Chapter 5
Data Link Layer

Computer Networking: A Top Down Approach Featuring the Internet,
2nd edition.
Jim Kurose, Keith Ross
Addison-Wesley, July 2002.

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:
- instantiation and implementation of various link layer technologies

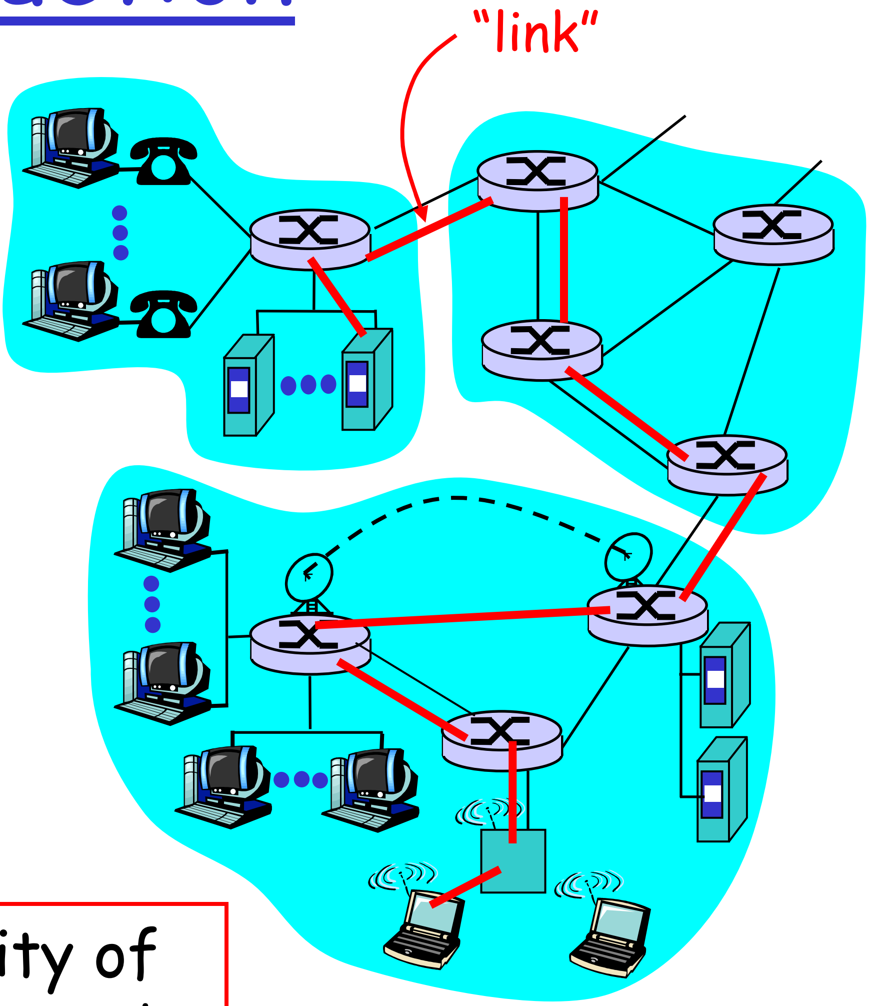
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple access protocols
- 5.4 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

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
- 2-PDU 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 Patna to Richardson TX
  - limo: Patna to Bombay
  - plane: Bombay to DFW
  - train: DFW to Richardson
- tourist = datagram
- transport segment = communication link
- transportation mode = link layer protocol
- travel agent = routing algorithm

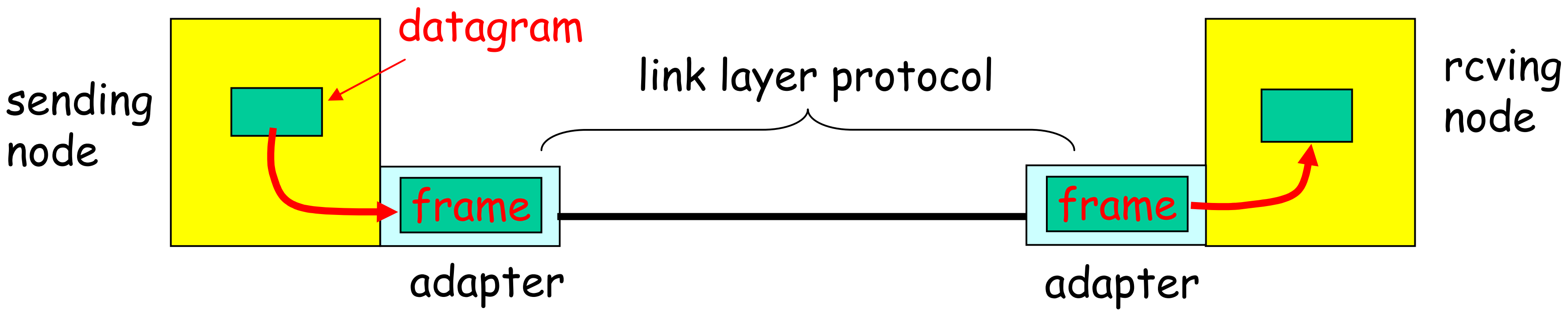
Link Layer Services

- Framing, link access:
  - encapsulate datagram into frame, adding header, trailer
  - channel access if shared medium
  - ‘physical addresses’ used in frame headers to identify source, dest
    - different from IP address!
- Reliable delivery between adjacent nodes
  - seldom used on low bit error link (fiber, some twisted pair)
  - wireless links: high error rates
    - Q: why both link-level and end-end reliability?

Link Layer Services (more)

- *Flow Control:*
  - pacing between adjacent sending and receiving nodes
- *Error Detection*:
  - errors caused by signal attenuation, noise.
  - receiver detects presence of errors:
    - signals sender for retransmission or drops frame
- Error Correction:
  - receiver identifies *and corrects* bit error(s) without resorting to retransmission
- *Half-duplex and full-duplex*
  - with half duplex, nodes at both ends of link can transmit, but not at same time

Adaptors Communicating



- link layer implemented in "adaptor" (aka NIC)
  - Ethernet card, 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 rcving node
- adapter is semi-autonomous
- link & physical layers

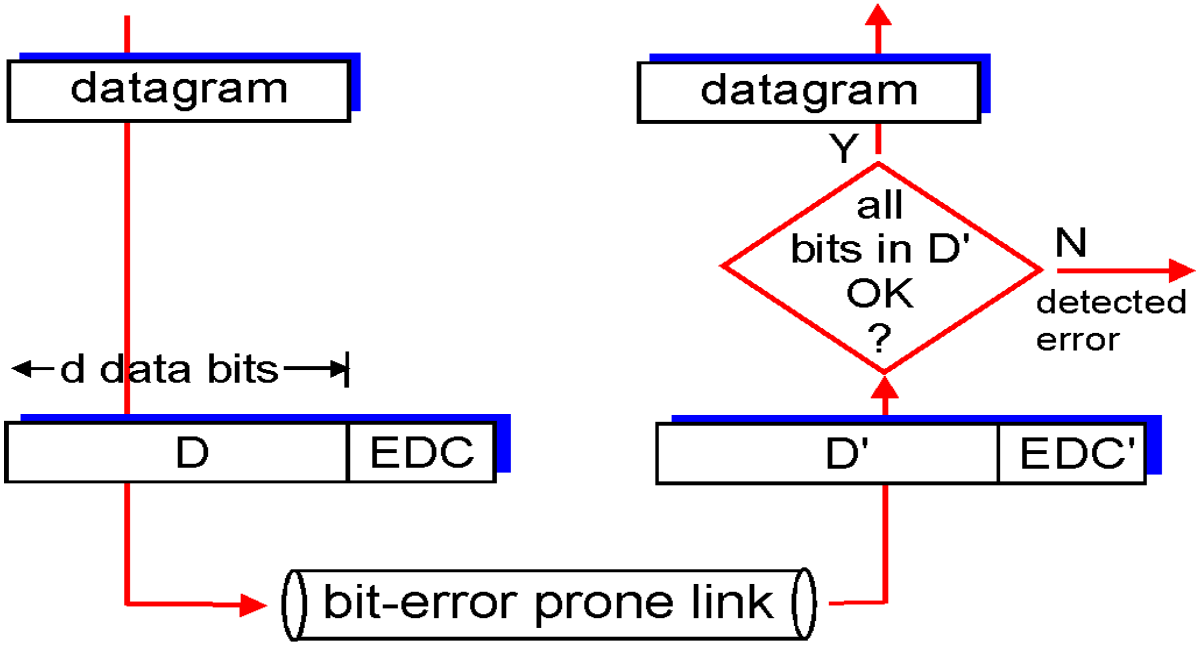
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple access protocols
- 5.4 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

Error Detection

EDC= Error Detection and Correction bits (redundancy)
D = Data protected by error checking, may include header fields

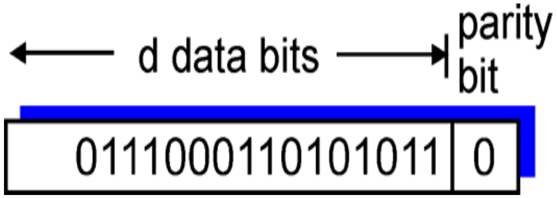
- Error detection not 100% reliable!
  - protocol may miss some errors, but rarely
  - larger EDC field yields better detection and correction



Parity Checking

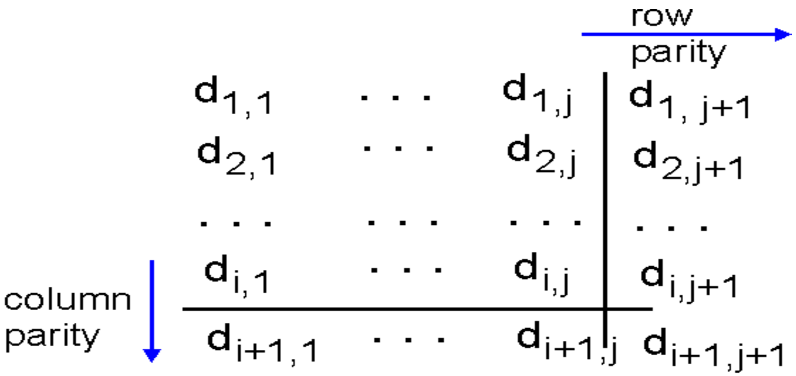
Single Bit Parity:

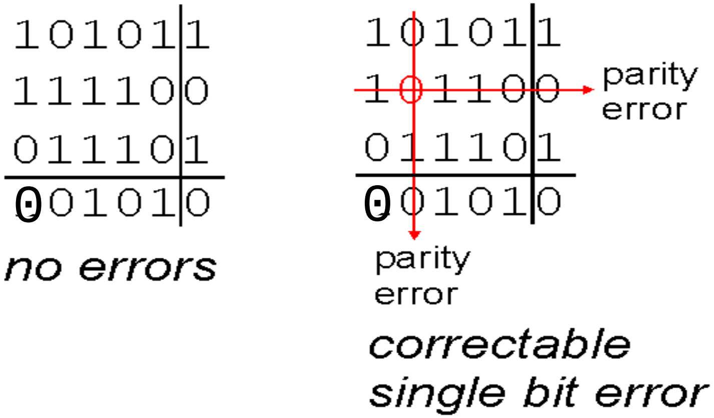
Detect single bit errors



Two Dimensional Bit Parity:

Detect *and correct* single bit errors


row parity
column parity
$d_{1,1}$ . . . $d_{1,j}$ $d_{1,\,j+1}$
$d_{2,1}$ . . . $d_{2,j}$ $d_{2,j+1}$
. . . .
$d_{i,1}$ . . . $d_{i,j}$ $d_{i,j+1}$
$d_{i+1,1}$ . . . $d_{i+1,j}$ $d_{i+1,j+1}$



```
101011
111100
011101
001010
```

no errors

```
101011
101100
011101
001010
```

correctable single bit error

Internet checksum

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

Sender:

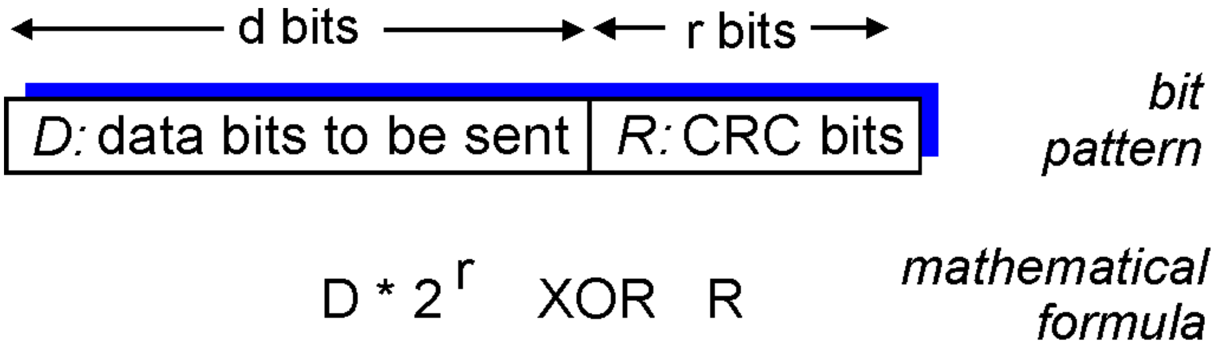
- treat segment contents as sequence of 16-bit integers
- checksum: addition (1's complement sum) of segment contents
- sender puts checksum value into UDP checksum field

Receiver:

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

Checksumming: Cyclic Redundancy Check

- view data bits, D, as a binary number
- choose r+1 bit pattern (generator), G
- goal: choose r CRC bits, R, such that
  - <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, HDCL)



CRC Example

Want:

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

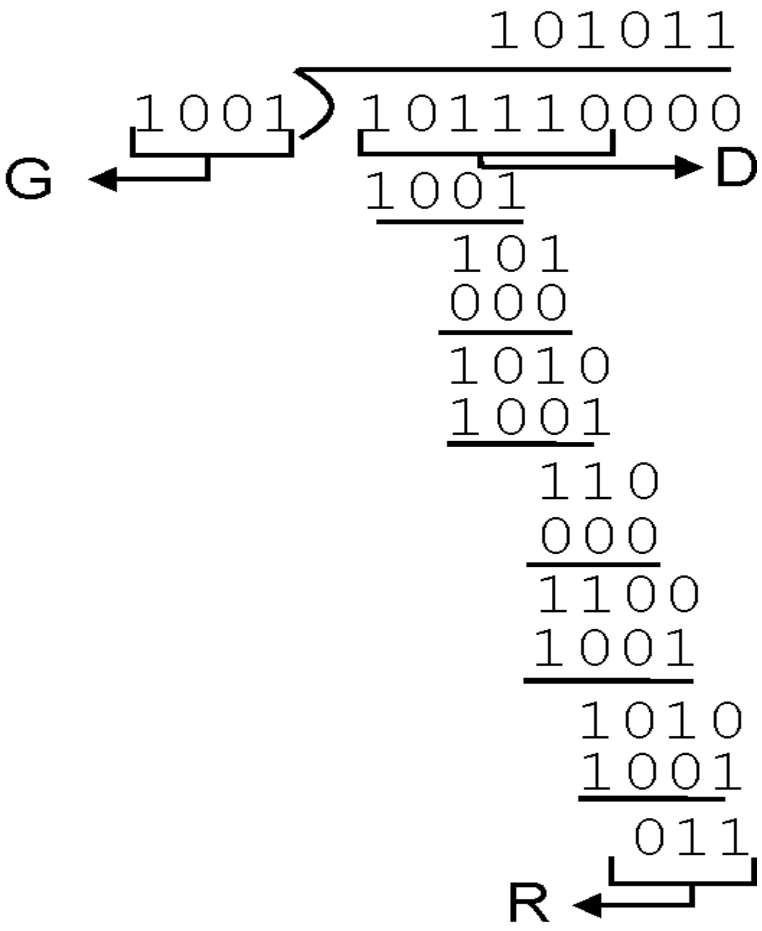
equivalently:

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

equivalently:

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

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



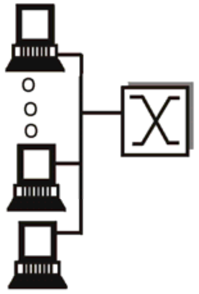
Chapter 5 outline

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

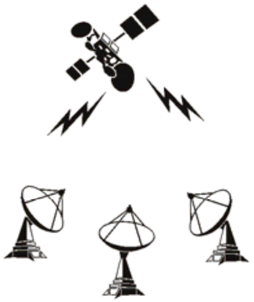
Two types of "links":

- point-to-point
  - PPP for dial-up access
  - point-to-point link between Ethernet switch and host
- broadcast (shared wire or medium)
  - traditional Ethernet
  - upstream HFC
  - 802.11 wireless LAN
- What is the difference between broadcast and multicast



shared wire
(e.g. Ethernet)

shared wireless
(e.g. Wavelan)



satellite

cocktail party

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! (what a paradox ☺)
- what to look for in multiple access protocols:

Ideal Mulitple 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

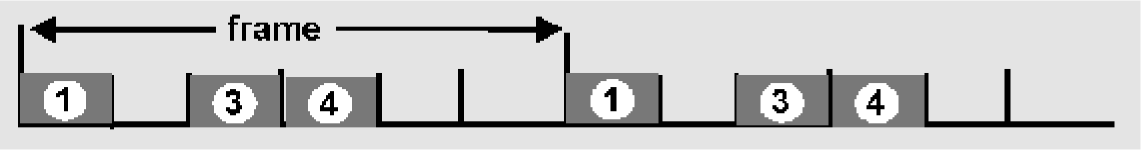
Three broad classes:

- 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"
  - tightly coordinate shared access to avoid collisions

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

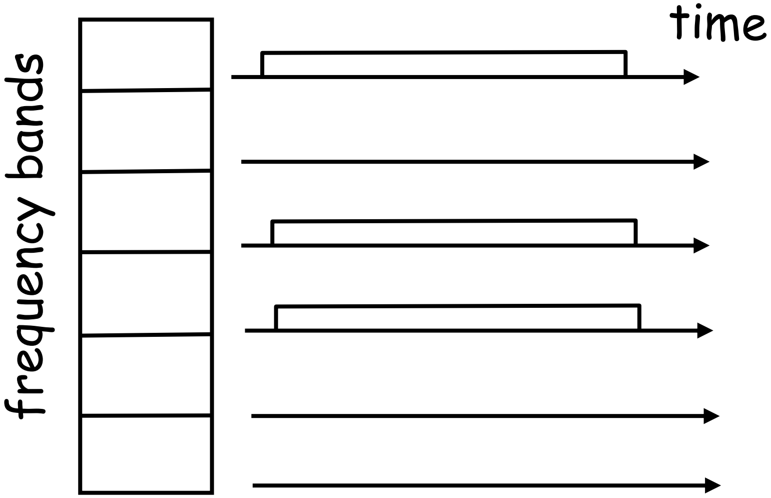
FDMA: frequency division multiple access

- 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

Q: to the class?

Is there a way to dynamically assign channel frequencies?

Such an algorithm would be called dynamic frequency allocation algorithm

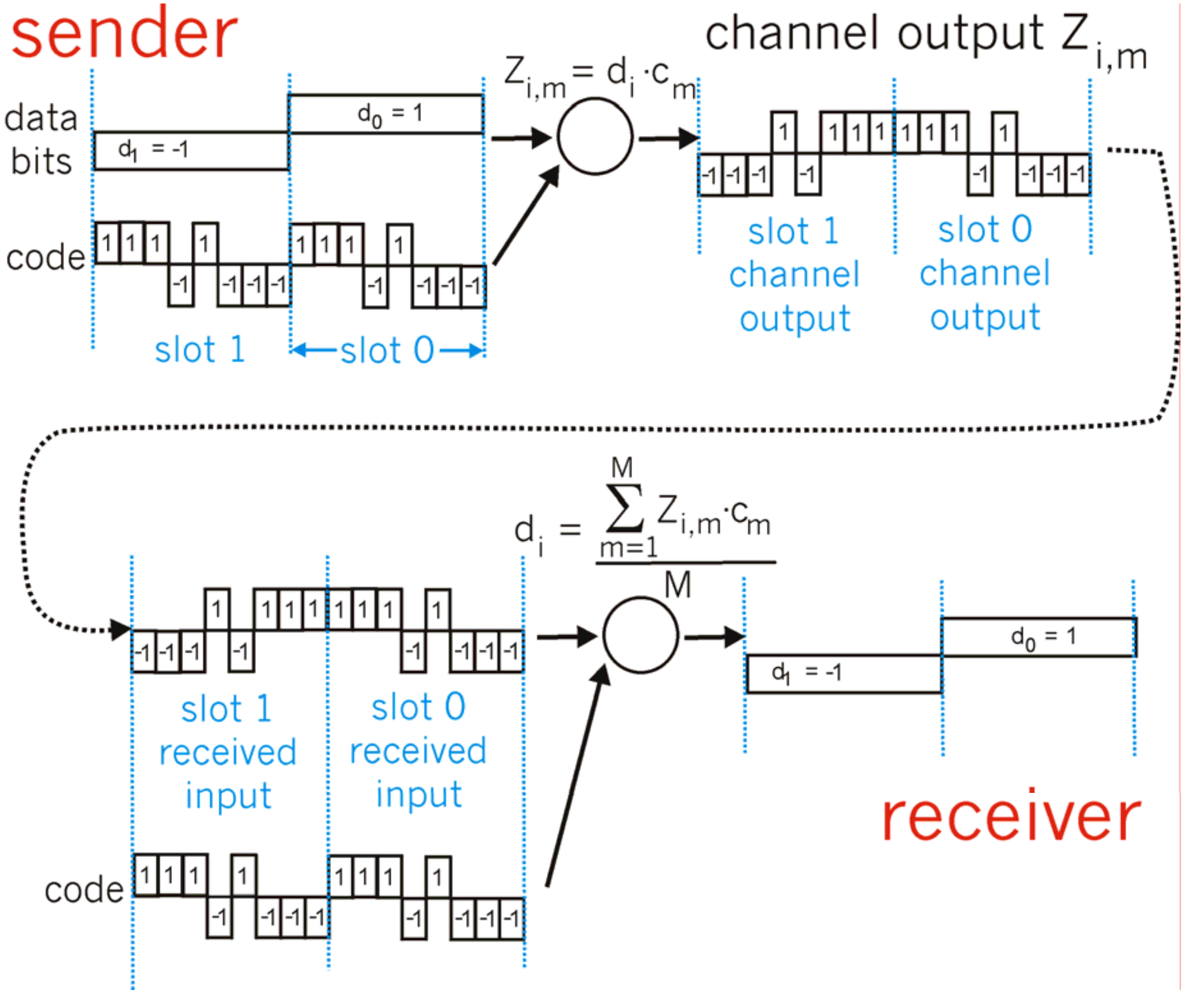


Channel Partitioning (CDMA)

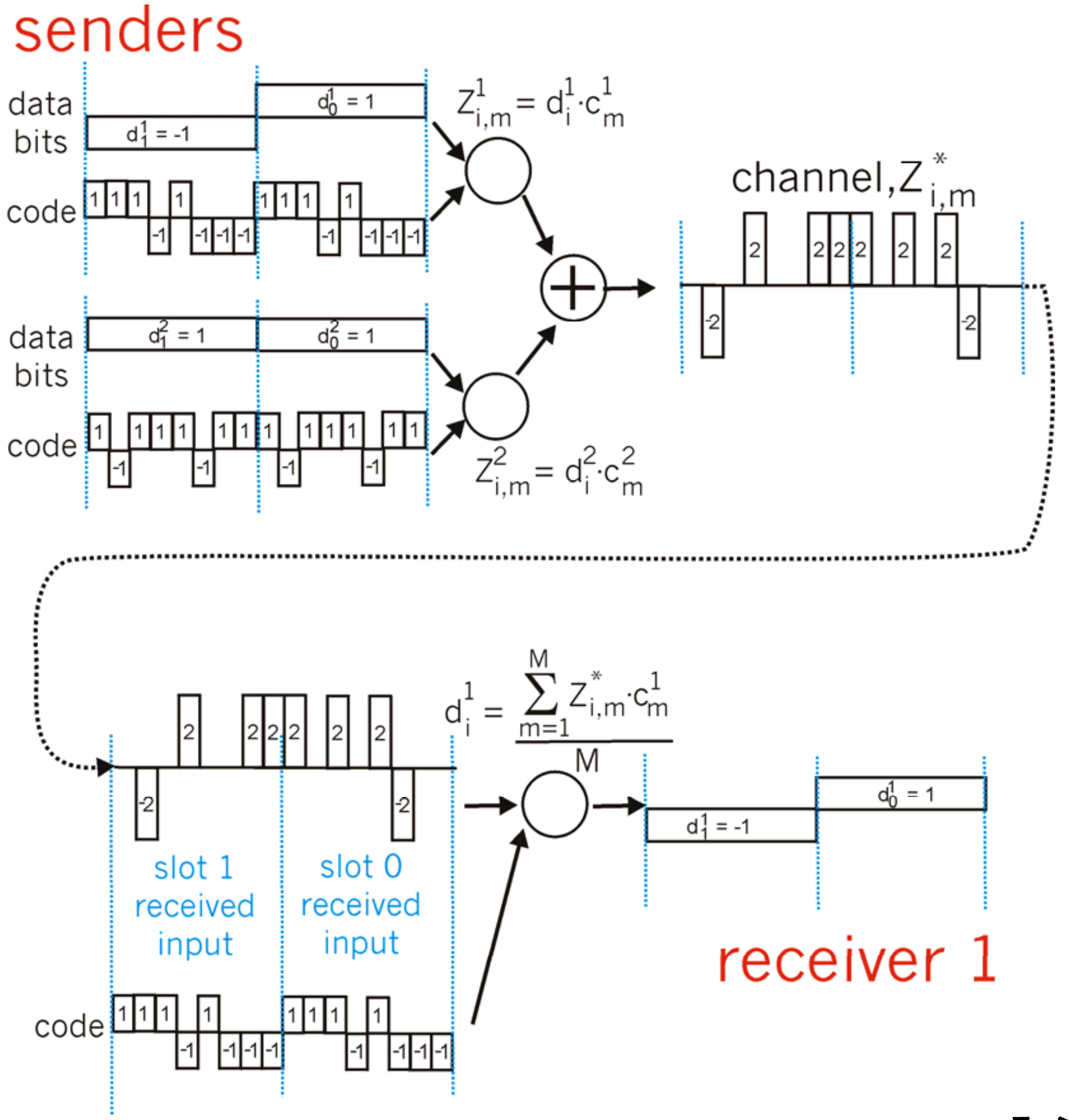
CDMA (Code Division Multiple Access)

- unique “code” assigned to each user; i.e., code set partitioning
- used mostly in wireless broadcast channels (cellular, satellite, etc)
- all users share same frequency, but each user has own “chipping” sequence (i.e., code) to encode data
- *encoded signal* = (original data) X (chipping sequence)
- *decoding:* inner-product of encoded signal and chipping sequence
- allows multiple users to “coexist” and transmit simultaneously with minimal interference (if codes are “orthogonal”)

CDMA Encode/Decode



CDMA: two-sender interference



Random Access Protocols

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

Slotted ALOHA

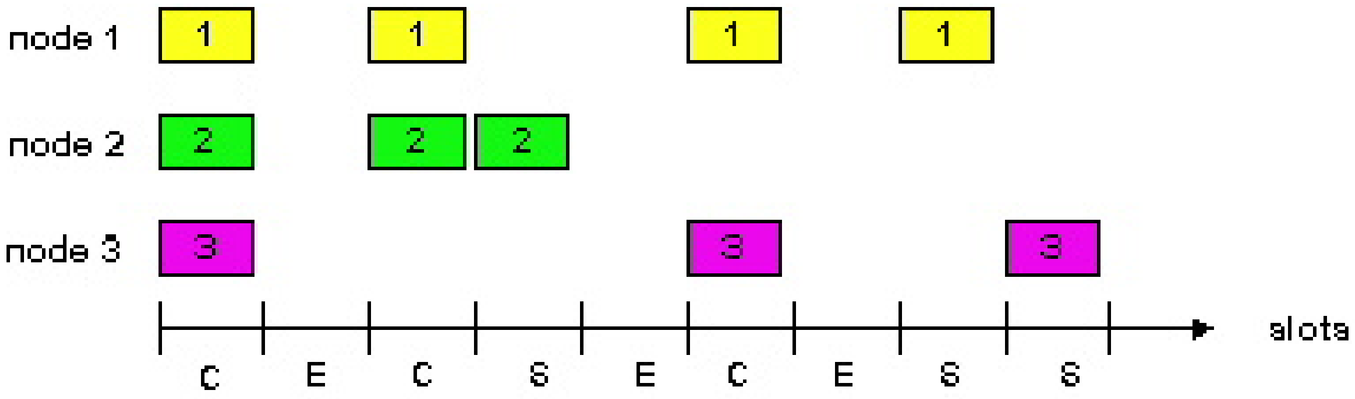
Assumptions

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

Operation

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

Slotted ALOHA



Pros

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

Cons

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

Slotted Aloha efficiency

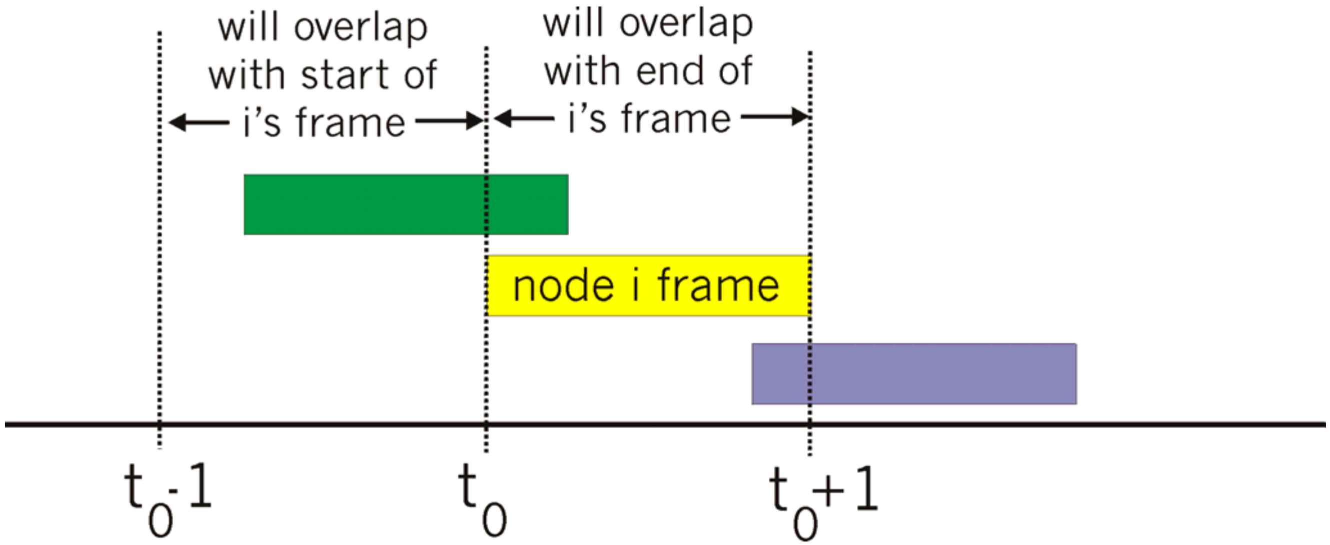
Efficiency is the long-run fraction of successful slots when there's many nodes, each with many frames to send

- Suppose N nodes with many frames to send, each transmits in slot with probability *p*
- prob that 1st node has success in a slot $= p(1-p)^{N-1}$
- prob that any node has a success $= Np(1-p)^{N-1}$
- For max efficiency with N nodes, find p* that maximizes $Np(1-p)^{N-1}$
- For many nodes, take limit of $Np^*(1-p^*)^{N-1}$ as N goes to infinity, gives 1/e = .37

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

Pure (unslotted) ALOHA

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


will overlap with start of i's frame
will overlap with end of i's frame
node i frame
t_0-1
t_0
t_0+1

Pure Aloha efficiency

P(success by given node) = P(node transmits) ·

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

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

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

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

… choosing optimum p and then letting n -> infty ...

Even worse !

= 1/(2e) = .18

CSMA (Carrier Sense Multiple Access)

CSMA: listen before transmit:

- If channel sensed idle: transmit entire frame
- If channel sensed busy, defer transmission

- Human analogy: don't interrupt others!

CSMA collisions

collisions *can* still occur:

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

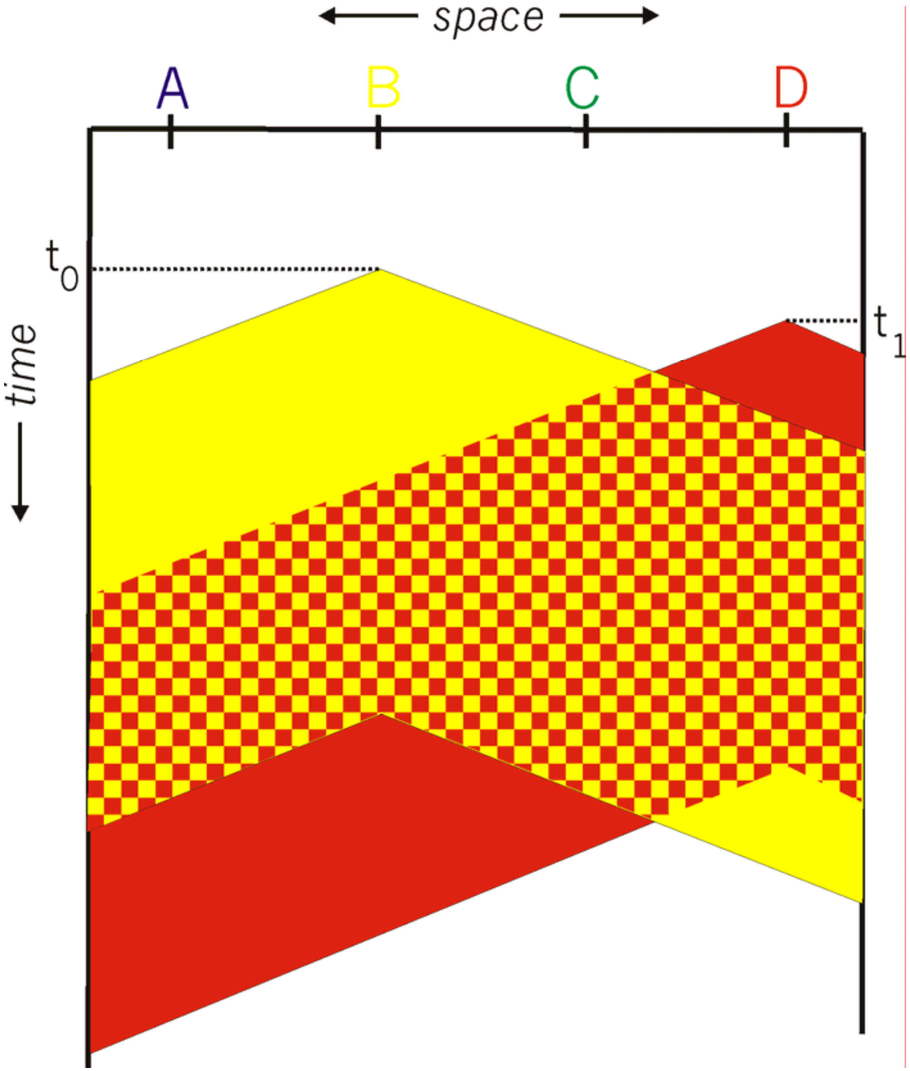
collision:

entire packet transmission time wasted

note:

role of distance & propagation delay in determining collision probability

spatial layout of nodes



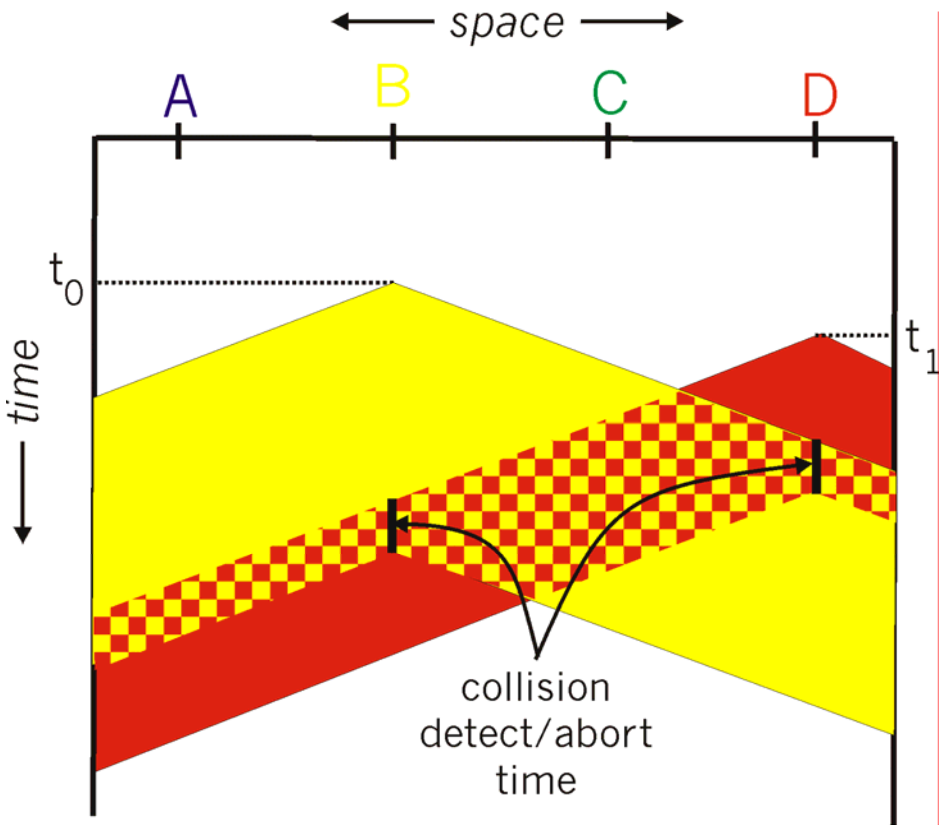
CSMA/CD (Collision Detection)

CSMA/CD: carrier sensing, deferral as in CSMA

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

- collision detection:
  - easy in wired LANs: measure signal strengths, compare transmitted, received signals
  - difficult in wireless LANs: receiver shut off while transmitting
- human analogy: the polite conversationalist

CSMA/CD collision detection



CSMA (Carrier-sense multiple access)

- If propagation time is much less than transmission time - all stations know that a transmission has started almost immediately
- First listen for clear medium (carrier sense)
  - If medium idle, transmit
- Collision occurs if another user starts transmitting within the time it takes for the first bit to reach this user (propagation delay)
- Collision detected by waiting round trip plus ACK contention
  - No ACK then retransmit
- Max utilization depends on propagation time (medium length) and frame length
  - Longer frame and shorter propagation gives better utilization

CSMA/CD

- With CSMA, collision occupies medium for duration of transmission
  - Even if the station next to transmitting station collided, collision will be detected after >= RTT
- Instead "CD"= collision detect:
  - Stations listen whilst transmitting
  - If medium idle, transmit
  - If busy, listen for idle, then transmit (and listen)
  - If collision detected, jam (send noise) then cease transmission
- After jam, wait random time then start again
  - Binary exponential back off

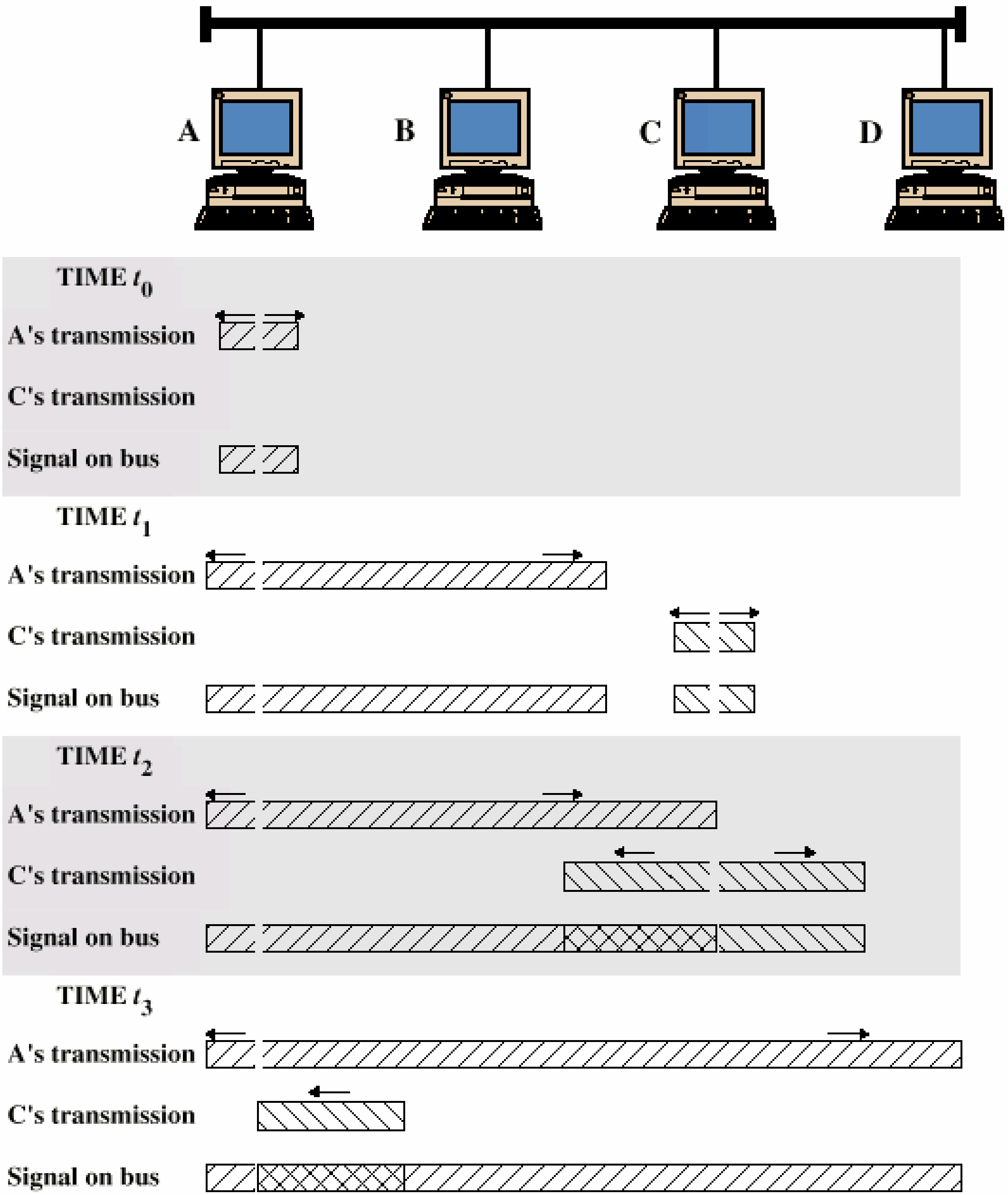
Collision Detection

- Collision produces much higher signal voltage than signal
- Collision detected if cable signal greater than single station signal
- Signal attenuated over distance
- Limit distance to 500m (10Base5) or 200m (10Base2)
- For twisted pair (star-topology) activity on more than one port is collision
  - Frames repeated, for CD to work

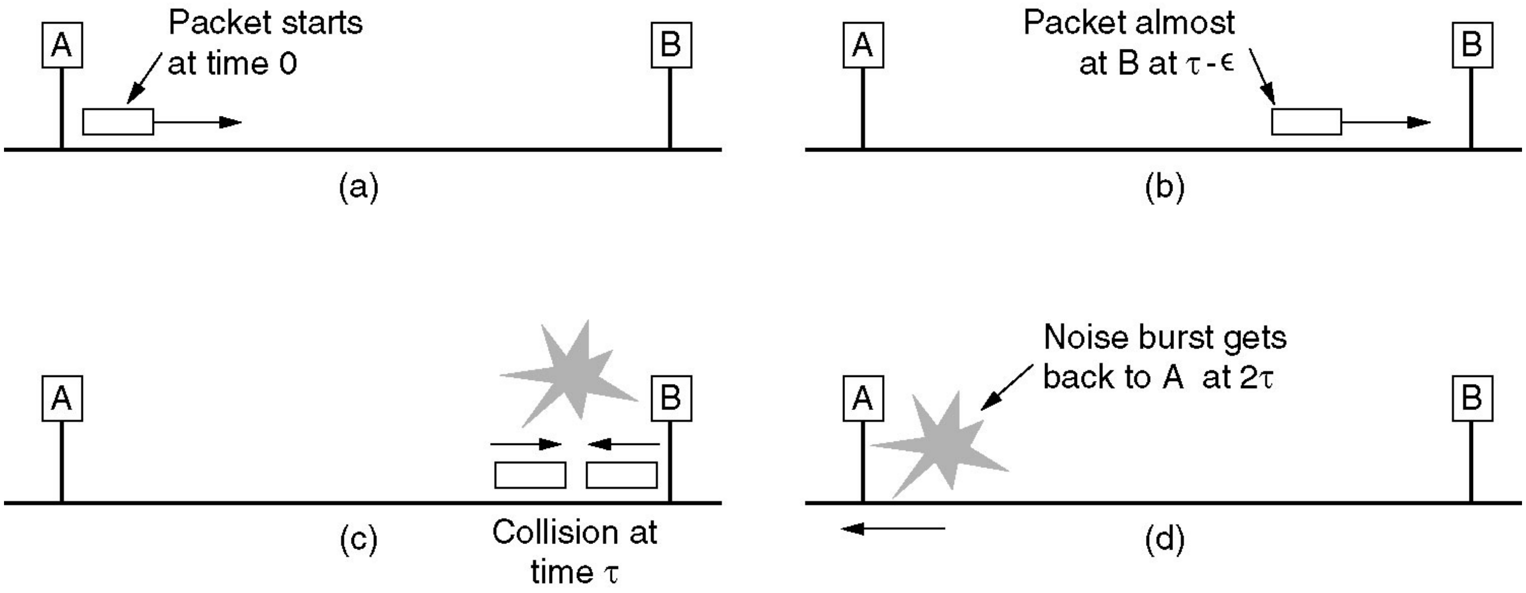
Why "Jam"?

- Tanenbaum: "to make sure the sender does not miss the collision" (48 bits)
- Halsall: "Ensure that the collision is detected by all stations involved"
- Stallings: "Assure all staitons know that there has been a collision"
- Keshav: "Sequence of 512 bits to ensure that every active station on the network knows that a collision happened and increments its backoff counter"; "to ensure that all colliding stations agree that a collision has happened"

CSMA/CD Operation



Collision detection



Collision detection can still take as long as 2τ

Collision detection

- Transmitting stations may detect collisions almost immediately, and stop transmission
  - Saves time and bandwidth
- Will improve upon just CSMA only if collision is detected during frame transmission
- This is possible if frames are long enough (and prop. Delay is short enough) so that collision is detected *while* transmission
  - Guideline used in IEEE 802.3

CSMA/CD efficiency

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

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

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

“Taking Turns” MAC protocols

channel partitioning MAC protocols:

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

Random access MAC protocols

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

“taking turns” protocols

look for best of both worlds!

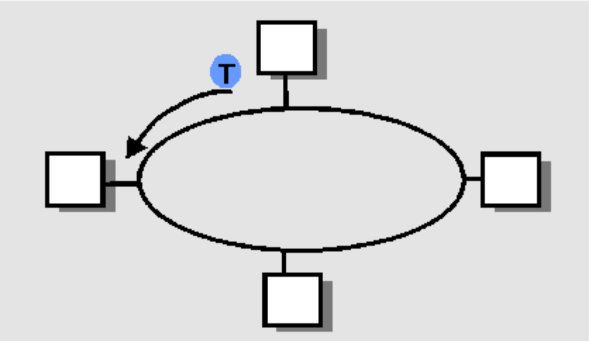
"Taking Turns" MAC protocols

Polling:

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

Token passing:

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



Summary of MAC protocols

- What do you do with a shared media?
  - Channel Partitioning, by time, frequency or code
    - Time Division,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

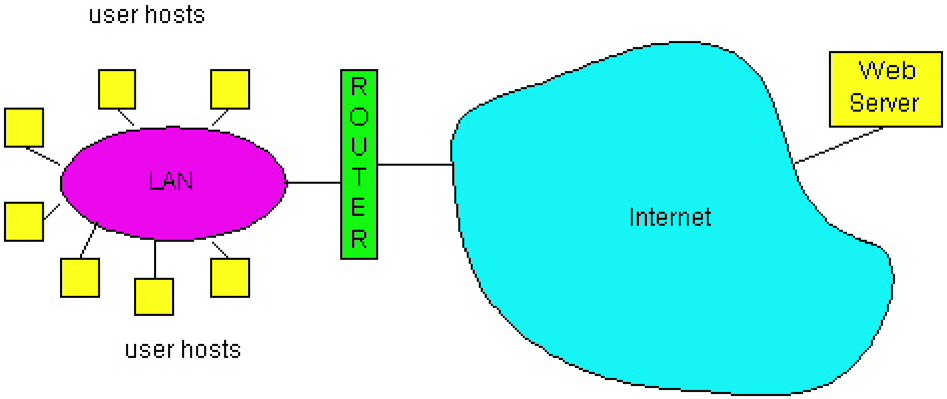
LAN technologies

Data link layer so far:

- services, error detection/correction, multiple access

Next: LAN technologies

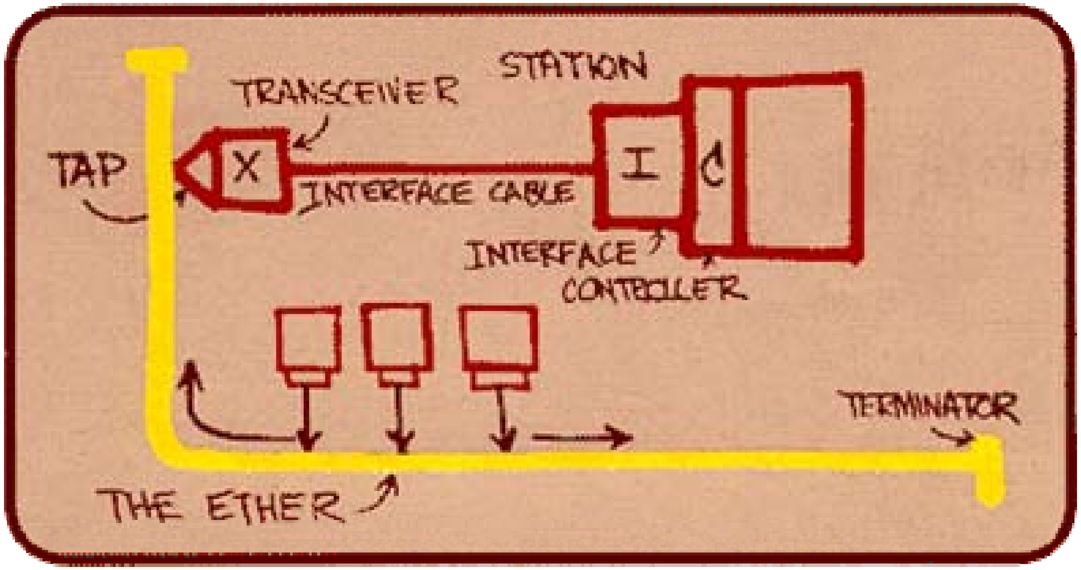
- addressing
- Ethernet
- hubs, bridges, switches
- 802.11
- PPP
- ATM



Ethernet

"dominant" LAN technology:

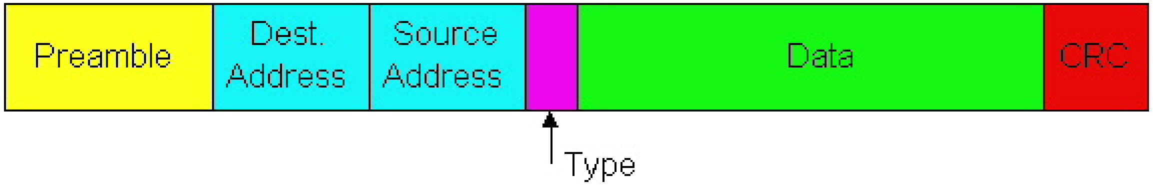
- cheap $20 for 100Mbs!
- first widely used LAN technology
- Simpler, cheaper than token LANs and ATM
- Kept up with speed race: 10, 100, 1000 Mbps
- Now we have 1 GigE and 10 Gige, we soon will have 100 GigE



Metcalfe's Ethernet sketch

Ethernet Frame Structure

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

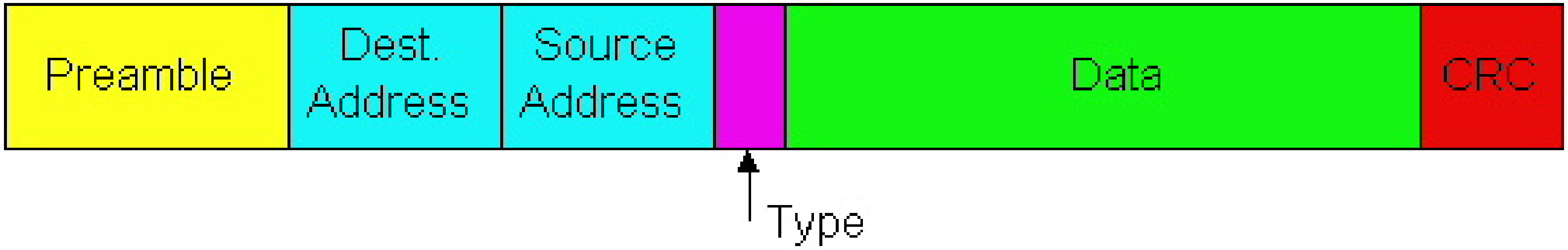


Preamble:

- 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, 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



Ethernet min frame length

- Min length needed for CD: for 2500m distance specification, RT prop delay is determined to be 50 μsec
  - Frame transmission time >= 50 μsec
  - At 10Mbps, bits transmitted in 50 μsec is 500 <= 512 = 64*8 bits = 64 bytes
- When transmission interrupted, "bits & pieces" of frames appear on the cable
  - Min frame length is one "filter" for valid frames

forward

Unreliable, connectionless service

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

Ethernet uses CSMA/CD

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

Ethernet CSMA/CD algorithm

1. Adaptor gets datagram from 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 mth collision, adapter chooses a K at random from $\{0,1,2,\ldots,2^m-1\}$. 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: .1 microsec for 10 Mbps Ethernet ;
for K=1023, wait time is about 50 msec

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}

Ethernet

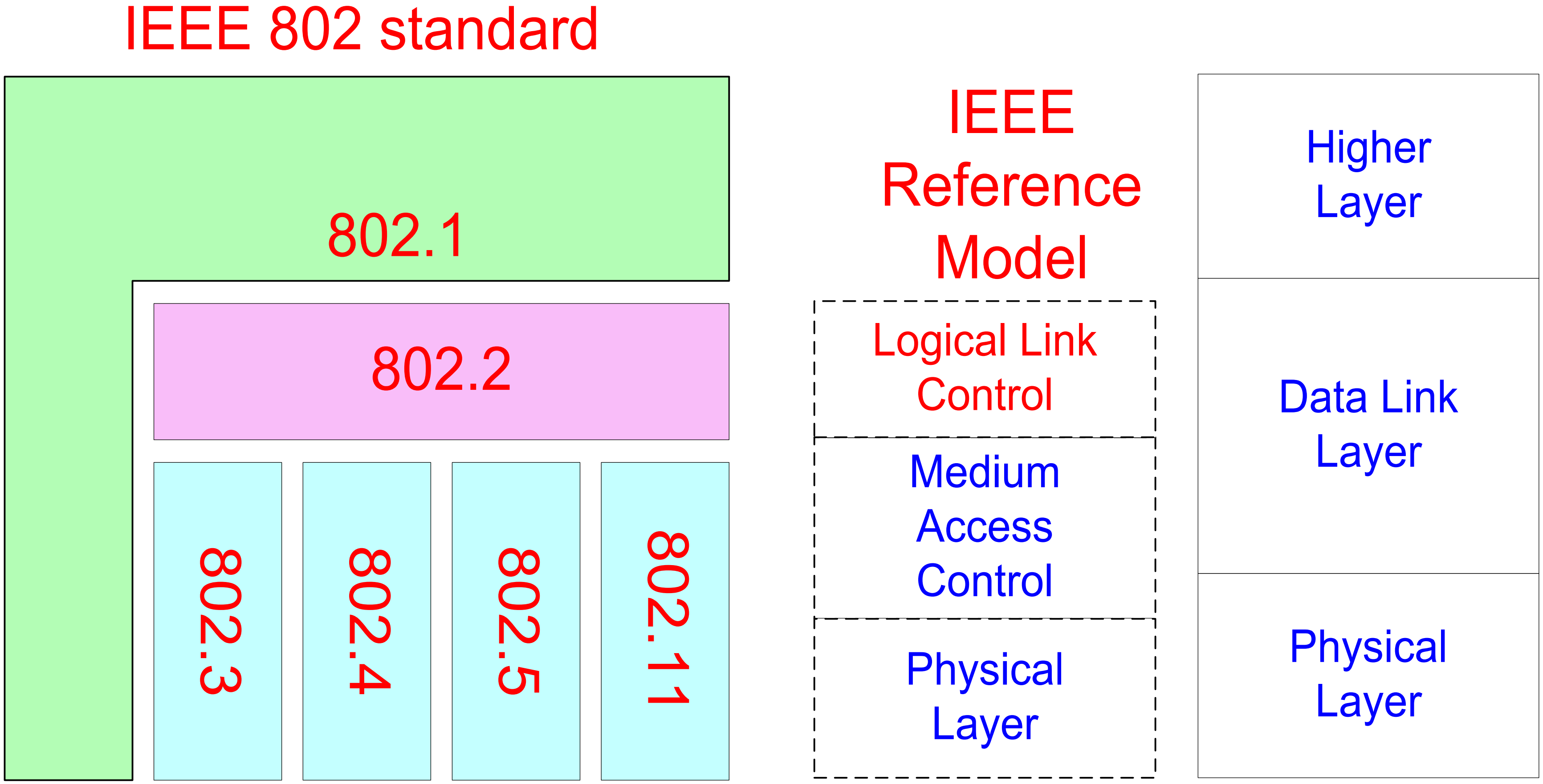
- Speed: 10Mbps -10 Gbps
- Standard: 802.3, Ethernet II (DIX)

- Most popular physical layers for Ethernet:

10Base5	**Thick Ethernet:** 10 Mbps coax cable
10Base2	**Thin Ethernet:** 10 Mbps coax cable
10Base-T	10 Mbps Twisted Pair
100Base-TX	100 Mbps over Category 5 twisted pair
100Base-FX	100 Mbps over Fiber Optics
1000Base-FX	1Gbps over Fiber Optics
10000Base-FX	1Gbps over Fiber Optics (for wide area links)

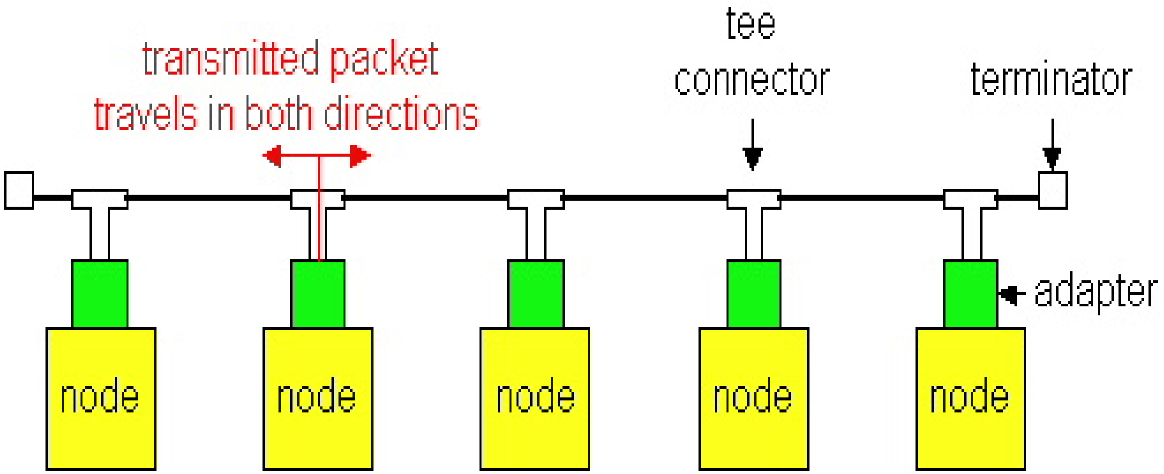
IEEE 802 Standards

- IEEE 802 is a family of standards for LANs, which defines an LLC and several MAC sublayers



Ethernet Technologies: 10Base2

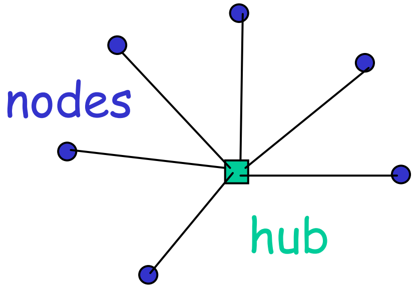
- 10: 10Mbps; 2: under 200 meters max cable length
- thin coaxial cable in a bus topology



- 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 are essentially physical-layer repeaters:
  - bits coming in one link go out all other links
  - no frame buffering
  - no CSMA/CD at hub: adapters detect collisions
  - provides net management functionality

Fast Ethernet

- Higher bit rate media (100 Mbps) is available.
  - Can it be used for Ethernet?
- Recall minimum frame length?
  - Set=512 bits by calculating time needed to detect collisions in Ethernets of upto 2.5km length, of 10Mbps bit rate
- Can higher bit rates be used *without* changing protocol specs, and still make it work?
  - Frame transmission time for 512 bit frame @100Mbps ~ 5μsec
  - 5 μsec >= twice prop. delay
  - Should be <= (1/10th) of 2.5 km => ~200m

This is what was Fast Ethernet: transmission media was available, Ethernet wires were anyway not stretching very far away -> perfect solution say, for e.g. “server room” LAN

Gbit Ethernet

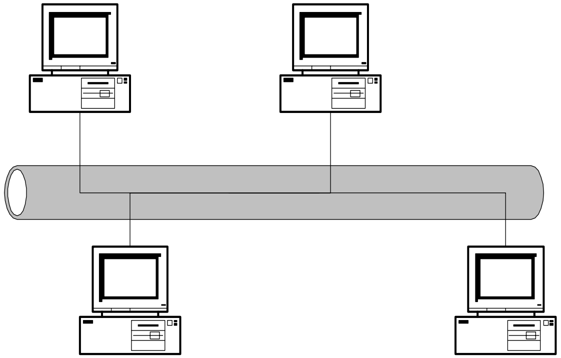
- use standard Ethernet frame format
- allows for point-to-point links and shared broadcast channels
- in shared mode, CSMA/CD is used; short distances between nodes to be efficient
- uses hubs, called here "Buffered Distributors"
- Full-Duplex at 1 Gbps for point-to-point links
- 10 Gbps now !

Gigabit Ethernet

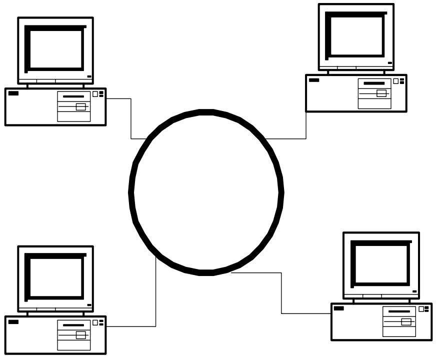
- 1000 Mbps transmission media available.
  - Cannot continue reducing max length
- Two enhancements to basic CSMA/CD
  - Carrier extension: Pad MAC frames to be at least 4096 bits
    - This means ~4 μsec frame transmission time
    - 2*Prop delay < 4 μsec : Length restrictions

Local Area Networks

- Local area networks (LANs) connect computers within a building or a enterprise network
- Almost all LANs are broadcast networks
- Typical topologies of LANs are **bus** or **ring**
-



Bus LAN



Ring LAN

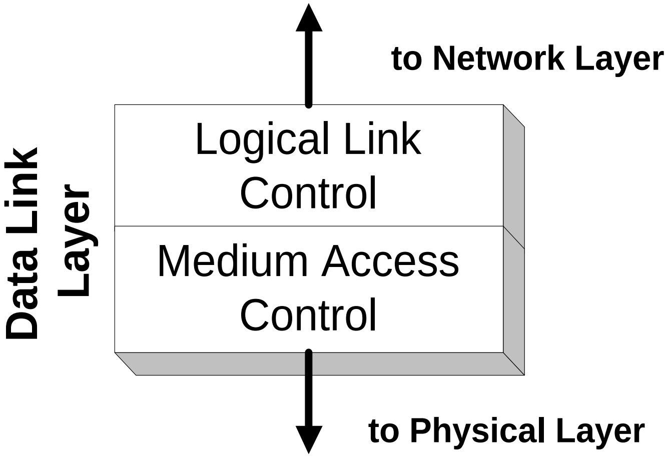
MAC and LLC

- In any broadcast network, the stations must ensure that only one station transmits at a time on the shared communication channel

- The protocol that determines who can transmit on a broadcast channel is called Medium Access Control (MAC) protocol

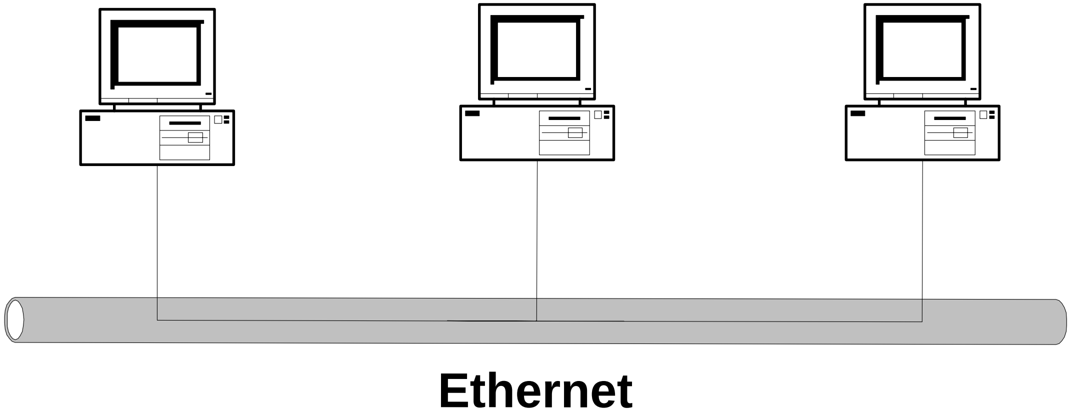
- The MAC protocol are implemented in the MAC sublayer which is the lower sublayer of the data link layer

- The higher portion of the data link layer is often called Logical Link Control (LLC)



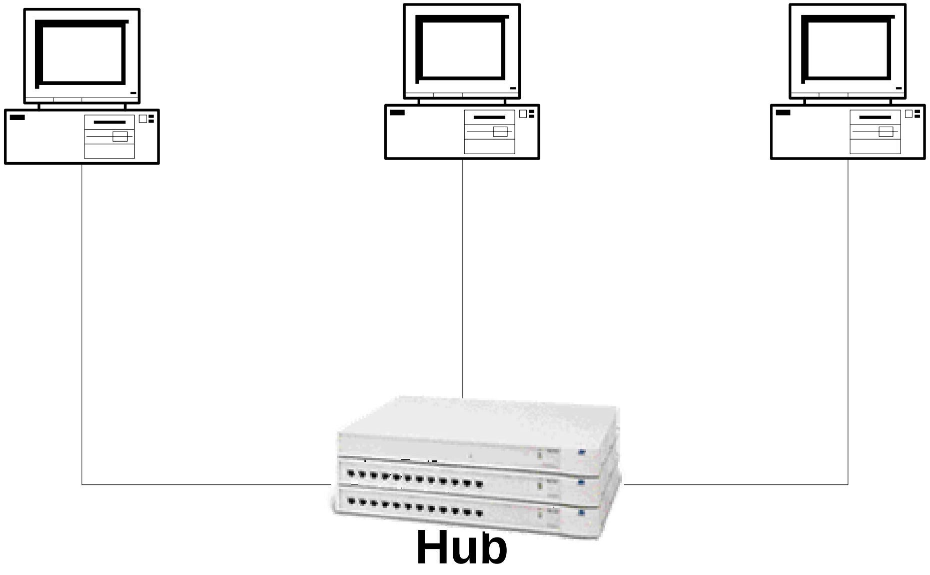
Bus Topology

- 10Base5 and 10Base2 Ethernets has a bus topology



Star Topology

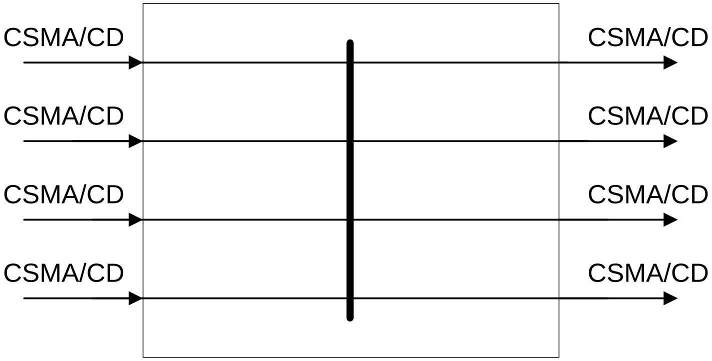
- Starting with 10Base-T, stations are connected to a hub in a star configuration

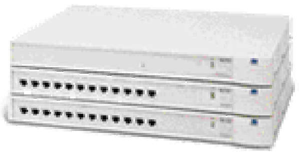


Ethernet Hubs vs. Ethernet Switches

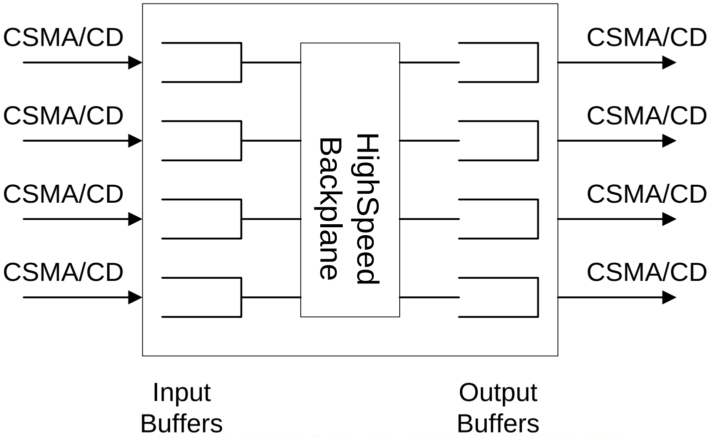
- An **Ethernet switch** is a packet switch for Ethernet frames
  - Buffering of frames prevents collisions.
  - Each port is isolated and builds its own collision domain
- An **Ethernet Hub** does not perform buffering:
  - Collisions occur if two frames arrive at the same time.

Hub





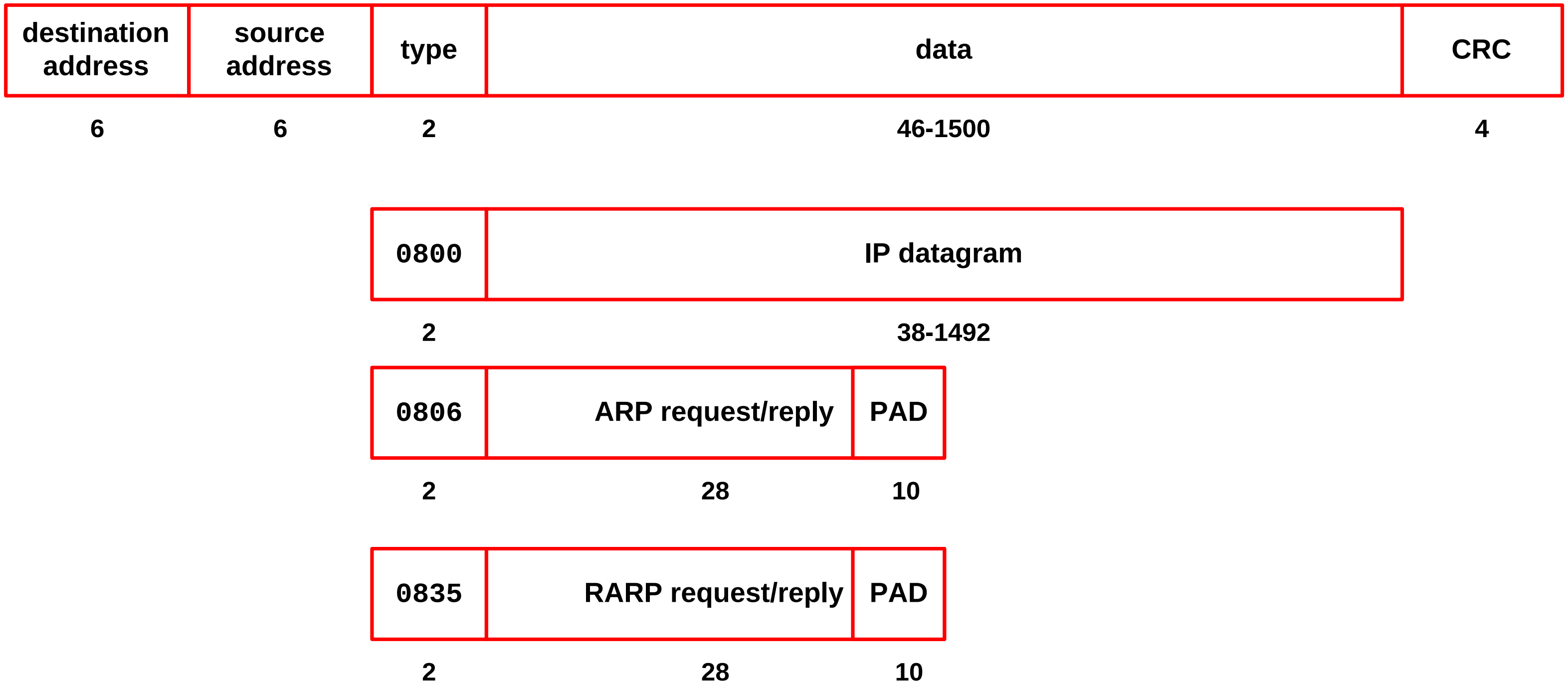
Switch



Ethernet and IEEE 802.3: Any Difference?

- There are two types of Ethernet frames in use, with subtle differences:
- ***"Ethernet" (Ethernet II, DIX)***
  - An industry standards from 1982 that is based on the first implementation of CSMA/CD by Xerox.
  - Predominant version of CSMA/CD in the US.
- **802.3:**
  - IEEE's version of CSMA/CD from 1985.
  - Interoperates with 802.2 (LLC) as higher layer.
- **Difference for our purposes:** Ethernet and 802.3 use different methods to encapsulate an IP datagram.

Ethernet II, DIX Encapsulation (RFC 894)



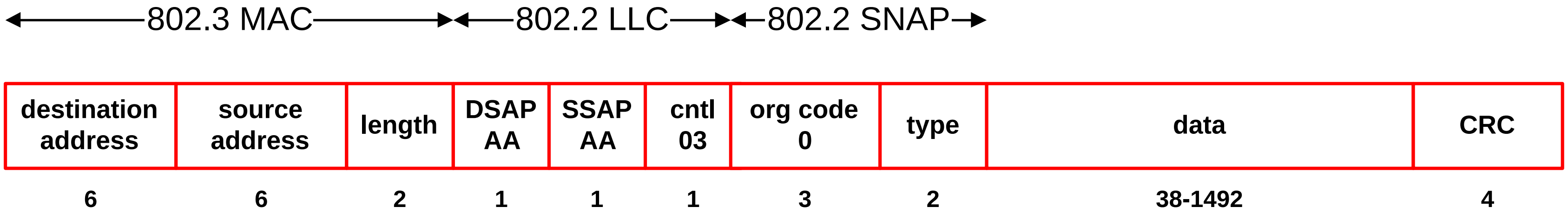
destination address	source address	type	data	CRC
6	6	2	46-1500	4

`0800`	IP datagram
2	38-1492

`0806`	ARP request/reply	PAD
2	28	10

`0835`	RARP request/reply	PAD
2	28	10

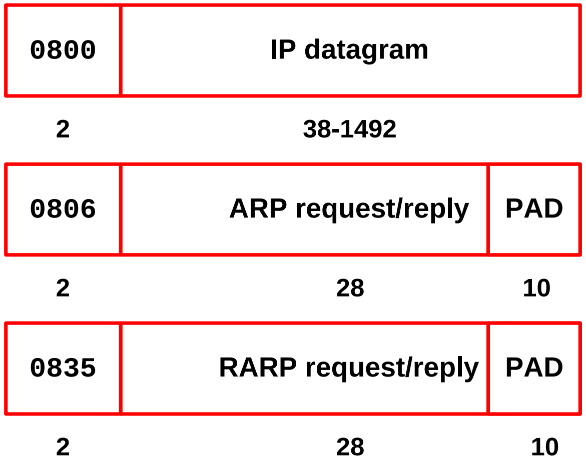
IEEE 802.2/802.3 Encapsulation (RFC 1042)



←802.3 MAC→←802.2 LLC→←802.2 SNAP→

destination address	source address	length	DSAP AA	SSAP AA	cntl 03	org code 0	type	data	CRC
6	6	2	1	1	1	3	2	38-1492	4

- **destination address, source address:** MAC addresses are 48 bit
- **length**: frame length in number of bytes
- **DSAP, SSAP**: always set to `0xaa`
- **Ctrl:** set to 3
- **org code:** set to 0
- **type field** identifies the content of the data field
- **CRC:** cylic redundancy check



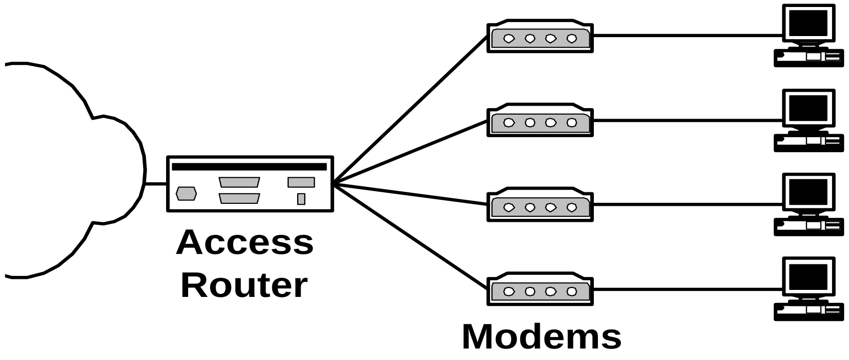
`0800`	IP datagram	
2	38-1492	

`0806`	ARP request/reply	PAD
2	28	10

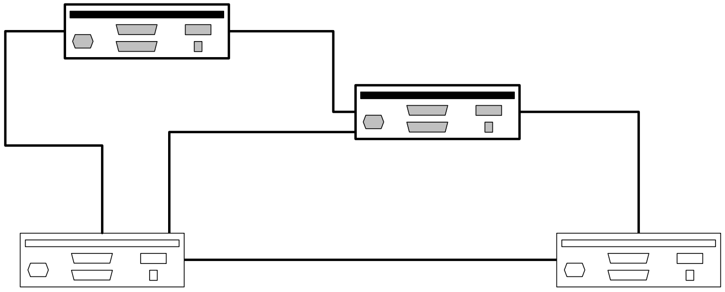
`0835`	RARP request/reply	PAD
2	28	10

Point-to-Point (serial) links

- Many data link connections are point-to-point serial links:
  - Dial-in or DSL access connects hosts to access routers
  - Routers are connected by high-speed point-to-point links
- Here, IP hosts and routers are connected by a serial cable
- Data link layer protocols for point-to-point links are simple:
  - Main role is encapsulation of IP datagrams
  - No media access control needed



Dial-Up Access



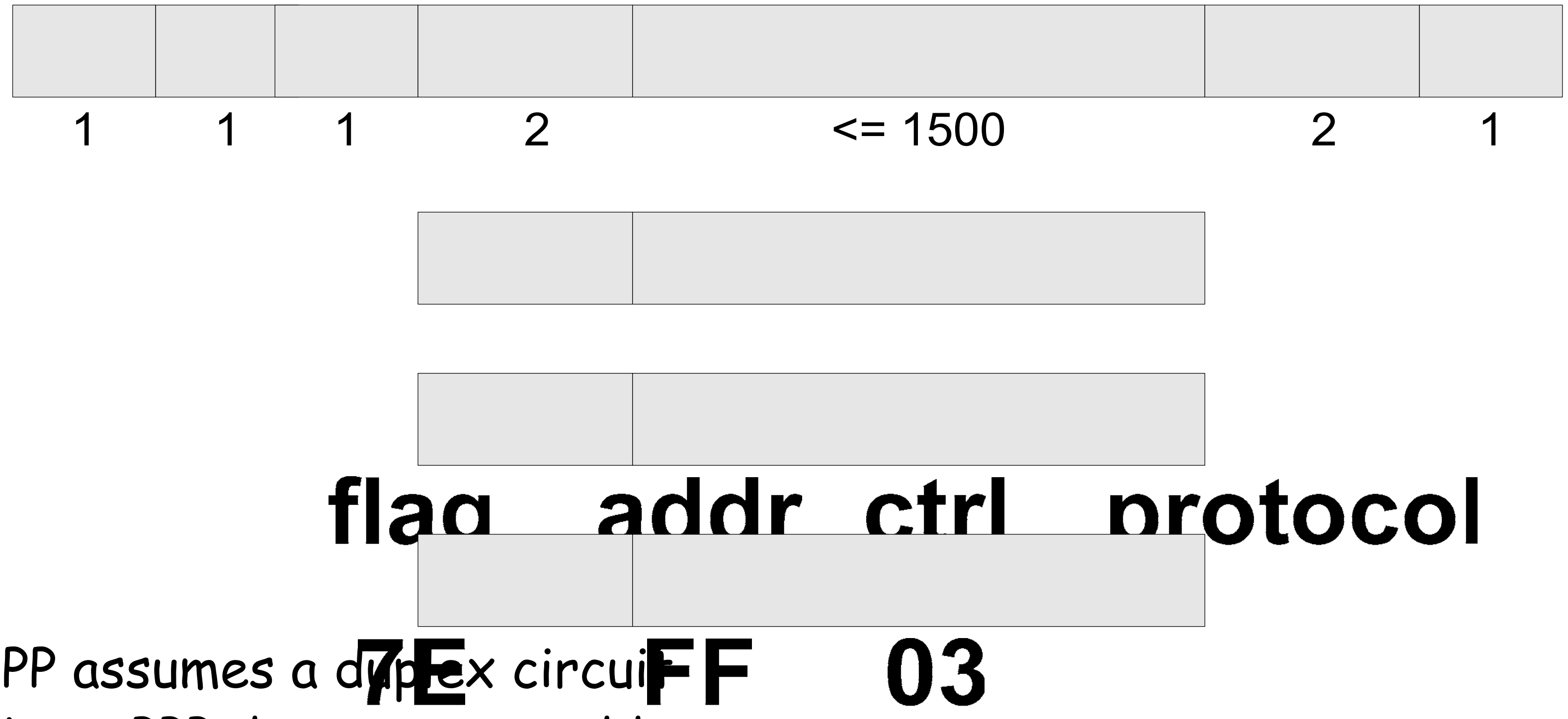
Point-to-Point Links

Data Link Protocols for Point-to-Point links

- SLIP (Serial Line IP)
  - First protocol for sending IP datagrams over dial-up links (from 1988)
  - Encapsulation, not much else
- PPP (Point-to-Point Protocol):
  - Successor to SLIP (1992), with added functionality
  - Used for dial-in and for high-speed routers
- HDLC (High-Level Data Link) :
  - Widely used and influential standard (1979)
  - Default protocol for serial links on Cisco routers
  - Actually, PPP is based on a variant of HDLC

PPP - IP encapsulation

- The frame format of PPP is similar to HDLC and the 802.2 LLC frame format:



- PPP assumes a duplex circuit
- Note: PPP does not use addresses
- Usual maximum frame size is 1500

Additional PPP functionality

- In addition to encapsulation, PPP supports:
  - multiple network layer protocols (protocol multiplexing)
  - Link configuration
  - Link quality testing
  - Error detection
  - Option negotiation
  - Address notification
  - Authentication

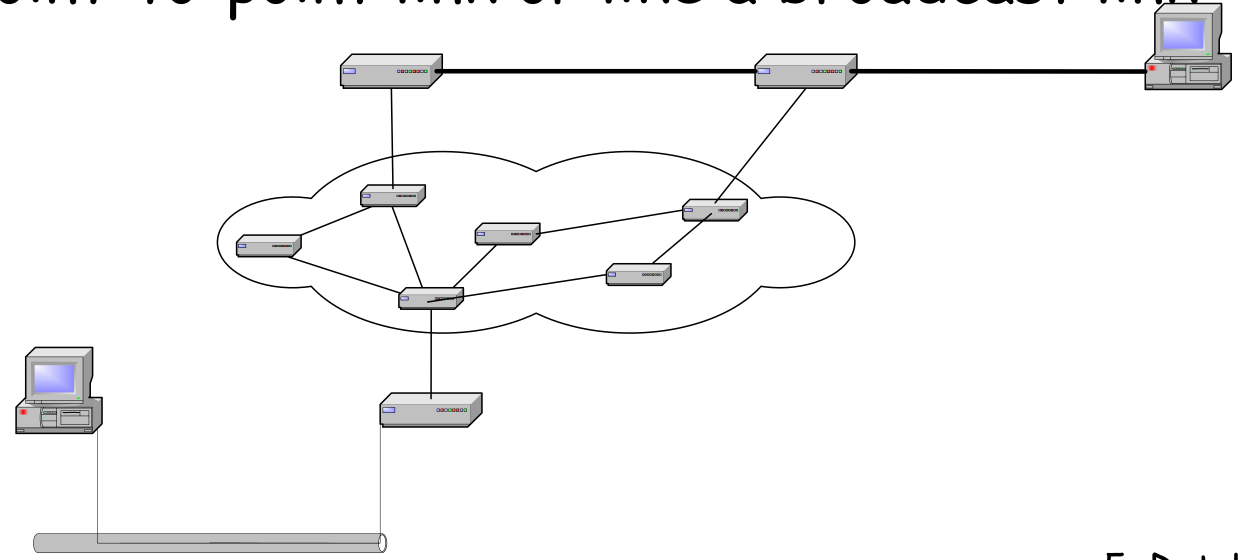
- The above functions are supported by helper protocols:
  - LCP
  - PAP, CHAP
  - NCP

PPP Support protocols

- **Link management:** The link control protocol (LCP) is responsible for establishing, configuring, and negotiating a data-link connection. LCP also monitors the link quality and is used to terminate the link.
- **Authentication:** Authentication is optional. PPP supports two authentication protocols: Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP).
- **Network protocol configuration:** PPP has network control protocols (NCPs) for numerous network layer protocols. The IP control protocol (IPCP) negotiates IP address assignments and other parameters when IP is used as network layer.

Switched networks

- Some data link technologies can be used to build complete networks, with their own addressing, routing, and forwarding mechanisms. These networks are often called switched networks.
- At the IP layer, a switched network may be like a point-to-point link or like a broadcast link



Switched networks

Data link layer technologies:

- Switched Ethernet
- ATM (Asynchronous Transfer Mode)
- Frame Relay
- Multiprotocol Label Switching (MPLS)

- Some switched networks are intended for enterprise networks (Switched Ethernet), wide area networks (MPLS, Frame Relay), or both (ATM)

LAN Addresses and ARP

32-bit IP address:

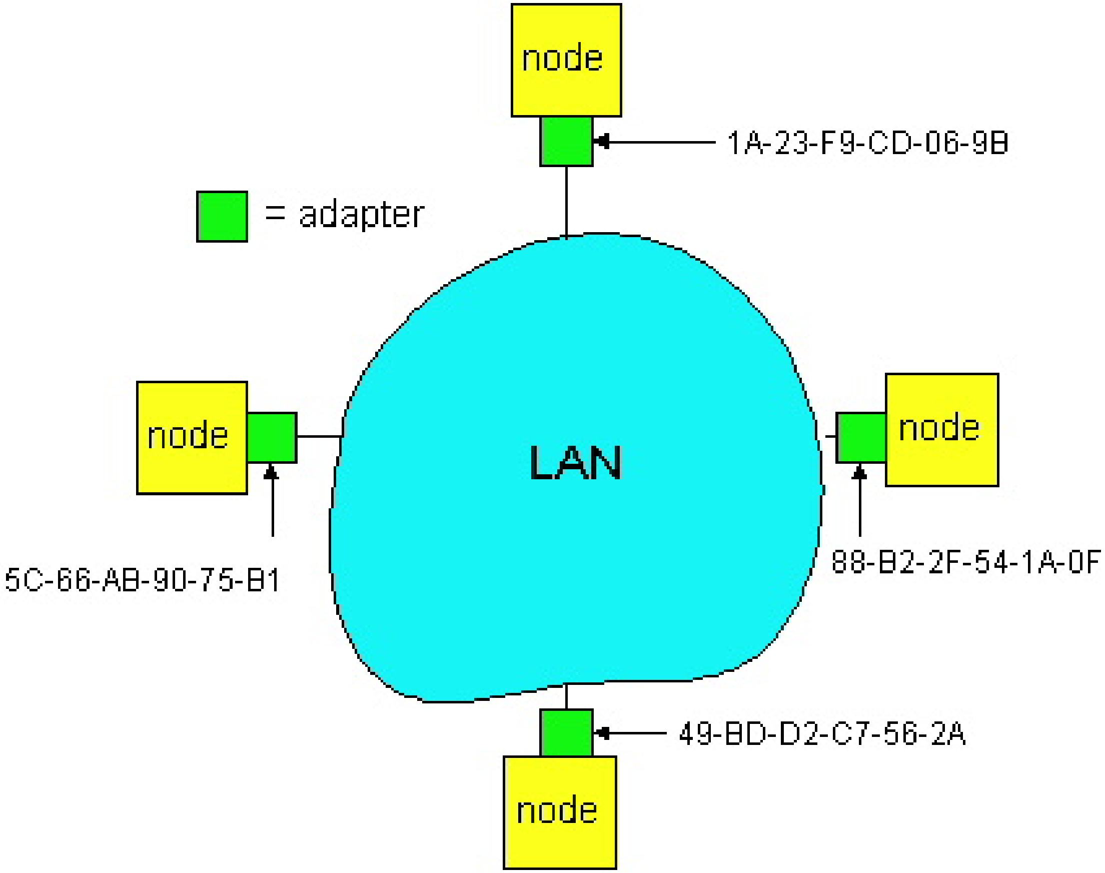
- *network-layer* address
- used to get datagram to destination IP network (recall IP network definition)

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



LAN Address (more)

- MAC address allocation administered by IEEE
- manufacturer buys portion of MAC address space (to assure uniqueness)
- Analogy:

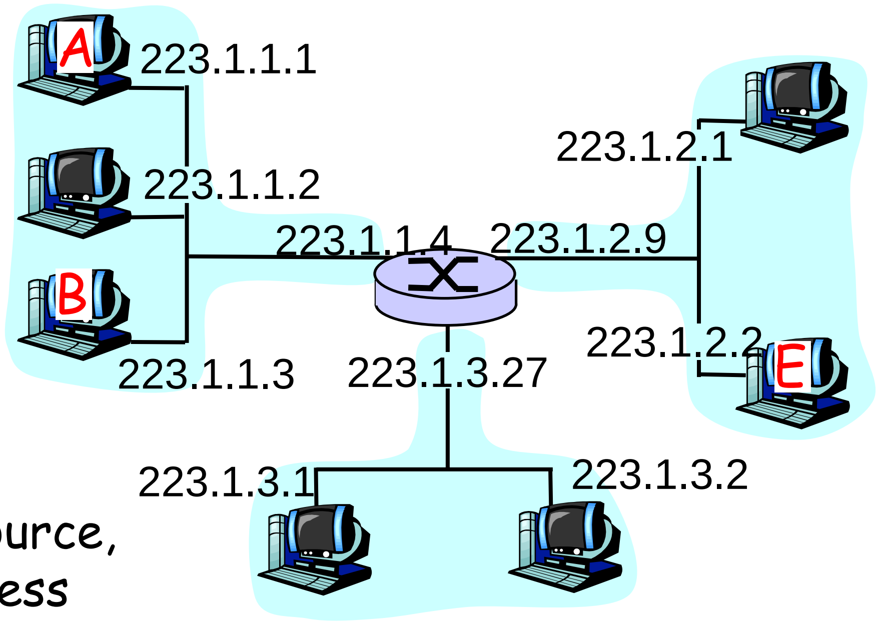
  (a) MAC address: like Social Security Number

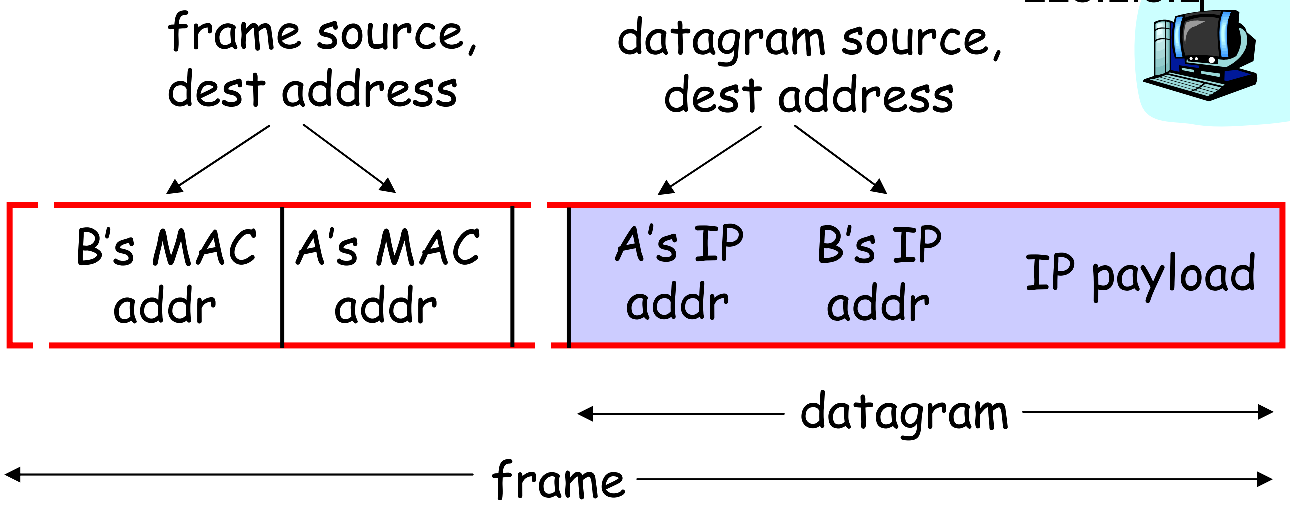
  (b) IP address: like postal address
- MAC flat address => portability
  - can move LAN card from one LAN to another
- IP hierarchical address NOT portable
  - depends on IP network to which node is attached

Recall earlier routing discussion

Starting at A, given IP datagram addressed to B:

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





frame source, dest address

datagram source, dest address

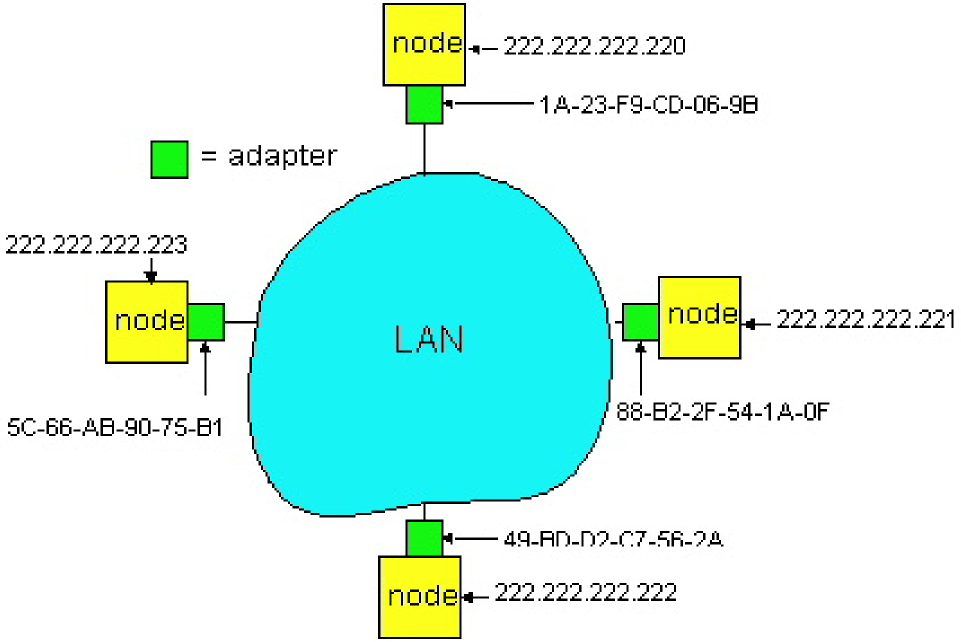
B's MAC addr	A's MAC addr	A's IP addr	B's IP addr	IP payload

datagram

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 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

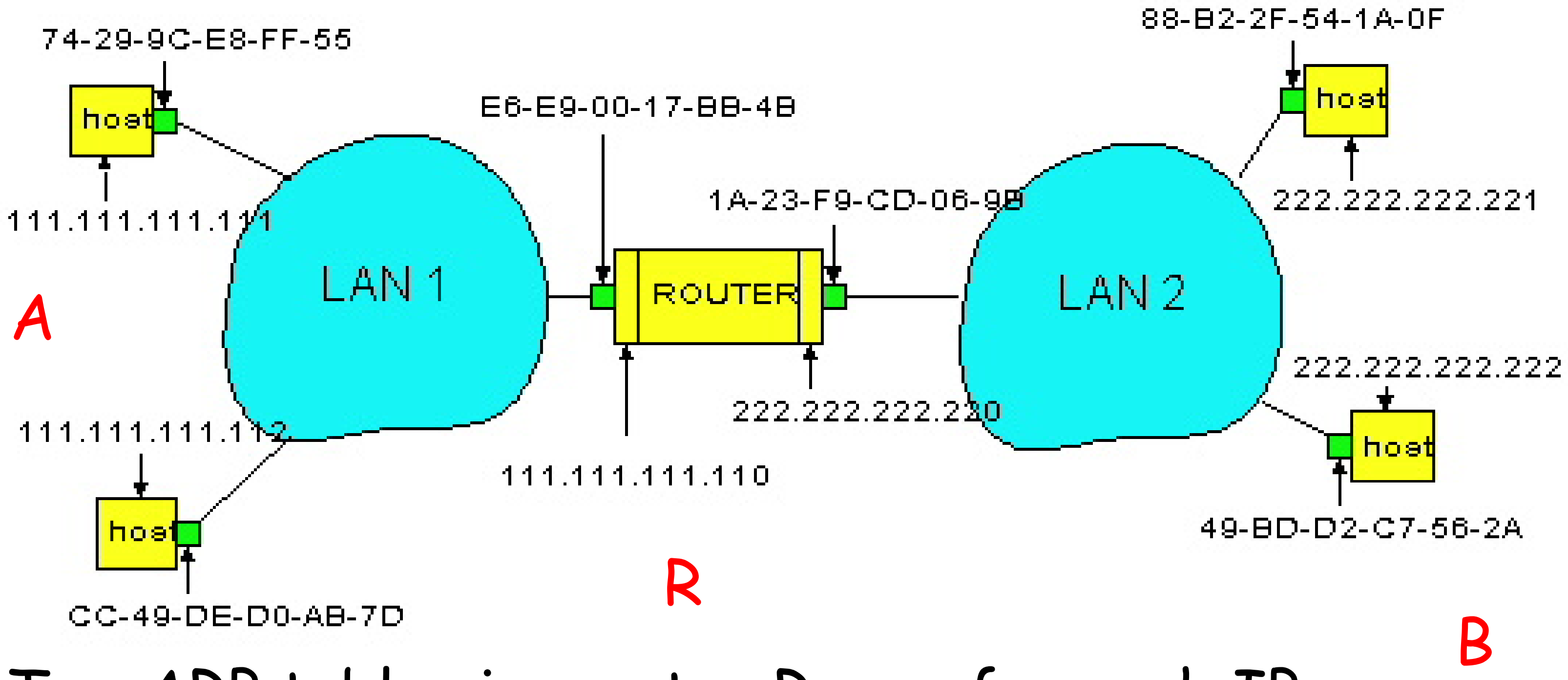
- A wants to send datagram to B, and A knows B's IP address.
- Suppose B's MAC address is not in A's ARP table.
- A broadcasts ARP query packet, containing B's IP address
  - all machines on LAN receive ARP query
- B receives ARP packet, replies to A with its (B's) MAC address
  - frame sent to A's MAC address (unicast)

- A caches (saves) IP-to-MAC address pair in its ARP table until information becomes old (times out)
  - soft state: information that times out (goes away) unless refreshed
- ARP is "plug-and-play":
  - nodes create their ARP tables without intervention from net administrator

Routing to another LAN

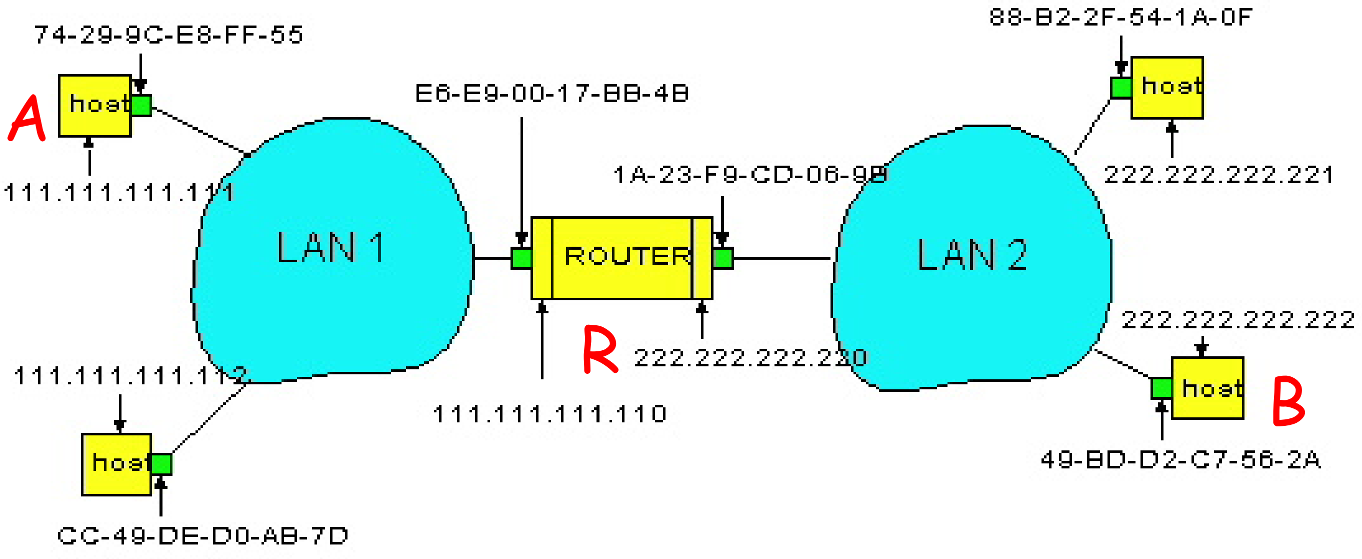
walkthrough: send datagram from A to B via R

assume A knows B IP address



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

- A creates datagram with source A, destination B
- A uses ARP to get R's MAC address for 111.111.111.110
- A creates link-layer frame with R's MAC address as dest, frame contains A-to-B IP datagram
- A's 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



Chapter 5 outline

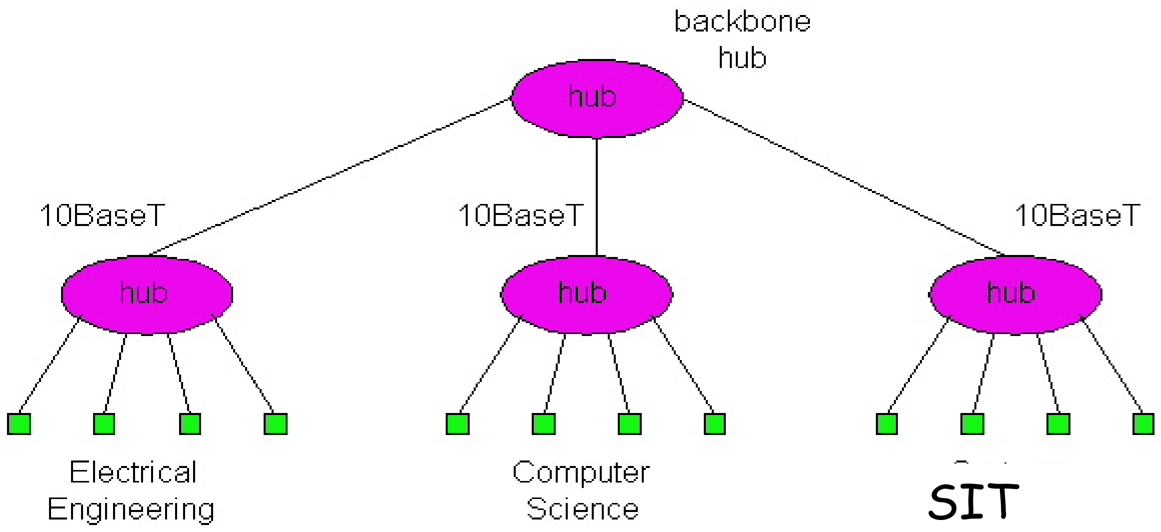
- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay

Interconnecting LAN segments

- Hubs
- Bridges
- Switches
    - Remark: switches are essentially multi-port bridges.
    - What we say about bridges also holds for switches!

Interconnecting with hubs

- Backbone hub interconnects LAN segments
- Extends max distance between nodes
- But individual segment collision domains become one large collision domain
  - if a node in CS and a node EE transmit at same time: collision
- Can't interconnect 10BaseT & 100BaseT

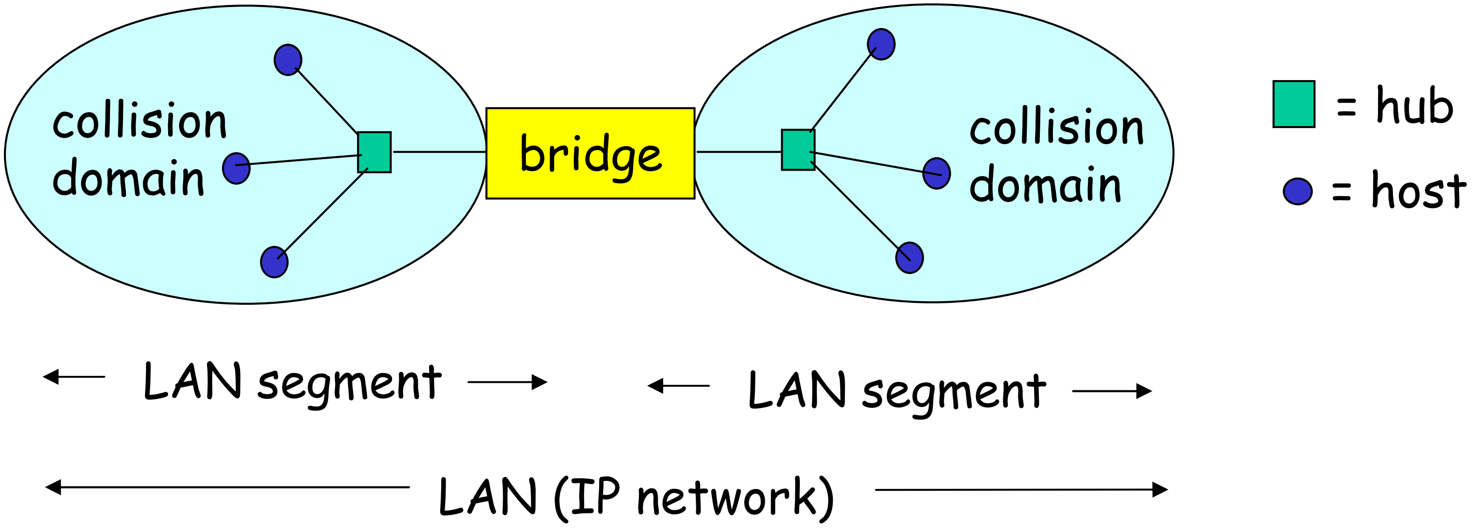


Bridges

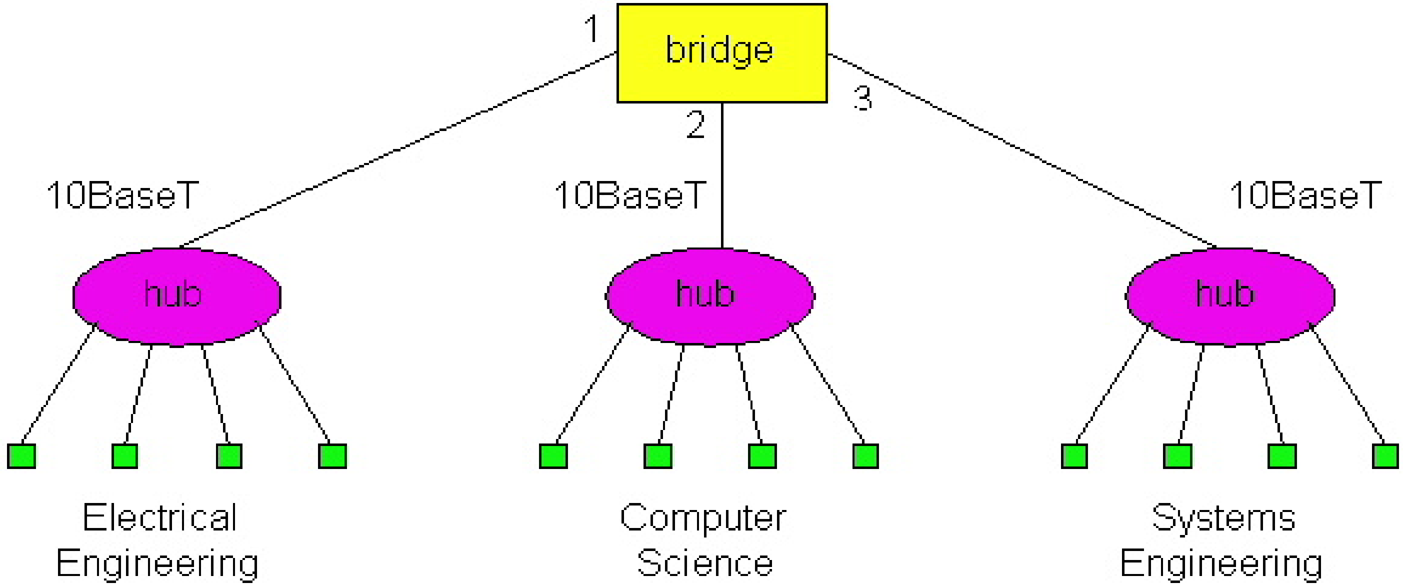
- Link layer device
  - stores and forwards Ethernet frames
  - examines frame header and selectively forwards frame based on MAC dest address
  - when frame is to be forwarded on segment, uses CSMA/CD to access segment
- transparent
  - hosts are unaware of presence of bridges
- plug-and-play, self-learning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

Bridges: traffic isolation

- Bridge installation breaks LAN into LAN segments
- bridges filter packets:
  - same-LAN-segment frames not usually forwarded onto other LAN segments
  - segments become separate collision domains



Forwarding



How to determine to which LAN segment to forward frame?

- Looks like a routing problem...

Self learning

- A bridge has a bridge table
- entry in bridge table:
  - (Node LAN Address, Bridge Interface, Time Stamp)
  - stale entries in table dropped (TTL can be 60 min)
- bridges *learn* which hosts can be reached through which interfaces
  - when frame received, bridge "learns" location of sender: incoming LAN segment
  - records sender/location pair in bridge table

Filtering/Forwarding

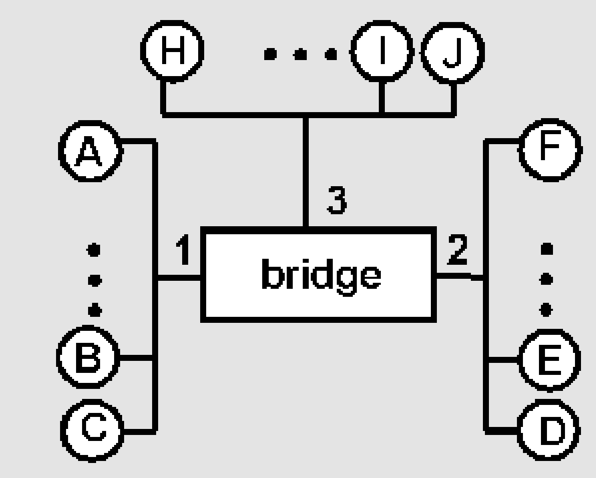
When bridge receives a frame:

```
index bridge table using MAC dest address
if  entry found for destination
    then{
        if  dest on segment from which frame arrived
            then drop the frame
            else forward the frame on interface indicated
        }
    else flood
```

forward on all but the interface on which the frame arrived

Bridge example

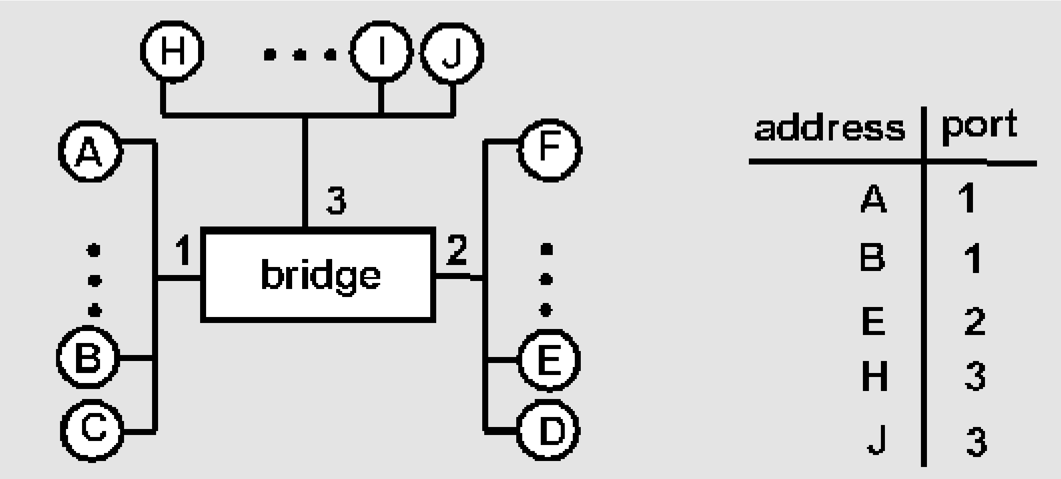
Suppose C sends frame to D and D replies back with frame to C.



address	port
A	1
B	1
E	2
H	3
J	3

- Bridge receives frame from from C
  - Its notes in the bridge table that C is on interface 1
  - because D is not yet in the table, the bridge sends a frame to interfaces 2 and 3
- frame received by D

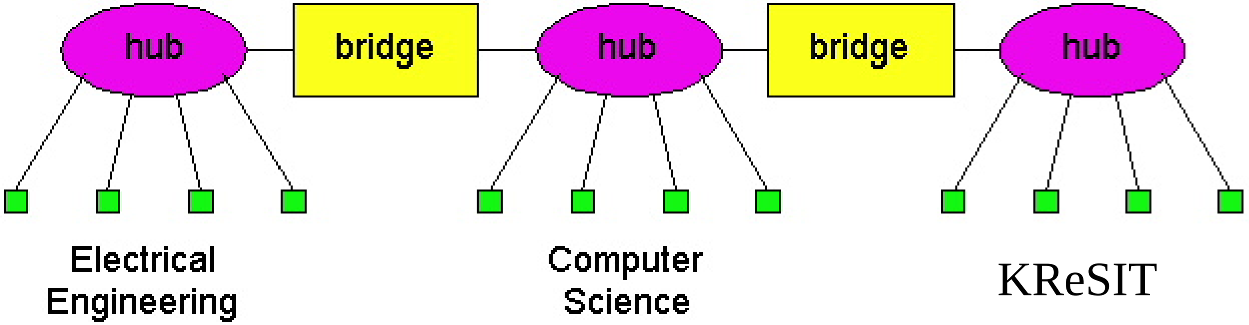
Bridge Learning: example



address	port
A	1
B	1
E	2
H	3
J	3

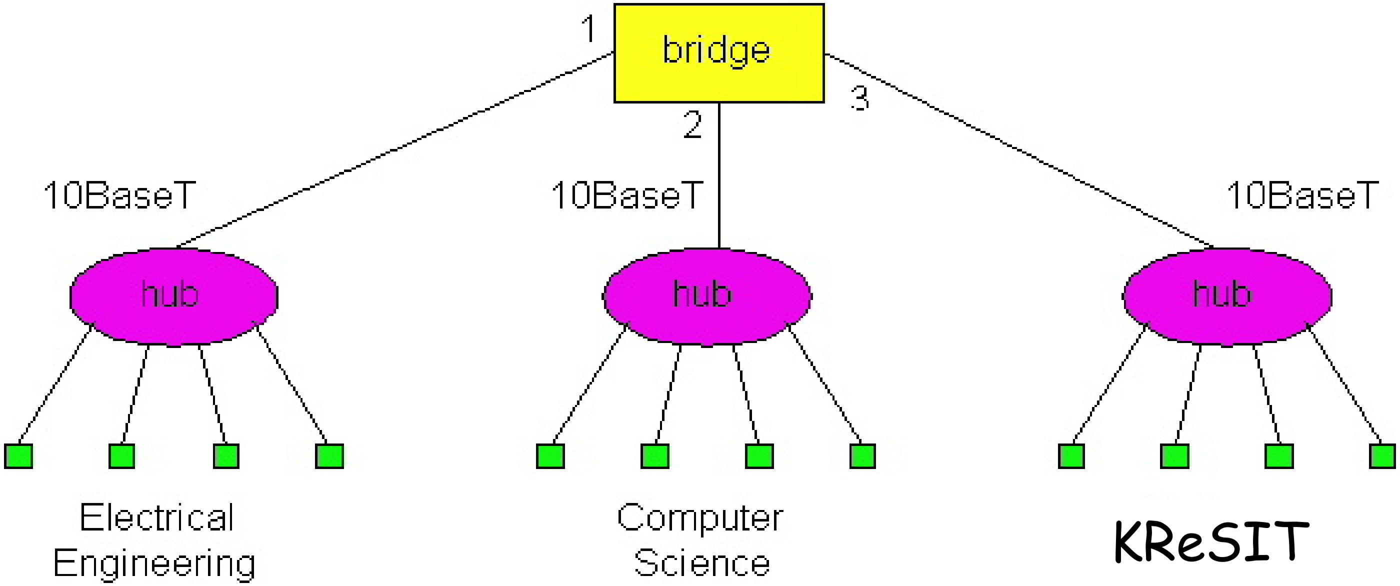
- D generates frame for C, and sends it
- bridge receives frame
  - notes in bridge table that D is on interface 2
  - bridge knows C is on interface 1, so ***selectively*** forwards frame to interface 1

Interconnection without backbone



- Not recommended for two reasons:
  - single point of failure at Computer Science hub
  - all traffic between EE and IT must path over CS segment

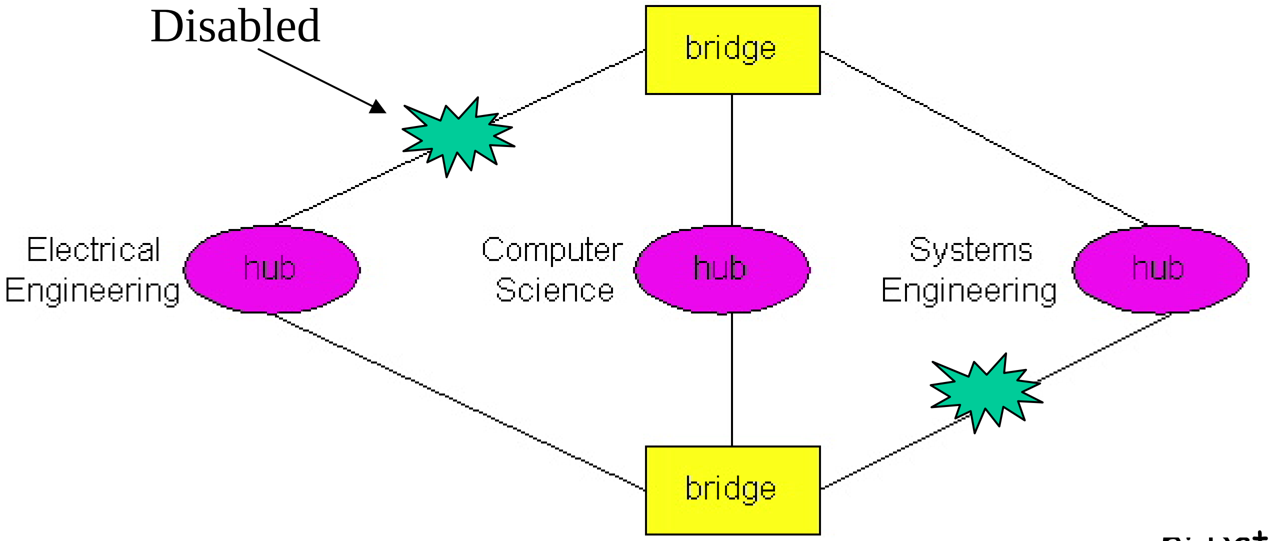
Backbone configuration (star)



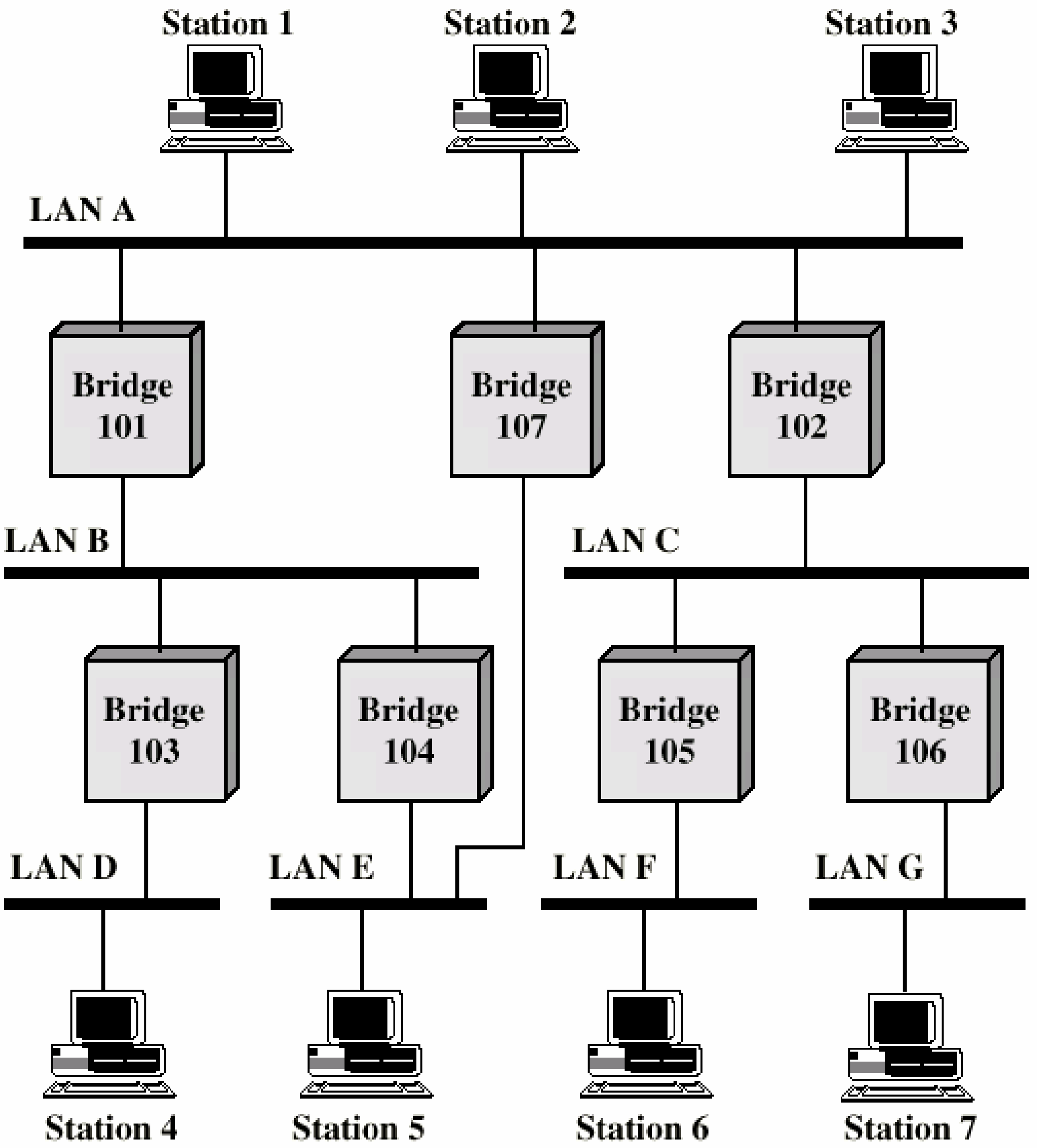
Recommended !

Bridges Spanning Tree

- for increased reliability, desirable to have redundant, alternative paths from source to dest
- with multiple paths, cycles result - bridges may multiply and forward frame forever
- solution: organize bridges in a spanning tree by disabling subset of interfaces



Multiple LANs



Needed: Routing

- Complex large LANs need alternative routes
  - Load balancing
  - Fault tolerance
- Bridge must decide whether to forward frame
  - Bridge must decide which LAN to forward frame on
- Routing selected for each source-destination pair of LANs
  - Done in configuration
  - Usually least hop route
  - Only changed when topology changes

Spanning Tree

- Bridge automatically develops routing table
- Automatically update in response to changes
  - Frame forwarding
  - Address learning
  - Loop resolution

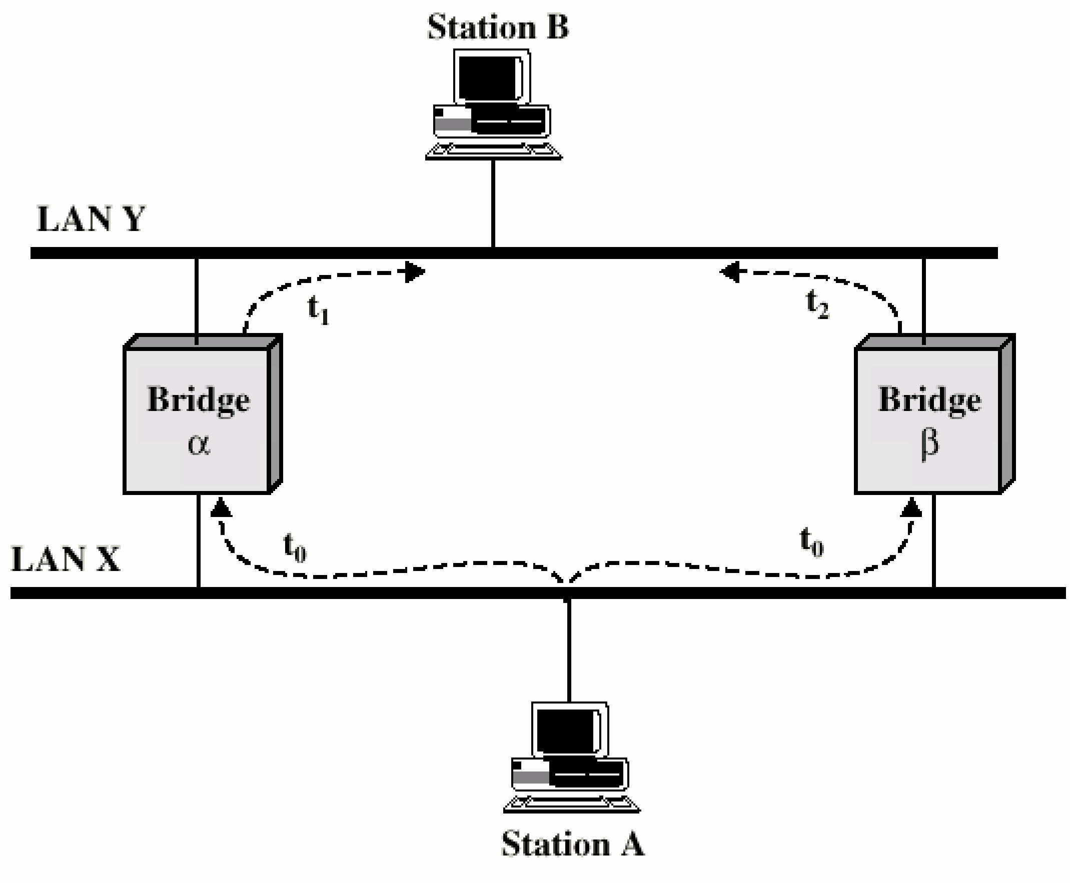
Frame forwarding

- Maintain forwarding database for each port
    - List station addresses reached through each port
- For a frame arriving on port X:
    - Search forwarding database to see if MAC address is listed for any port except X
    - If address not found, forward to all ports except X
    - If address listed for port Y, check port Y for blocking or forwarding state
        - Blocking prevents port from receiving or transmitting
    - If not blocked, transmit frame through port Y

Address Learning

- When frame arrives at port X, it has come form the LAN attached to port X
- Use the source address to update forwarding database for port X to include that address
- Timer on each entry in database (reset whenever frame received)
- Each time frame arrives, source address checked against forwarding database

Loop of Bridges



Spanning Tree Algorithm

- Creates a logical, or “active“ topology that behaves like a spanning tree
  - Makes alternate bridges redundant
  - Is run periodically, so will discover failures and use alternate bridges if necessary

Reference: Fred Halsall: “Data Communications, Computer Networks and Open Systems”, 4th Edition.

Spanning Tree Algorithm

- Variables:
  1. Each bridge has a **Priority Value** and a unique **Identifier (ID)**
  2. Each LAN segment has a Designated Cost (DC) inversely proportional to the bit rate
  3. Each port of a bridge has a Path Cost (PC) = DC of the LAN segment to which it is attached

Spanning Tree Algorithm

- Working: Bridges regularly exchange frames known as Bridge Protocol Data Units (BPDUs). This exchange does the following:

1. Bridge with highest priority and smallest ID is selected as root bridge.
2. Each bridge determines for each port, the least cost path from root bridge to this port. This is the Root Path Cost (RPC) for *this* port.
   a) Select the port which has the least RPC and designate it as the Root Port (RP). This is the port which will be used for communicating with the root.
3. Once a root port is determined, one bridge port is selected for each LAN segment as the designated bridge port (DP) to which frames will be sent for that LAN segment.
   a) This is a port (**which is NOT a root port**) which has the least path cost to the root
   b) The ports of the root bridge are always DPs for the LAN segments connected to the root bridge
4. The state of the bridge ports can be set either to forwarding or blocking.
   a) All ports that are either RPs or DPs are forwarding, the rest are blocking.

Topology Initialization

- BPDUs are sent to a broadcast MAC address of all bridges on the LAN
- Each BPDU contains (self ID, root ID, transmitting port ID, RPC of this port)
- If necessary,
  - Update root ID based on received BPDUs
  - Add path cost of the port on which frame was received to the RPC in the frame
  - Sends out this new info on all other ports with all updated Ids
  - Procedure repeated by all bridges
    - Will determine RPCs of each port
    - Will select Root Ports based on this
  - Two or more bridges on the same segment will exchange BPDUs so that designated bridge-port can be seleted

Topology Change

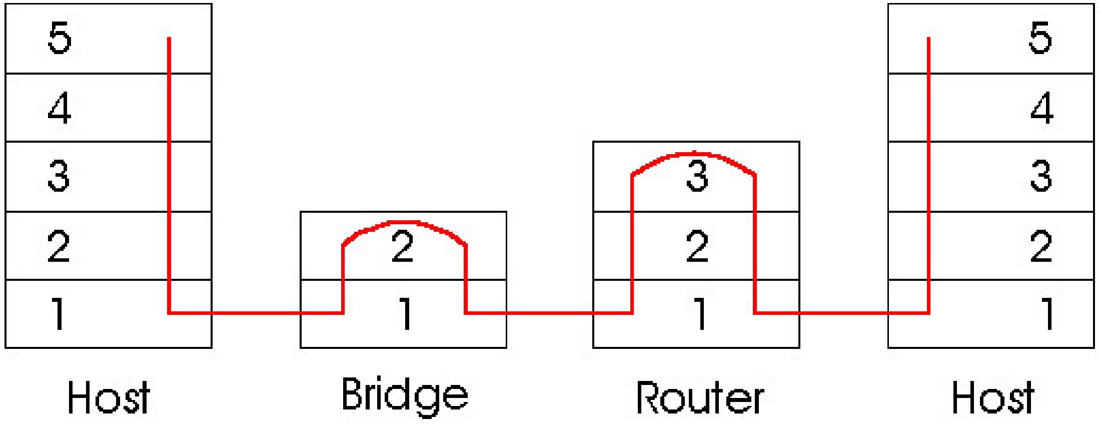
- Root bridge regularly transmits BPDUs, forwarded by all bridges on all ports
- Bridges will keep timers associated with each of its forwarding ports
- When timers expire, procedure similar to topology initialization is done
  - Details…

Some bridge features

- Isolates collision domains resulting in higher total max throughput
- limitless number of nodes and geographical coverage
- Can connect different Ethernet types (though not preferable)
- Transparent ("plug-and-play"): no configuration necessary

Bridges vs. Routers

- both store-and-forward devices
  - routers: network layer devices (examine network layer headers)
  - bridges are link layer devices
- routers maintain routing tables, implement routing algorithms
- bridges maintain bridge tables, implement filtering, learning and spanning tree algorithms



Routers vs. Bridges

Bridges + and -

+ Bridge operation is simpler requiring less packet processing

+ Bridge tables are self learning

- All traffic confined to spanning tree, even when alternative bandwidth is available

- Bridges do not offer protection from broadcast storms

Routers vs. Bridges

Routers + and -

+ arbitrary topologies can be supported, cycling is limited by TTL counters (and good routing protocols)

+ provide protection against broadcast storms

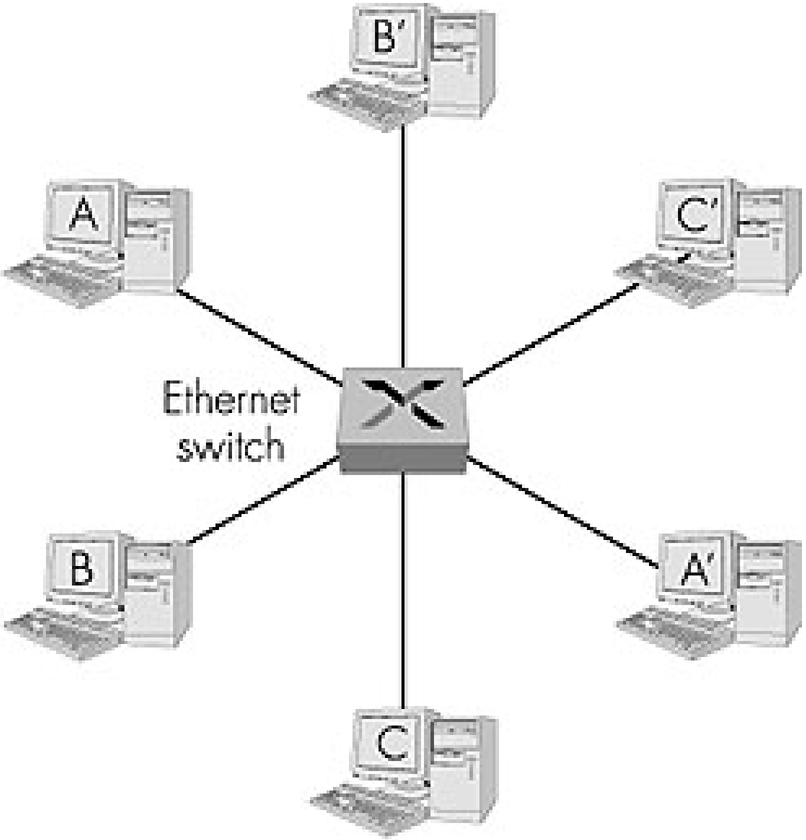
- require IP address configuration (not plug and play)

- require higher packet processing

- ❒ bridges do well in small (few hundred hosts) while routers used in large networks (thousands of hosts)

Ethernet Switches

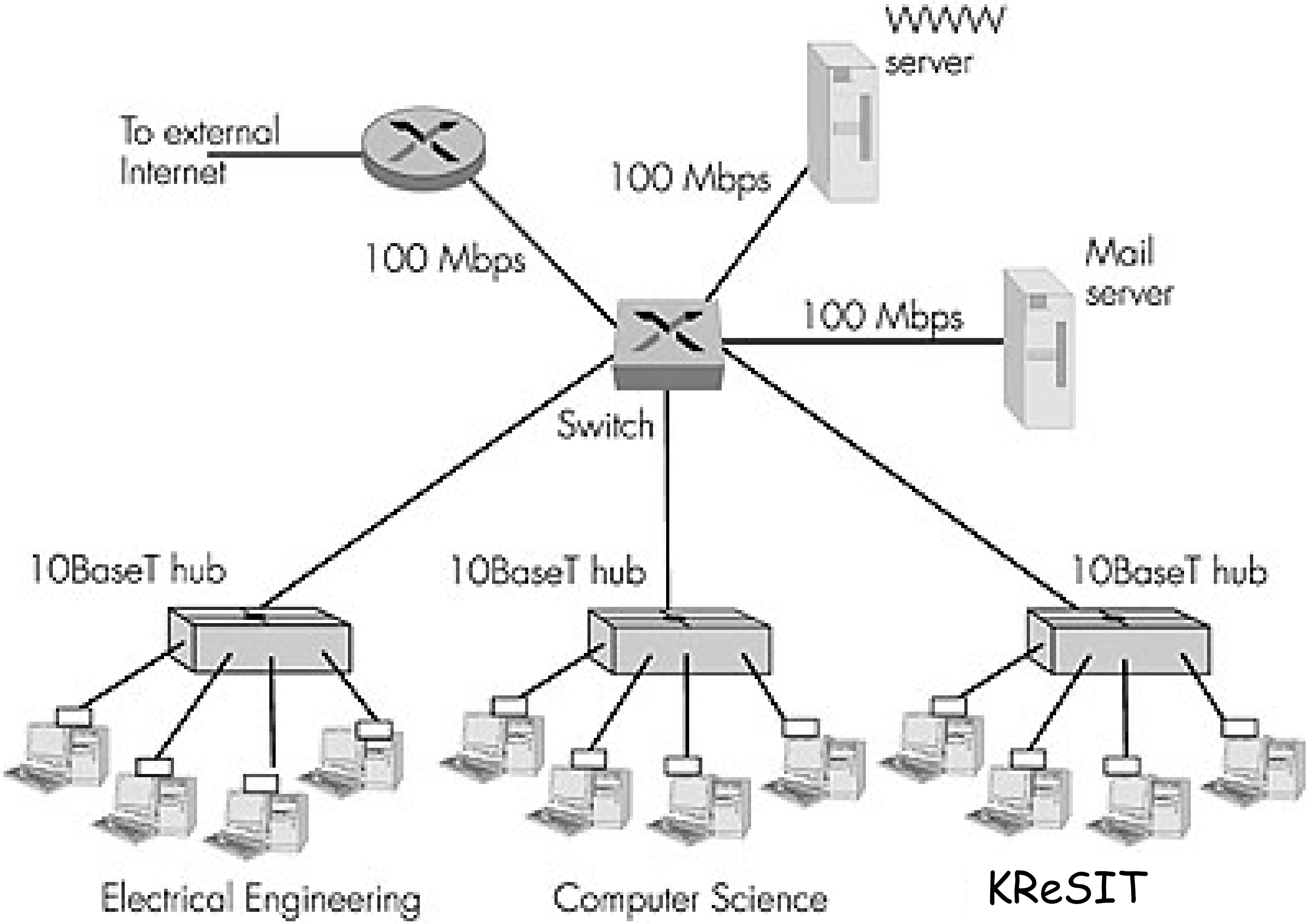
- Essentially a multi-interface bridge
- layer 2 (frame) forwarding, filtering using LAN addresses
- Switching: A-to-A' and B-to-B' simultaneously, no collisions
- large number of interfaces
- often: individual hosts, star-connected into switch
  - Ethernet, but no collisions!



Ethernet Switches

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

Not an atypical LAN (IP network)



Summary comparison

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

Chapter 5 outline

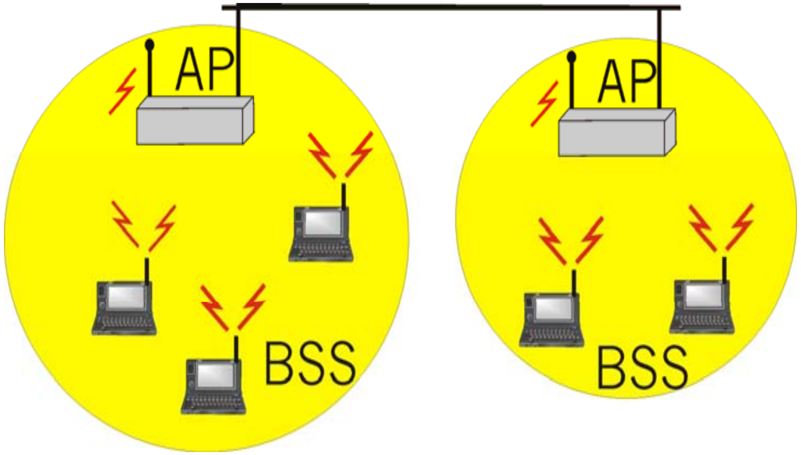
- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay

IEEE 802.11 Wireless LAN

- 802.11b
  - 2.4-5 GHz unlicensed radio spectrum
  - up to 11 Mbps
  - direct sequence spread spectrum (DSSS) in physical layer
    - all hosts use same chipping code
  - widely deployed, using base stations
- 802.11a
  - 5-6 GHz range
  - up to 54 Mbps
- 802.11g
  - 2.4-5 GHz range
  - up to 54 Mbps
- All use CSMA/CA for multiple access
- All have base-station and ad-hoc network versions

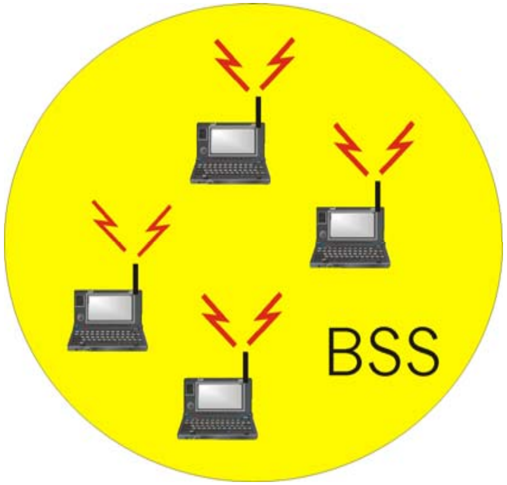
Base station approach

- Wireless host communicates with a base station
  - base station = access point (AP)
- Basic Service Set (BSS) (a.k.a. "cell") contains:
  - wireless hosts
  - access point (AP): base station
- BSSs combined to form distribution system (DS)



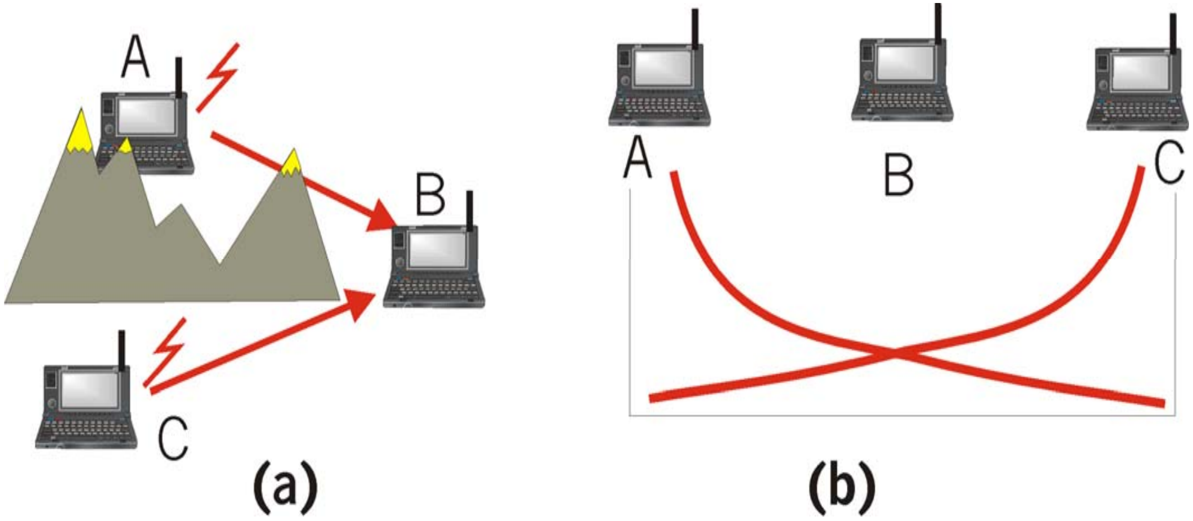
Ad Hoc Network approach

- No AP (i.e., base station)
- wireless hosts communicate with each other
  - to get packet from wireless host A to B may need to route through wireless hosts X,Y,Z
- Applications:
  - "laptop" meeting in conference room, car
  - interconnection of "personal" devices
  - battlefield
- IETF MANET (Mobile Ad hoc Networks) working group



IEEE 802.11: multiple access

- Collision if 2 or more nodes transmit at same time
- CSMA makes sense:
  - get all the bandwidth if you're the only one transmitting
  - shouldn't cause a collision if you sense another transmission
- Collision detection doesn't work: hidden terminal problem



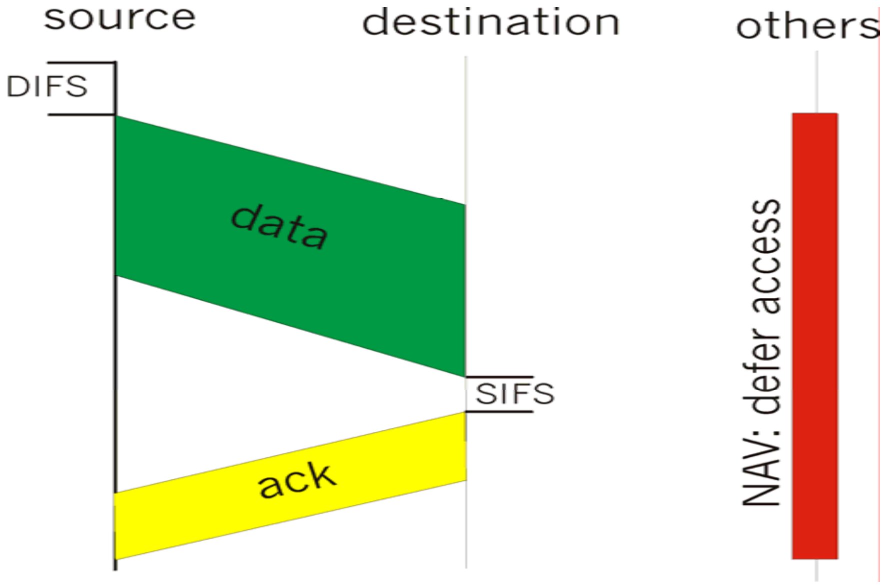
IEEE 802.11 MAC Protocol: CSMA/CA

802.11 CSMA: sender

- if sense channel idle for **DISF** sec.
  then transmit entire frame (no collision detection)
- -if sense channel busy
  then binary backoff

802.11 CSMA receiver

- if received OK
  return ACK after **SIFS**
  (ACK is needed due to hidden terminal problem)



DIFS: Distributed interframe space

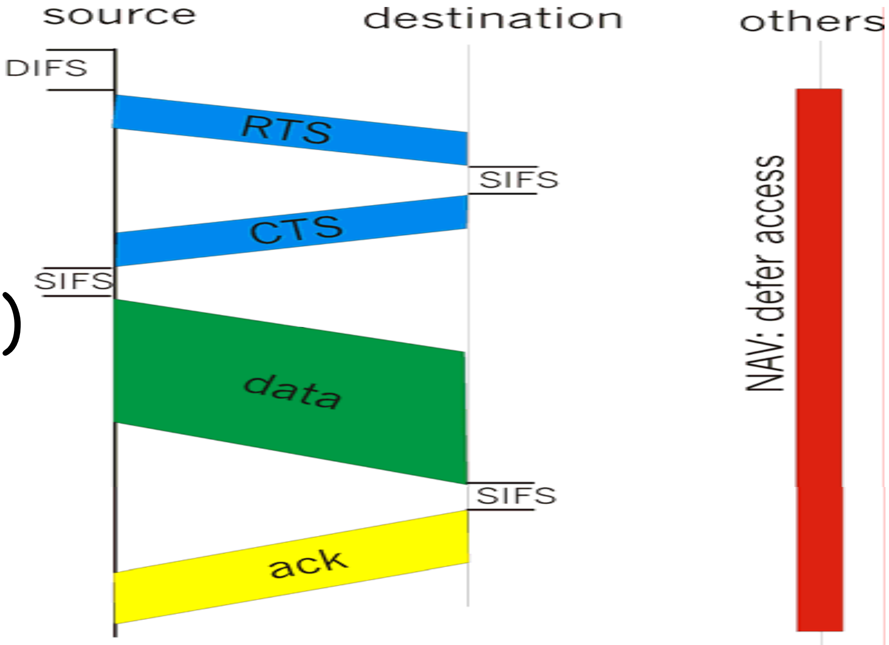
SIFS: Short Interframe space

Collision avoidance mechanisms

- Problem:
  - two nodes, hidden from each other, transmit complete frames to base station
  - wasted bandwidth for long duration !
- Solution:
  - small reservation packets
  - nodes track reservation interval with internal "network allocation vector" (NAV)

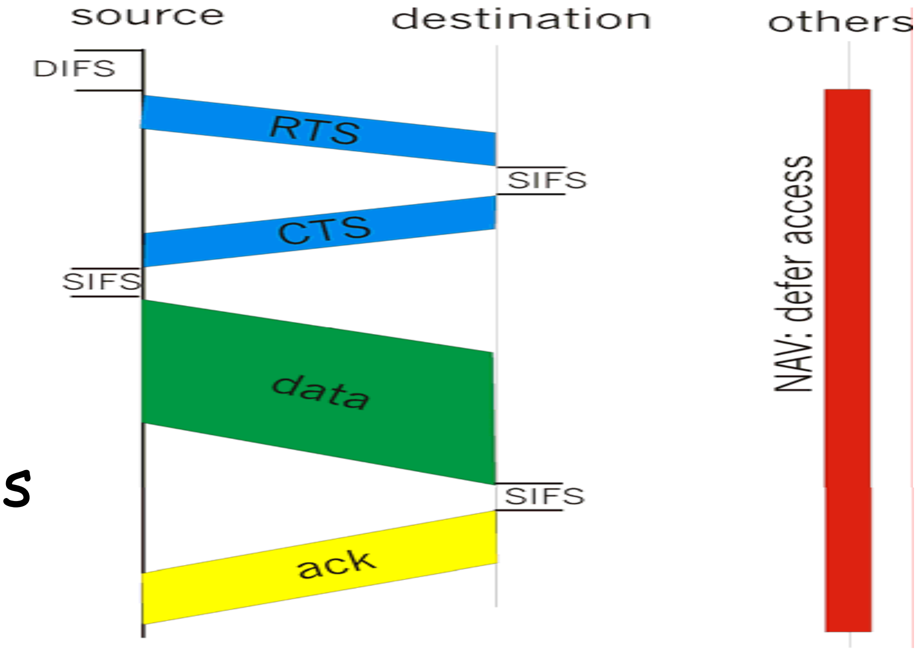
Collision Avoidance: RTS-CTS exchange

- sender transmits short RTS (request to send) packet: indicates duration of transmission
- receiver replies with short CTS (clear to send) packet
  - notifying (possibly hidden) nodes
- hidden nodes will not transmit for specified duration



Collision Avoidance: RTS-CTS exchange

- RTS and CTS short:
  - collisions less likely, of shorter duration
  - end result similar to collision detection
- IEEE 802.11 allows:
  - CSMA
  - CSMA/CA: reservations
  - polling from AP



A word about Bluetooth

- Low-power, small radius, wireless networking technology
  - 10-100 meters
- omnidirectional
  - not line-of-sight infrared
- Interconnects gadgets
- 2.4-2.5 GHz unlicensed radio band
- up to 721 kbps

- Interference from wireless LANs, digital cordless phones, microwave ovens:
  - frequency hopping helps
- MAC protocol supports:
  - error correction
  - ARQ
- Each node has a 12-bit address

Chapter 5 outline

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

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

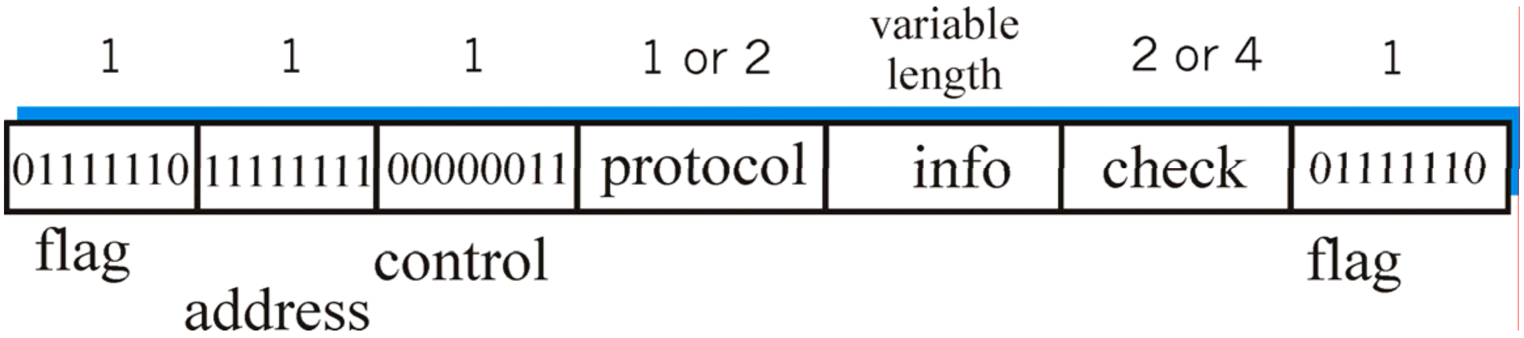
PPP non-requirements

- 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

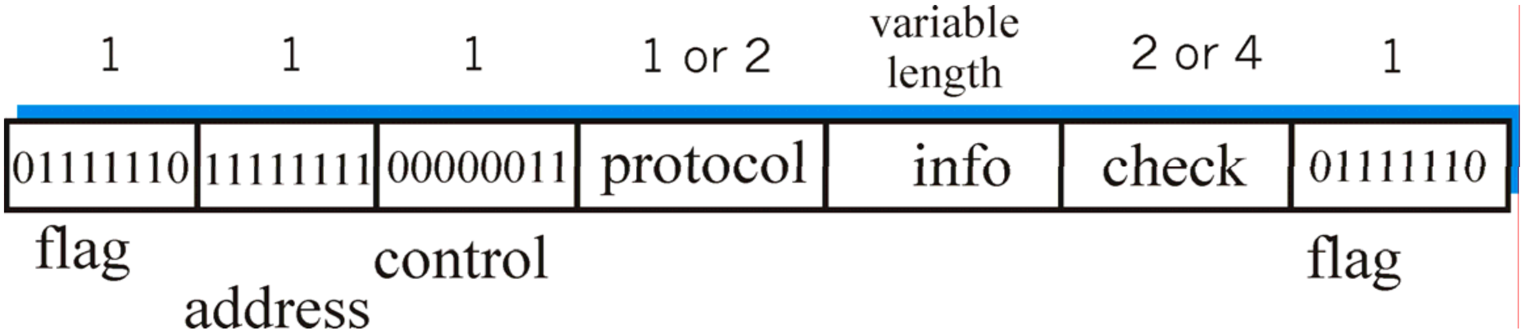
- Flag: delimiter (framing)
- Address: does nothing (only one option)
- Control: does nothing; in the future possible multiple control fields
- Protocol: upper layer protocol to which frame delivered (eg, PPP-LCP, IP, IPCP, etc)



1	1	1	1 or 2	variable length	2 or 4	1
01111110	11111111	00000011	protocol	info	check	01111110
flag	address	control				flag

PPP Data Frame

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



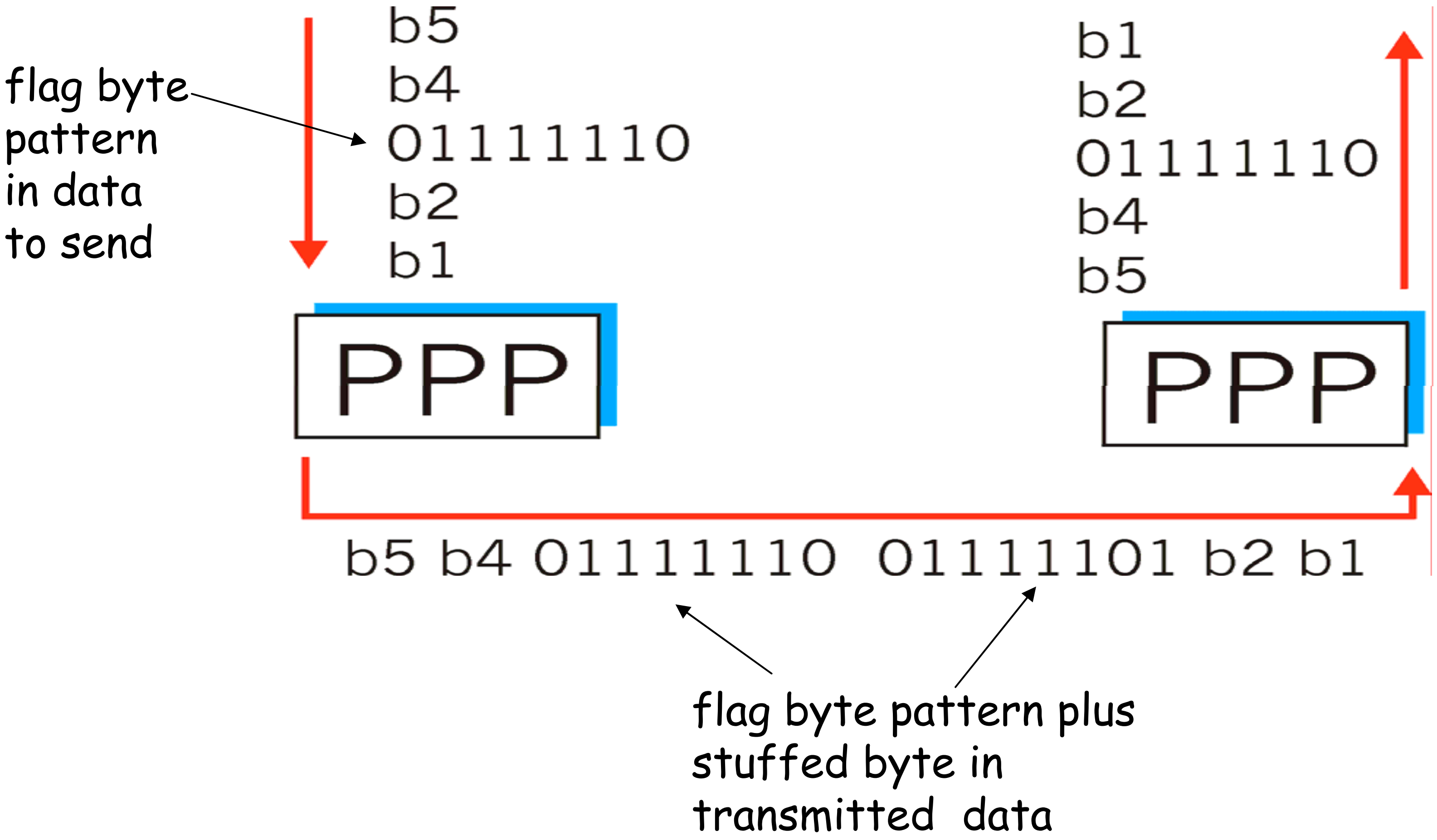
1	1	1	1 or 2	variable length	2 or 4	1
01111110	11111111	00000011	protocol	info	check	01111110
flag	address	control				flag

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



flag byte pattern in data to send

b5
b4
01111110
b2
b1

PPP

b1
b2
01111110
b4
b5

PPP

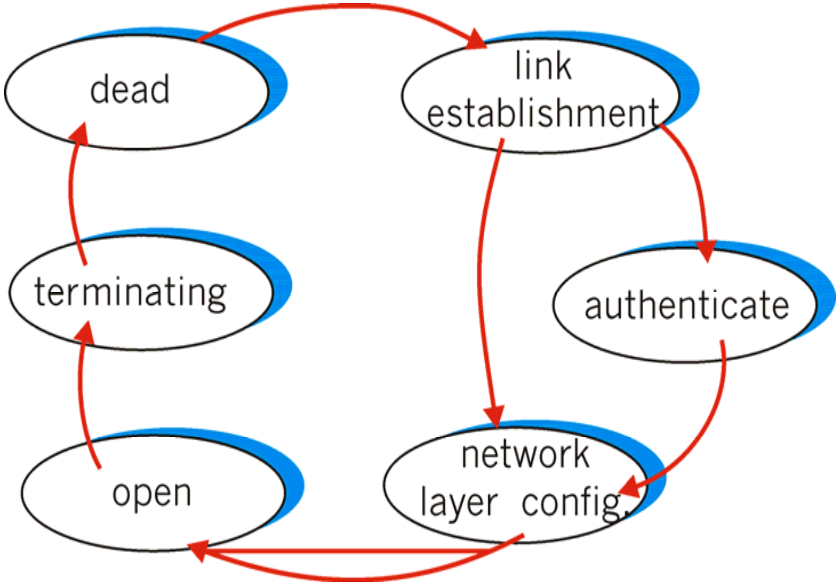
b5 b4 01111110 01111101 b2 b1

flag byte pattern plus stuffed byte in transmitted data

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



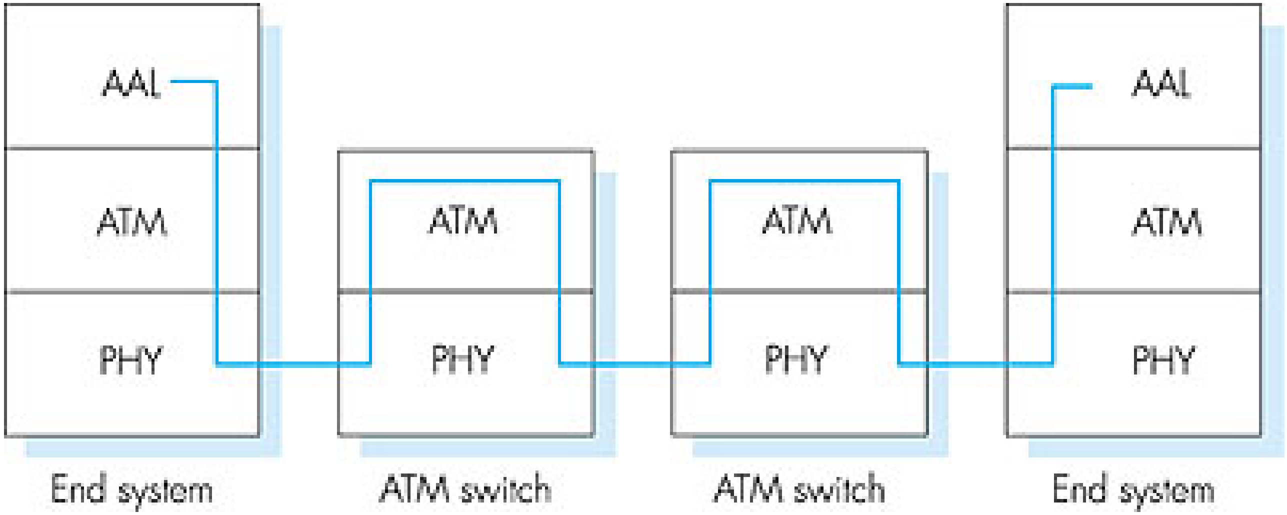
Chapter 5 outline

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

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 analogous to Internet transport layer
- ATM layer: “network” layer
  - cell switching, routing
- physical layer

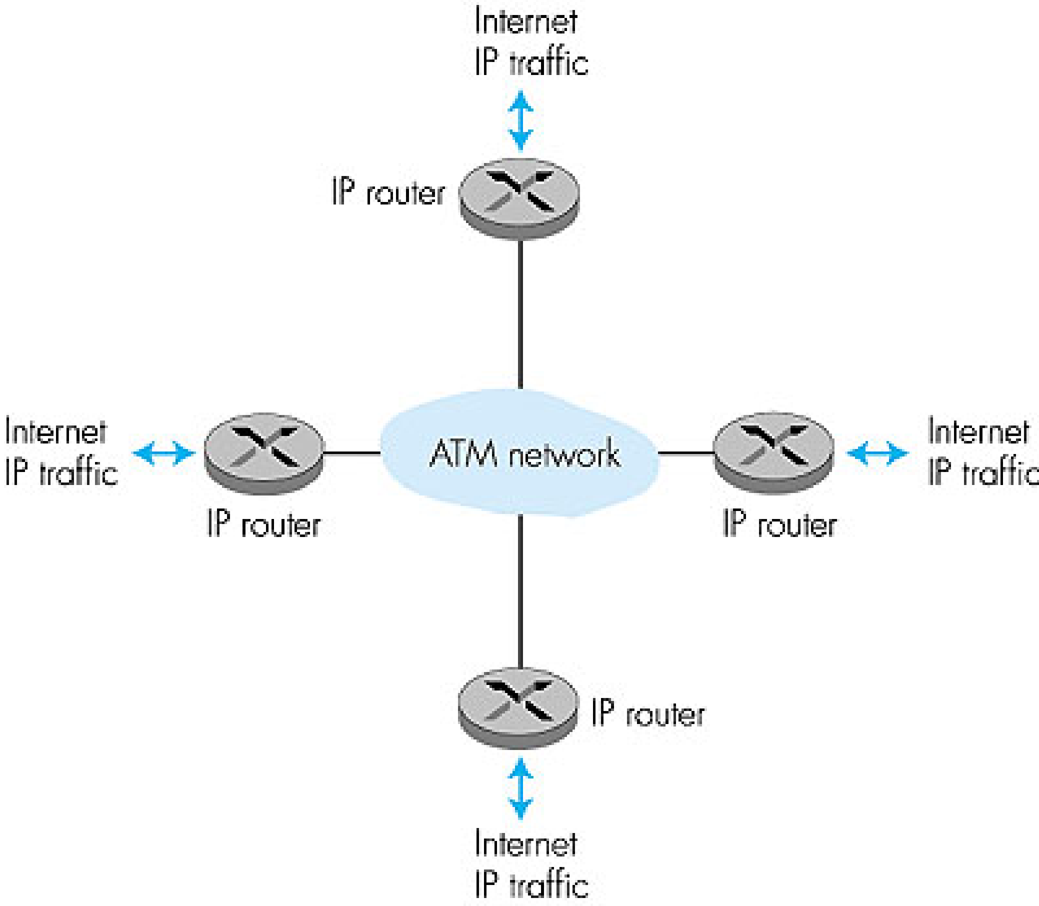
ATM: network or link layer?

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

- ATM *is* a network technology

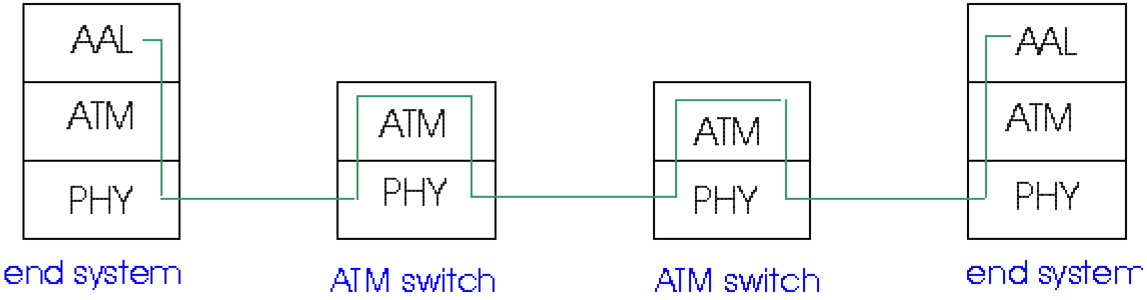
Reality: used to connect IP backbone routers

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



ATM Adaptation Layer (AAL)

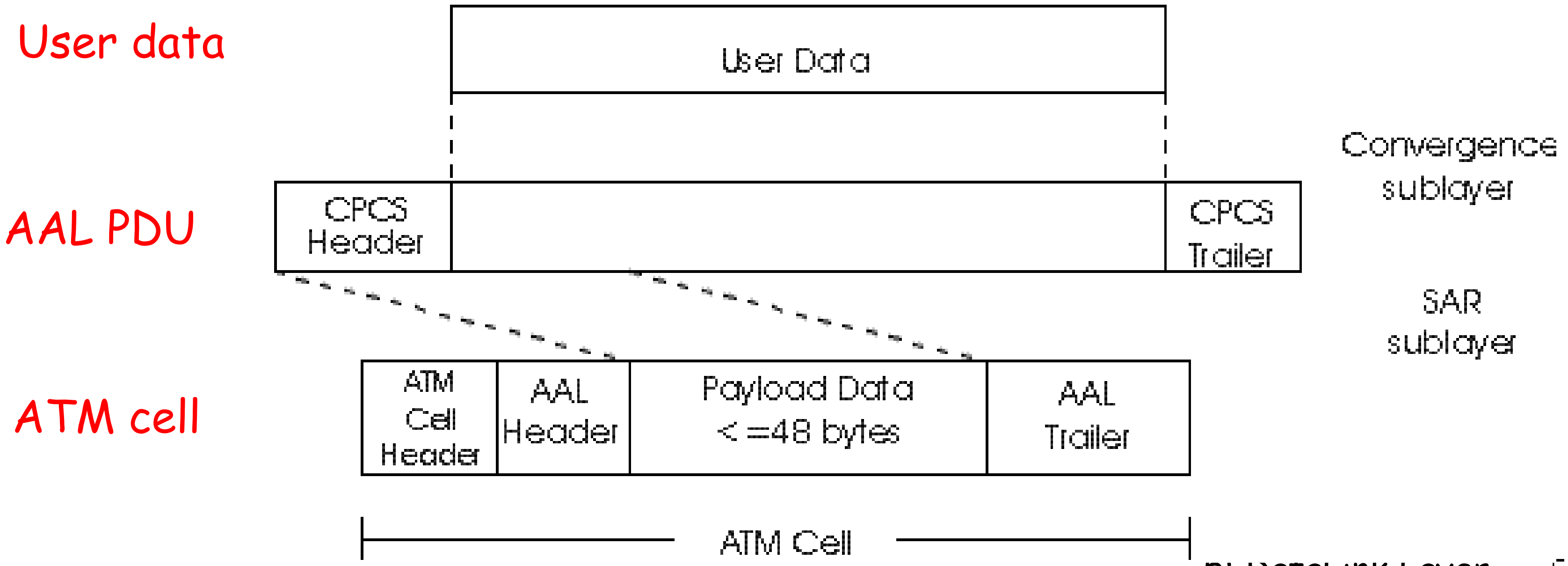
- ATM **Adaptation Layer** (AAL): "adapts" upper layers (IP or native ATM applications) to ATM layer below
- AAL present **only in end systems**, not in switches
- AAL layer segment (header/trailer fields, data) fragmented across multiple ATM cells
    - analogy: TCP segment in many IP packets



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 (e.g., IP datagrams)



AAL5 - Simple And Efficient AL (SEAL)

- **AAL5: low overhead** AAL used to carry IP datagrams
  - 4 byte cyclic redundancy check
  - PAD ensures payload multiple of 48bytes
  - large AAL5 data unit to be fragmented into 48-byte ATM cells

CPCS-PDU payload	PAD	Length	CRC
0-65535	0-47	2	4

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	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 two IP routers
- Switched VCs (SVC):
  - dynamically set up on per-call basis

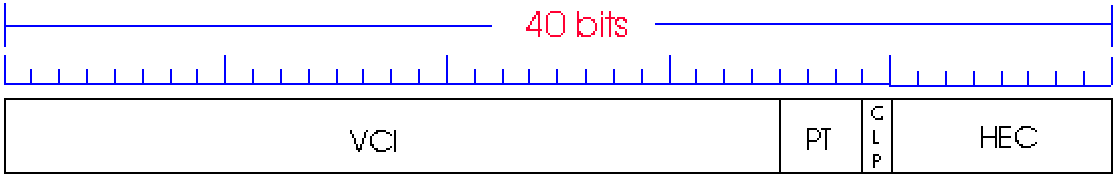
ATM VCs

- 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

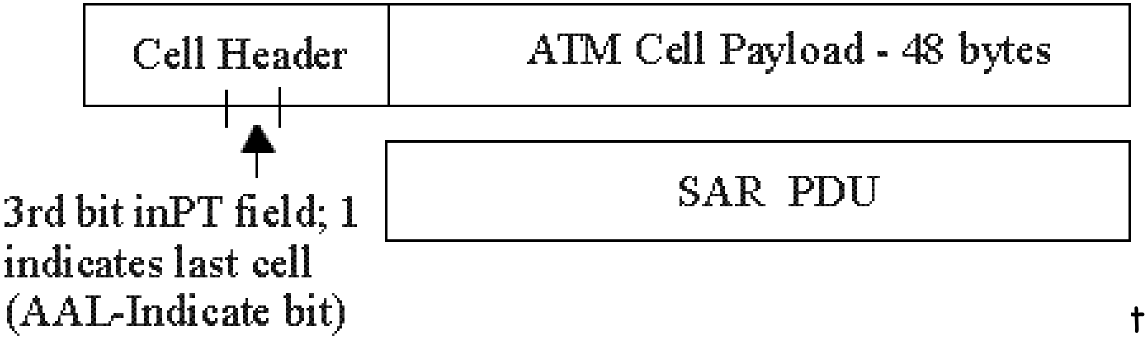
ATM Layer: ATM cell

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

Cell header

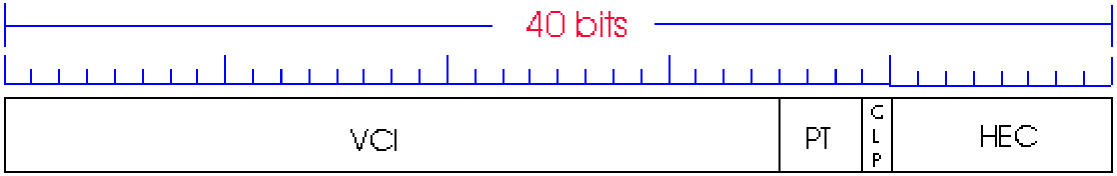


Cell format


Cell Header
ATM Cell Payload - 48 bytes
SAR PDU
3rd bit inPT field; 1 indicates last cell (AAL-Indicate bit)

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

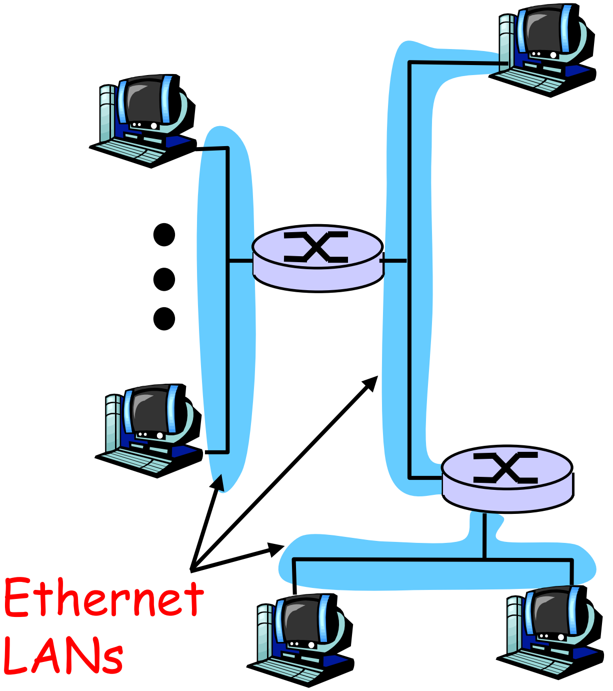
Physical Medium Dependent (PMD) sublayer

- **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

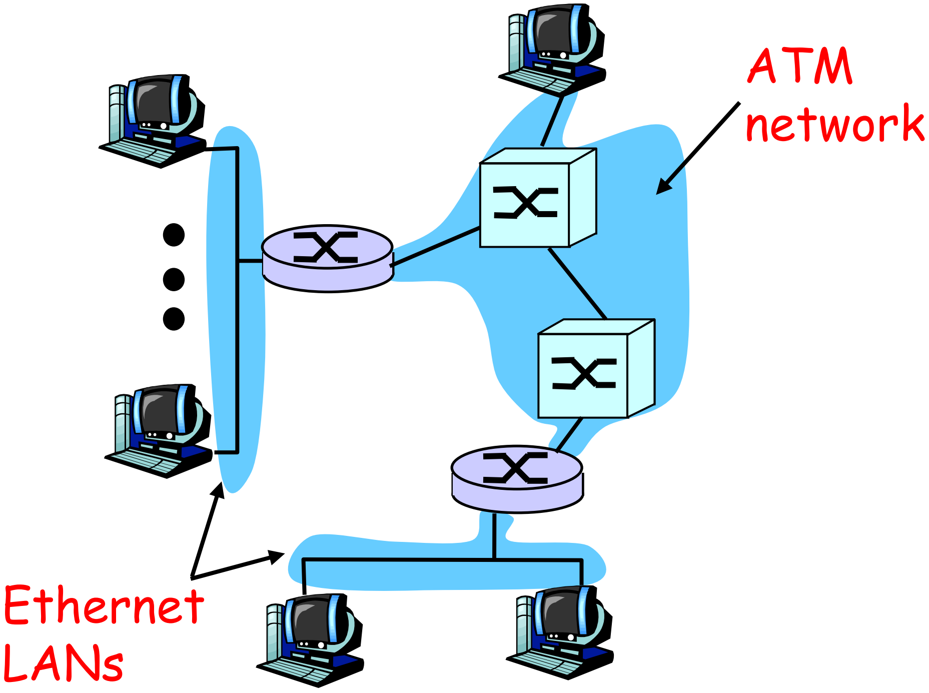
Classic IP only

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



IP over ATM

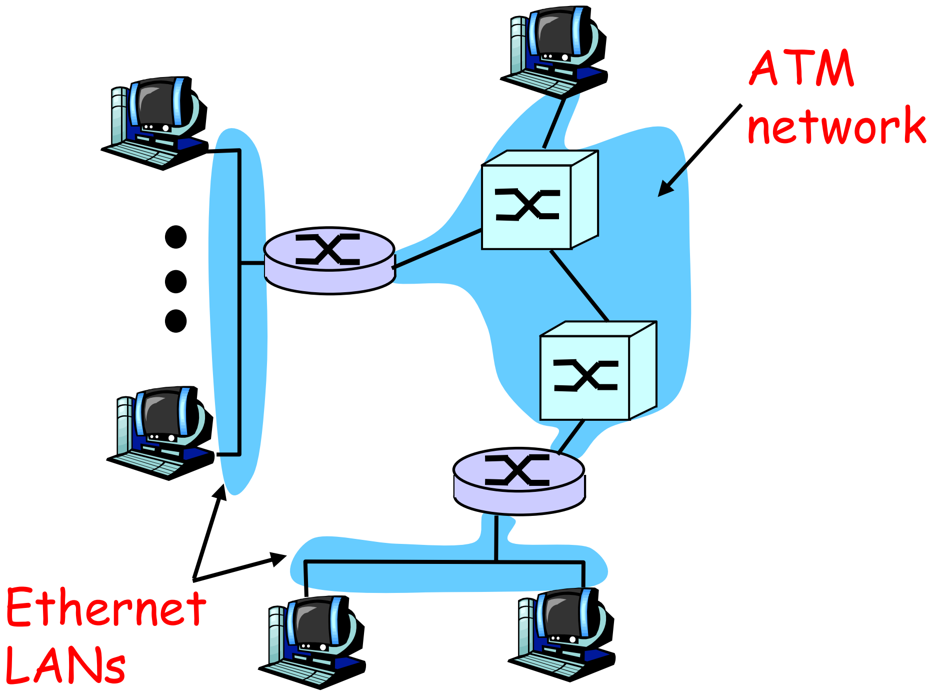
- replace “network” (e.g., LAN segment) with ATM network
- ATM addresses, IP addresses



IP-Over-ATM

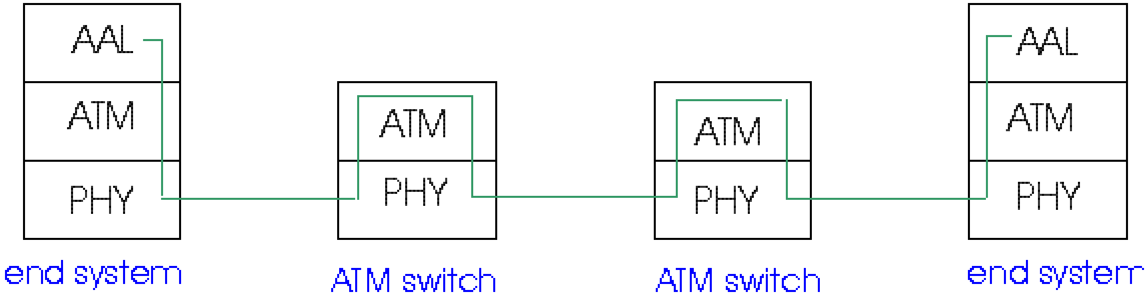
Issues:

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



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



Chapter 5 outline

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

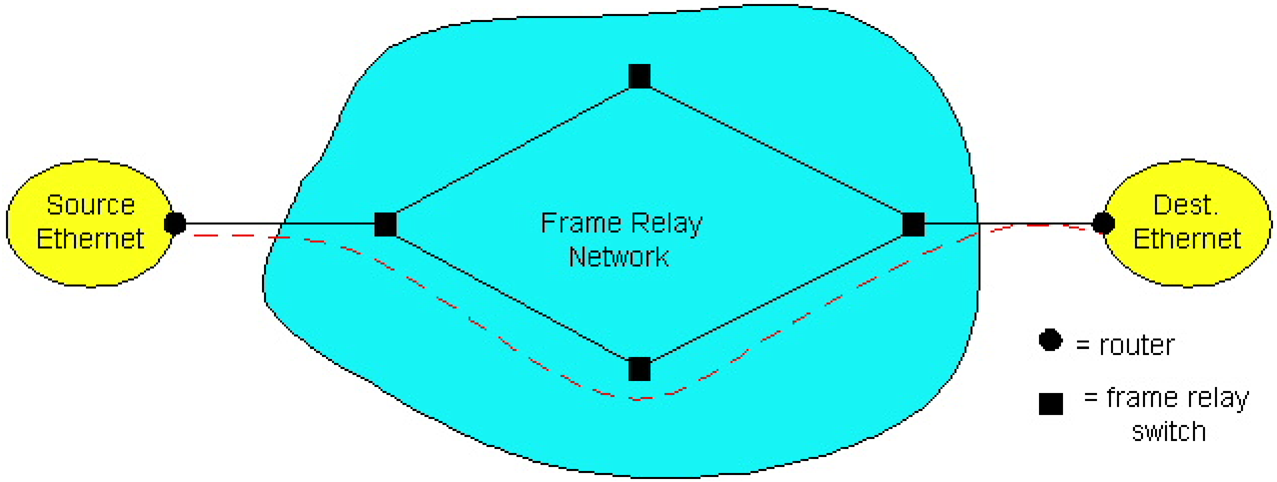
Frame Relay

Like ATM:

- wide area network technologies
- Virtual-circuit oriented
- origins in telephony world
- can be used to carry IP datagrams
  - can thus be viewed as link layers by IP protocol

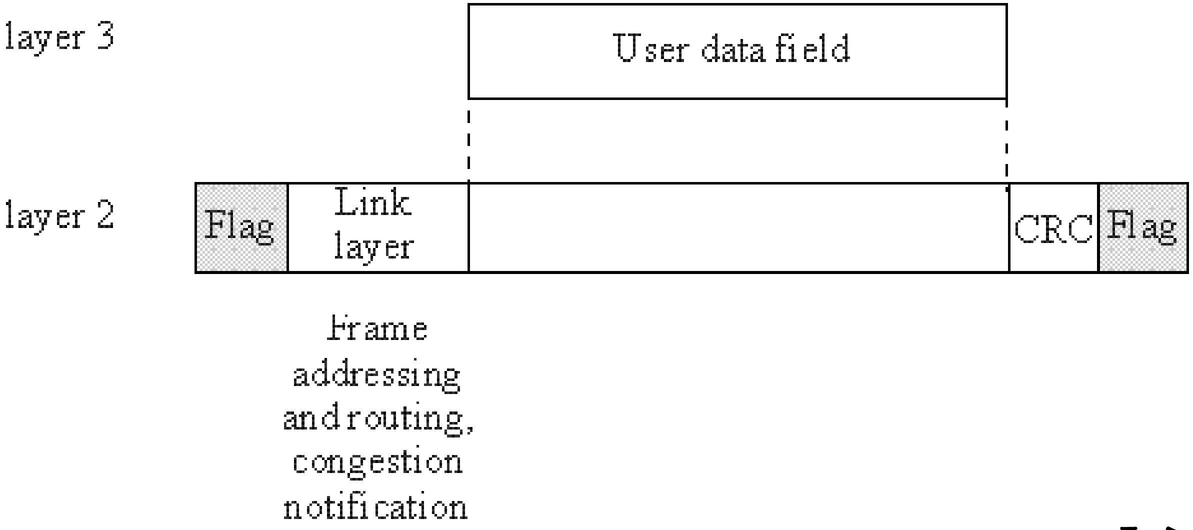
Frame Relay

- Designed in late '80s, widely deployed in the '90s
- Frame relay service:
  - no error control
  - end-to-end congestion control

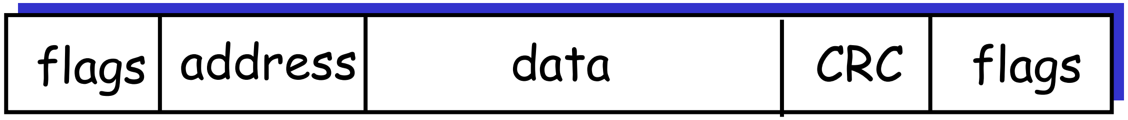


Frame Relay (more)

- Designed to **interconnect** corporate customer LANs
  - typically permanent VC's: "**pipe**" carrying aggregate traffic between two routers
  - switched VC's: as in ATM
- corporate customer **leases** FR service from public Frame Relay network (e.g., Sprint, ATT)



Frame Relay (more)



- Flag bits, 01111110, delimit frame
- address:
  - 10 bit VC ID field
  - 3 congestion control bits
    - FECN: forward explicit congestion notification (frame experienced congestion on path)
    - BECN: congestion on reverse path
    - DE: discard eligibility

Frame Relay -VC Rate Control

- Committed Information Rate (CIR)
  - defined, "guaranteed" for each VC
  - negotiated at VC set up time
  - customer pays based on CIR

- DE bit: Discard Eligibility bit
  - Edge FR switch measures traffic rate for each VC; marks DE bit
  - DE = 0: high priority, rate compliant frame; deliver at "all costs"
  - DE = 1: low priority, eligible for congestion discard

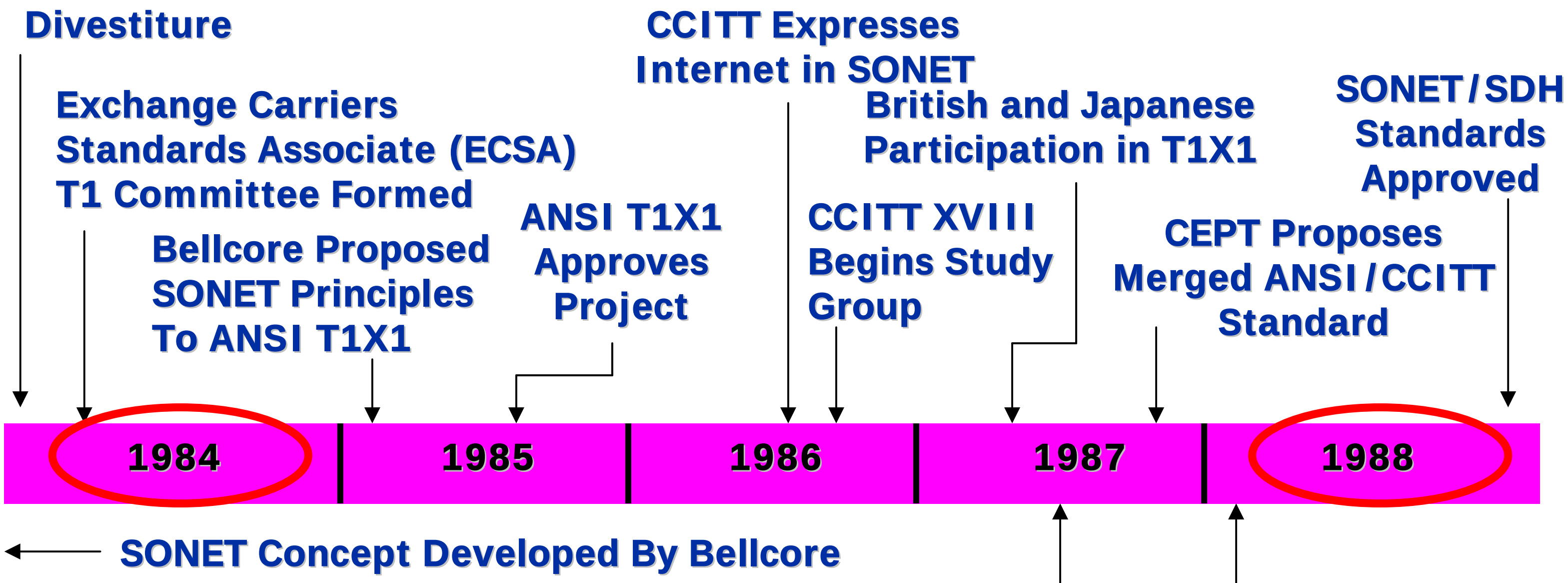
Frame Relay - CIR & Frame Marking

- **Access Rate**: rate **R** of the access link between **source router** (customer) and **edge FR switch** (provider); 64Kbps < **R** < 1,544Kbps
- Typically, **many VCs** (one per destination router) multiplexed on the same access trunk; each VC has own **CIR**
- Edge FR switch **measures** traffic rate for each VC; it **marks** (i.e. DE = 1) frames which **exceed** CIR (these may be later dropped)
- Internet's more recent differentiated service uses similar ideas

Chapter 5: Summary

- principles behind data link layer services:
  - error detection, correction
  - sharing a broadcast channel: multiple access
  - link layer addressing, ARP
- link layer technologies: Ethernet, hubs, bridges, switches,IEEE 802.11 LANs, PPP, ATM, Frame Relay
- journey down the protocol stack now *OVER!*
  - next stops: multimedia, security, network management

The LONG Standards Process



Divestiture

Exchange Carriers Standards Associate (ECSA) T1 Committee Formed

Bellcore Proposed SONET Principles To ANSI T1X1

ANSI T1X1 Approves Project

CCITT Expresses Internet in SONET

CCITT XVIII Begins Study Group

British and Japanese Participation in T1X1

CEPT Proposes Merged ANSI/CCITT Standard

SONET/SDH Standards Approved

1984

1985

1986

1987

1988

SONET Concept Developed By Bellcore

US T1X1 Accepts Modifications

ANSI Approves SYNTRAN

Standard That Almost Wasn't

- >400 Technical Proposals
- Rate Discussions AT&T vs. Bellcore
- International Changes For Byte/Bit Interleaving, Frames, Data Rates
- Phase I, II, III Separate APS, etc.

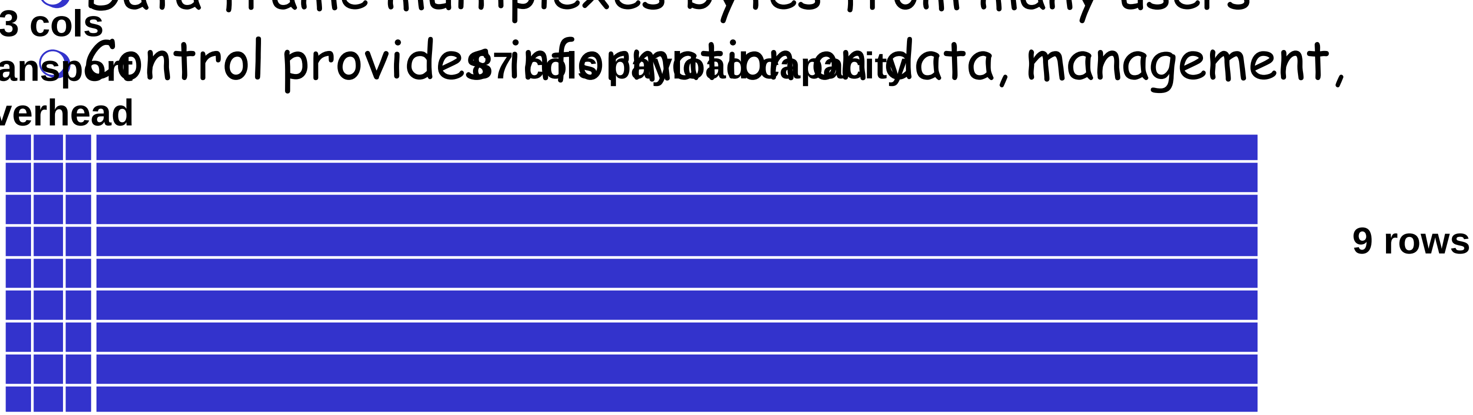
SONET Defined

- Synchronous Optical Network
- Set of Layer 1 Standards For Communication over Fiber Optic (and Electrical) Links In Order To Facilitate:

	Benefits and/or Direct Results
Transport Standard	Allow easier carrier interconnects
Survivability and Networking Flexibility	Rings and protected linear add/drops
Performance Monitoring and Alarming	Enhanced service degradation and trouble isolation
Remote Operations, Administration, Maintenance and Provisioning (OAM&P)	Minimize truck rolls or out-of-band data communication network (in-band DCC)
Timing Synchronization	Minimize network “slips”
Transport Scalability	OC-3/12/48/192 vs. async DS3 FOT
Transport of present and future services	From DS1, DS3, OC-N to GigE and 10GE

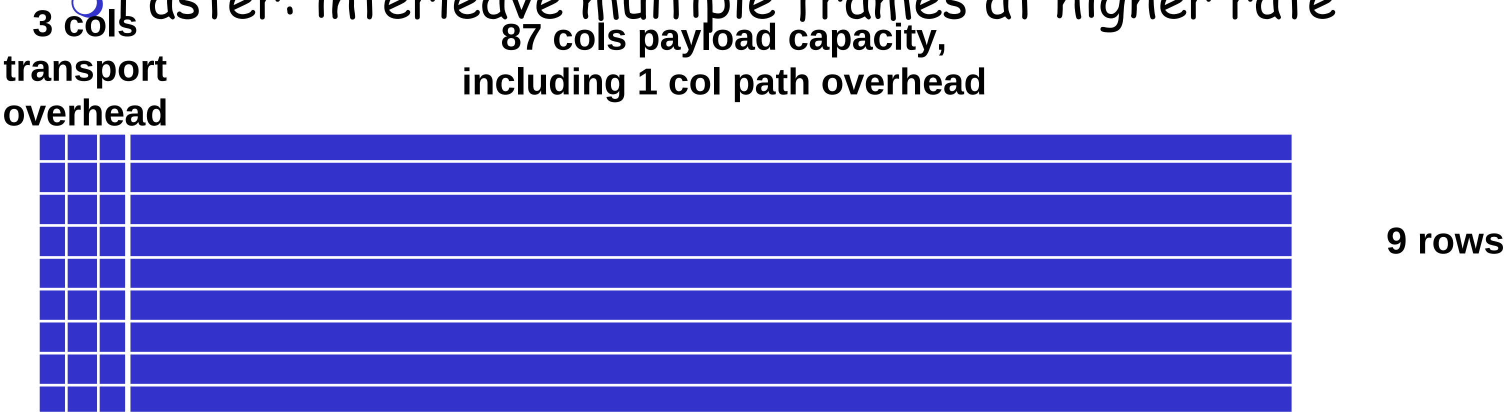
Synchronous Data Transfer

- Sender and receiver are always synchronized.
  - Frame boundaries are recognized based on the clock
  - No need to continuously look for special bit sequences
- SONET frames contain room for control and data.
  - Data frame multiplexes bytes from many users
  - Control provides information on data, management,



SONET Framing

- Base channel is STS-1 (Synchronous Transport System).
  - Takes 125 μsec and corresponds to 51.84 Mbps
  - 1 byte corresponds to a 64 Kbs channel (PCM voice)
  - Also called OC-1 = optical carrier
- Standard ways of supporting slower and faster channels.
  - Slower: select a set of bytes in each frame
  - Faster: interleave multiple frames at higher rate

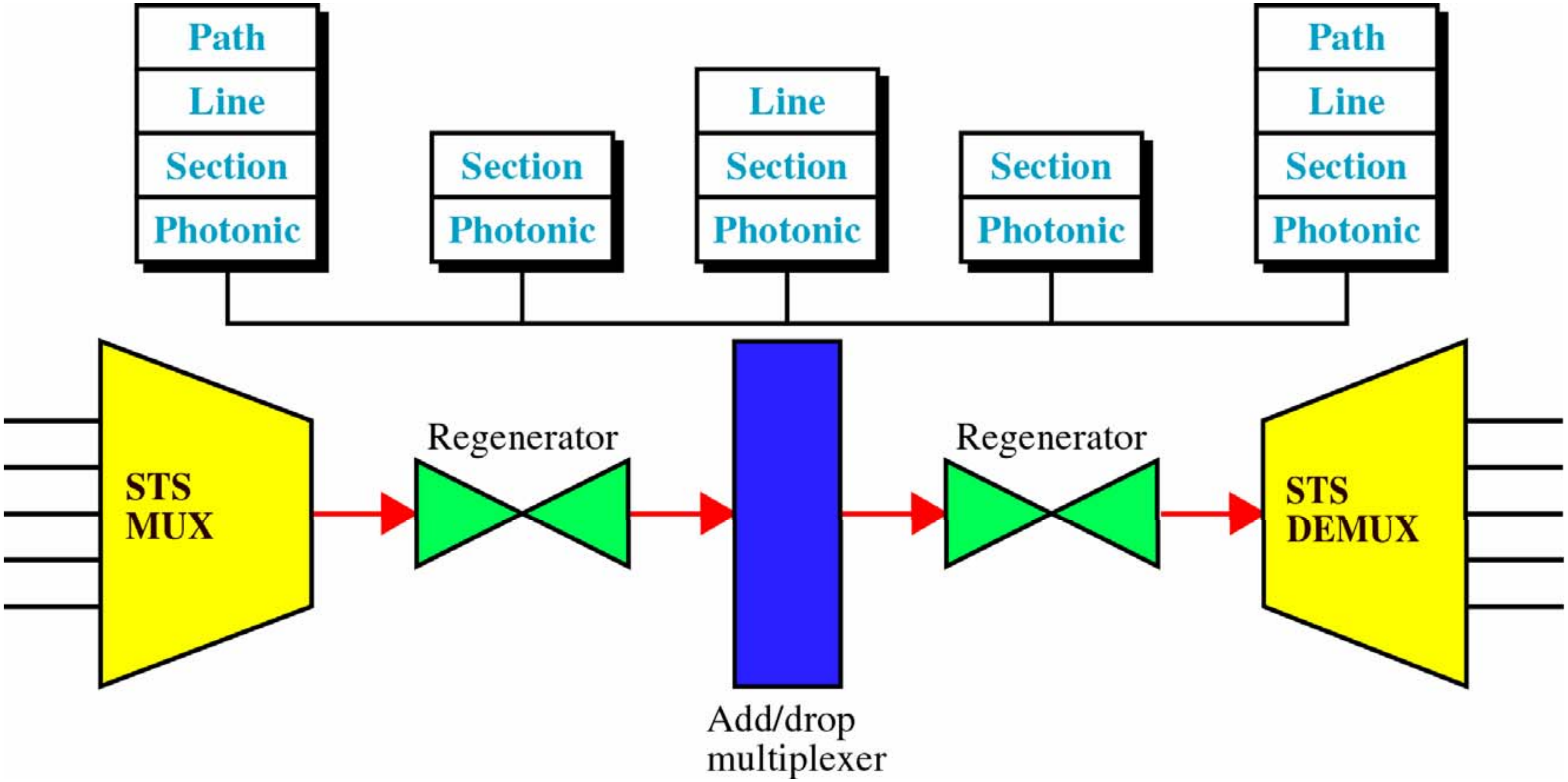

3 cols transport overhead
87 cols payload capacity, including 1 col path overhead
9 rows

Know Your Signal Line Rates

Signal Type	Line Rate	Asynchronous Payload Carrying Capacity		
		# of DS0	# of DS1	# of DS3
DS0 (POTS eq.)	64,000 bps	-	-	-
DS1	1.544 Mbps	24	-	-
DS3	44.736 Mbps	672	28	-
EC-1 (STS-1E)	51.84 Mbps	672	28	-
OC-3	155 Mbps	2,016	84	3
OC-12	622 Mbps	8,064	336	12
OC-48	2.49 Gbps	32,256	1,344	48
OC-192	9.95 Gbps	129,024	5,376	192
OC-768	39.8 Gbps	516,096	21,504	768

Figure 20-4

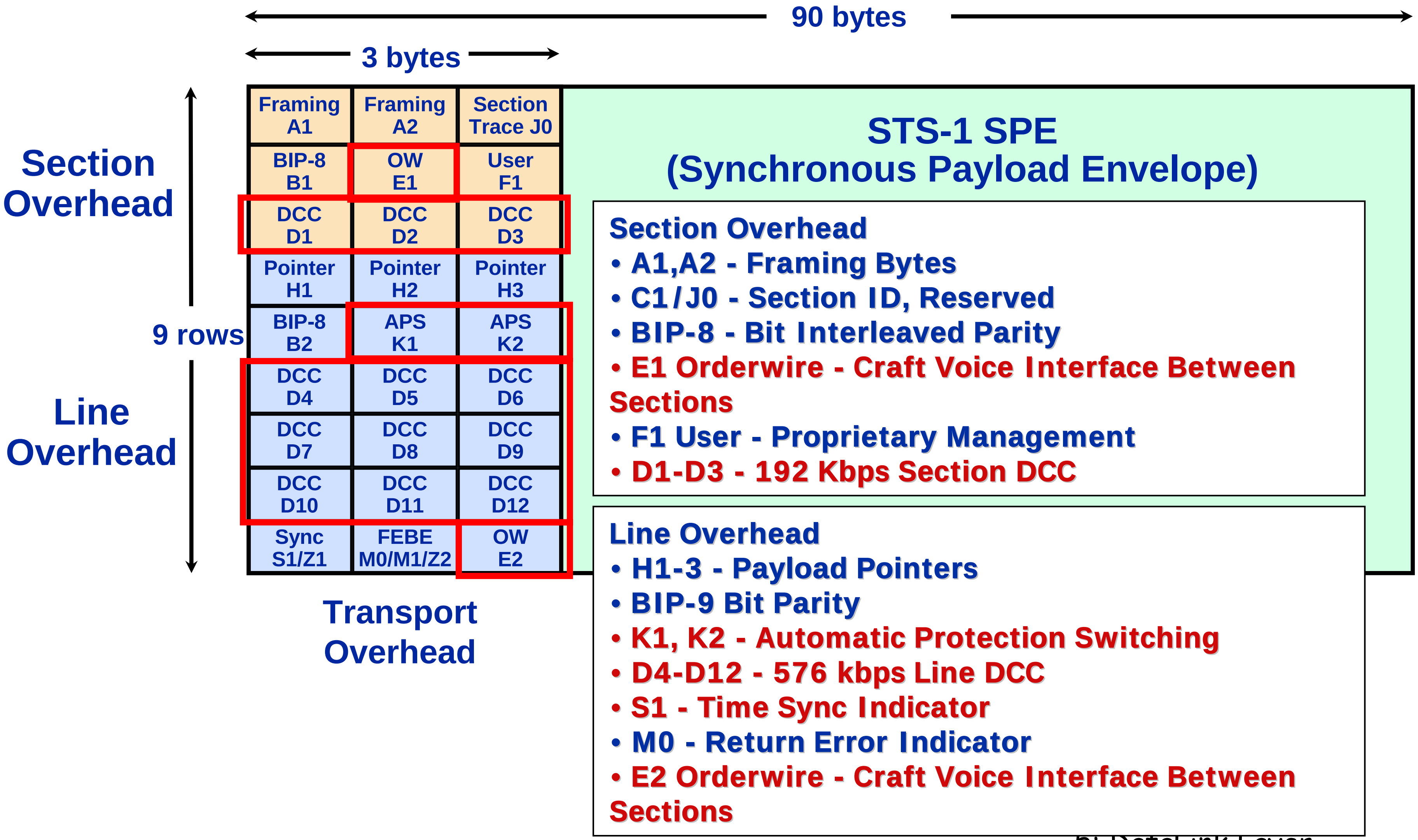
SONET Device Layers



- Section Termination (STE) - Span between regens
- Line Termination (LTE) - Span(s) between muxes
- Path Termination (PTE) - SONET path ends

Transport Overhead

Section and Line



Framing A1	Framing A2	Section Trace J0
BIP-8 B1	OW E1	User F1
DCC D1	DCC D2	DCC D3
Pointer H1	Pointer H2	Pointer H3
BIP-8 B2	APS K1	APS K2
DCC D4	DCC D5	DCC D6
DCC D7	DCC D8	DCC D9
DCC D10	DCC D11	DCC D12
Sync S1/Z1	FEBE M0/M1/Z2	OW E2

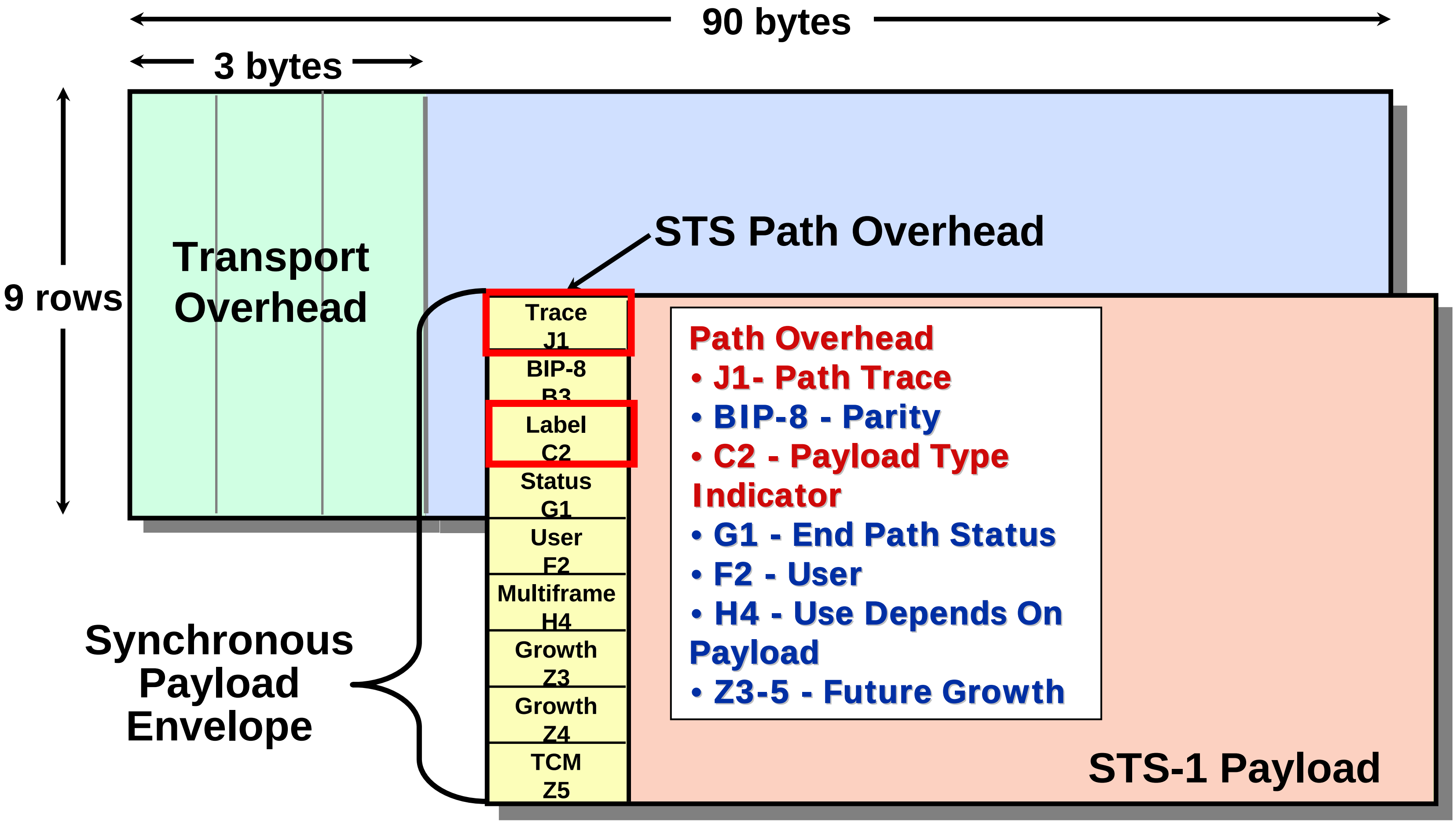
Section Overhead

- A1,A2 - Framing Bytes
- C1/J0 - Section ID, Reserved
- BIP-8 - Bit Interleaved Parity
- E1 Orderwire - Craft Voice Interface Between Sections
- F1 User - Proprietary Management
- D1-D3 - 192 Kbps Section DCC

Line Overhead

- H1-3 - Payload Pointers
- BIP-9 Bit Parity
- K1, K2 - Automatic Protection Switching
- D4-D12 - 576 kbps Line DCC
- S1 - Time Sync Indicator
- M0 - Return Error Indicator
- E2 Orderwire - Craft Voice Interface Between Sections

Path Overhead

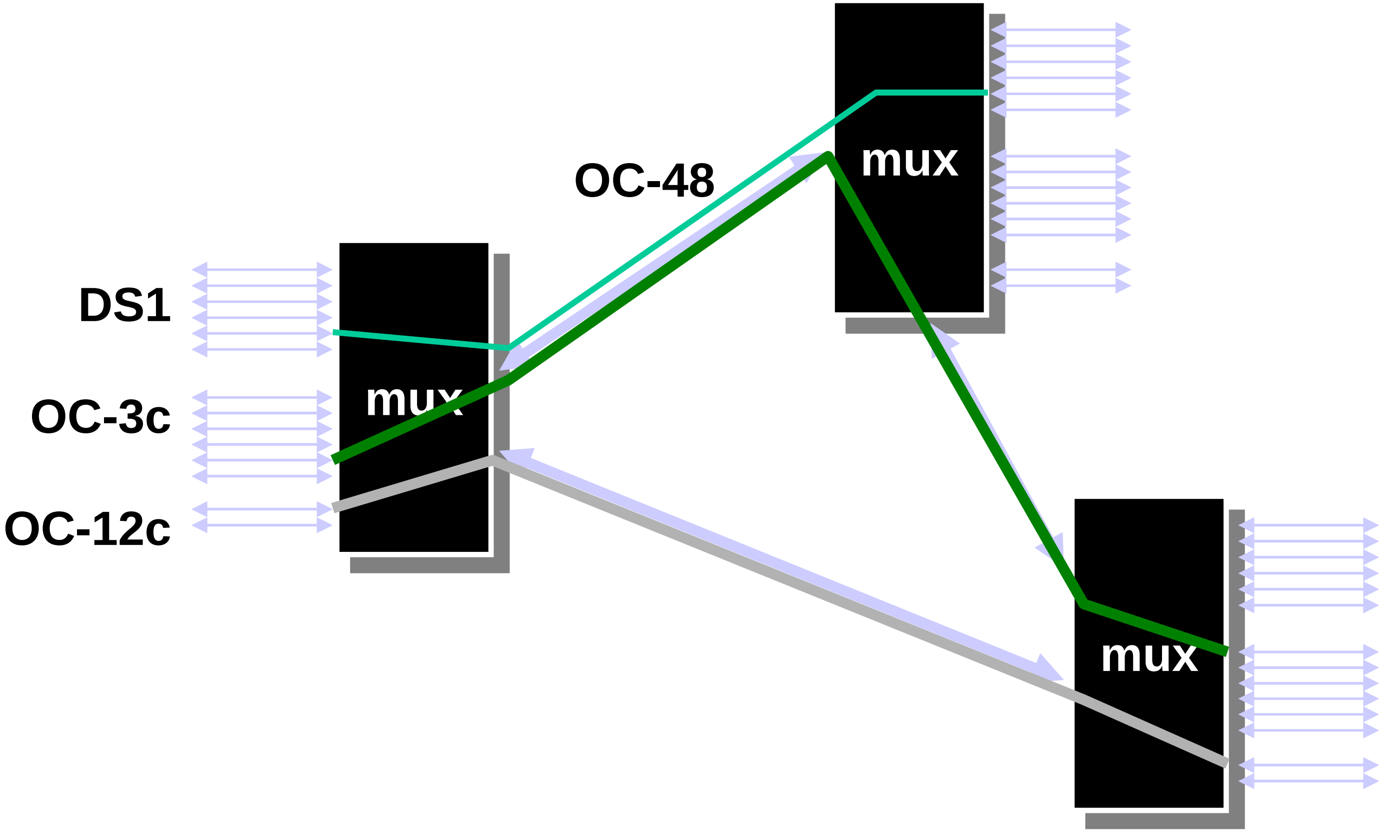


Path Overhead

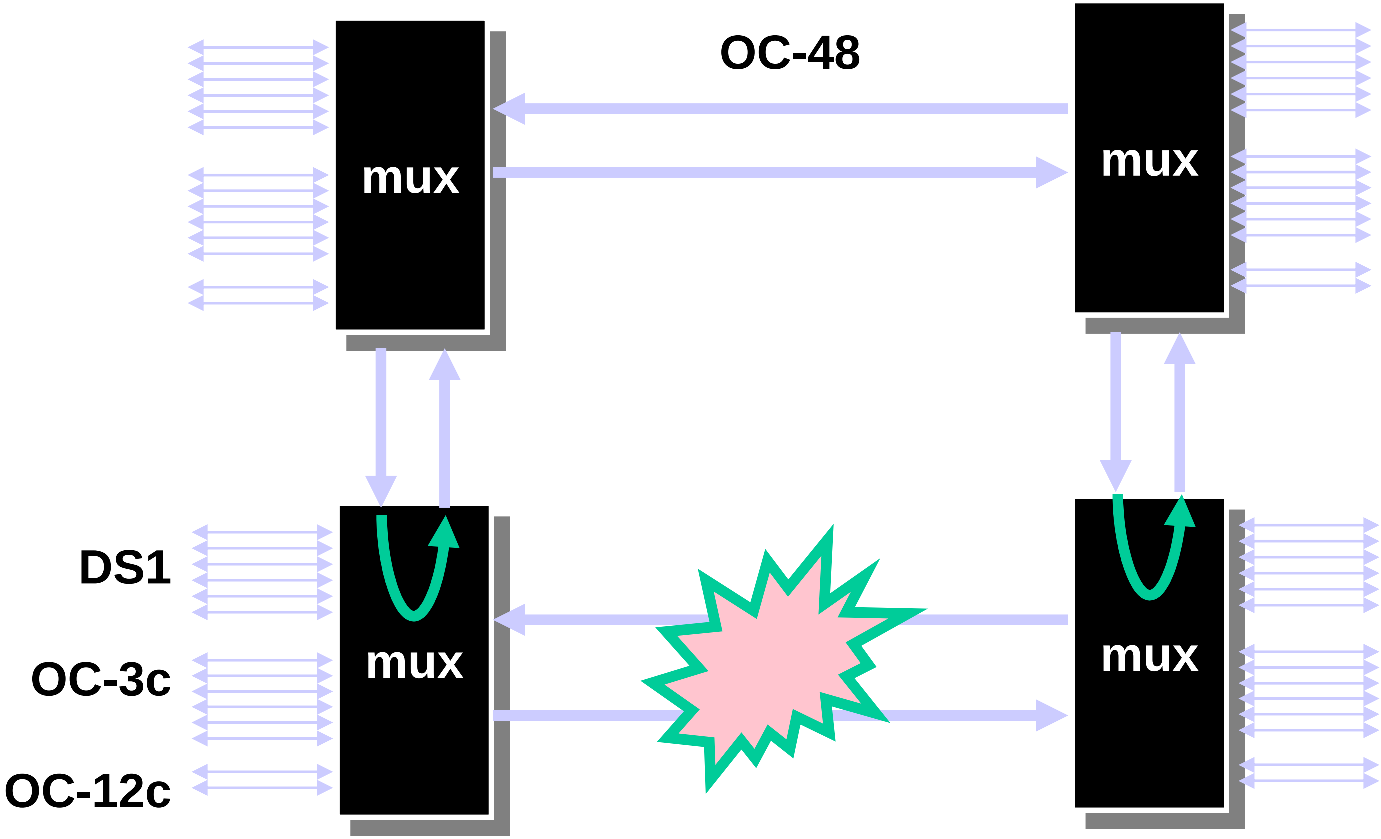
- J1- Path Trace
- BIP-8 - Parity
- C2 - Payload Type Indicator
- G1 - End Path Status
- F2 - User
- H4 - Use Depends On Payload
- Z3-5 - Future Growth

Using SONET in Networks

Add-drop capability allows soft configuration of networks, usually managed manually.



Self-Healing SONET Rings


OC-48
mux
mux
DS1
OC-3c
OC-12c
mux
mux