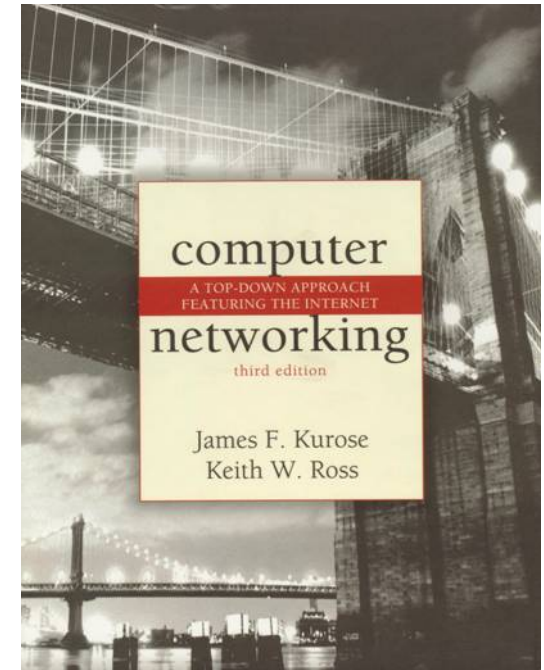


Chapter 5

Link Layer and LANs



*Computer Networking:
A Top Down Approach
Featuring the Internet,
3rd edition.*

Jim Kurose, Keith Ross
Addison-Wesley, July
2004.

Chapter 5: The Data Link Layer

Our goals:

- ❑ understand principles behind data link layer services:
 - error detection, correction
 - sharing a broadcast channel: multiple access
 - link layer addressing
 - reliable data transfer, flow control: *done!*
- ❑ instantiation and implementation of various link layer technologies

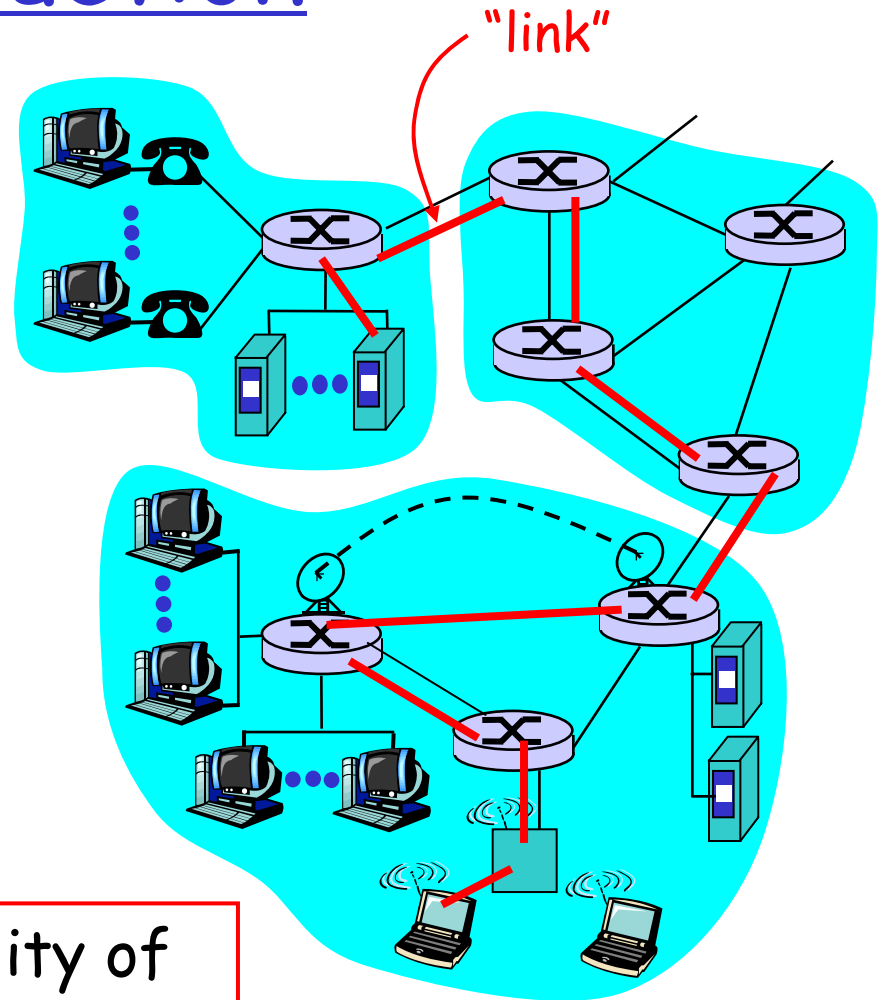
Link Layer

- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Hubs and switches

Link Layer: Introduction

Some terminology:

- ❑ hosts and routers are **nodes**
- ❑ communication channels that connect adjacent nodes along communication path are **links**
 - wired links
 - wireless links
 - LANs
- ❑ layer-2 packet is a **frame**, encapsulates datagram



data-link layer has responsibility of transferring datagram from one node to adjacent node over a link

Link layer: context

- ❑ Datagram transferred by different link protocols over different links:
 - e.g., Ethernet on first link, frame relay on intermediate links, 802.11 on last link
- ❑ Each link protocol provides different services
 - e.g., may or may not provide rdt over link

transportation analogy

- ❑ trip from Princeton to Lausanne
 - limo: Princeton to JFK
 - plane: JFK to Geneva
 - train: Geneva to Lausanne
- ❑ tourist = **datagram**
- ❑ transport segment = **communication link**
- ❑ transportation mode = **link layer protocol**
- ❑ travel agent = **routing algorithm**

Link Layer Services

❑ Framing, link access:

- encapsulate datagram into frame, adding header, trailer
- channel access if shared medium
- "MAC" addresses used in frame headers to identify source, dest
 - different from IP address!

❑ Reliable delivery between adjacent nodes

- we learned how to do this already (chapter 3)!
- seldom used on low bit error link (fiber, some twisted pair)
- wireless links: high error rates
 - Q: why both link-level and end-end reliability?

Link Layer Services (more)

❑ *Flow Control:*

- pacing between adjacent sending and receiving nodes

❑ *Error Detection:*

- errors caused by signal attenuation, noise.
- receiver detects presence of errors:
 - signals sender for retransmission or drops frame

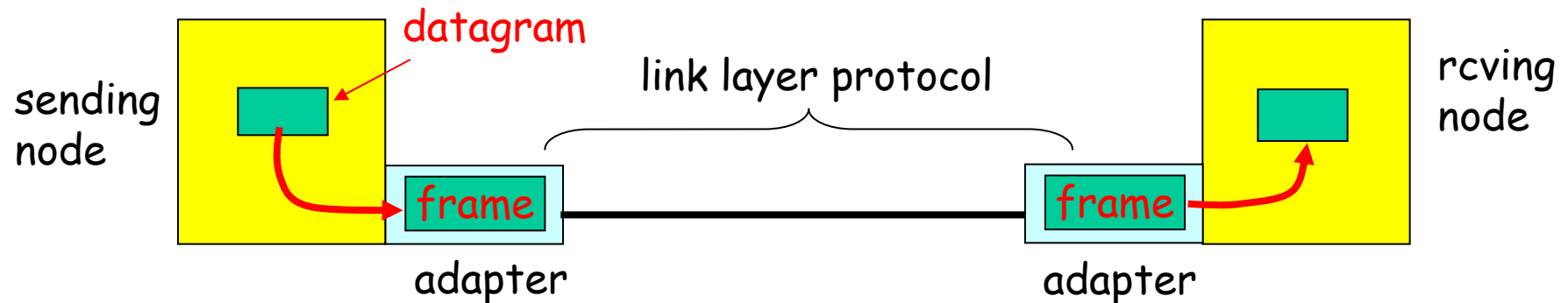
❑ *Error Correction:*

- receiver identifies *and corrects* bit error(s) without resorting to retransmission

❑ *Half-duplex and full-duplex*

- with half duplex, nodes at both ends of link can transmit, but not at same time

Adaptors Communicating



- ❑ link layer implemented in "adaptor" (aka NIC)
 - Ethernet card, PCMCIA card, 802.11 card
- ❑ sending side:
 - encapsulates datagram in a frame
 - adds error checking bits, rdt, flow control, etc.
- ❑ receiving side
 - looks for errors, rdt, flow control, etc
 - extracts datagram, passes to rcvng node
- ❑ adapter is semi-autonomous
- ❑ link & physical layers

Link Layer

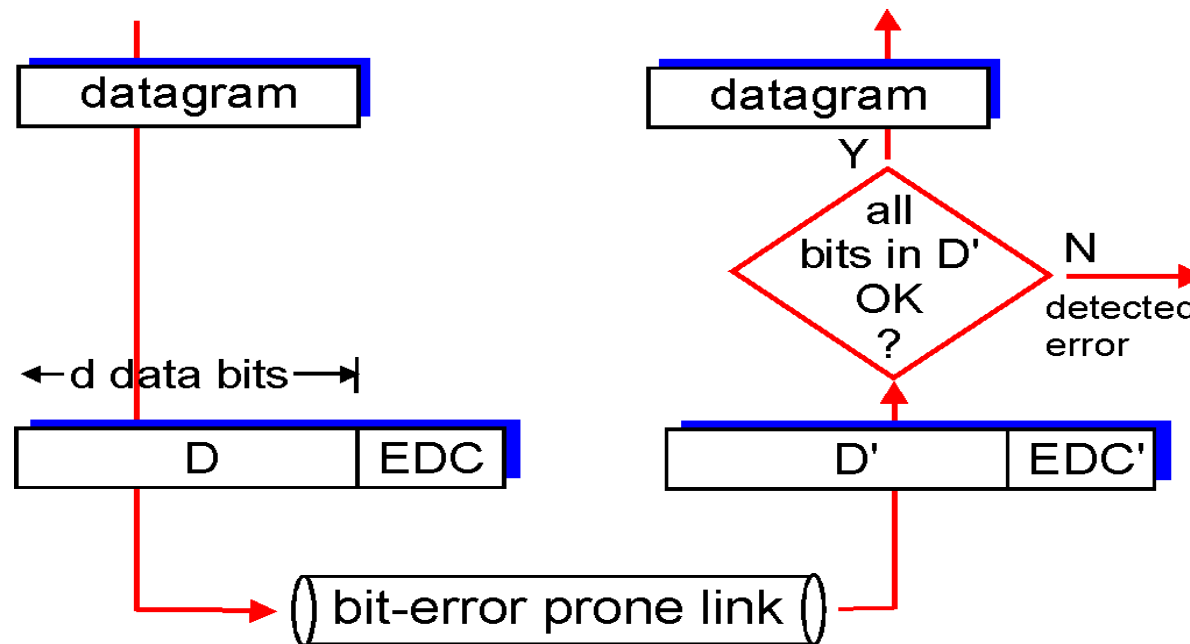
- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Hubs and switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM

Error Detection

EDC= Error Detection and Correction bits (redundancy)

D = Data protected by error checking, may include header fields

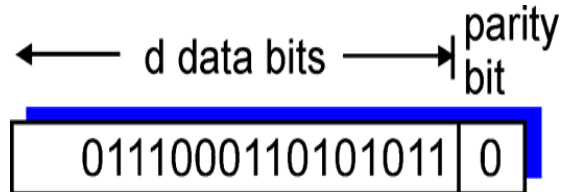
- Error detection not 100% reliable!
 - protocol may miss some errors, but rarely
 - larger EDC field yields better detection and correction



Parity Checking

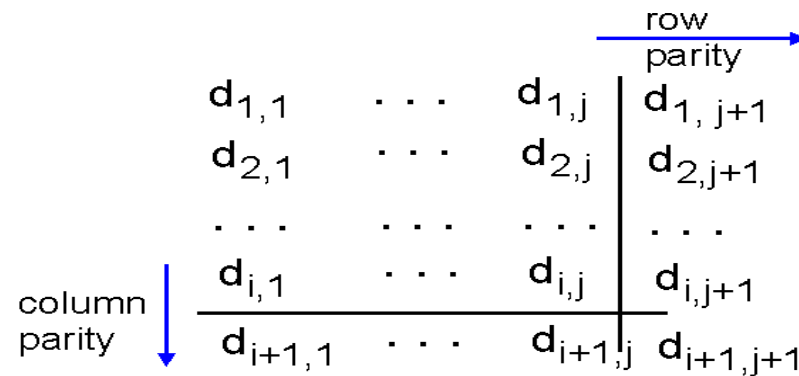
Single Bit Parity:

Detect single bit errors



Two Dimensional Bit Parity:

Detect *and correct* single bit errors



1	0	1	0	1	1
1	1	1	1	0	0
0	1	1	1	0	1
0	0	1	0	1	0

no errors

1	0	1	0	1	1
1	1	1	1	0	0
0	1	1	1	0	1
0	0	1	0	1	0

*correctable
single bit error*

Internet checksum

Goal: detect "errors" (e.g., flipped bits) in transmitted segment (note: used at transport layer *only*)

Sender:

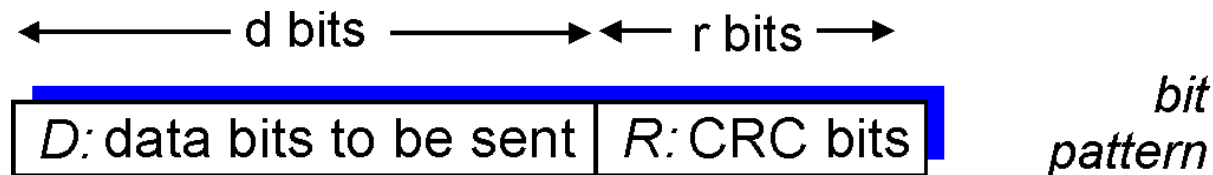
- ❑ treat segment contents as sequence of 16-bit integers
- ❑ checksum: addition (1's complement sum) of segment contents
- ❑ sender puts checksum value into UDP checksum field

Receiver:

- ❑ compute checksum of received segment
 - ❑ check if computed checksum equals checksum field value:
 - NO - error detected
 - YES - no error detected. *But maybe errors nonetheless?*
- More later

Checksumming: Cyclic Redundancy Check

- ❑ view data bits, **D**, as a binary number
- ❑ choose $r+1$ bit pattern (generator), **G**
- ❑ goal: choose r CRC bits, **R**, such that
 - $\langle D, R \rangle$ exactly divisible by G (modulo 2)
 - receiver knows G , divides $\langle D, R \rangle$ by G . If non-zero remainder: error detected!
 - can detect all burst errors less than $r+1$ bits
- ❑ widely used in practice (ATM, HDLC)



$$D * 2^r \text{ XOR } R$$

mathematical formula

CRC Example

Want:

$$D \cdot 2^r \text{ XOR } R = nG$$

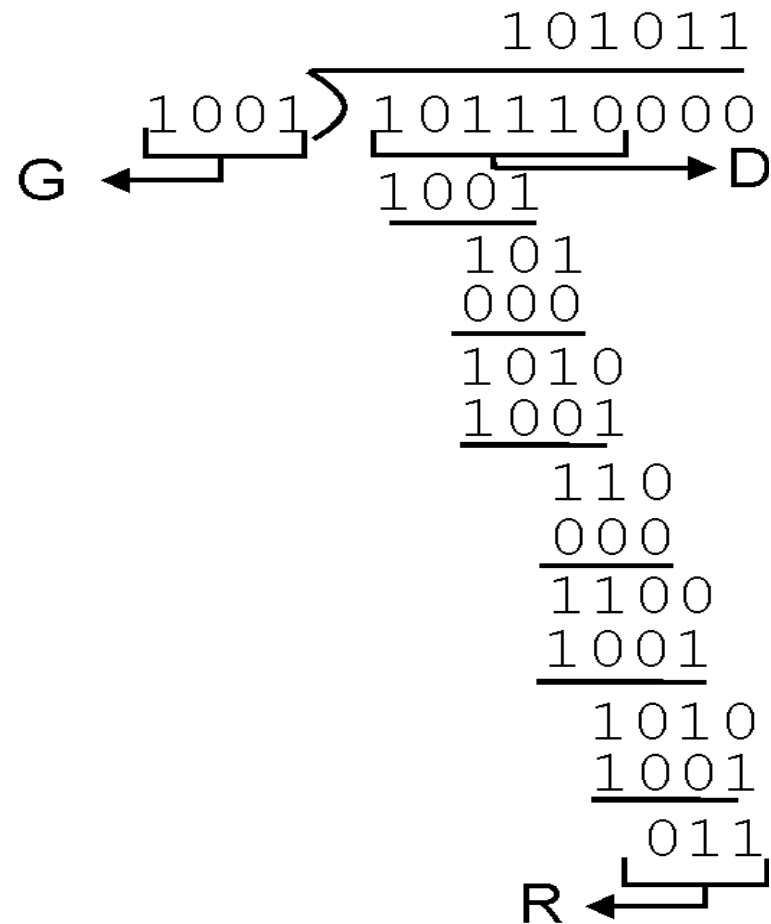
equivalently:

$$D \cdot 2^r = nG \text{ XOR } R$$

equivalently:

if we divide $D \cdot 2^r$ by G , want remainder R

$$R = \text{remainder} \left[\frac{D \cdot 2^r}{G} \right]$$



Link Layer

- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Hubs and switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM

Multiple Access Links and Protocols

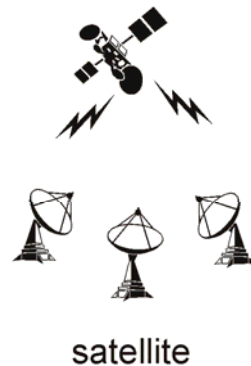
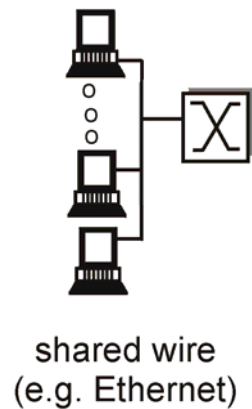
Two types of "links":

❑ point-to-point

- PPP for dial-up access
- point-to-point link between Ethernet switch and host

❑ **broadcast** (shared wire or medium)

- Old-fashioned Ethernet
- upstream HFC
- 802.11 wireless LAN



Multiple Access protocols

- ❑ single shared broadcast channel
- ❑ two or more simultaneous transmissions by nodes:
interference
 - collision if node receives two or more signals at the same time

multiple access protocol

- ❑ distributed algorithm that determines how nodes share channel, i.e., determine when node can transmit
- ❑ communication about channel sharing must use channel itself!
 - no out-of-band channel for coordination

Ideal Multiple Access Protocol

Broadcast channel of rate R bps

1. When only one node wants to transmit, it can send at rate R .
2. When M nodes want to transmit, each can send at average rate R/M
3. Fully decentralized:
 - no special node to coordinate transmissions
 - no synchronization of clocks, slots
4. Simple

MAC Protocols: a taxonomy

Three broad classes:

❑ Channel Partitioning

- divide channel into smaller "pieces" (time slots, frequency, code)
- allocate piece to node for exclusive use

❑ Random Access

- channel not divided, allow collisions
- "recover" from collisions

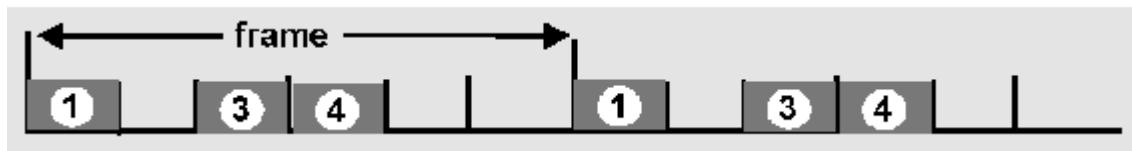
❑ "Taking turns"

- Nodes take turns, but nodes with more to send can take longer turns

Channel Partitioning MAC protocols: TDMA

TDMA: time division multiple access

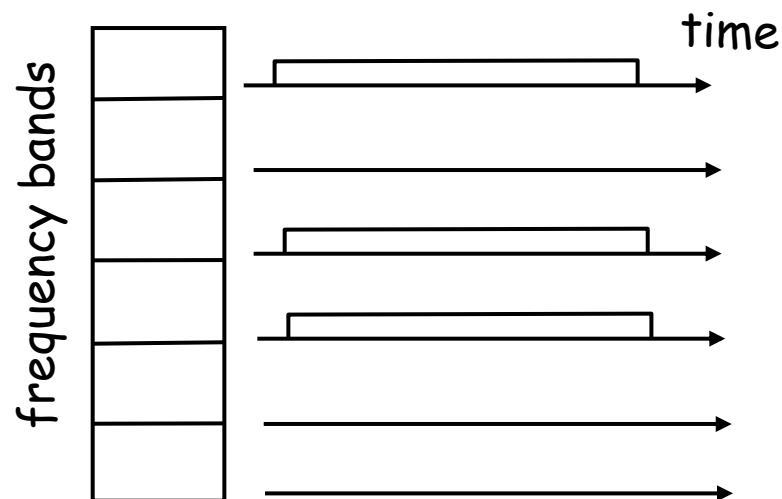
- ❑ access to channel in "rounds"
- ❑ each station gets fixed length slot (length = pkt trans time) in each round
- ❑ unused slots go idle
- ❑ example: 6-station LAN, 1,3,4 have pkt, slots 2,5,6 idle



Channel Partitioning MAC protocols: FDMA

FDMA: frequency division multiple access

- ❑ channel spectrum divided into frequency bands
- ❑ each station assigned fixed frequency band
- ❑ unused transmission time in frequency bands go idle
- ❑ example: 6-station LAN, 1,3,4 have pkt, frequency bands 2,5,6 idle



Random Access Protocols

- ❑ When node has packet to send
 - transmit at full channel data rate R .
 - no *a priori* coordination among nodes
- ❑ two or more transmitting nodes → “collision”,
- ❑ **random access MAC protocol** specifies:
 - how to detect collisions
 - how to recover from collisions (e.g., via delayed retransmissions)
- ❑ Examples of random access MAC protocols:
 - slotted ALOHA
 - ALOHA
 - CSMA, CSMA/CD, CSMA/CA

Slotted ALOHA

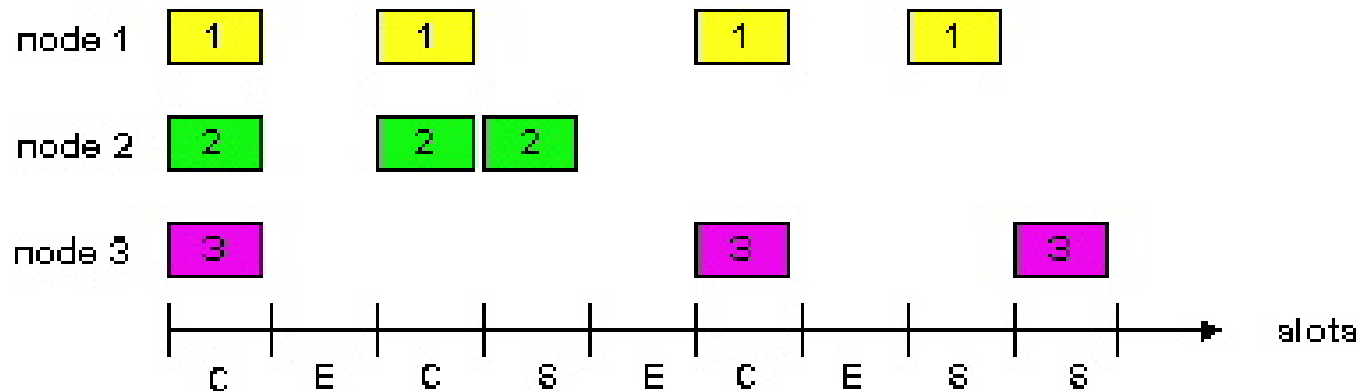
Assumptions

- ❑ all frames same size
- ❑ time is divided into equal size slots, time to transmit 1 frame
- ❑ nodes start to transmit frames only at beginning of slots
- ❑ nodes are synchronized
- ❑ if 2 or more nodes transmit in slot, all nodes detect collision

Operation

- ❑ when node obtains fresh frame, it transmits in next slot
- ❑ no collision, node can send new frame in next slot
- ❑ if collision, node retransmits frame in each subsequent slot with prob. p until success

Slotted ALOHA



Pros

- ❑ single active node can continuously transmit at full rate of channel
- ❑ highly decentralized: only slots in nodes need to be in sync
- ❑ simple

Cons

- ❑ collisions, wasting slots
- ❑ idle slots
- ❑ nodes may be able to detect collision in less than time to transmit packet
- ❑ clock synchronization

Slotted Aloha efficiency

Efficiency is the long-run fraction of successful slots when there are many nodes, each with many frames to send

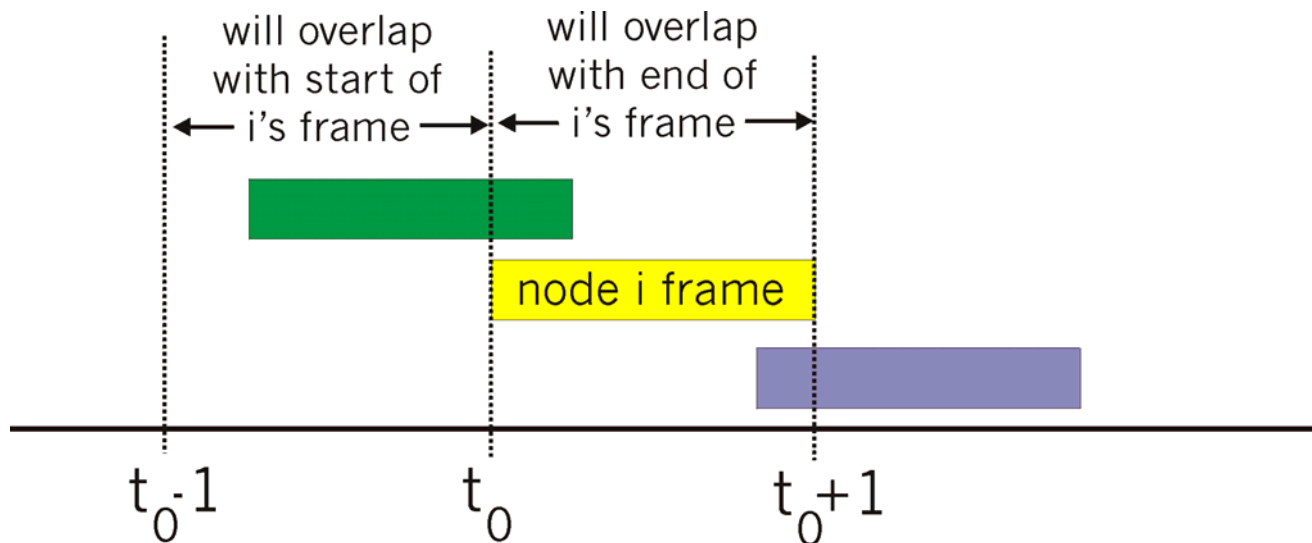
- Suppose N nodes with many frames to send, each transmits in slot with probability p
- prob that node 1 has success in a slot
 $= p(1-p)^{N-1}$
- prob that any node has a success $= Np(1-p)^{N-1}$

- For max efficiency with N nodes, find p^* that maximizes $Np(1-p)^{N-1}$
- For many nodes, take limit of $Np^*(1-p^*)^{N-1}$ as N goes to infinity, gives $1/e = .37$

At best: channel used for useful transmissions 37% of time!

Pure (unslotted) ALOHA

- ❑ unslotted Aloha: simpler, no synchronization
- ❑ when frame first arrives
 - transmit immediately
- ❑ collision probability increases:
 - frame sent at t_0 collides with other frames sent in $[t_0-1, t_0+1]$



Pure Aloha efficiency

$P(\text{success by given node}) = P(\text{node transmits}) \cdot$

$P(\text{no other node transmits in } [t_0-1, t_0] \cdot$

$P(\text{no other node transmits in } [t_0, t_0+1]$

$$= p \cdot (1-p)^{N-1} \cdot (1-p)^{N-1}$$

$$= p \cdot (1-p)^{2(N-1)}$$

... choosing optimum p and then letting $n \rightarrow \infty$...

Even worse ! $= 1/(2e) = .18$

CSMA (Carrier Sense Multiple Access)

CSMA: listen before transmit:

If channel sensed idle: transmit entire frame

- ❑ If channel sensed busy, defer transmission

- ❑ Human analogy: don't interrupt others!

CSMA collisions

collisions *can* still occur:

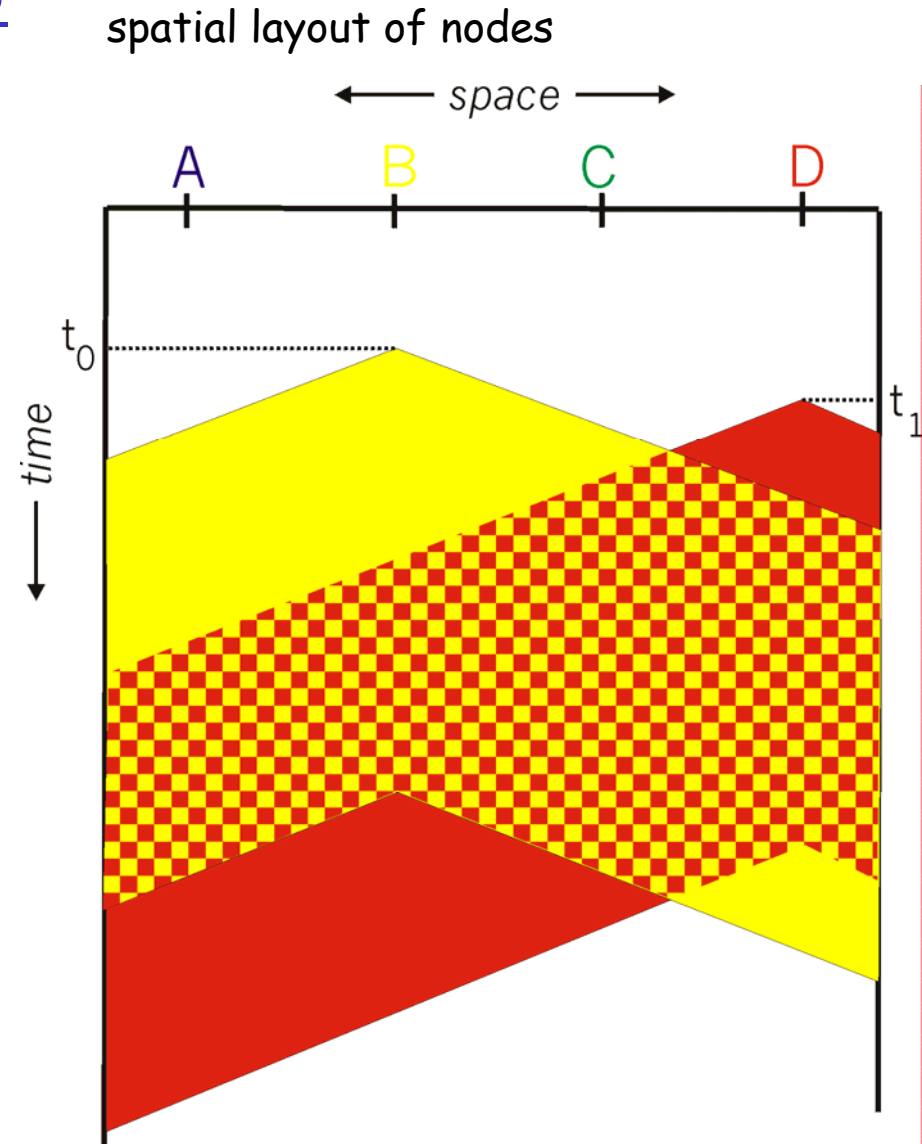
propagation delay means
two nodes may not hear
each other's transmission

collision:

entire packet transmission
time wasted

note:

role of distance & propagation
delay in determining collision
probability



CSMA/CD (Collision Detection)

CSMA/CD: carrier sensing, deferral as in CSMA

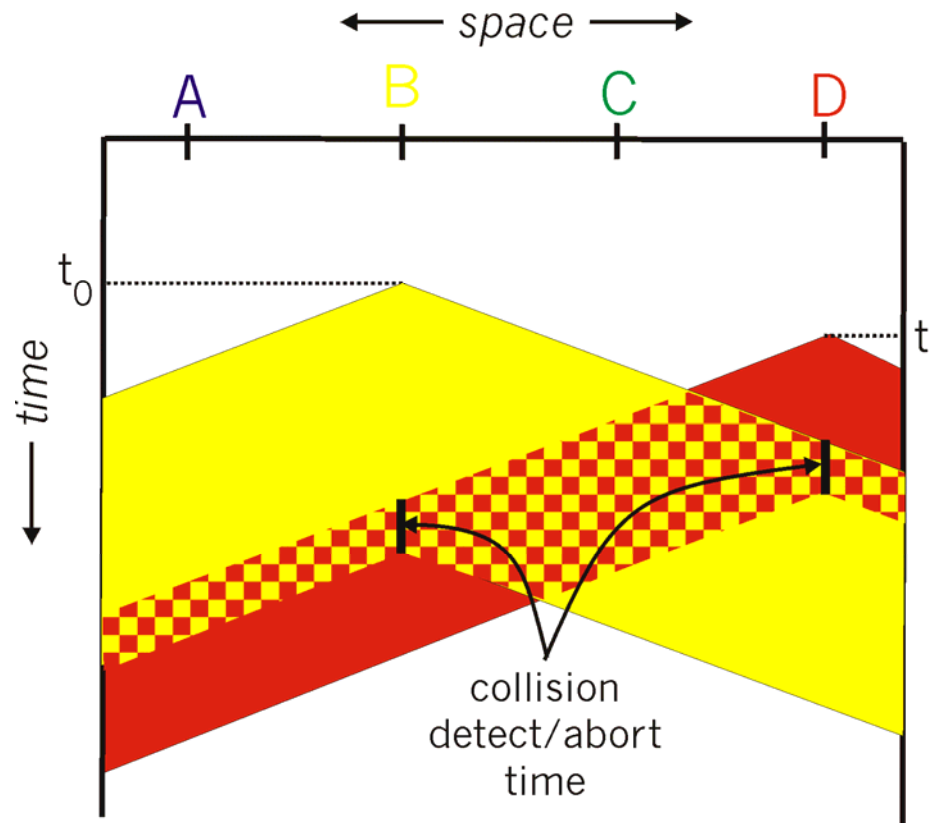
- collisions *detected* within short time
- colliding transmissions aborted, reducing channel wastage

□ collision detection:

- easy in wired LANs: measure signal strengths, compare transmitted, received signals
- difficult in wireless LANs: receiver shut off while transmitting

□ human analogy: the polite conversationalist

CSMA/CD collision detection



"Taking Turns" MAC protocols

channel partitioning MAC protocols:

- share channel efficiently and fairly at high load
- inefficient at low load: delay in channel access, $1/N$ bandwidth allocated even if only 1 active node!

Random access MAC protocols

- efficient at low load: single node can fully utilize channel
- high load: collision overhead

"taking turns" protocols

look for best of both worlds!

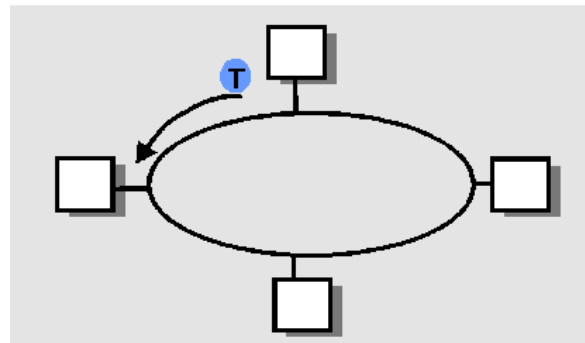
"Taking Turns" MAC protocols

Polling:

- ❑ master node
 - "invites" slave nodes to transmit in turn
- ❑ concerns:
 - polling overhead
 - latency
 - single point of failure (master)

Token passing:

- ❑ control **token** passed from one node to next sequentially.
- ❑ token message
- ❑ concerns:
 - token overhead
 - latency
 - single point of failure (token)



Summary of MAC protocols

- ❑ What do you do with a shared media?
 - Channel Partitioning, by time, frequency or code
 - Time Division, Frequency Division
 - Random partitioning (dynamic),
 - ALOHA, S-ALOHA, CSMA, CSMA/CD
 - carrier sensing: easy in some technologies (wire), hard in others (wireless)
 - CSMA/CD used in Ethernet
 - CSMA/CA used in 802.11
 - Taking Turns
 - polling from a central site, token passing

LAN technologies

Data link layer so far:

- services, error detection/correction, multiple access

Next: LAN technologies

- addressing
- Ethernet
- hubs, switches
- PPP

Link Layer

- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Hubs and switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM

MAC Addresses and ARP

□ 32-bit IP address:

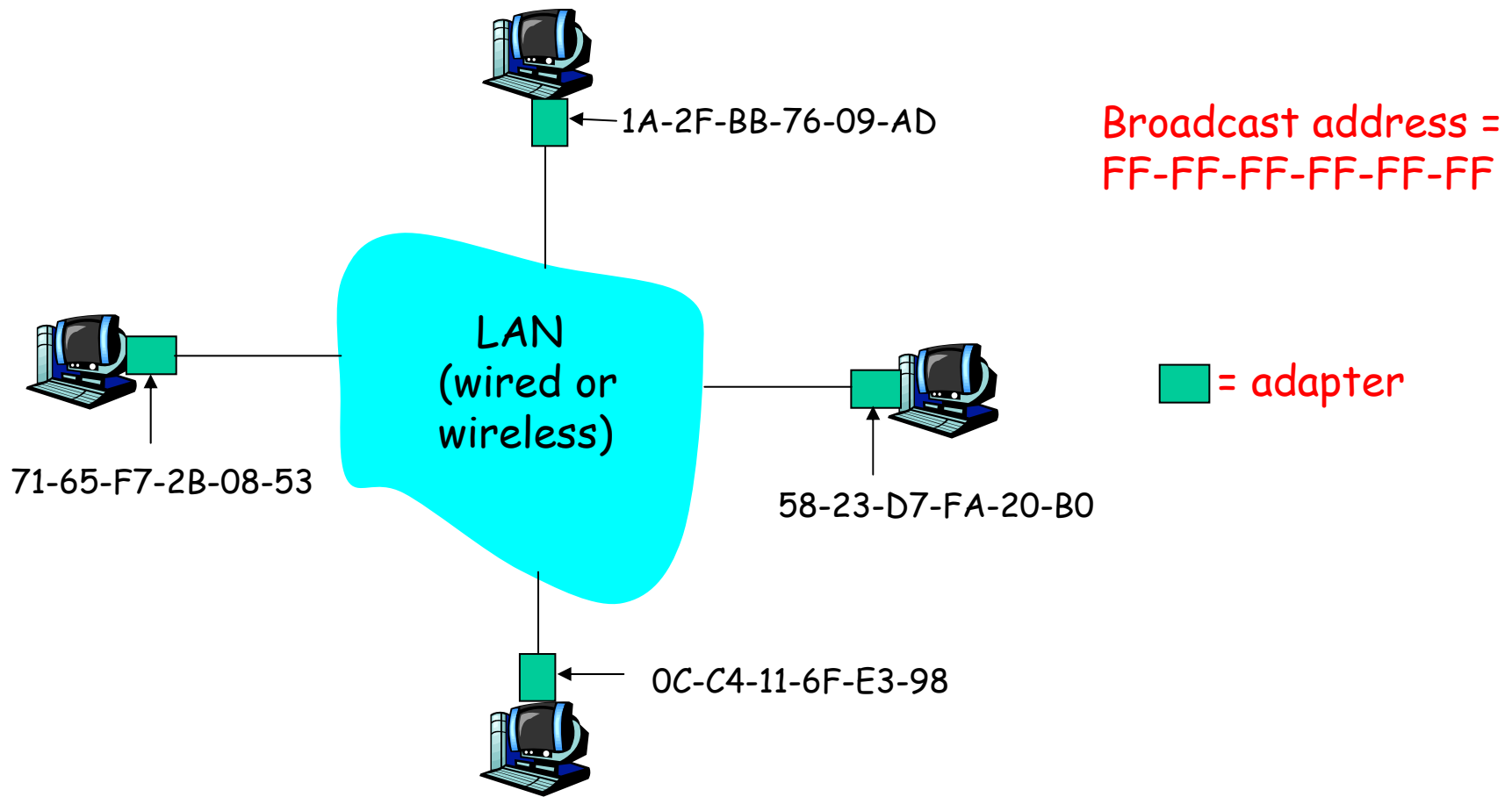
- *network-layer* address
- used to get datagram to destination IP subnet

□ MAC (or LAN or physical or Ethernet) address:

- used to get frame from one interface to another physically-connected interface (same network)
- 48 bit MAC address (for most LANs)
burned in the adapter ROM

LAN Addresses and ARP

Each adapter on LAN has unique LAN address



LAN Address (more)

- ❑ MAC address allocation administered by IEEE
- ❑ manufacturer buys portion of MAC address space (to assure uniqueness)
- ❑ Analogy:
 - (a) MAC address: like Social Security Number
 - (b) IP address: like postal address
- ❑ MAC flat address → portability
 - can move LAN card from one LAN to another
- ❑ IP hierarchical address NOT portable
 - depends on IP subnet to which node is attached

ARP: Address Resolution Protocol

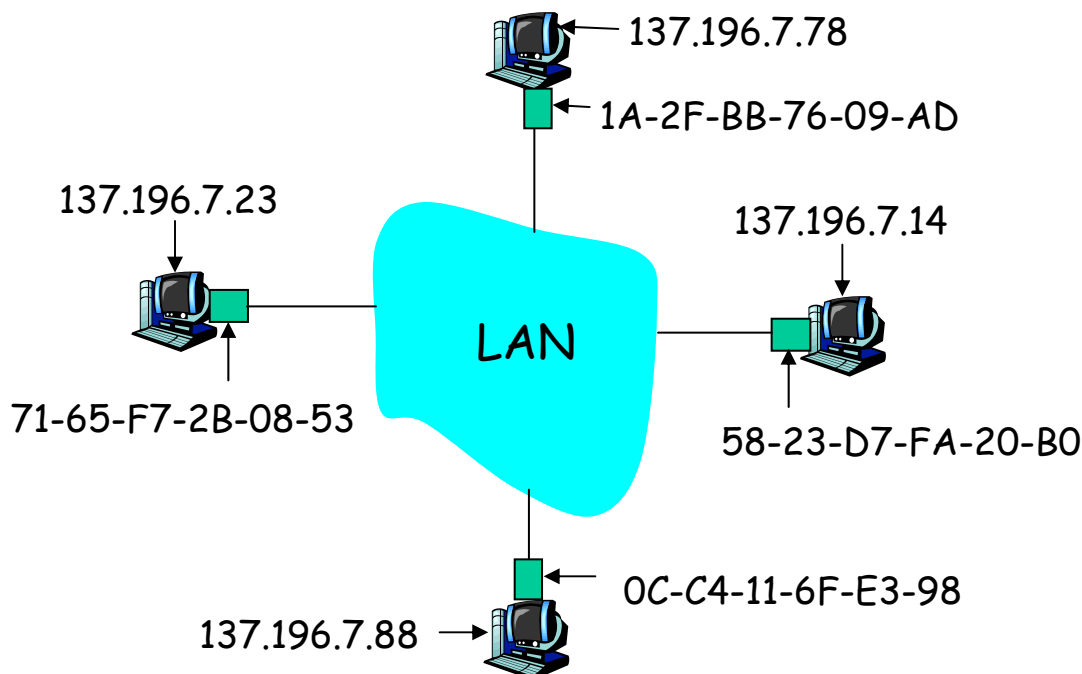
Question: how to determine MAC address of B knowing B's IP address?

- Each IP node (Host, Router) on LAN has **ARP** table

- ARP Table: IP/MAC address mappings for same LAN nodes

< IP address; MAC address; TTL >

- TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min)

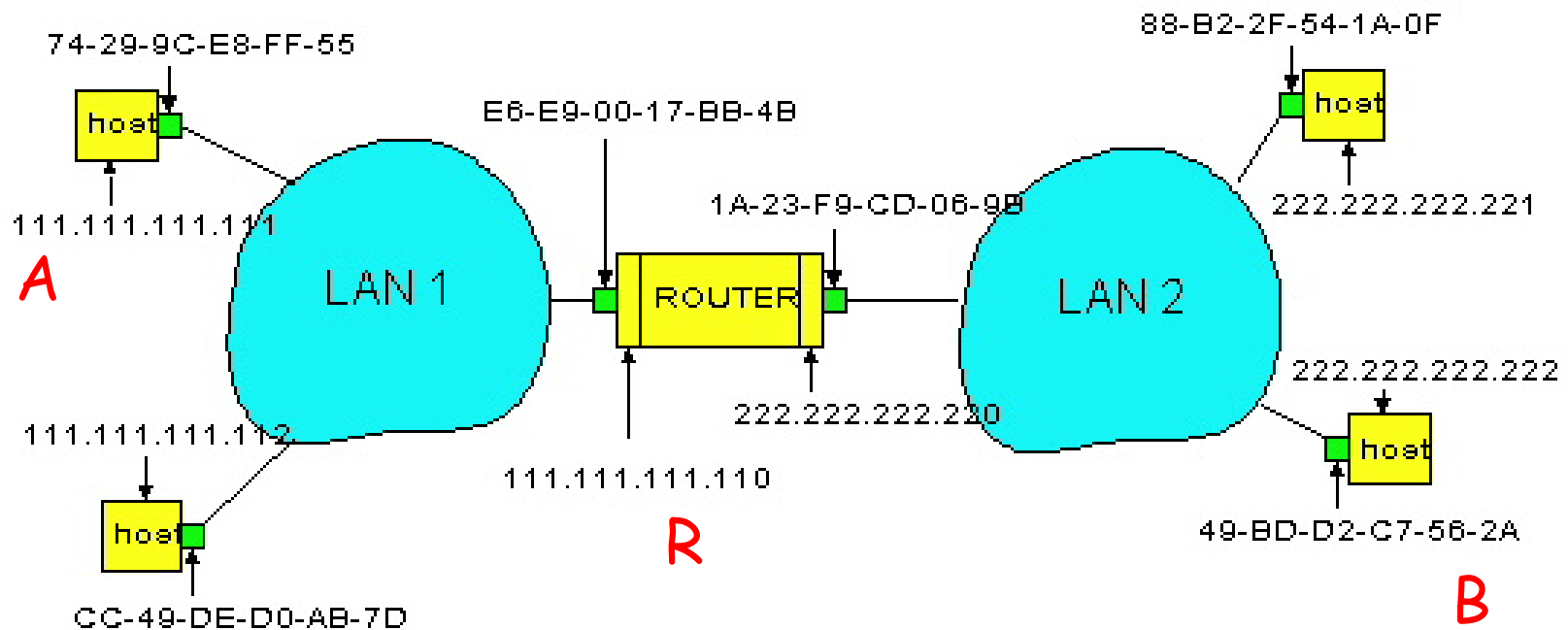


ARP protocol: Same LAN (network)

- ❑ A wants to send datagram to B, and B's MAC address not in A's ARP table.
- ❑ A **broadcasts** ARP query packet, containing B's IP address
 - Dest MAC address = FF-FF-FF-FF-FF-FF
 - all machines on LAN receive ARP query
- ❑ B receives ARP packet, replies to A with its (B's) MAC address
 - frame sent to A's MAC address (unicast)
- ❑ A caches (saves) IP-to-MAC address pair in its ARP table until information becomes old (times out)
 - soft state: information that times out (goes away) unless refreshed
- ❑ ARP is "plug-and-play":
 - nodes create their ARP tables without intervention from net administrator

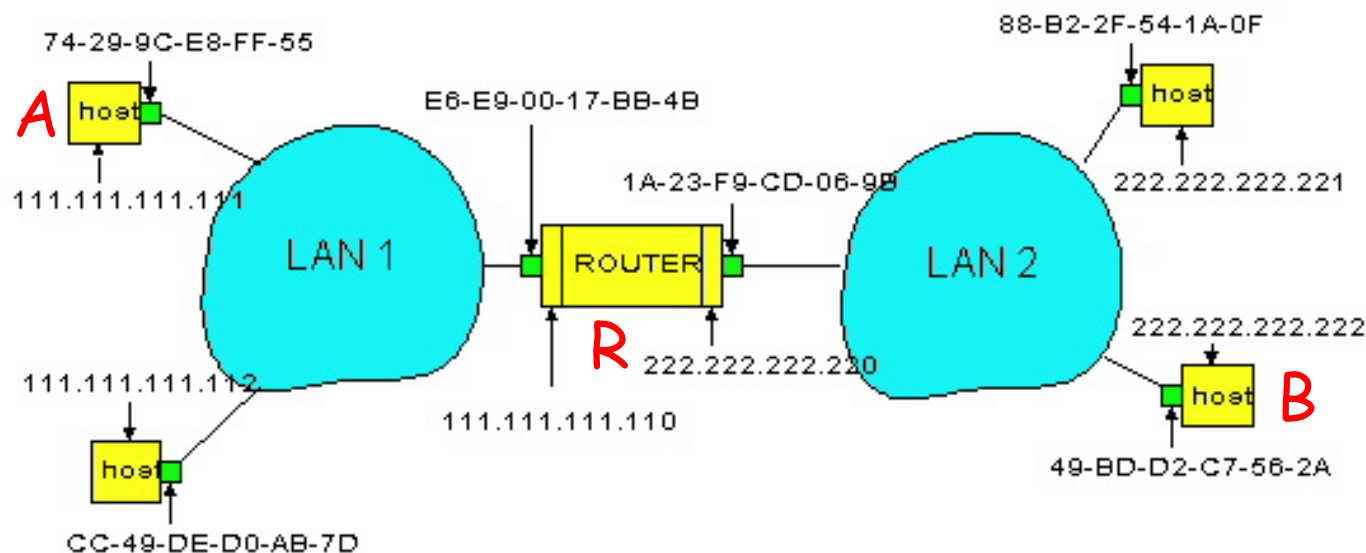
Routing to another LAN

walkthrough: **send datagram from A to B via R**
assume A knows B's IP address



- ❑ Two ARP tables in router R, one for each IP network (LAN)

- ❑ A creates datagram with source A, destination B
- ❑ A uses ARP to get R's MAC address for 111.111.111.110
- ❑ A creates link-layer frame with R's MAC address as dest, frame contains A-to-B IP datagram
- ❑ A's adapter sends frame
- ❑ R's adapter receives frame
- ❑ R removes IP datagram from Ethernet frame, sees its destined to B
- ❑ R uses ARP to get B's MAC address
- ❑ R creates frame containing A-to-B IP datagram sends to B



DHCP: Dynamic Host Configuration Protocol

Goal: allow host to *dynamically* obtain its IP address from network server when it joins network

Can renew its lease on address in use

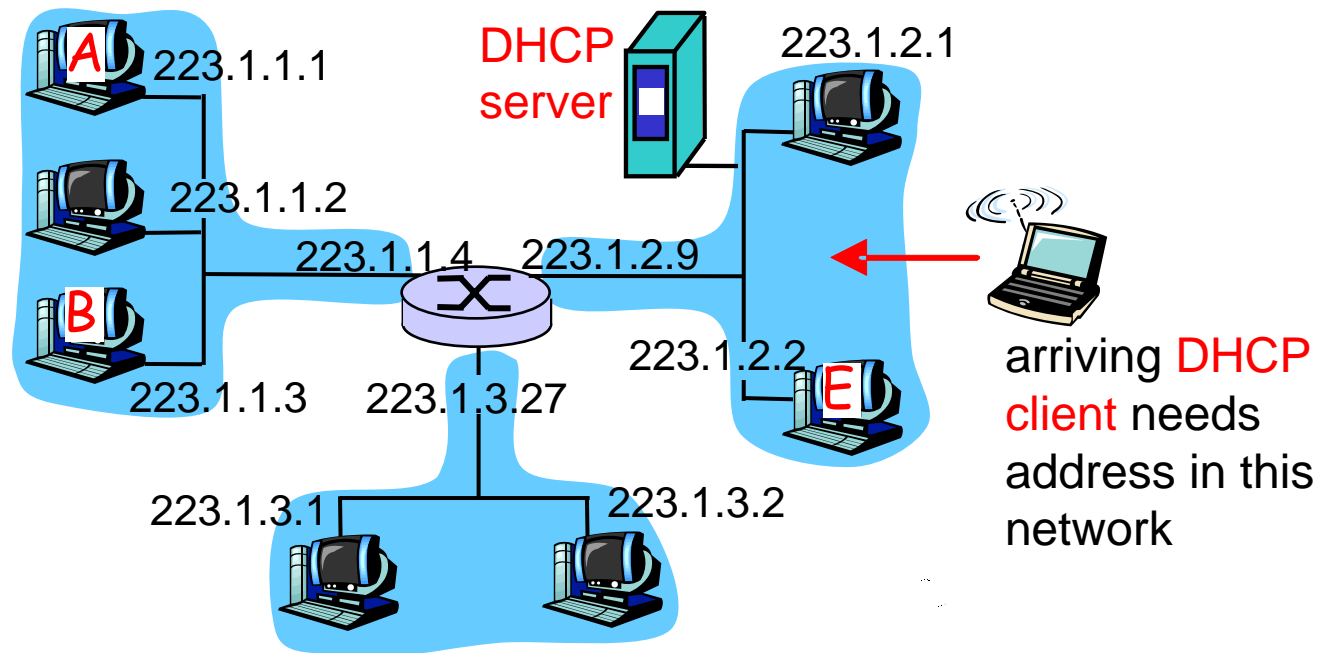
Allows reuse of addresses (only hold address while connected an "on")

Support for mobile users who want to join network (more shortly)

DHCP overview:

- host broadcasts "DHCP discover" msg
- DHCP server responds with "DHCP offer" msg
- host requests IP address: "DHCP request" msg
- DHCP server sends address: "DHCP ack" msg

DHCP client-server scenario



DHCP client-server scenario

DHCP server: 223.1.2.5

DHCP discover

src : 0.0.0.0, 68
dest.: 255.255.255.255, 67
yiaddr: 0.0.0.0
transaction ID: 654

arriving
client



DHCP offer

src: 223.1.2.5, 67
dest: 255.255.255.255, 68
yiaddr: 223.1.2.4
transaction ID: 654
Lifetime: 3600 secs

DHCP request

src: 0.0.0.0, 68
dest.: 255.255.255.255, 67
yiaddr: 223.1.2.4
transaction ID: 655
Lifetime: 3600 secs

DHCP ACK

src: 223.1.2.5, 67
dest: 255.255.255.255, 68
yiaddr: 223.1.2.4
transaction ID: 655
Lifetime: 3600 secs

time

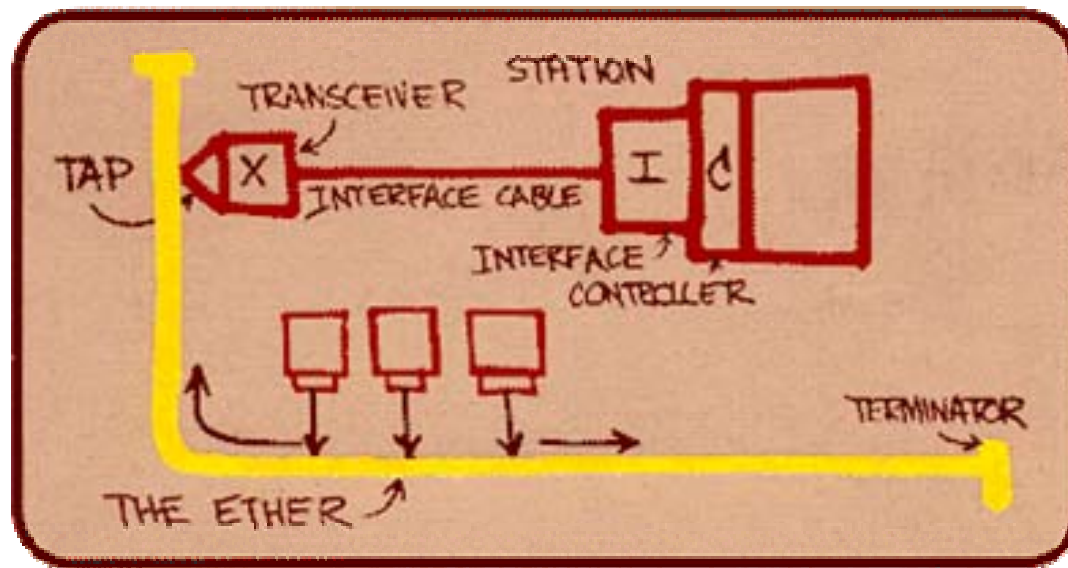
Link Layer

- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Hubs and switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM

Ethernet

"dominant" wired LAN technology:

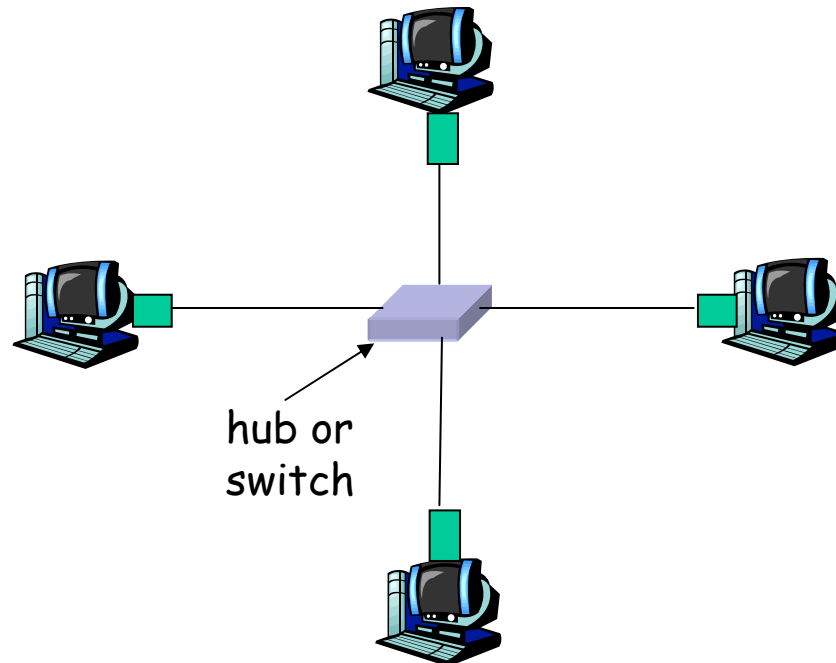
- ❑ cheap \$20 for 100Mbps!
- ❑ first widely used LAN technology
- ❑ Simpler, cheaper than token LANs and ATM
- ❑ Kept up with speed race: 10 Mbps - 10 Gbps



Metcalfe's Ethernet sketch

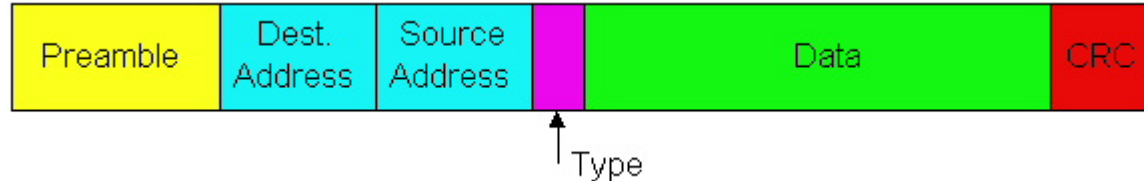
Star topology

- ❑ Bus topology popular through mid 90s
- ❑ Now star topology prevails
- ❑ Connection choices: hub or switch (more later)



Ethernet Frame Structure

Sending adapter encapsulates IP datagram (or other network layer protocol packet) in **Ethernet frame**

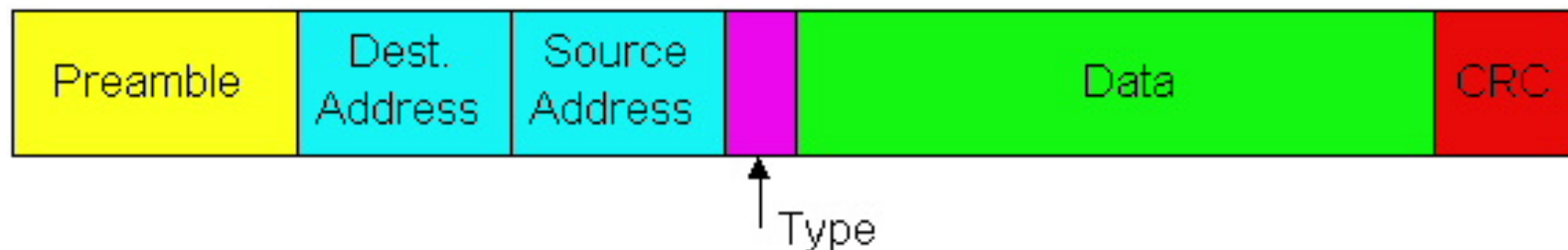


Preamble:

- ❑ 7 bytes with pattern 10101010 followed by one byte with pattern 10101011
- ❑ used to synchronize receiver, sender clock rates

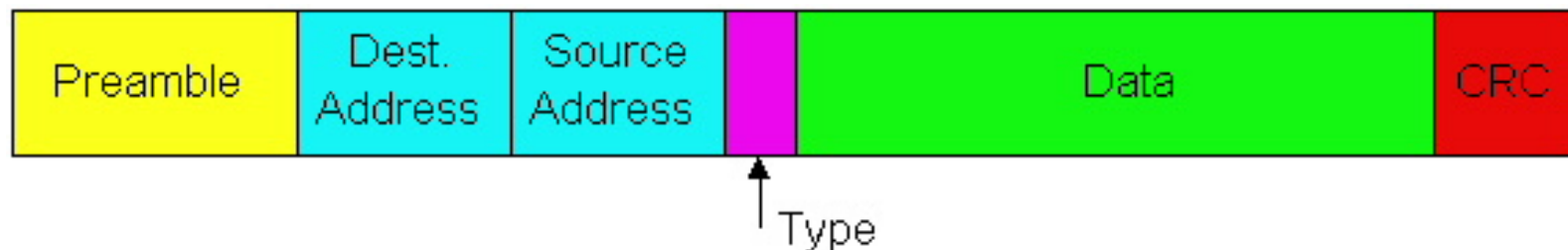
Ethernet Frame Structure (more)

- ❑ **Addresses:** 6 bytes dest. address and 6 bytes source address
 - if adapter receives frame with matching destination address, or with broadcast address (eg ARP packet), it passes data in frame to net-layer protocol
 - otherwise, adapter discards frame
- ❑ **Type:** 2 bytes indicates the higher layer protocol (mostly IP but others may be supported such as Novell IPX and AppleTalk)
- ❑ **CRC :** 4 bytes checked at receiver, if error is detected, the frame is simply dropped



Ethernet Frame Structure (more)

- ❑ **Data:** 46 to 1500 bytes
- ❑ Minimum frame size: 8 preamble + 6 dest. Add. + 6 source add. + 2 type + 46 data + 4 CRC = 72 bytes .
64 bytes (512 bits) excluding 8 bytes preamble
- ❑ Maximum frame size: 1500 data + 26 = 1526 bytes



Unreliable, connectionless service

- ❑ **Connectionless:** No handshaking between sending and receiving adapter.
- ❑ **Unreliable:** receiving adapter doesn't send acks or nacks to sending adapter
 - stream of datagrams passed to network layer can have gaps
 - gaps will be filled if app is using TCP
 - otherwise, app will see the gaps

Ethernet uses CSMA/CD

- ❑ No slots
- ❑ adapter doesn't transmit if it senses that some other adapter is transmitting, that is, **carrier sense**
- ❑ transmitting adapter aborts when it senses that another adapter is transmitting, that is, **collision detection**
- ❑ Before attempting a retransmission, adapter waits a random time, that is, **random access**

Ethernet CSMA/CD algorithm

1. Adaptor receives datagram from net layer & creates frame
2. If adapter senses channel idle for 96 bit times, it starts to transmit frame. If it senses channel busy, waits until channel idle for 96 bit times and then transmits
3. If adapter transmits entire frame without detecting another transmission, the adapter is done with frame !
4. If adapter detects another transmission while transmitting, aborts and sends 48-bit jam signal
5. After aborting, adapter enters **exponential backoff**: after the n -th collision, adapter chooses a K at random from $\{0, 1, 2, \dots, 2^m - 1\}$, where $m = \min(n, 10)$. Adapter waits $K \cdot 512$ bit times and returns to Step 2

Ethernet's CSMA/CD (more)

Jam Signal: make sure all other transmitters are aware of collision; 48 bits

Bit time: .1 microsec for 10 Mbps Ethernet ;
for $K=1023$, wait time is about 50 msec

See/interact with Java applet on AWL Web site: highly recommended !

Exponential Backoff:

- ❑ *Goal:* adapt retransmission attempts to estimated current load
 - heavy load: random wait will be longer
- ❑ first collision: choose K from $\{0,1\}$; delay is $K \cdot 512$ bit transmission times
- ❑ after second collision: choose K from $\{0,1,2,3\}$...
- ❑ after ten collisions, choose K from $\{0,1,2,3,4,\dots,1023\}$

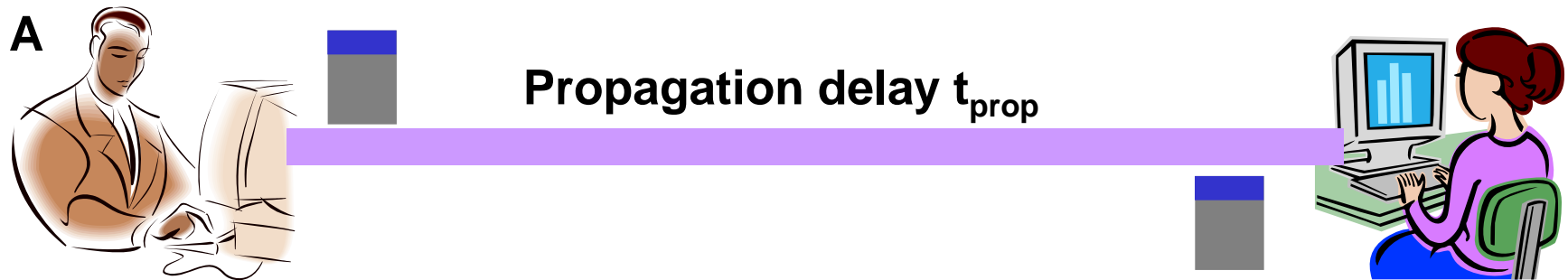
CSMA/CD efficiency

- T_{prop} = max prop between 2 nodes in LAN
- t_{trans} = time to transmit max-size frame

$$\text{efficiency} = \frac{1}{1 + 5t_{\text{prop}} / t_{\text{trans}}}$$

- Efficiency goes to 1 as t_{prop} goes to 0
- Goes to 1 as t_{trans} goes to infinity
- Much better than ALOHA, but still decentralized, simple, and cheap

Propagation delay & frame size



- Suppose A sends a packet at time 0
And B sees an idle channel just before $t = t_{prop}$
so B happily starts transmitting a frame
- B detects a collision, but A doesn't see collision till $t = 2 t_{prop}$
- A needs to wait for time $2 t_{prop}$ to detect collision
So, A should keep transmitting during this period
... and keep an eye out for a possible collision

Propagation delay & frame size

Propagation delay determines min. frame size to prevent undetected collisions

❑ Trans. time of minimum frame $> 2 t_{prop}$

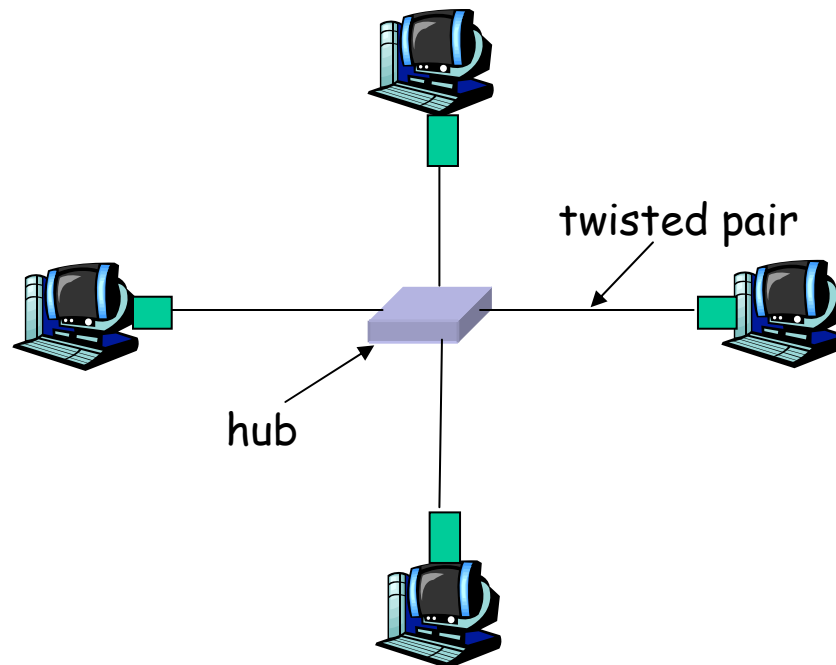
❑ Modern 10Mb Ethernet

○ Minimum frame size calculation

- 500m maximum segment length
- Can add repeaters up to a maximum 5 segments (2500m)
- c in cable = 60% * c in vacuum = 1.8×10^8 m/s
- ~ 12.5us one-way delay
- Add repeater and transceiver delay
- To be safe IEEE specifies a 512 “bit-time” slot for Ethernet = 51.2us
- 512 bits = 64 bytes (data payload = 46 bytes)

10BaseT and 100BaseT

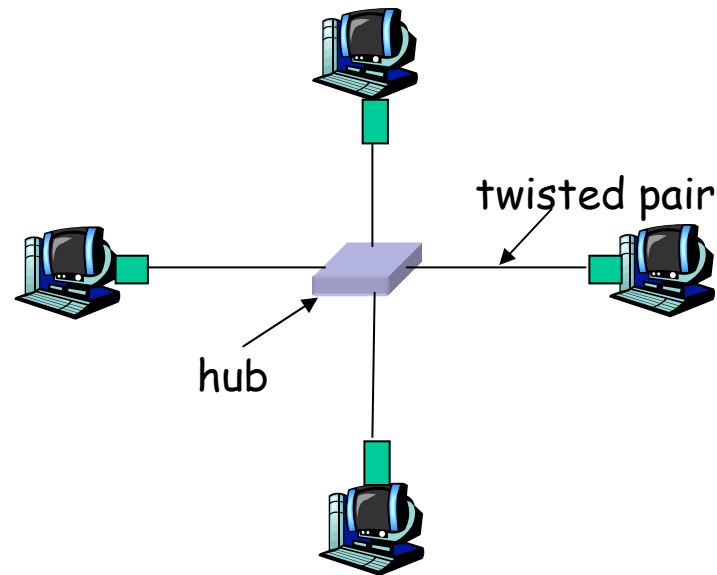
- ❑ 10/100 Mbps rate; latter called "fast ethernet"
- ❑ T stands for Twisted Pair
- ❑ Nodes connect to a hub: "star topology"; 100 m max distance between nodes and hub



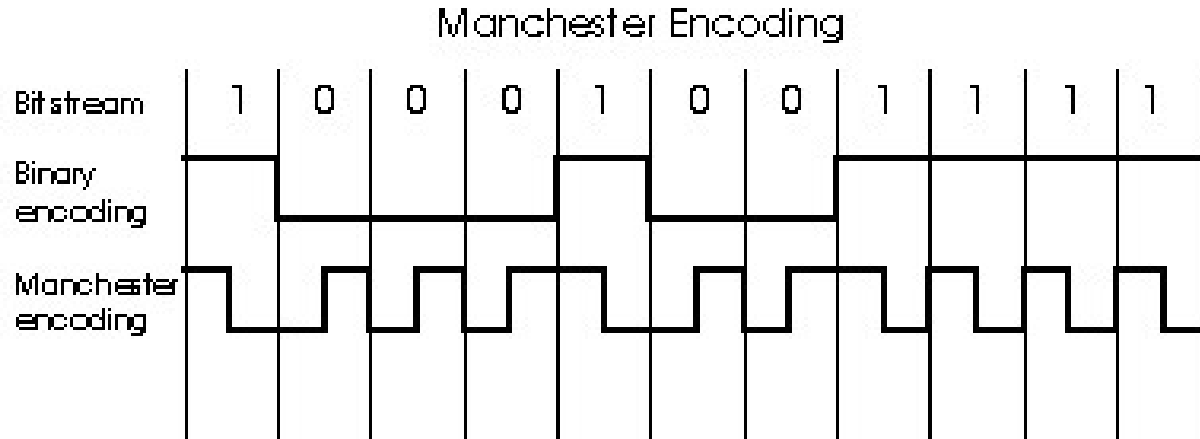
Hubs

Hubs are essentially physical-layer repeaters:

- bits coming from one link go out all other links
- at the same rate
- no frame buffering
- no CSMA/CD at hub: adapters detect collisions
- provides net management functionality



Manchester encoding



- ❑ Used in 10BaseT
- ❑ Each bit has a transition
- ❑ Allows clocks in sending and receiving nodes to synchronize to each other
 - no need for a centralized, global clock among nodes!
- ❑ Hey, this is physical-layer stuff!

Gbit Ethernet

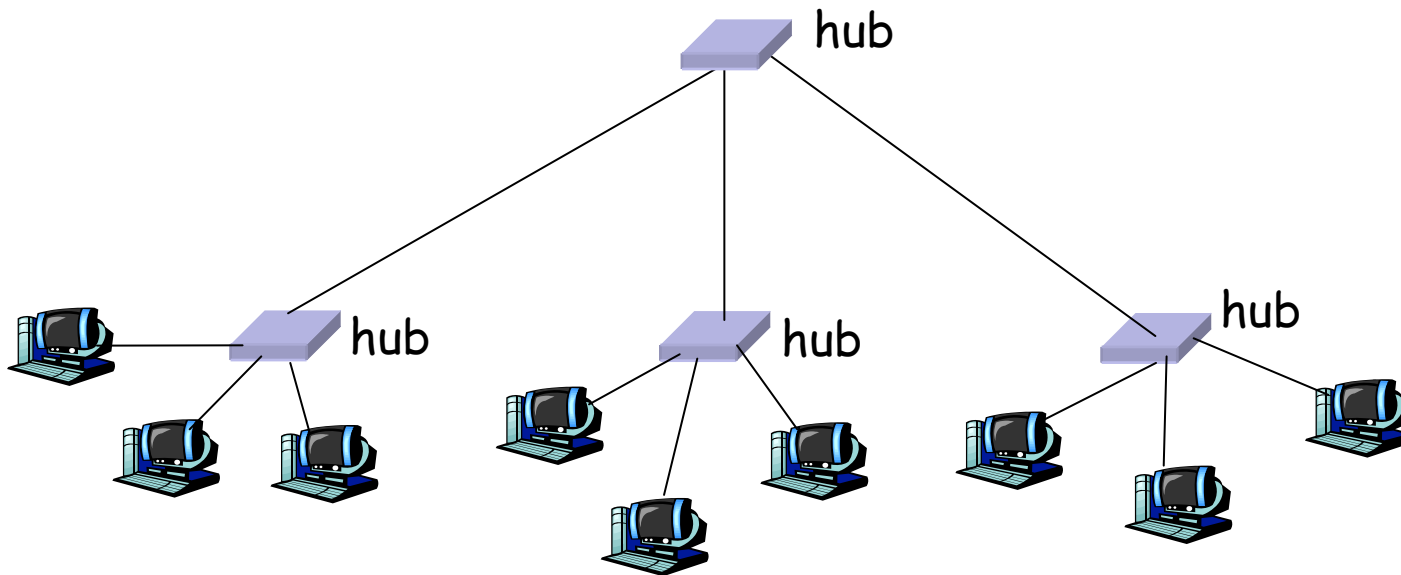
- ❑ uses standard Ethernet frame format
- ❑ allows for point-to-point links and shared broadcast channels
- ❑ in shared mode, CSMA/CD is used; short distances between nodes required for efficiency
- ❑ uses hubs, called here "Buffered Distributors"
- ❑ Full-Duplex at 1 Gbps for point-to-point links
- ❑ 10 Gbps now !

Link Layer

- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Interconnections: Hubs and switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM

Interconnecting with hubs

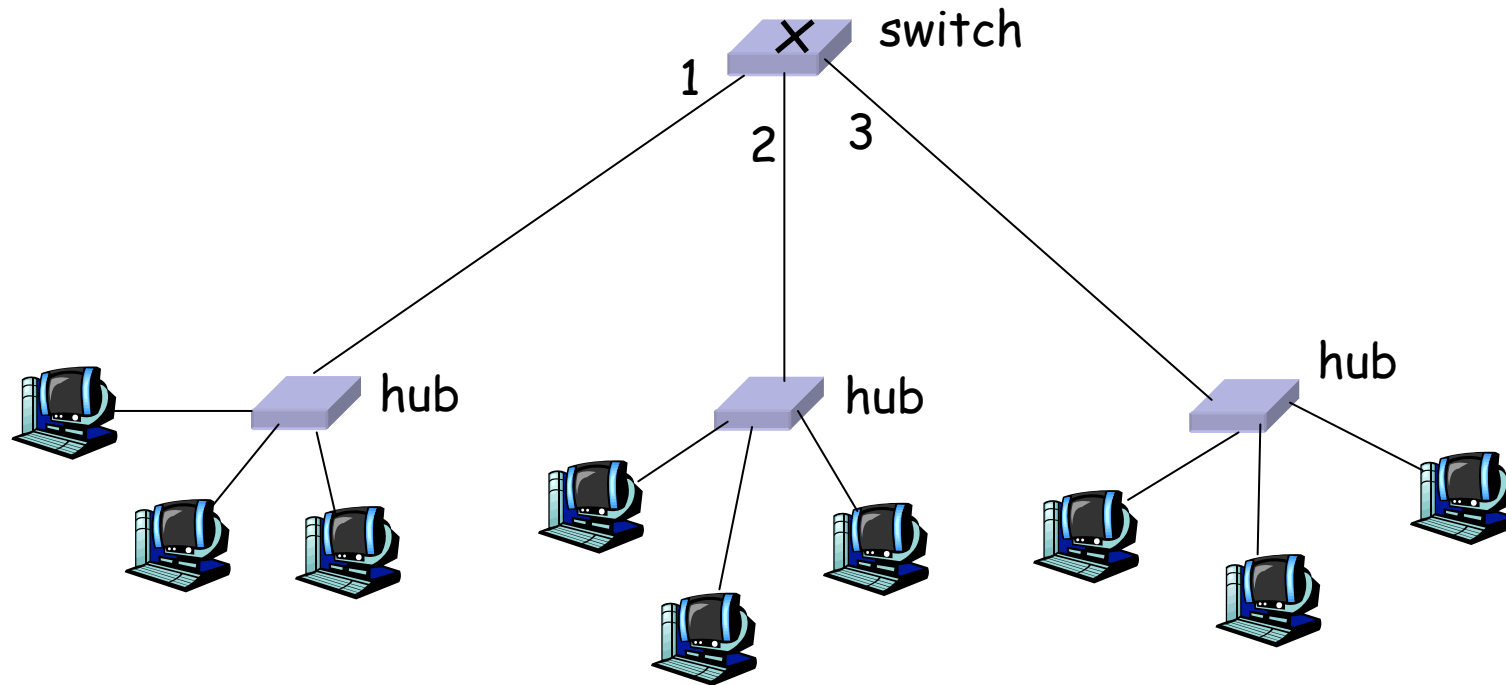
- ❑ Backbone hub interconnects LAN segments
- ❑ Extends max distance between nodes
- ❑ But individual segment collision domains become one large collision domain
- ❑ Can't interconnect 10BaseT & 100BaseT



Switch

- ❑ Link layer device
 - stores and forwards Ethernet frames
 - examines frame header and **selectively** forwards frame based on MAC dest address
 - when frame is to be forwarded on segment, uses CSMA/CD to access segment
- ❑ transparent
 - hosts are unaware of presence of switches
- ❑ plug-and-play, self-learning
 - switches do not need to be configured

Forwarding



- How do determine onto which LAN segment to forward frame?
- Looks like a routing problem...

Self learning

- ❑ A switch has a **switch table**
- ❑ entry in switch table:
 - (MAC Address, Interface, Time Stamp)
 - stale entries in table dropped (TTL can be 60 min)
- ❑ switch **learns** which hosts can be reached through which interfaces
 - when frame received, switch “learns” location of sender: incoming LAN segment
 - records sender/location pair in switch table

Filtering/Forwarding

When switch receives a frame:

index switch table using MAC dest address

if entry found for destination
then{

if dest on segment from which frame arrived
 then drop the frame

else forward the frame on interface indicated

}

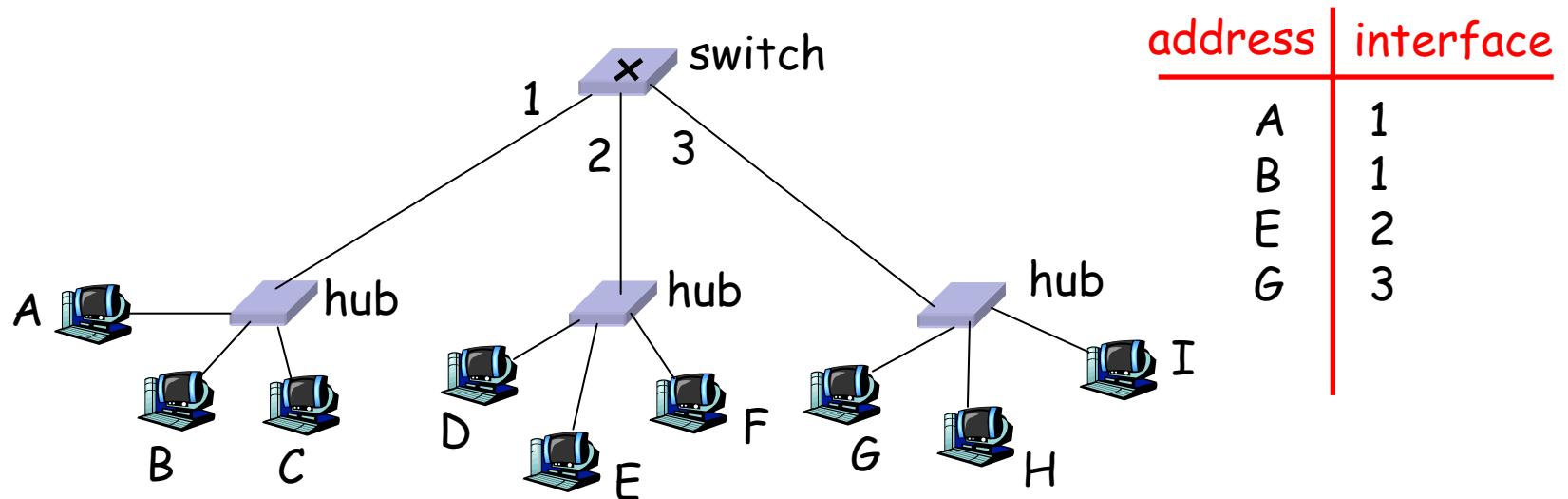
else flood



*forward on all but the interface
on which the frame arrived*

Switch example

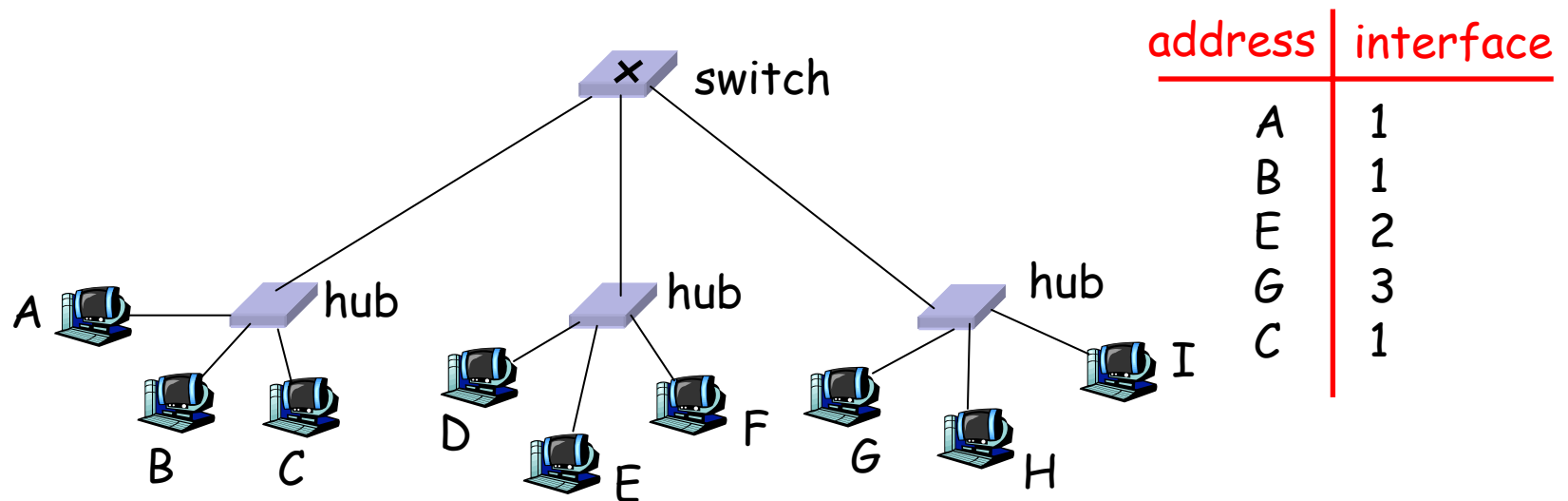
Suppose C sends frame to D



- ❑ Switch receives frame from from C
 - entry in switch table that C is on interface 1
 - because D is not in table, switch forwards frame into interfaces 2 and 3
- ❑ frame received by D

Switch example

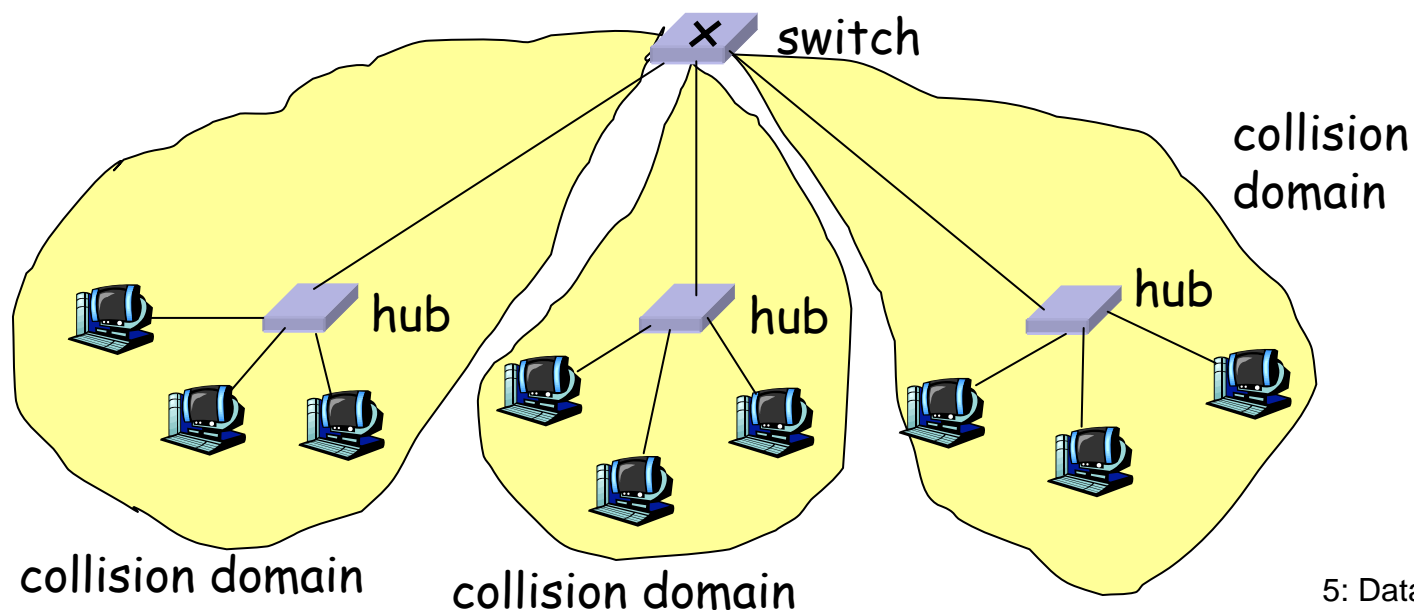
Suppose D replies back with frame to C.



- ❑ Switch receives frame from from D
 - Entry in switch table says that D is on interface 2
 - because C is in table, switch forwards frame only to interface 1
- ❑ frame received by C

Switch: traffic isolation

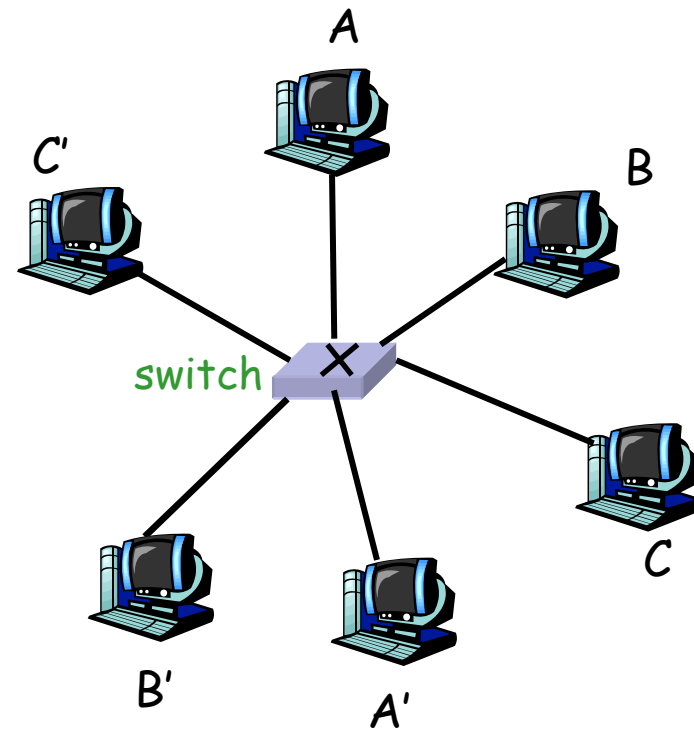
- ❑ switch installation breaks subnet into LAN segments
- ❑ switch **filters** packets:
 - same-LAN-segment frames not usually forwarded onto other LAN segments
 - segments become separate **collision domains**



Switches: dedicated access

- ❑ Switch with many interfaces
- ❑ Hosts have direct connection to switch
- ❑ No collisions; full duplex

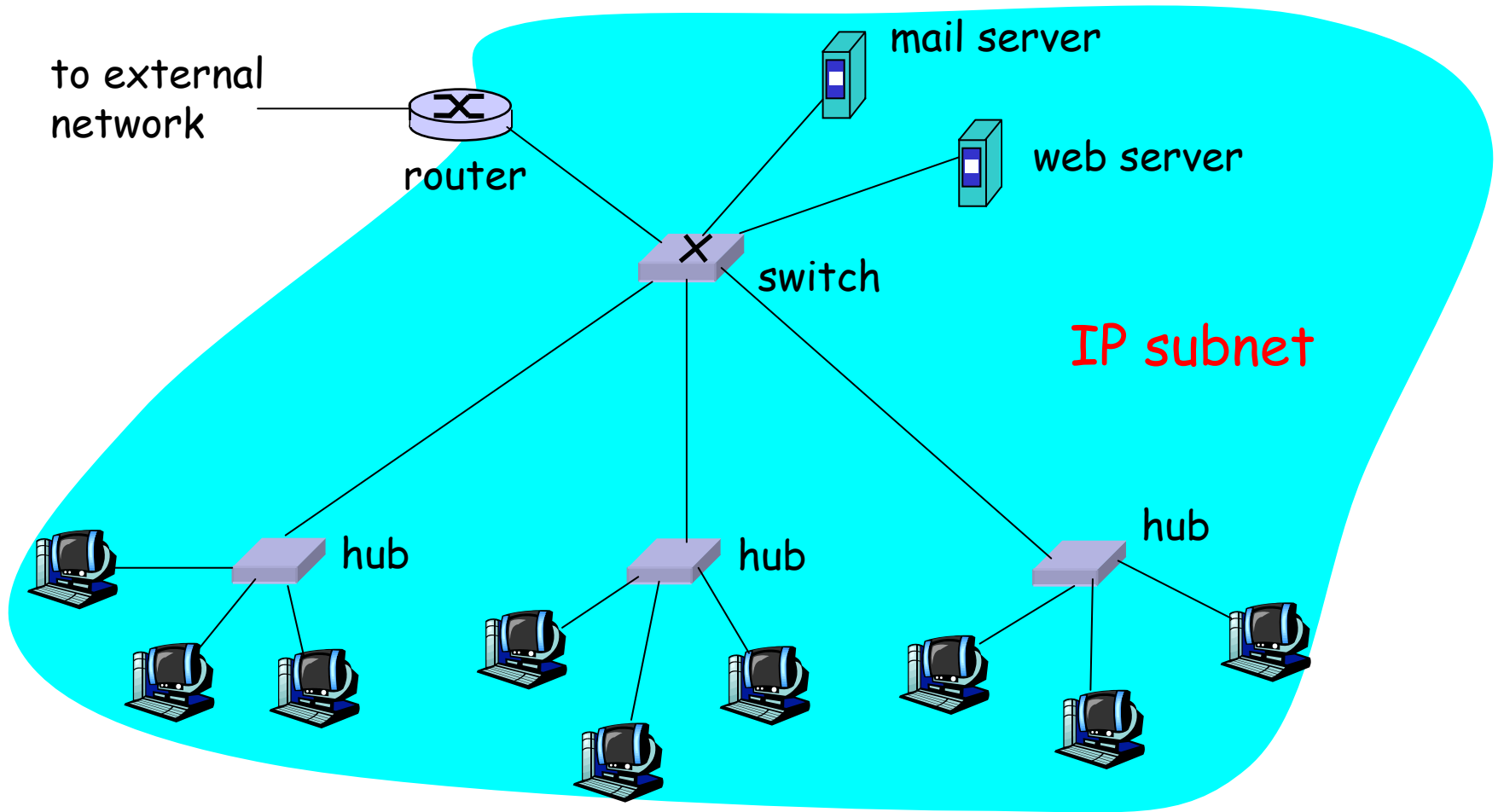
Switching: A-to-A' and B-to-B' simultaneously, no collisions



More on Switches

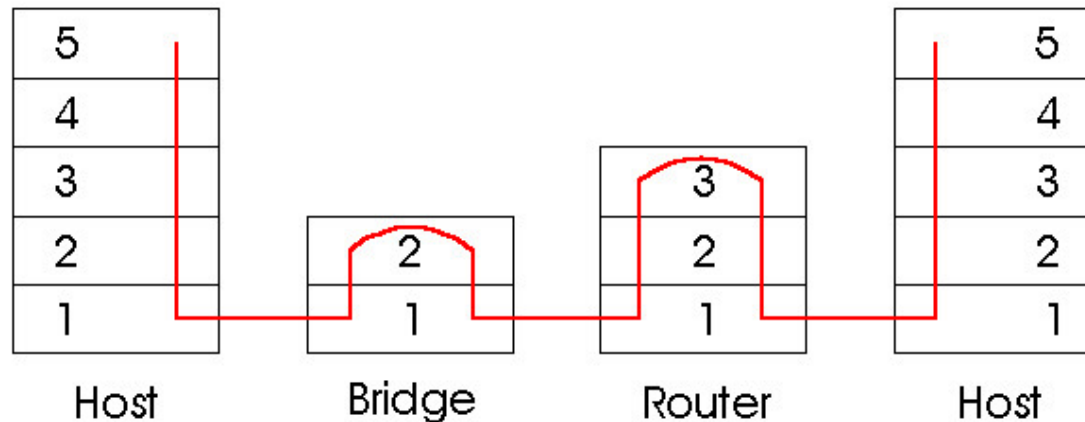
- ❑ **cut-through switching:** frame forwarded from input to output port without first collecting entire frame
 - slight reduction in latency
- ❑ combinations of shared/dedicated, 10/100/1000 Mbps interfaces

Institutional network



Switches vs. Routers

- ❑ both store-and-forward devices
 - routers: network layer devices (examine network layer headers)
 - switches are link layer devices
- ❑ routers maintain routing tables, implement routing algorithms
- ❑ switches maintain switch tables, implement filtering, learning algorithms



Switches vs. Routers (more)

□ switches

- Pros: plug-and-play; relatively high packet filtering and forwarding rates
- Cons: topology restricted to a spanning tree; large switched networks require large ARP tables; no protection against broadcast storms

□ Routers

- Pros: No spanning tree restriction; intelligent routing; firewall protection against broadcast storms
- Cons: not plug-and-play; larger per-packet processing time

Summary comparison

	<u>hubs</u>	<u>routers</u>	<u>switches</u>
traffic isolation	no	yes	yes
plug & play	yes	no	yes
optimal routing	no	yes	no
cut through	yes	no	yes