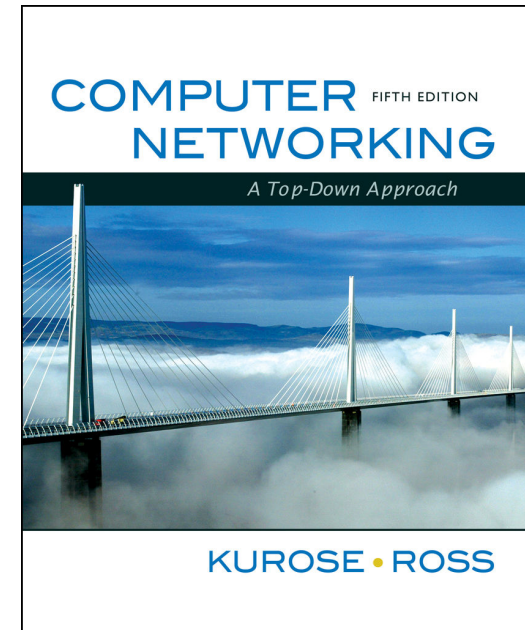


# Chapter 5

## Link Layer and LANs



### 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

All material copyright 1996-2009  
J.F Kurose and K.W. Ross, All Rights Reserved

*Computer Networking:  
A Top Down Approach  
5<sup>th</sup> edition.*

*Jim Kurose, Keith Ross  
Addison-Wesley, April  
2009.*

# 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

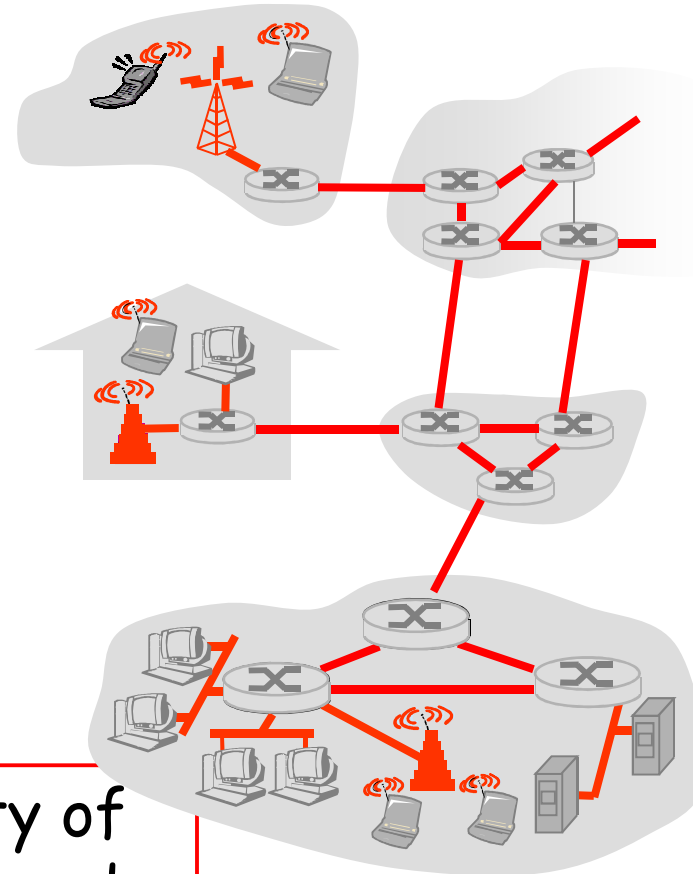
# Link Layer

- ❑ 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 Link-layer switches
- ❑ 5.7 PPP
- ❑ 5.8 Link virtualization: ATM, MPLS

# 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:*

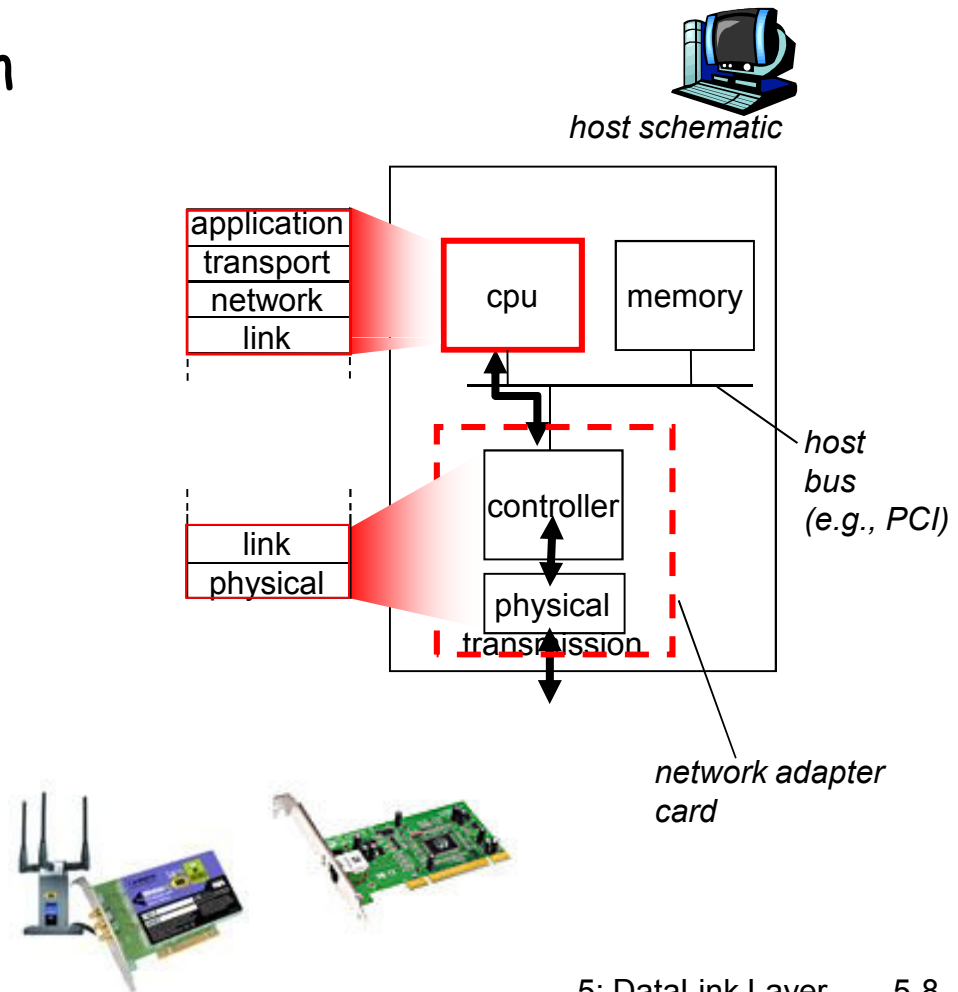
- receiver identifies *and corrects* bit error(s) without resorting to retransmission

## ❑ *half-duplex and full-duplex*

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

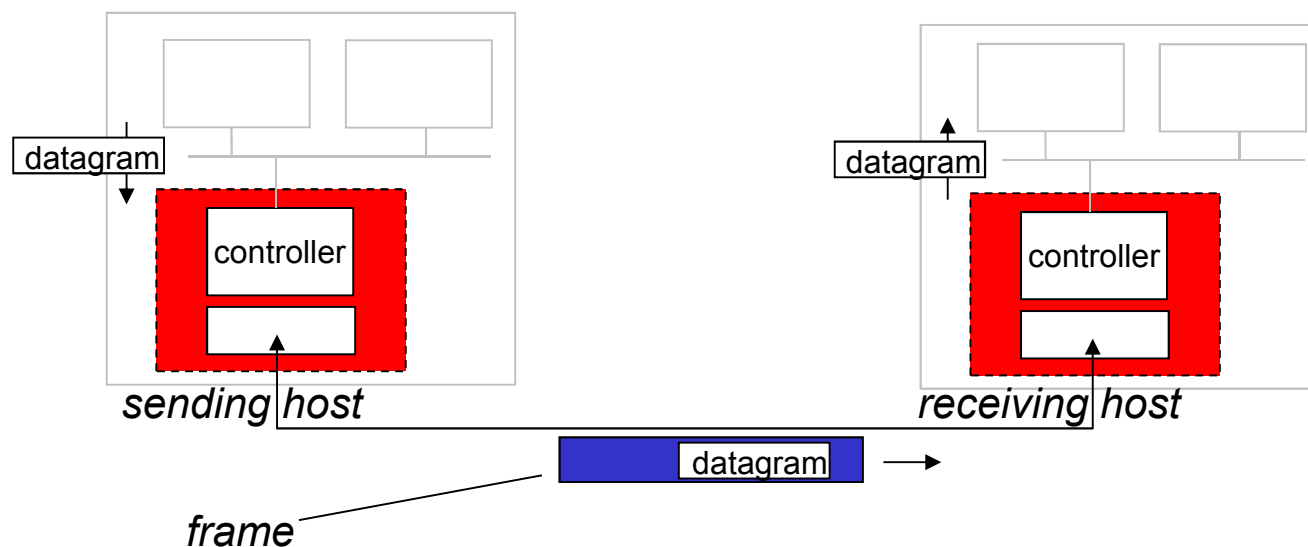
# Where is the link layer implemented?

- ❑ in each and every host
- ❑ link layer implemented in "adaptor" (aka *network interface card* NIC)
  - Ethernet card, PCMCIA card, 802.11 card
  - implements link, physical layer
- ❑ attaches into host's system buses
- ❑ combination of hardware, software, firmware





# Adaptors Communicating



## □ sending side:

- encapsulates datagram in frame
- adds error checking bits, rdt, flow control, etc.

## □ receiving side

- looks for errors, rdt, flow control, etc
- extracts datagram, passes to upper layer at receiving side

# Link Layer

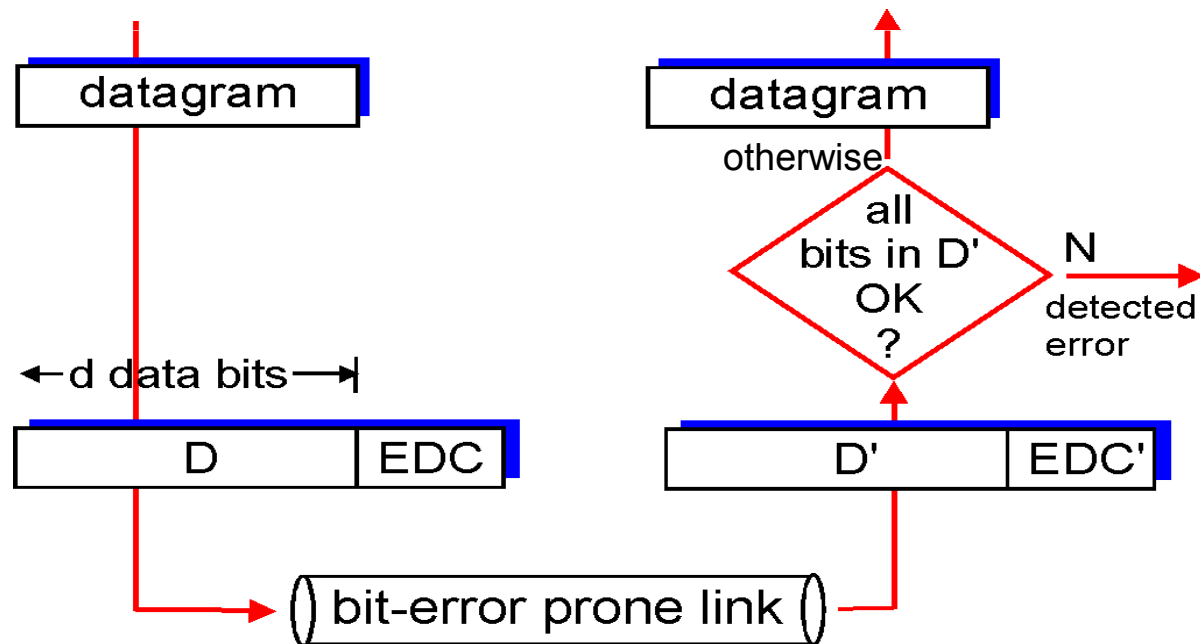
- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Link-layer switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM. MPLS

# Error Detection

EDC= Error Detection and Correction bits (redundancy)

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

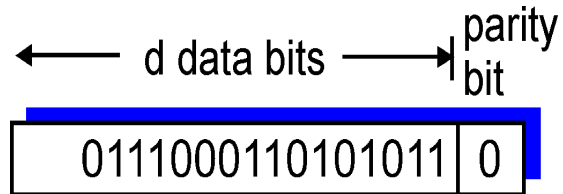
- Error detection not 100% reliable!
  - protocol may miss some errors, but rarely
  - larger EDC field yields better detection and correction



# Parity Checking

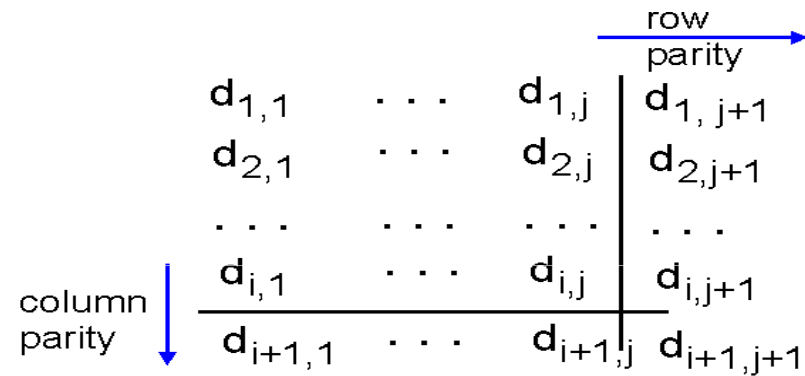
## Single Bit Parity:

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

parity error

*correctable  
single bit error*

# Internet checksum (review)

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

## Sender:

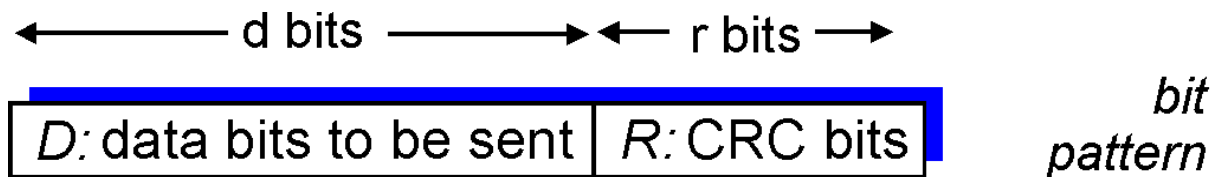
- ❑ treat segment contents as sequence of 16-bit integers
- ❑ checksum: addition (1's complement sum) of segment contents
- ❑ sender puts checksum value into UDP checksum field

## 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?*

# 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 (Ethernet, 802.11 WiFi, ATM)



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

*mathematical  
formula*



# Link Layer

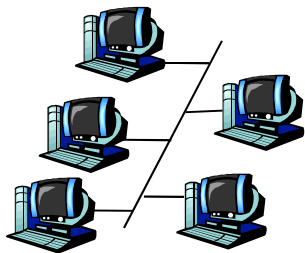
- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Link-layer switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM, MPLS



# Multiple Access Links and Protocols

Two types of "links":

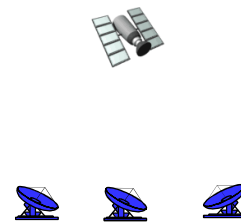
- ❑ point-to-point
  - PPP for dial-up access
  - point-to-point link between Ethernet switch and host
- ❑ **broadcast** (shared wire or medium)
  - old-fashioned Ethernet
  - upstream HFC
  - 802.11 wireless LAN



shared wire (e.g.,  
cabled Ethernet)



shared RF  
(e.g., 802.11 WiFi)



shared RF  
(satellite)



humans at a  
cocktail party  
(shared air, acoustical)

# 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 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

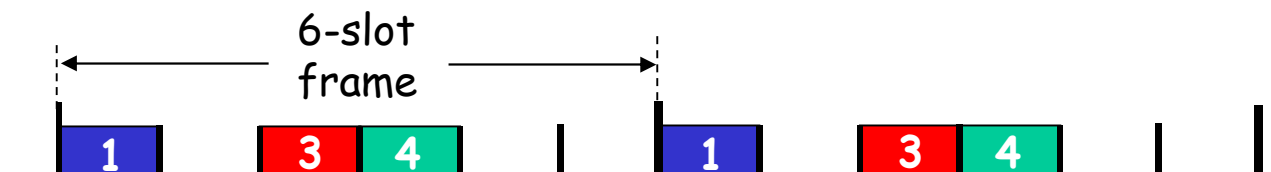
## ❑ "Taking turns"

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

# Channel Partitioning MAC protocols: TDMA

## TDMA: time division multiple access

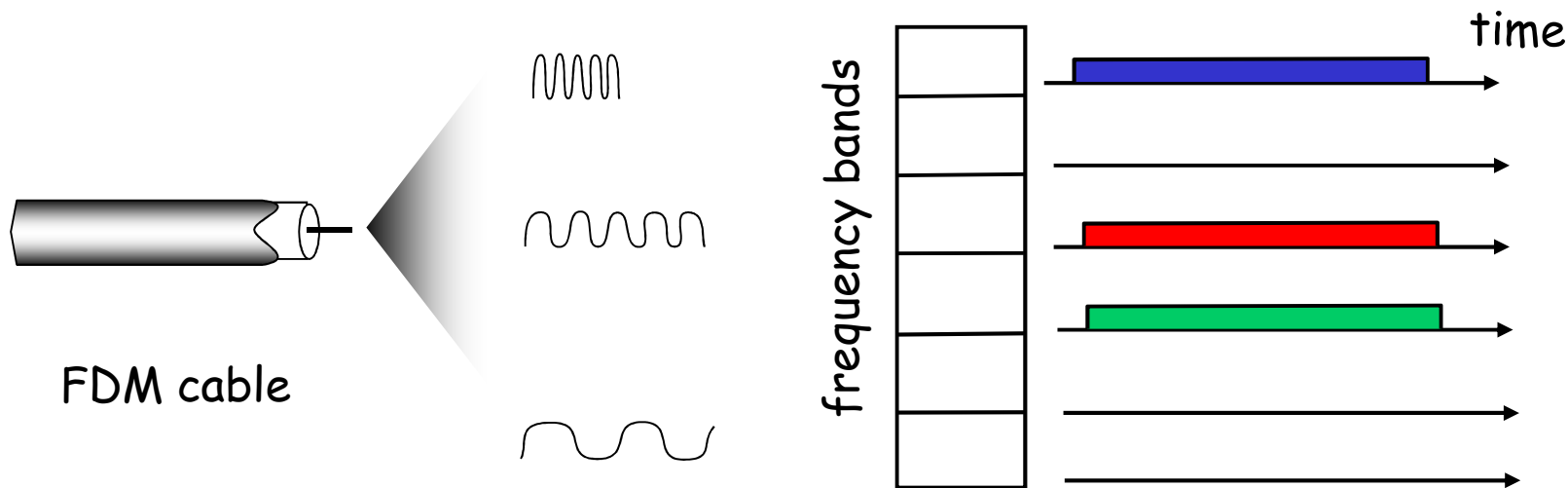
- ❑ access to channel in "rounds"
- ❑ each station gets fixed length slot (length = pkt trans time) in each round
- ❑ unused slots go idle
- ❑ example: 6-station LAN, 1,3,4 have pkt, slots 2,5,6 idle



# 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:

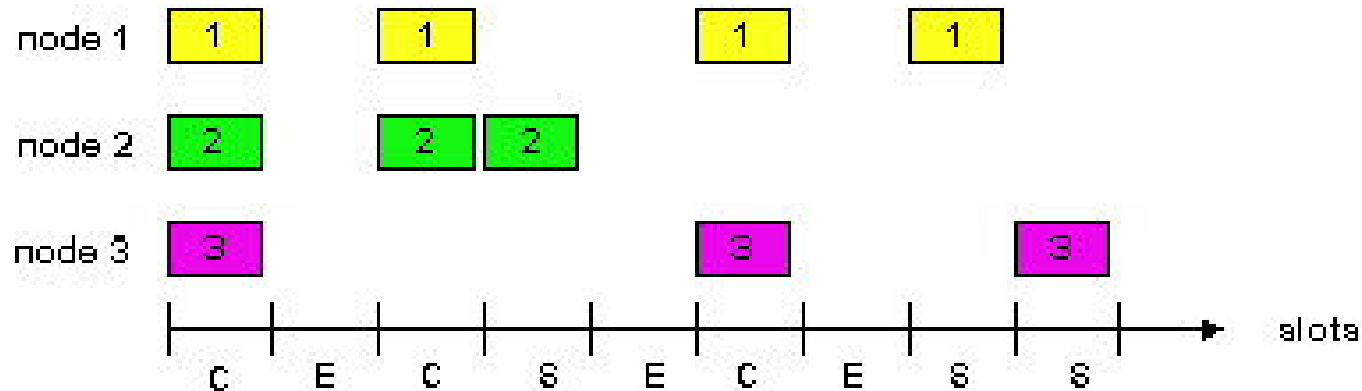
- ❑ all frames same size
- ❑ time divided into equal size slots (time to transmit 1 frame)
- ❑ nodes start to transmit only slot beginning
- ❑ nodes are synchronized
- ❑ if 2 or more nodes transmit in slot, all nodes detect collision

## Operation:

- ❑ when node obtains fresh frame, transmits in next slot
  - *if 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** : long-run fraction of successful slots (many nodes, all with many frames to send)

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

- max efficiency: 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:

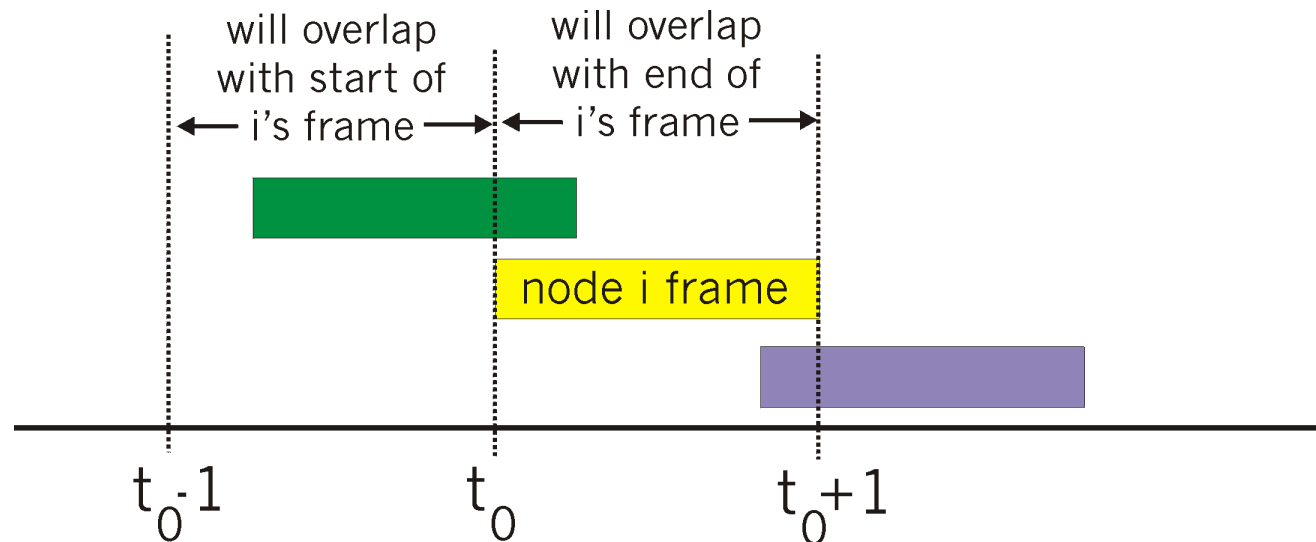
$$\text{Max efficiency} = 1/e = .37$$

***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

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

$P(\text{no other node transmits in } [p_0-1, p_0]) \cdot$

$P(\text{no other node transmits in } [p_0-1, p_0])$

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

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

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

$$= 1/(2e) = .18$$

*even worse than slotted Aloha!*

# CSMA (Carrier Sense Multiple Access)

**CSMA**: listen before transmit:

If channel sensed idle: transmit entire frame

❑ If channel sensed busy, defer transmission

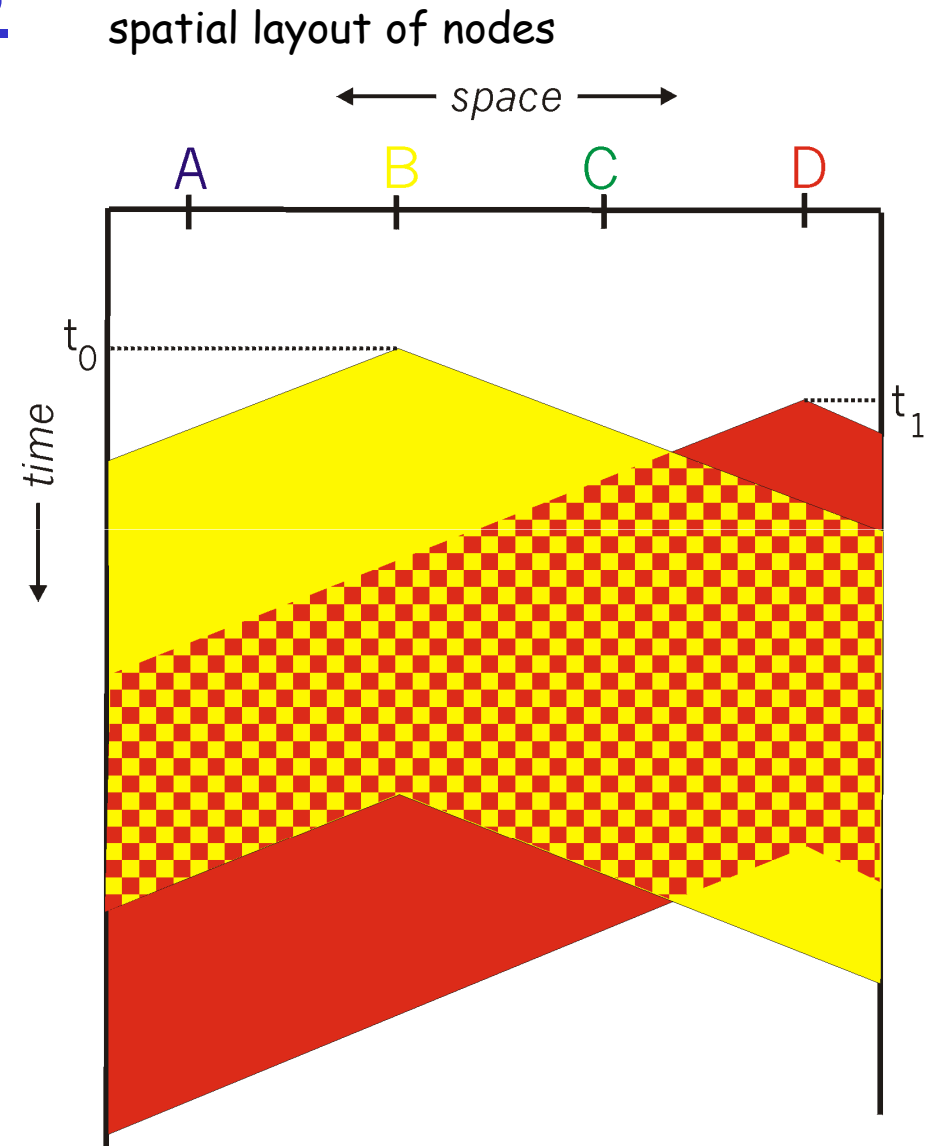
❑ human analogy: don't interrupt others!

# 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

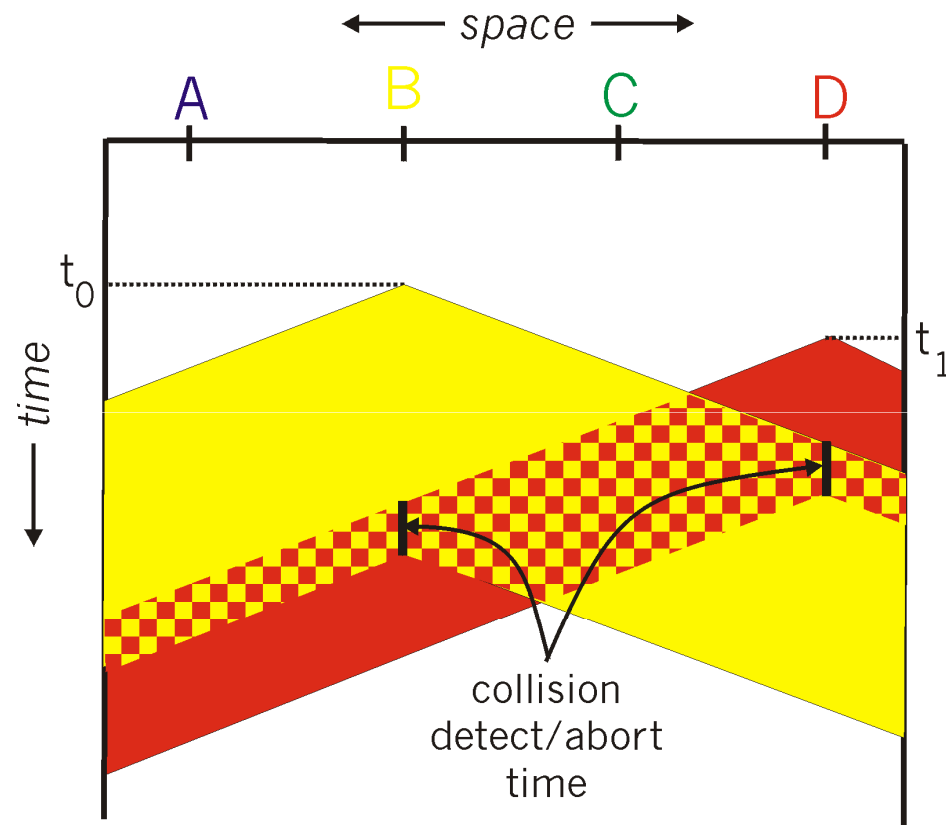


# CSMA/CD (Collision Detection)

**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: received signal strength overwhelmed by local transmission strength
- 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

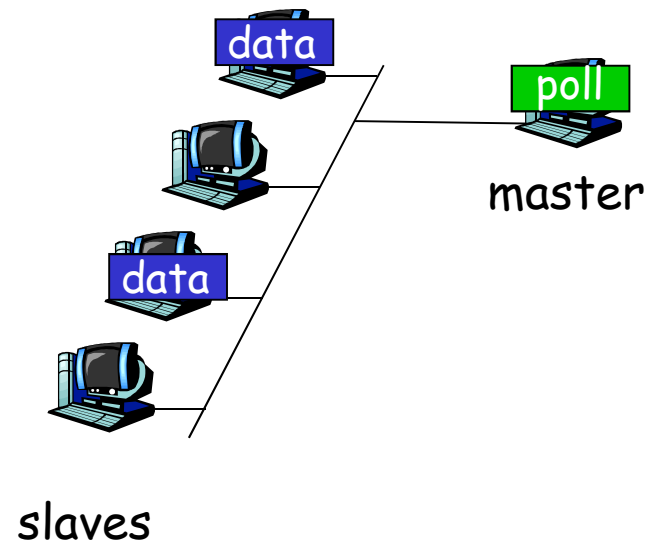
"taking turns" protocols

look for best of both worlds!

# "Taking Turns" MAC protocols

## Polling:

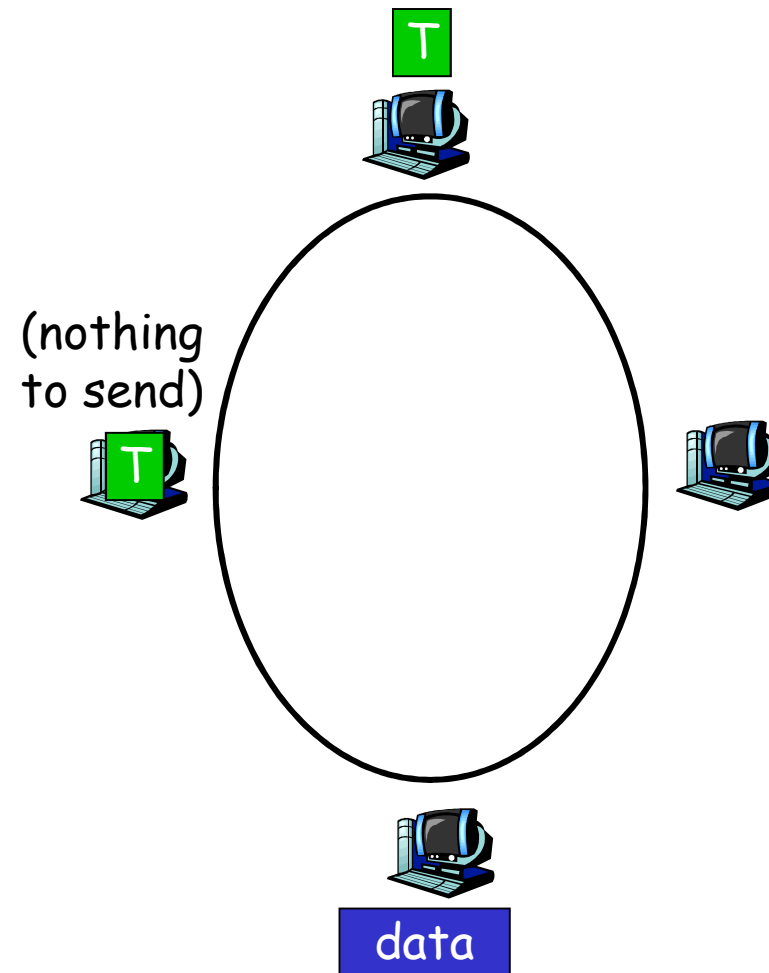
- ❑ master node
  - "invites" slave nodes to transmit in turn
- ❑ typically used with "dumb" slave devices
- ❑ concerns:
  - polling overhead
  - latency
  - single point of failure (master)



# "Taking Turns" MAC protocols

## 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

- ❑ *channel partitioning*, by time, frequency or code
  - Time Division, Frequency Division
- ❑ *random access* (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 central site, token passing
  - Bluetooth, FDDI, IBM Token Ring

# Link Layer

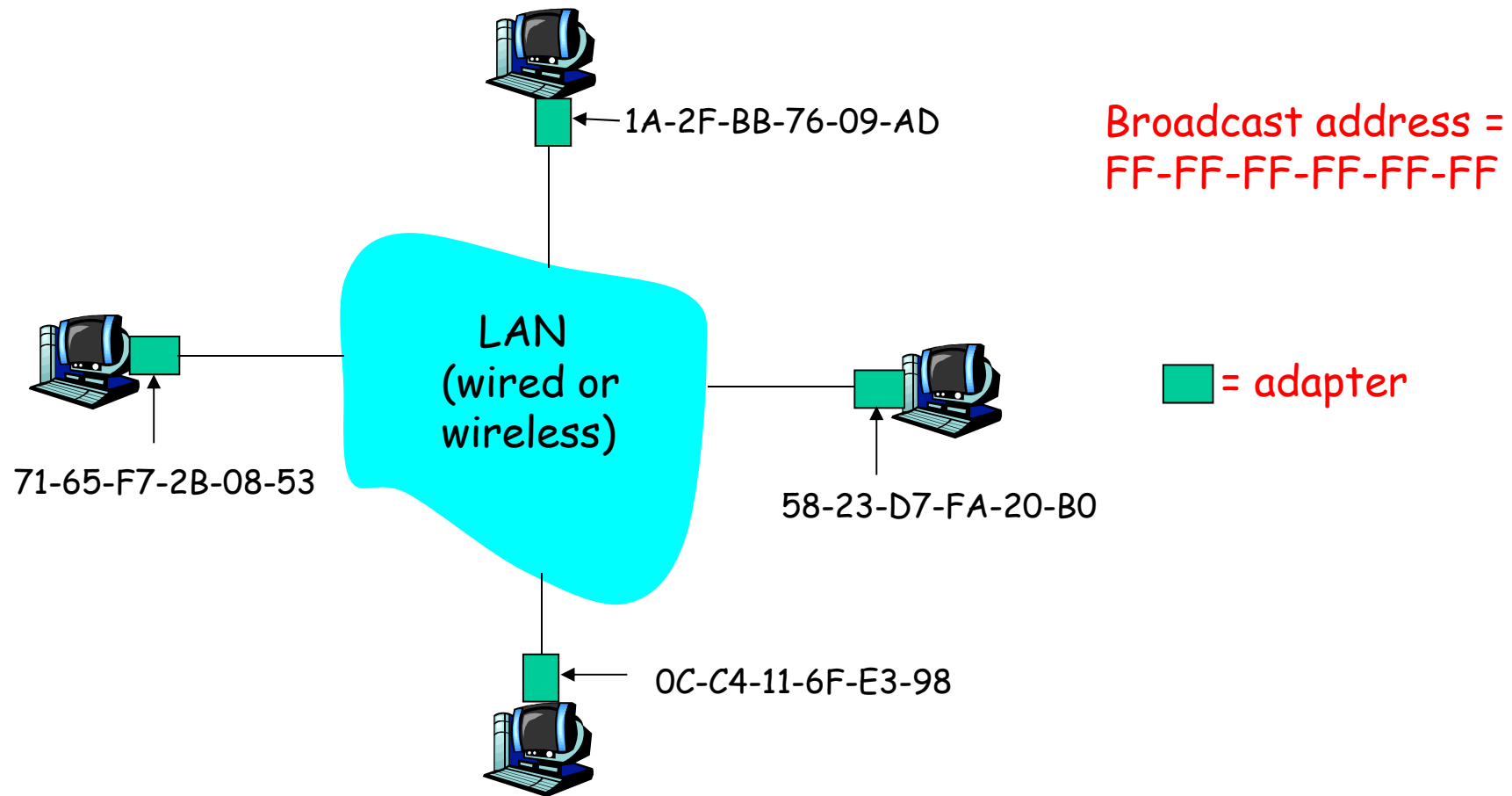
- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-Layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Link-layer switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM, MPLS

# 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:
  - function: *get frame from one interface to another physically-connected interface (same network)*
  - 48 bit MAC address (for most LANs)
    - burned in NIC ROM, also sometimes software settable

# LAN Addresses and ARP

Each adapter on LAN has unique LAN address



Broadcast address =  
FF-FF-FF-FF-FF-FF

■ = adapter

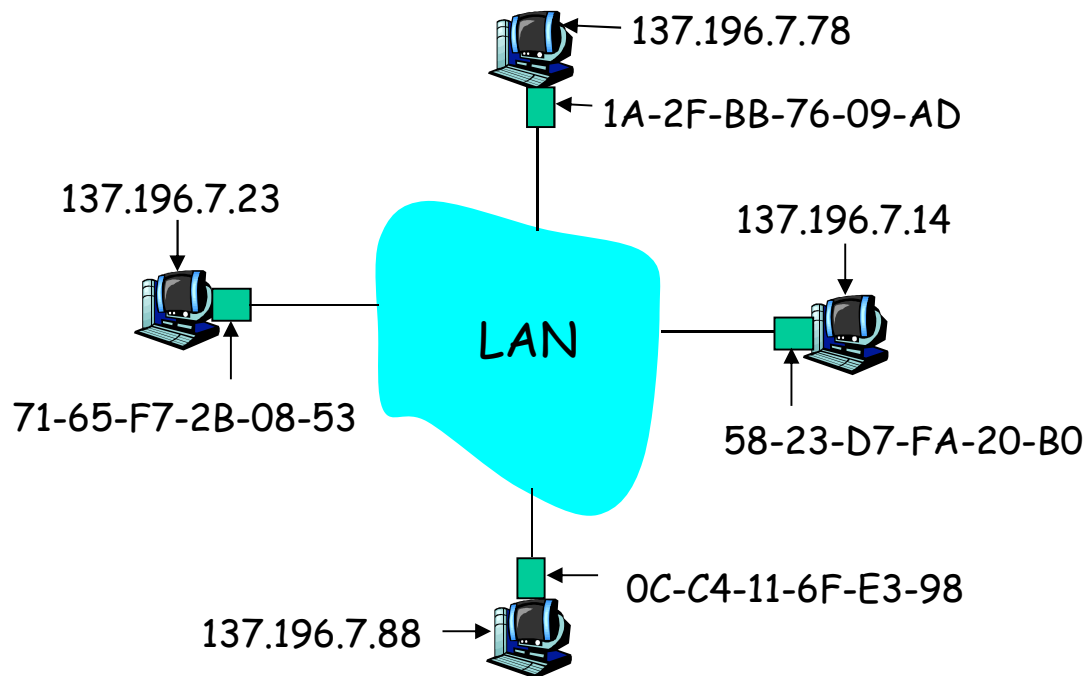
# LAN Address (more)

- ❑ 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
  - address 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 some LAN nodes
  - < IP address; MAC address; TTL >
    - TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min)

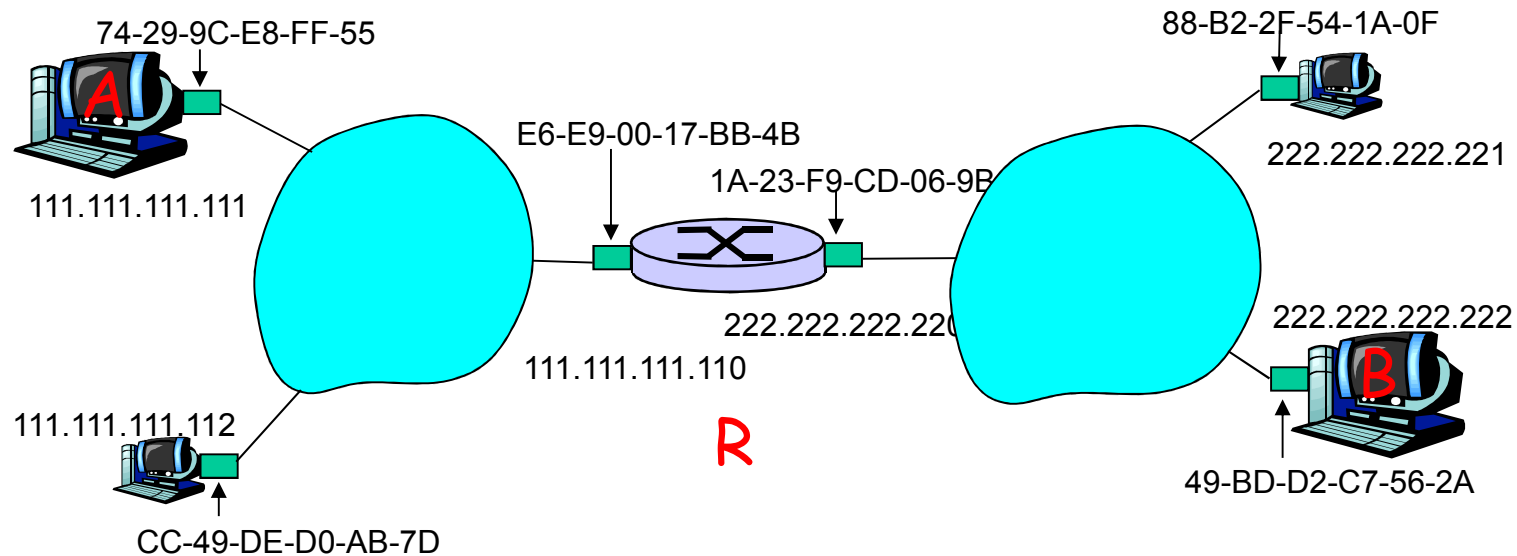
# ARP protocol: Same LAN (network)

- 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*

# Addressing: routing to another LAN

walkthrough: **send datagram from A to B via R**

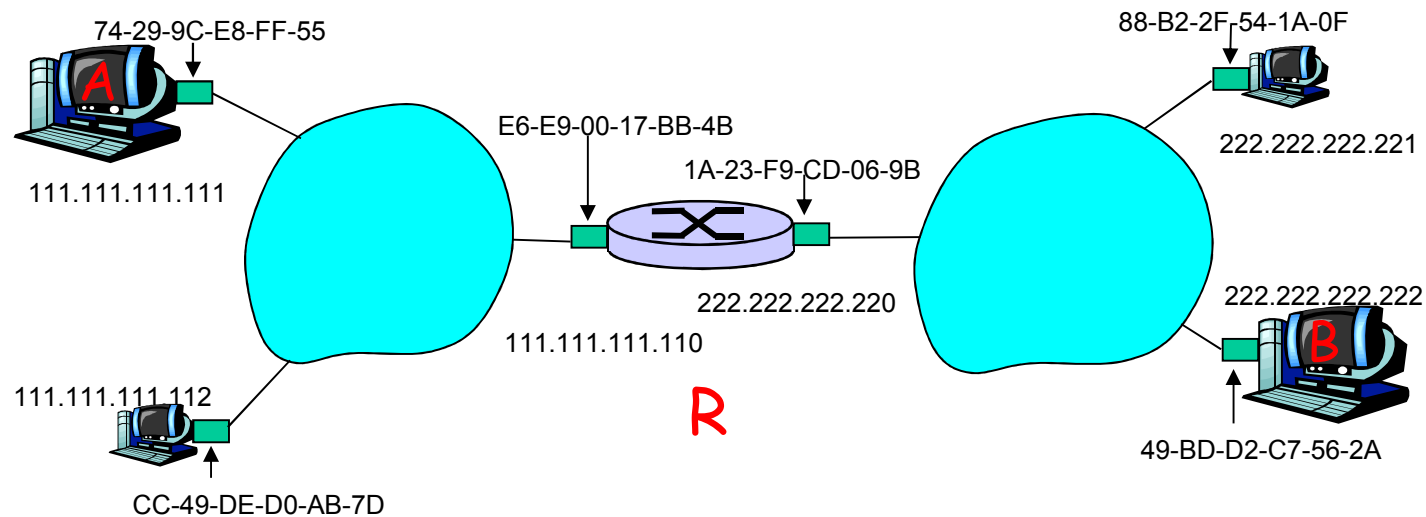
assume A knows B's IP address



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

- ❑ A creates IP 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 NIC sends frame
- ❑ R's NIC 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

This is a **really** important example - make sure you understand!



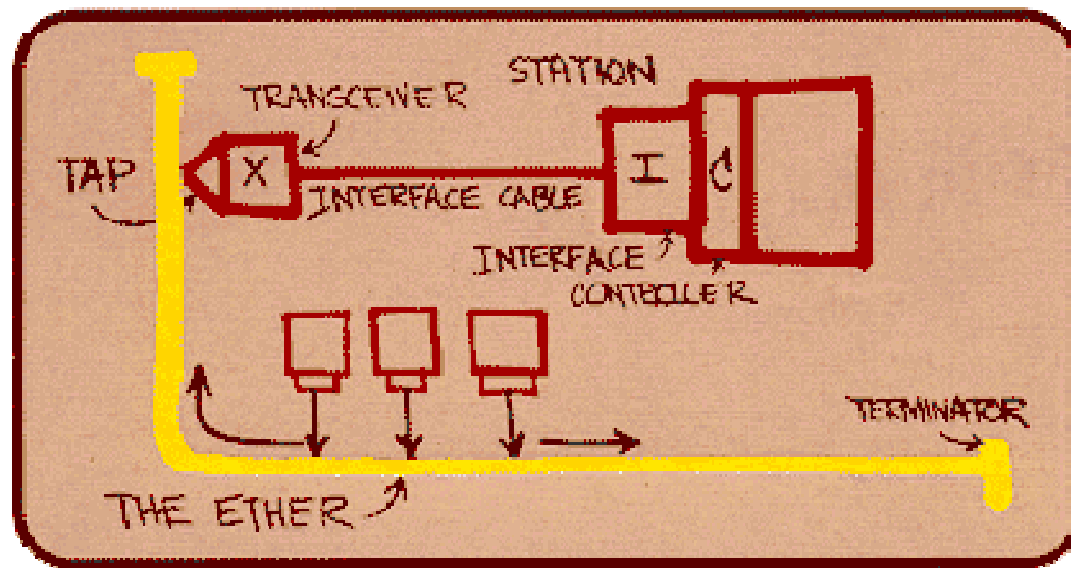
# Link Layer

- ❑ 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 Link-layer switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM and MPLS

# Ethernet

"dominant" wired LAN technology:

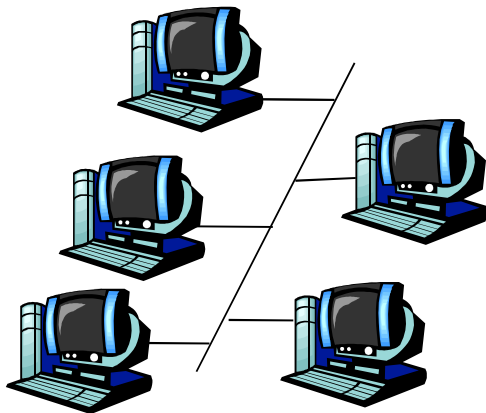
- ❑ cheap \$20 for NIC
- ❑ first widely used LAN technology
- ❑ simpler, cheaper than token LANs and ATM
- ❑ kept up with speed race: 10 Mbps - 10 Gbps



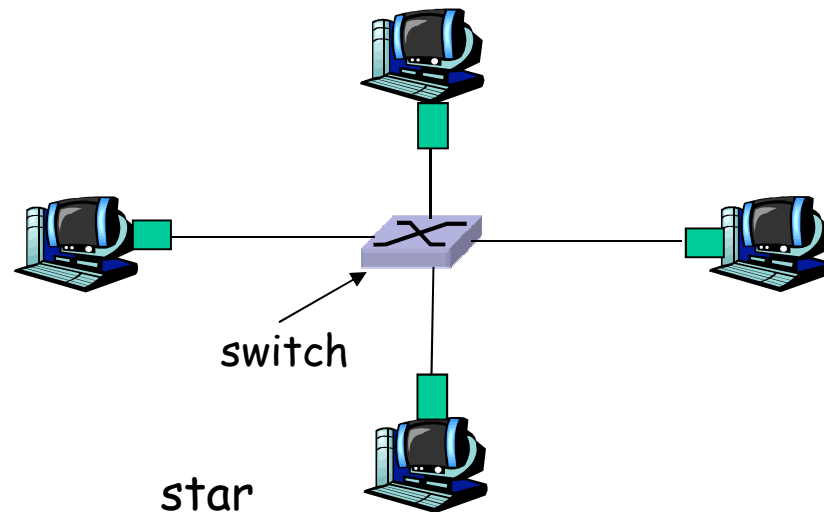
Metcalfe's Ethernet sketch

# Star topology

- ❑ bus topology popular through mid 90s
  - all nodes in same collision domain (can collide with each other)
- ❑ today: star topology prevails
  - active *switch* in center
  - each "spoke" runs a (separate) Ethernet protocol (nodes do not collide with each other)

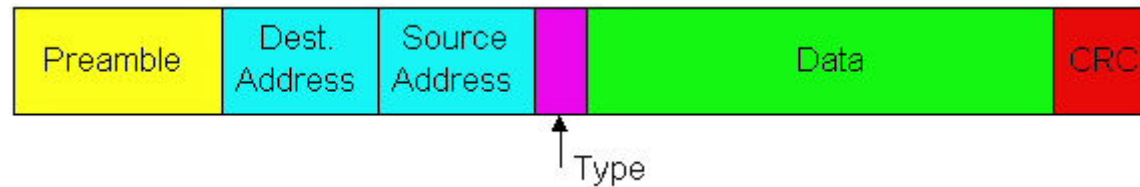


bus: coaxial cable



# Ethernet Frame Structure

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



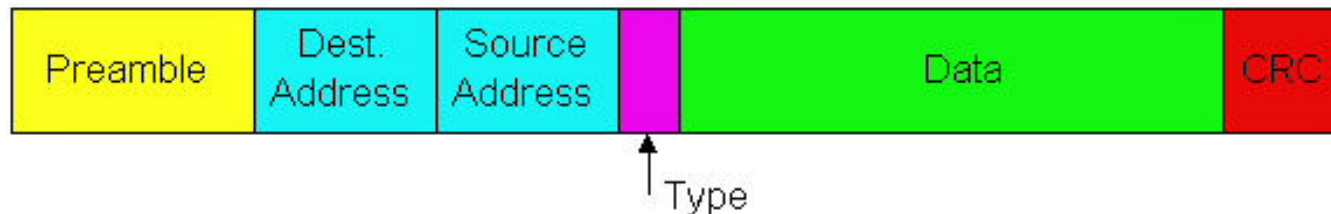
## Preamble:

- ❑ 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
  - if adapter receives frame with matching destination address, or with broadcast address (eg ARP packet), it passes data in frame to network layer protocol
  - otherwise, adapter discards frame
- ❑ **Type:** indicates higher layer protocol (mostly IP but others possible, e.g., Novell IPX, AppleTalk)
- ❑ **CRC:** checked at receiver, if error is detected, frame is dropped



# Ethernet: Unreliable, connectionless

- ❑ **connectionless**: No handshaking between sending and receiving NICs
- ❑ **unreliable**: receiving NIC doesn't send acks or nacks to sending NIC
  - stream of datagrams passed to network layer can have gaps (missing datagrams)
  - gaps will be filled if app is using TCP
  - otherwise, app will see gaps
- ❑ Ethernet's MAC protocol: unslotted **CSMA/CD**

# Ethernet CSMA/CD algorithm

1. NIC receives datagram from network layer, creates frame
2. If NIC senses channel idle, starts frame transmission  
If NIC senses channel busy, waits until channel idle, then transmits
3. If NIC transmits entire frame without detecting another transmission, NIC is done with frame !
4. If NIC detects another transmission while transmitting, aborts and sends jam signal
5. After aborting, NIC enters **exponential backoff**: after  $m$ th collision, NIC chooses  $K$  at random from  $\{0,1,2,\dots,2^m-1\}$ . NIC waits  $K \cdot 512$  bit times, 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

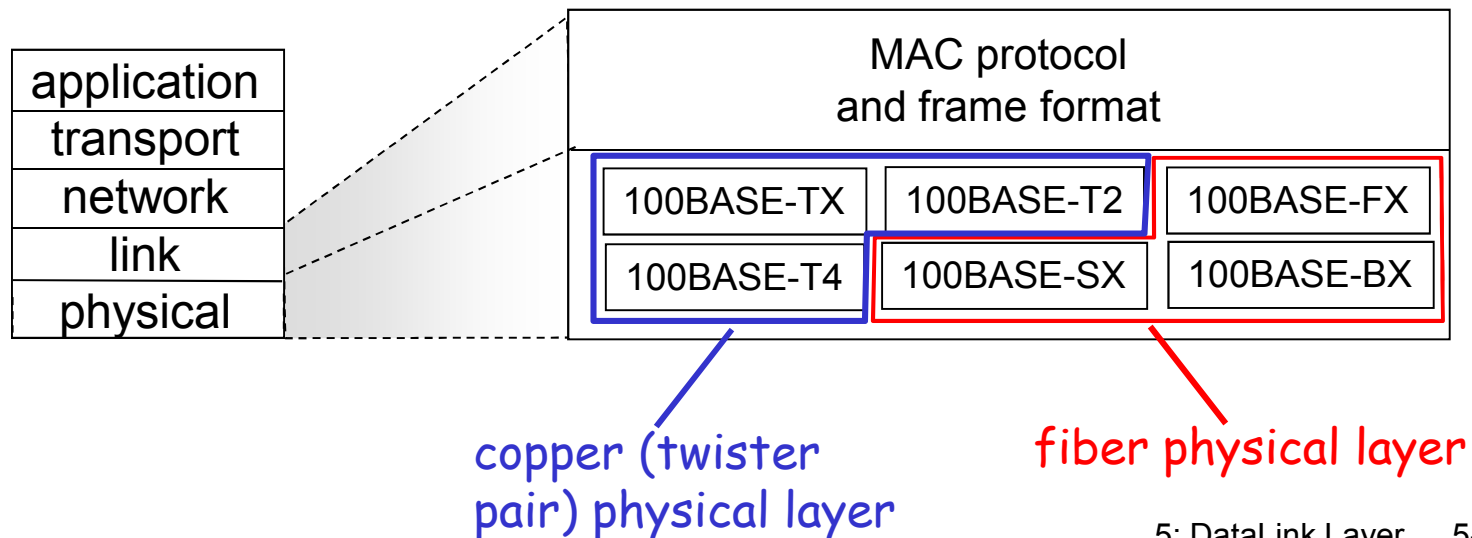
- $T_{prop}$  = max prop delay between 2 nodes in LAN
- $t_{trans}$  = time to transmit max-size frame

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

- efficiency goes to 1
  - as  $t_{prop}$  goes to 0
  - as  $t_{trans}$  goes to infinity
- better performance than ALOHA: and simple, cheap, decentralized!

# 802.3 Ethernet Standards: Link & Physical Layers

- *many* different Ethernet standards
  - common MAC protocol and frame format
  - different speeds: 2 Mbps, 10 Mbps, 100 Mbps, 1Gbps, 10G bps
  - different physical layer media: fiber, cable



# 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!

# Link Layer

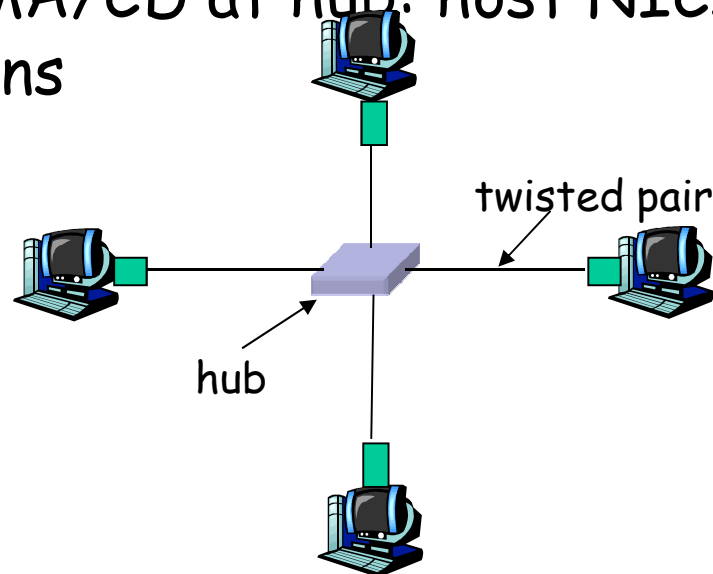
- ❑ 5.1 Introduction and services
- ❑ 5.2 Error detection and correction
- ❑ 5.3 Multiple access protocols
- ❑ 5.4 Link-layer Addressing
- ❑ 5.5 Ethernet
- ❑ 5.6 Link-layer switches
- ❑ 5.7 PPP
- ❑ 5.8 Link Virtualization: ATM, MPLS



# Hubs

... physical-layer ("dumb") repeaters:

- bits coming in one link go out *all* other links at same rate
- all nodes connected to hub can collide with one another
- no frame buffering
- no CSMA/CD at hub: host NICs detect collisions

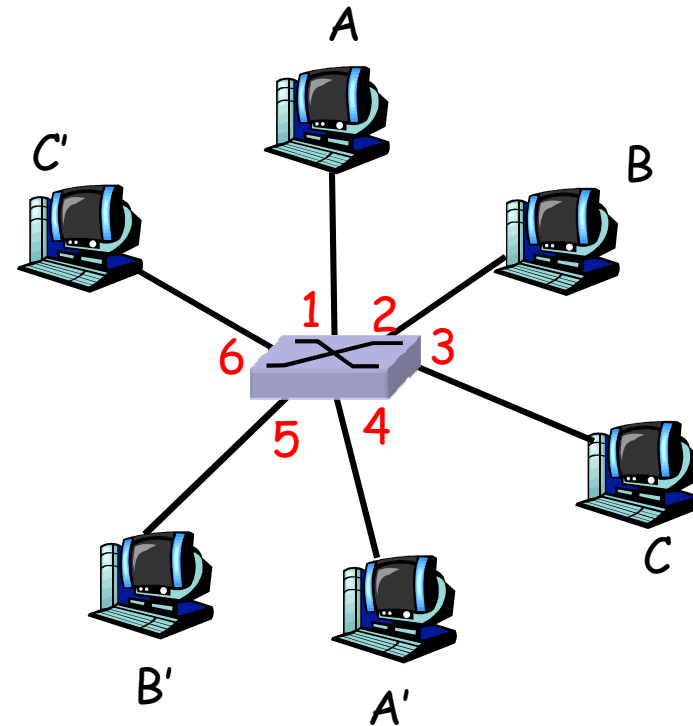


# Switch

- link-layer device: smarter than hubs, take *active role*
  - store, forward Ethernet frames
  - examine incoming frame's MAC address, *selectively* forward frame to one-or-more outgoing links 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

# Switch: allows multiple simultaneous transmissions

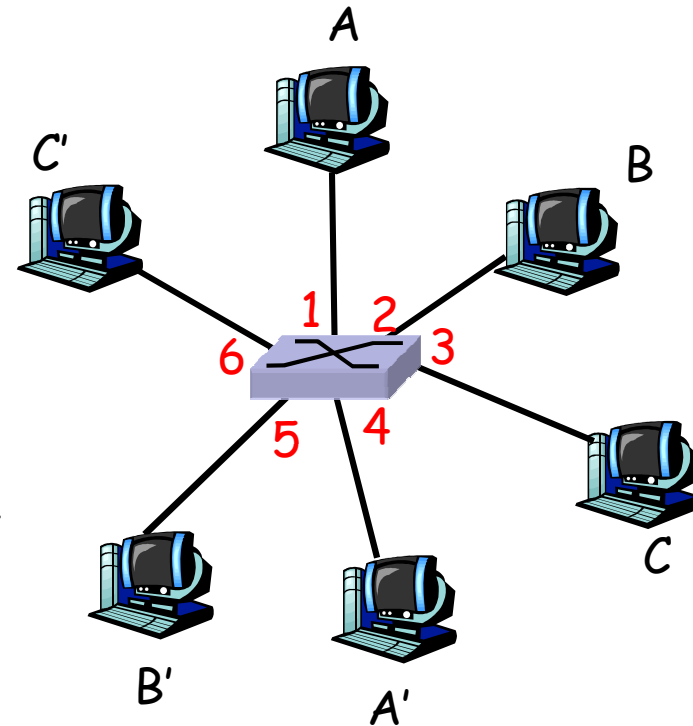
- ❑ hosts have dedicated, direct connection to switch
- ❑ switches buffer packets
- ❑ Ethernet protocol used on *each* incoming link, but no collisions; full duplex
  - each link is its own collision domain
- ❑ *switching*: A-to-A' and B-to-B' simultaneously, without collisions
  - not possible with dumb hub



*switch with six interfaces  
(1,2,3,4,5,6)*

# Switch Table

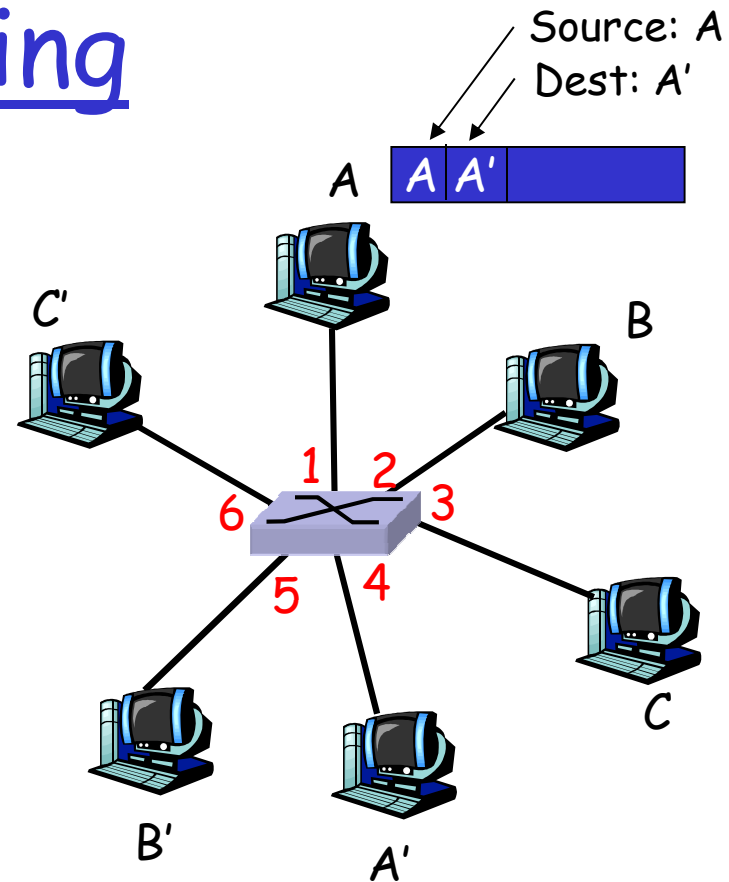
- ❑ **Q:** how does switch know that A' reachable via interface 4, B' reachable via interface 5?
- ❑ **A:** each switch has a **switch table**, each entry:
  - (MAC address of host, interface to reach host, time stamp)
- ❑ looks like a routing table!
- ❑ **Q:** how are entries created, maintained in switch table?
  - something like a routing protocol?



*switch with six interfaces  
(1,2,3,4,5,6)*

# Switch: self-learning

- 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



MAC addr	interface	TTL
A	1	60

*Switch table  
(initially empty)*

# Switch: frame filtering/forwarding

## When frame received:

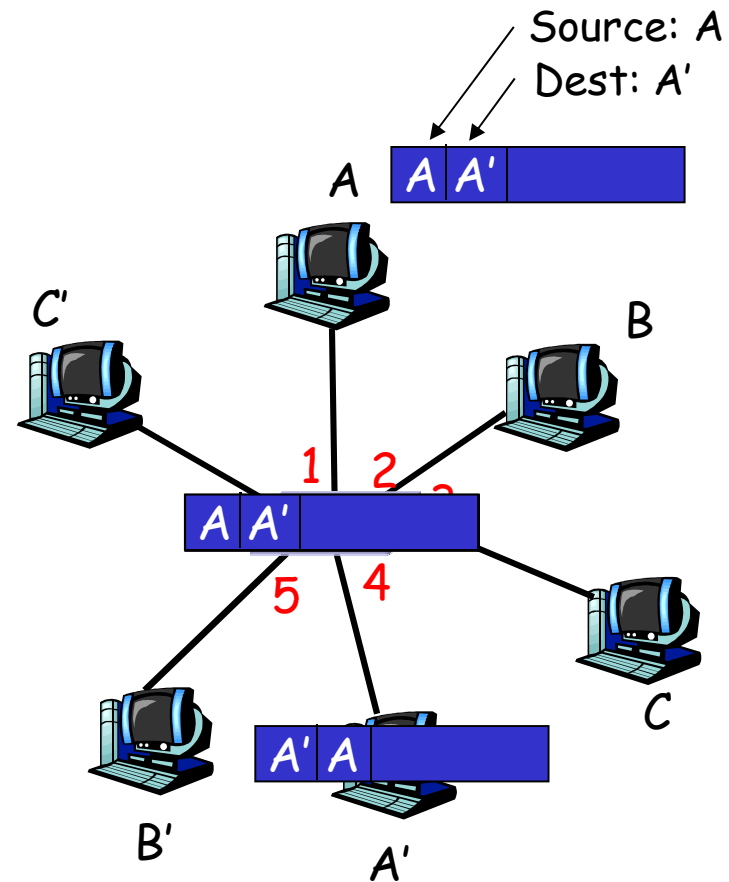
1. record link associated with sending host
2. index switch table using MAC dest address
3. **if** entry found for destination  
**then** {  
    **if** dest on segment from which frame arrived  
    **then** drop the frame  
    **else** forward the frame on interface indicated  
}

**else** flood

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

# Self-learning, forwarding: example

- ❑ frame destination unknown: *flood*
- ❑ destination A location known: *selective send*

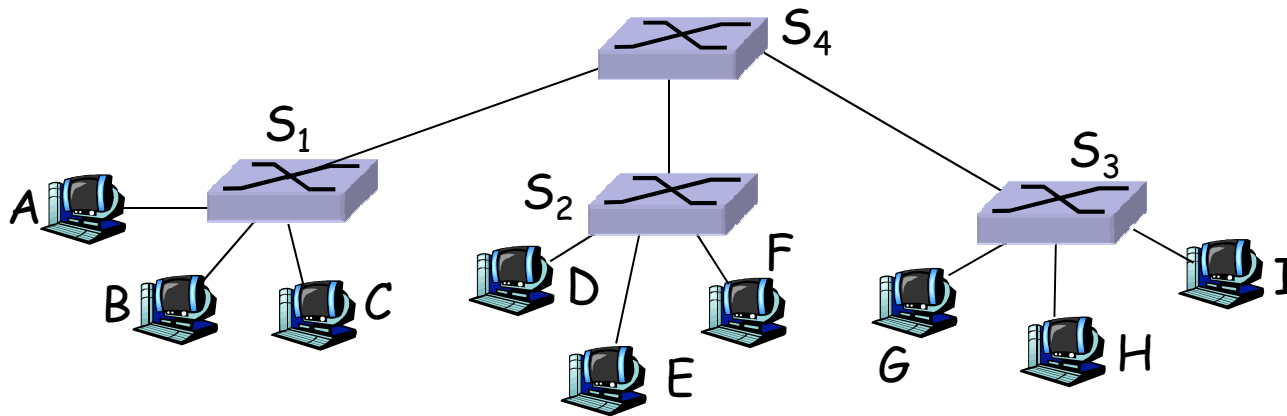


MAC addr	interface	TTL
A	1	60
A'	4	60

*Switch table  
(initially empty)*

# Interconnecting switches

- switches can be connected together

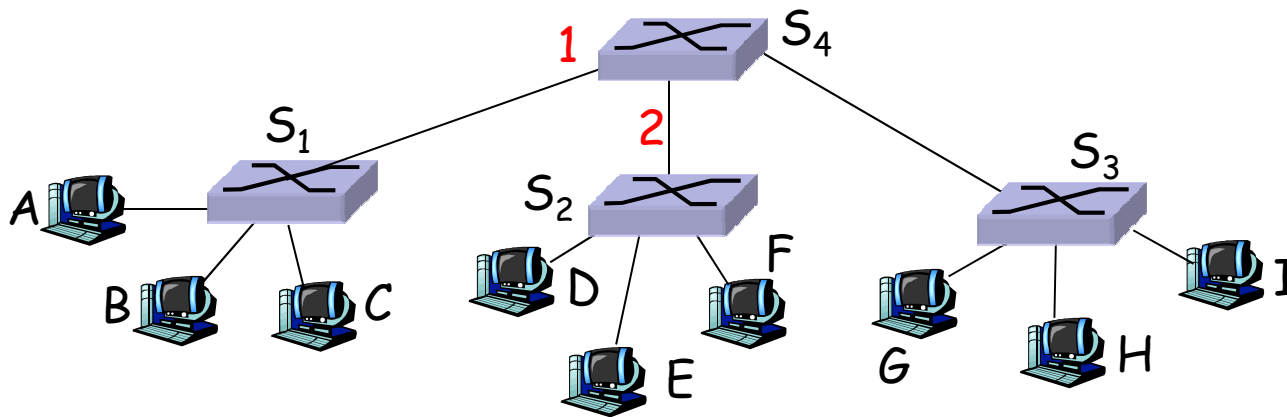


- Q: sending from A to G - how does S<sub>1</sub> know to forward frame destined to F via S<sub>4</sub> and S<sub>3</sub>?
- A: self learning! (works exactly the same as in single-switch case!)



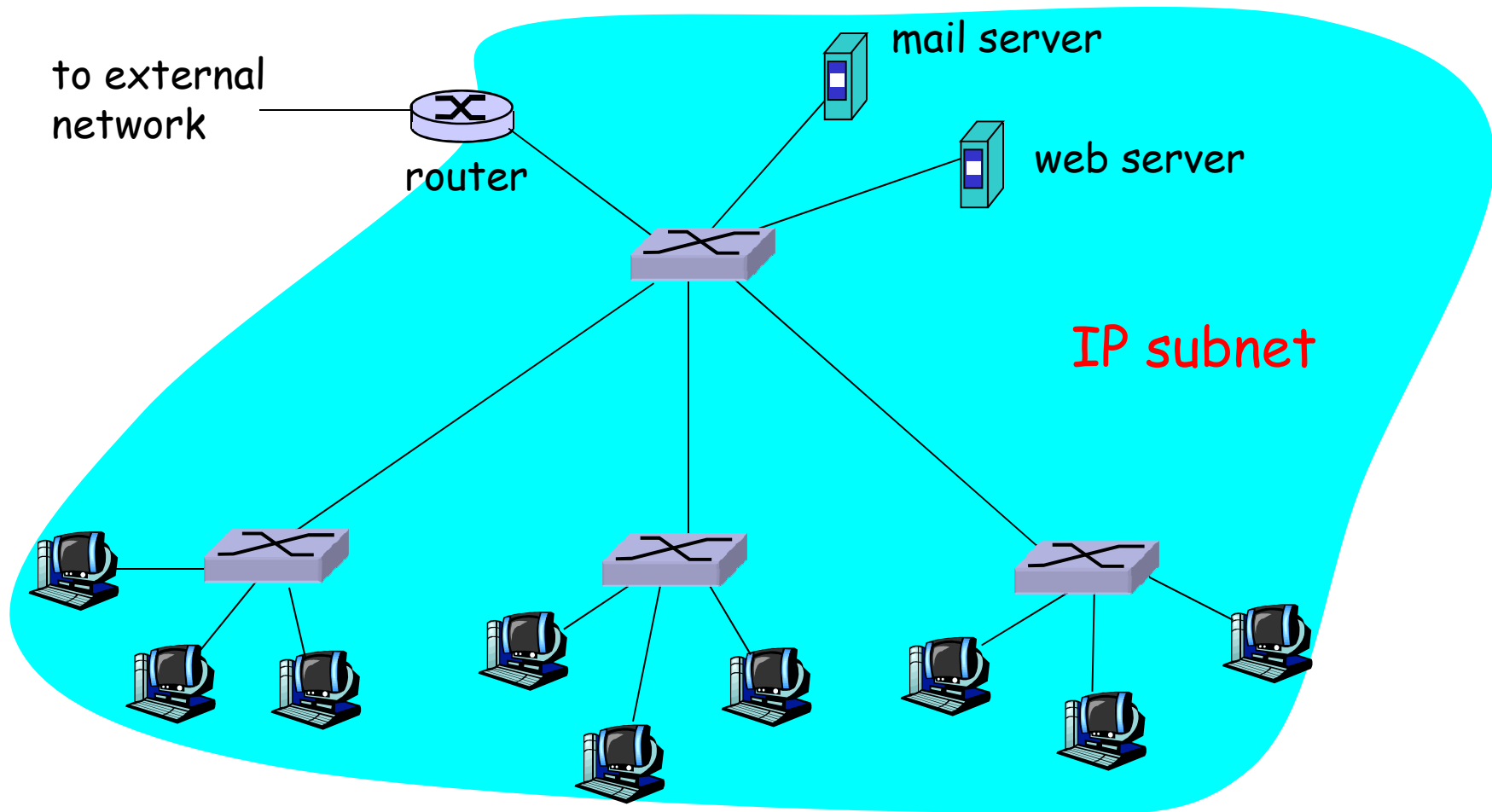
# Self-learning multi-switch example

Suppose C sends frame to I, I responds to C



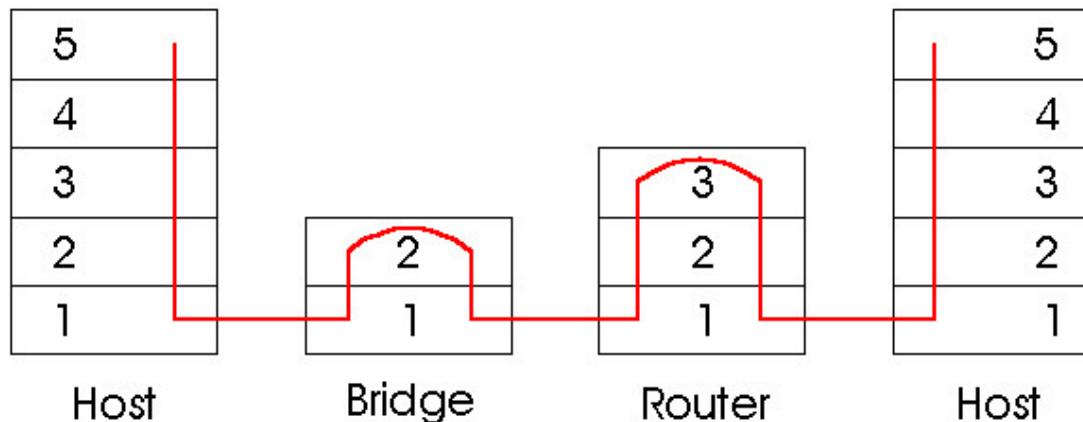
- Q: show switch tables and packet forwarding in S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>, S<sub>4</sub>

# 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



# 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

# 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 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, signal link failure to network layer
- ❑ **network layer address negotiation:** endpoint can learn/configure each other's network address

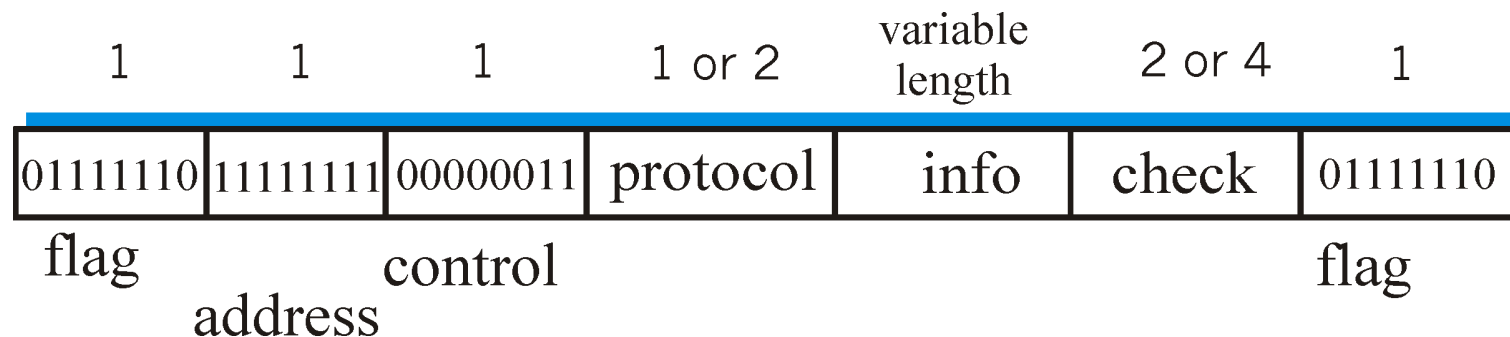
# PPP non-requirements

- ❑ 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

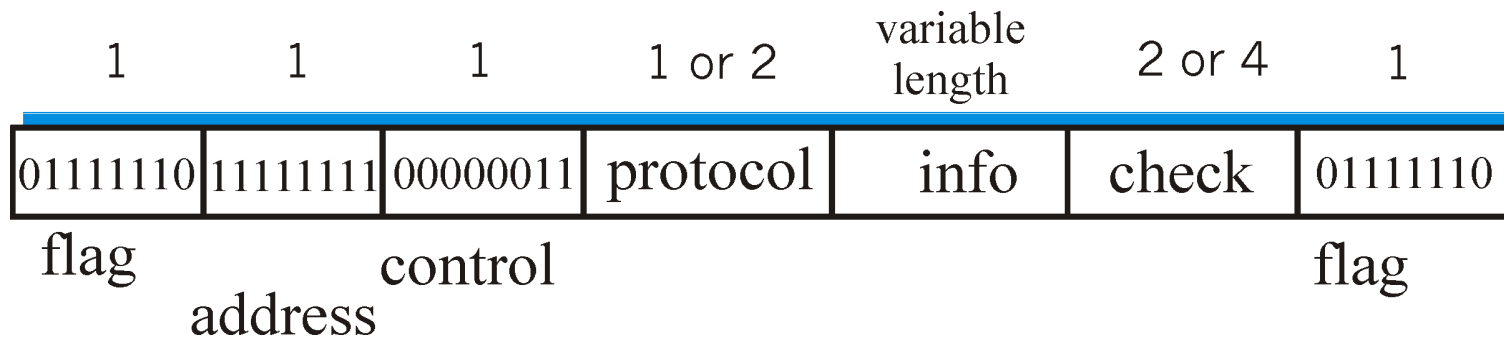
- ❑ **Flag:** delimiter (framing)
- ❑ **Address:** does nothing (only one option)
- ❑ **Control:** does nothing; in the future possible multiple control fields
- ❑ **Protocol:** upper layer protocol to which frame delivered (eg, PPP-LCP, IP, IPCP, etc)





# PPP Data Frame

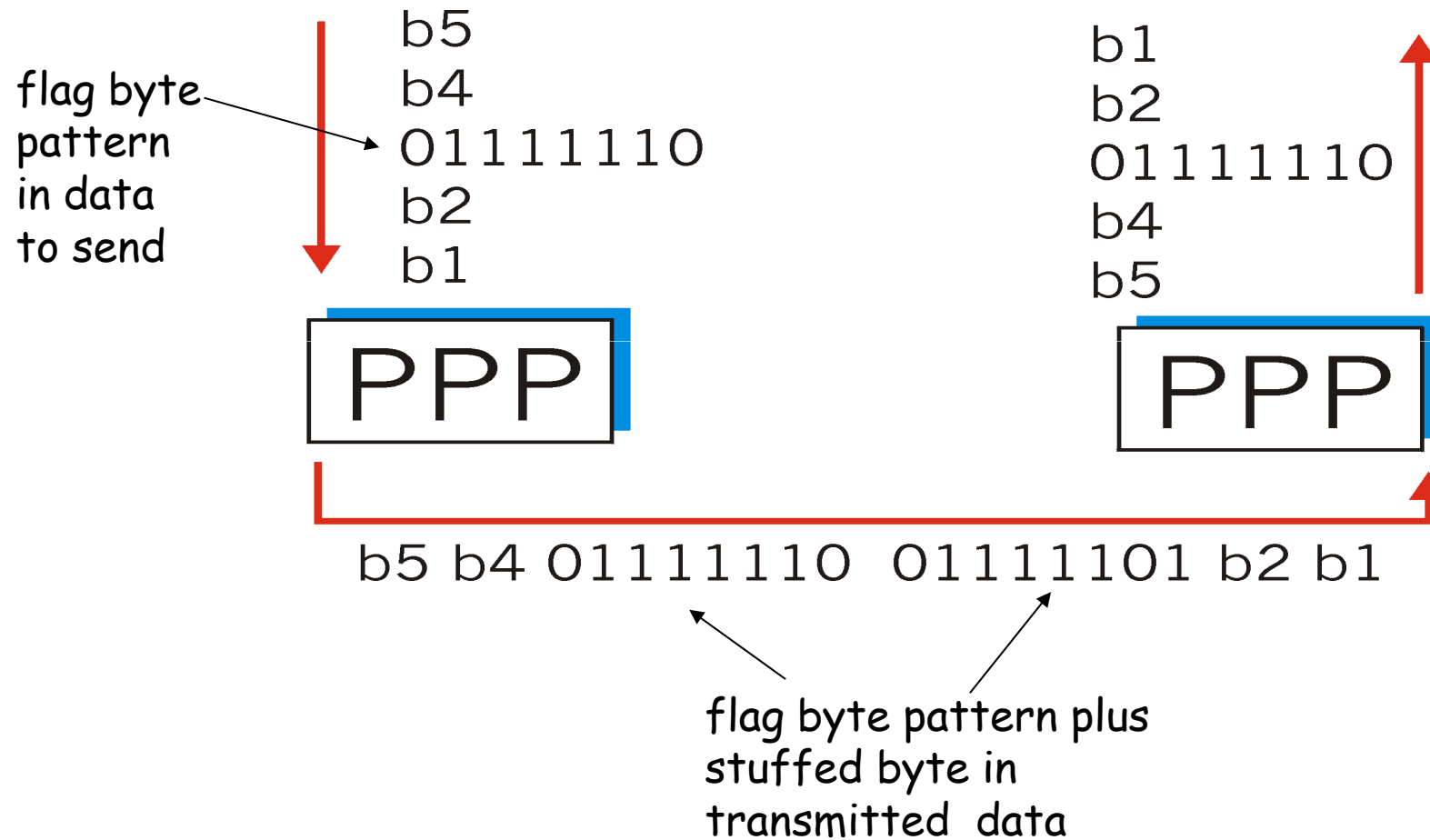
- **info**: upper layer data being carried
- **check**: cyclic redundancy check for error detection



# 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 < 01111110> byte after each < 01111110> *data* byte
- **Receiver**:
  - two 01111110 bytes in a row: discard first byte, continue data reception
  - single 01111110: flag byte

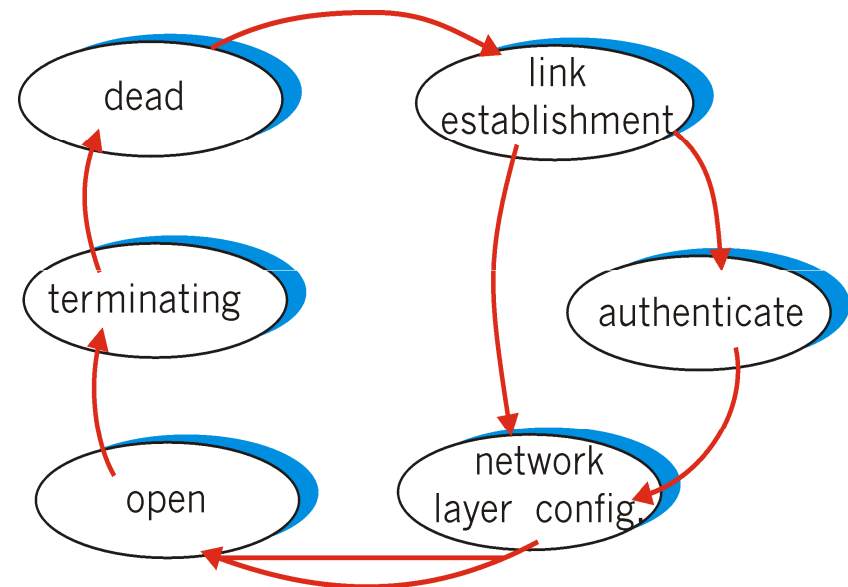
# Byte Stuffing



# PPP Data Control Protocol

Before exchanging network-layer data, data link peers must

- ❑ **configure PPP link** (max. frame length, authentication)
- ❑ **learn/configure network layer information**
  - for IP: carry IP Control Protocol (IPCP) msgs (protocol field: 8021) to configure/learn IP address



# 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 and MPLS

# Virtualization of networks

Virtualization of resources: 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

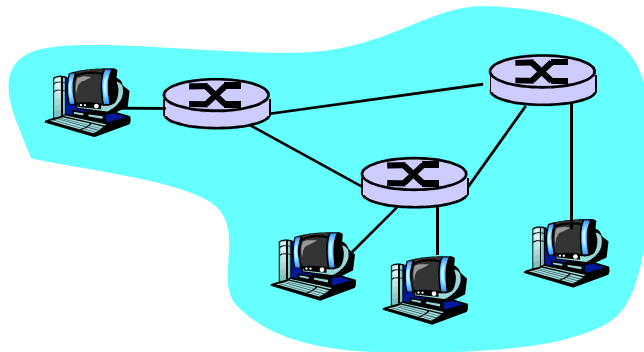
# The Internet: virtualizing networks

1974: multiple unconnected nets

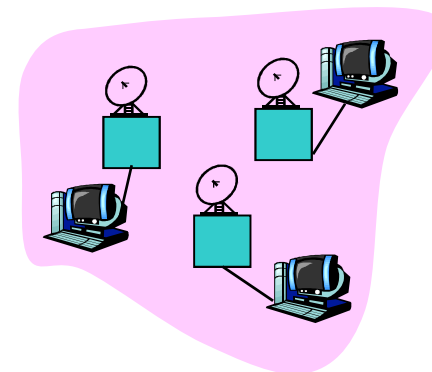
- ARPAnet
- data-over-cable networks
- packet satellite network (Aloha)
- packet radio network

... 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.

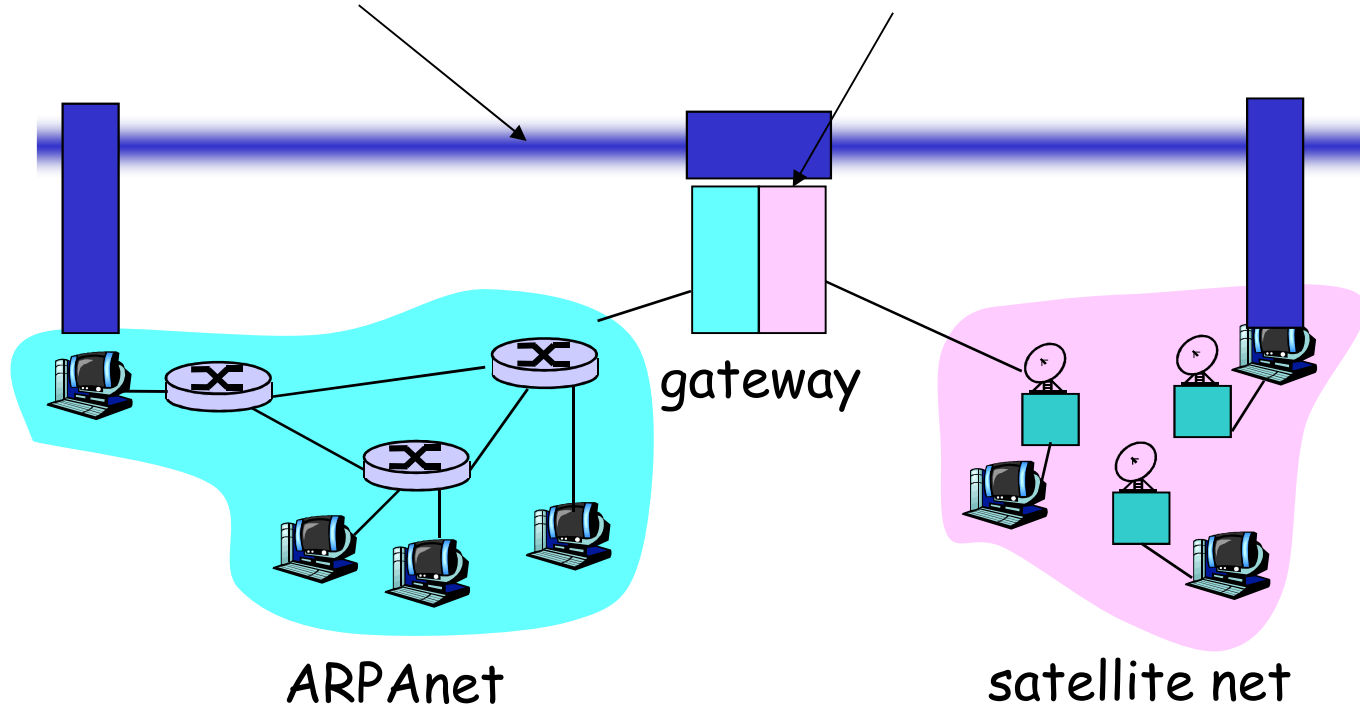
# The Internet: virtualizing networks

Internetwork layer (IP):

- addressing: internetwork appears as 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!

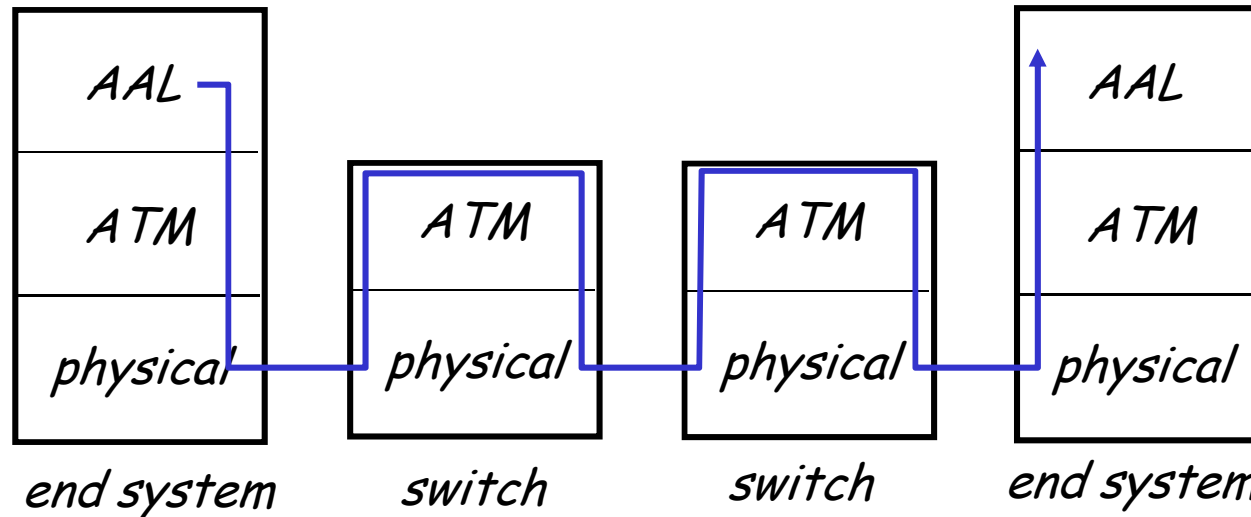
# ATM and MPLS

- ❑ 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, MPLS: 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-end transport of carry 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



- ❑ **adaptation layer:** only at edge of ATM network
  - data segmentation/reassembly
  - roughly analagous to Internet transport layer
- ❑ **ATM layer:** "network" layer
  - cell switching, routing
- ❑ **physical layer**

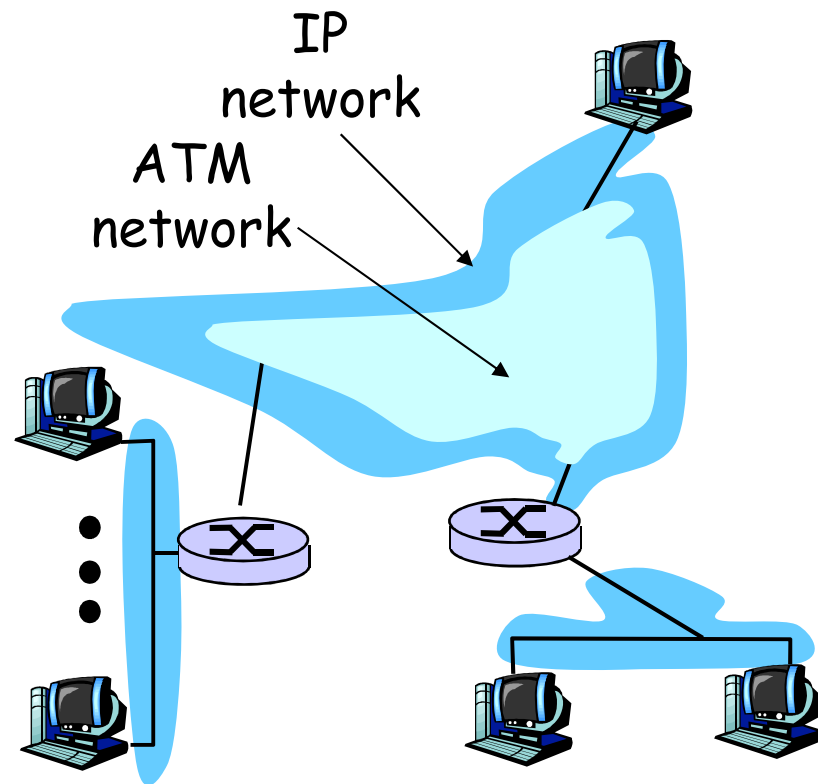
# ATM: network or link layer?

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

- ATM is a network technology

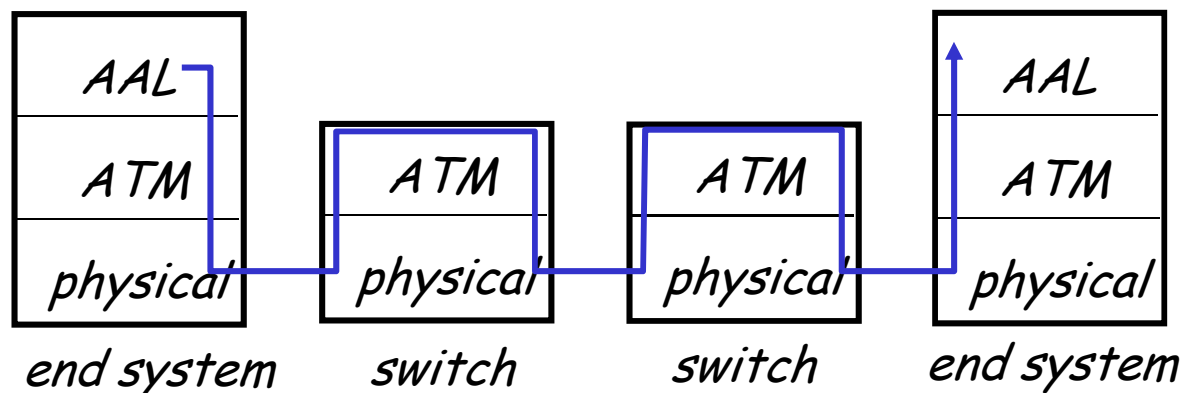
Reality: used to connect IP backbone routers

- "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) fragmented across multiple ATM cells
  - analogy: TCP segment in many IP packets



# ATM Adaptation Layer (AAL) [more]

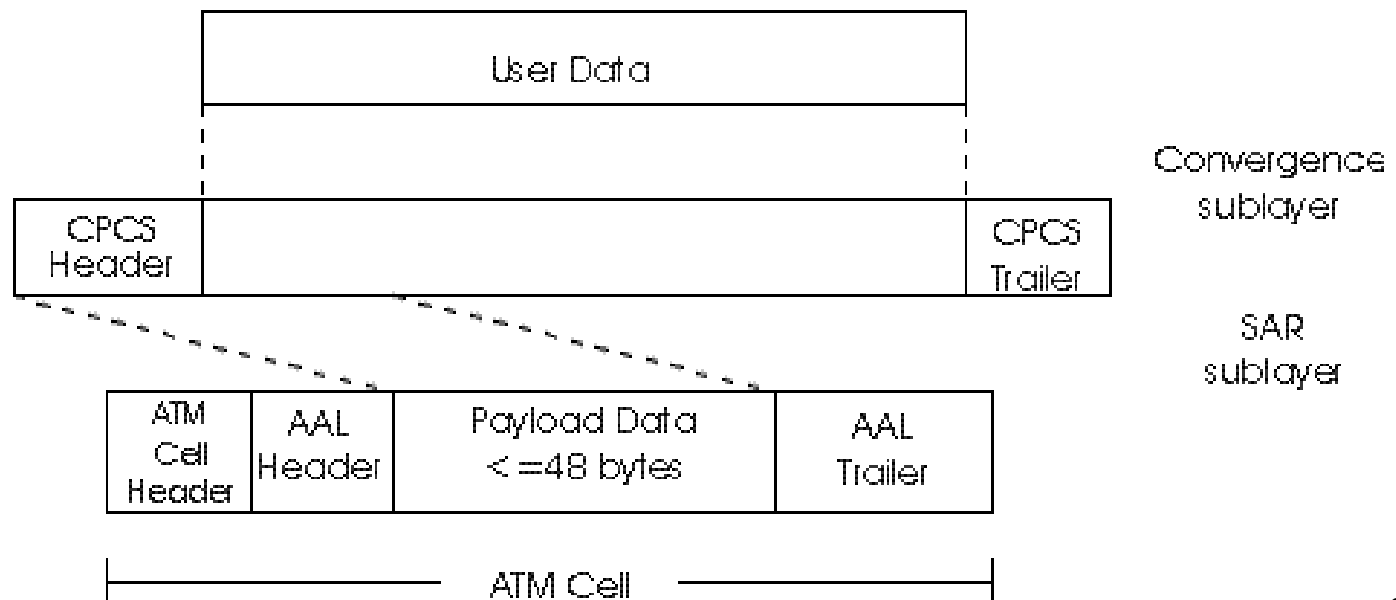
Different versions of AAL layers, depending on ATM service class:

- ❑ **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)

User data

AAL PDU

ATM cell



# 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 ?			Congestion feedback	
		Bandwidth	Loss	Order Timing		
Internet	best effort	none	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 Circuits

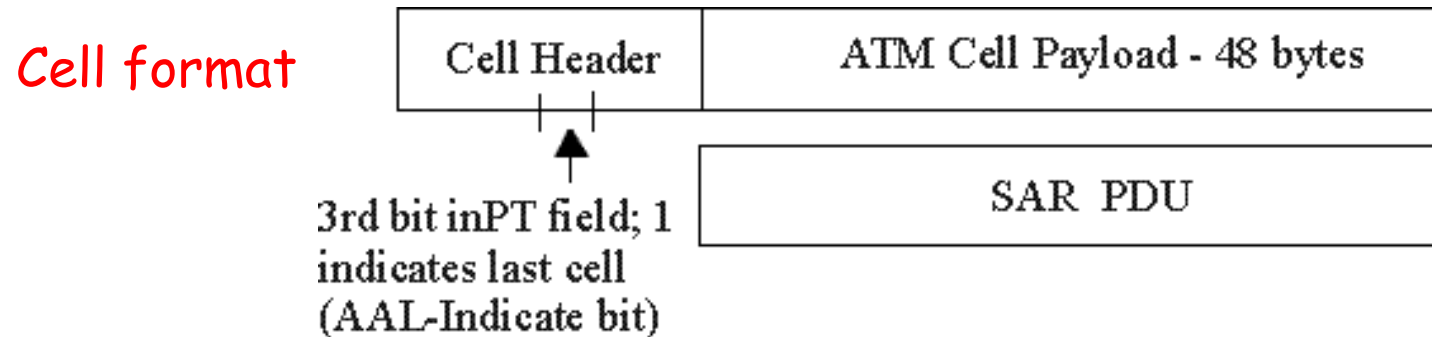
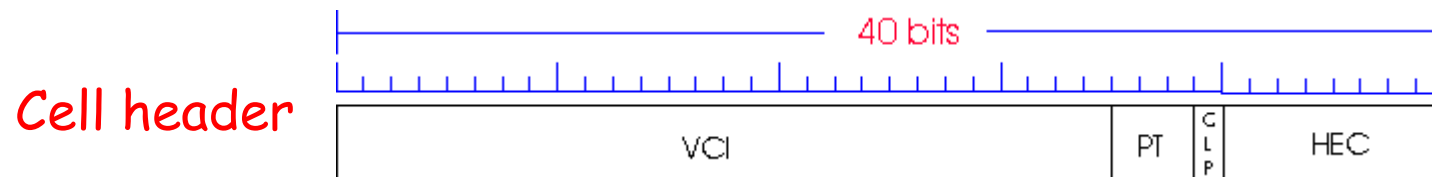
- ❑ **VC transport:** cells carried on VC from source to dest
  - call setup, teardown for each call *before* data can flow
  - each packet carries VC identifier (not destination ID)
  - *every* switch on source-dest path maintain "state" for each passing connection
  - link, switch resources (bandwidth, buffers) may be *allocated* to VC: to get circuit-like perf.
- ❑ **Permanent VCs (PVCs)**
  - long lasting connections
  - typically: "permanent" route between to 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/dest pair) does not scale ( $N^2$  connections needed)
  - SVC introduces call setup latency, processing overhead for short lived connections

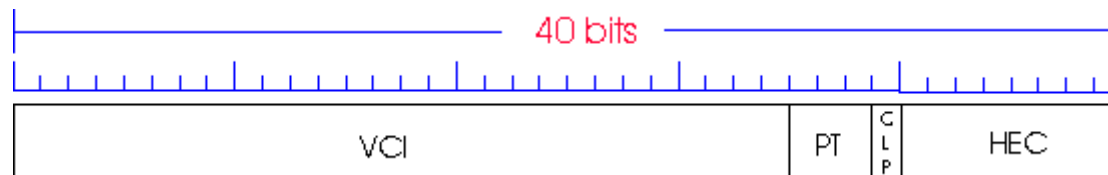
# ATM Layer: ATM cell

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



# ATM cell header

- ❑ **VCI:** virtual channel ID
  - will *change* from link to link thru 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 (more)

*Two pieces (sublayers) of physical layer:*

- ❑ **Transmission Convergence Sublayer (TCS):** adapts ATM layer above to PMD sublayer below
- ❑ **Physical Medium Dependent:** depends on physical medium being used

**TCS Functions:**

- Header **checksum** generation: 8 bits CRC
- Cell **delineation**
- With "unstructured" PMD sublayer, transmission of **idle cells** when no data cells to send

# ATM Physical Layer

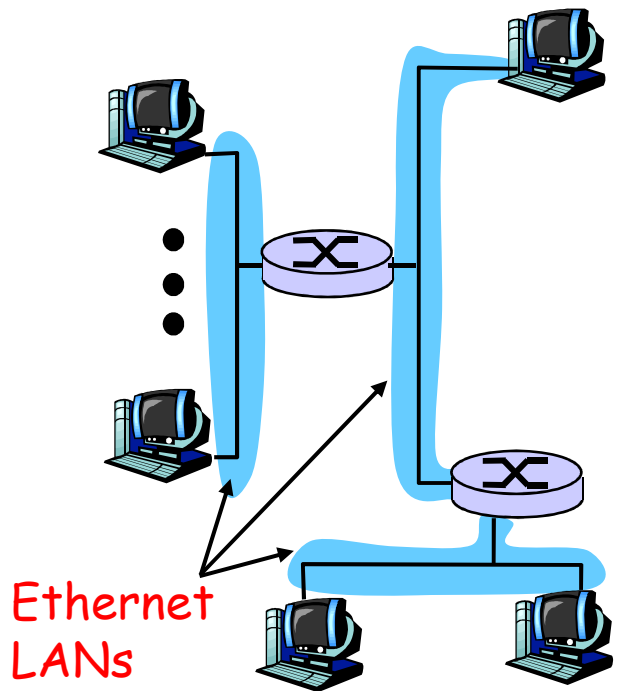
## Physical Medium Dependent (PMD) sublayer

- ❑ **SONET/SDH:** transmission frame structure (like a container carrying bits);
  - bit synchronization;
  - bandwidth partitions (TDM);
  - several speeds: OC3 = 155.52 Mbps; OC12 = 622.08 Mbps; OC48 = 2.45 Gbps, OC192 = 9.6 Gbps
- ❑ **TI/T3:** transmission frame structure (old telephone hierarchy): 1.5 Mbps/ 45 Mbps
- ❑ **unstructured:** just cells (busy/idle)

# IP-Over-ATM

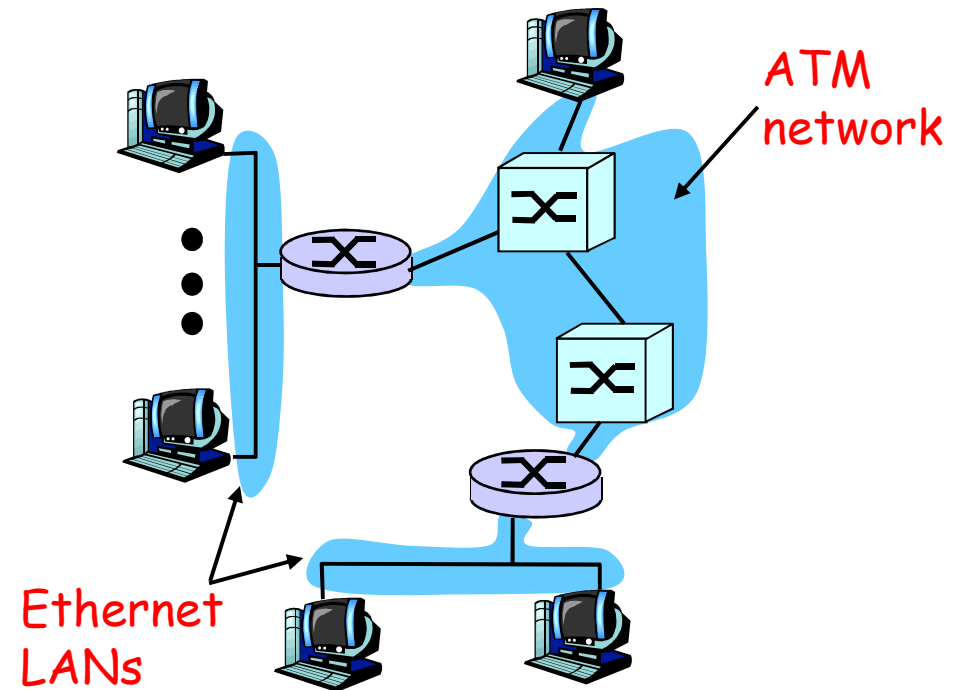
## Classic IP only

- 3 "networks" (e.g., LAN segments)
- MAC (802.3) and IP addresses

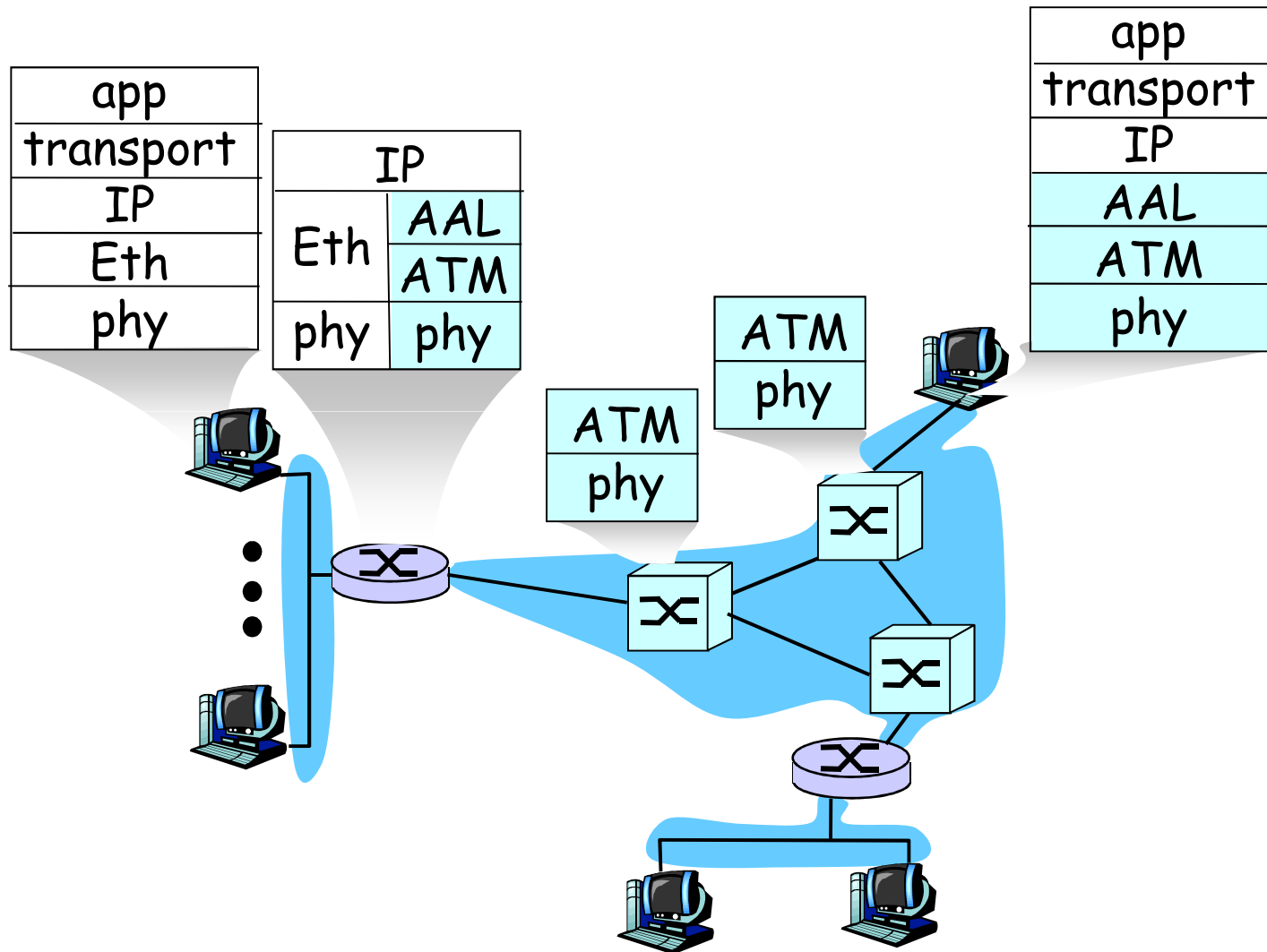


## IP over ATM

- replace "network" (e.g., LAN segment) with ATM network
- ATM addresses, IP addresses



# IP-Over-ATM





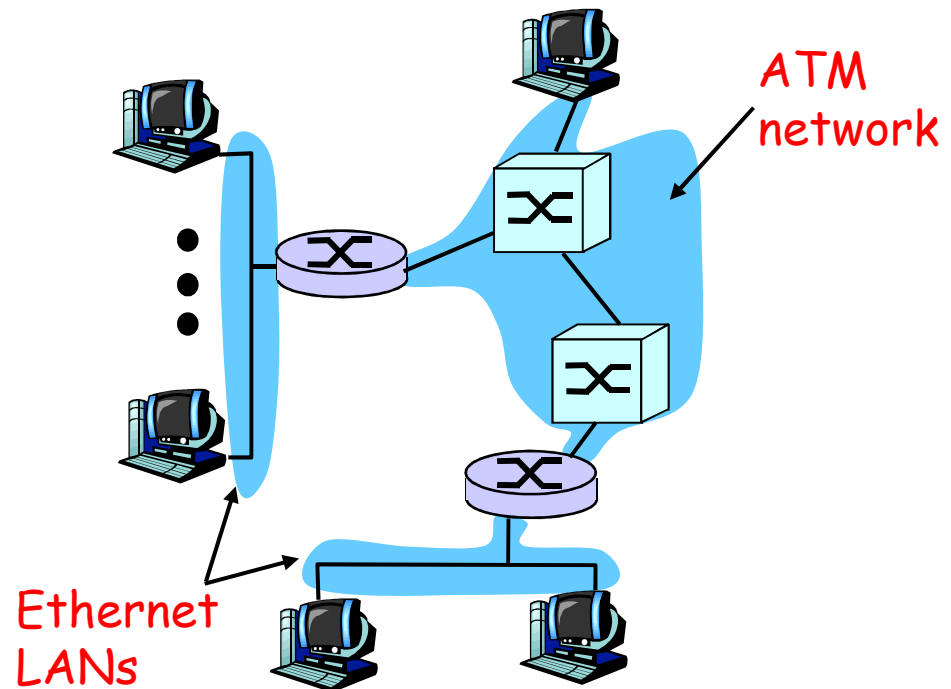
# Datagram Journey in IP-over-ATM Network

- ❑ **at Source Host:**
  - IP layer maps between IP, ATM dest 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 Destination Host:**
  - AAL5 reassembles cells into original datagram
  - if CRC OK, datagram is passed to IP

# IP-Over-ATM

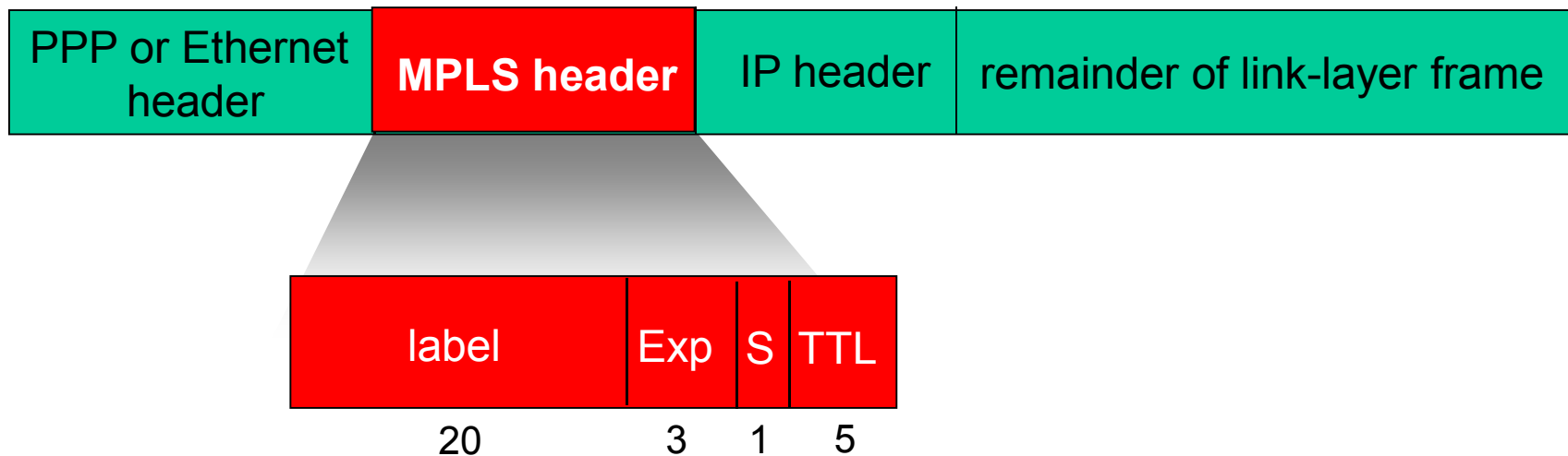
## Issues:

- ❑ IP datagrams into ATM AAL5 PDUs
- ❑ from IP addresses to ATM addresses
  - just like IP addresses to 802.3 MAC addresses!



# Multiprotocol label switching (MPLS)

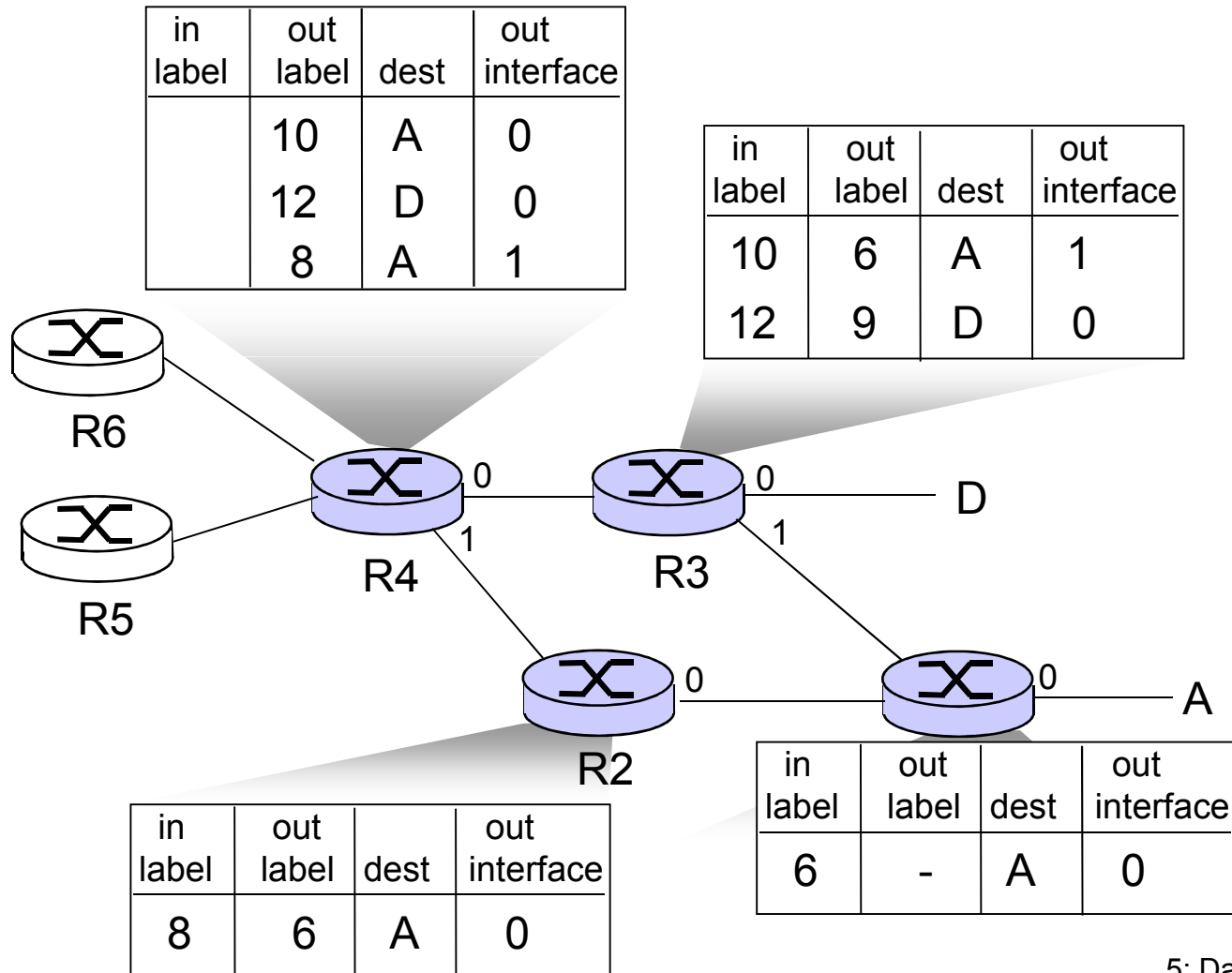
- initial goal: speed up IP forwarding by using fixed length label (instead of IP address) to do forwarding
  - borrowing ideas from Virtual Circuit (VC) approach
  - but IP datagram still keeps IP address!



# 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
  - RSVP-TE
  - 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



# 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

# Chapter 5: let's take a breath

- ❑ journey down protocol stack *complete*  
(except PHY)
- ❑ solid understanding of networking principles,  
practice
- ❑ ..... could stop here .... but *lots* of interesting  
topics!
  - wireless
  - multimedia
  - security
  - network management