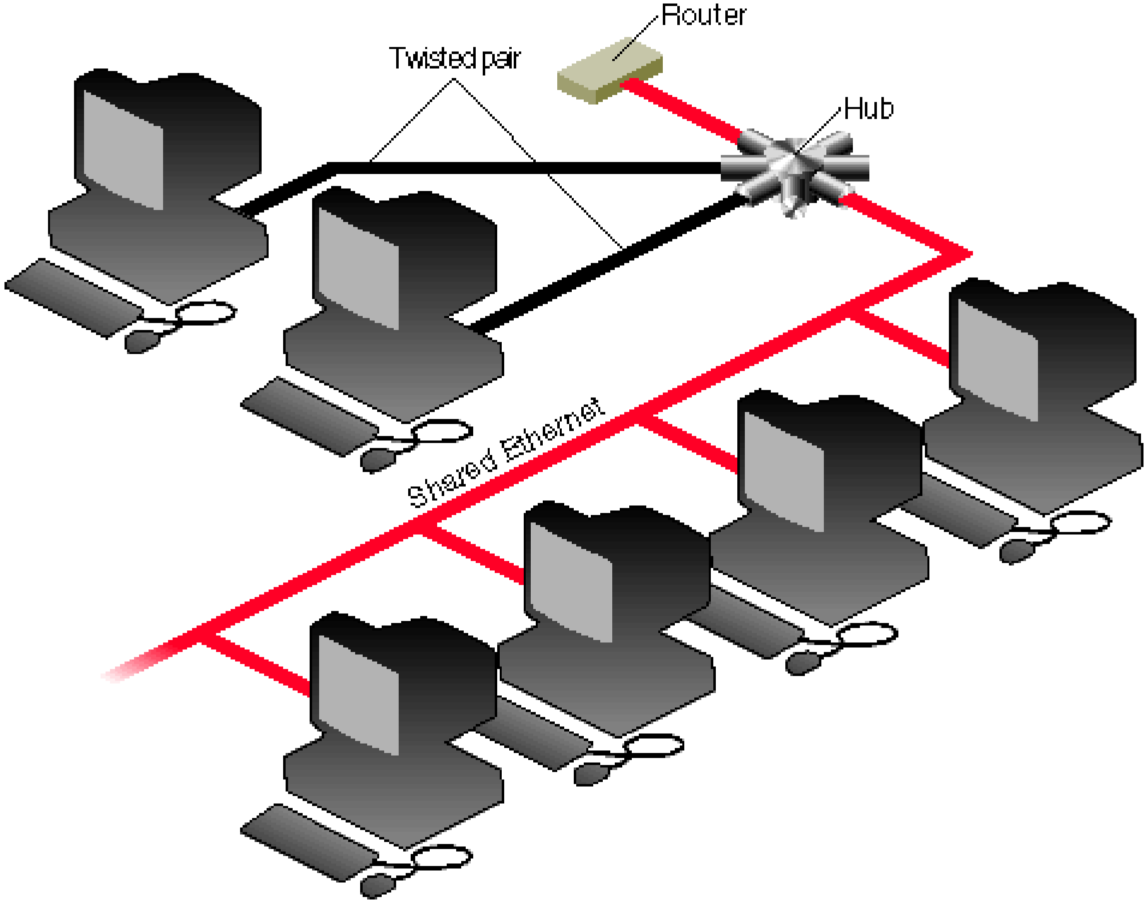
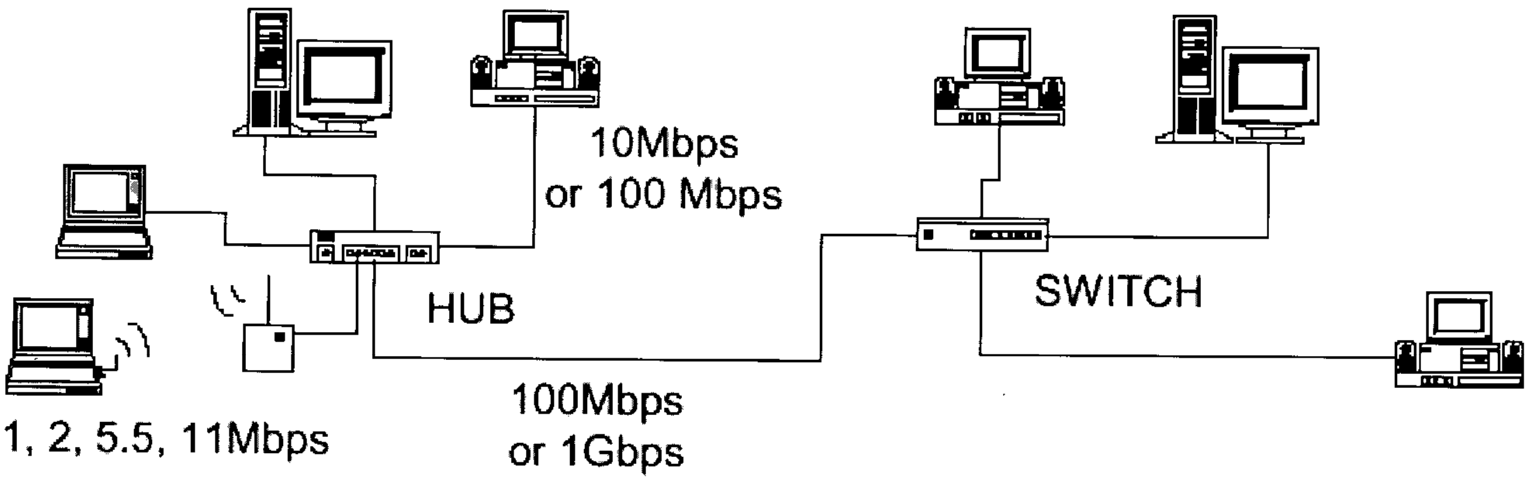
Ethernet

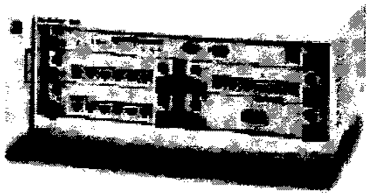


Contents

- Typical Setup
- Physical Layer
- Names
- 10Base5
- CSMA/CD
- Frame
- Fast Ethernet; Gigabit Ethernet
- Perspective

Ethernet - Typical Setup



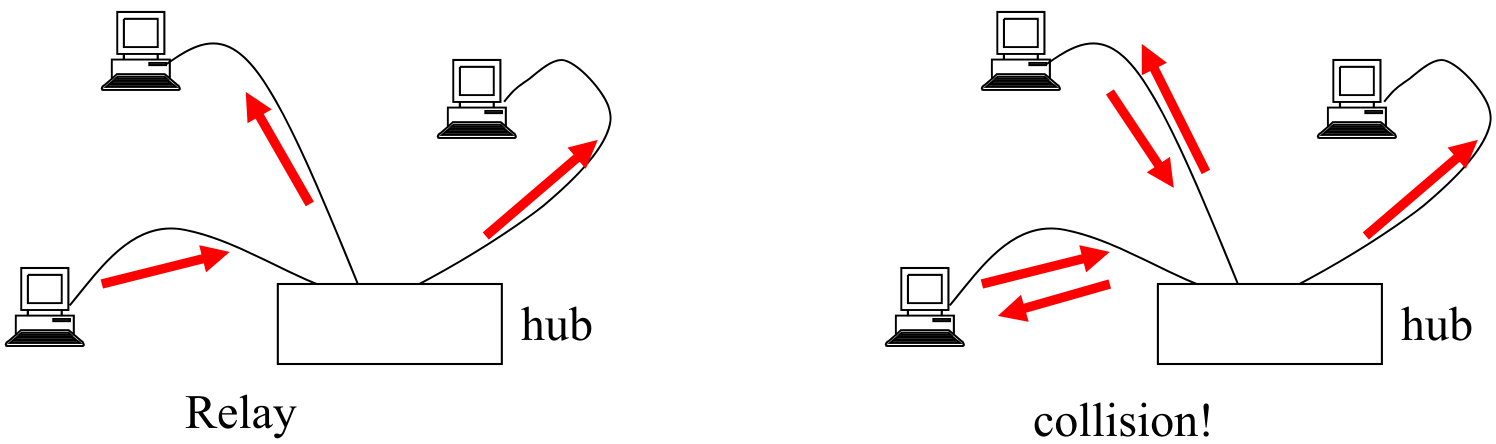


Ethernet Physical Layer

- UTP
  - Unshielded twisted pair
  - Up to 110m
- Fiber
  - 100Mbps: 2000m
  - Gbps: 220m, 500m, 5000m
- Wireless
  - 2.4Ghz DSSS: 1Mbps, 2Mbps, 5.5Mbps, and 11Mbps
  - 25m – 200m

Ethernet Hub - Physical Layer

Hub: Single Collision Domain

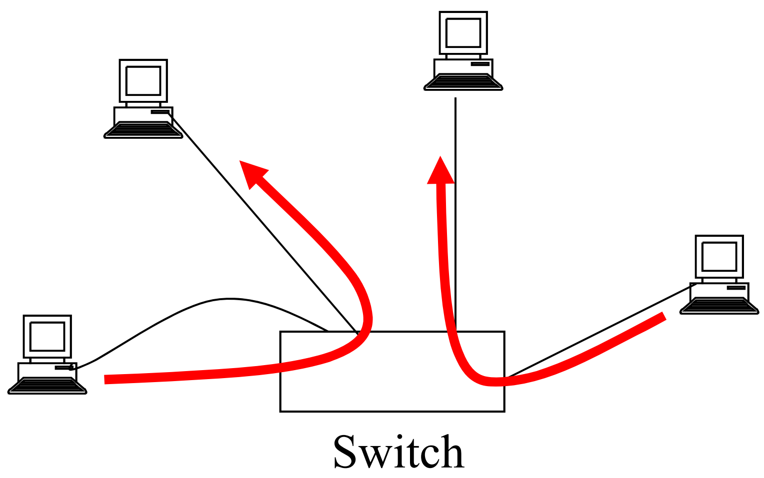


MAC Protocol: CSMA/CD

- Wait until silent (carrier sense)
- Transmit
- If collision, wait random time and repeat

Switched Ethernet - Physical Layer

Switch: No Collisions



Multiple transmissions are possible without collision

Switch stores packets that wait for same output

Ethernet

- Standardizes as IEEE 802.3
- 1-persistent CSMA/CD with exponential backoff on wired LAN
  - Once the channel is sensed idle by a station, tx takes place with probability 1
- "Classical" Ethernet is 10Mb/s over 50-ohm Coax wiring
- Newer standards cover UTP wiring, 100Mb/s operation, etc

Names for Ethernet

Names of form

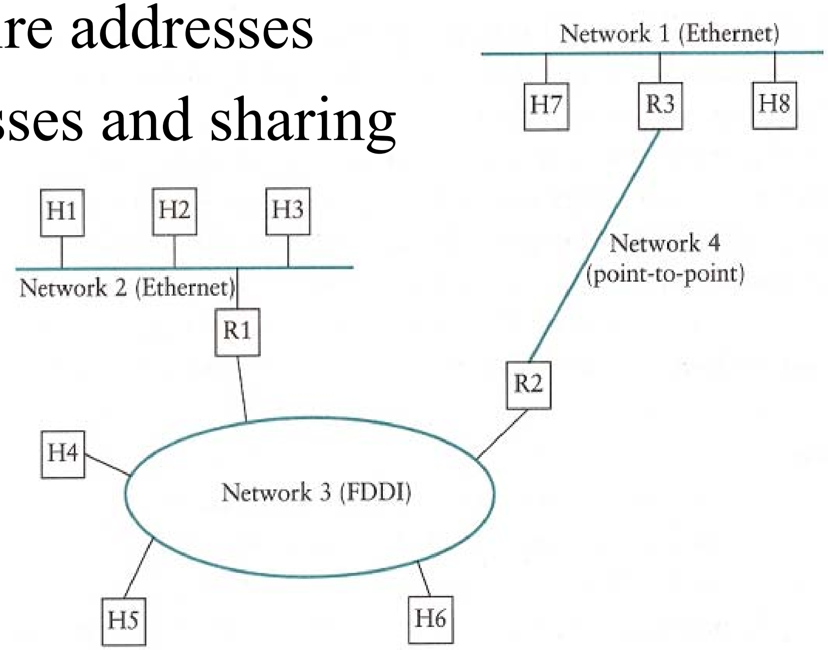
[rate][modulation][media or distance]

Examples:

- 10base2 (10Mb/s, baseband, thin coax, 200m): Thinnet
- 10base5 (10Mb/s, baseband, thick coax, 500m): Thicknet
- 10base-T (10Mb/s, baseband, twisted pair): star from hub
- 100base-TX (100Mb/s, baseband, 2 pair): star from hub/switch
- 100base-FX (100Mb/s, baseband, 2 fibers): star from hub/switch

Link Layer Networks

- Achieve station-to-station connectivity
- May be point-to-point or multi-access
  - Point-to-point may not require addresses
  - Multi-access requires addresses and sharing
- Error detection
- Addressing



Sharing

- How to share a broadcast media (e.g. wire, air)?
  - What to do with two simultaneous speakers
  - Approaches: centralized and distributed
- Centralized (polling)
- Decentralized (speaking when media idle)

Comparing Approaches

- Centralized Approach
  - Polling requires speaker to await moderator, even if others idle
  - Problems if moderator's connection fails
- Decentralized Approach
  - No moderator wait, but subject to collisions
  - Collisions could continue forever
- The **Multiple-Access** Problem

CSMA Type Networks

- CSMA – Carrier Sense Multiple Access
  - Detect when medium is idle/busy
  - Persistent: Once the channel is sensed idle by a station, tx takes place with probability 1
  - Non-persistent (send some time soon)
- Approach to collisions
  - *p*-Persistence: tx takes place with probability p or tx is deferred one unit of time with probability 1-p: Wireless LAN
  - **Detection and backoff: Ethernet**

p-Persistent CSMA

p = Prob(send|idle)

E(# stations tx after idle) = *np* [n: # total stations ready to send]

If *np* > 1, likely secondary collision, so want *p* < 1/*n*

n increases with system load, so want smaller *p* with high load

Smaller *p* affects message delay

Ethernet is 1-persistent

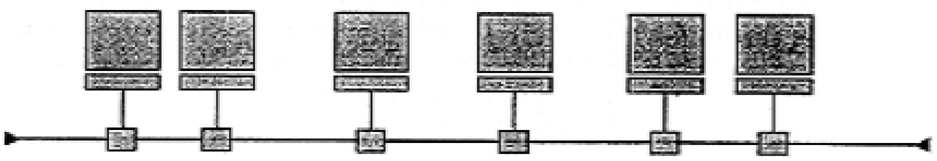
Ethernet Properties - 10base5 (Thicknet)

"Classical" Ethernet

Single segment up to 500m; with up to 4 repeaters gives 2500m max length

Baseband signals broadcast, Manchester encoding, 32-bit CRC for error detection

Max 1024 stations/Ethernet



Ethernet Bus

Collision Detection and Backoff

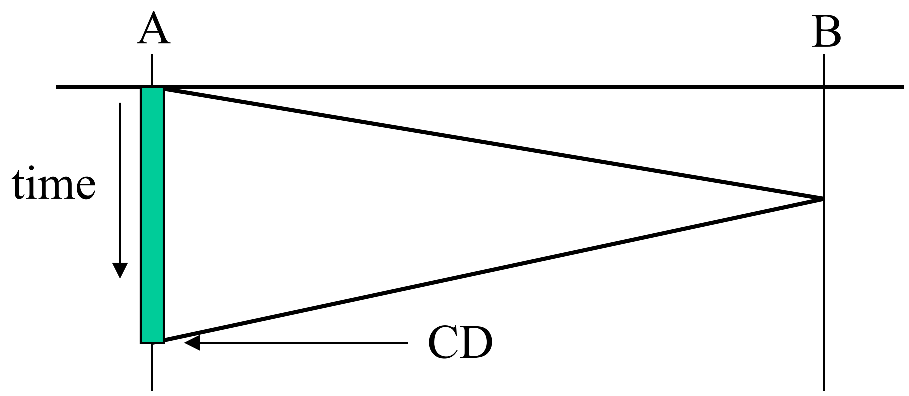
- Determine if frame transmitted successfully, if not, wait
- Collision detection
- Wait using **exponential backoff**
  - Wait random on interval [0, 2(max prop. delay)]
  - Double on each successive collision

Collision Detection

CD circuit operates by looking for voltage exceeding a transmitted voltage

Want to ensure that a station dose not complete transmission prior to CD

Time to CD takes up to 2(max prop. delay)



Minimum frame size

Speed of light is about $3x10^8$m/s in vacuum and about $2x10^8$m/s in copper

So, with 4 repeaters (2500m 10base5) max Ethernet signal prop time is about 12.5µsec, or 25µsec RTT

With repeaters, etc. 802.3 requires 51.2µsec, corresponding to 512 bit-times

Thus, minimum frame size is 512 bits (64 bytes); also called **slot time**

Slot time and minimum frame size

Collision undetected:

A frame tx time < Round-trip propagation time (slot time)

Collision detection:

tx time of minimum frame length > slot time

Max frame size

- 1500 byte limitation on maximum frame transmission size (MTU)
- Limits maximum buffers at receiver
- Allows for other stations to send
  - Also requires 96 bit Inter-Packet-Gap (IPG)

Transmitter

When ready and line idle, await IPG (96 bit times) and send while listening (CD)

If CD true, the adapter (NIC) generates max 48-bit jamming sequence and do exponential backoff

Jamming sequence used to inform all stations that a collision has occurred

Exponential Backoff

For retransmission n ($1 \le n \le 10$)

- Choose k at random on U(0, 2^n-1)
- Wait k* (51.2μsec) to retransmit
- Send on idle; repeat on another collision
- For ($11 \le n \le 15$), use U(0, 1023)
- If n = 16, drop frame

Longer wait implies lower priority (strategy is not "fair")

CSMA/CD (IEEE 802.3 specification)

1. If no carrier, then it tx its information
2. If a carrier is present, it waits until carrier clears, waits an additional time designated as the Inter-Packet Gap (or IPG: inserted after every successful tx) which is 9.6 μS long, and then tx

CSMA/CD (IEEE 802.3 specification)

3. If more than one station tx at the same time, the collision will be detected before any of the stations complete their tx

The minimum frame length of 64 bytes insures that simultaneous tx will be detected before completion, even in the worst case scenario of two stations at the limits of the segments colliding

When a collision is detected, the transmitting station ceases tx of information and tx a 48-bit jam signal to indicate to other stations on the network that a collision has occurred

CSMA/CD (IEEE 802.3 specification)

4. After collision, the transmitting station executes the backoff algorithm

The backoff algorithm basically consists of the transmitting station waiting a random interval before re-transmitting

The transmitting station selects a random number from a limited range of random numbers specified by the equation:

$$\text{Max Range} = 2^n - 1$$

where n is the number of collisions that the transmitting station has experienced

CSMA/CD (IEEE 802.3 specification)

The random number selected from the specified range is then multiplied by a constant referred to as a time slot, which is derived from the bit length of the medium (512 bits at 10 Mbps or 51.2 μS)

The number n is incremented each time a collision occurs, until its value is 10, corresponding to a possible maximum waiting interval of 51.2 μS × 1023

This range of numbers is used for the next 5 collisions, after 15 collisions the algorithm will report an error to the higher level application

Capture Effect

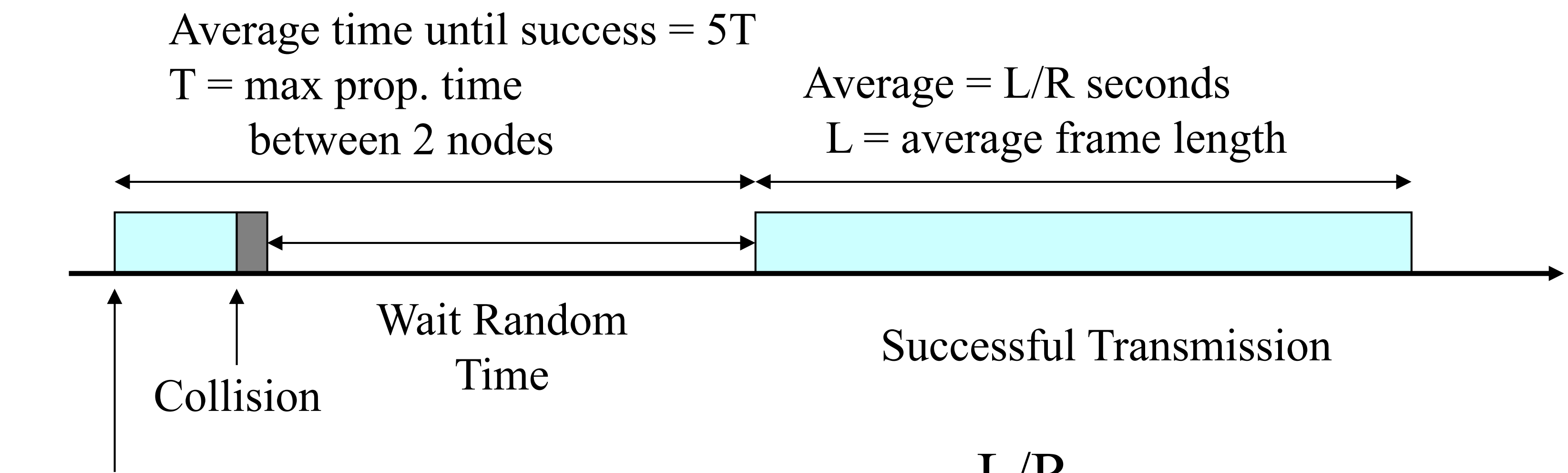
Given two stations A and B, unfair strategy can cause A to continue to “win”

Assume A and B always ready to send:

- If busy, both wait, then send and collide
- Suppose A wins, B backs off
- Next time, B’s chances of winning are halved
- To winner every collision is the 1st one

CSMA/CD

Typical Sequence of Events:



$$\text{Efficiency} = \frac{L/R}{L/R + 5T} = 1/(1 + 5a)$$

$$a = T/(L/R) = RT/L$$

Efficiency of CSMA/CD

a = RT/L
eff = 1/(1 + 5a)

a impacts what happens during simultaneous transmisstion:

a small: early collision detection, efficient

a large: late detection, inefficient

Example 1: 10Mbps, 1000m

=> T = (1km)(4μs/km) = 4μs; RT = 40 bits

L = 4000 bits

5a = 200/4000 = 0.05 => efficiency 95%

Example 2: 1Gbps, 200m

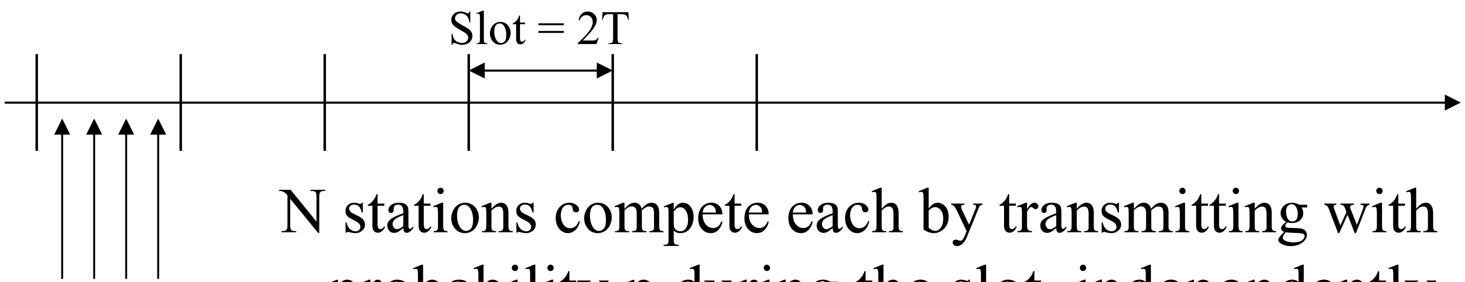
=> T = (0.2km)(4μs/km) = 0.8μs; RT = 800bits

L = 4000 bits

5a = 4000/4000 = 1 => efficiency = 50%

Efficiency of CSMA/CD – Analysis

Model:



N stations compete each by transmitting with probability p during the slot, independently

If success => transmit L bits

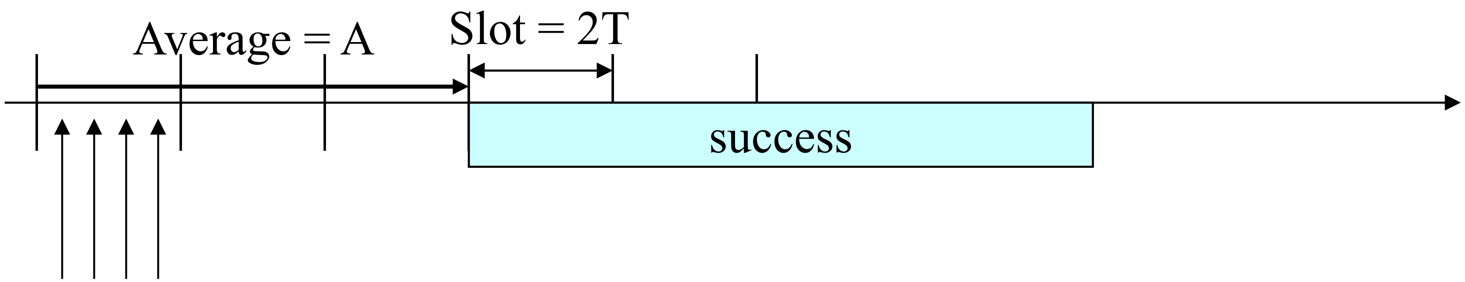
If failure (idle or collision), try next slot

P(success) = P(exactly 1 out of N transmits) = $Np(1-p)^{N-1}$

Indeed: N possibilities of station that transmits

P(one given station transmits, others do not) = $p(1-p)^{N-1}$

Efficiency of CSMA/CD – Analysis



P(success) = $Np(1-p)^{N-1}$

Maximum when p = 1/N

=> P(success) = 1/e = 0.36

Average time until success:

$$A = 0.36(0) + 0.64(2T + A)$$

=> A = 1.28T/0.36 = 3.5T

In practice, backoff not quite optimal => 5T

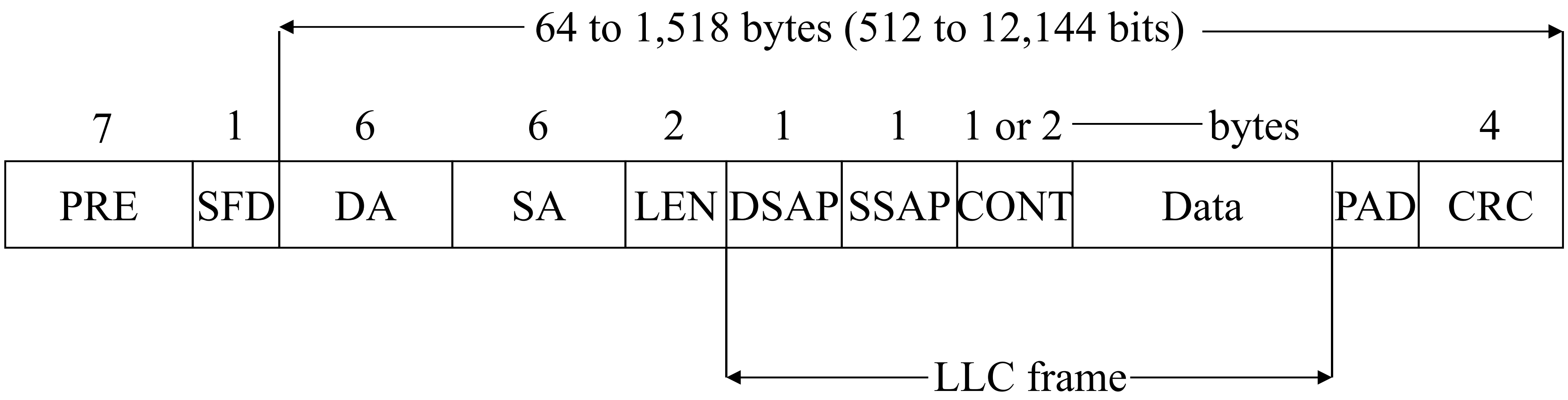
Network Interface Card (NIC)

Each node has a NIC

NIC is responsible for tx and rx data frames

Each NIC has a different address: MAC address

Ethernet Frame



Frame Structure

- 7 byte preamble: alternating 0/1 combination producing 10Mhz square wave for 5.6 μsec, used for receiver sync
- 1byte SFD (start of frame delimiter) 10101011
- 6 byte destination address, 6 byte source address, 2 byte type/length overloaded field
- Variable sized data portion followed by 4-byte CRC-32
- Sends low-order bit first for 802.3

Ethernet Frame

Length/Type field:

- Type (Ethernet V2)
  - Indicates type of data contained in payload
  - Issue: what is the length?
- Length field (IEEE 802.3)
  - Type info follows frame header
- So, is it the type or length?
  - “Ethernet”: types have values above 2048 (RFC894 for IP)
  - 802.3: length value below 05DC (RFC1042 for IP)
- IF length, next headers are LLC and SNAP (for IP)
  - LLC (3 bytes): DSAP, SSAP, CTL
  - SNAP (5 bytes): org code, type (above)

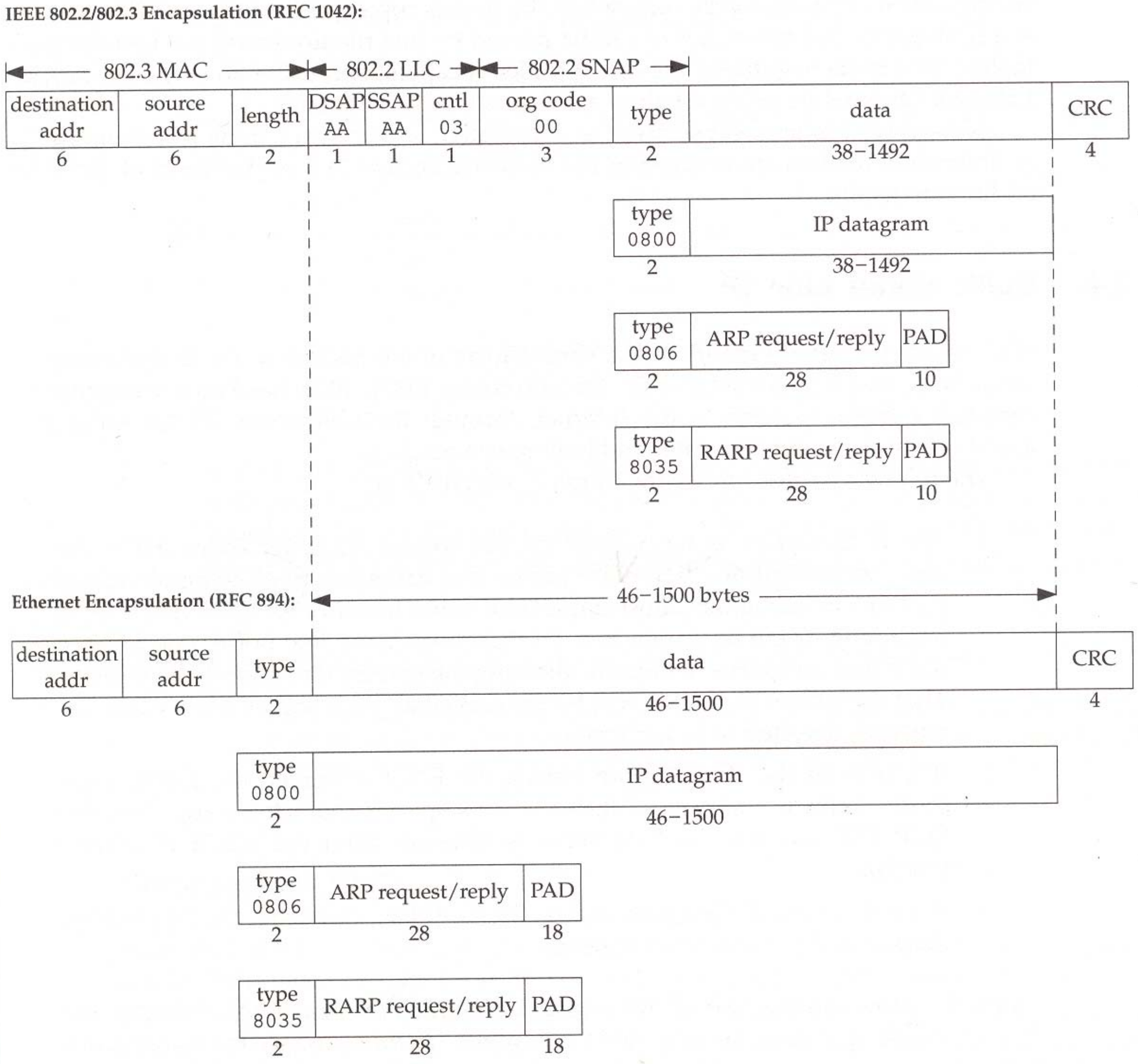
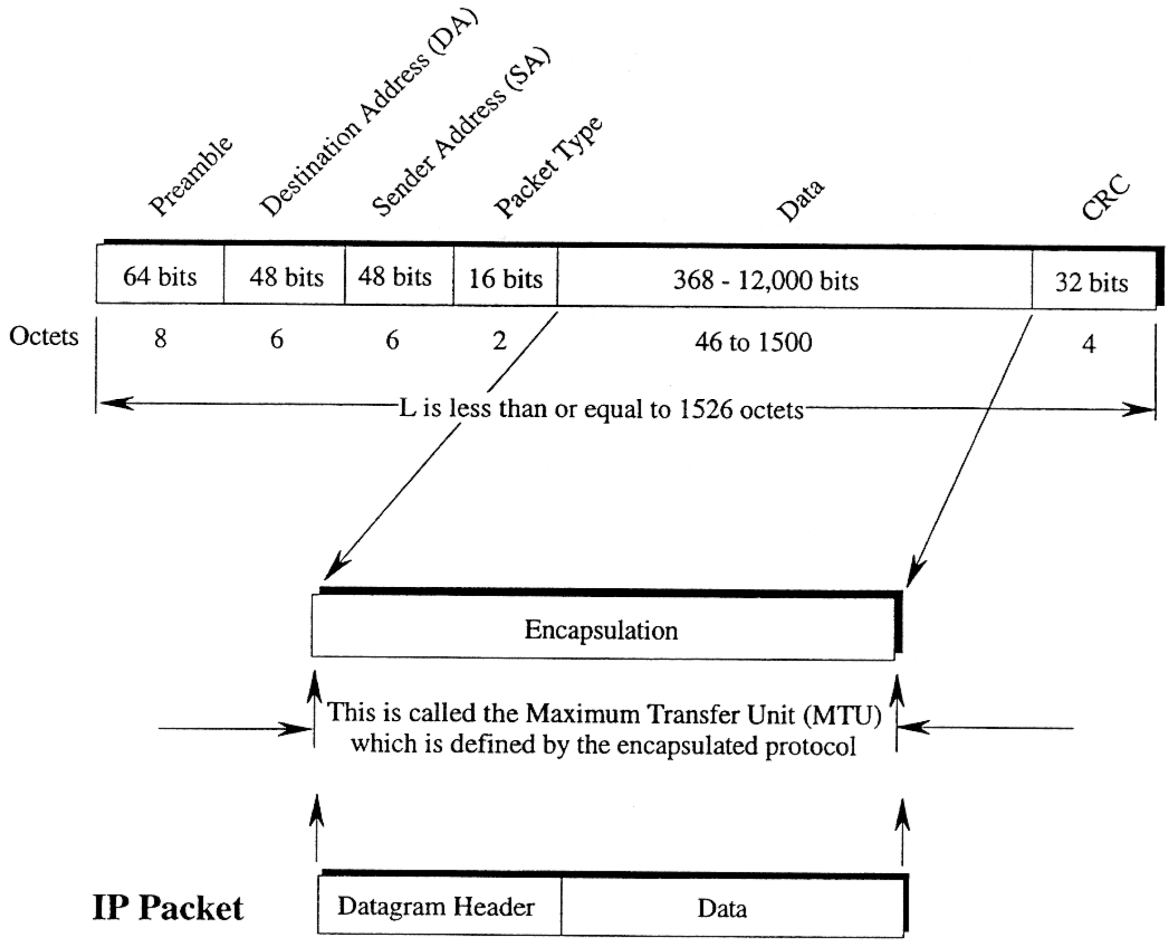


Figure 2.1 IEEE 802.2/802.3 encapsulation (RFC 1042) and Ethernet encapsulation (RFC 894).

← Most commonly used

DATA ENCAPSULATION IN AN ETHERNET FRAME



Ethernet Addressing

48 bit Ethernet/MAC/Hardware Addresses in NIC

00:60:8C:01:02:03

Prefix assigned per-vendor by IEEE

Unique per-adapter, burned into ROM

Ethernet Address 00:60:8C:01:02:03

byte 1	00000000	00	00:60:8C is the Vendor Code (this is a 3Com address)
byte 2	01100000	60	
byte 3	10001100	8C	
byte 4	00000001	01	01:02:03 is the Device Code
byte 5	00000010	02	
byte 6	00000011	03	

Multicast Addressing

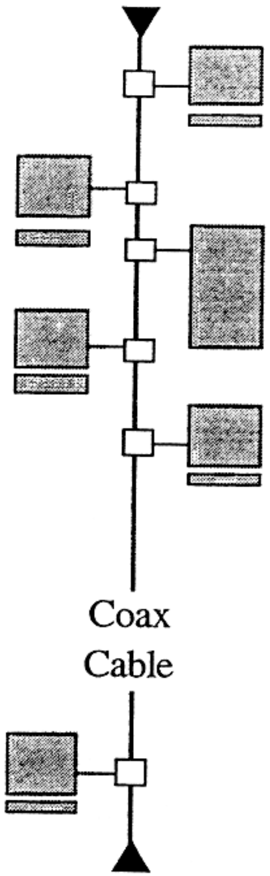
- Multicast (least significant bit of 1st byte = 1) and **Broadcast (all 1's)** addresses (used for ARP)
- Many adapters support *promiscuous* mode, i.e., delivers all received frames to the host
- Each vendor assignment supports 2^{24} individual and group (multicast) addresses
- Each adapter supports multiple group "subscriptions"
  - Usually implemented as hash table
  - Thus, software may have to filter at higher layer

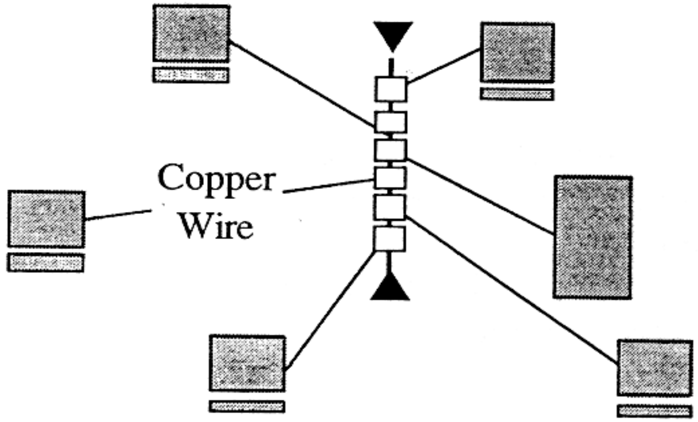
IEEE 802.3u 100 Mb/s

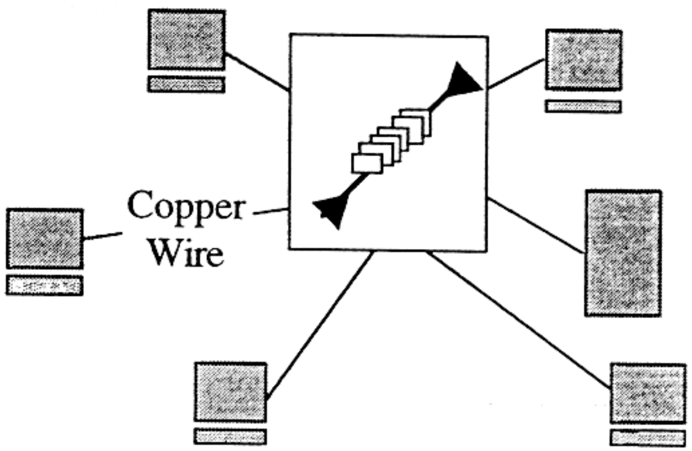
"Fast Ethernet" (1995) adds:

- 10x speed increase
- 100m max cable length retains min 64 byte frames between hub and stations
- Replace Manchester with 4B/5B
- Full-duplex operation using switches:
  - non-shared link
  - CSMA/CD disabled
- Speed and duplex auto-negotiation

Hubbing with Ethernet

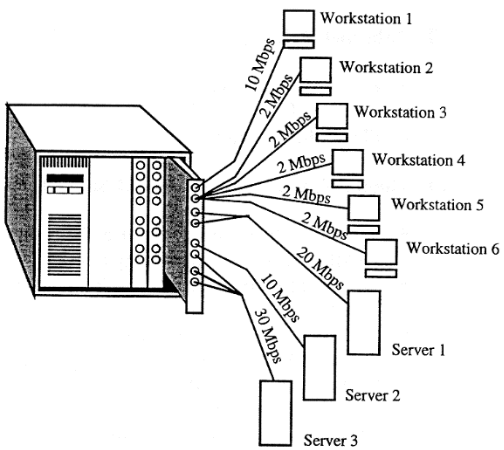


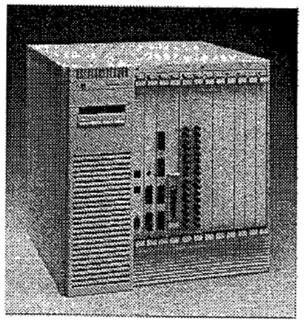




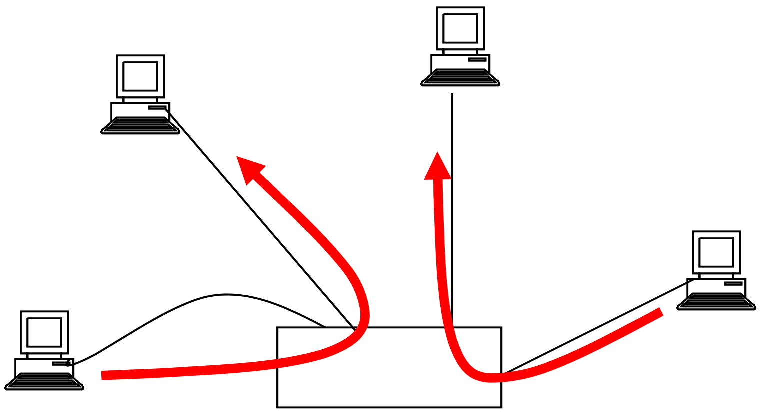
Ethernet Switch

SWITCHED ETHERNET HUBS (CON'D)





3Com Switched Ethernet Hub, The Linkbuilder 3GH



LAN Hub and Switch

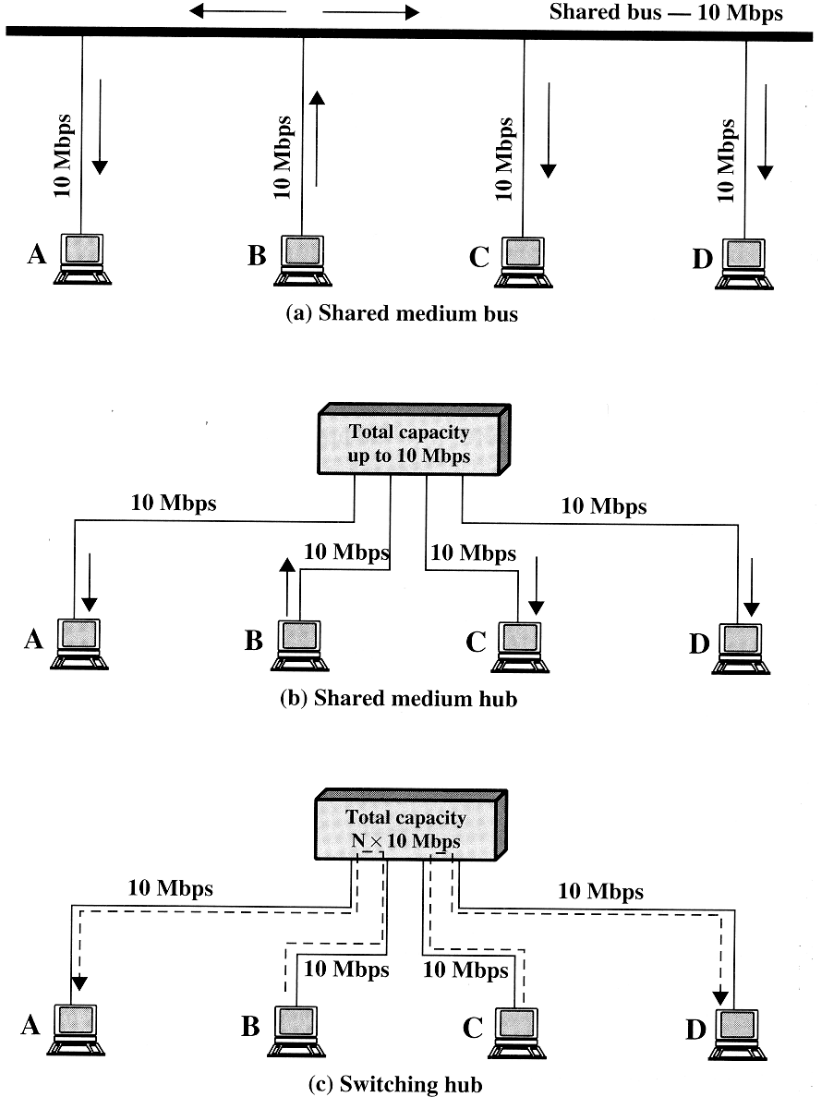


Figure 13.10 LAN Hubs and Switches

IEEE 802.3{z,ab} 1000 Mb/s

"Gigabit Ethernet" (1998, 9) adds:

- 100x speed increase
- Carrier extension (invisible padding…)
  - >=4096 bit-times, up from the minimum 512 bit-times
- Frame bursting
  - multiple short frames tx consecutively without CSMA/CD between frames

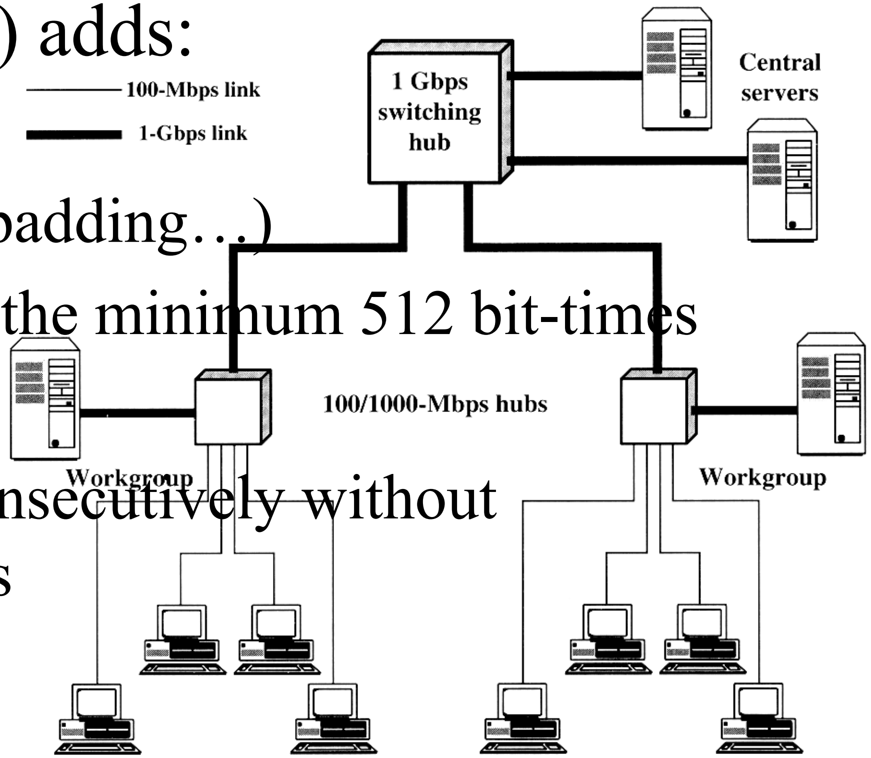
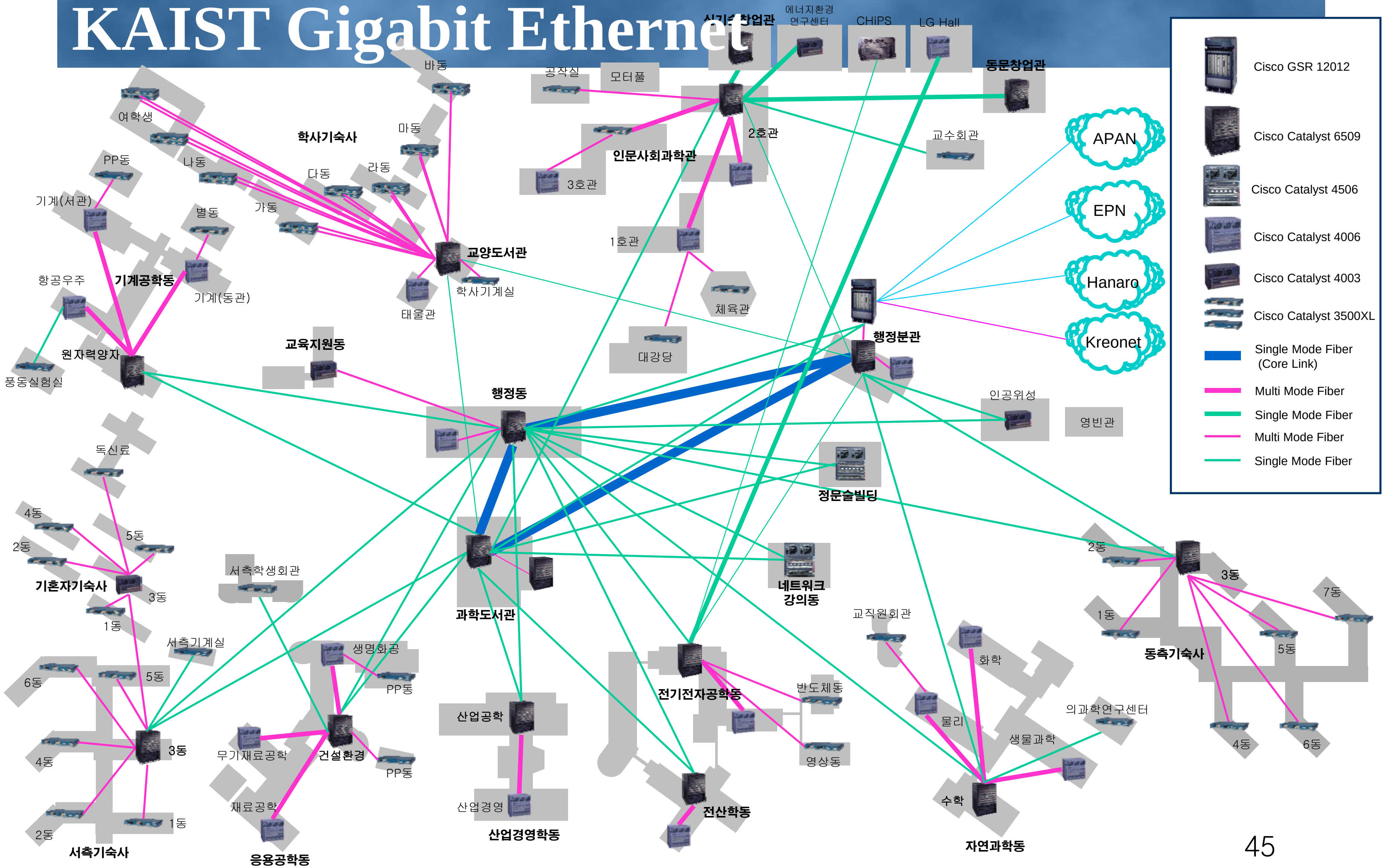


Figure 14.4 Example Gigabit Ethernet Configuration

KAIST Gigabit Ethernet



Other LAN Technologies

- Ring networks generally more complex
  - IBM 4/16 Mb/s token ring
  - FDDI
- Connection-oriented
  - ATM, HIPPI

Perspective

Ethernet is wildly successful, partly due to low cost;
Compare with FDDI or Token Ring

Some issues:

- Nondeterministic service
- No priorities
- Min frame size may be large

Summary

- Ethernet: Physical/Link layer protocol
- 1-persistent CSMA/CD with exponential backoff on wired LAN
- Minimum frame size is 512 bits (64 bytes) to detect collision
- Ethernet Frame tr/rx by NIC
- 48 bit Ethernet/MAC/Hardware Addresses in NIC
- Fast/Gigabit Ethernet using hubs and switches