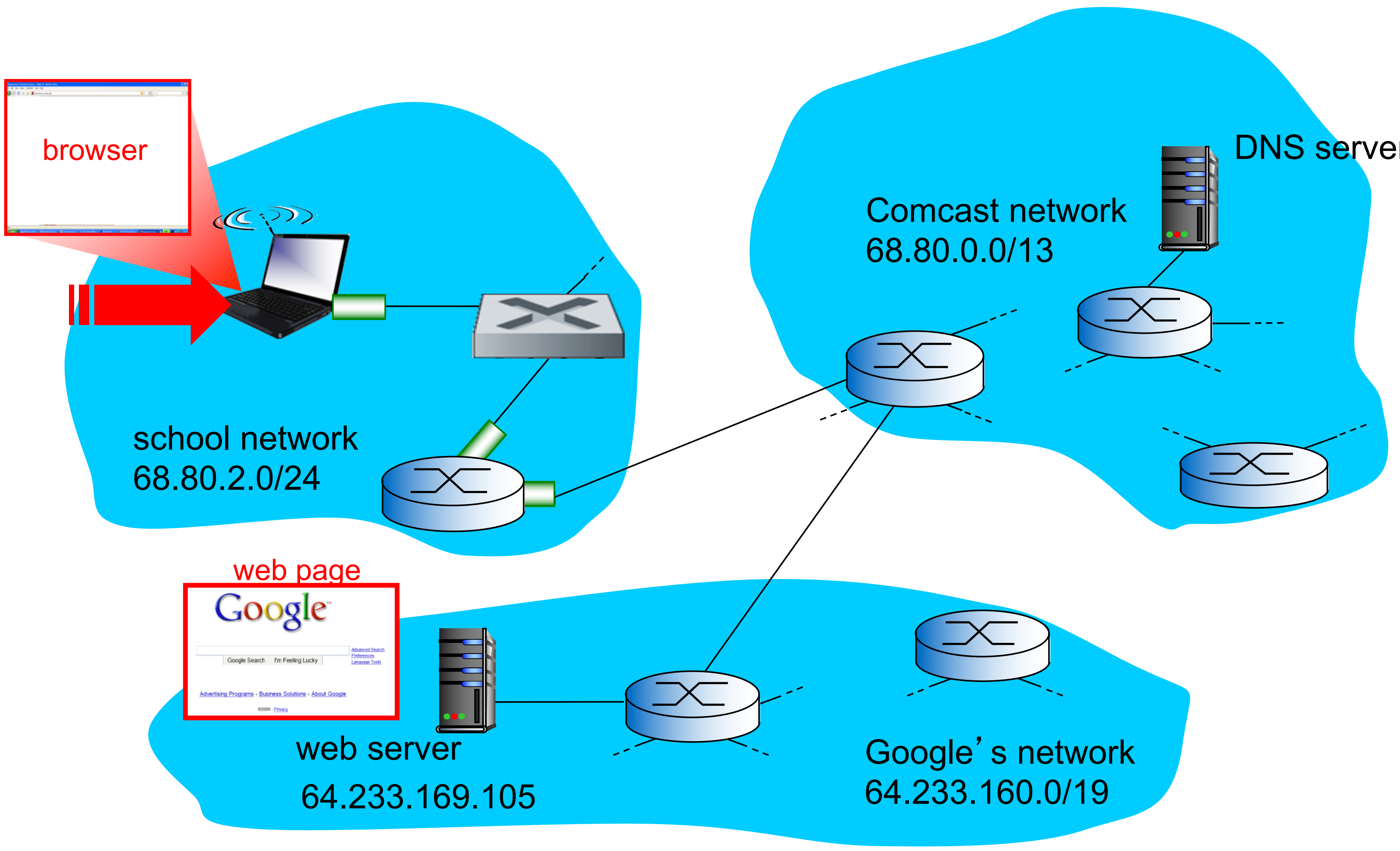


Recap

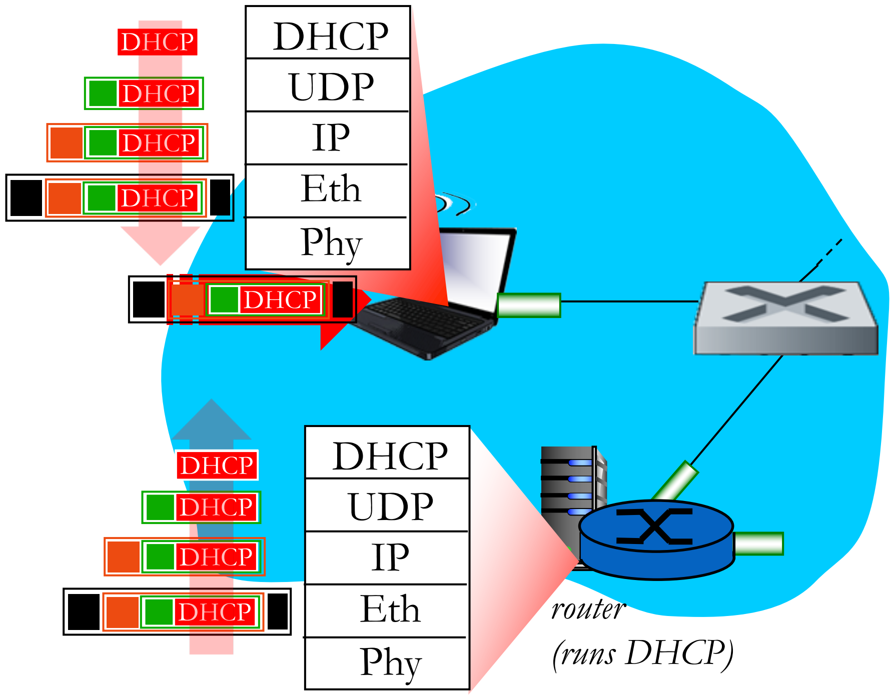
- Link layer handles error detection and correction: **Parity**, **Checksum**, and Cyclic Redandancy Check (**CRC**).
- Ethernet adds **MAC addresses** to identify src/dst on a shared link.
  - **ARP** uses Ethernet broadcast to find *IP address* → *MAC address* mapping
- **DHCP** requests are sent by Ethernet broadcast (to FF:FF:FF:FF:FF:FF)
- Old Ethernet **hubs** broadcasted data to all ports.
- Ethernet **switches** learn/remember which MAC addresses are reachable on each port and relay traffic only to the appropriate ports.
  - Reduce broadcast traffic and eliminate collisions.
- **VLANs** create multiple isolated LANs/subnets on one switch.
- **Data Centers** & **Supercomputers** demand fast local networks.

That concludes our trip down the networking stack!

What happens when visiting www.google.com?

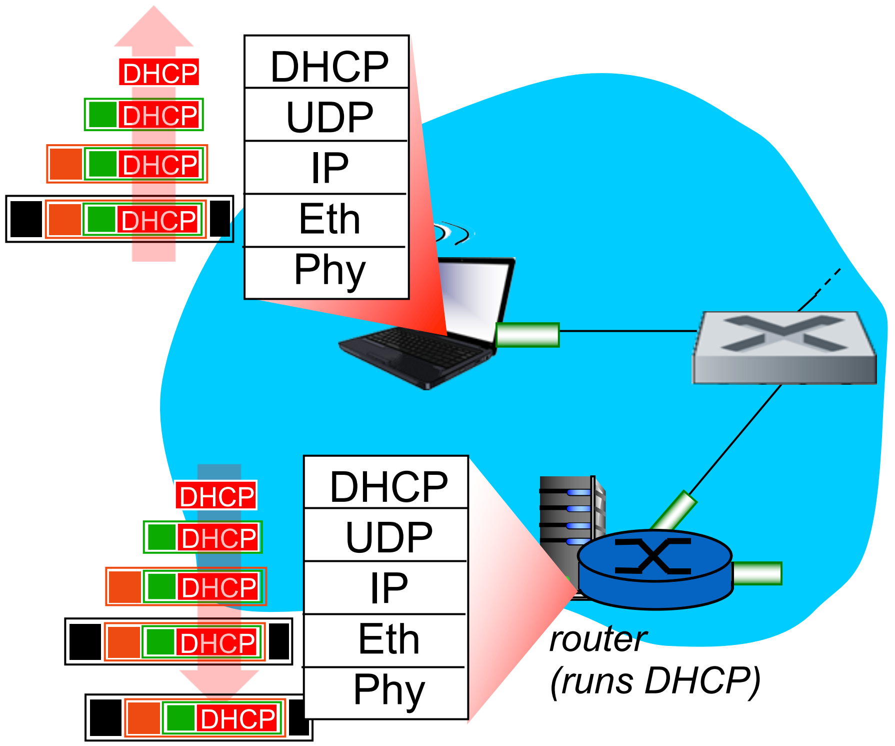


Step 1: Connect to the network



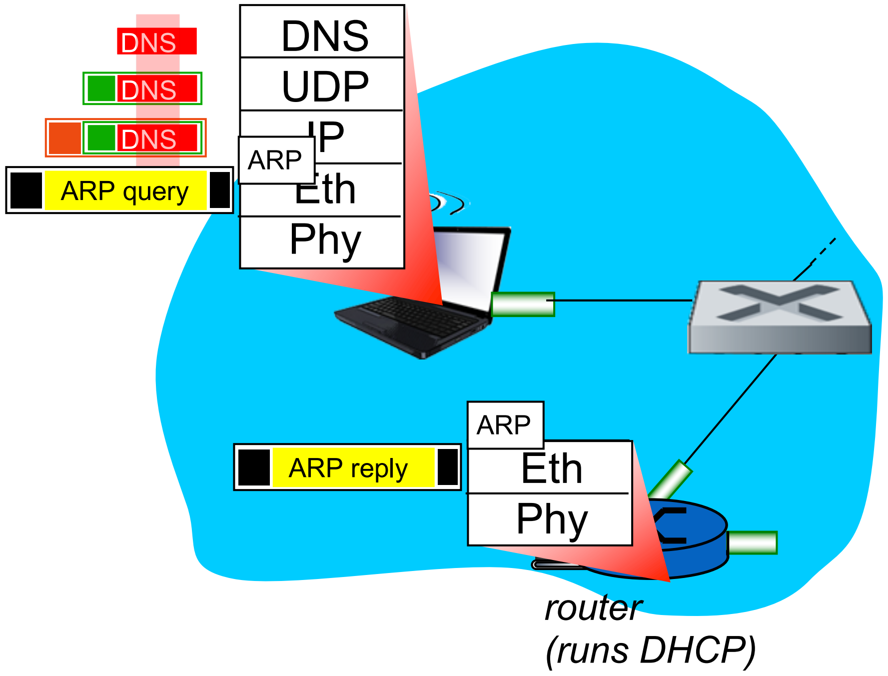
- connecting laptop needs to get its own IP address, addr. of first-hop router, addr. of DNS server: use *DHCP*
- DHCP request encapsulated in UDP, encapsulated in IP, encapsulated in 802.3 Ethernet.
- Ethernet frame broadcast (dest: FF:FF:FF:FF:FF:FF) on LAN, received by router running DHCP server
- Ethernet unpacked to get IP, unpacked to get UDP, unpacked to get DHCP

Connecting to network (continued)



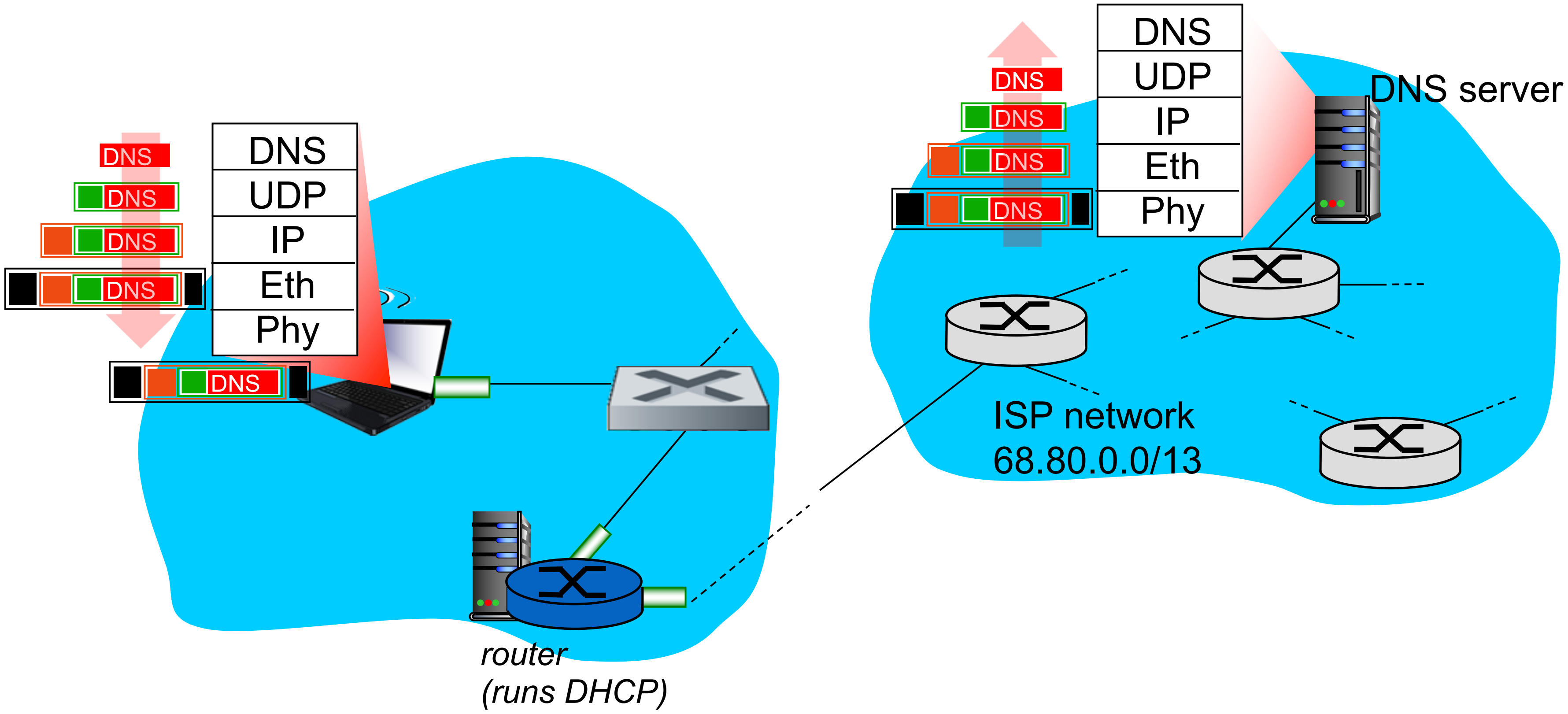
- DHCP server creates DHCP ACK containing client's IP address, IP address of first-hop router for client, subnet mask, & IP address of DNS server
- Encapsulation at DHCP server, frame forwarded (switch learning) through LAN, demultiplexing at client
- Client receives DHCP ACK

ARP (before DNS, before HTTP)



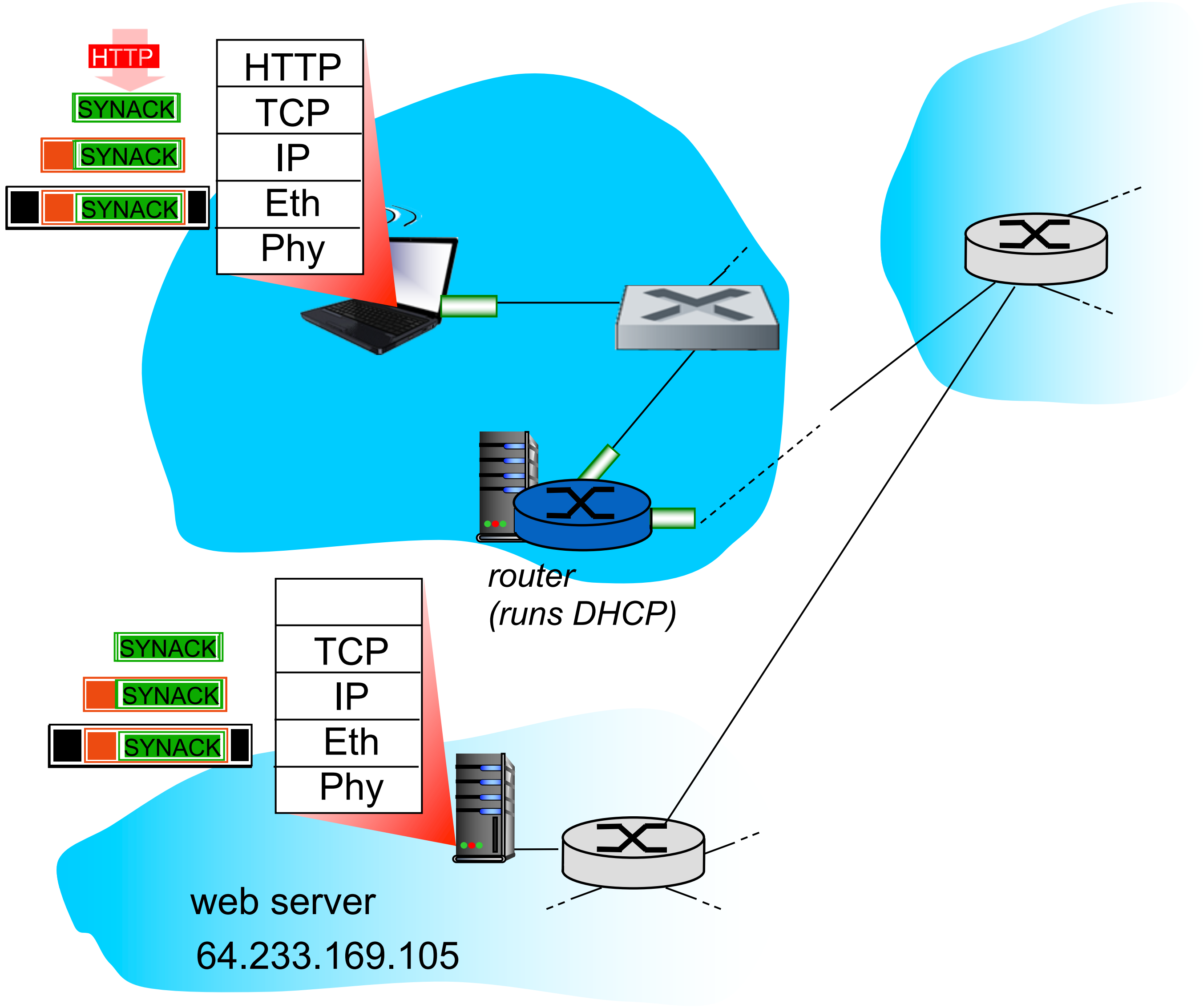
- Before sending HTTP request, we need to send a DNS request to get IP address of www.google.com
- Create DNS request, inside UDP segment, inside IP datagram, inside *Ethernet frame*, but we don't yet have the MAC address of the router to set as the first-hop Ethernet destination.
- Client broadcasts *ARP query* listing the router's IP address. Router replies with its MAC address (on that subnet).

DNS



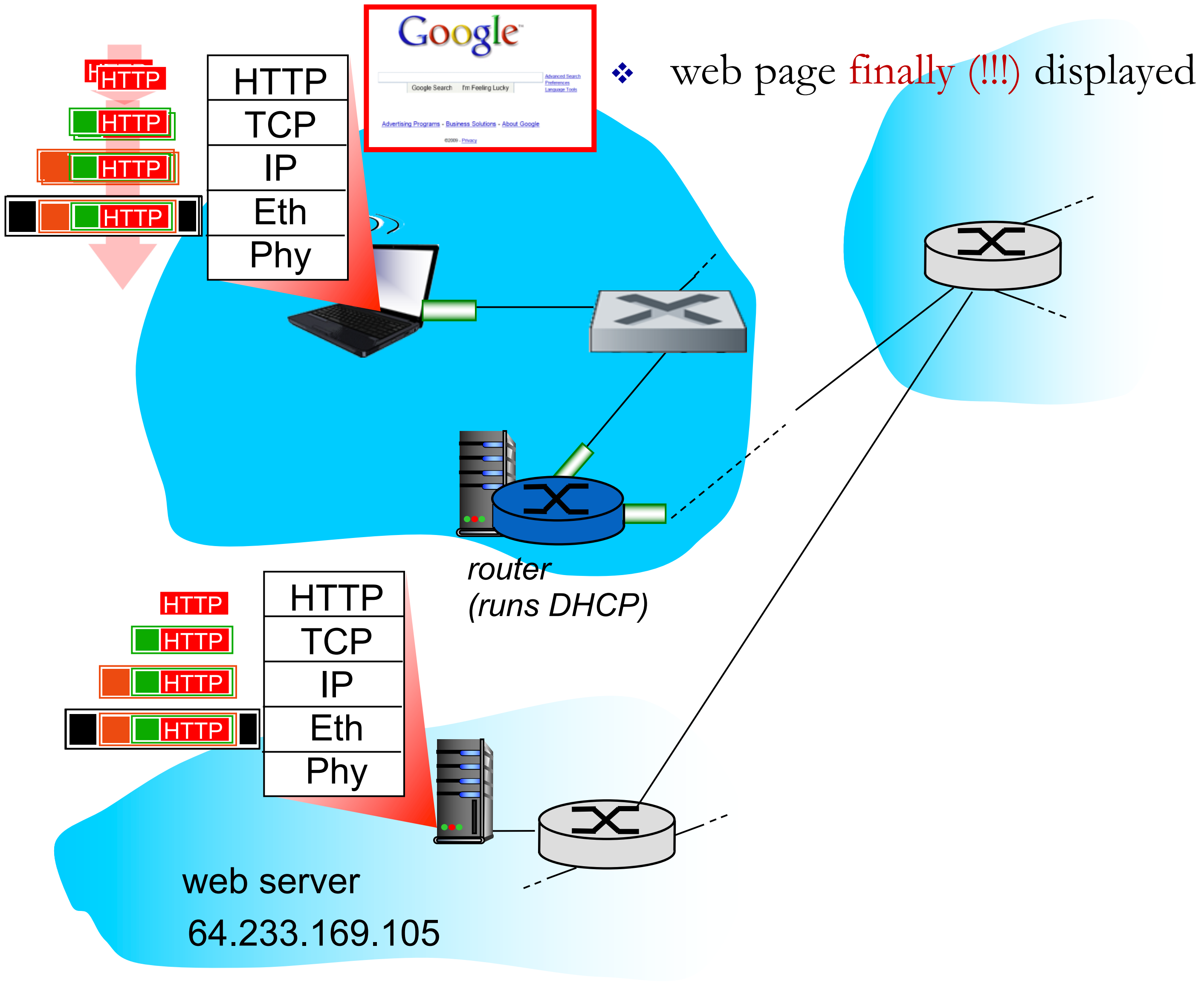
- IP datagram containing DNS query forwarded via LAN switch from client to 1st hop router. Router unwraps ethernet frame and rewraps with different MAC addresses (for next hop).
- IP datagram forwarded from campus network into ISP network, routed (tables already created by RIP, OSPF, IS-IS and/or BGP routing protocols) to DNS server.
- DNS server replies to client with IP address of www.google.com

TCP



- To send HTTP request, client opens TCP socket to server IP address, port 80.
- TCP SYN packet sent -- step 1 of 3-way handshake.
- Server sends SYN-ACK – step 2 of 3-way handshake
- TCP connection is established!

HTTP Request/Reply



- *HTTP request* sent into TCP socket.
  - IP datagram containing request routed to www.google.com.
- Web server replies with HTTP response (containing web page HTML).
  - IP datagram containing response routed back to client.