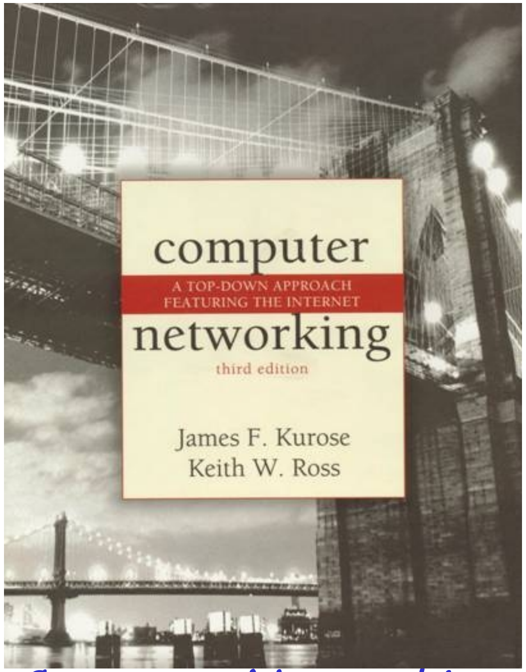
Chapter 5
Data Link Layer



Computer Networking: A Top Down Approach Featuring the Internet,

3rd edition.
Jim Kurose, Keith Ross
Addison-Wesley, July 2004.

A note on the use of these ppt slides:

We're making these slides freely available to all (faculty, students, readers). They're in PowerPoint form so you can add, modify, and delete slides (including this one) and slide content to suit your needs. They obviously represent a *lot* of work on our part. In return for use, we only ask the following:

- If you use these slides (e.g., in a class) in substantially unaltered form, that you mention their source (after all, we'd like people to use our book!)
- If you post any slides in substantially unaltered form on a www site, that you note that they are adapted from (or perhaps identical to) our slides, and note our copyright of this material.

Thanks and enjoy! JFK/KWR

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

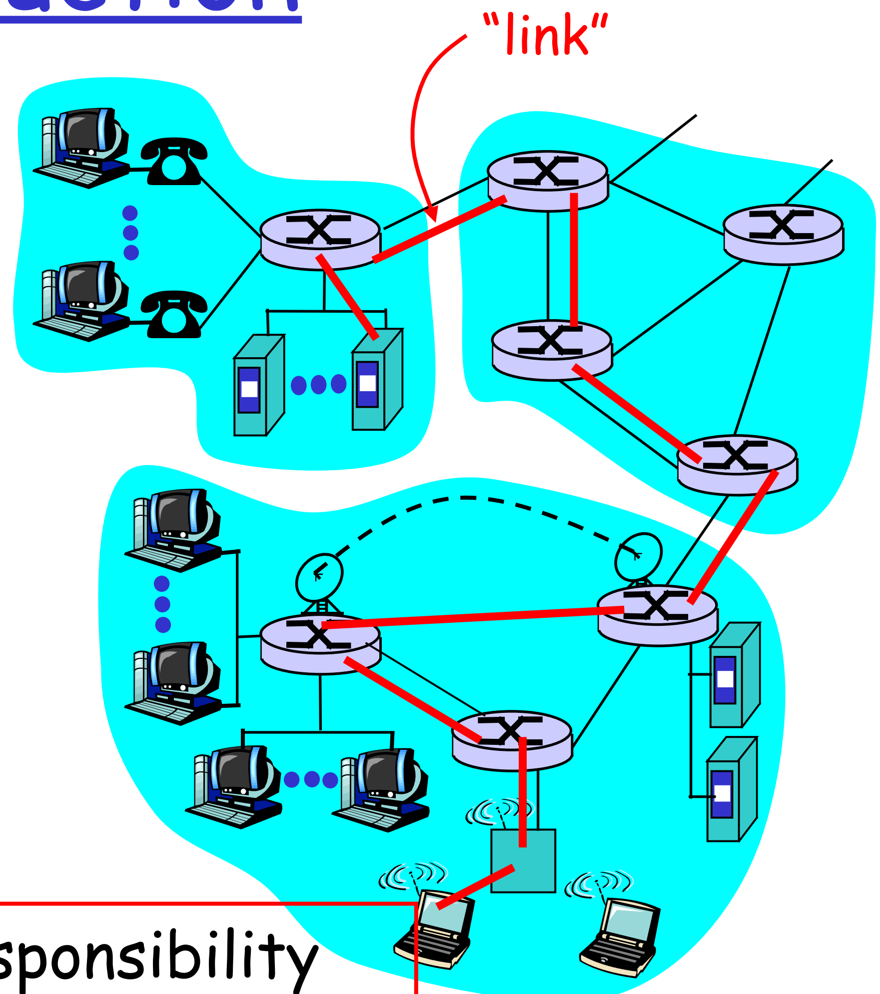
Chapter 5 outline

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

Link Layer: Introduction

Some terminology:

- hosts and routers are **nodes** (bridges and switches too)
- communication channels that connect adjacent nodes along communication path are **links**
  - wired links
  - wireless links
  - LANs
- Link-layer PDU is a **frame**, encapsulates a network-layer datagram



Link-layer protocol has the 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 reliable data transfer 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

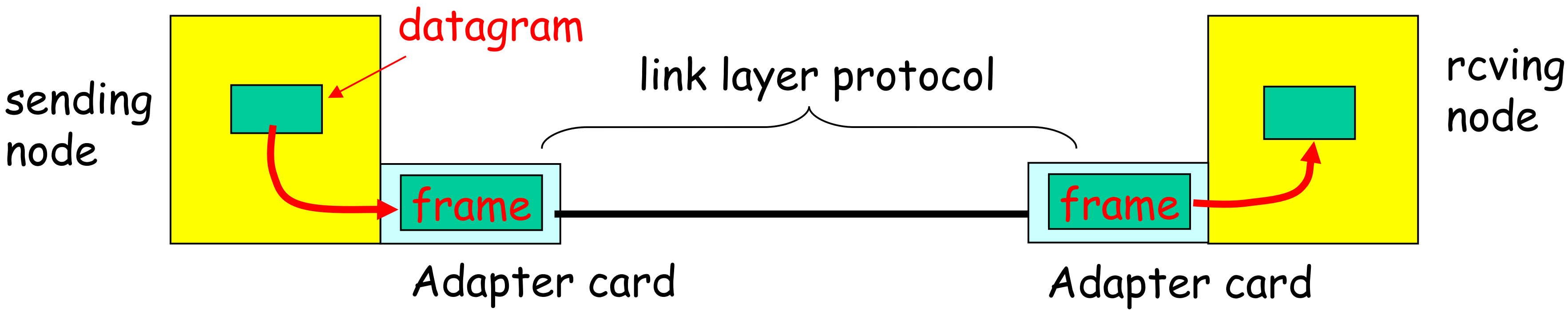
Link Layer Services

- Framing:
  - encapsulate datagram into frame, adding header, trailer
  - 'physical addresses' used in frame headers to identify source, destination
    - different from IP address!
- Link access
  - Media access control (MAC) protocol
  - Coordinate the frame transmissions of many nodes if multiple nodes share a medium
- 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)
  - Used on wireless links: high error rates
    - Correct an error locally at link level

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, PCMCI 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 receiving node
- adapter is semi-autonomous
- link & physical layers

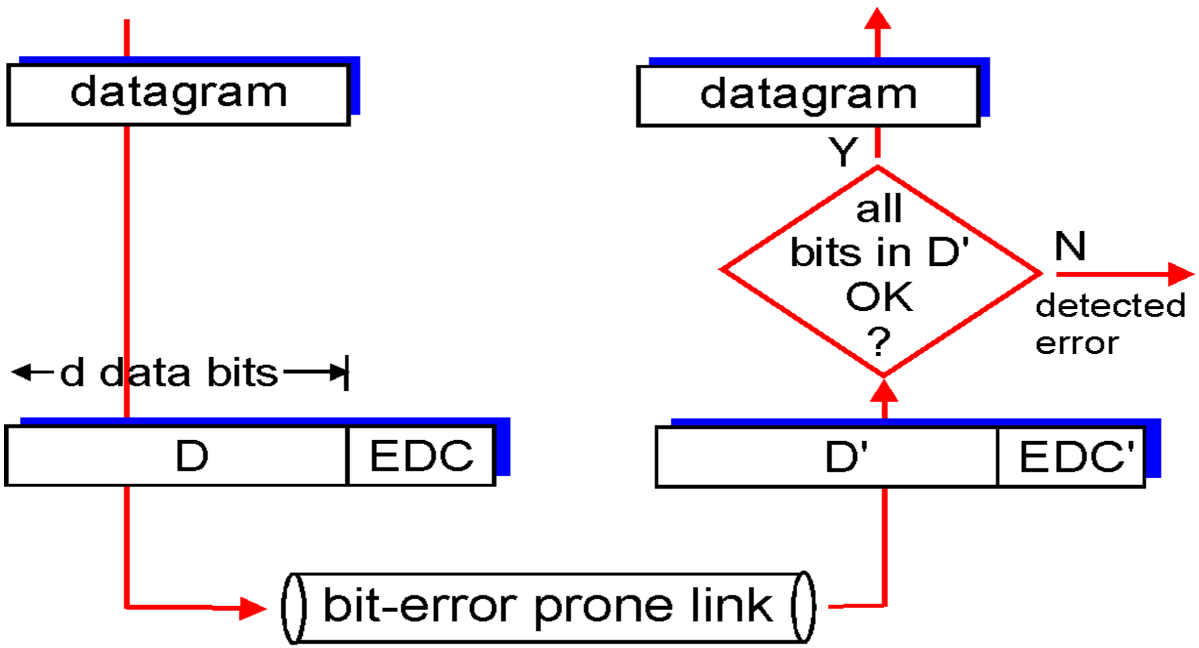
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple 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

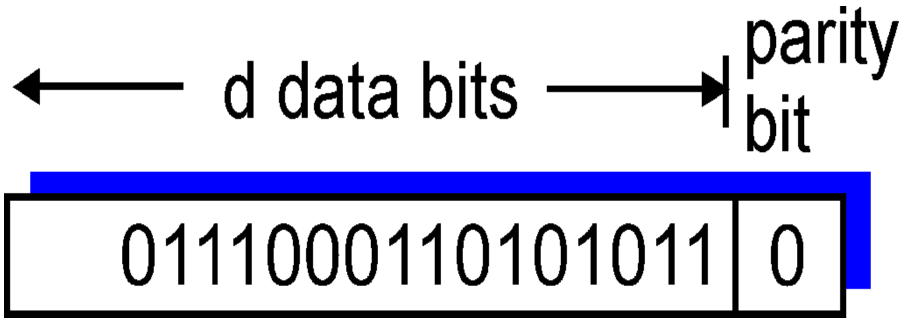


Techniques for Error Detection

- Parity checks
- Checksumming methods
- Cyclic redundancy checks

Parity Checks

Single Bit Parity: **Detect single bit errors**

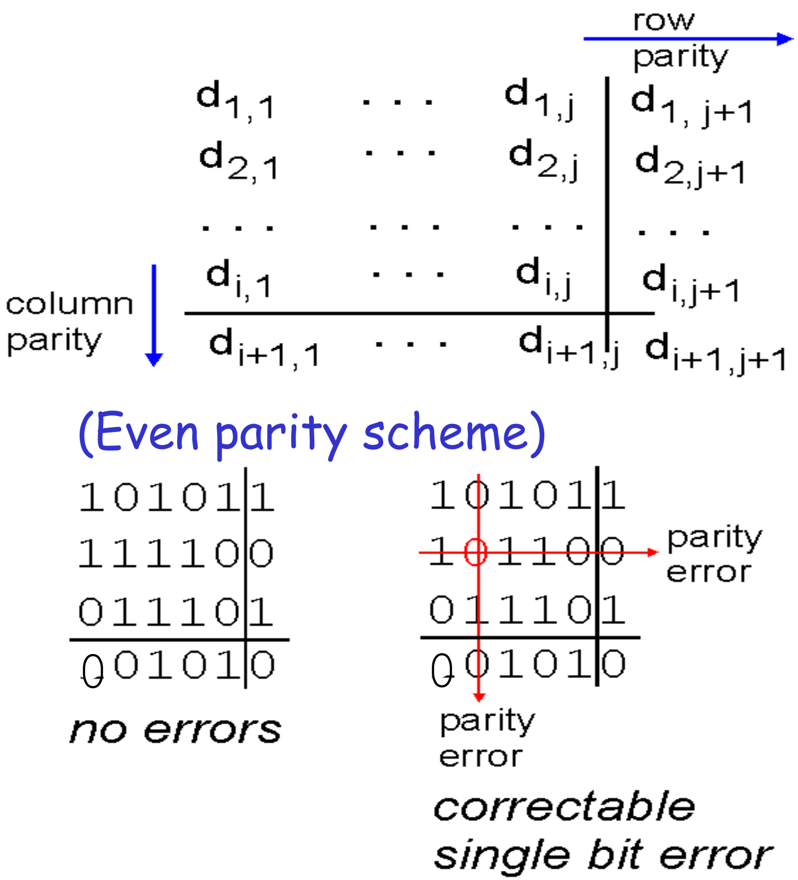


- Even parity scheme: choose the value of the parity bit such that the total number of 1s in the d+1 bits is even

- Odd parity scheme: choose the value of the parity bit such that the total number of 1s in the d+1 bits is odd

Parity Checks (Cont.)

Two Dimensional Bit Parity: **Detect *and correct* single bit errors**



Checksumming Methods

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

Internet checksum:

Sender:

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

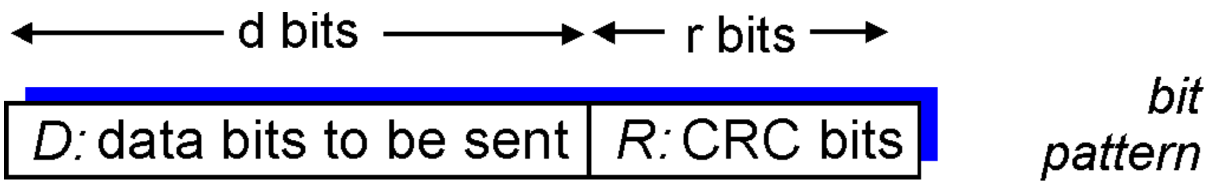
Receiver:

- 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 ….

- Checksum is easy and fast to compute
- Typically used in software implemented protocols (e.g. ,TCP and UDP)

Cyclic Redundancy Check

- view data bits, D, as a binary number
- choose r+1 bit pattern (generator), G (both sender and receiver know G)
- sender chooses r CRC bits, R, such that
  - <D,R> exactly divisible by G (modulo 2)
- receiver knows G, divides <D,R> by G.
  - If non-zero remainder: error detected!
  - can detect all burst errors less than r+1 bits
- widely used in practice (ATM, HDLC)



$$\underbrace{D * 2^r}_{\text{Left shifts r bits}} \quad \text{XOR} \quad R$$

mathematical formula

CRC Example

Want to find R such that:

$D \cdot 2^r$ XOR R = nG

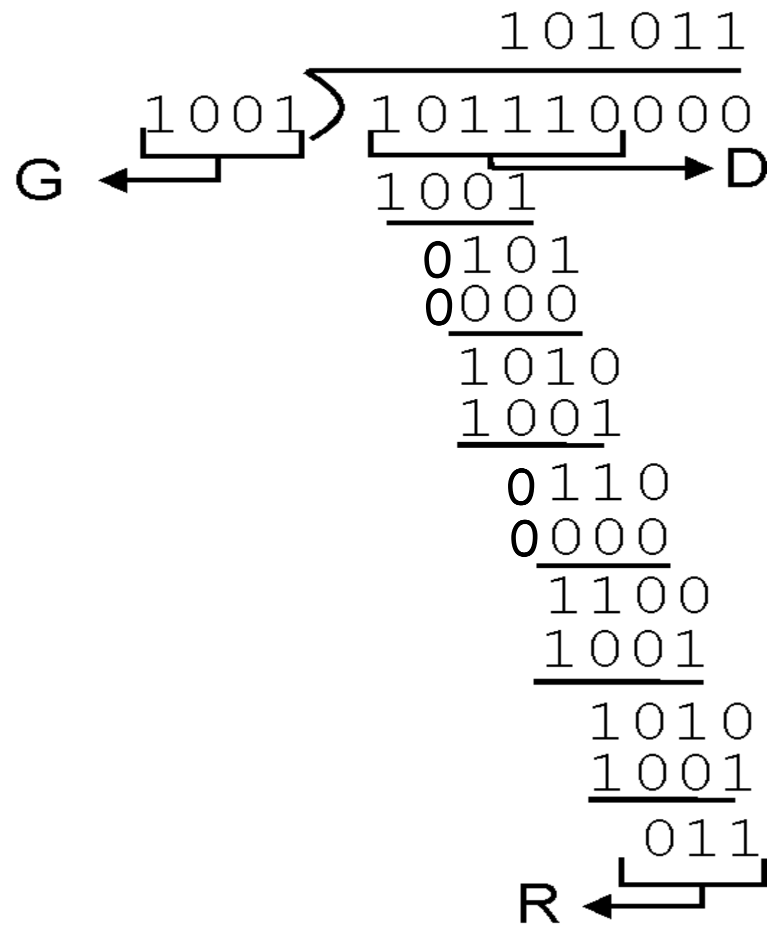
XOR R to the right of both sides :

$D \cdot 2^r$ = nG XOR R

equivalently:

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

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



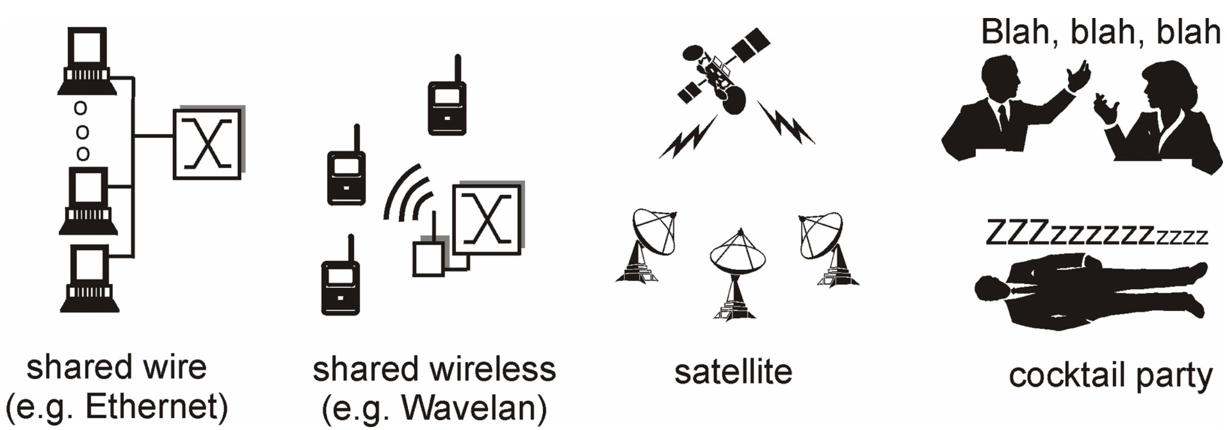
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3 Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay

Multiple Access Links and Protocols

Two types of "links":

- point-to-point
  - PPP (point-to-point protocol) for dial-up access
  - point-to-point link between Ethernet switch and host
- broadcast (shared wire or medium)
  - traditional Ethernet
  - upstream HFC (Hybrid fiber coaxial cable)
  - 802.11 wireless LAN



Multiple Access protocols

- single shared broadcast channel
- two or more simultaneous transmissions by nodes: interference
  - only one node can send successfully at a 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 Mulitple Access Protocol

What to look for in multiple access protocols?

Broadcast channel of rate R bps

1. When 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

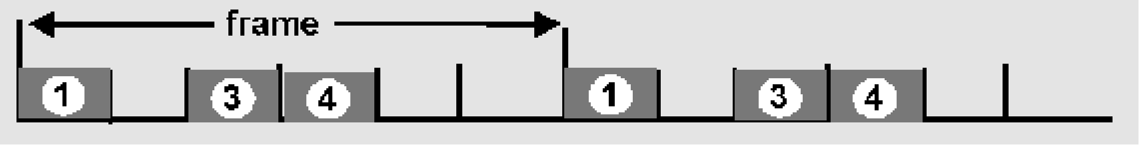
Three broad classes:

- Channel Partitioning protocols
  - divide channel into smaller "pieces" (time slots, frequency, code)
  - allocate piece to node for exclusive use
- Random Access protocols
  - channel not divided, allow collisions
  - "recover" from collisions
- Taking-turns protocols
  - tightly coordinate shared access to avoid collisions

Channel Partitioning MAC protocols: TDMA

TDMA: time division multiple access

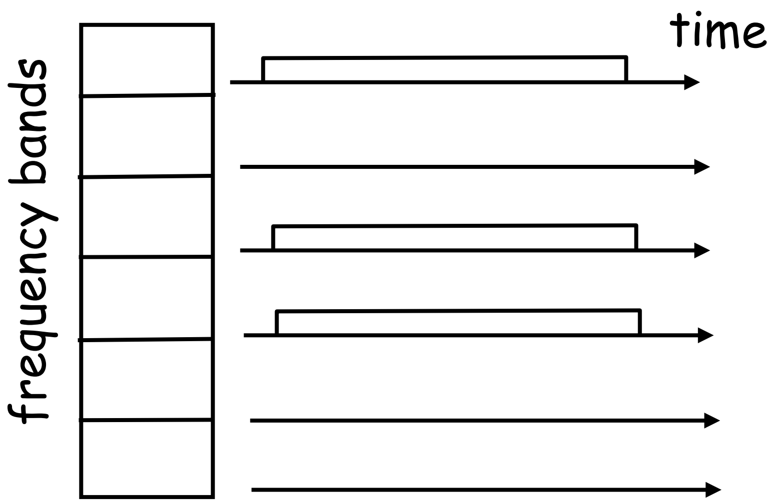
- channel divided into N time slots, one per user
- access to channel in "rounds"
- each station gets fixed length slot (length = packet trans time) in each round
- unused slots go idle
- inefficient with low duty cycle users and at light load
- example: 6-station LAN, 1,3,4 have packets, slots 2,5,6 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 packets, 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

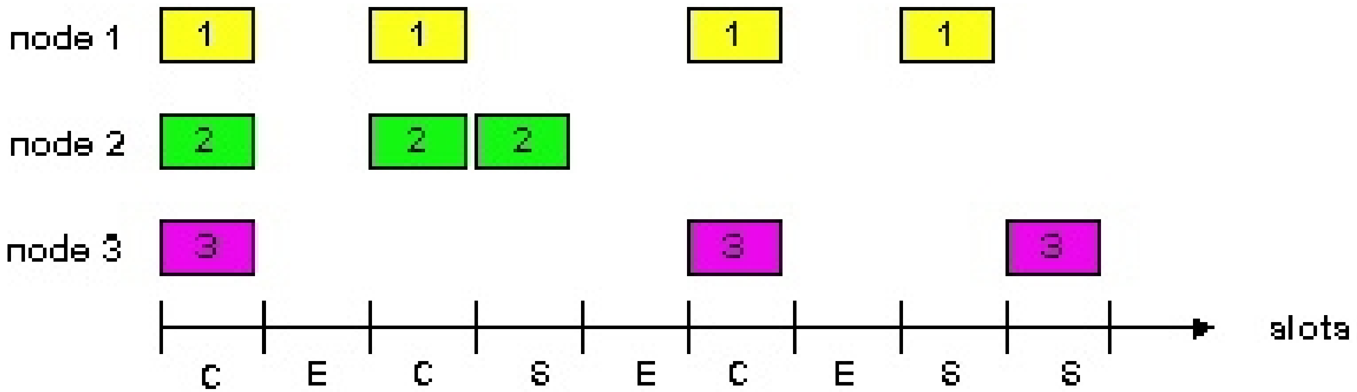
Assumptions

- all frames same size
- time is divided into equal size slots (length of a slot equals 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 a slot, all nodes detect collision

Operation

- when a node has a fresh frame to send , it transmits in the next slot
- If no collision, the frame is transmitted successfully
- if collision, the node retransmits the frame in each subsequent slot with probability 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 due to probabilistic retransmission
- nodes may be able to detect collision in a time interval of length less than the time to transmit a packet

Slotted Aloha efficiency

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

To derive the maximum efficiency

- Modified protocol: each node attempts to transmit a fresh frame in each slot with probability p
- Suppose N nodes with many frames to send
- Probability that 1st node has success in a slot = $p(1-p)^{N-1}$
- Probability that any node has a success = $Np(1-p)^{N-1}$

Slotted Aloha efficiency (Cont.)

- 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

$$E'(p) = 0 \Rightarrow p^* = \frac{1}{N} \quad , \quad E(p^*) = N\frac{1}{N}(1-\frac{1}{N})^{N-1} = (1-\frac{1}{N})^{N-1} = \frac{(1-\frac{1}{N})^N}{1-\frac{1}{N}}$$

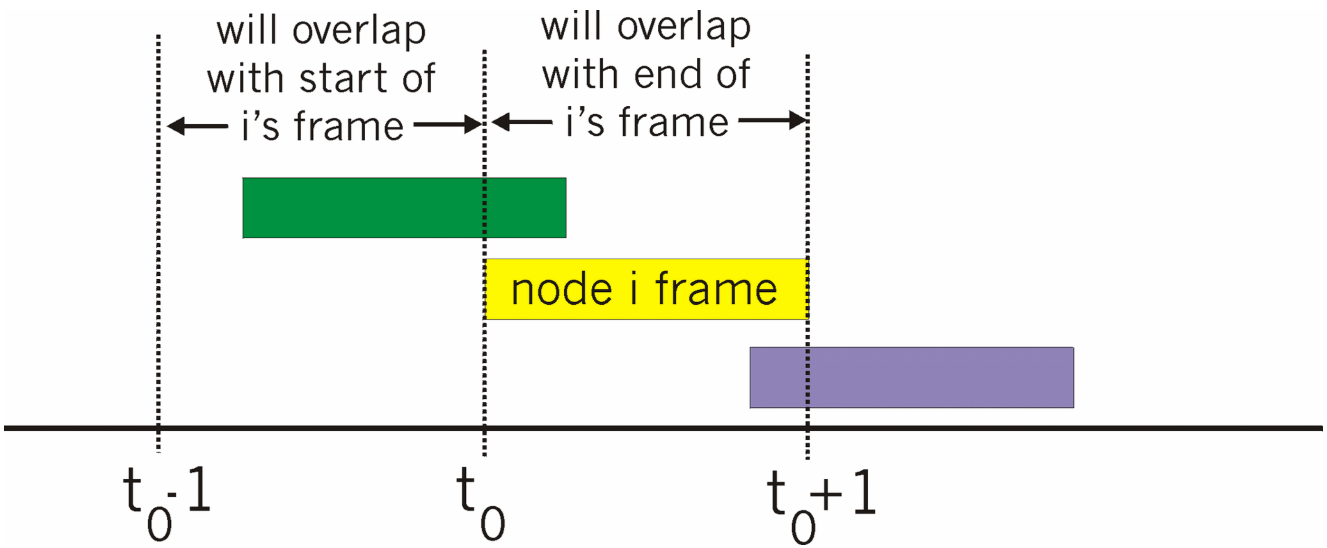
$$\lim_{N\to\infty}(1-\frac{1}{N}) = 1 \quad , \quad \lim_{N\to\infty}(1-\frac{1}{N})^N = \frac{1}{e}$$

$$\lim_{N\to\infty} E(p^*) = \frac{1}{e}$$

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

Pure (unslotted) ALOHA

- unslotted Aloha: simpler, no synchronization
- when frame first arrives
  - transmit immediately
  - If collision, retransmits with probability p, or waits for another frame With probability 1-p
- collision probability increases:
  - frame sent at t_0 collides with other frames sent in $[t_0-1,t_0+1]$



Pure Aloha efficiency

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

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

P(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 -> infinity ...

maximum efficiency = 1/(2e) = .18

Even worse !

CSMA (Carrier Sense Multiple Access)

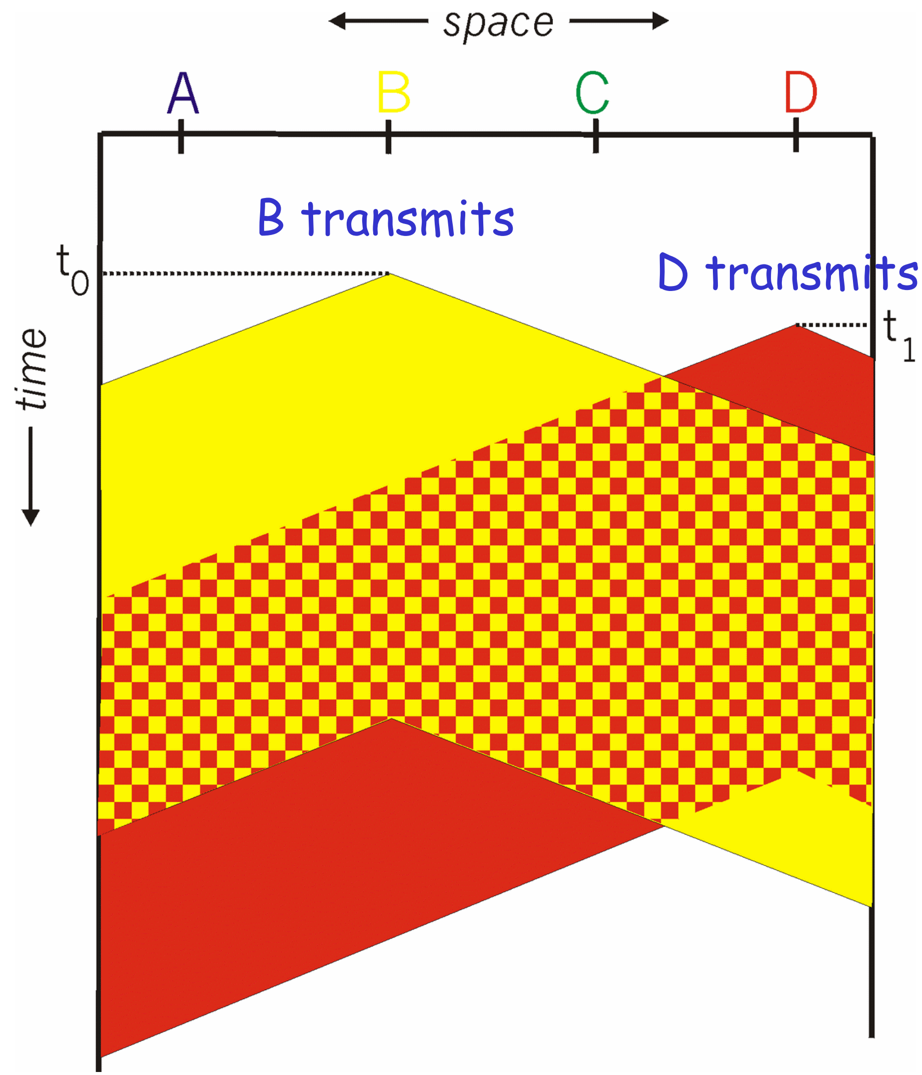
CSMA: listen before transmit:

- If channel sensed idle: transmit entire frame
- If channel sensed busy, defer transmission for a random amount of time

- Human analogy: don't interrupt others!

CSMA collisions

spatial layout of nodes



collisions *can* still occur:
propagation delay means two nodes may not hear each other's transmission

collision:
entire packet transmission time wasted

note:
The larger the end-to-end propagation delay, the larger the chance that a node is not able to sense a transmission that has already begun at another node

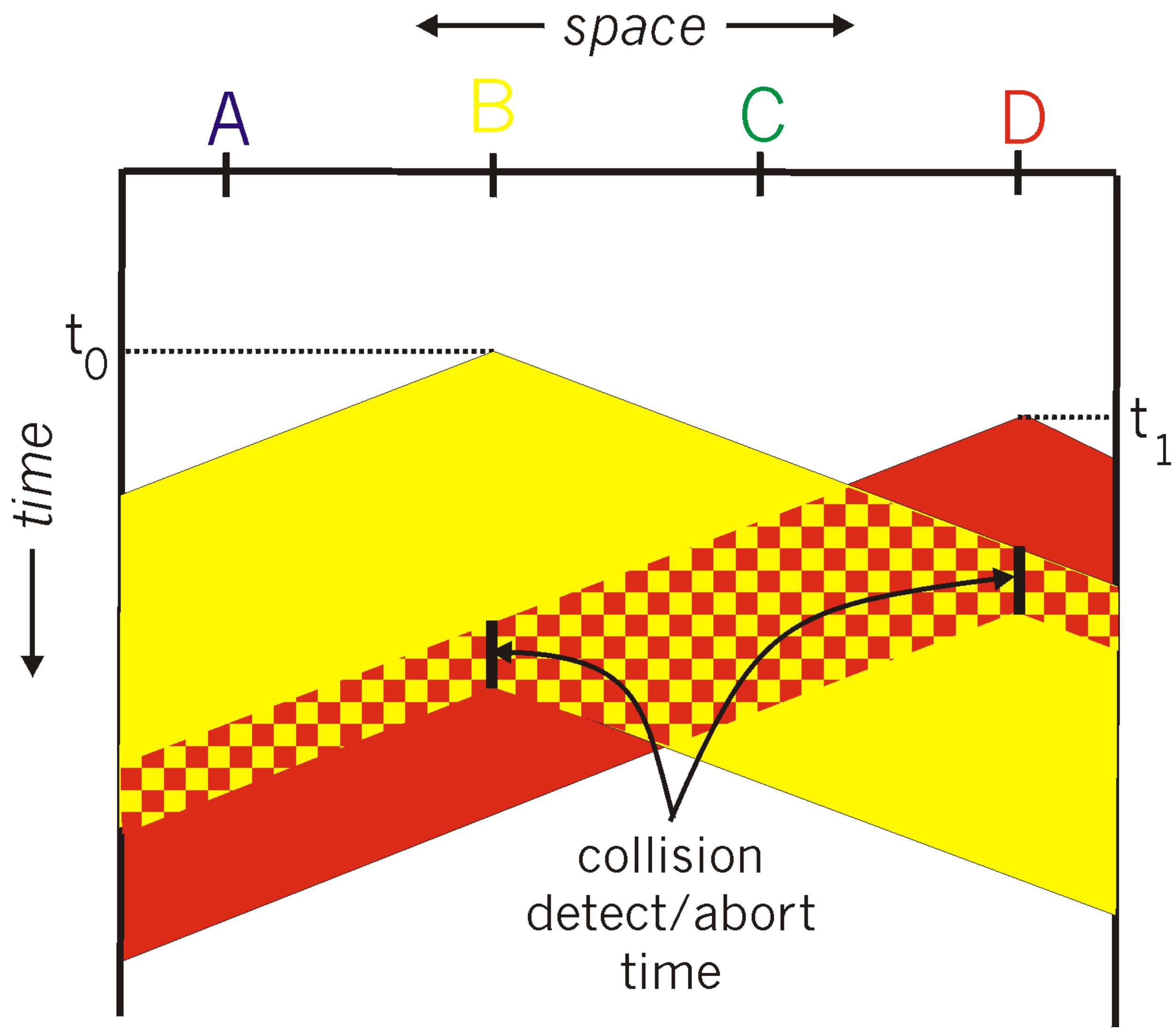
CSMA/CD (Collision Detection)

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 and received signals
  - difficult in wireless LANs: receiver shut off while transmitting; i.e., cannot transmit and receive at the same time
- 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: 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!

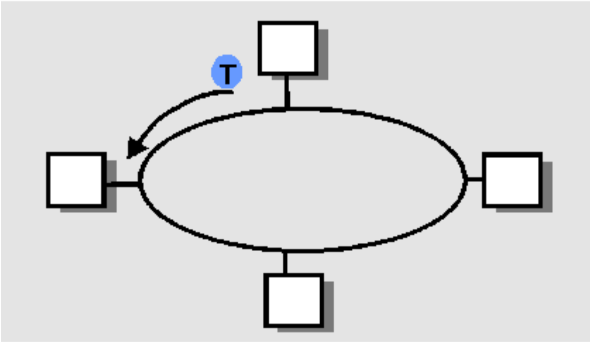
"Taking Turns" MAC protocols

Polling:

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

Token passing:

- control **token** passed from one node to next sequentially.
- When a node receives a token, it can transmits up to a maximum number of frames
- 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, Code 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
  - 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
- 5.3Multiple access protocols
- 5.4 Link-Layer Addressing
- 5.5 Ethernet
- 5.6 Hubs and switches
- 5.7 PPP
- 5.8 Link Virtualization: ATM

LAN Addresses and ARP

32-bit IP address:

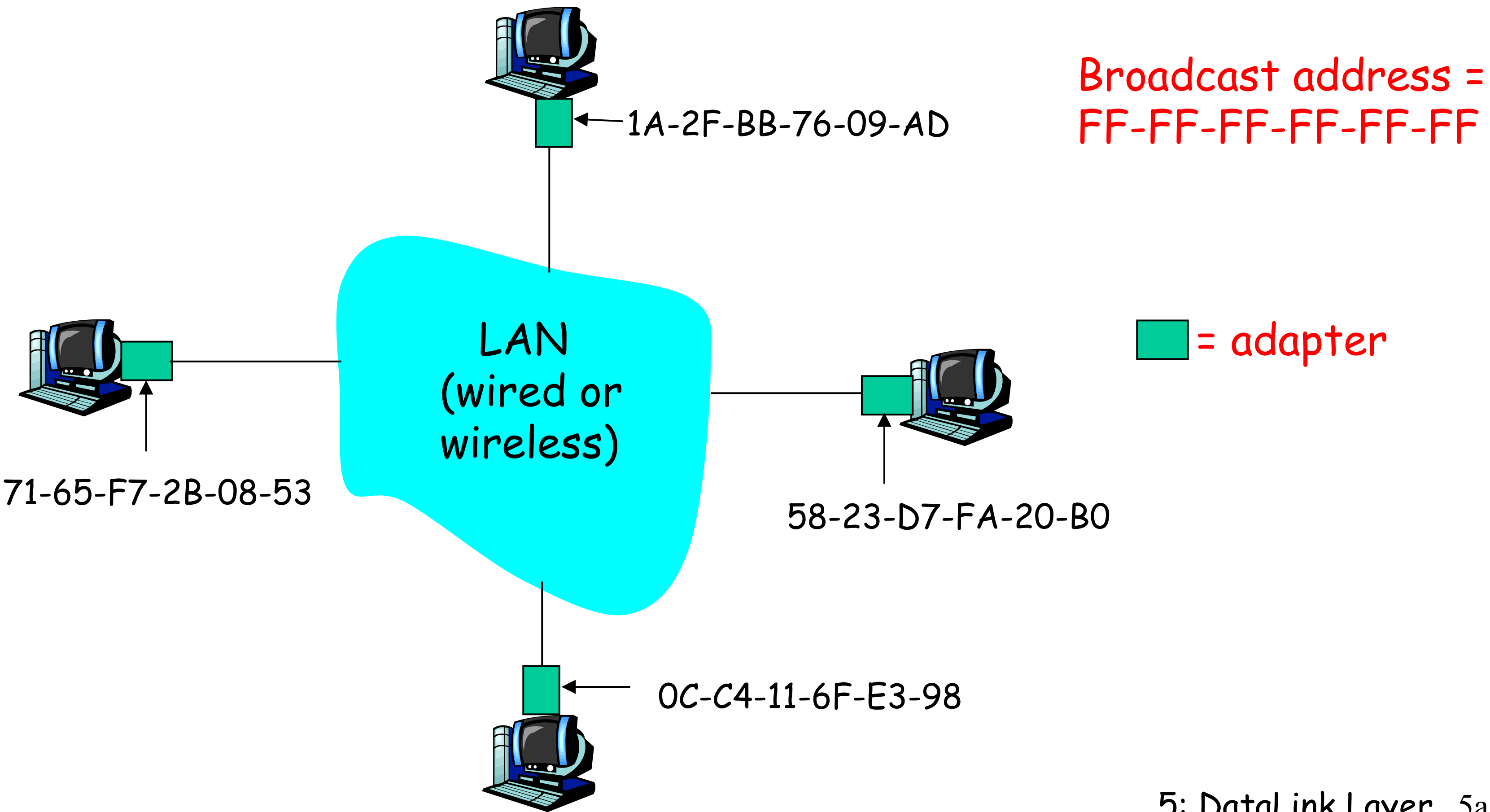
- *network-layer* address
- used to get datagram to destination IP network (recall IP network definition)

LAN (or MAC or physical or Ethernet) address:

- used to get datagram 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

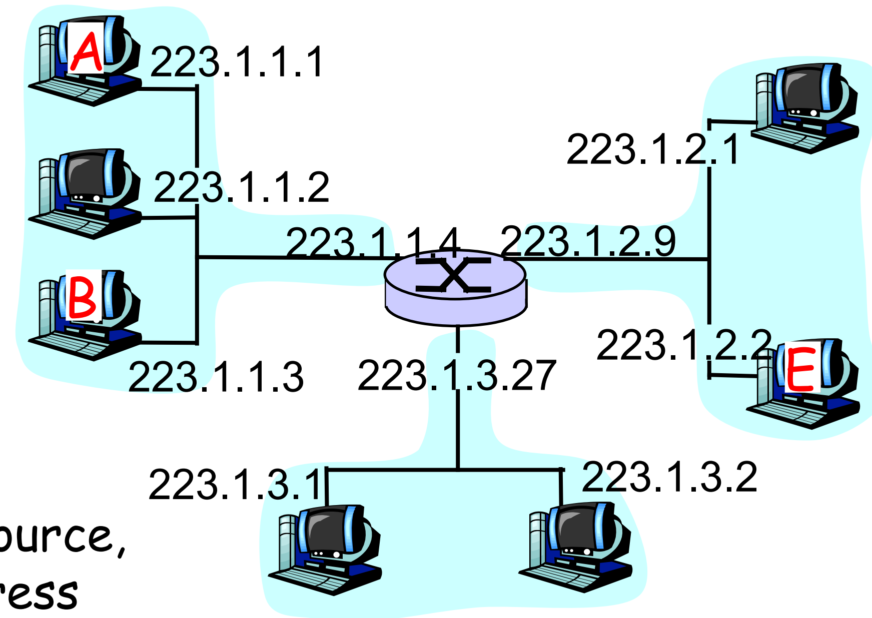
- Each adapter on LAN has unique LAN address
- Six bytes
- Expressed in hexadecimal notation



LAN Address (more)

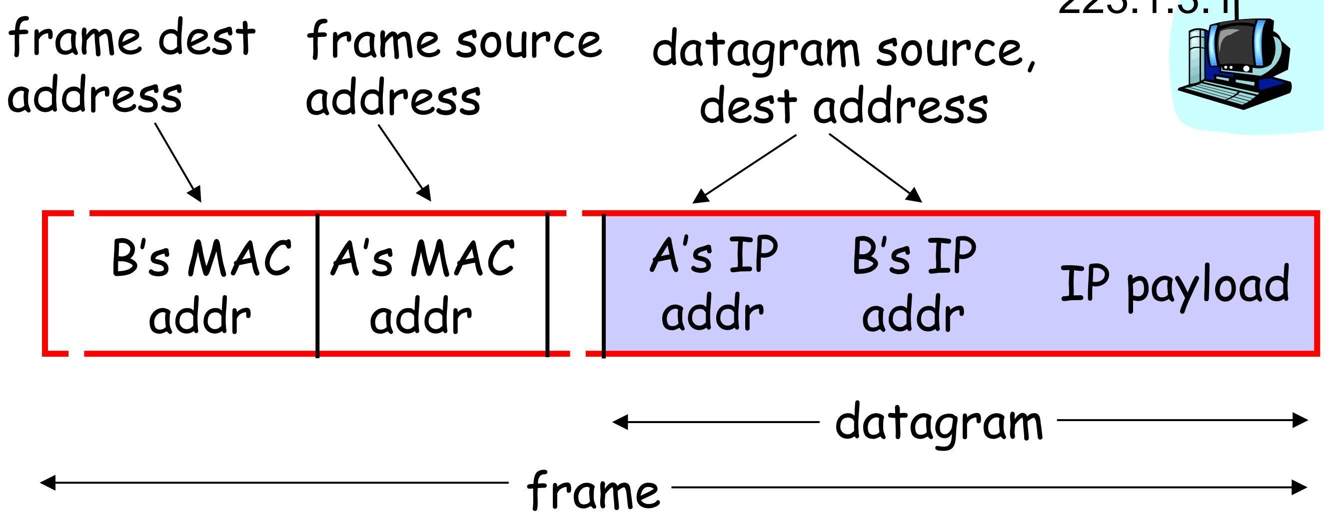
- MAC address allocation administered by IEEE
- manufacturer buys portion of MAC address space (to assure uniqueness)
- Analogy:
  - (a) MAC address: like Social Security Number
  - (b) IP address: like postal address
- MAC flat address => portability
  - MAC address of an adapter card does not change when it is moved from one LAN to another
- IP hierarchical address NOT portable
  - depends on IP network to which node is attached

Recall earlier routing discussion



Starting at A, given IP datagram addressed to B:

- look up network address of B, find B on same network as A
- link layer send datagram to B inside link-layer frame



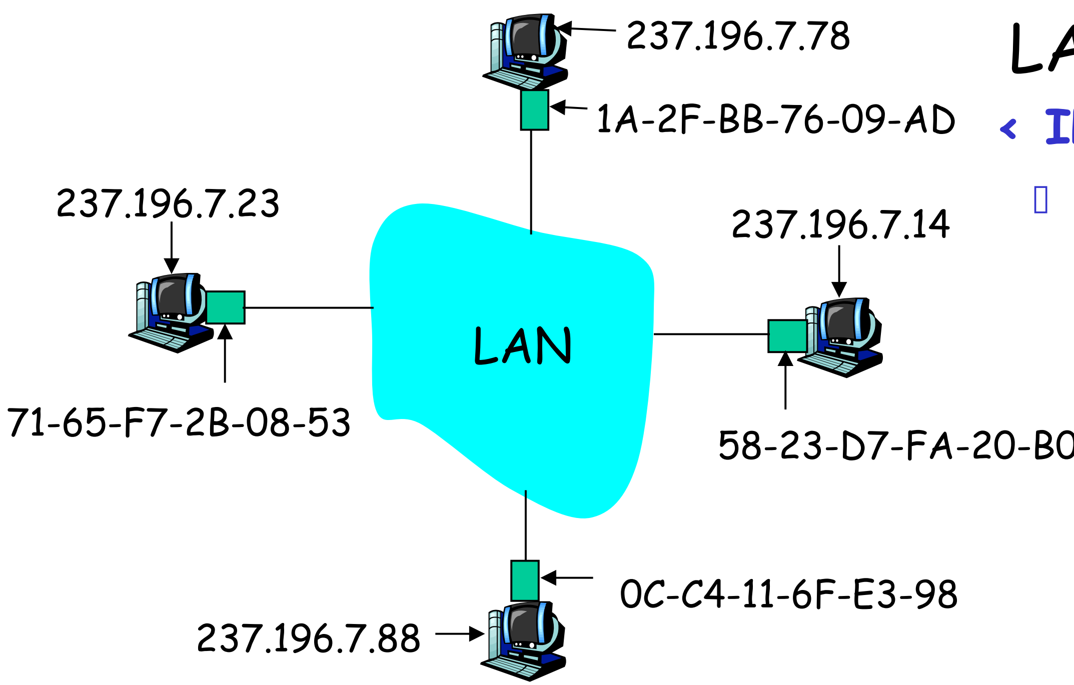
ARP: Address Resolution Protocol

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

- Each IP node (Host, Router) on LAN has an ARP table
- ARP Table: IP/MAC address mappings for some LAN nodes

  < IP address; MAC address; TTL>

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

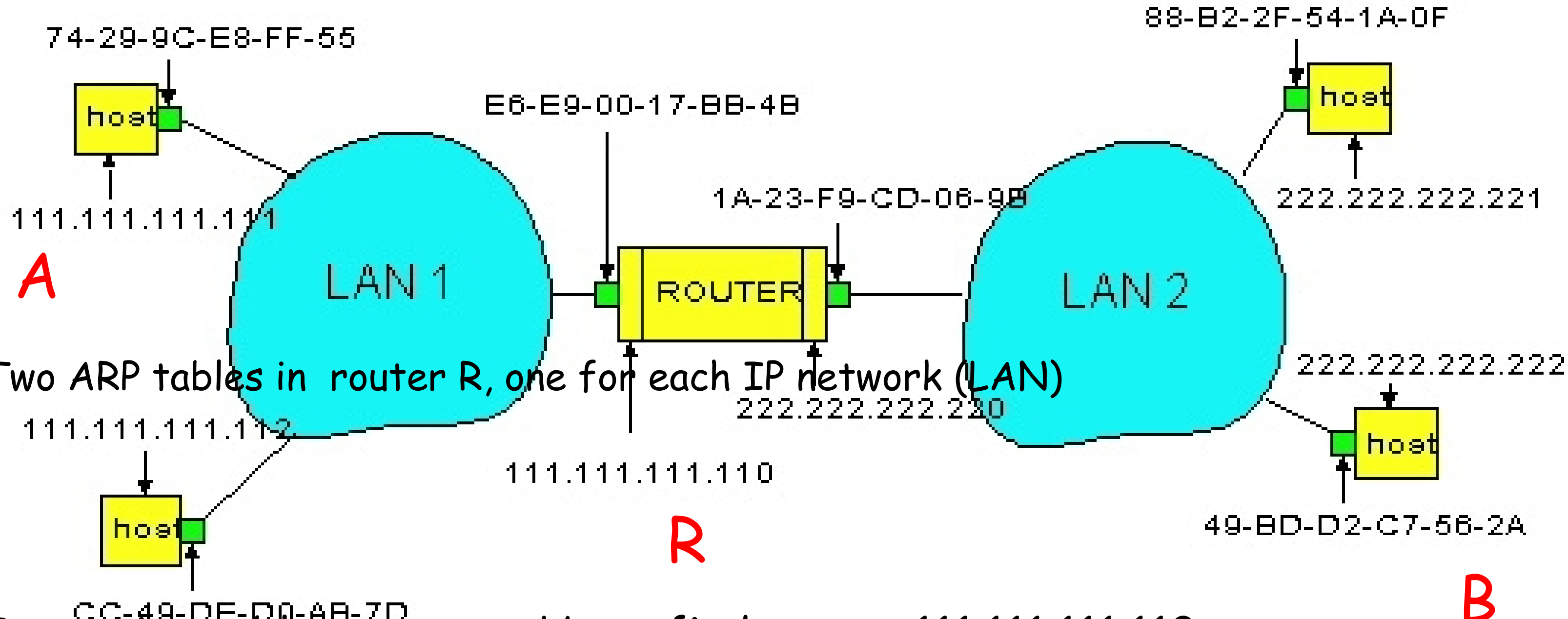


ARP protocol: Same LAN (network)

- 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

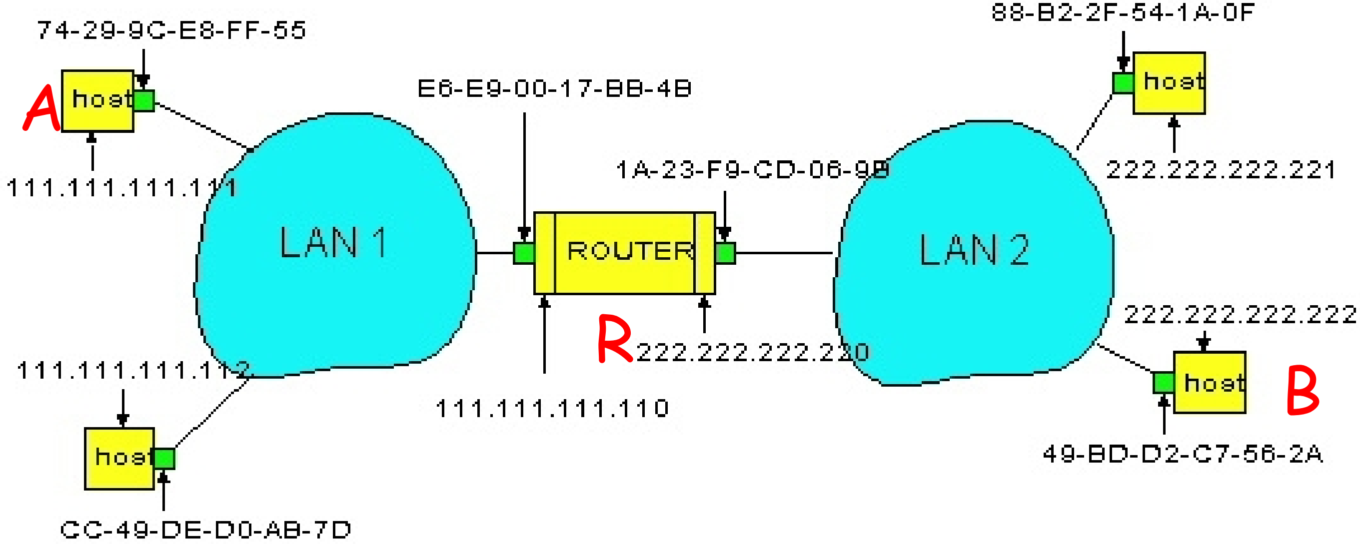
Routing to another LAN

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



- Two ARP tables in router R, one for each IP network (LAN)
- In routing table at source Host, find router 111.111.111.110
- In ARP table at source, find MAC address E6-E9-00-17-BB-4B, etc

- 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 destination, frame contains A-to-B IP datagram
- A's data link layer sends frame
- R's data link layer receives frame
- R removes IP datagram from Ethernet frame, sees its destined to B
- R uses ARP to get B's physical layer address
- R creates frame containing A-to-B IP datagram sends to B



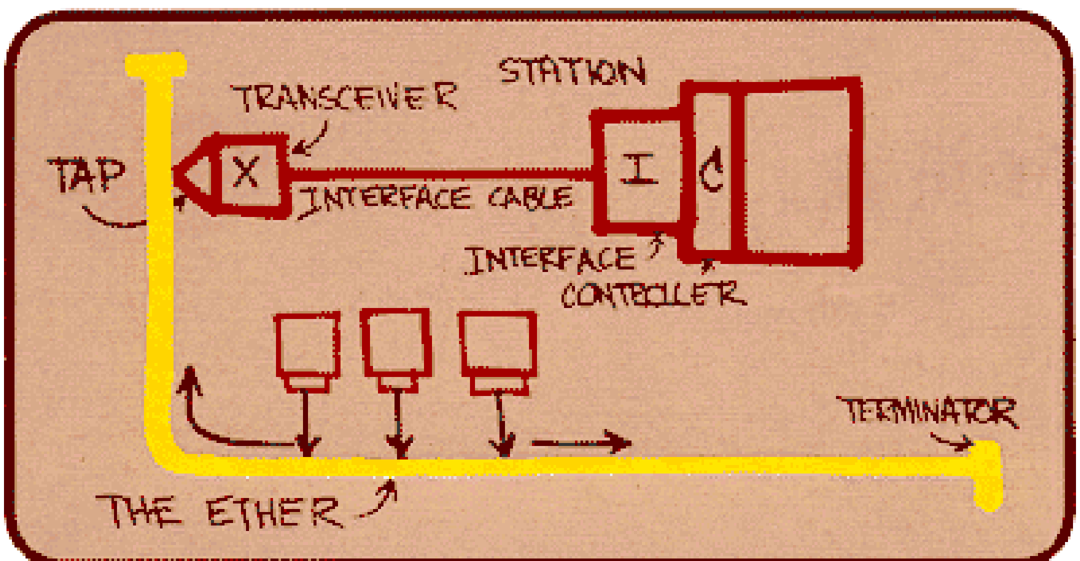
Link Layer

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple 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:

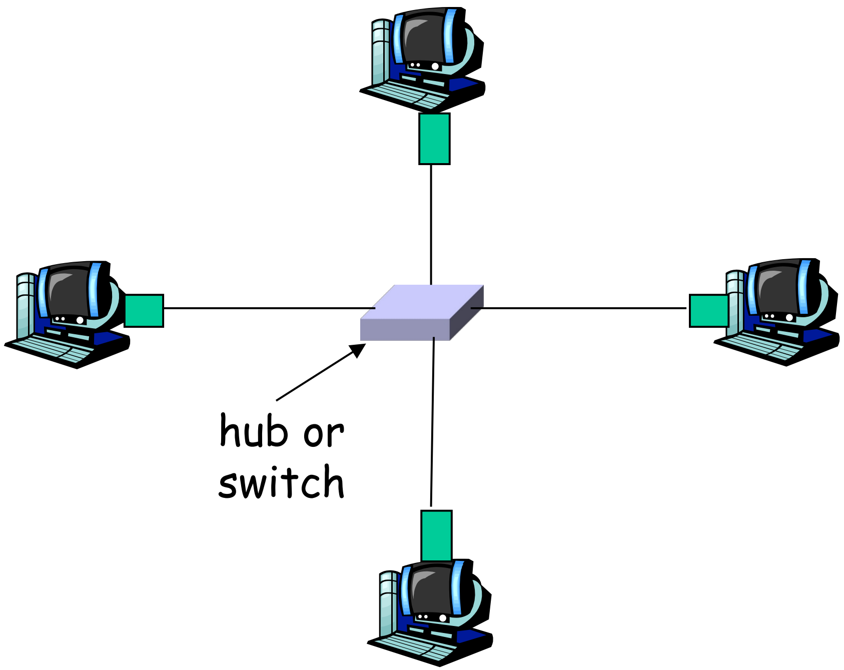
- cheap $20 for 100Mbs!
- first widely used LAN technology
- Simpler, cheaper than token LANs and ATM
- Kept up with speed race: 10, 100, 1000 Mbps



Metcalfe's Ethernet sketch

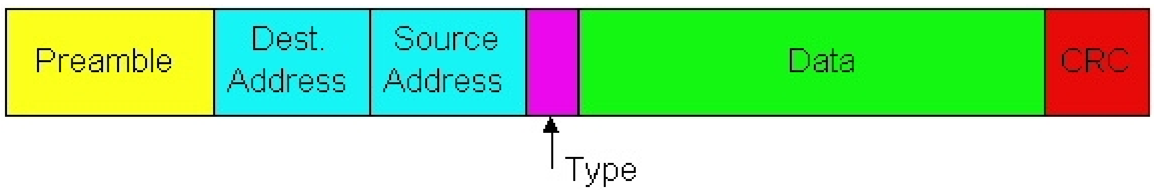
Star topology

- 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

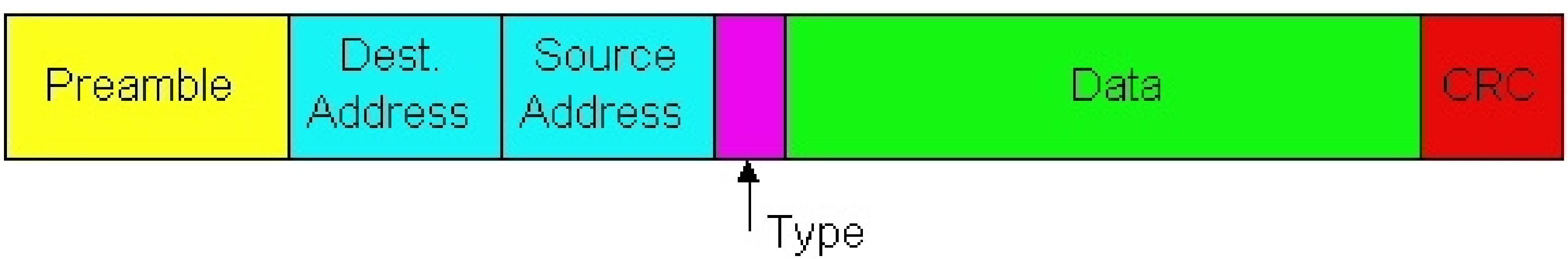


Preamble:

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

Ethernet Frame Structure (more)

- **Data:** 46 to 1500 bytes
- **Addresses:** 6 bytes
  - 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:** indicates the higher layer protocol (mostly IP but others may be supported such as Novell IPX and AppleTalk)
- **CRC:** checked at receiver, if error is detected, the frame is simply dropped



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 data gaps due to discarded fames if the application is using UDP
  - data gaps will be filled by retransmissions if application is using TCP
  - otherwise, application will see the gaps

Ethernet uses CSMA/CD

- adapter may begin to transmit at anytime, i.e., no slots are used
- 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 also 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 network layer and creates frame
2. If adapter senses channel idle, it starts to transmit frame.

   If it senses channel busy, waits until channel idle 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 jam signal
5. After aborting, adapter enters **exponential backoff**: after the nth collision, adapter chooses a K at random from $\{0,1,2,\ldots,2^m-1\}$ where $m = \min(n, 10)$. Adapter waits K*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: 0.1 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 x 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,…,1023}

CSMA/CD efficiency

- T_{prop} = max propagation delay between 2 nodes in LAN
- t_{trans} = time to transmit max-size frame
- Efficiency: the long-run fraction of time during which frames are being transmitted on the channel without collisions when there are a large number of active nodes

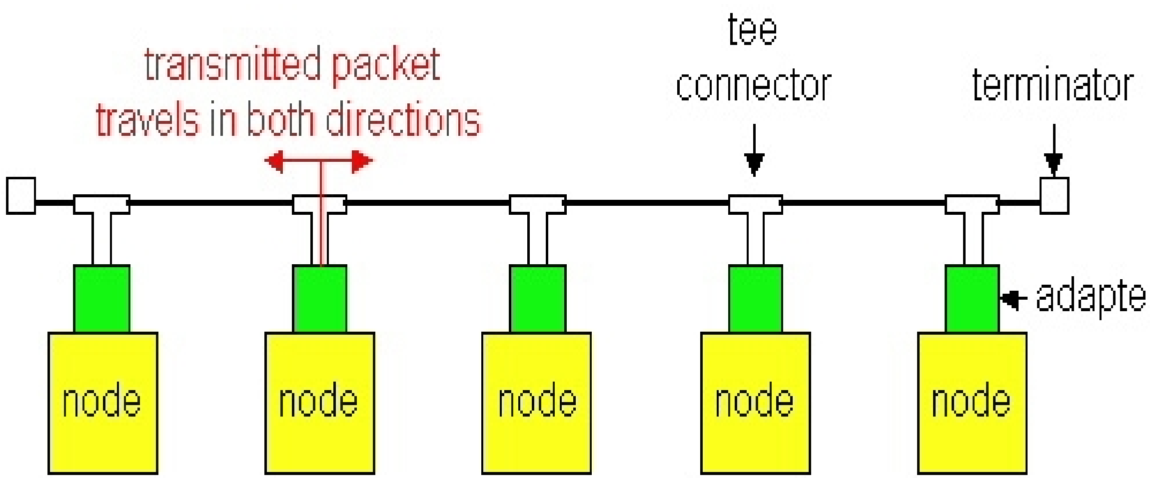
$$\text{efficiency} = \frac{1}{1 + 5t_{prop} / t_{trans}}$$

[Lam 1980, Bertsekas 1991]

- 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

Ethernet Technologies: 10Base2

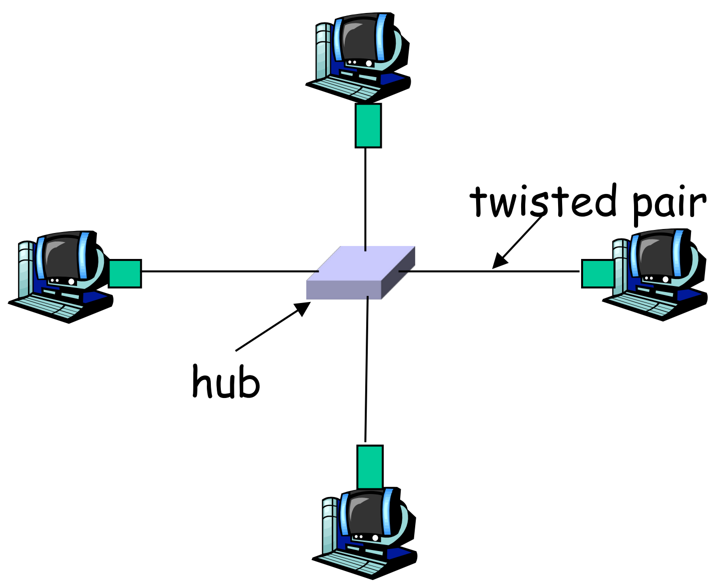
- 10: 10Mbps;
- 2: under 200 meters max cable length
- thin coaxial cable in a bus topology



- repeaters used to connect up to multiple segments
- repeater repeats bits it hears on one interface to its other interfaces: physical layer device only!
- has become a legacy technology

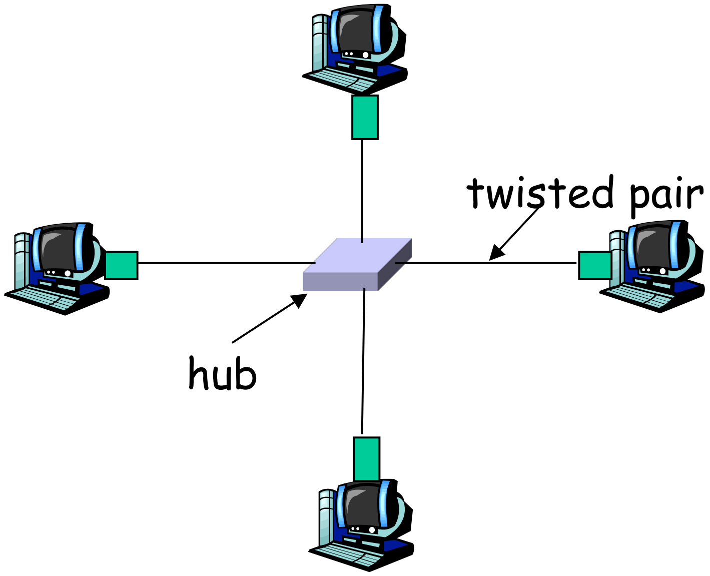
10BaseT and 100BaseT

- 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

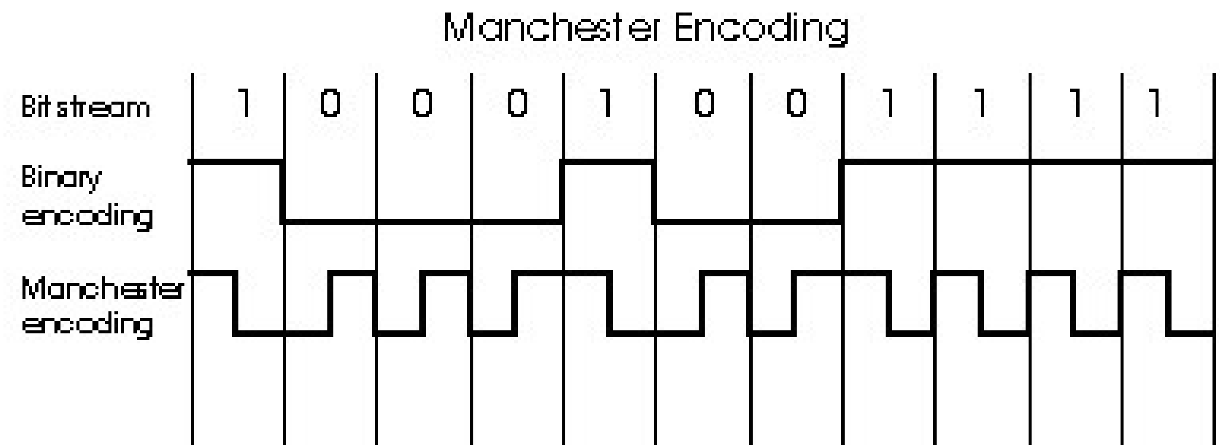


Hubs

- 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



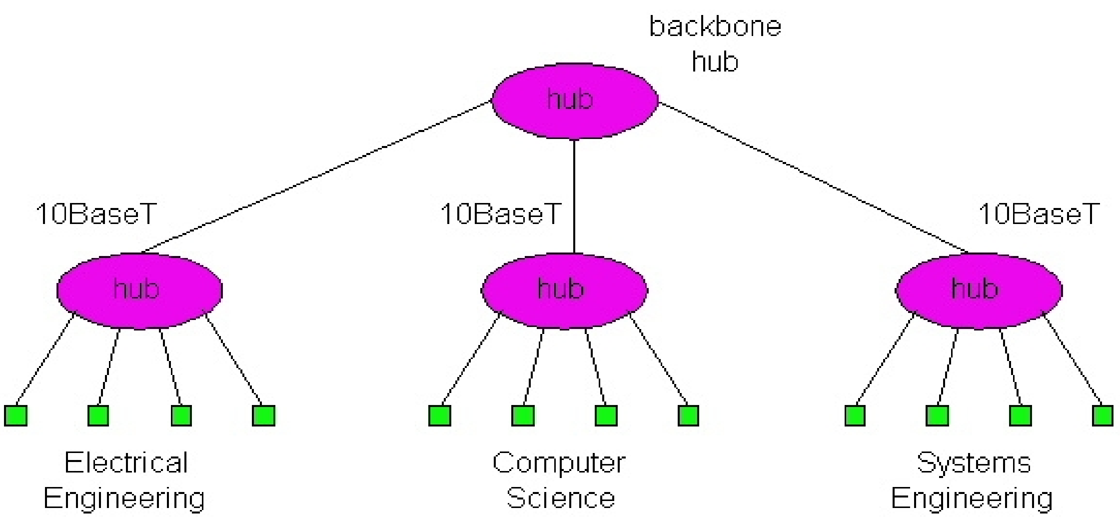
- Used in 10BaseT, 10Base2
- Each bit has a transition – 1: up to down, 0: down to up
- Allows clocks in sending and receiving nodes to synchronize to each other
  - no need for a centralized, global clock among nodes!
- Hey, this is physical-layer stuff!

Gbit Ethernet

- use standard Ethernet frame format
- allows for point-to-point links as well as shared broadcast channels
- Point-to-point links use switches
- Shared broadcast channels use hubs called "Buffered Distributors"
- in shared broadcast channels, CSMA/CD is used; short distances between nodes to be efficient
- 10 Gbps now !

Interconnecting with hubs

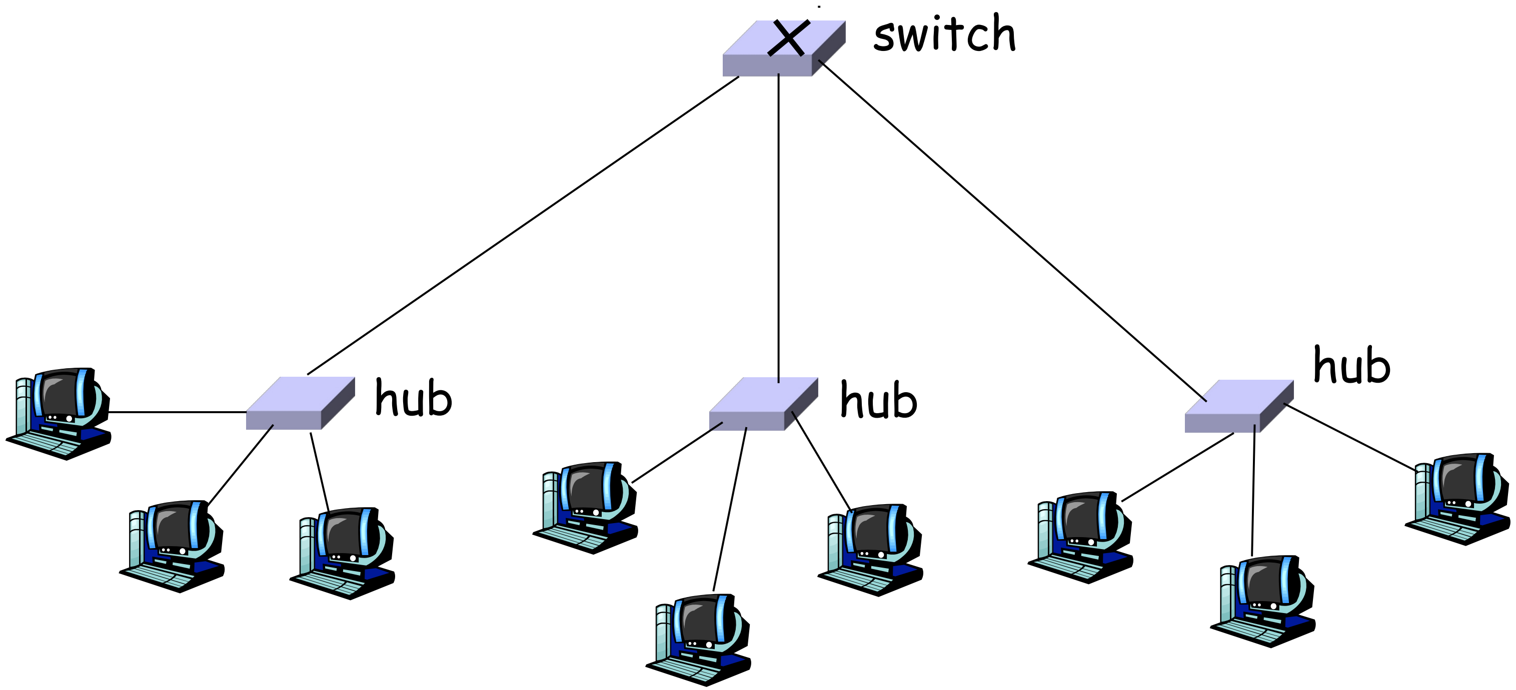
- Backbone hub interconnects LAN segments
- Extends max distance between nodes
- Limitations:
  - But individual segment collision domains become one large collision domain – all hosts share 10Mbps
    - if a node in CS and a node EE transmit at same time: collision
  - Can't interconnect 10BaseT & 100BaseT
  - A collision domain has restrictions on the maximum allowable number of nodes, the maximum distance between two hosts, the maximum number of tiers in a multi-tier design



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 switches determine to 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 of a node, Switch 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 interface
  - records sender/interface pair in switch table

Filtering/Forwarding

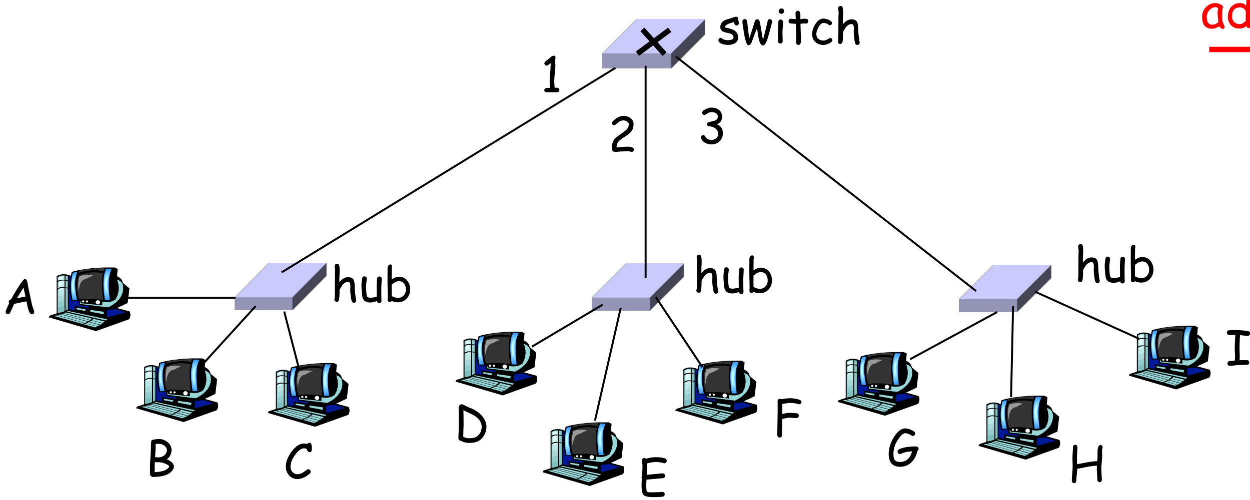
When switch receives a frame:

index switch table using MAC destination address
if entry found for destination
  then {
    if destination on interface 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 and D replies back with frame to C.

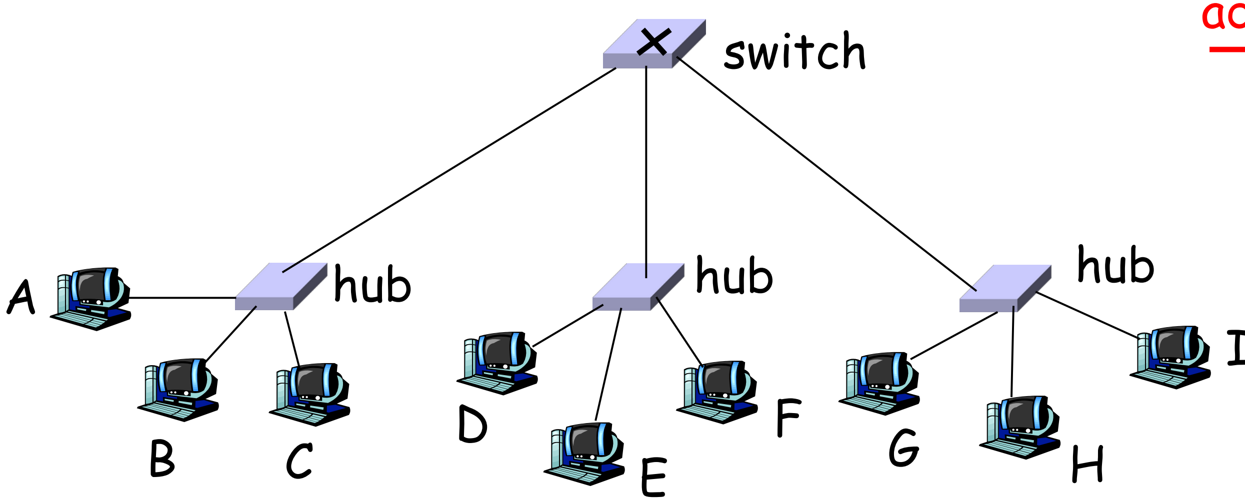


address	interface
A	1
B	1
E	2
G	3

- Switch receives frame from C
  - records 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.

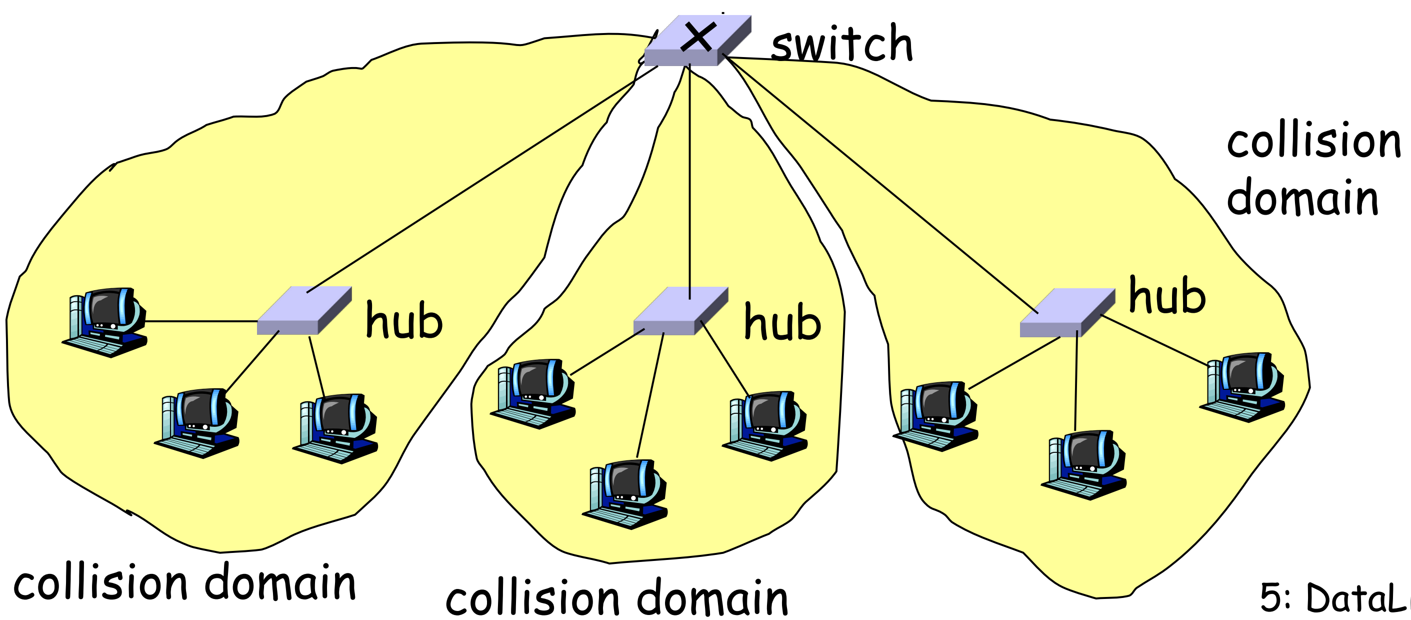


address	interface
A	1
B	1
E	2
G	3
C	1

- Switch receives frame from from D
  - records in switch table 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

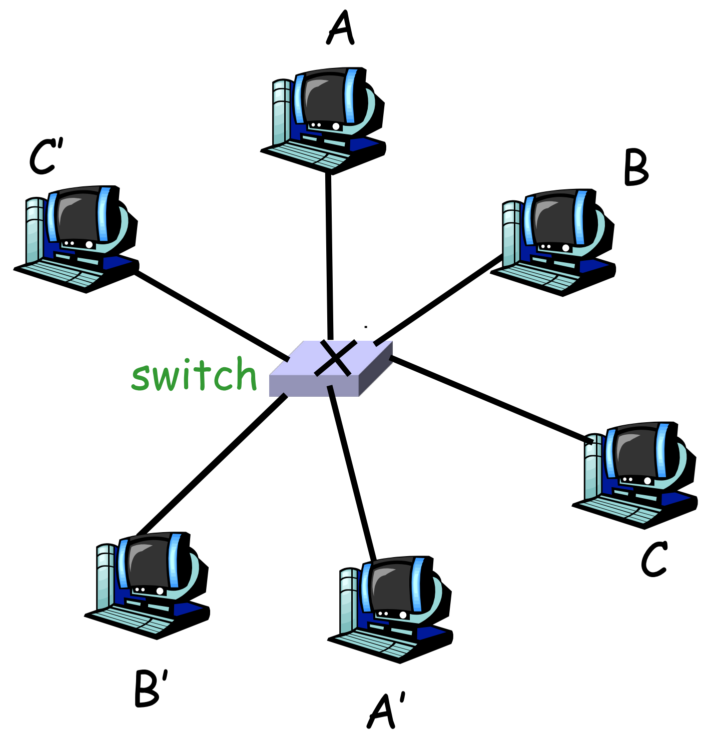
- switch installation breaks subnet into LAN segments
- switch **filters** packets:
  - same-LAN-segment frames not usually forwarded onto other LAN segments
  - segments become separate **collision domains**



Switches: dedicated access

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

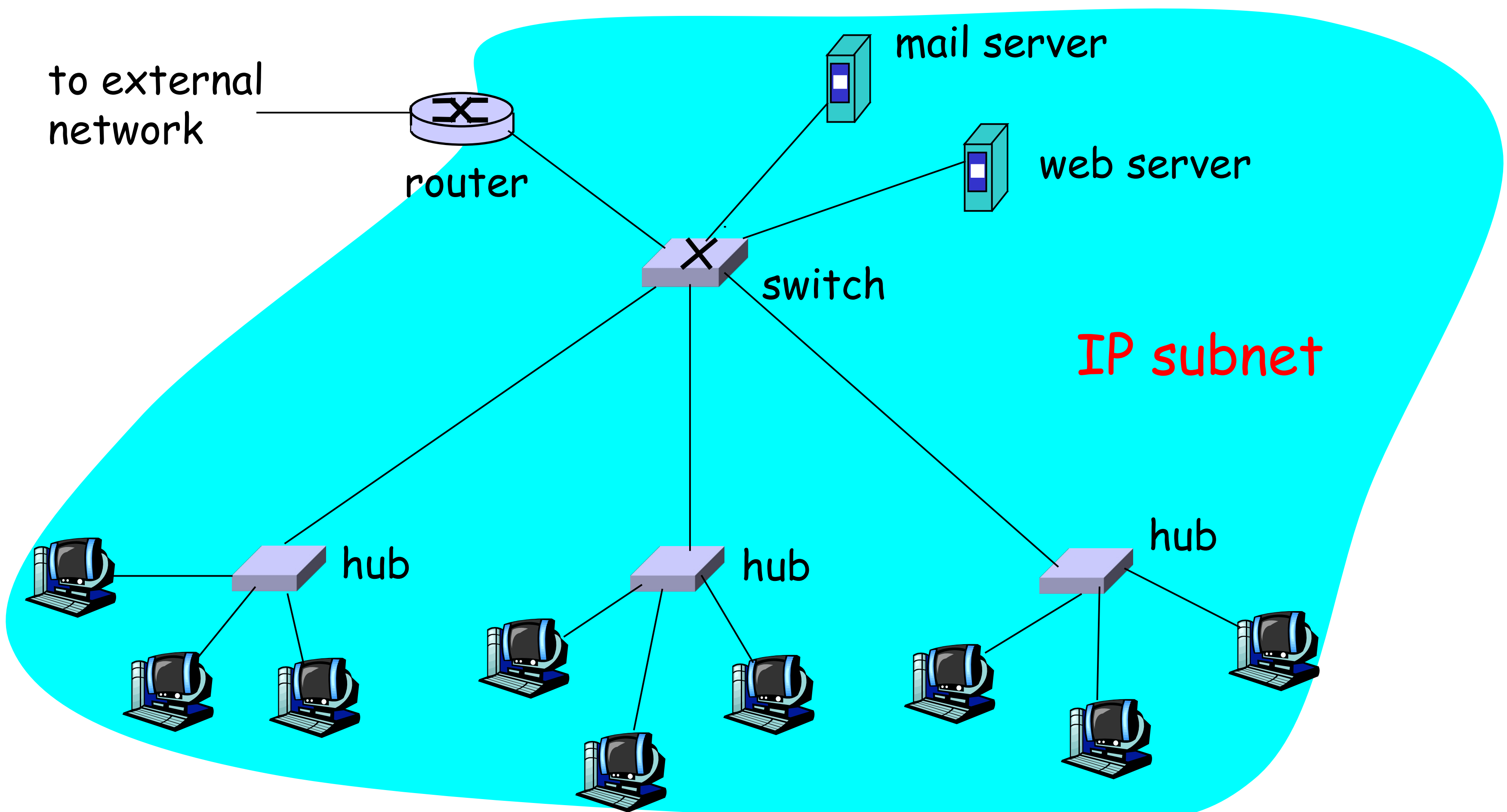
Switching: A-to-A' and B-to-B' simultaneously, no collisions



More on Switches

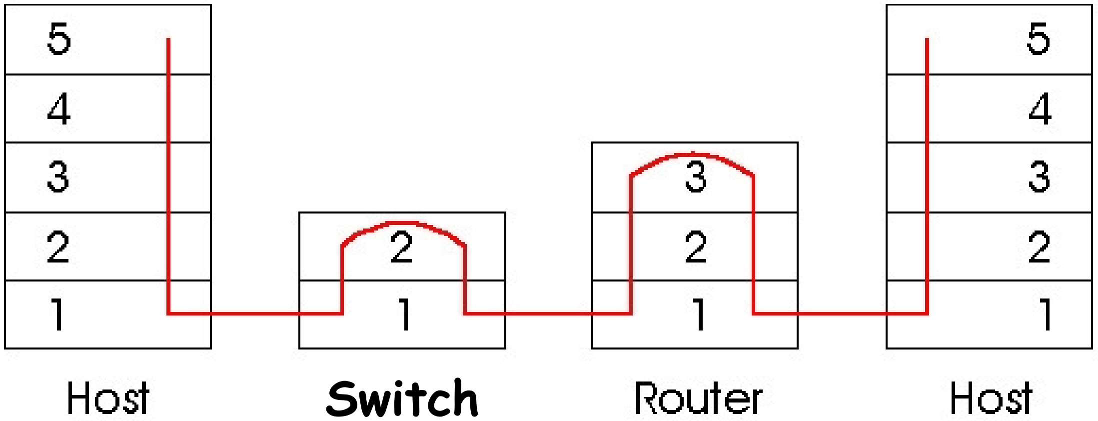
- cut-through switching: when the output buffer is empty, a frame forwarded from input to output port without first collecting entire frame
  - slight reduction in latency
- combinations of shared/dedicated, 10/100/1000 Mbps interfaces

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



Summary comparison

	hubs	routers	switches
traffic isolation	no	yes	yes
plug & play	yes	no	yes
optimal routing	no	yes	no
cut through	yes	no	yes

Link Layer

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

Point to Point Data Link Control

- one sender, one receiver, one link: easier than broadcast link:
  - no Media Access Control
  - no need for explicit MAC addressing
  - e.g., dialup link, ISDN line
- popular point-to-point Data Link Control (DLC) protocols:
  - PPP (point-to-point protocol)
  - HDLC: High level data link control (Data link used to be considered "high layer" in protocol stack!)

PPP Design Requirements [RFC 1557]

- packet framing: encapsulation of network-layer datagram in data link frame
  - carry network layer data of any network layer protocol (not just IP) *at same time*
  - ability to demultiplex upwards
- bit transparency: must carry any bit pattern in the data field
- error detection (no correction)
- connection liveness: detect a link failure, signal link failure to network layer
- network layer address negotiation: endpoint can learn/configure each other's network address

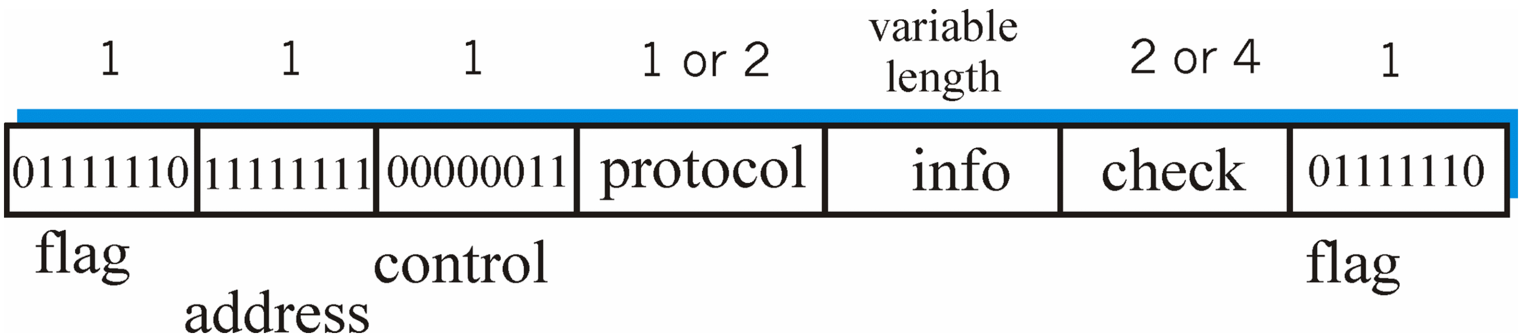
PPP non-requirements

- no error correction/recovery
- no flow control
- out of order delivery OK
- no need to support multipoint links (e.g., polling)

Error recovery, flow control, data re-ordering all relegated to higher layers!

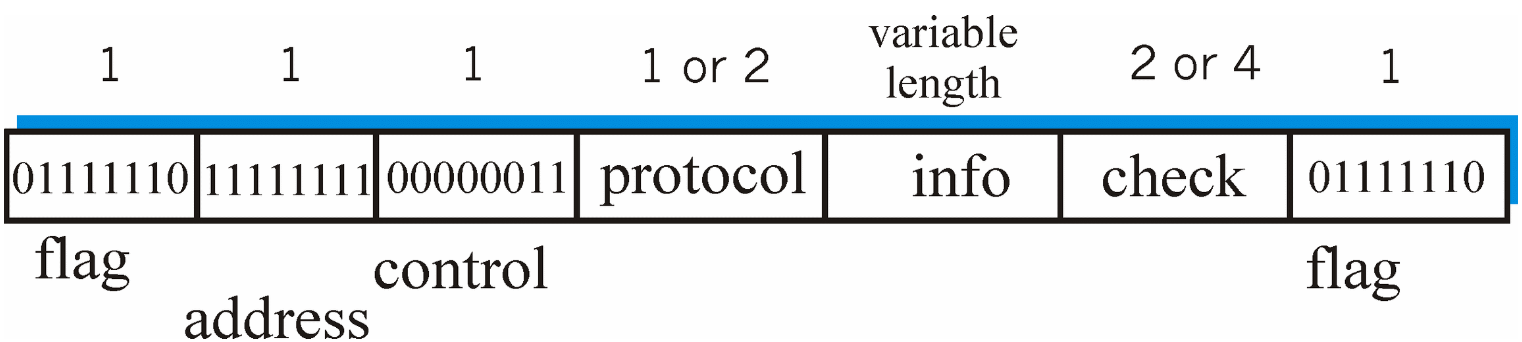
PPP Data Frame

- Flag: delimiter (framing)
- Address: does nothing (only one option)
- Control: does nothing; in the future possible multiple control fields
- Protocol: upper layer protocol to which frame delivered (eg, PPP-LCP, IP, IPCP, etc)



PPP Data Frame

- info: upper layer data being carried, default maximum length = 1500 bytes
- check: cyclic redundancy check for error detection

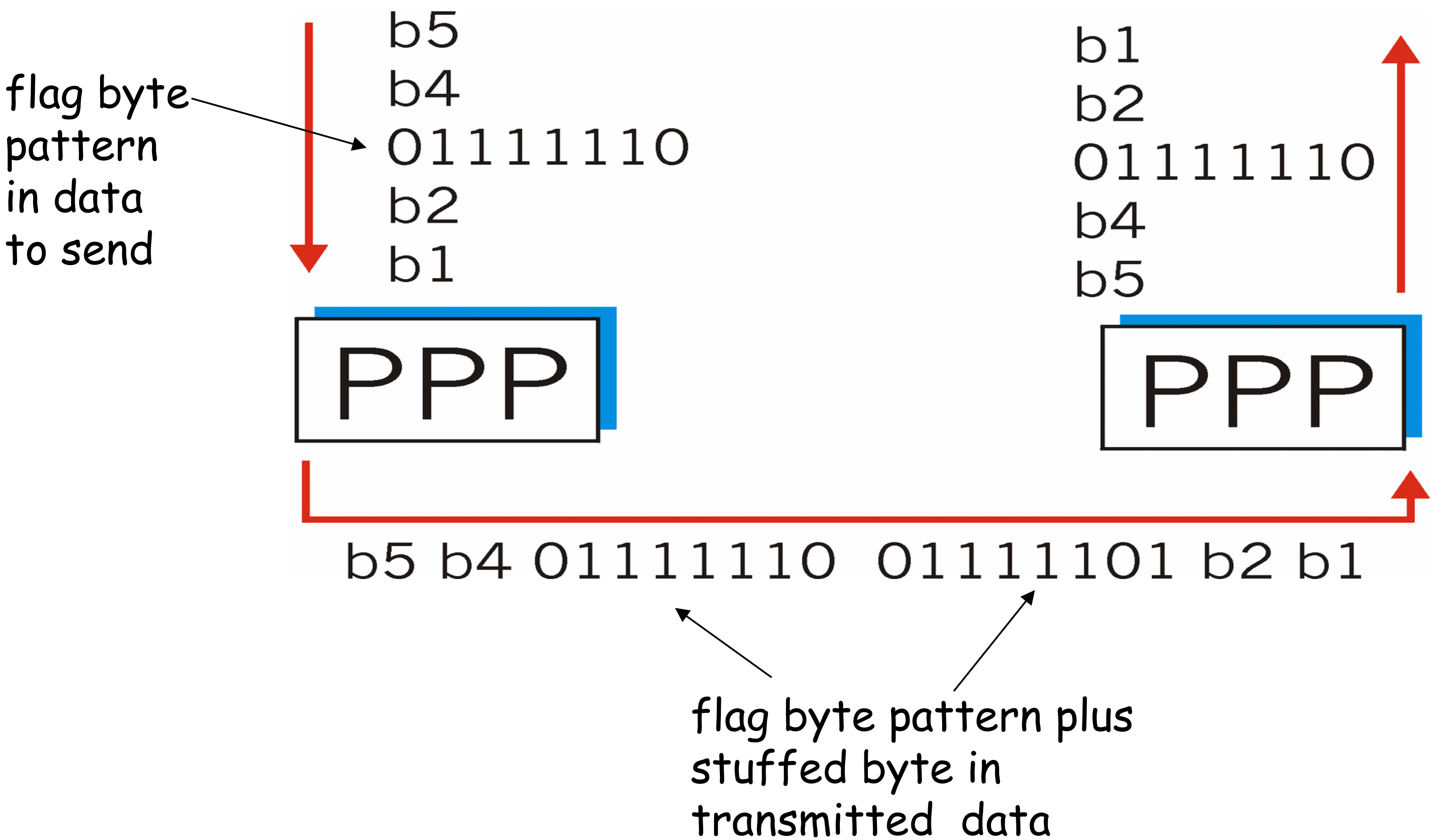


Byte Stuffing

- "data transparency" requirement: data field must be allowed to include flag pattern <01111110>
  - Q: is received <01111110> data or flag?

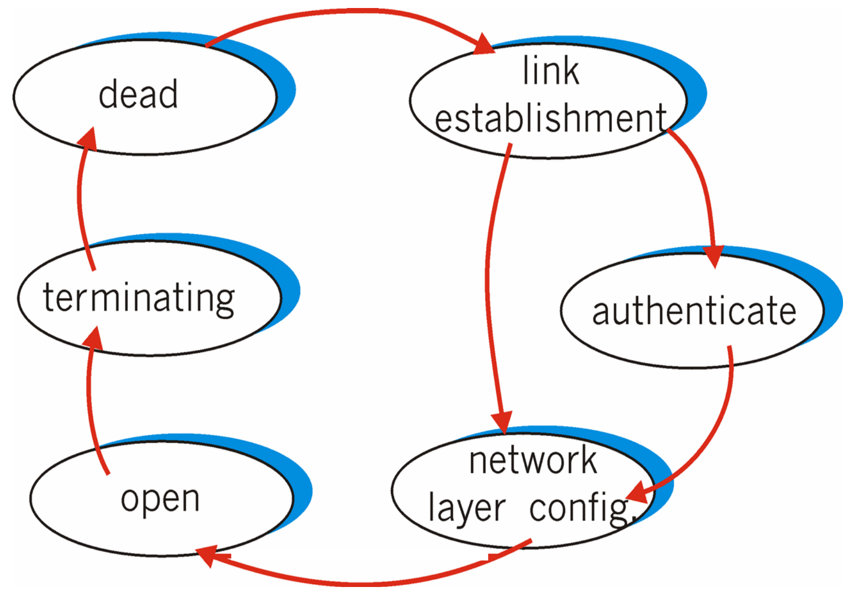
- Sender:
  - adds ("stuffs") extra < 01111101> byte before each < 01111110> *data* byte
  - adds ("stuffs") extra < 01111101> byte before each < 01111101> *data* byte
- Receiver:
  - single 01111101 byte: discard 01111101
  - two 01111101 bytes in a row: discard first byte, continue data reception
  - single 01111110: flag byte

Byte Stuffing



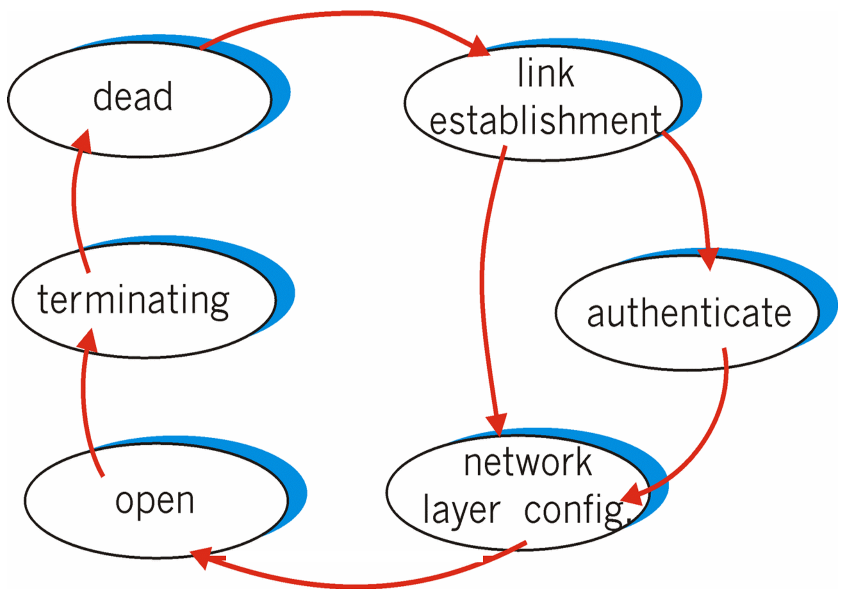
PPP Control Protocol

- Begins and ends in the dead state
- Enters link establishment state when the physical layer is present and ready to be used
- In the link establishment state, PPP link-control protocol (LCP) is used to negotiate link configuration options such as maximum frame size, authentication protocol (if any) to be used, etc.



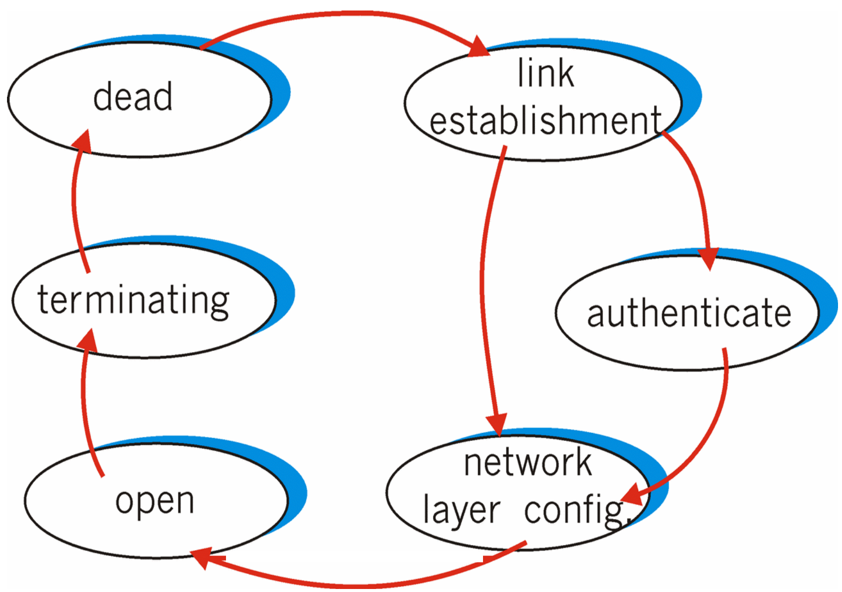
PPP Control Protocol (Cont.)

- Then, the end points enter the network layer configuration state to learn/configure network layer information using a network-control protocol
- The network-control protocol to be used depends on the specific network layer protocol
  - for IP: IP Control Protocol (IPCP) (protocol field: 8021) is used to configure/learn IP address
- Once the network layer has been configured, PPP enters the open state and may begin sending network layer datagrams



PPP Control Protocol (Cont.)

- The LCP echo-request frame and echo reply frame can be exchanged between Two PPP endpoints in order to check the status of the link
- To terminate the link, one end of the PPP link sends a terminate-request LCP frame and the other end replies with a terminate-ack LCP frame
- The link enter the dead state



Link Layer

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

Virtualization of networks

Virtualization of resources: a powerful abstraction in systems engineering:

- computing examples: virtual memory, virtual devices
  - Virtual machines: e.g., java
  - IBM VM os from 1960's/70's
- layering of abstractions: don't sweat the details of the lower layer, only deal with lower layers abstractly

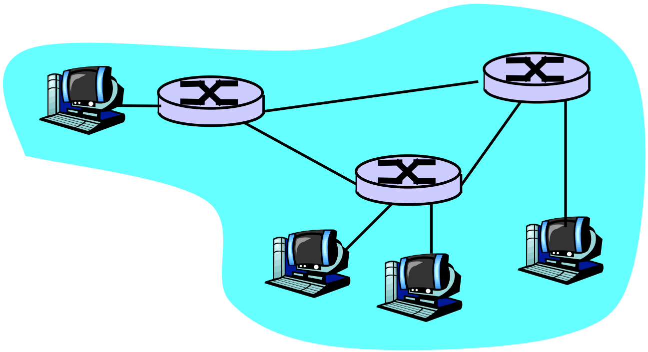
The Internet: virtualizing networks

1974: multiple unconnected nets

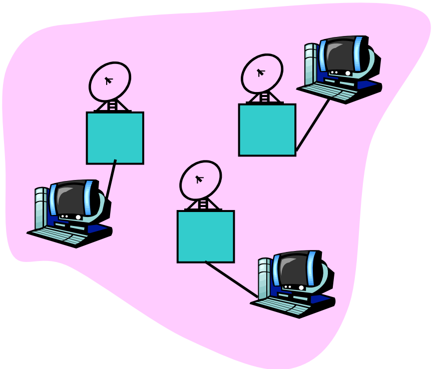
- ARPAnet
- data-over-cable networks
- packet satellite network (Aloha)
- packet radio network

… differing in:

- addressing conventions
- packet formats
- error recovery
- routing



ARPAnet



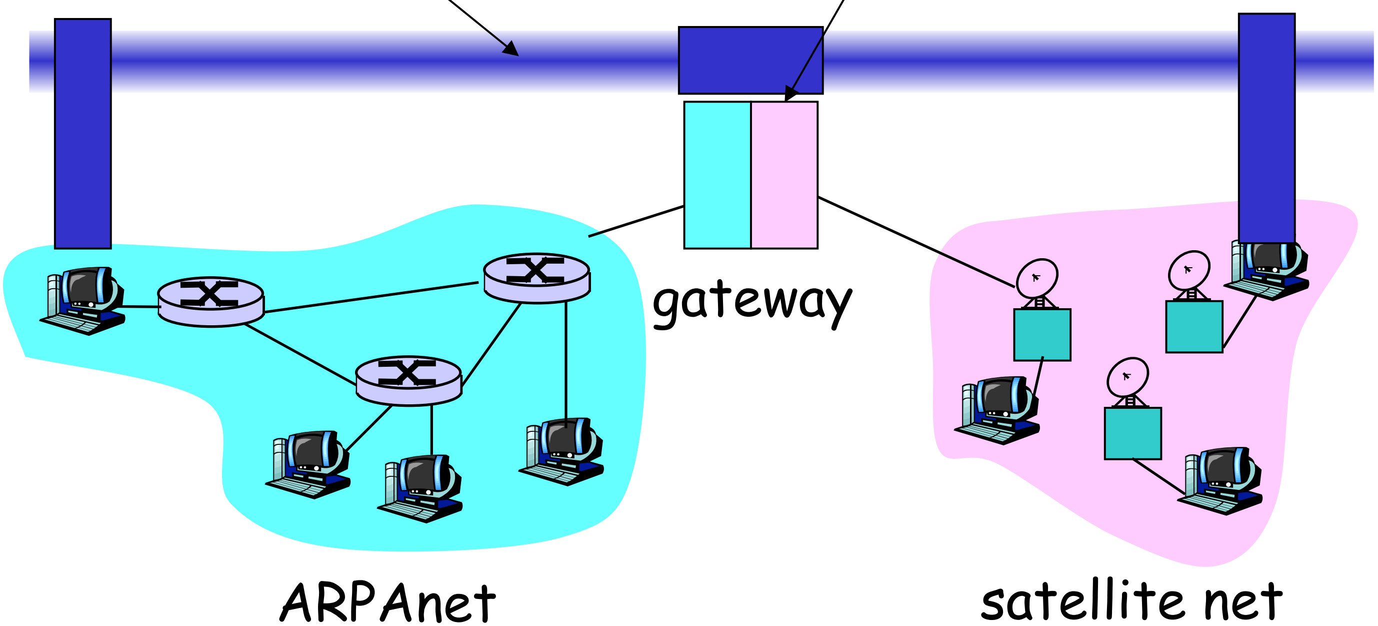
satellite net

"A Protocol for Packet Network Intercommunication",
V. Cerf, R. Kahn, IEEE Transactions on Communications,
May, 1974, pp. 637-648.

The Internet: virtualizing networks

Internetwork layer (IP):
- addressing: internetwork appears as a single, uniform entity, despite underlying local network heterogeneity
- network of networks

Gateway:
- "embed internetwork packets in local packet format or extract them"
- route (at internetwork level) to next gateway



Cerf & Kahn's Internetwork Architecture

What is virtualized?

- two layers of addressing: internetwork and local network
- new layer (IP) makes everything homogeneous at internetwork layer
- underlying local network technology
  - cable
  - satellite
  - 56K telephone modem
  - today: ATM, MPLS

… "invisible" at internetwork layer. Looks like a link layer technology to IP!

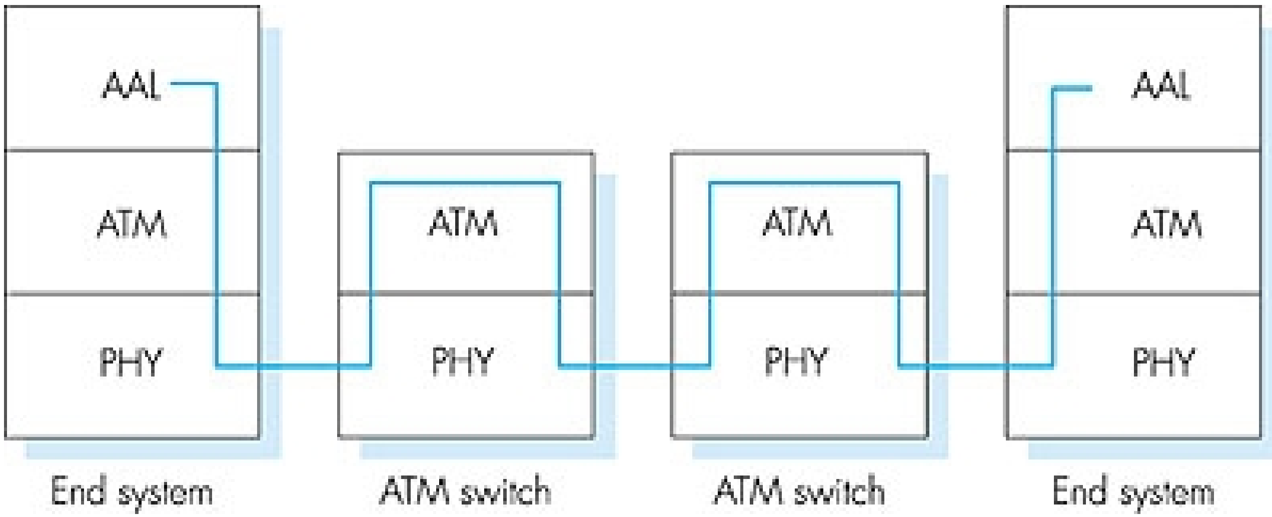
ATM and MPLS

- ATM, MPLS separate networks in their own right
  - different service models, addressing, routing from Internet
- viewed by Internet as logical link connecting IP routers
  - just like dialup link is really part of separate network (telephone network)
- ATM, MPSL: of technical interest in their own right

Asynchronous Transfer Mode: ATM

- **1990's/00 standard for high-speed** (155Mbps to 622 Mbps and higher) *Broadband Integrated Service Digital Network* architecture
- Goal: *integrated, end-to-end transport for carrying voice, video, data*
  - meeting timing/QoS requirements of voice, video (versus Internet best-effort model)
  - "next generation" telephony: technical roots in telephone world
  - packet-switching (fixed length packets, called "cells") using virtual circuits

ATM architecture



The ATM protocol stack consists of three layers:

- adaptation layer: only at edge of ATM network
  - data segmentation/reassembly
  - roughly analagous to Internet transport layer
  - Several different types of AALs to support different types of services
- ATM layer: the core of the ATM standard
  - cell switching, routing
- physical layer

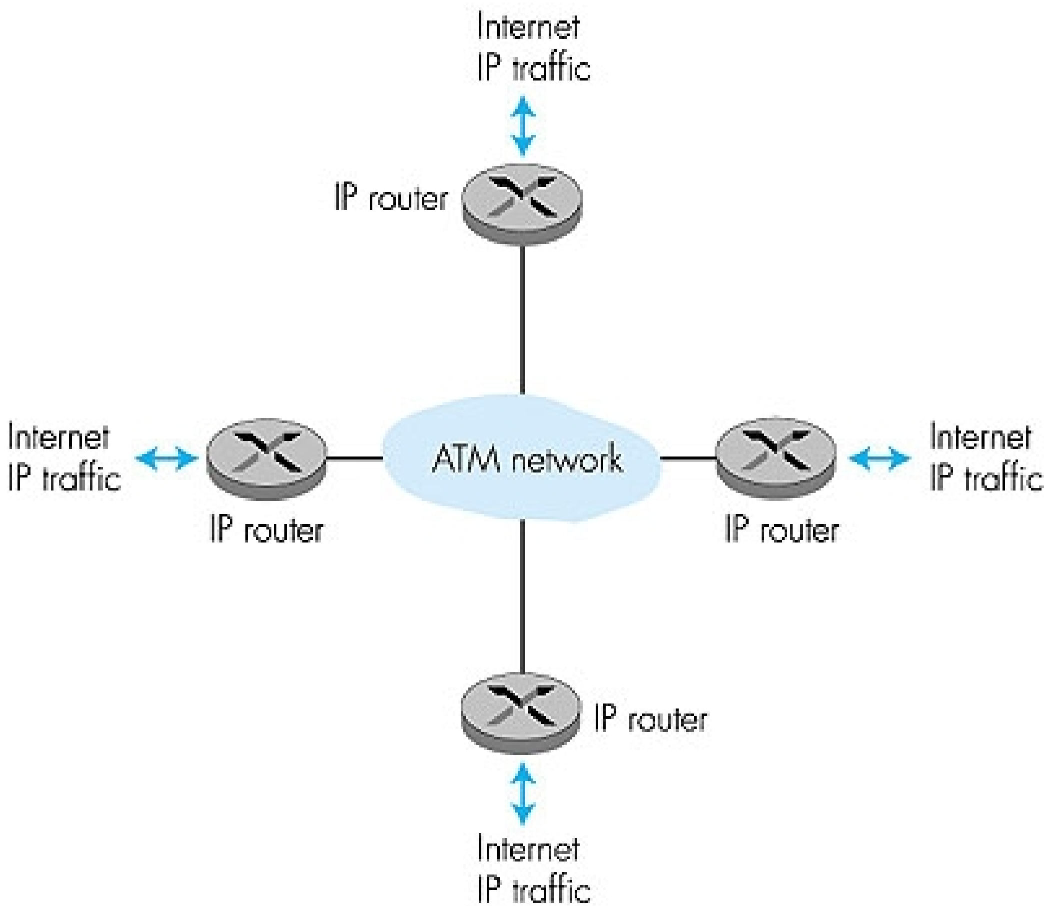
ATM: network or link layer?

Vision: end-to-end transport: "ATM from desktop to desktop"

- ATM *is* a network technology

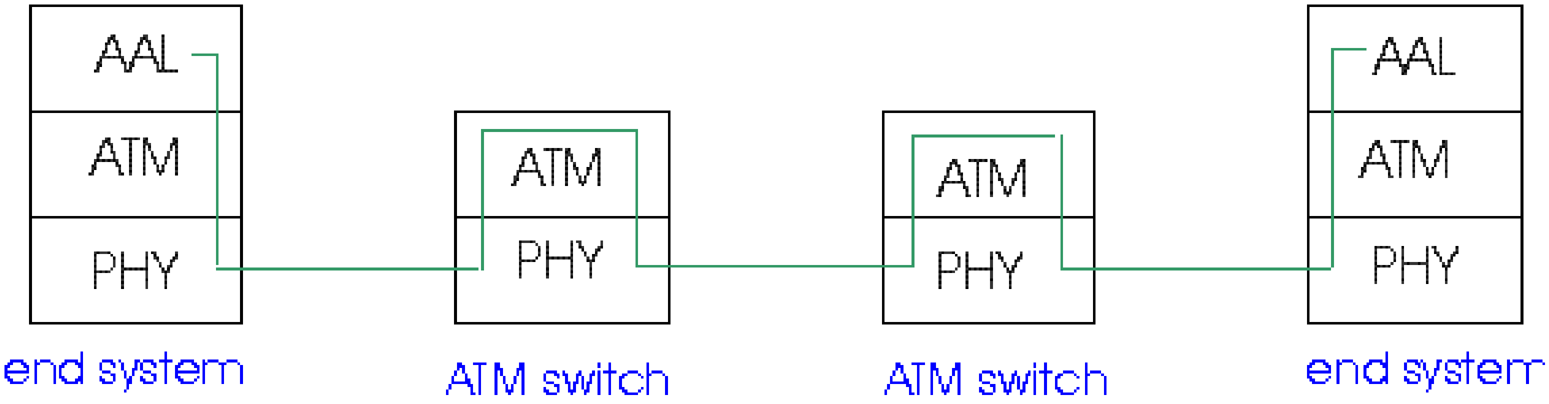
Reality: used to connect IP backbone routers

- "IP over ATM"
- ATM as switched link layer, connecting IP routers



ATM Adaptation Layer (AAL)

- ATM **Adaptation Layer** (AAL): "adapts" upper layers (IP or native ATM applications) to ATM layer below
- AAL present **only in end systems**, not in switches
- AAL layer segment (header/trailer fields, data) is fragmented across multiple ATM cells
  - analogy: TCP segment is fragmented in many IP packets



ATM Adaptation Layer (AAL) [more]

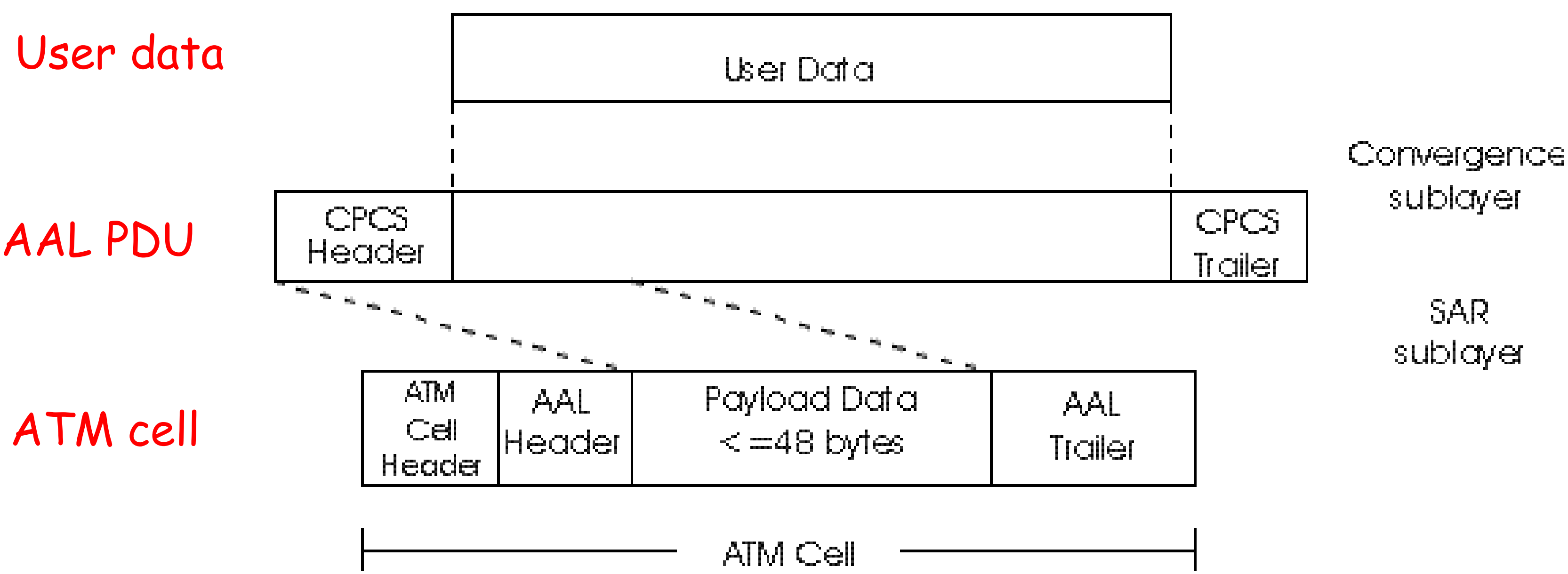
Different versions of AAL layers, depending on ATM service class:

- AAL1: for CBR (Constant Bit Rate) services, e.g. circuit emulation
- AAL2: for VBR (Variable Bit Rate) services, e.g., MPEG video
- AAL5: for data (eg, IP datagrams)

ATM Adaptation Layer (AAL) [more]

AAL has two sublayers:

- Convergence sublayer: higher-layer data are encapsulated in a common part convergence sublayer (CPCS)
- Segmentation and reassembly (SAR) sublayer: segments the CPCS-PDU and adds AAL header and trailer bits to form the payloads of the ATM



ATM Layer

- Service: transport cells across ATM network
- analogous to IP network layer
- very different services than IP network layer

Network Architecture	Service Model	Guarantees ? Bandwidth	Loss	Order	Timing	Congestion feedback
Internet	best effort	none	no	no	no	no (inferred via loss)
ATM	CBR	constant rate	yes	yes	yes	no congestion
ATM	VBR	guaranteed rate	yes	yes	yes	no congestion
ATM	ABR	guaranteed minimum	no	yes	no	yes
ATM	UBR	none	no	yes	no	no

ATM Layer: Virtual Channels

- VC transport: cells carried on VC from source to dest
  - call setup for each call *before* data can flow
  - each packet carries a virtual channel identifier (VCI)
  - *every* switch on source-dest path maintain "state" for each passing connection
  - link, switch resources (bandwidth, buffers) may be *allocated* to VC: to get circuit-like performance
- Two types of VCs
  - Permanent VCs (PVCs)
    - long lasting connections
    - typically: "permanent" route between IP routers
  - Switched VCs (SVC):
    - dynamically set up on per-call basis

ATM VCs

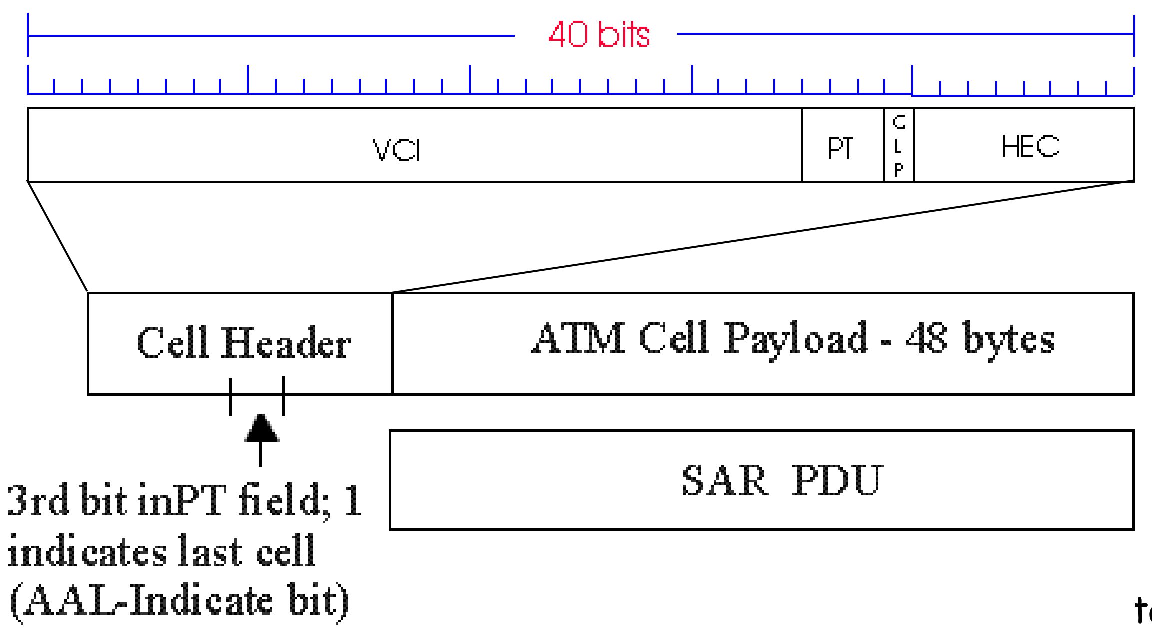
- Advantages of ATM VC approach:
  - QoS performance guarantee for connection mapped to VC (bandwidth, delay, delay jitter)
- Drawbacks of ATM VC approach:
  - Inefficient support of datagram traffic
  - one PVC between each source/destination pair does not scale (N*2 connections needed)
  - SVC introduces call setup latency, processing overhead for short lived connections

ATM Layer: ATM cell

- 5-byte ATM cell header
- 48-byte payload
  - Why?: small payload -> short cell-creation delay for digitized voice
  - halfway between 32 and 64 (compromise!)

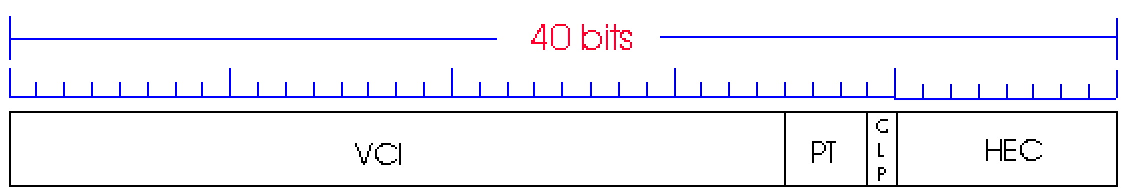
Cell header

Cell format



ATM cell header

- **VCI:** virtual channel ID
  - will *change* from link to link through net
- **PT:** Payload type (e.g. RM cell versus data cell)
- **CLP:** Cell Loss Priority bit
  - CLP = 1 implies low priority cell, can be discarded if congestion
- **HEC:** Header Error Checksum
  - cyclic redundancy check



ATM Physical Layer

Two classes of physical layer:

- Structured: have a transmission frame structure (TDM like frame)
- Unstructured: do not have frame structure

Two sublayers of physical layer:

- Transmission Convergence Sublayer (TCS):
  - Accept ATM cells from the ATM layer and prepare them for transmission
  - Group bits arriving from the physical medium into cells and pass the cells to the ATM layer
- Physical Medium Dependent (PMD) Sublayer:
  - depends on physical medium being used
  - Generates and delineating bits

ATM Physical Layer (more)

Transmission Convergence Sublayer (TCS)

- At the transmit side: generates header checksum (HEC) byte -- 8 bits CRC
- If the Physical Medium Dependent (PMD) sublayer is cell-based with no frames, TCS sends idle cells when ATM layer has not provided data cells to send
- At the receive side, uses the HEC byte to correct all one-bit errors and some multiple-bit errors in the header
- At the receive side, delineates cells by running the HEC on all contiguous sets of 40 bits (When a match occurs, a cell is delineated)

ATM Physical Layer (more)

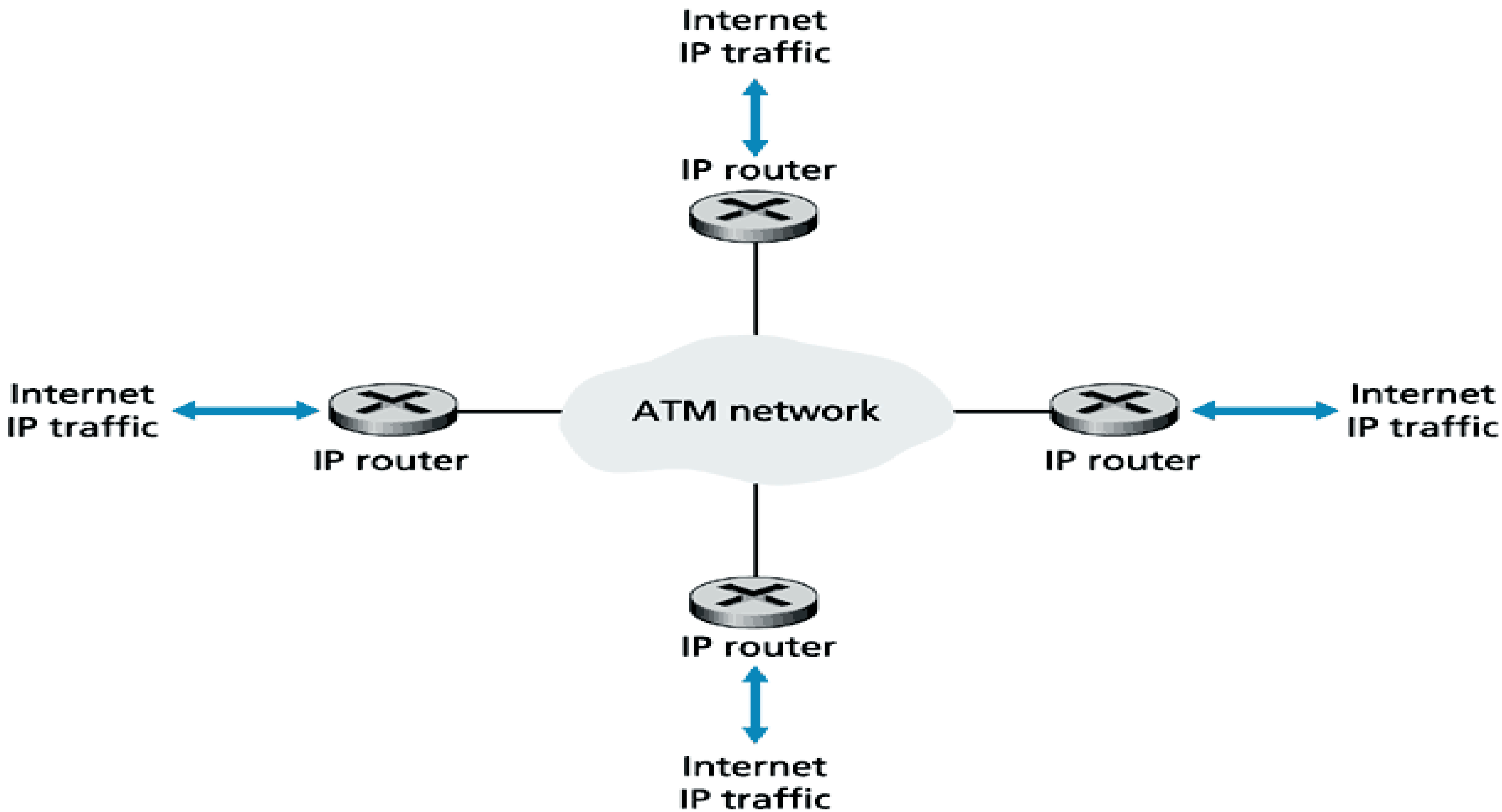
Physical Medium Dependent (PMD) sublayer

Some possible PMD sublayers:

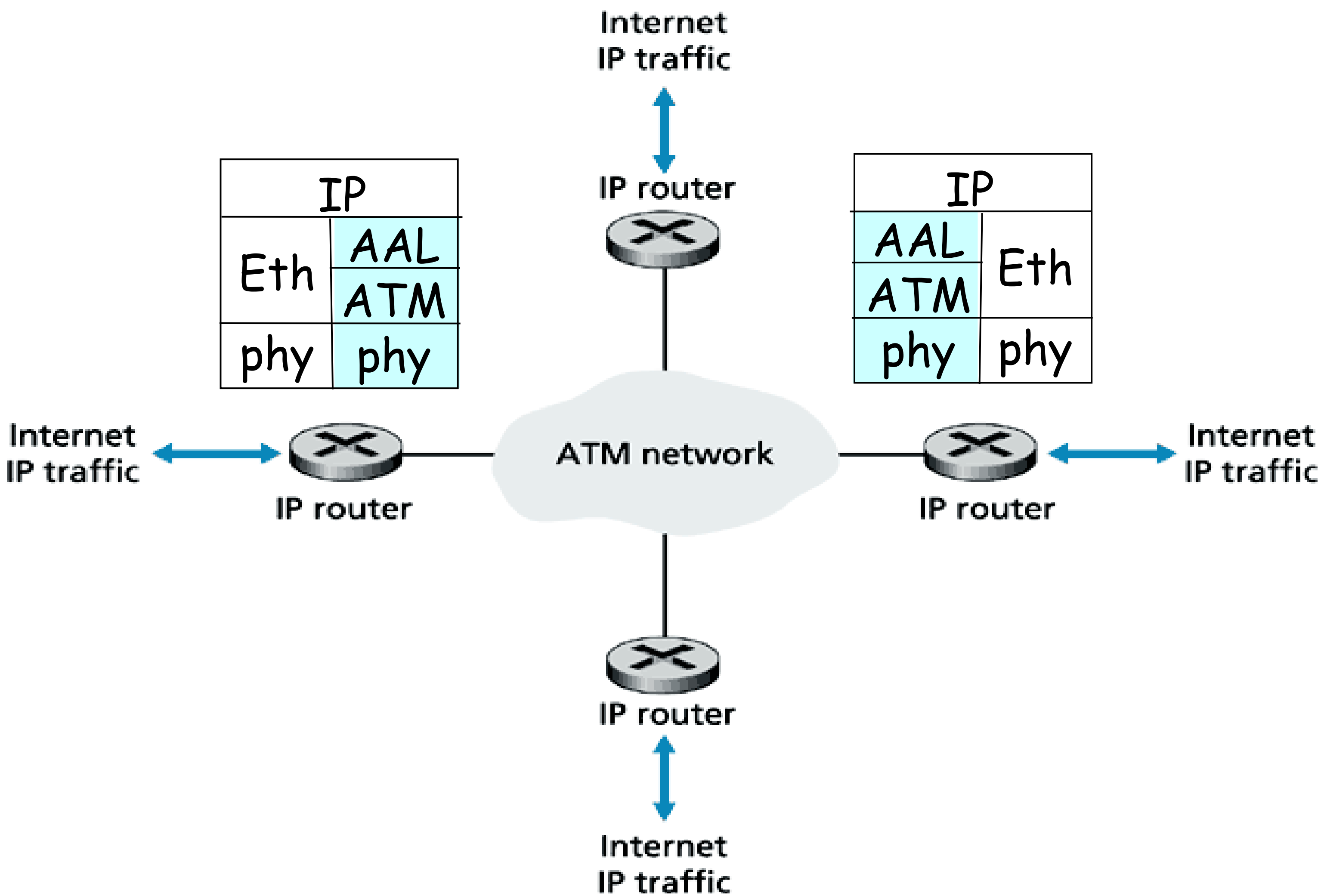
- SONET/SDH (synchronous optical network/synchronous digital hierarchy) : have transmission frame structure (like a container carrying bits);
  - bit synchronization;
  - Generates and delineates frames
  - bandwidth partitions (TDM);
  - several speeds: OC3 = 155.52 Mbps; OC12 = 622.08 Mbps; OC48 = 2.45 Gbps, OC192 = 9.6 Gbps
- T1/T3: have transmission frame structure (old telephone hierarchy): T1 = 1.5Mbps/ T3 = 45Mbps
- Cell-based with no frames: just cells (busy/idle cells)

IP-Over-ATM

- replace “network” with ATM network
- ATM addresses, IP addresses

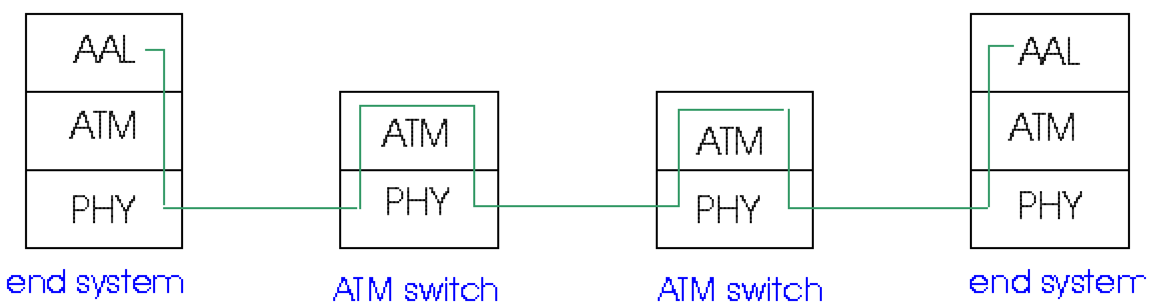


IP-Over-ATM


Internet
IP traffic
IP router
IP
Eth
AAL
ATM
phy
phy
IP
AAL
ATM
Eth
phy
phy
Internet
IP traffic
IP router
ATM network
IP router
Internet
IP traffic
IP router
Internet
IP traffic

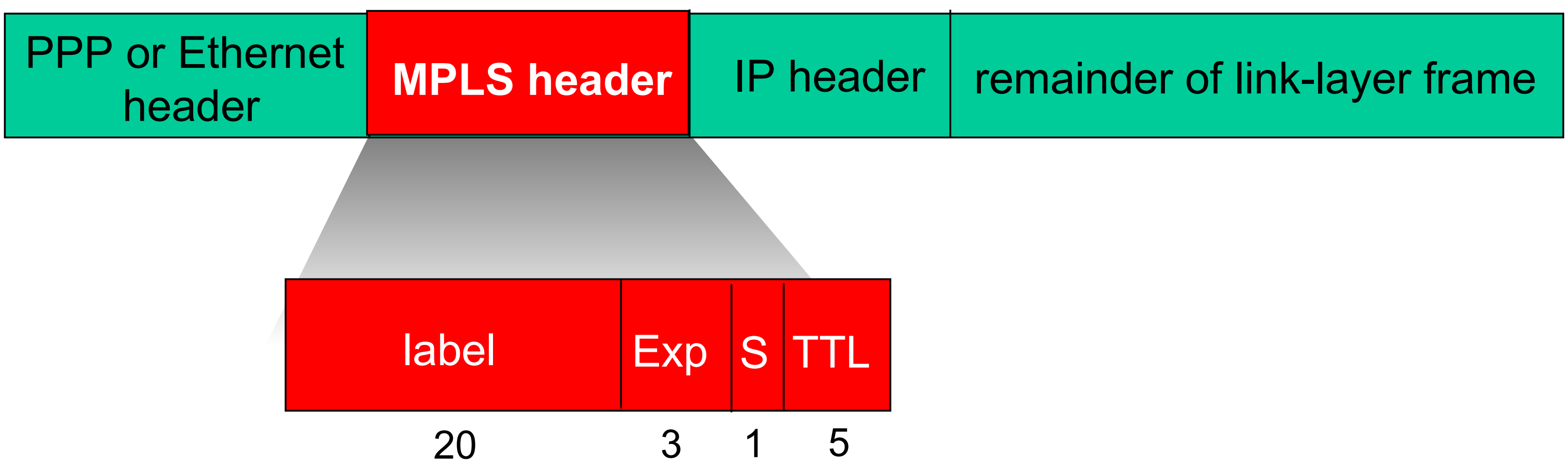
Datagram Journey in IP-over-ATM Network

- at entry router:
  - maps between IP destination address and ATM destination address (using ARP)
  - passes datagram to AAL5
  - AAL5 encapsulates data, segments cells, passes to ATM layer
- ATM network: moves cell along VC to destination
- at exit router:
  - AAL5 reassembles cells into original datagram
  - if CRC OK, datagram is passed to IP



Multiprotocol label switching (MPLS)

- initial goal: speed up IP forwarding by using fixed length label (instead of IP address) to do forwarding
  - borrowing ideas from Virtual Circuit (VC) approach
  - but IP datagram still keeps IP address!



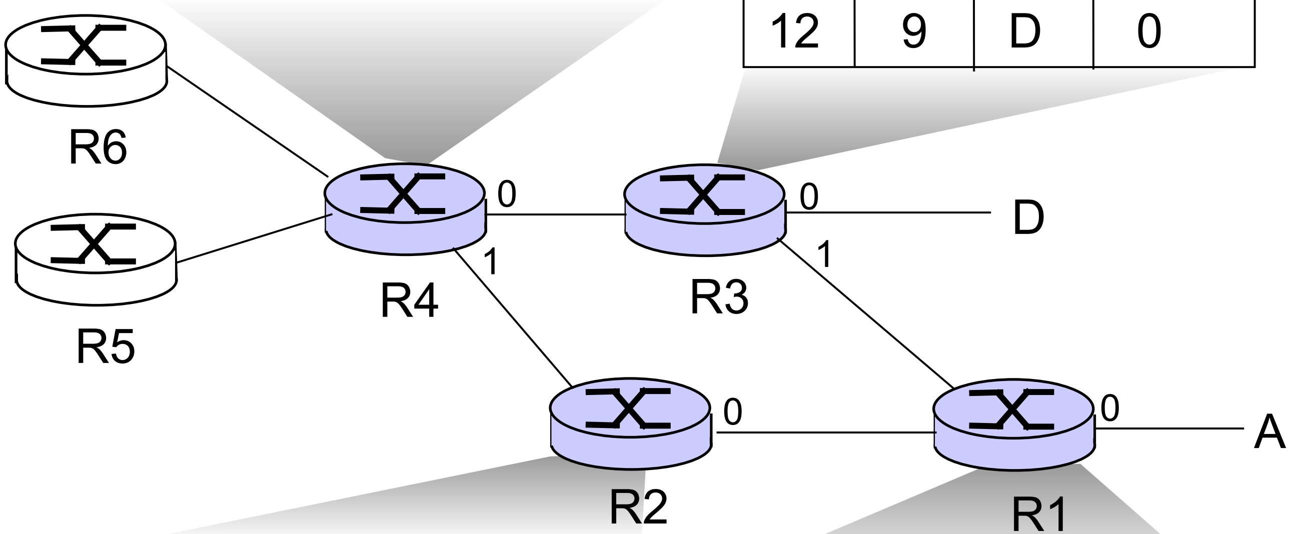
MPLS capable routers

- a.k.a. label-switched router
- forwards packets to outgoing interface based only on label value (don't inspect IP address)
  - MPLS forwarding table distinct from IP forwarding tables
- signaling protocol needed to set up forwarding table
  - RSVP-TE (RFC 3209)
  - forwarding possible along paths that IP alone would not allow (e.g., source-specific routing) !!
  - use MPLS for traffic engineering
- must co-exist with IP-only routers

MPLS forwarding tables

in label	out label	dest	out interface
	10	A	0
	12	D	0
	8	A	1

in label	out label	dest	out interface
10	6	A	1
12	9	D	0



in label	out label	dest	out interface
8	6	A	0

in label	out label	dest	out interface
6	-	A	0

Chapter 5: Summary

- principles behind data link layer services:
  - error detection, correction
  - sharing a broadcast channel: multiple access
  - link layer addressing
- instantiation and implementation of various link layer technologies
  - Ethernet
  - switched LANS
  - PPP
  - virtualized networks as a link layer: ATM, MPLS