IEEE 802.3 EtherNET

- Basics for MAC
- IEEE 802.3 Ethernet
- MAC address

MAC protocols: taxonomy

Three broad classes:

- channel partitioning
  - divide channel into smaller "pieces" (time slots, frequency, code)
  - allocate piece to node for exclusive use
  - FDMA(Cellular 1G), TDMA(Cellular 2G), CDMA(Cellular 2G, 3G), OFDMA+TDMA(Cellular 4G/5G)

    Lecture-10 (CDMA)

    Lecture-10 (OFDMA)
- random access
  - channel not divided, allow collisions
  - "recover" from collisions
  - CSMA/CD (Ethernet), CSMA/CA (WiFi)

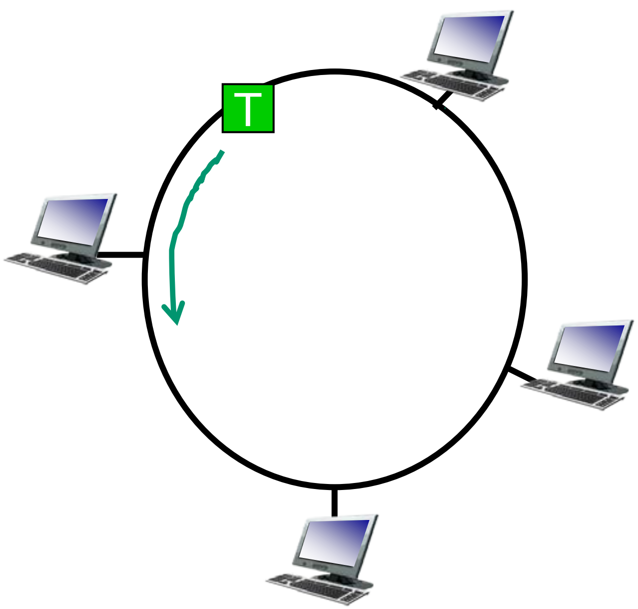
    Lecture-8 (CSMA/CD)

    Lecture-9 (CSMA/CA)
- "taking turns"
  - nodes take turns, but nodes with more to send can take longer turns

"Taking turns" MAC

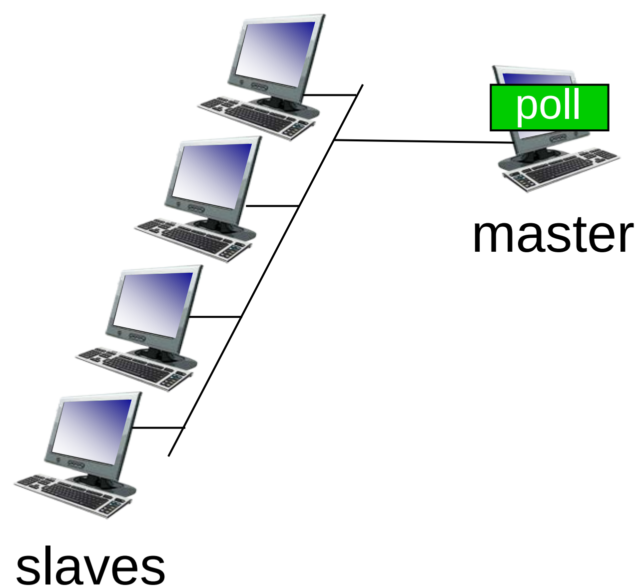
token passing:

- A special control frame (token) is passed from one node to next sequentially.
- concerns:
  - token overhead
  - latency
  - token failure



polling: (HDLC NRM)

- The master node "invites" slave nodes to transmit in turn
- concerns:
  - polling overhead
  - latency
  - master failure

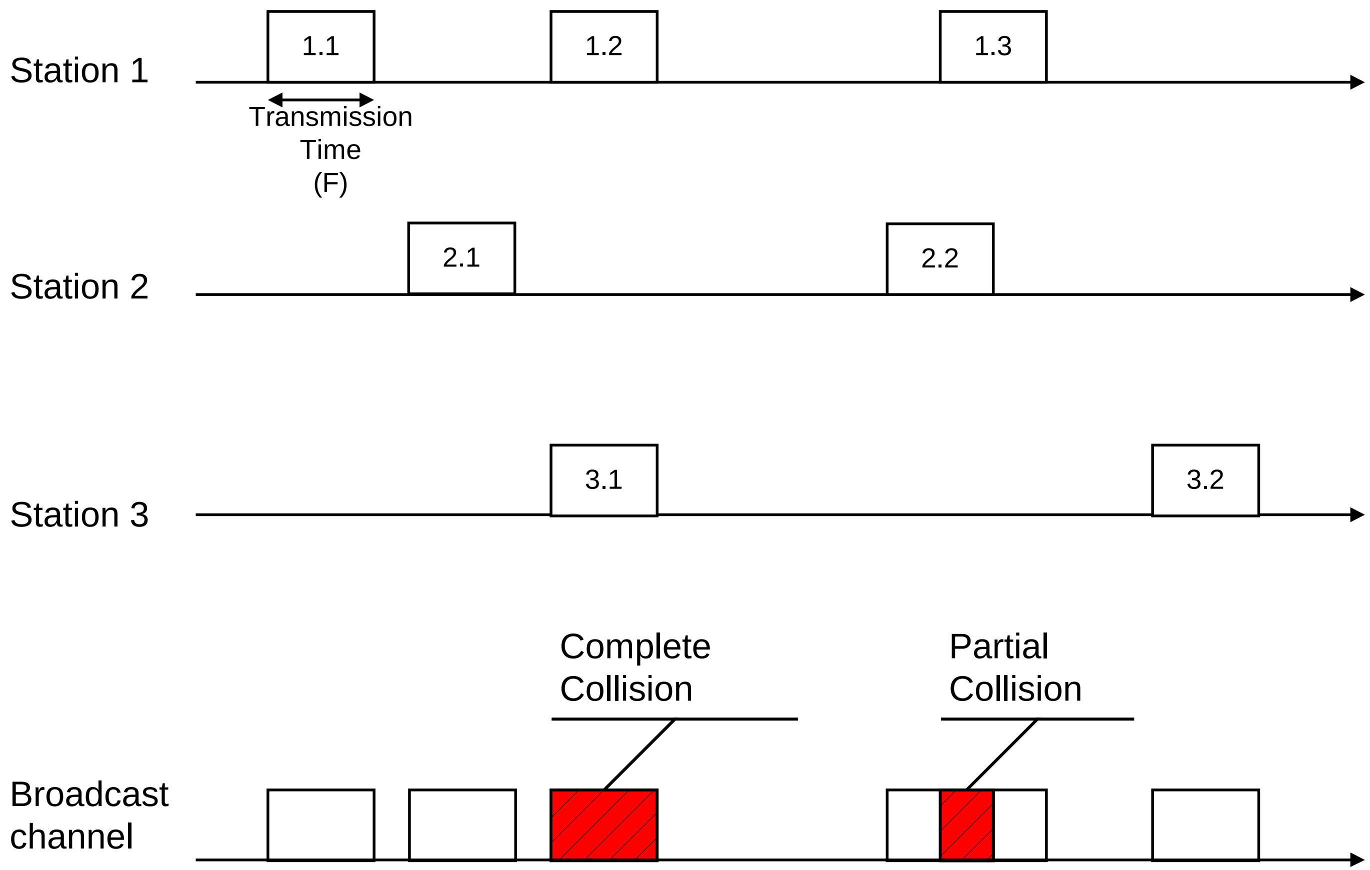


Random Access MAC

- Precursors of IEEE 802.3 MAC
  - Pure ALOHA
  - Slotted-ALOHA (S-ALOHA)
  - Carrier Sense Multiple Access (CSMA)
- IEEE 802.3 MAC
  - CSMA-CD (collision detection)
- IEEE 802.11 MAC
  - CSMA-CA (collision avoidance)

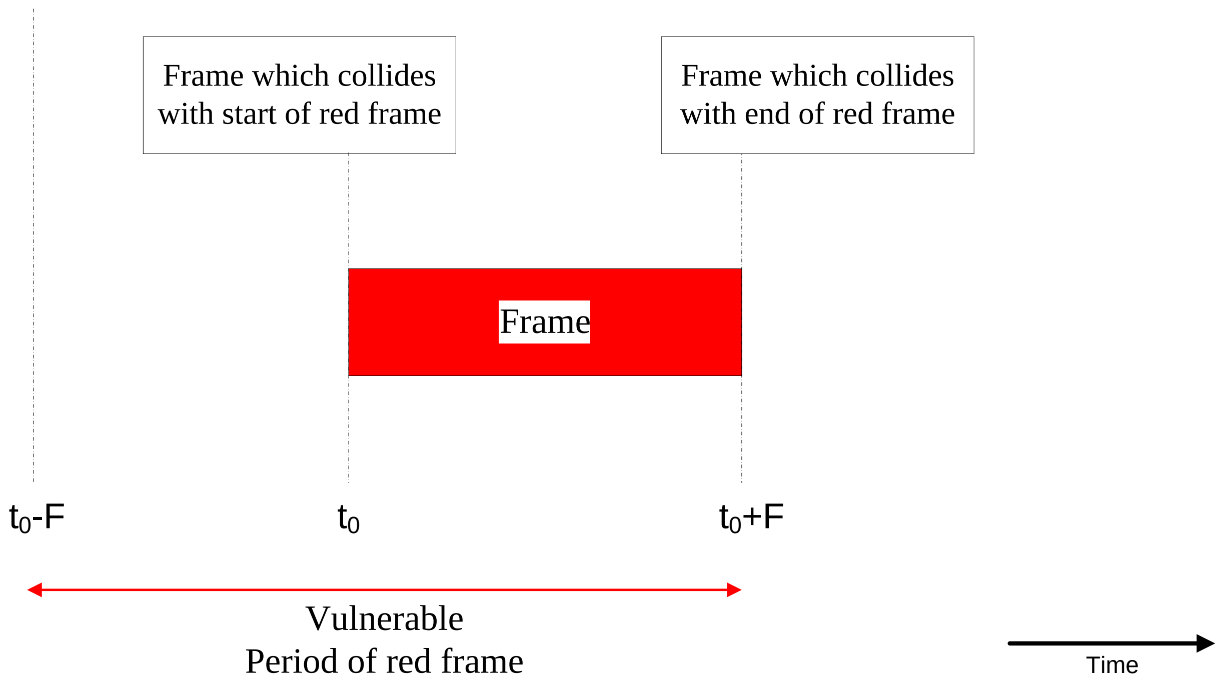
Pure ALOHA (1/2)

- In pure ALOHA, frames are transmitted at completely arbitrary times.



Pure ALOHA (2/2)

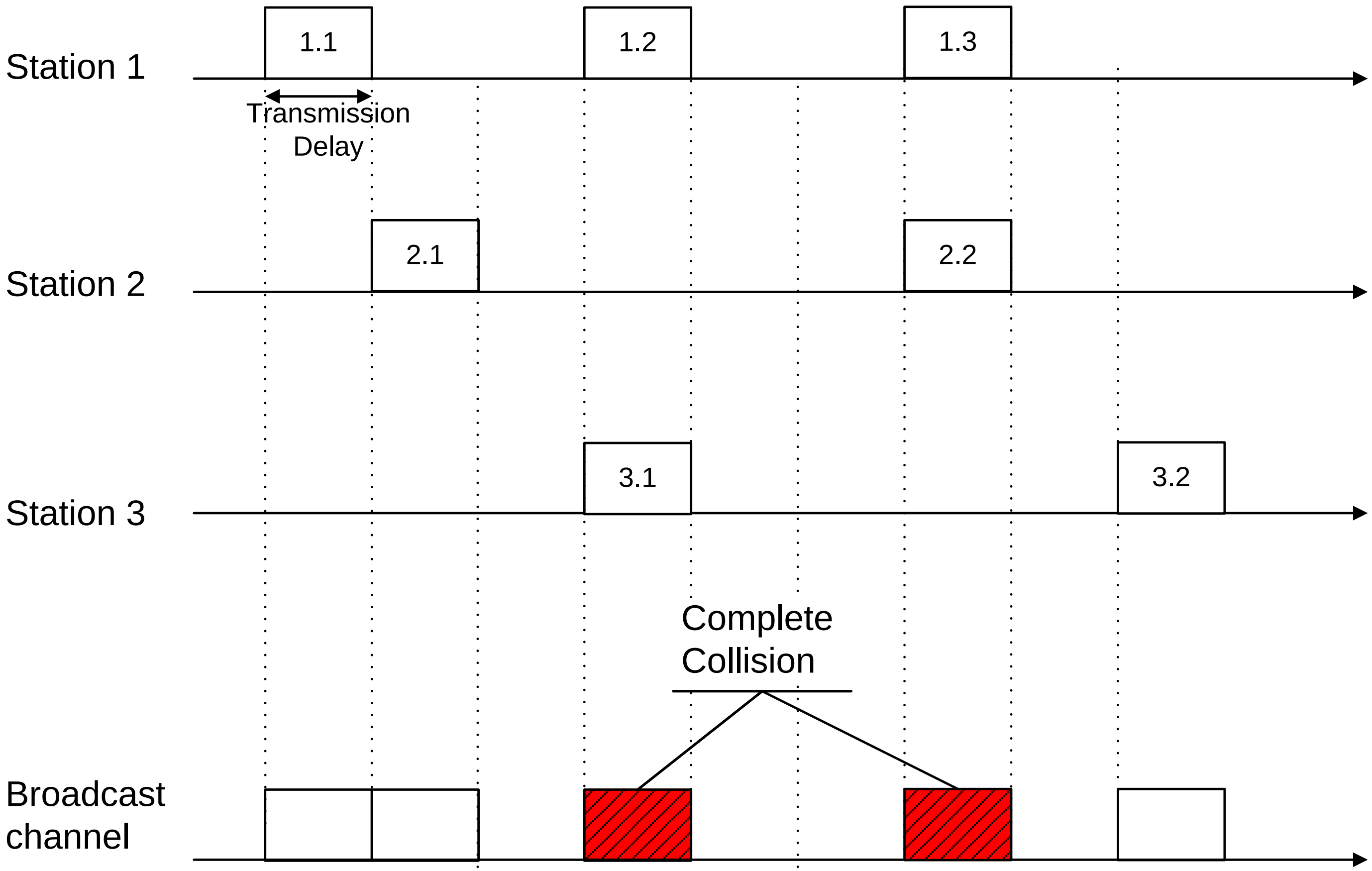
- Vulnerable period for the red frame



  - A frame (red frame) will be in a collision if and only if another transmission begins in the vulnerable period of the frame
  - Vulnerable period has the length of 2 frame times

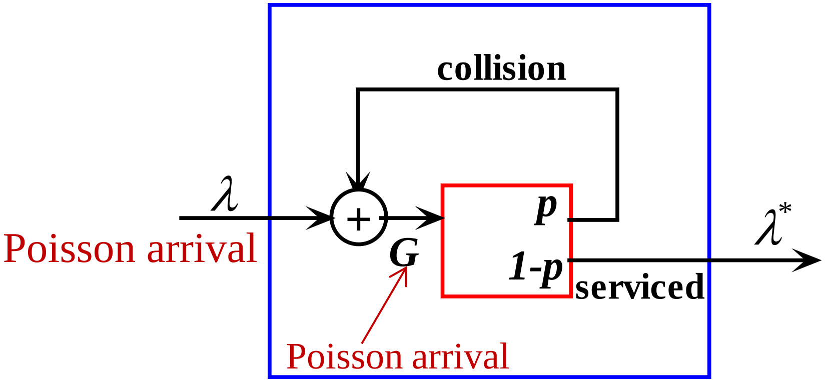
S-ALOHA

- ALOHA with an additional constraint
- Time is divided into discrete time intervals (slot)
- A station can transmit a frame only at the beginning of a slot



Performance of ALOHA system (1/2)

- Analysis Model



$p \equiv$ probability of collision

$= 1-$ Pr{no arrival during vulnerable period}

$G \equiv$ total carried load (incl. Retx)

$\Rightarrow G = \lambda + pG$

- ALOHA: $p = 1 - (2G)^0 e^{-2G}/0! = 1 - e^{-2G}$

$$\lambda = \lambda^* = G(1-p) = Ge^{-2G}$$

- S-ALOHA: $p = 1 - G^0 e^{-G}/0! = 1 - e^{-G}$

$$\lambda^* = G(1-p) = Ge^{-G}$$

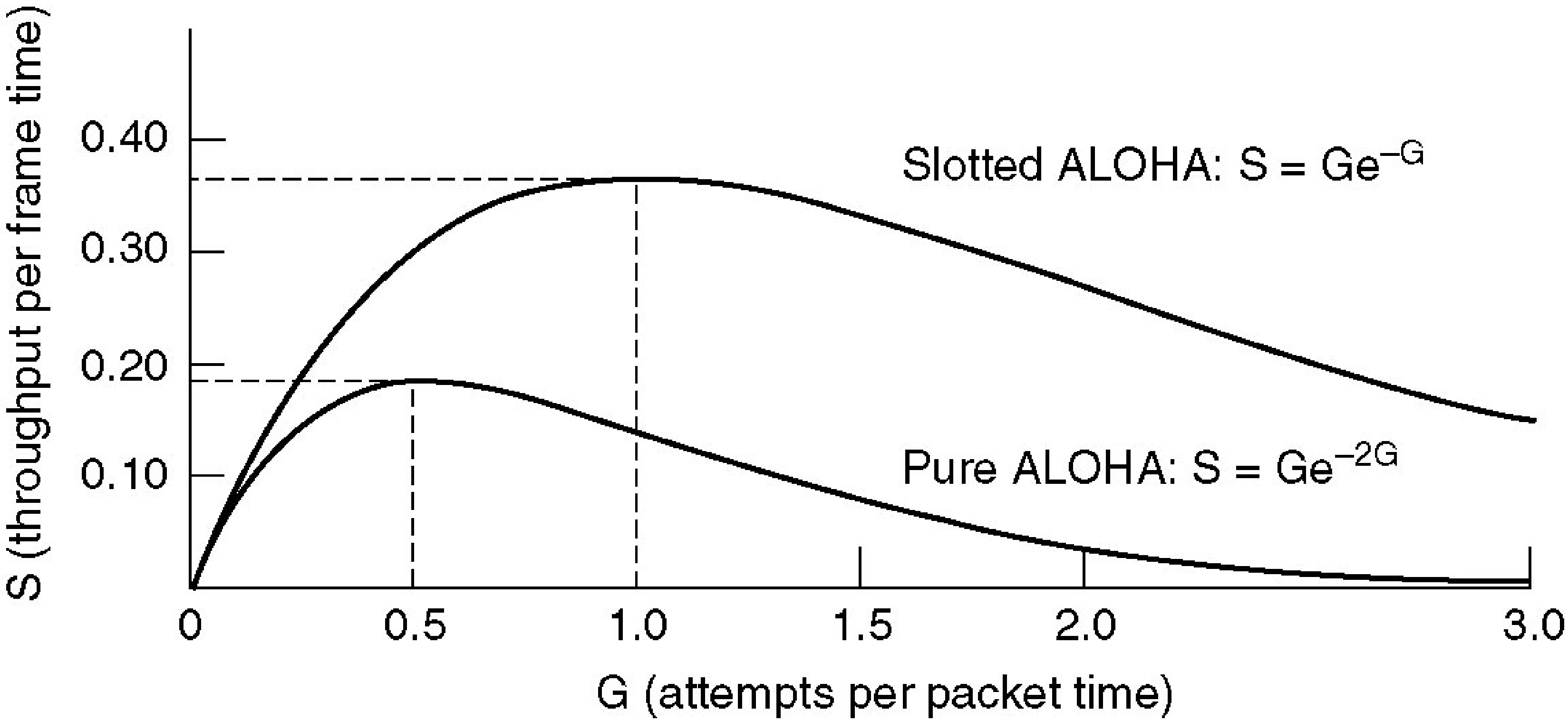
Poisson with rate α

$$\frac{(\alpha t)^k e^{-\alpha t}}{k!}$$

Probability of k arrivals during t

Performance of ALOHA systems (2/2)

- Throughput versus offered traffic



CSMA (1/2)

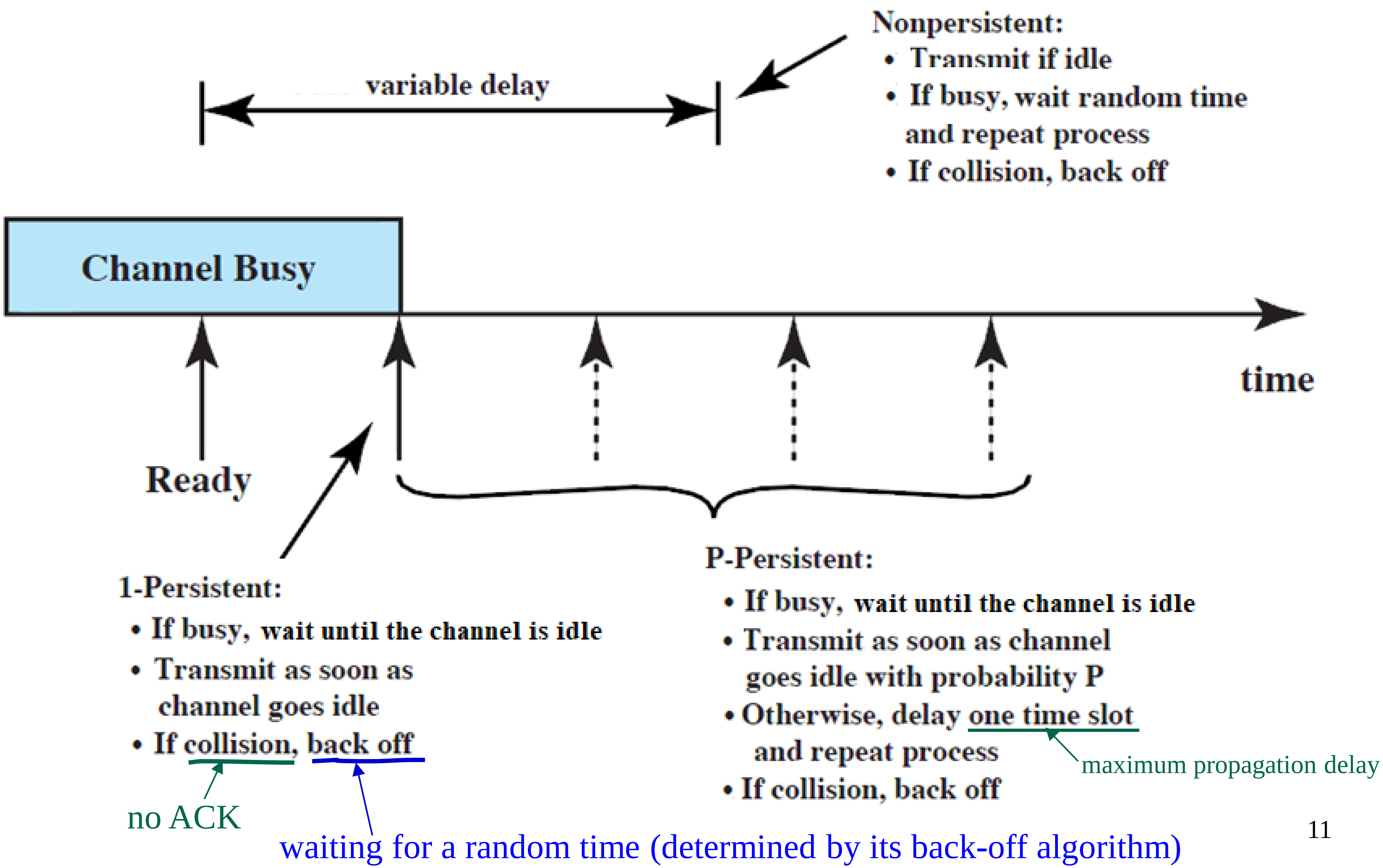
- Random access based on contention
- "Listen Before Talk" discipline
  - When maximum propagation time (t_p) << Transmission time (t_d), more improvement is possible.

$t_p < t_d$:

$t_p > t_d$:

- Three CSMA schemes
  - Non-persistent CSMA
  - *p*-persistent CSMA
  - 1-persistent CSMA

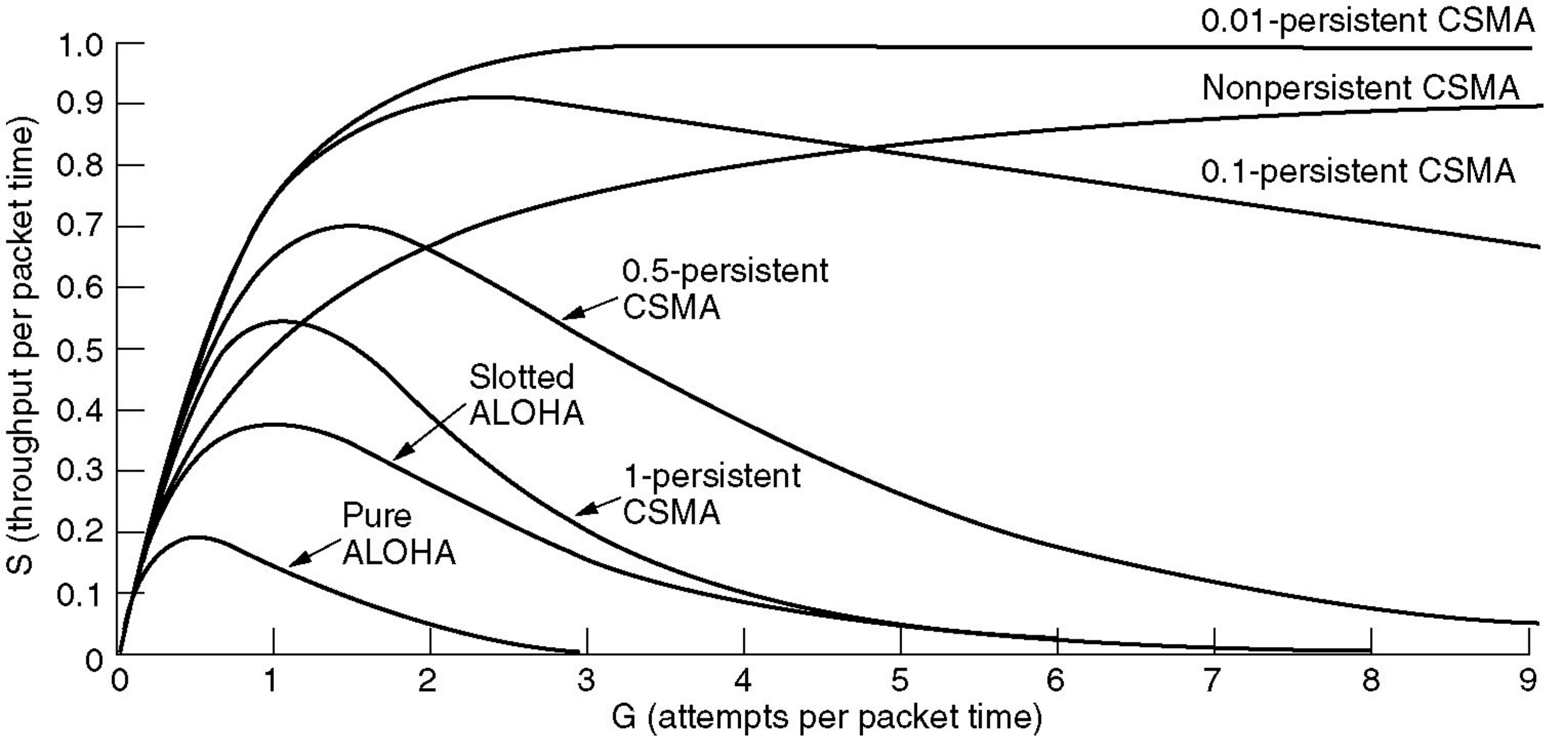
CSMA (2/2)



Nonpersistent:
• Transmit if idle
• If busy, wait random time and repeat process
• If collision, back off
variable delay
Channel Busy
time
Ready
1-Persistent:
• If busy, wait until the channel is idle
• Transmit as soon as channel goes idle
• If collision, back off
no ACK
waiting for a random time (determined by its back-off algorithm)
P-Persistent:
• If busy, wait until the channel is idle
• Transmit as soon as channel goes idle with probability P
• Otherwise, delay one time slot and repeat process
maximum propagation delay
• If collision, back off

Performance Comparison

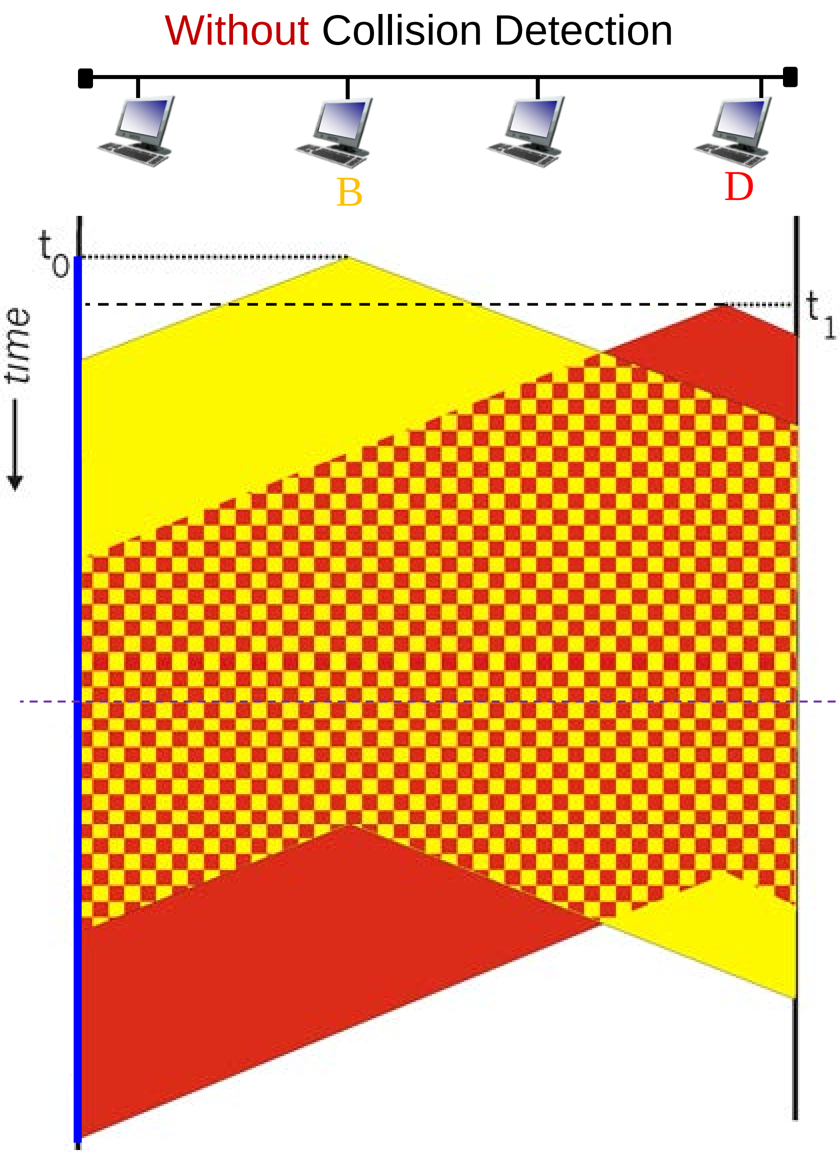
- Throughput versus offered traffic load
  - Throughput: Mean number of packets which are transmitted successfully for a packet time

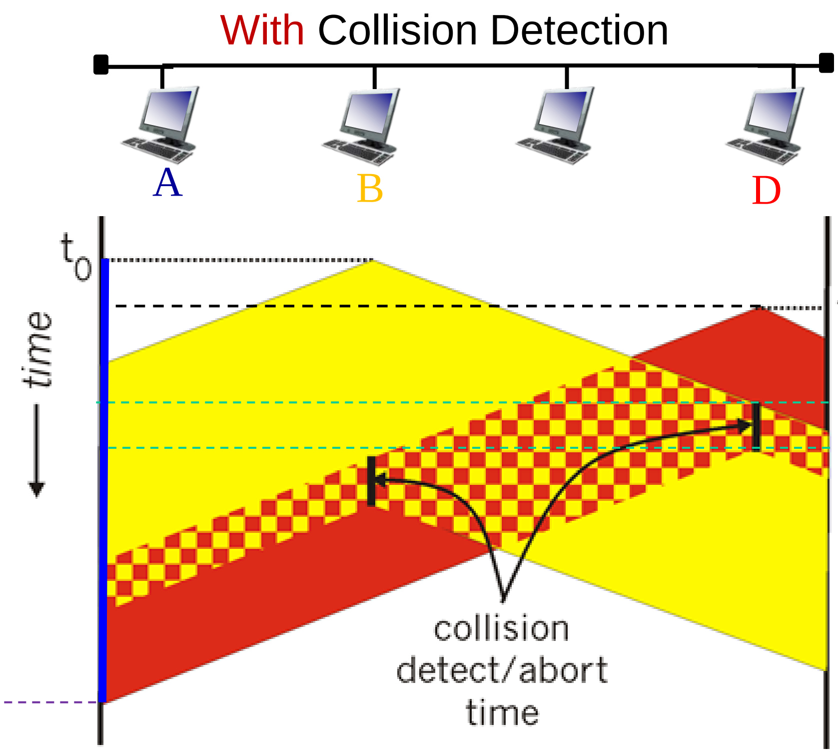


IEEE 802.3 MAC: CSMA/CD (1/3)

- 1-persistent CSMA + Collision Detection
  - Listen before transmission till channel is free (CSMA)
  - If the channel idle, transmit (1-persistent)
  - Additionally continue to monitor the channel during transmission, and if collision is detected, immediately abort transmission (collision detection)
  - If the transmission is aborted due to collision, try to retransmit after the delay which is determined according to the binary exponential backoff scheme

Collision Detection





- Maximum collision detection time: $2 \times T_{prop}$
- Packet length should be so that a packet transmission time is at least twice the propagation delay: $a \leq 0.5$

IEEE 802.3 CSMA/CD (2/3)

- Binary exponential backoff
  - The delay is an integral multiple of slot time.
  - The number of slot time to delay before the *n*th retransmission attempt is chosen as an uniformly distributed random integer r in the range $1 \le r \le 2^K$, where K=min(n,10).

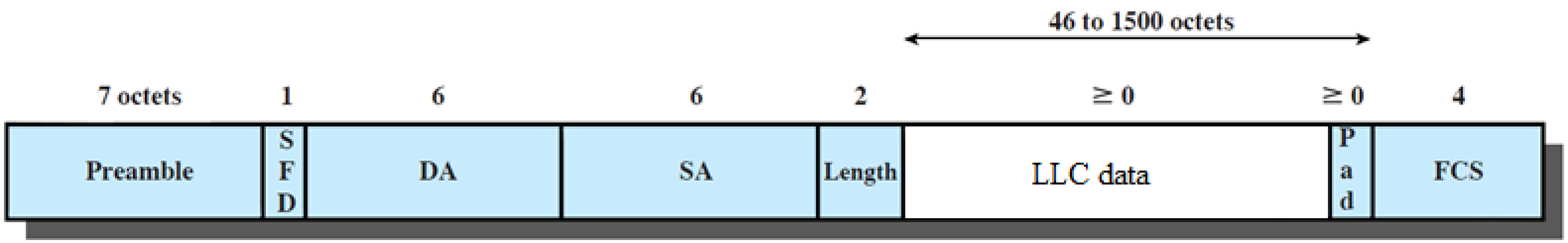
range length: 2^K; *mean delay* = $(2^K+1)/2 \approx 2^K/2$

1st attempt: the range [1, 2] : mean delay =1.5
2nd attempt: the range [1, 2, 3, 4]: mean dela=2.5
⋮
9th attempt: the range [1, 2, …, 512]: mean delay=256.5
10th attempt: the range [1, 2, …, 1024]: mean delay=512.5
⋮
14th attempt: the range [1, 2, …, 1024]: mean delay=512.5

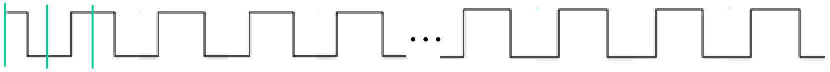
IEEE 802.3 CSMA/CD (3/3)

- Short transmission delay at light (low) loads.
  - 1-persistent CSMA
- reduces "bandwidth waste"
  - by aborting the transmission immediately at collision detection
- improves utilization at high loads
  - by using the binary exponential backoff scheme, which doubles mean delay at each collision (adaptive control according to network load condition)

IEEE 802.3 MAC Frame Format



- Preamble: A 7-octet pattern of alternating 0s and 1s used by the receiver to establish bit synchronization (establishes the rate at which bit are sampled)



- Start frame delimiter (SFD): Special pattern 10101011 indicating the start of a frame (frame sync)
- Length: Length of the LLC data field
- LLC data
- Pad: Octets added to ensure that the frame is long enough for proper CD operation
- FCS: Error checking using 32-bit CRC
  - For erroneous frame, just discard the frame.

IEEE 802.3 10-Mbps Spec. (Ethernet)

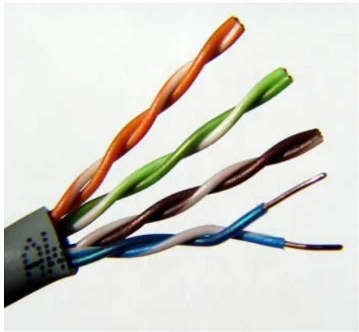
- MAC: CSMA/CD
- Many alternative physical configurations

	10BASE5	10BASE2	10BASE-T	10BASE-FP
Transmission medium	Coaxial cable (50 ohm)	Coaxial cable (50 ohm)	Unshielded twisted pair	850-nm optical fiber pair
Signaling technique	Baseband (Manchester)	Baseband (Manchester)	Baseband (Manchester)	Manchester/on-off
Topology	Bus	Bus	Star	Star
Maximum segment length (m)	500	185	100	500
Nodes per segment	100	30	—	33
Cable diameter (mm)	10	5	0.4 to 0.6	62.5/125 μm

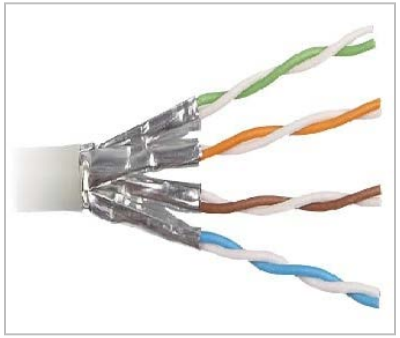
Transmission Medium (1)

- Separately insulated
- Twisted together
- Often "bundled" into cables
- Usually installed in buliding when built

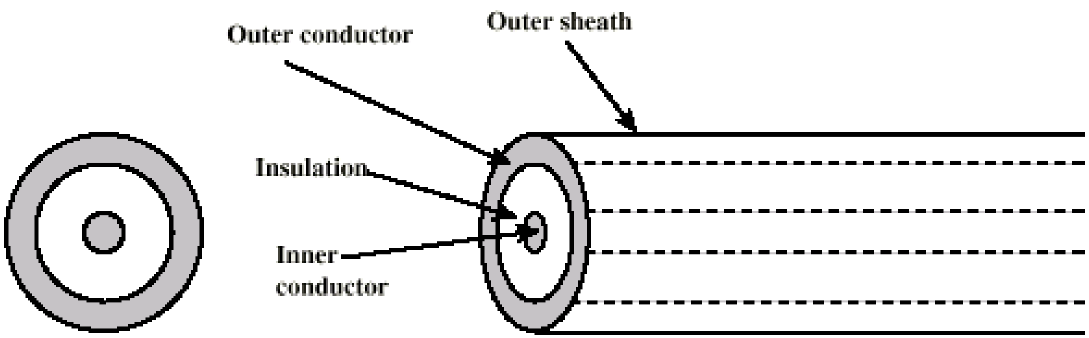
(a) Twisted pair



UTP



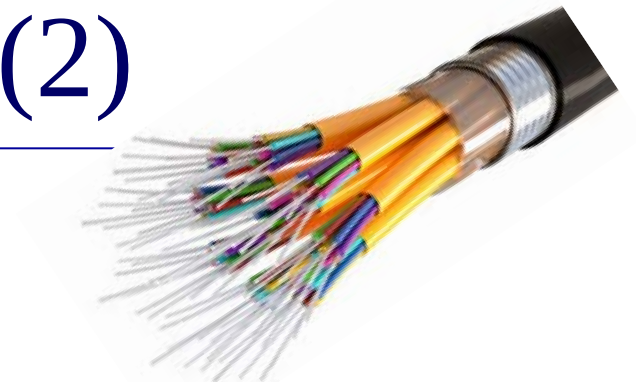
STP



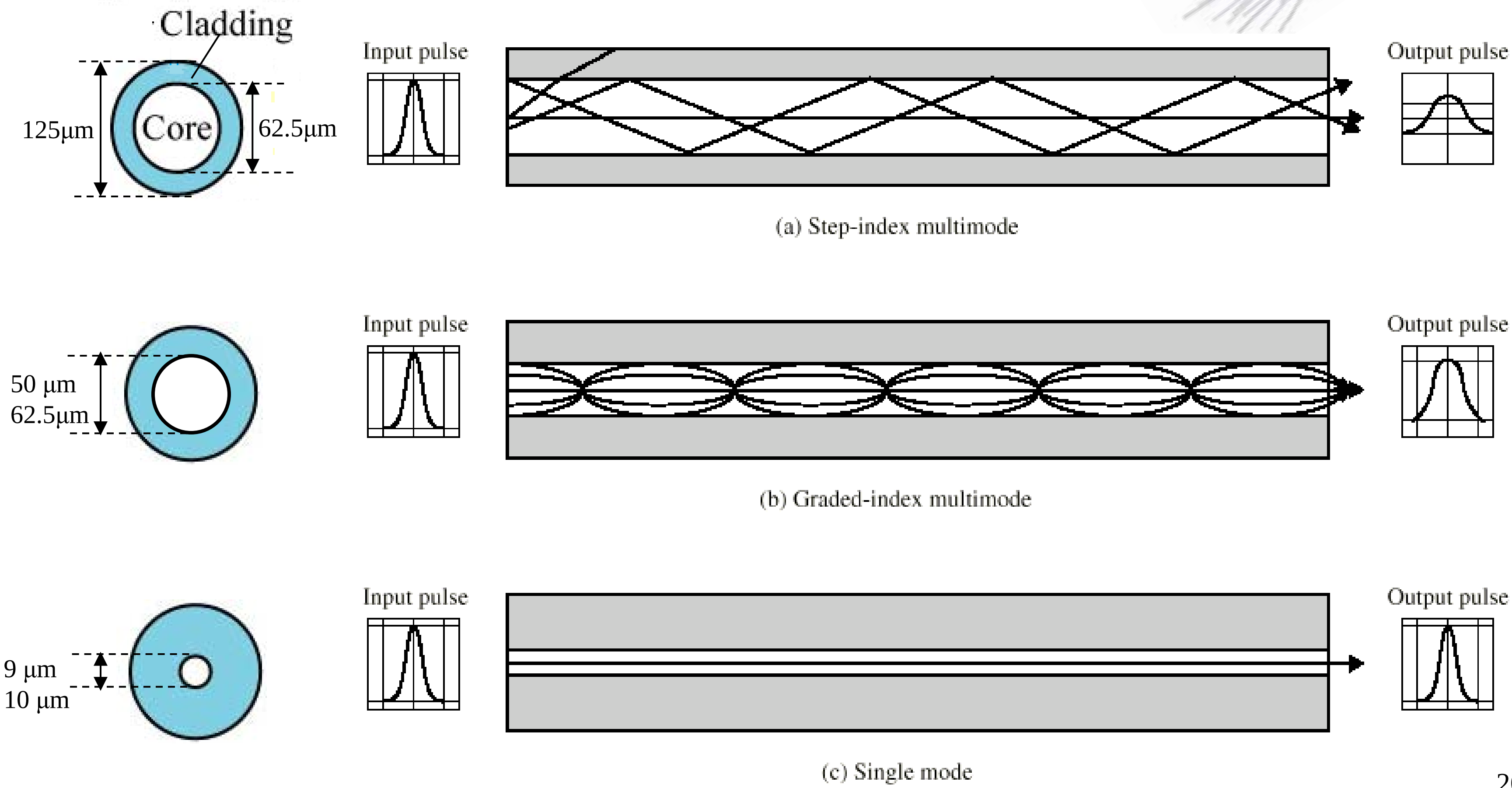
- Outer conductor is braided shield
- Inner conductor is solid metal
- Separated by insulating material
- Covered by padding

(b) Coaxial cable

Transmission Medium (2)



- Optical Fiber



Transmission Medium (3)

- Optical Fiber

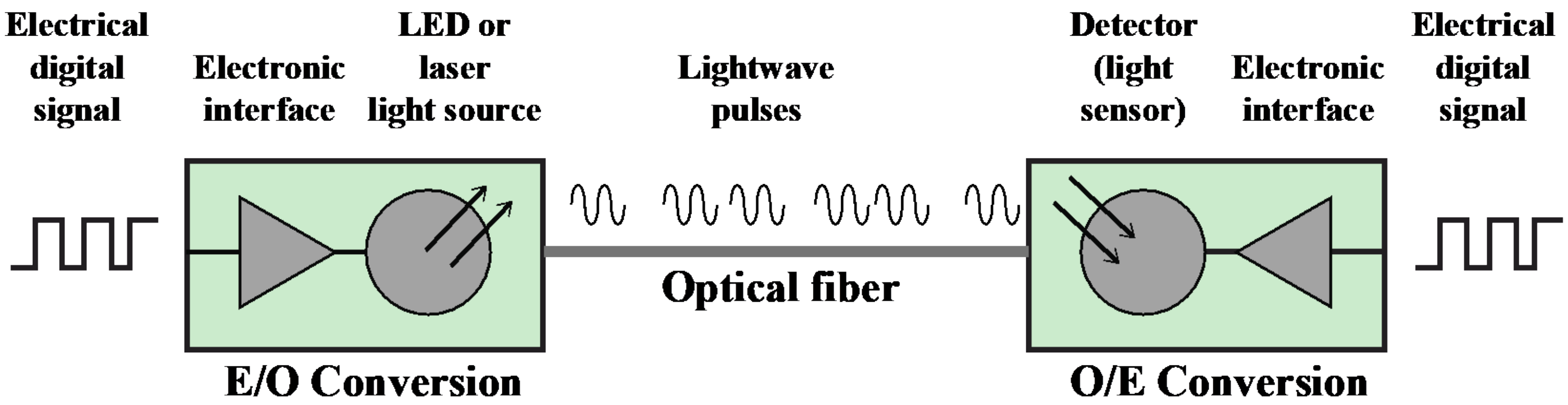
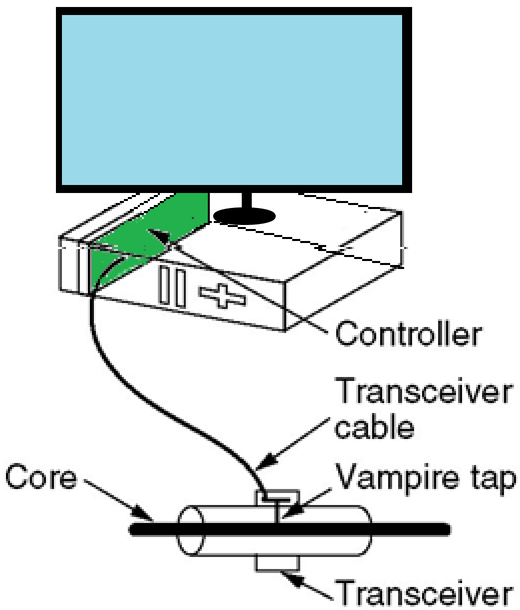
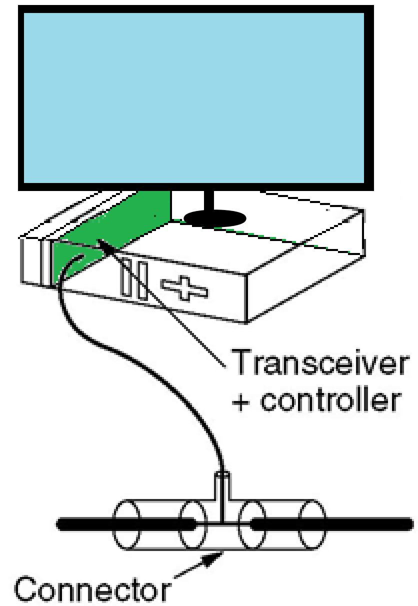


Figure 4.6 Optical Communication

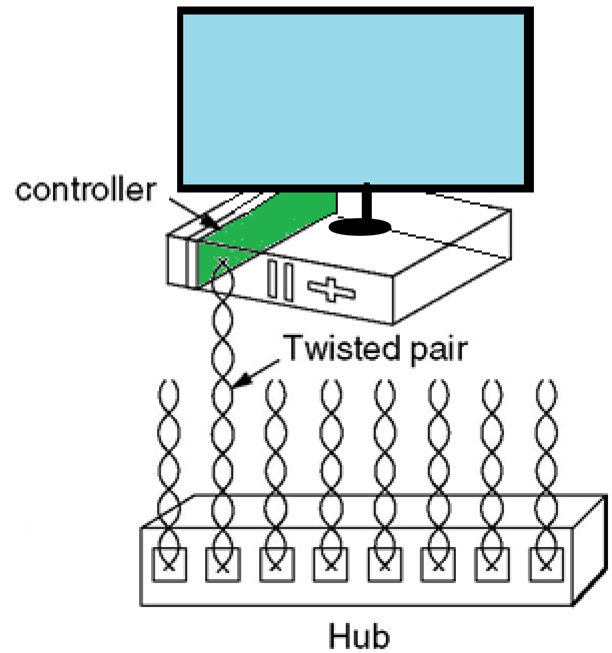
Ethernet Cabling



(a) 10Base5

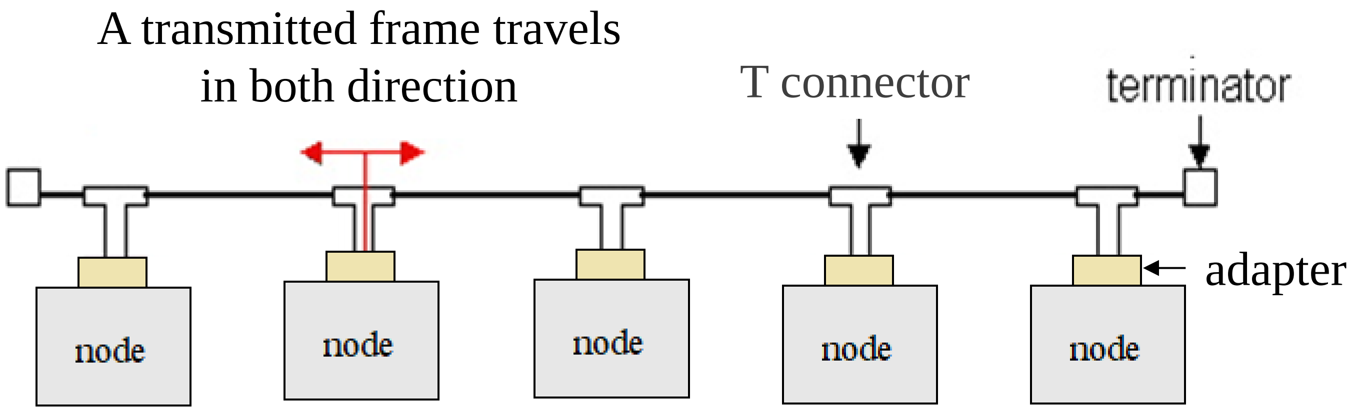


(b) 10Base2

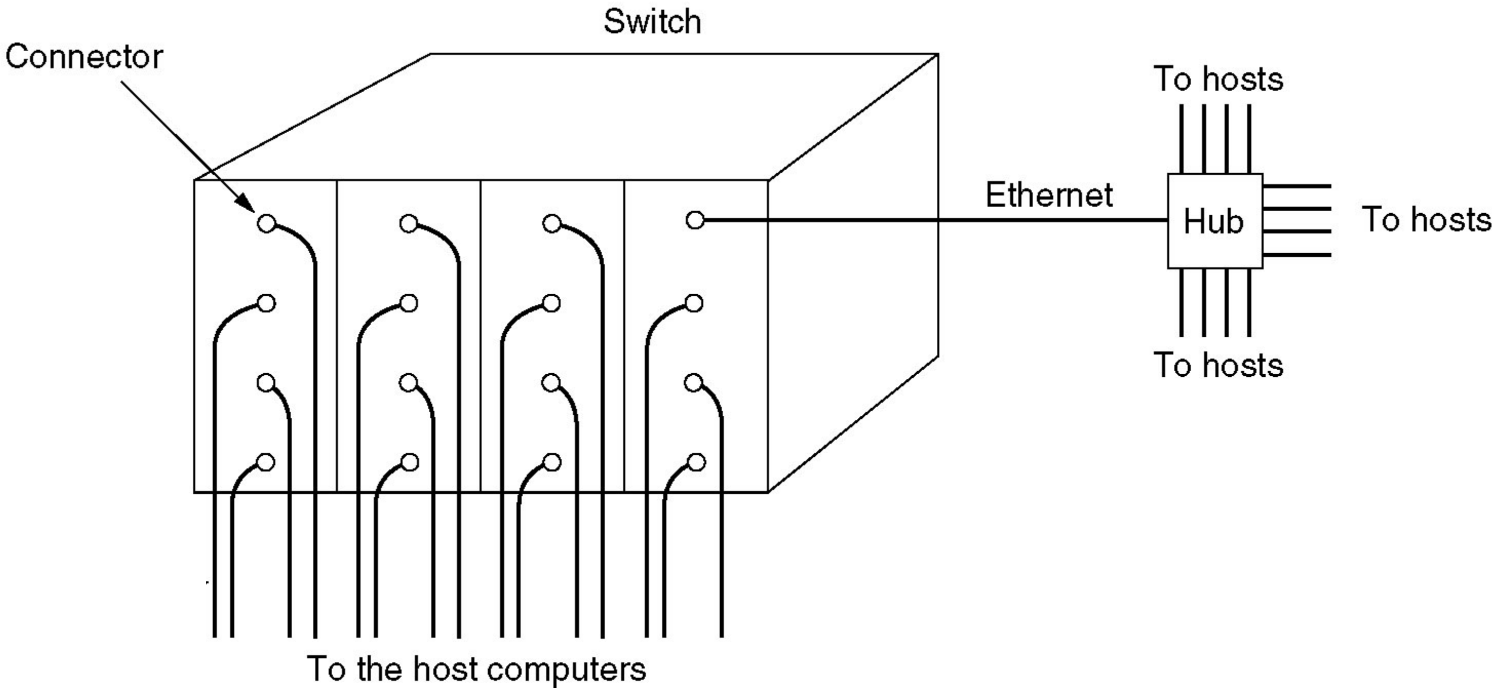


(c) 10Base-T

10BASE2


A transmitted frame travels
in both direction
T connector
terminator
adapter
node
node
node
node
node

Ethernet Switch

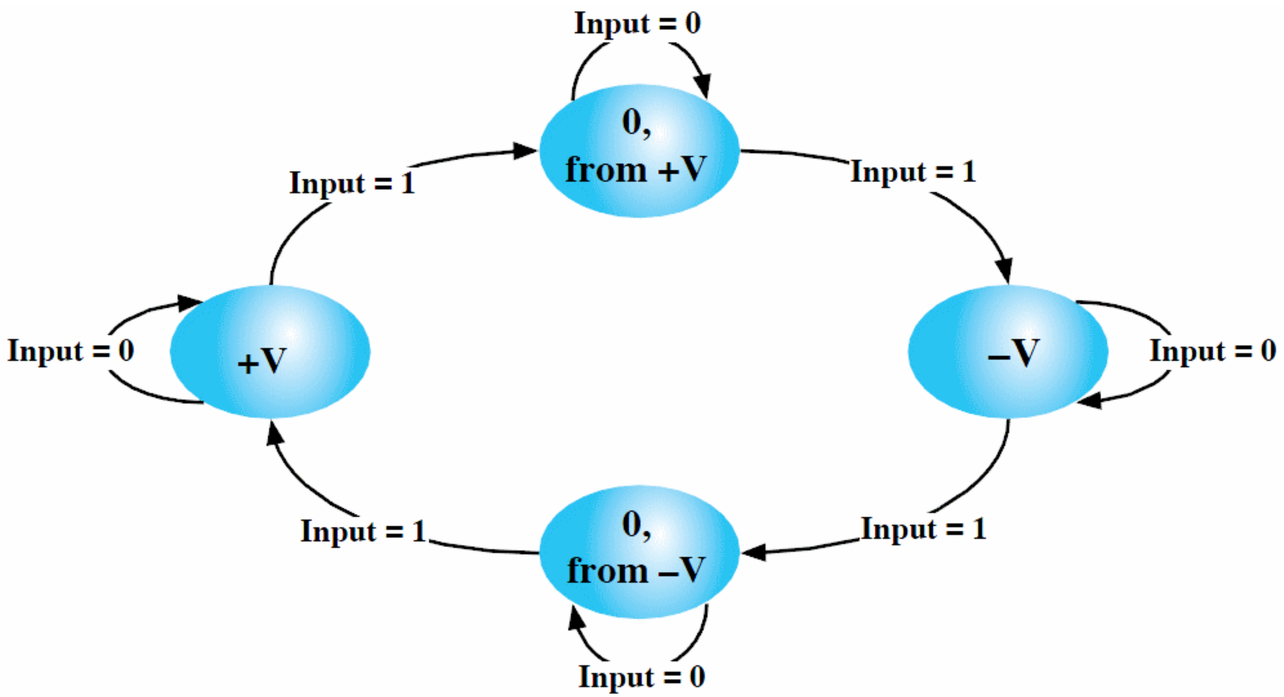


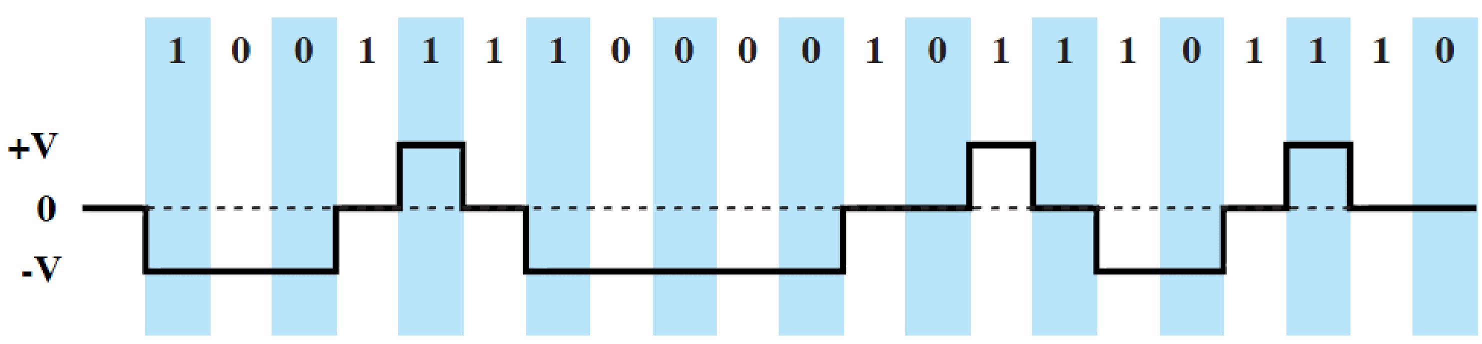
IEEE 802.3 100Mbps Spec. (Fast Ethernet)

- developed to provide a low-cost (10Mbps) Ethernet compatible LAN operating at 100 Mbps
- MAC: CSMA/CD
- Physical layer: Medium alternatives

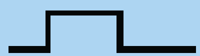
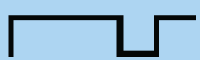
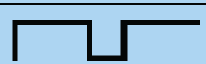
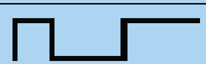
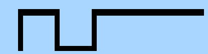
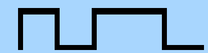
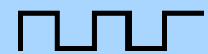
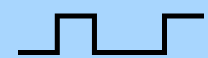
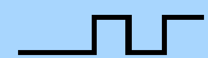
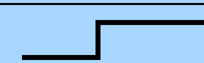
	100BASE-TX		100BASE-FX	100BASE-T4
Transmission medium	2 pair, STP	2 pair, Category 5 UTP	2 optical fibers	4 pair, Category 3, 4, or 5 UTP
Signaling technique	MLT-3	MLT-3	4B5B, NRZI	8B6T, NRZ
Data rate	100 Mbps	100 Mbps	100 Mbps	100 Mbps
Maximum segment length	100 m	100 m	100 m	100 m
Network span	200 m	200 m	400 m	200 m

MLT-3 Encoding

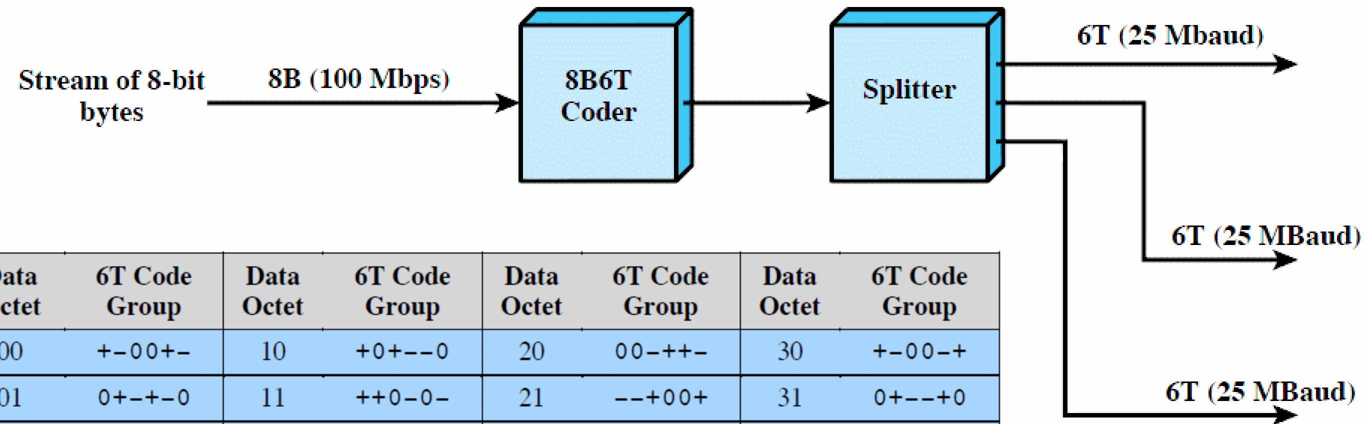




4B5B + NRZI

Data Input (4 bits)	Code Group (5 bits)	NRZI pattern	Interpretation
0000	11110		Data 0
0001	01001		Data 1
0010	10100		Data 2
0011	10101		Data 3
0100	01010		Data 4
0101	01011		Data 5
0110	01110		Data 6
0111	01111		Data 7
1000	10010		Data 8
1001	10011		Data 9
1010	10110		Data A
1011	10111		Data B
1100	11010		Data C
1101	11011		Data D
1110	11100		Data E
1111	11101		Data F
	11111		Idle
	11000		Start of stream delimiter, part 1
	10001		Start of stream delimiter, part 2
	01101		End of stream delimiter, part 1
	00111		End of stream delimiter, part 2
	00100		Transmit error
	other		invalid codes

8B6T Transmitter of 100BASE-T4



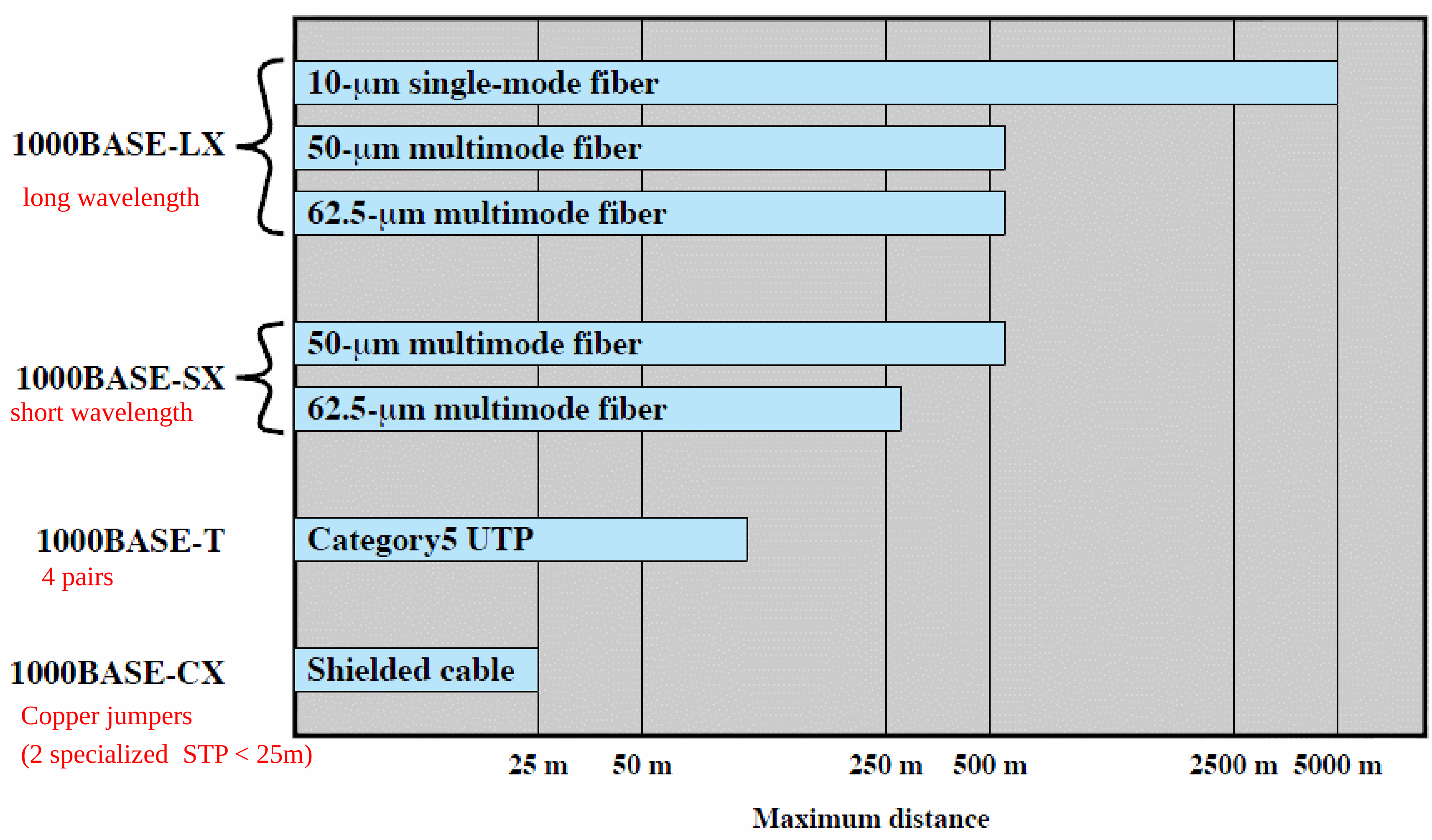
Data Octet	6T Code Group	Data Octet	6T Code Group	Data Octet	6T Code Group	Data Octet	6T Code Group
00	+-00+-	10	+0+--0	20	00-++-	30	+-00-+
01	0+-+-0	11	++0-0-	21	--+00+	31	0+--+0
02	+-0+-0	12	+0+-0-	22	++-0+-	32	+-0-+0
03	-0++-0	13	0++-0-	23	++-0-+	33	-0+-+0
04	-0+0+-	14	0++--0	24	00+0-+	34	-0+0-+
05	0+--0+	15	++00--	25	00+0+-	35	0+-+0-
06	+-0-0+	16	+0+0--	26	00-00+	36	+-0+0-
07	-0+-0+	17	0++0--	27	--+++-	37	-0++0-
08	-+00+-	18	0+-0+-	28	-0-++0	38	-+00-+
09	0-++-0	19	0+-0-+	29	--0+0+	39	0-+-+0
0A	-+0+-0	1A	0+-++-	2A	-0-+0+	3A	-+0-+0
0B	+0-+-0	1B	0+-00+	2B	0--+0+	3B	+0--+0
0C	+0-0+-	1C	0-+00+	2C	0--++0	3C	+0-0-+
0D	0-+-0+	1D	0-+++-	2D	--00++	3D	0-++0-
0E	-+0-0+	1E	0-+0-+	2E	-0-0++	3E	-+0+0-
0F	+0--0+	1F	0-+0+-	2F	0--0++	3F	+0-+0-

Four Pairs
: one Tx, one Rx,
two Tx/Rx pairs

Gigabit Ethernet: MAC

- CDMA/CD
- Two enhancements to the basic CDMA/CD
  - Carrier extension
    - So that the frame transmission time gets longer than the round-trip propagation time at 1 Gbps
    - At least 4096 bit-times long (512 bit-times for 10/100 Mbps)
  - Frame bursting
    - allows for multiple short frames to be transmitted consecutively, without relinquishing control for CSMA/CD between frames
    - avoids the overhead of carrier extension when a single station has a number of small frames

Gigabit Ethernet: PHY



Gigabit Ethernet Configuration Example

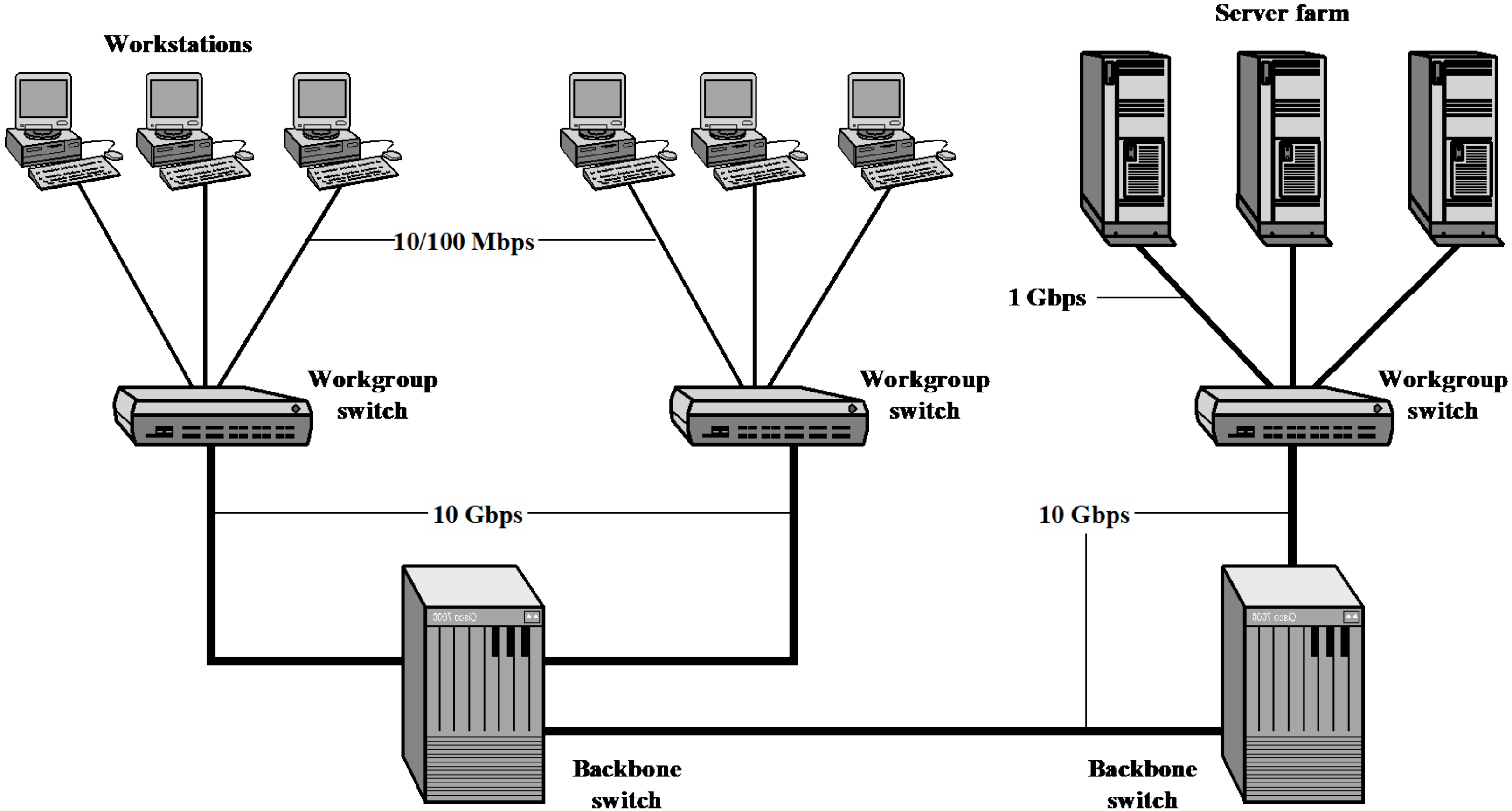


Figure 12.6 Example 10 Gigabit Ethernet Configuration

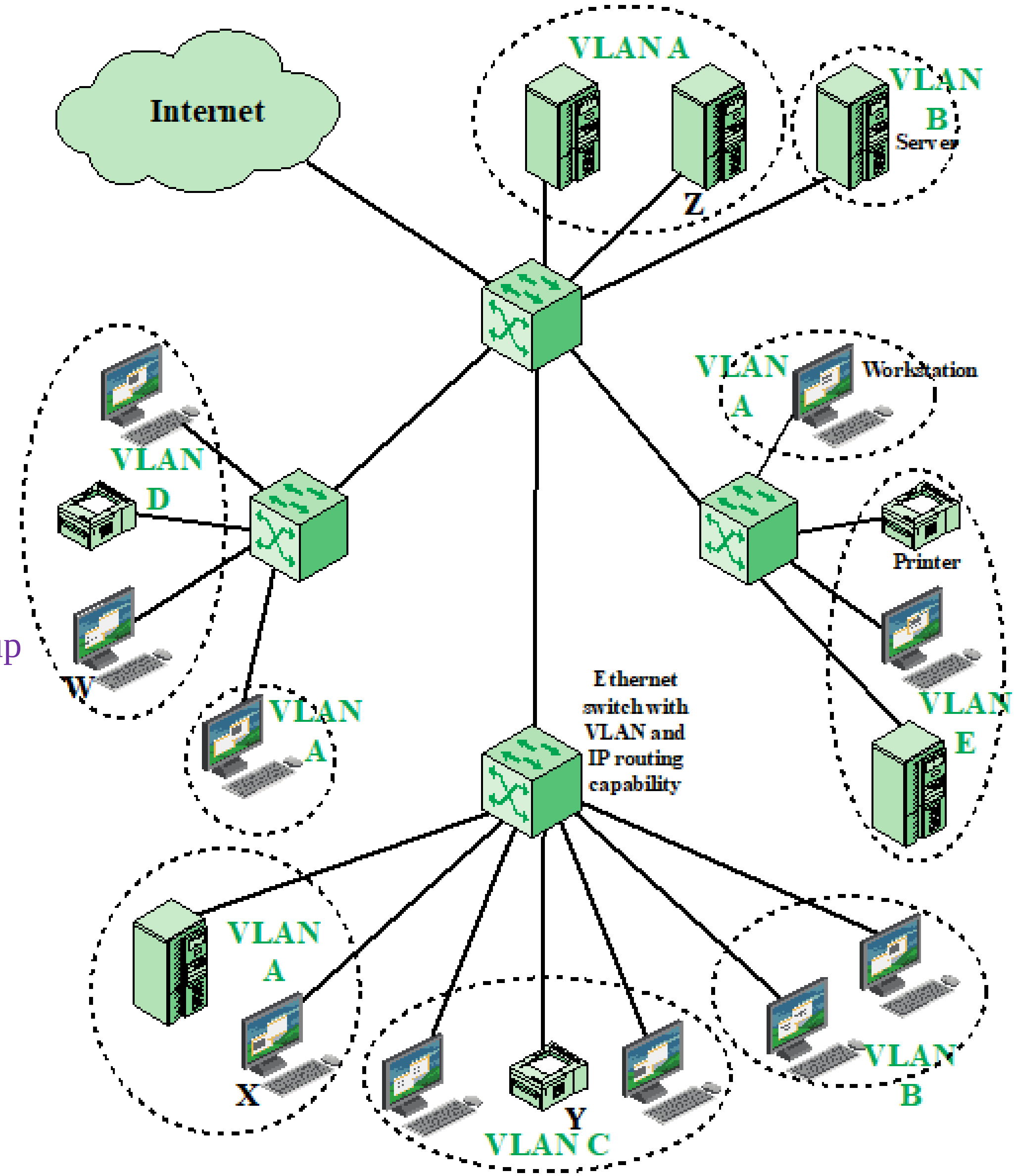
Virtual LAN

- Broadcast domain
  - A group of end stations that receive broadcast frames from each other

    - DA: FFFFFFFFFFFF (broadcast addr)
    - used for network management, alert

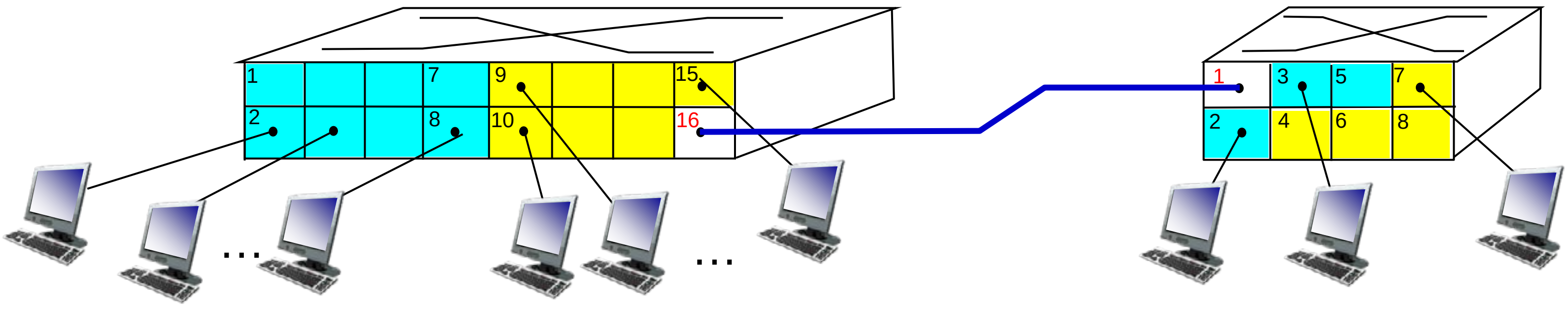
* Addressing
  - unicast: a single destination
  - multicast: some stations called a multicast group
  - broadcast: all stations

- VLAN
  - Broadcast domain consisting of a group of end stations not limited by physical location
  - Communicate as if they were on a common LAN



VLAN Membership (1/2)

- Membership by:
  - Port–based
  - MAC address-based
  - Protocol information (IP addr, Layer-4 protocol, application)–based



< Port-based membership >

VLAN Membership (2/2)

- Switches need to know VLAN membership
  - Configure information manually
  - Network management signaling protocol
- Frame tagging (IEEE802.1Q)

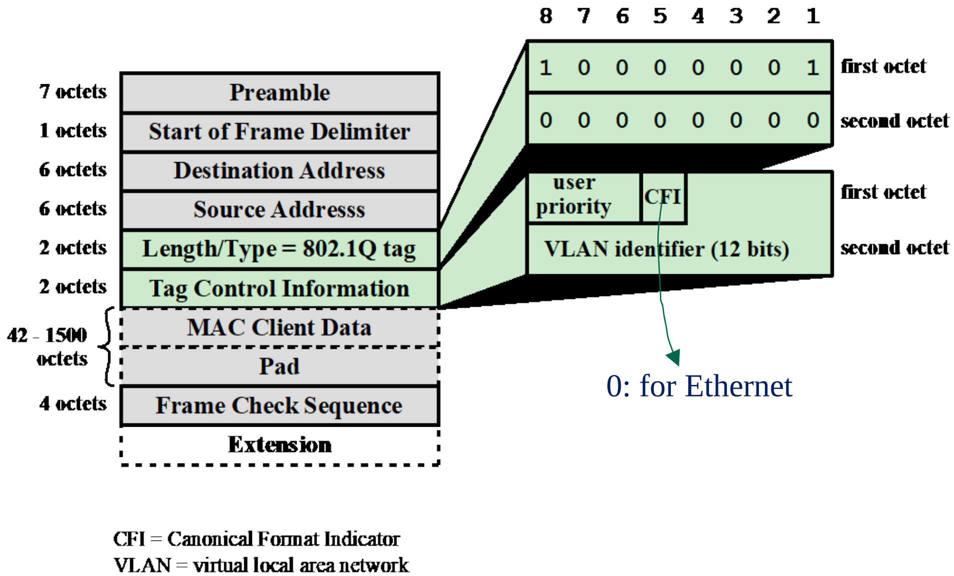


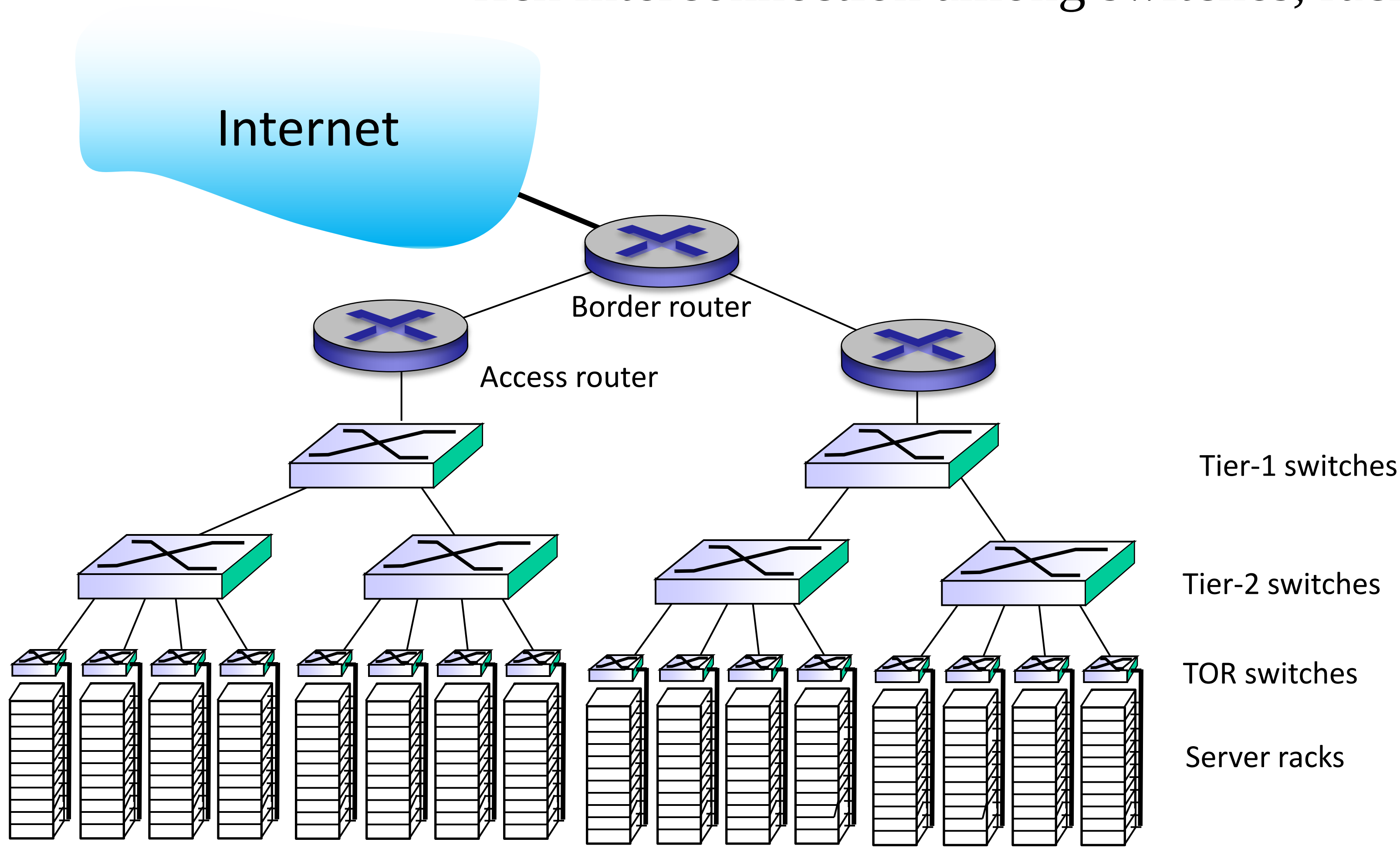
Figure 12.10 Tagged IEEE 802.3 MAC Frame Format

Data center networks (1/3)

- tens to hundreds of thousands of hosts, often closely coupled, in close proximity:
  - e-business (e.g. Amazon)
  - content-servers (e.g., YouTube, Apple, Microsoft)
  - search engines, data mining (e.g., Google)

Data center networks (2/3)

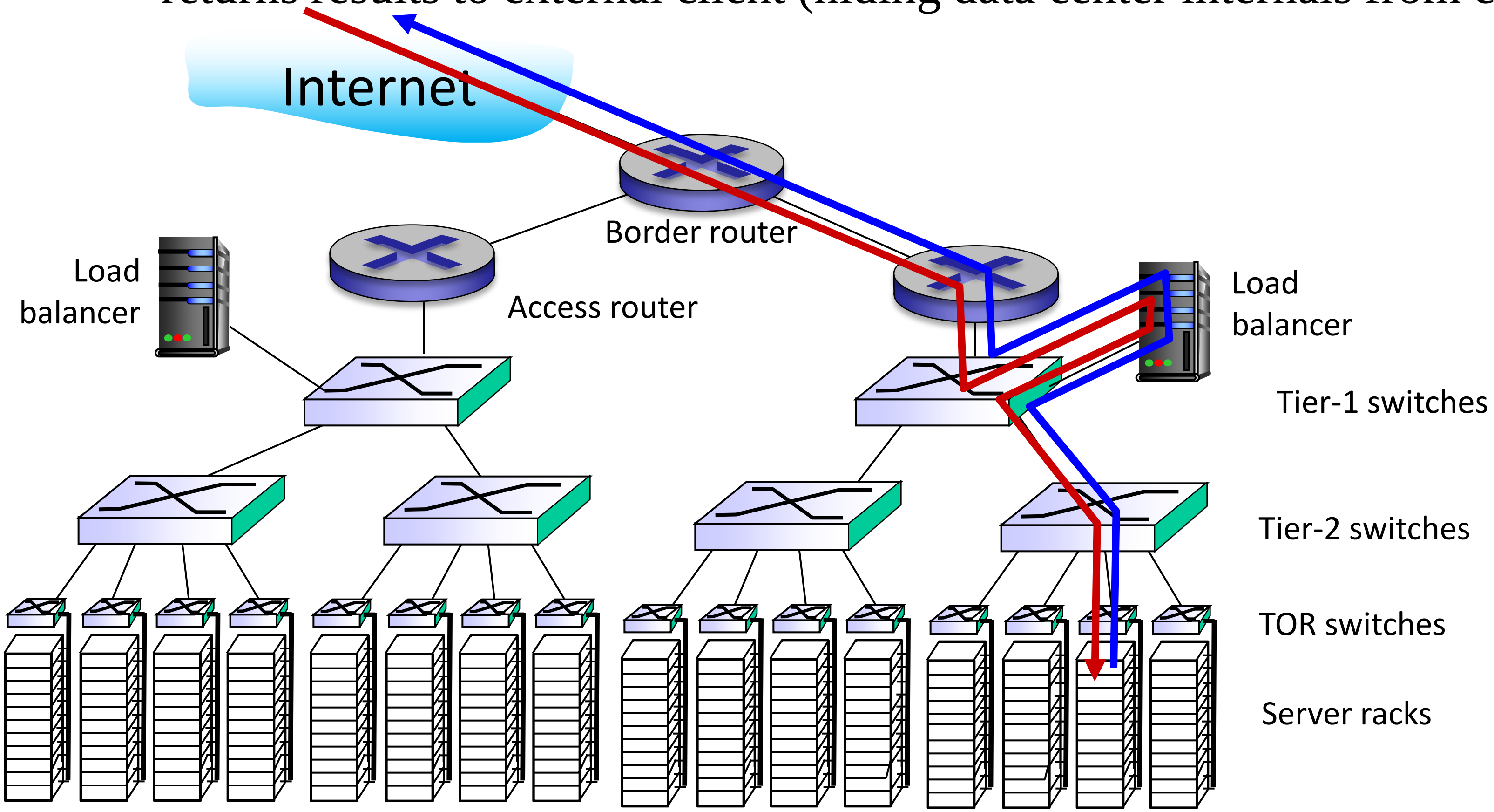
- rich interconnection among switches, racks:



Data center networks (3/3)

load balancer: application-layer routing

- receives external client requests
- directs workload within data center
- returns results to external client (hiding data center internals from client)



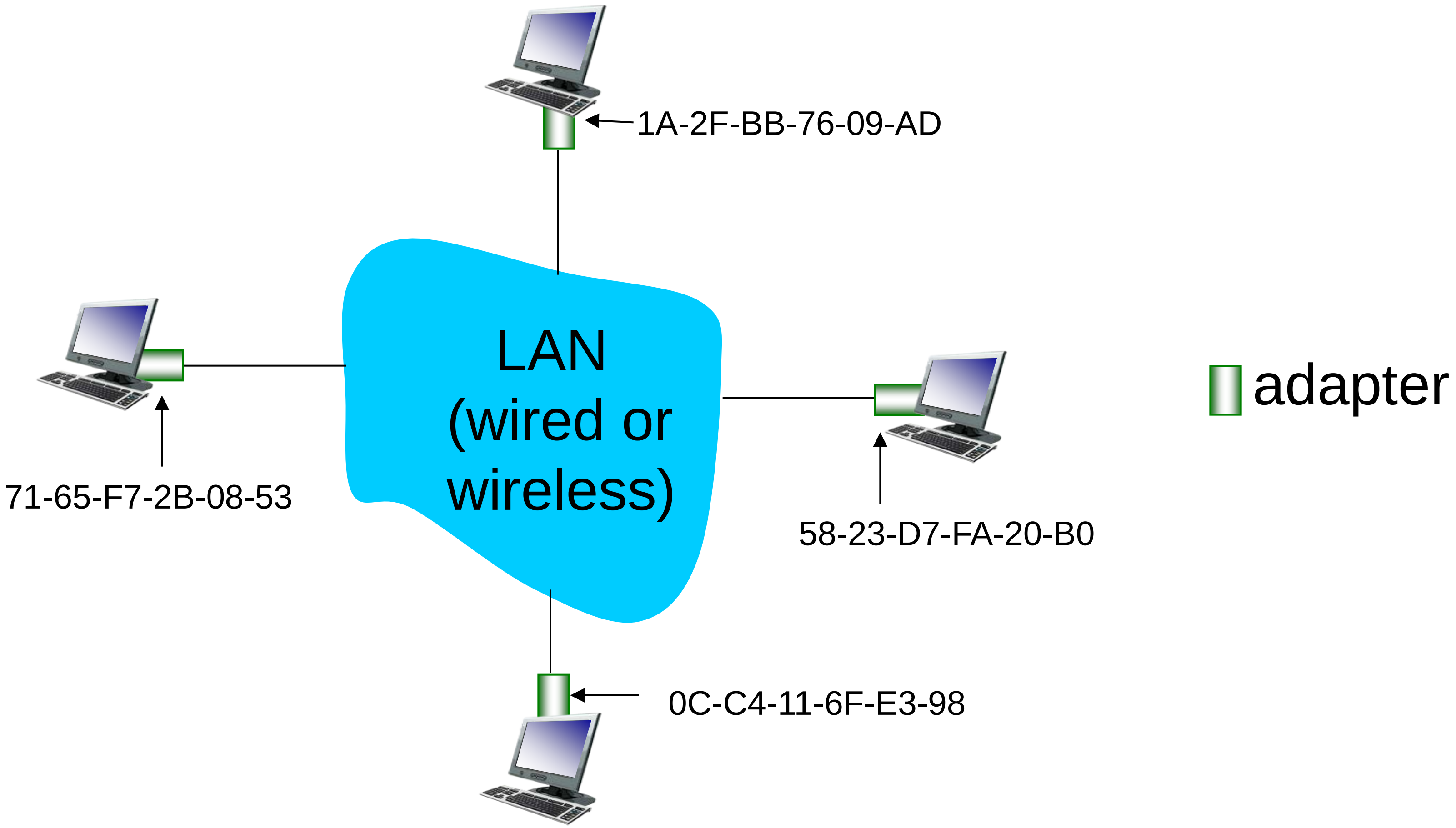
MAC address (1/3) : Kurose Chap6

- 32-bit IP address (IPv4):
  - network-layer address (subnet addr + host addr)
  - used for layer 3 forwarding
- MAC (or LAN or physical) address:
  - used ‘locally” to forward frame, in access network
  - 48 bit MAC address (for most LANs) burned in NIC ROM, also sometimes software settable
  - e.g.: 1A-2F-BB-76-09-AD

    hexadecimal notation

MAC address (2/3)

Each adapter on LAN has a unique MAC address

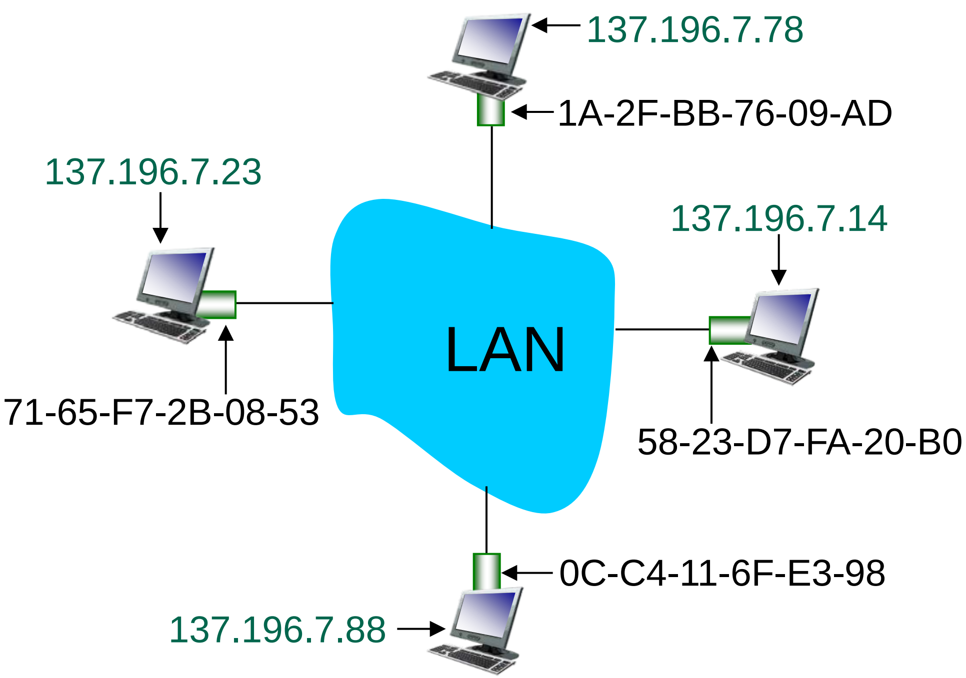


MAC address (3/3)

- MAC address allocation is administered by IEEE
- manufacturer buys portion of MAC address space (to assure uniqueness)
- MAC address has a flat structure ➜ portability
  - can move LAN card from one LAN to another
- IP hierarchical address *not* portable
  - address depends on IP subnet to which node is attached

ARP: address resolution protocol

How to determine interface's MAC address, when knowing its IP address?



ARP table:

Each IP node (host, router) on LAN has table

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

ARP protocol: same LAN

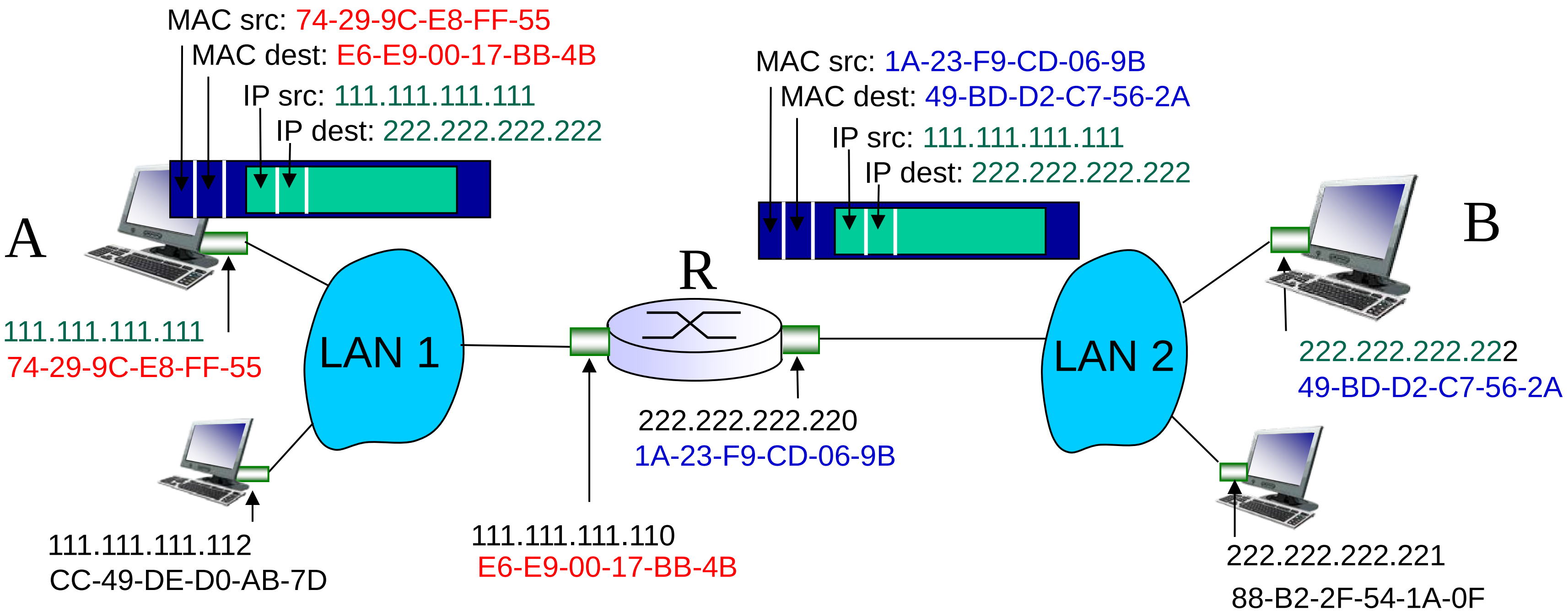
- A wants to send datagram to B (A knows IP address of B)
  - B's MAC address is not in A's ARP table (A does not know MAC add ress of B).
- A broadcasts ARP query packet, containing B's IP address
  - destination MAC address = FF-FF-FF-FF-FF-FF (broadcast addr)
  - all nodes on LAN receive the ARP query
- B receives the ARP packet, replies to A with its (B's) MAC address
  - frame sent to A's MAC address (unicast)
- A caches (saves) an IP-to-MAC address pair in its ARP table until information becomes old (times out)
- ARP is "plug-and-play":
  - nodes create their ARP tables *without intervention from net administrator*

Addressing: routing to another LAN (1/5)

walkthrough: send datagram from A to B via R

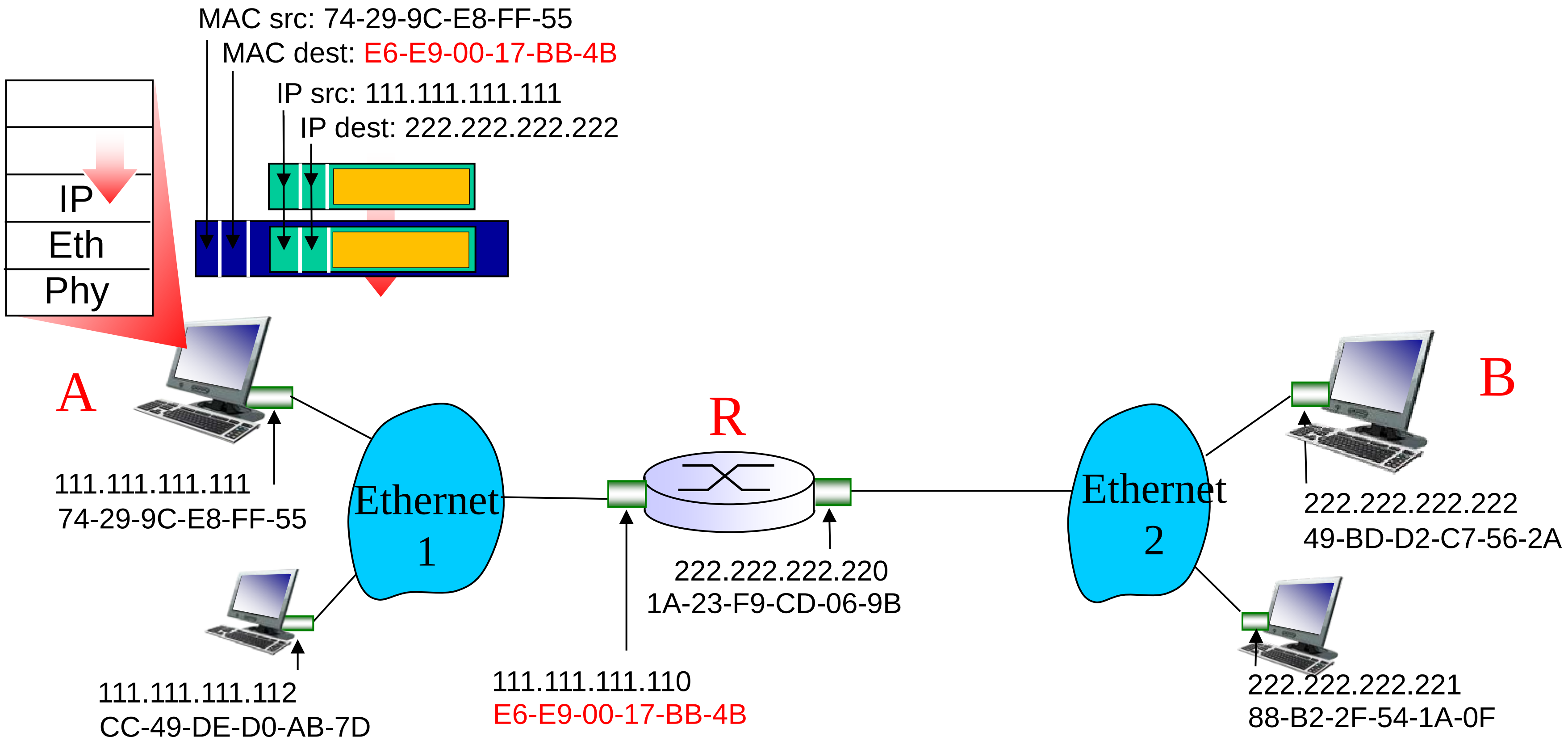
- focus on addressing – at IP (datagram) and MAC layer (frame)
- assume A knows B's IP address
- assume A knows MAC address of first hop router R, or R's IP addr

ARP



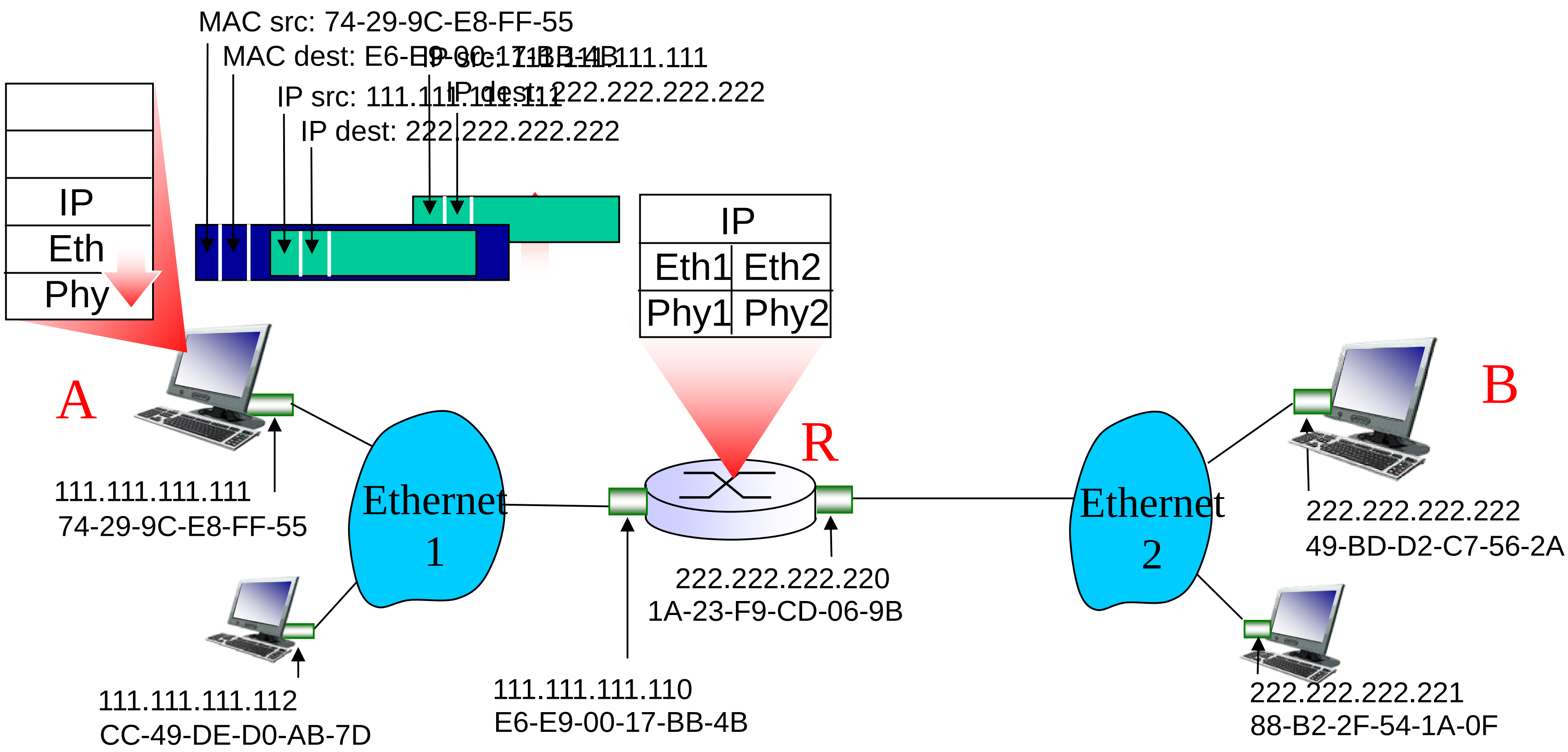
Addressing: routing to another LAN (2/5)

- A creates an IP datagram with IP source A, destination B
- Routing decision: router R via Ethernet-1
- A creates a link-layer frame with R's MAC address as destination address, containing the A-to-B IP datagram



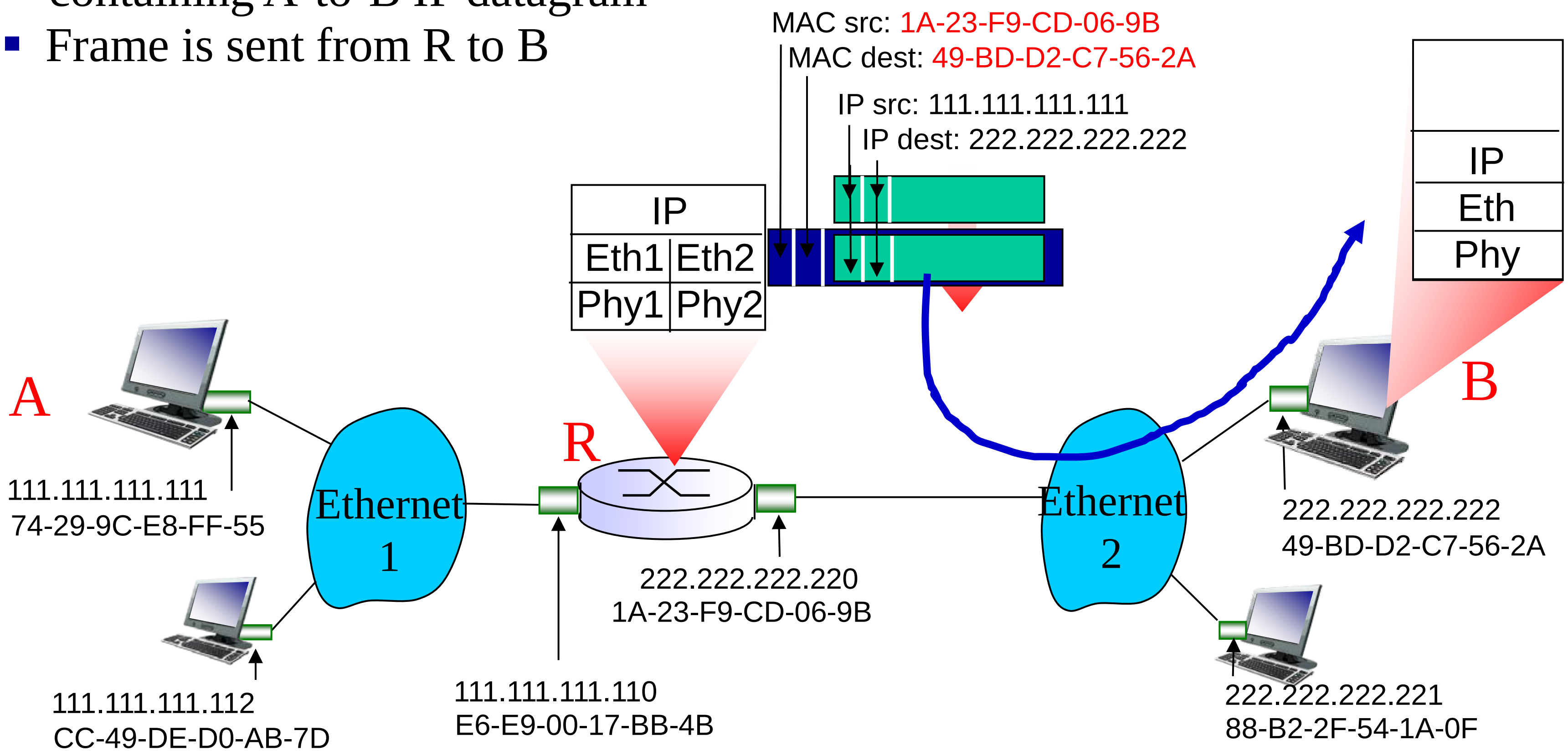
Addressing: routing to another LAN (3/5)

- frame is sent from A to R; the frame is received at R (destination at LAN)
- Datagram is de-capsulated and passed up to IP
- Routing Decision: Destination B via Ethernet-2



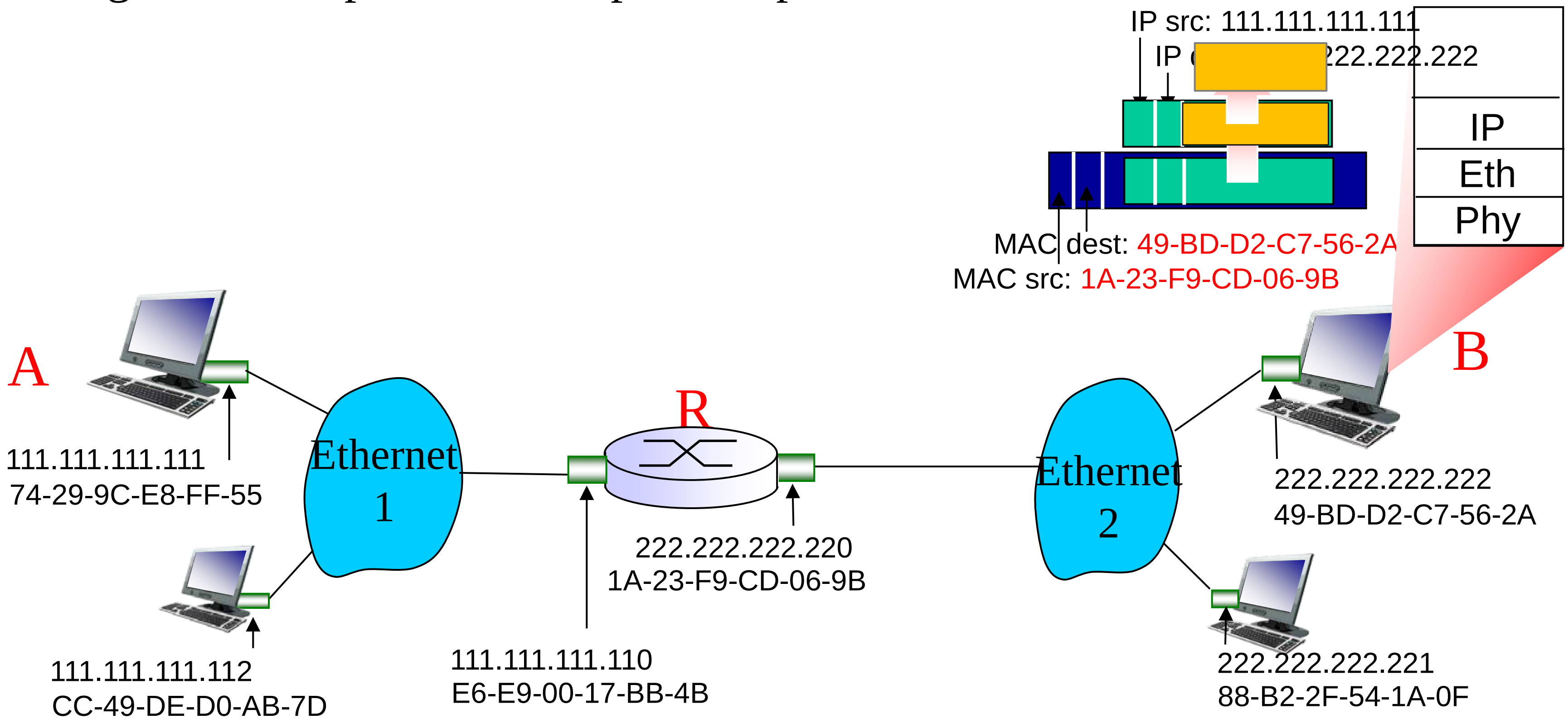
Addressing: routing to another LAN (4/5)

- R passes the IP datagram to Ethernet-2 link layer
- R creates a link-layer frame with
  its MAC address as source addr, B's MAC address as destination addr, containing A-to-B IP datagram
- Frame is sent from R to B

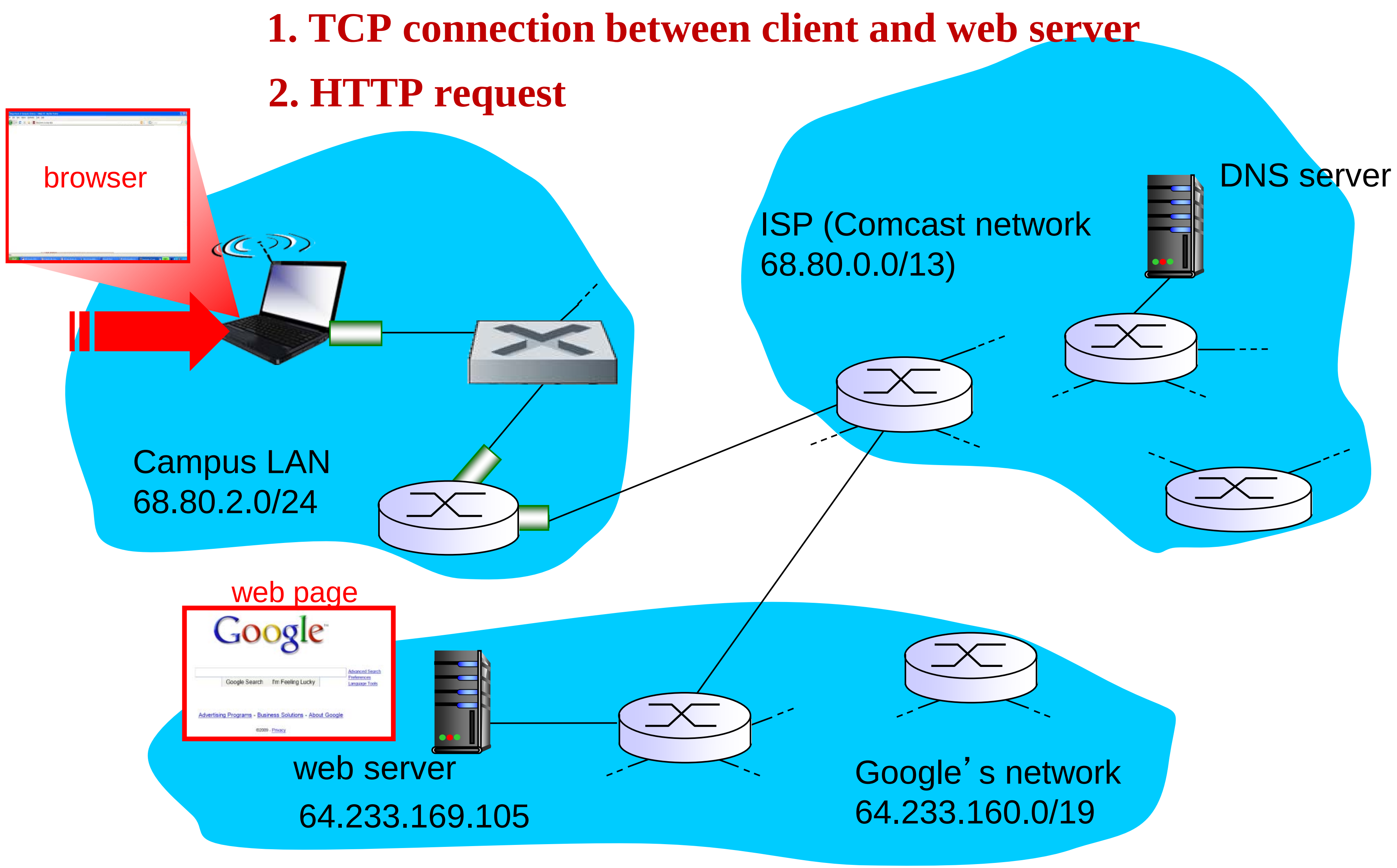


Addressing: routing to another LAN (5/5)

- The frame is received at B (destination MAC addr)
- Datagram is de-capsulated and passed up to IP
- IP at B identifies its IP addr (Destination host)
- Segment de-capsulated and passed up to TCP/UDP



Example Scenario: A student attaches laptop to campus network and requests/receives www.google.com



Additional Operations for this work

- This connecting laptop needs to get its own IP address, IP address of first-hop router, IP address of DNS server
  – access the DHCP server

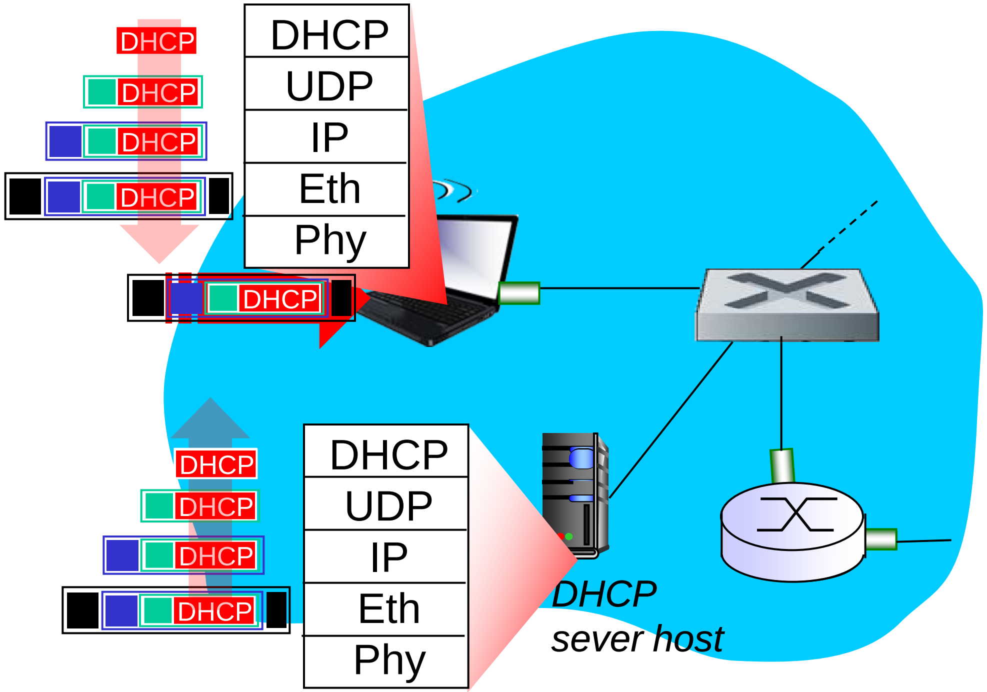
- Before sending HTTP request, this connecting laptop should know the IP address of www.google.com
  – access the DNS server

- Since the DNS server locates outside, this connecting laptop should get the MAC address of first-hop router
  – make ARP query containing the IP address of first-hop router

1. DHCP
2. DNS (ARP) DNS
3. TCP connection
4. Http request

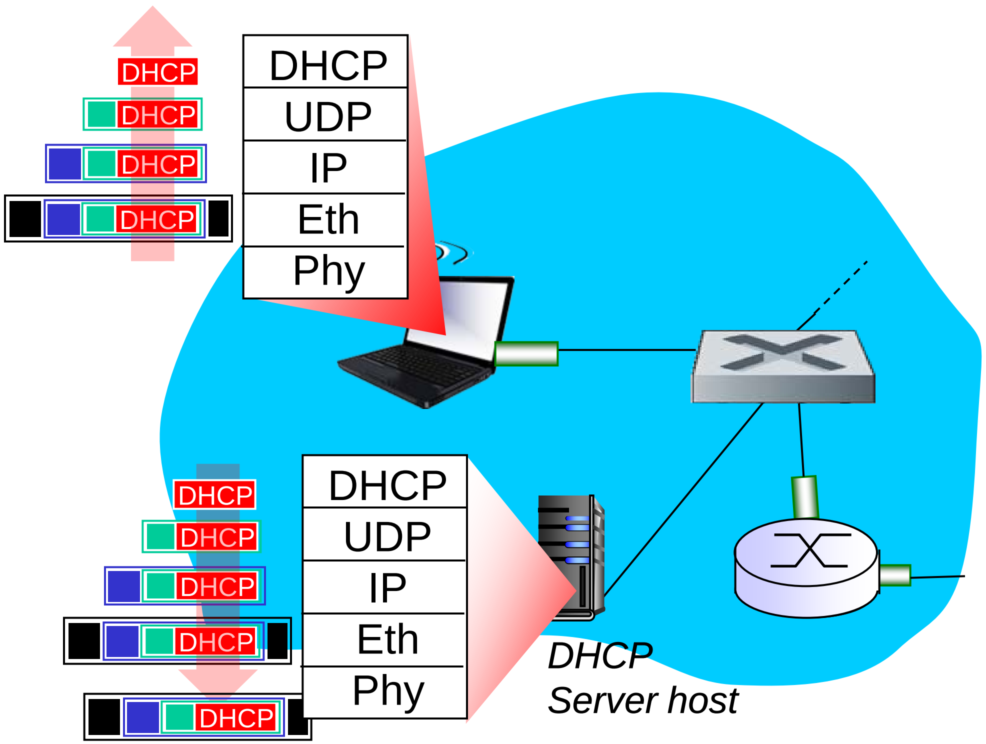
1. Connecting to the Internet (1/2)

- Connecting laptop gets its own IP address, IP addr of first-hop router, IP addr of DNS server, by using DHCP



- DHCP request is encapsulated in UDP, encapsulated in IP, encapsulated in 802.3 Ethernet
- Ethernet frame broadcasts (dest : FFFFFFFFFFFF) on LAN, is received at DHCP server
- The frame is transferred to DHCP server process via IP and UDP

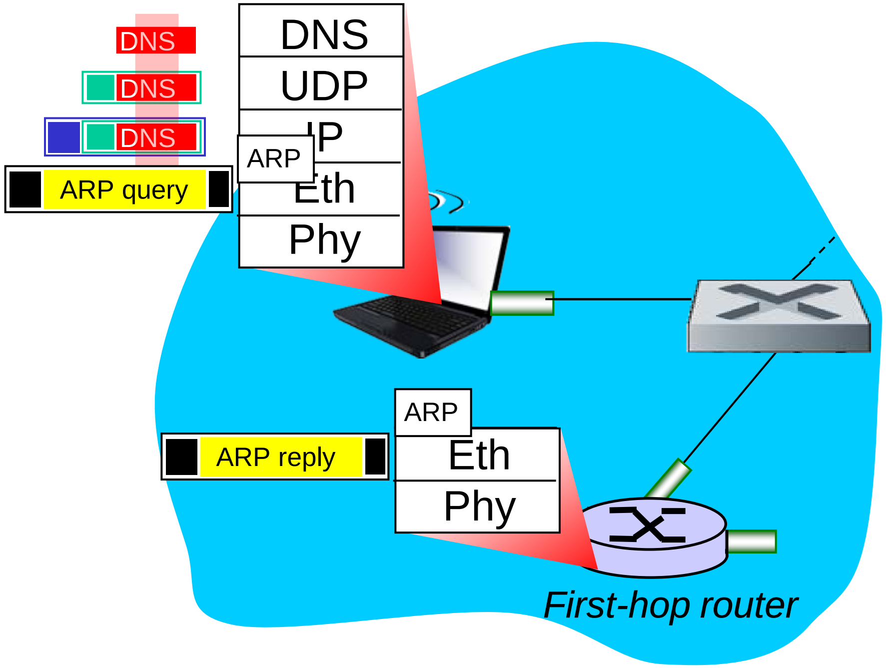
1. Connecting to the Internet (2/2)



- DHCP server process formulates DHCP ACK containing client's IP address, IP address of first-hop router for client, name & IP address of DNS server host
- encapsulation at DHCP server host, frame forwarded (Ethernet switch: address learning) through LAN to client host, finally delivered to DHCP client process
- DHCP client process receives DHCP ACK

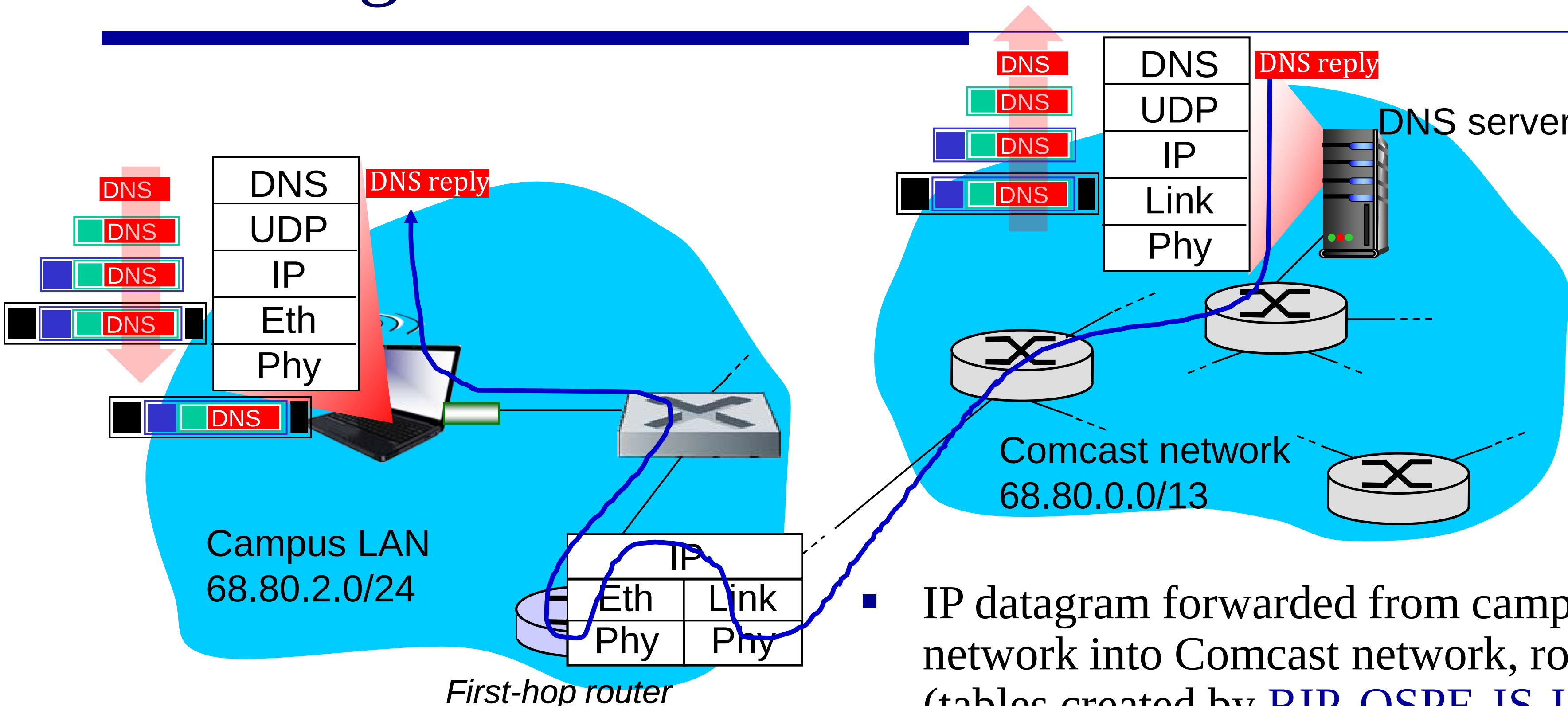
Client now has IP address, knows name & addr of DNS server, IP address of its first-hop router

2. ARP (before DNS, before HTTP)



- Before sending HTTP request, need IP address of www.google.com: DNS
- DNS query created, encapsulated in UDP and encapsulated in IP. To send frame to router, need MAC address of router interface: ARP
- ARP query broadcast, received by router, which replies with ARP reply giving MAC address of router interface
- client now knows MAC address of first hop router, so can now send frame containing DNS query

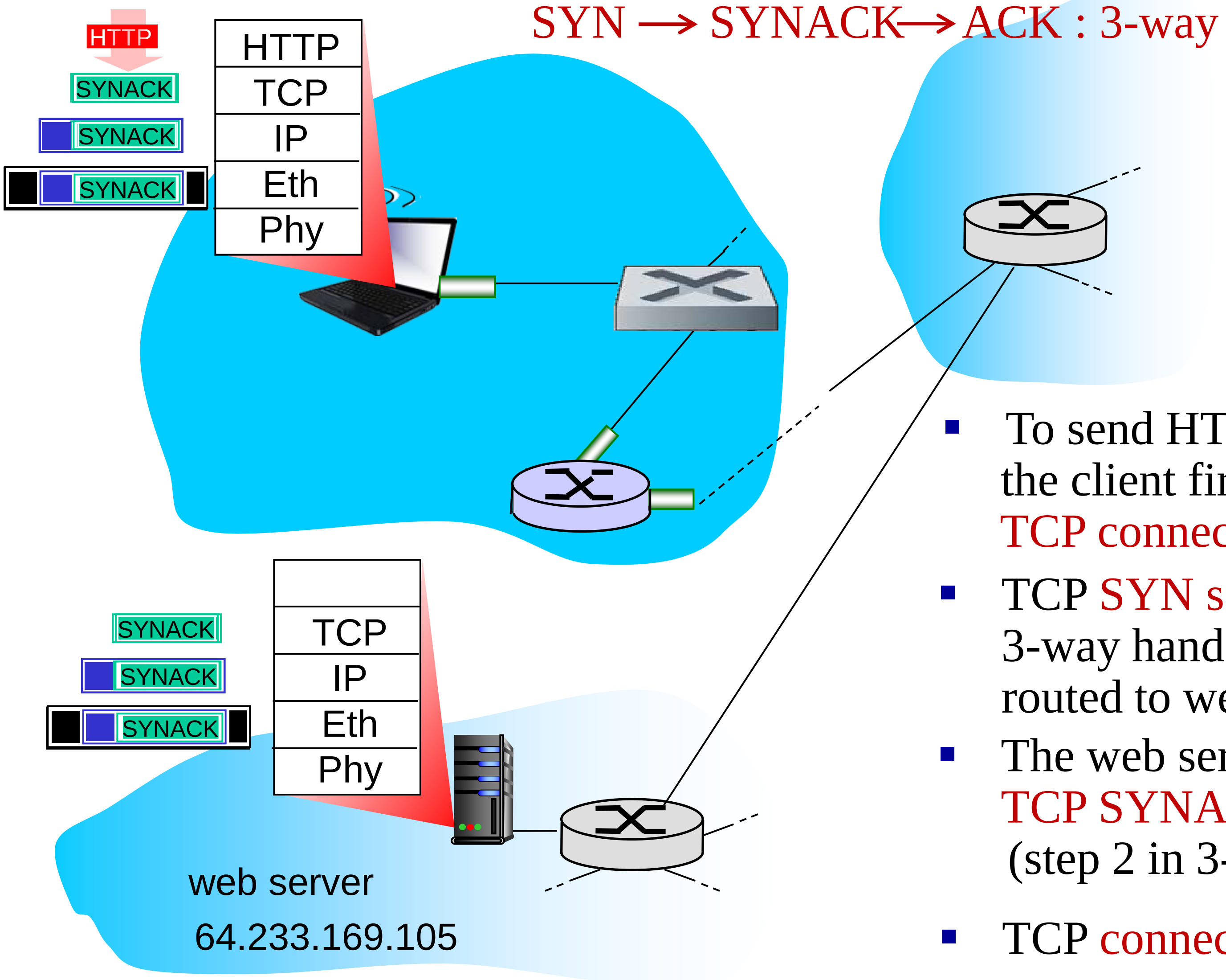
3. Using DNS



- IP datagram containing DNS query forwarded via LAN switch from client to 1st hop router
- IP datagram forwarded from campus network into Comcast network, routed (tables created by RIP, OSPF, IS-IS and/or BGP routing protocols) to DNS server
- delivered to DNS server
- DNS server process replies to client with IP address of www.google.com

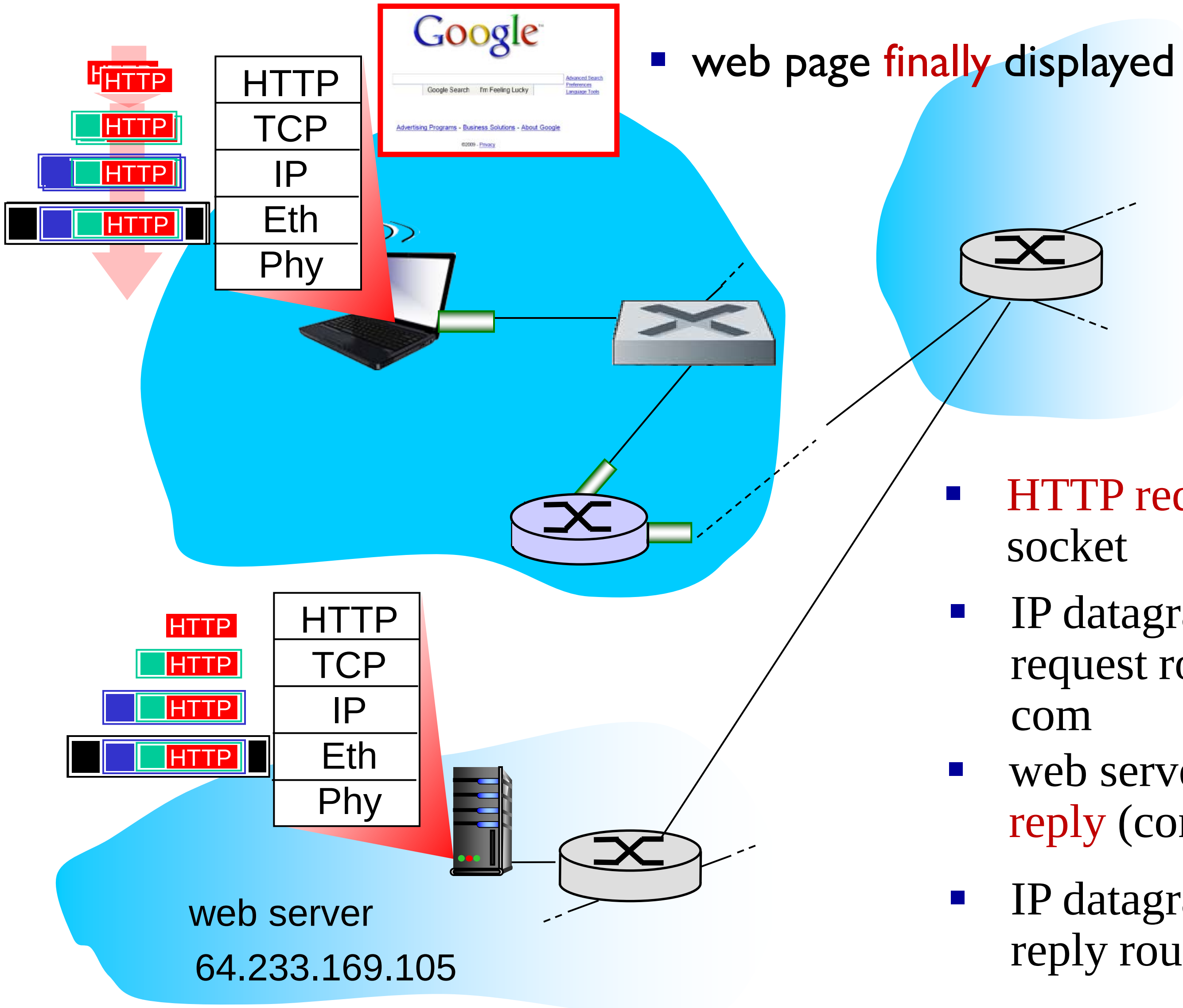
4. TCP connection carrying HTTP

SYN → SYNACK → ACK : 3-way handshaking



- To send HTTP request, the client first establishes TCP connection to web server
- TCP SYN segment (step 1 in 3-way handshake) inter-domain routed to web server
- The web server responds with TCP SYNACK (step 2 in 3-way handshake)
- TCP connection established!

5. HTTP request/reply



- web page finally displayed

- HTTP request sent into TCP socket
- IP datagram containing HTTP request routed to www.google.com
- web server responds with HTTP reply (containing web page)
- IP datagram containing HTTP reply routed back to client