Ethernet, Fast Ethernet, Switching, Routing Options

4 December 2024
Lecture 5

Some Slides Credits: Steve Zdancewic (UPenn)

Topics for Today

- Ethernet
- Fast Ethernet
  - 100 Mbps
  - 1 Gbps
- Switching
  - Layer 2 (Link)
  - Layer 3 (Network)
    - Datagram
    - Virtual Circuit

- Source: PD 2.8, 3.2, 4.1

Open Systems Interconnection (OSI)

End Host	Reference model – not actual implementation.
Application	Transmits *messages* (e.g. FTP or HTTP)
Presentation	Data format issues (e.g. big- vs. little-endian)
Session	Manages multiple streams of data
Transport	Process to process protocols
Network	Routes *packets* among nodes in network
Data Link	Packages bit streams into *frames*
Physical	Transmits raw bits over link

IEEE 802 network standards

The IEEE 802 committee produces standards & specifications for Local Area Networks (LAN):

- **802.3 CSMA/CD Networks (Ethernet)**
- 802.4 Token Bus Networks
- 802.5 Token Ring Networks
- 802.6 Metropolitan Area Networks
- **802.11 Wireless LAN (Wifi)**

Ethernet (802.3)

- A standard for local area networks (LAN)

- Developed in mid-70's at Xerox PARC
  - Descendent of Aloha, a U. of Hawaii radio packet network
  - DEC, Intel, and Xerox standard: 1978 for 10Mbps
  - IEEE 802.3 standard grew out of that

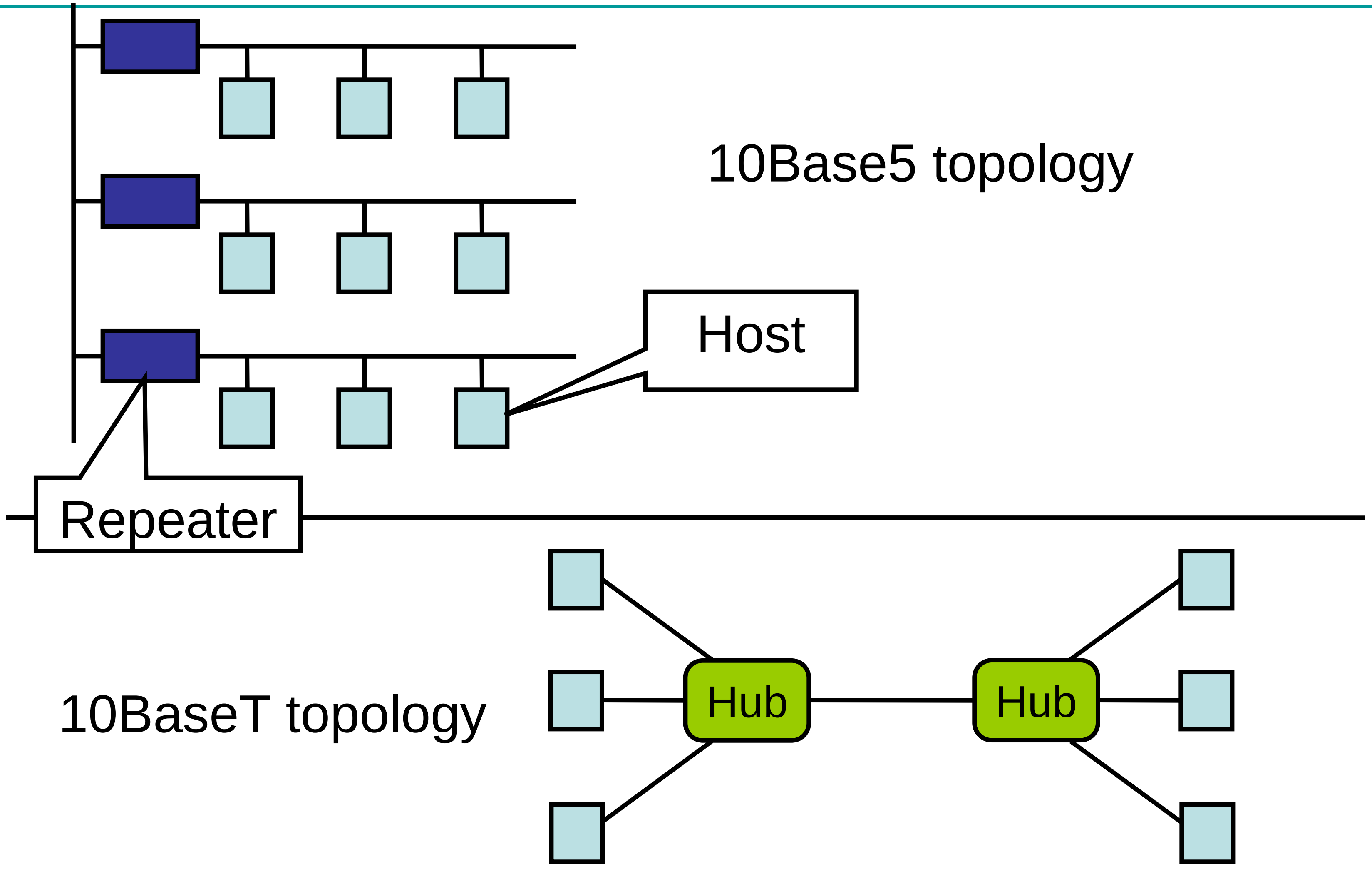
- Physical implementations:
  - 10Base5, 10Base2, 10BaseT, 10BaseF, 100BaseT, 1000BaseT…
  - Speed: 10Mbps,100Mbps, 1000Mbps, …

Name	Cable	Max Seg. Length	Nodes per Seg.	Advantages
10Base5	Thick coaxial	500 m	100	Original cable; obsolete
10Base2	Thin coaxial	185 m	30	No hub needed
10Base-T	Twisted pair	100 m	1024	Cheapest system
10Base-F	Fiber optics	2000 m	1024	Best between buildings, secure

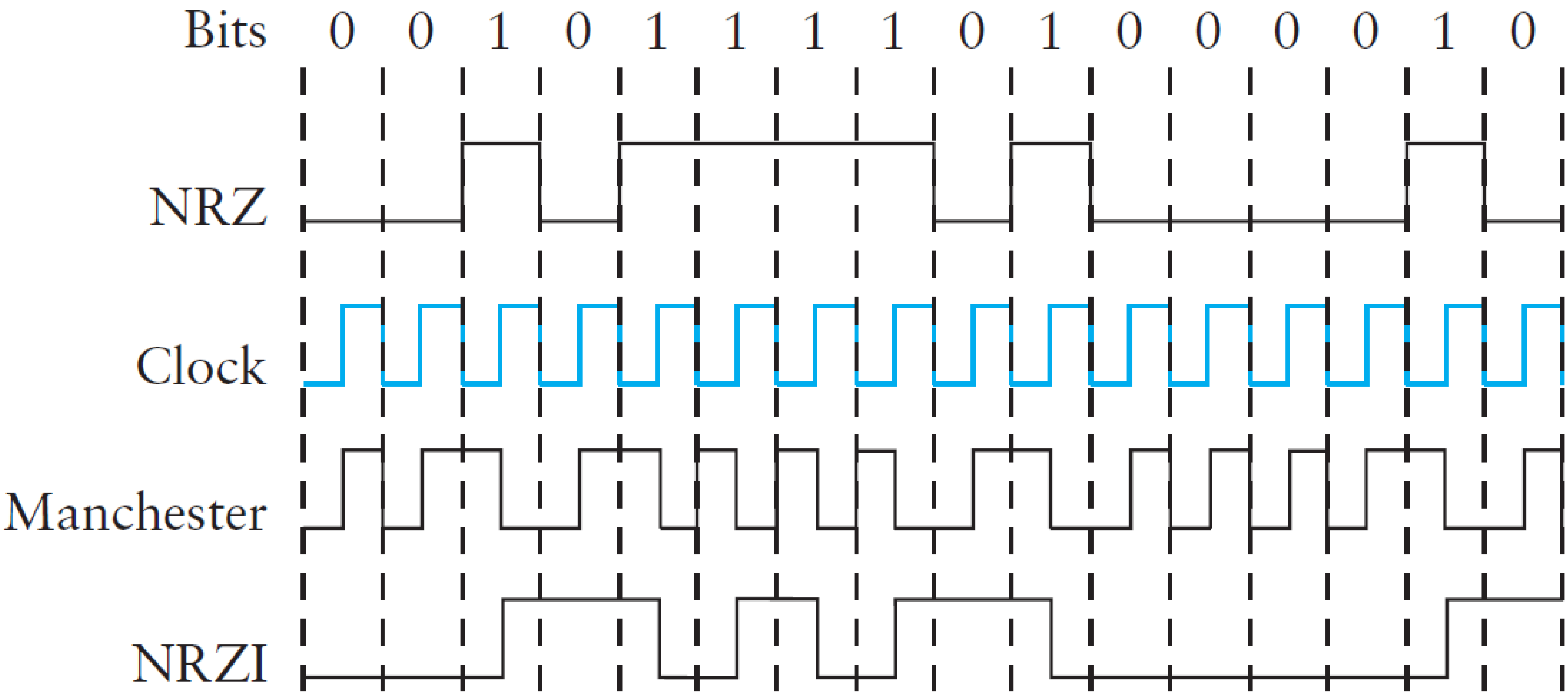
Ethernet Physical links

- Originally used “Thick-net” 10Base5
  - $10 = 10Mbps$
  - 5 = maximum of 500 meters segments
  - Up to 4 repeaters between two hosts $= 2500m$ max
- More common: 10BaseT
  - $10 = 10Mbps$
  - T = Twisted pair (typically Category 5), Maximum of 100 meter segments
  - Connected via *hubs* (still $2500m$ max)
- Today’s standards: 100BaseT, 1000BaseT

Ethernet topologies



Ethernet Encoding: Machester



Ethernet basics

The Ethernet link is *shared*

- Signal transmitted by one host reaches *all* hosts

Method of operation: **CSMA/CD**

- Carrier Sense, Multiple Access, with Collision Detection

Hosts competing for the same link are said to be in the same **collision domain**

Good news: easy to exchange data

Bad news: must regulate link access

Protocol: **Media Access Control (MAC)**

Ethernet Addresses

- Every adapter manufactured has a unique address

6 bytes (48 bits) usually written in hexadecimal

Examples

- `00-40-50-B1-39-69`
- `8:0:2b:e4:b1:2`

Each manufacturer is assigned 24bit prefix

Manufacturer ensures unique suffixes

- https://www.wireshark.org/tools/oui-lookup.html

Ethernet Frame Format

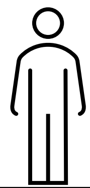
64	48	48	16		32
Preamble	Dest	Src	Type	Body	CRC

- Preamble – repeating pattern of 0's & 1's
  - Used by receiver to synchronize on signal
  - In Manchester this is $10MHz$ square wave for $6.4\ \mu s$
- Destination and Source – Ethernet Addresses
- Type – demultiplexing key
  - Identifies higher-level protocol
- Body – payload
  - Minimum $46\ Bytes$
  - Maximum $1500\ Bytes$

Addresses in an Ethernet frame

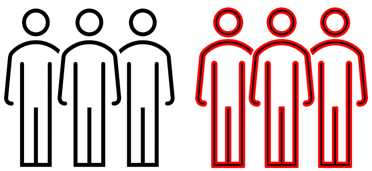
First bit = 0 indicates *unicast address*

- Sent to one receiver



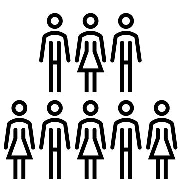
First bit = 1 indicates *multicast address*

- Sent to a group of receivers



All bits = 1 indicates a *broadcast address*

- Sent to all adapters



Ethernet Adapter Algorithm

Always

- Frames addressed to the broadcast address
- Frames addressed to its own address

If programmed to

- Frames sent to multicast addresses

All frames

- If the adapter has been put into *promiscuous mode*

Ethernet Transmitter Algorithm

If the link is idle transmit the frame immediately

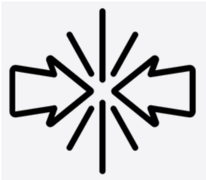
- Upper bound on frame size means adapter can't monopolize link

If the link is busy

- Wait for the line to go idle
- Wait for $9.6\mu s$ after end of last frame (sentinel)
- Transmit the frame

Two (or more) frames may *collide*

- Simultaneously sent frames interfere

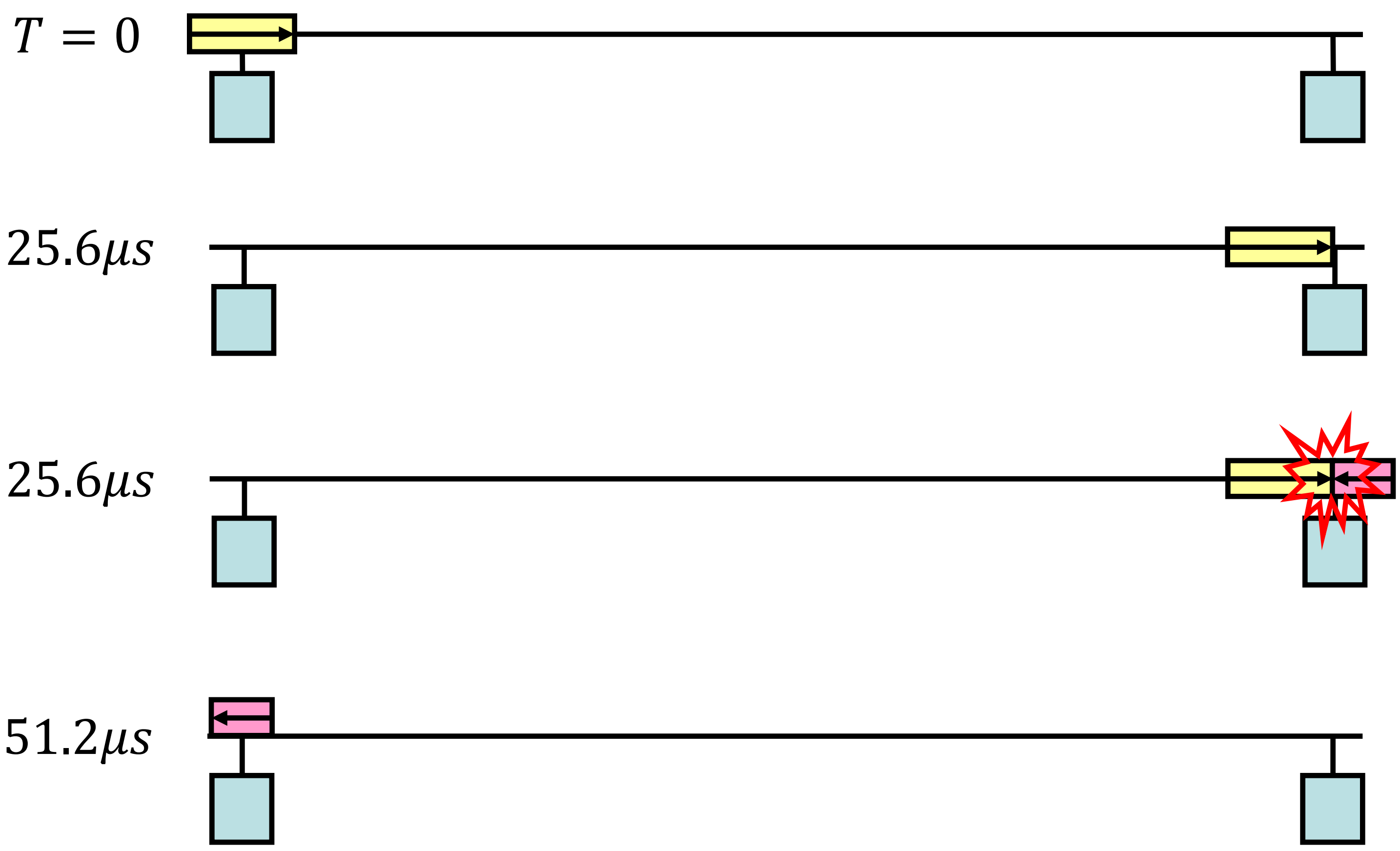


Collision Detection

- When an adapter detects a collision
  - Immediately sends 32-bit *jamming signal* (alternating 1's and 0's)
  - Stops transmitting

- A $10Mbps$ adapter may need to send 512 bits to detect a collision. Why?
  - $2500m + 4$ repeaters gives RTT of $51.2\mu s$
  - $51.2\mu s$ at $10Mbps = 512$ bits
  - Fortunately, minimum frame (excluding preamble) is 512 bits = 64 Bytes
    - 46 Bytes data + 14 Bytes header + 4 Bytes CRC

Ethernet Collision (Worst Case)



Exponential Backoff

- After it detects 1st collision
  - Adapter waits either 0 or $51.2\mu s$ before retrying
  - Selected randomly
- After 2nd failed transmission attempt
  - Adapter randomly waits $0, 51.2, 102.4,$ or $153.6\mu s$
- After nth failed transmission attempt
  - Pick k in 0 ... $2^n - 1$
  - Wait $k \times 51.2\mu s$
  - Give up after 16 retries

    (but cap n at 10)

Ethernet Efficiency

- How efficient is Ethernet? Compare it to “Just send”
- Model: *k* computers, each sends with probability *p*
- Chance one of them gets to send: $p(1-p)^{k-1}$
- Chance **<u>any</u>** one gets to send: $A = kp(1-p)^{k-1}$
- *A* is maximized when $p = \frac{1}{k}$
  - Then $k \to \infty \Rightarrow A \to \frac{1}{e}$
- Chance *j* slots until someone gets to send: $A(1-A)^{j-1}$
- Mean number of slots until some gets to: $\sum_{j=1}^{\infty} jA(1-A)^{j-1} = \frac{1}{A}$
- Since optimal *A* is $\frac{1}{e}$, optimal number of slots: e
- Assume a one-way time of τ
- The slot time is 2τ
- So the average wait time in optimal conditions is: $2\tau e$

Channel Efficiency

- Assume it takes *P* seconds to transmit:

$$\text{Channel efficiency} = \frac{P}{P+\frac{2\tau}{A}}$$

  - In ideal conditions, $2\tau e$

- For frame length F, bandwidth B, cable length L, speed of light c, ideal conditions ($A = \frac{1}{e}$) :
  - Note that $P = \frac{F}{B}$

$$\text{Channel efficiency} = \frac{1}{1+\frac{2BLe}{cF}}$$

What does that mean?

$$\text{Channel efficiency} = \frac{1}{1+\frac{2BLe}{cF}}$$

- Increase the frame size → higher efficiency
- Increase the bandwidth→ lower efficiency
- Increase the length → lower efficiency

Problems?

- We keep increasing the bandwidth (and length)

So Far

- Ethernet
- Fast Ethernet
  - 100 Mbps
  - 1 Gbps
- Switching
  - Layer 2 (Link)
  - Layer 3 (Network)
    - Datagram
    - Virtual Circuit

Fast Ethernet: Problem

In 1995, IEEE reconvened the 802.3 group to improve Ethernet

The Goal: Get Ethernet up to 100 Mbps

But –

- Same frame format
- Same/Similar physical layer
- Same protocol properties (minimum frame size, maximum frame size)

How?

Fast Ethernet: Solution

Fundamental ideas:

- Cut the maximum distance from $2500m$ to about $200m$ (reduce the length to 10%)
- Increase the baud of the line (this got tricky)

The details:

- Three physical media were approved, each with pluses and minuses

Name	Cable	Max. Seg. Length	Advantages
100Base-T4	Twisted pair	100 m	Uses category 3 UTP
100Base-TX	Twisted pair	100 m	Full duplex at 100Mbps (Cat 5 UTP)
100Base-FX	Fiber optics	2000 m	Full duplex at 100Mbps; long runs

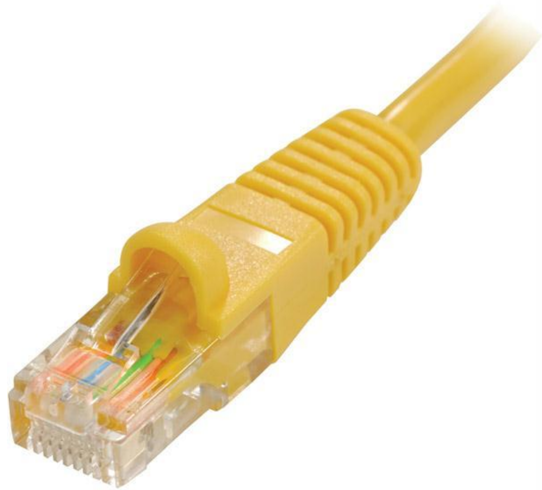
Category 5 UTP

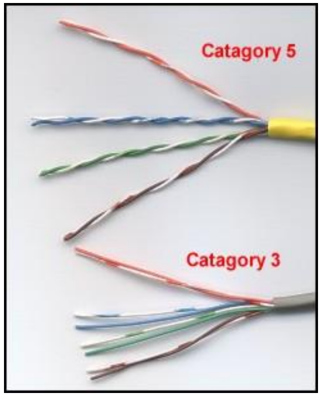
Cat 5 can handle $125\ MHz$ at $100m$

– We don't need the tricks for Cat 3

<u>Solution:</u>

- Forget Manchester (50% efficient) - use 4B/5B instead (80% efficient)
  – 4B/5B only uses some 5 bit combinations for data
  – The rest are forbidden or used for in-band signaling
- Use one line upstream, one downstream
  – $1\ bit \times 100\ MHz = 100\ Mbps$





Fiber Optics

Fiber can easily handle $100\ Mbps$ at full duplex

The length is more than $200m$, so all fiber cables **must go through switches** – no hubs

- Means no collisions, so no CSMA/CD

Gigabit Ethernet: Problem

In 1999, IEEE released Gigabit Ethernet – 1 $Gbps$

Same requirements:

- Similar hardware
- Same packet format
- Same protocol properties (minimum, maximum packet sizes)

How?

Gigabit Ethernet: Solution

In “Normal” Gigabit Ethernet, everything goes via switches

- No more hubs, no more collisions, no more CSMA/CD

What if you want to use hubs?

Non-Solution:

- Reduce maximum distance to $25\ m$

Solution:

- Leave the maximum distance at $200m$
- Raise the minimum packet size to 512 Bytes (!)
  - **Frame Bursting** allows multiple frames to be sent at once if they are less than 512 Bytes.
  - The Ethernet NIC silently pads packets with 0s if they are less than 512 (**Carrier Extension**)
    - What’s the efficiency for an old minimum packet?

Gigabit Ethernet: Cabling

Fiber Optics have special requirements

- LEDs couldn't go on/off at $1ns$, so used lasers
- Maximum distance affected by wavelength and thickness of the fiber
  - Different laser wavelengths:
    - Short wave length 1000Base-SX
    - Long wave length 1000Base-LX
- Uses 8B/10B encoding scheme

Twisted pair options too

- Using Cat 5 cables (all four internal wires are active)
- Cat 5e and Cat 6 cables (more insulation)

Gigabit Ethernet: Flow Control

Gigabit Ethernet is fast:

- In $1ms$, 1953 frames can arrive
- If the processor is busy for a few milliseconds, the buffers may be overrun

Gigabit Ethernet standard uses flow control

- Receiver can send a PAUSE frame to the sender
- Control frames have special type (`0x8808`)
- Pause command tells how many $512\ nsec$ intervals to wait (can make the sender wait up to $33.6\ msec$)

Next generations

- 2002 - 10 Gigabit Ethernet
  - Uses only switches, so no CSMA/CD
- $25/50\ Gbps$ Ethernet now out for data centers
- Power over Ethernet
- 2010 - $40\ Gbps$ and $100\ Gbps$ Ethernet
  - $40\ Gbps$ for use in data centers
  - $100\ Gbps$ for long haul backbone use ($4 \times 25 Gbps$)
- Now: Working on $400\ Gbps$ and $1\ Terabit$ Ethernet
  - Want to preserve Ethernet frame format and frame size
  - Up to 10KM runs

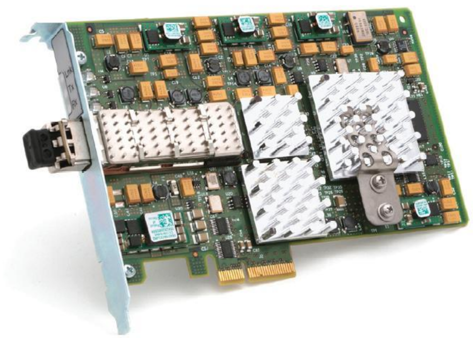
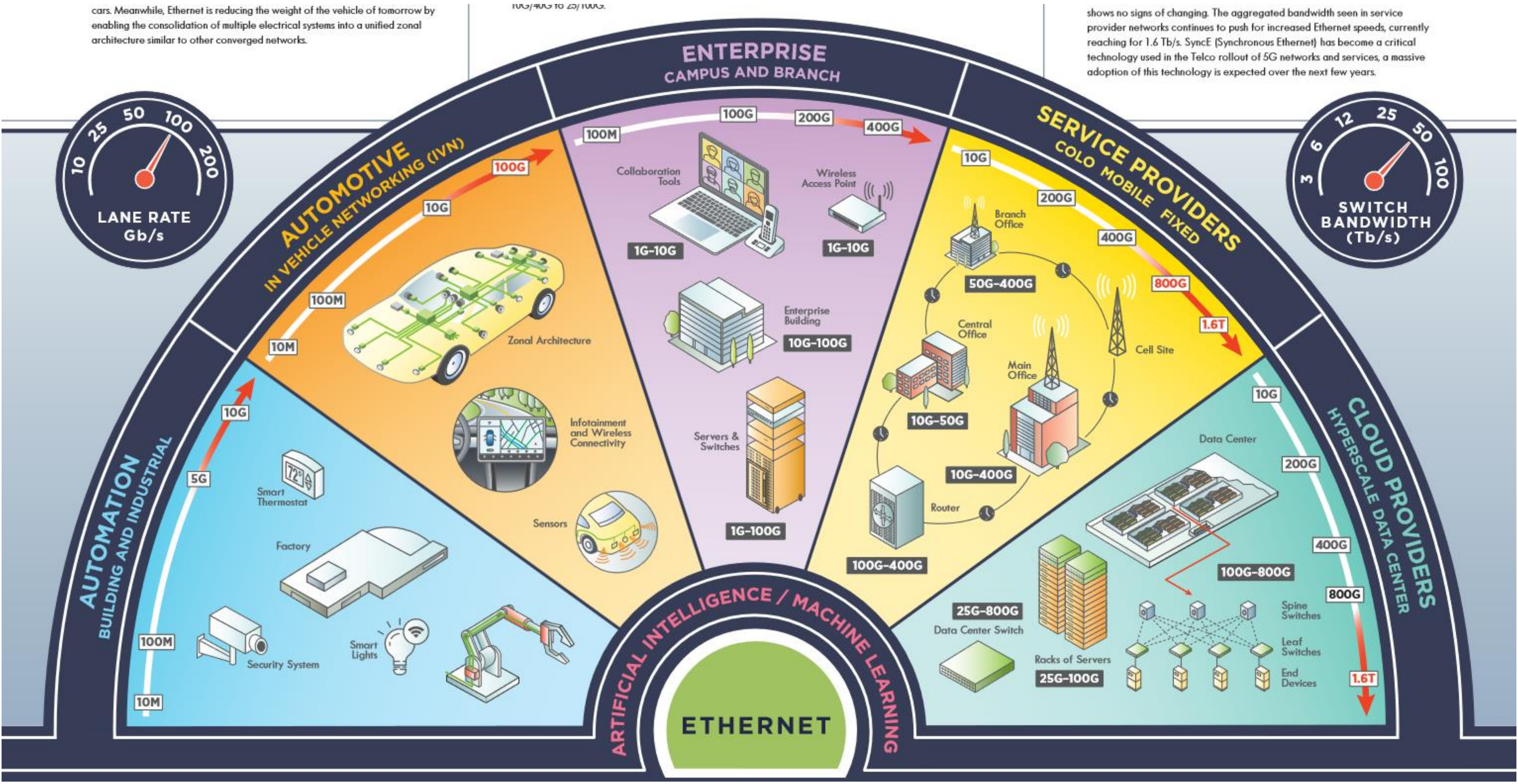


Figure 1. Cisco ASR 9000 Series 2-Port 100 Gigabit Ethernet Line Card

Ethernet Alliance 2024 Roadmap



So Far

- Ethernet
- Fast Ethernet
  - 100 Mbps
  - 1 Gbps
- Switching
  - Layer 2 (Link)
  - Layer 3 (Network)
    - Datagram
    - Virtual Circuit

Switches

- A switch is a box of electronics with $4 - 32$ *cards*
- Each card contains $1 - 8$ ports
  - The cards are connected by a high speed ($Gbps$) backplane
  - Each port connects to one computer (or a hub)

Two kinds of cards:

1. Each card has a buffer
   - Means only one computer can talk on the card at a time
   - Cuts down collision domains to just the ports on the same card
2. Each port has a buffer
   - Means each port has its own send/receive buffer
   - No collisions possible (since each port is independent)

Switched Ethernet

- A *switch* adds
  - Queuing
  - A *routing table*
  - Selective forwarding

- Switches have the following rules when a frame from Y destined for X arrives
  - If X has an entry in the routing table
    - If the frame came from X's *outbound* link, drop it
    - Otherwise, forward it on X's outbound link
  - If X does not have an entry in the routing table
    - Send out the frame on all links ***except the one it arrived on***
  - If Y does not have an entry in the routing table
    - Add the *inbound* link as the *outbound* link for Y

Switch Advantages

- We get:
  - Divided up collision domains (dividing Ethernets)
  - Queuing
  - Adaptation to multiple subnet speeds

- Switches are a bit more expensive than hubs
  - 10/100/1000 switch costs $25-$100
  - 10/100 hub costs is basically impossible to find now to buy

Ethernet Security Issues

Image copyright Sony Pictures Classics and Sony Pictures Entertainment

Promiscuous mode

- *Packet sniffer* detects all Ethernet frames

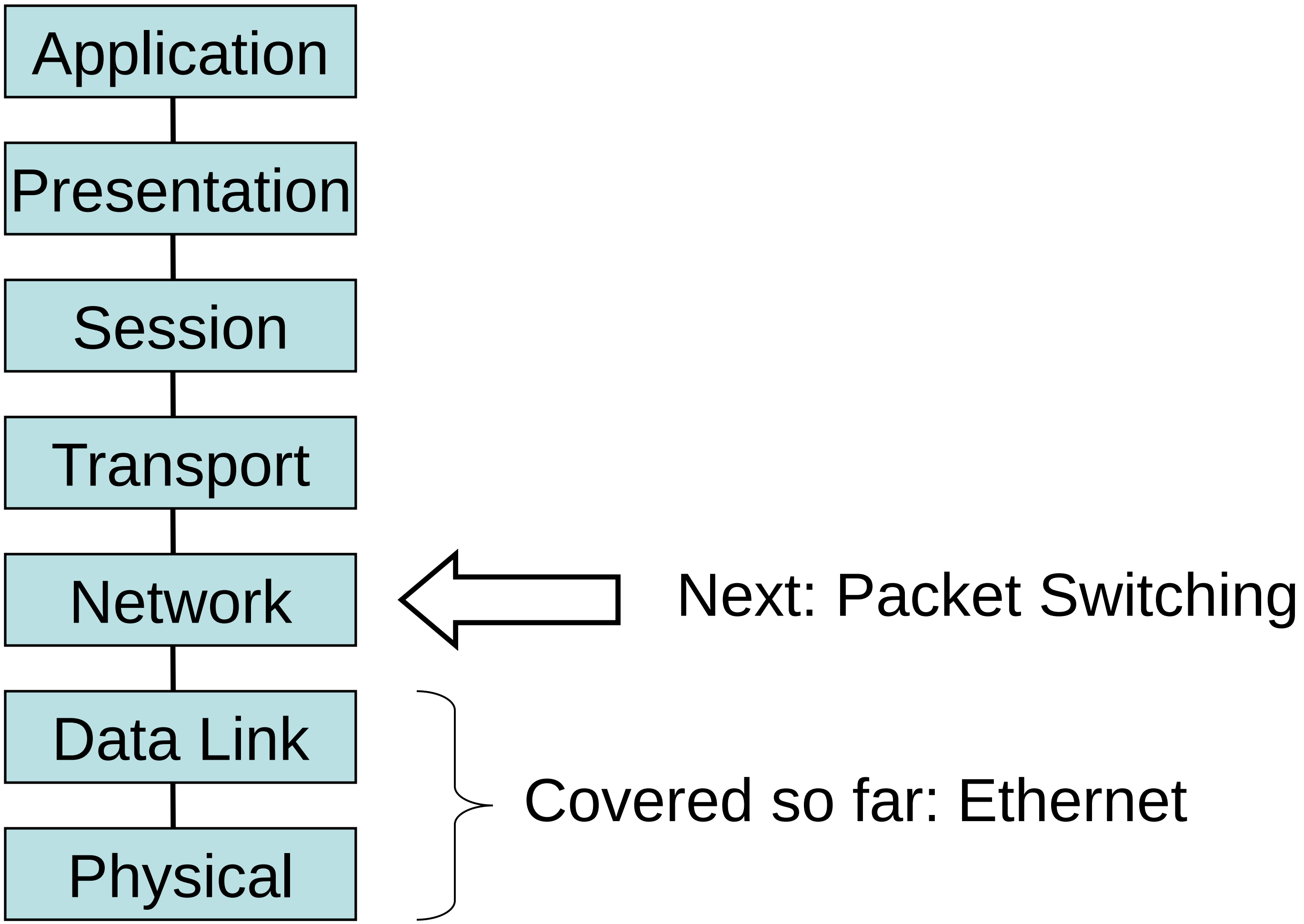
Less of a problem in *switched* Ethernet

- Why?

So Far

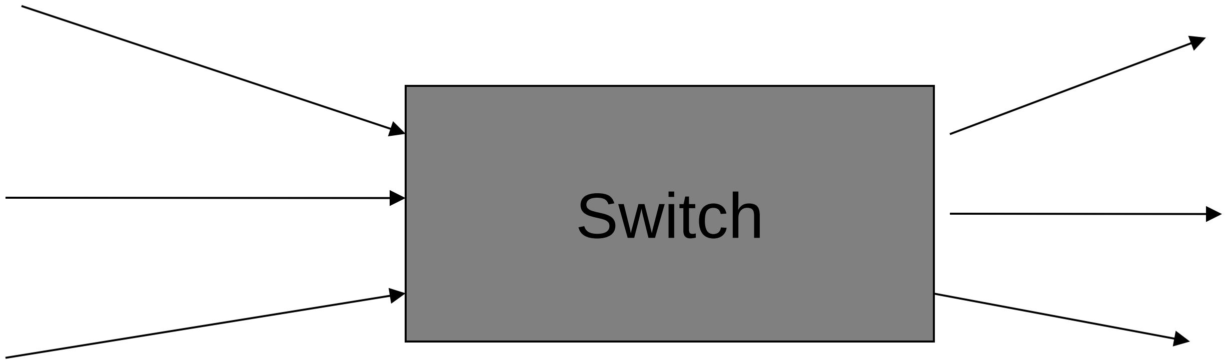
- Fast Ethernet
  - 100 Mbps
  - 1 Gbps
- Switching
  - Layer 2 (Link)
  - Layer 3 (Network)
    - Datagram
    - Virtual Circuit

OSI Reference Model



Packet Switching

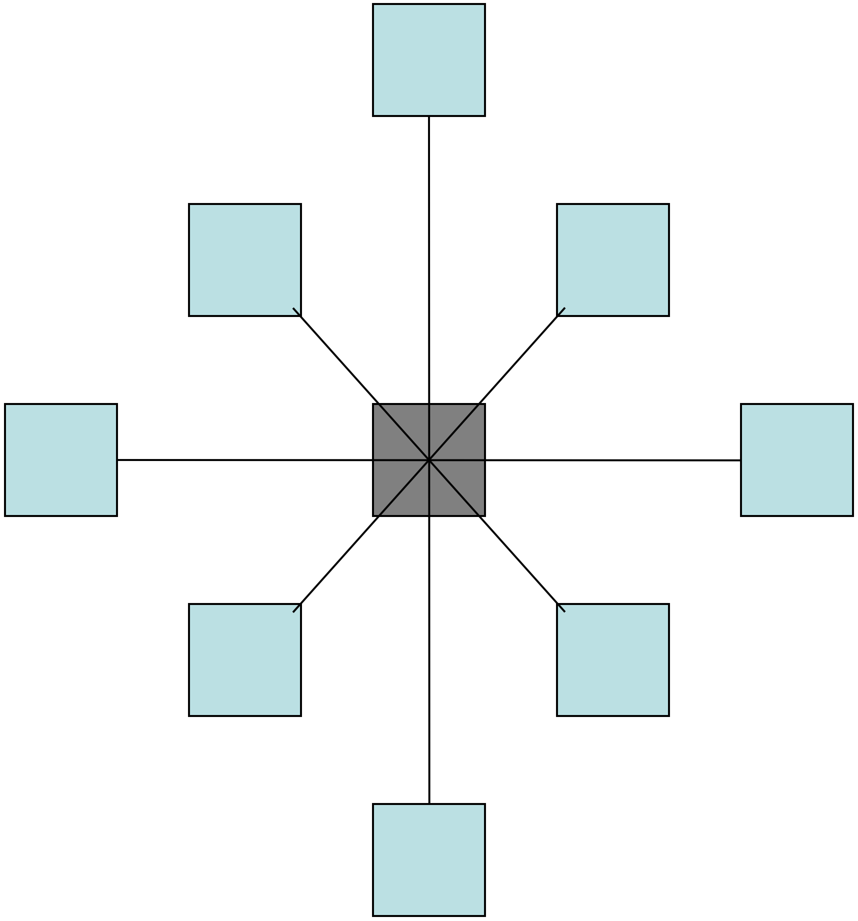
- A *switch*
  - Has many inputs and many outputs
  - Takes packets that arrive on an input and forwards them to the right output



- Key problem: finite output bandwidth

Star Topology

- Scalability
  - Large networks can be built by interconnecting switches.
  - Can connect via high bandwidth point-to-point links = large distances.
  - Adding a new host to a switch doesn't necessarily degrade performance.



Switching Issues

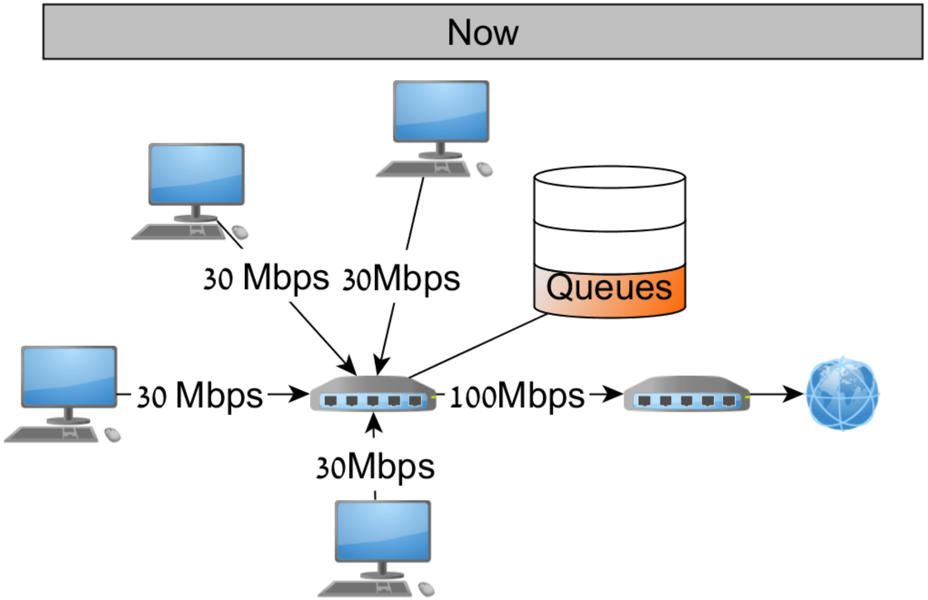
Contention

- Arrival rate of packets going to the same output exceeds output capacity
- Switch buffers packets

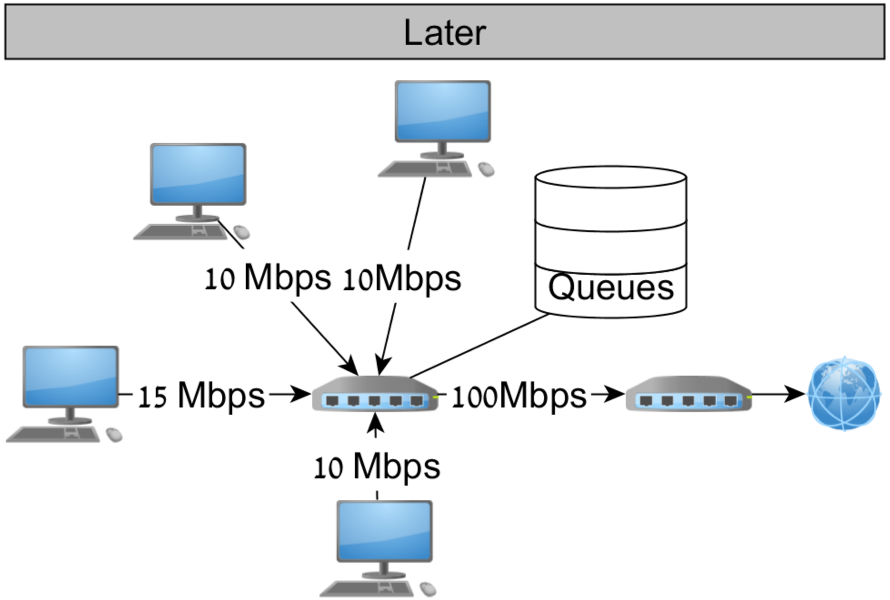
Congestion

- Switch runs out of buffer space
- Forces packets to be dropped

Contention

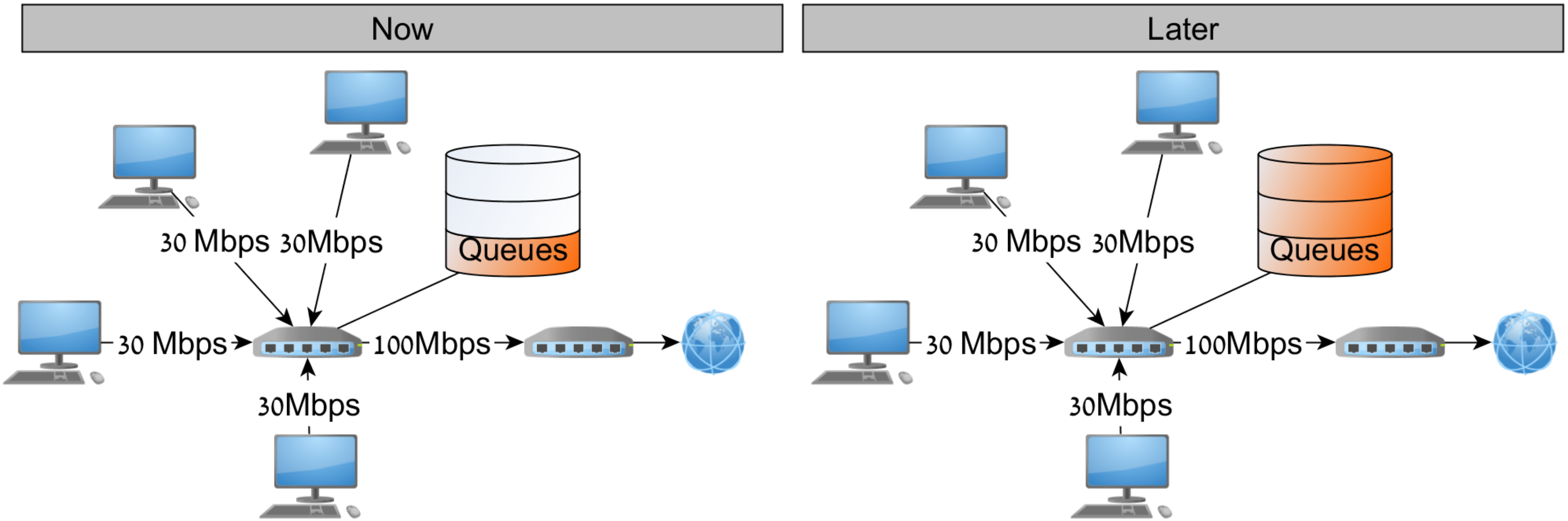


Now
30 Mbps
30Mbps
Queues
30 Mbps
100Mbps
30Mbps



Later
10 Mbps
10Mbps
Queues
15 Mbps
100Mbps
10 Mbps

Congestion



Now
30 Mbps
30Mbps
Queues
30 Mbps
100Mbps
30Mbps
Later
30 Mbps
30Mbps
Queues
30 Mbps
100Mbps
30Mbps

Forwarding Decision

- How does the switch know where to forward a packet?
  – Looks at the packet header to make the decision

- Common approaches

Datagram (or *connectionless*)	Virtual Circuit (or *connection-oriented*)	Source routing
• e.g. IP	• e.g. Frame Relay, ATM	• Less common

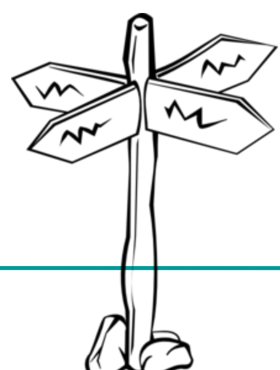
Routing: Drive to Afula

Option 1:

The road authority puts a sign at each intersection to show which way to go

If there are 200 potential destinations, there are 200 arrows at each intersection

At each intersection I must search through 200 arrows to find the one for Afula.

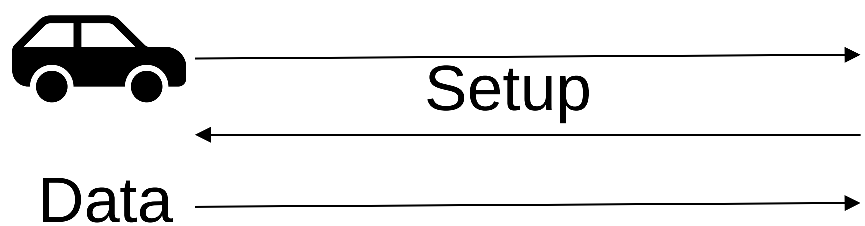


Option 2:

First send someone who knows the way. He puts a sign at each intersection with the name “Michael” and the direction to travel.

There only have to be as many signs at each intersection as there are travelers.

Route setup costs one round trip



Drive Kinneret to Afula

Problem: What if more than one person named “Michael” need to travel through a given intersection?

Solution:

Ensure that each person has a unique name at each intersection (doesn't need to be global).

At each intersection put three pieces of information: (a) name, (b) direction, (c) name for the next intersection.

Example:

At צומת צמח: Michael, go straight. Next name: Noah

At צומת כנרת : Noah, turn left. Next name: Otto.

At צומת אלומות: Otto, go straight. Next name: Paul.

At צומת יבנאל : Paul, go right. Next name: Quentin.

Datagram approach

Every packet contains a complete destination address

- Enough information so that any switch can decide where the packet should go.

Features of datagram approach

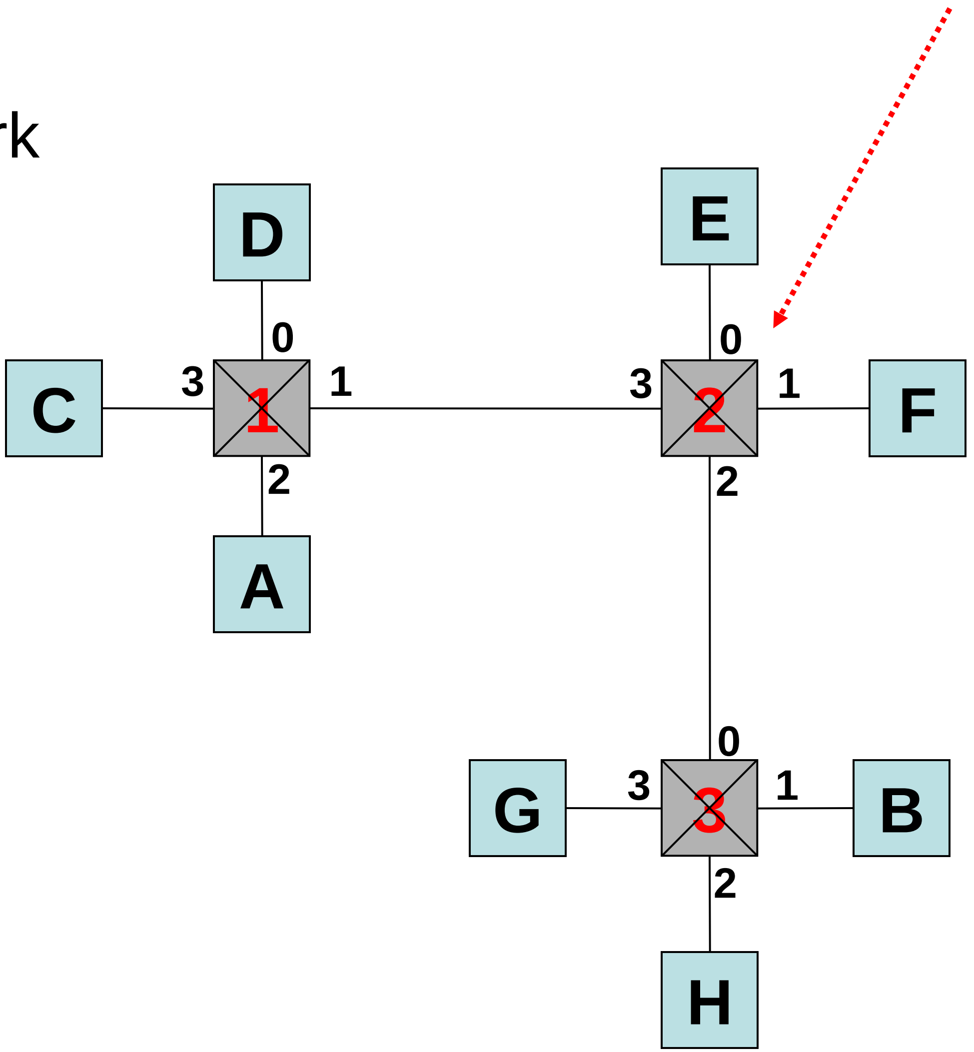
- Packets can be sent *anywhere at any time*
- Sender doesn't know if network can deliver the packet (or if destination host is available)
- Each packet is forwarded independently (two packets may take different routes)
- Possible to route around switch or link failures

Forwarding Tables

- Provide route information
- Easy to determine if network in known (and unchanging)

Forwarding table for switch 2

Dest.	Port
A	3
B	2
C	3
D	3
E	0
F	1
G	2
H	2



Virtual circuit approach

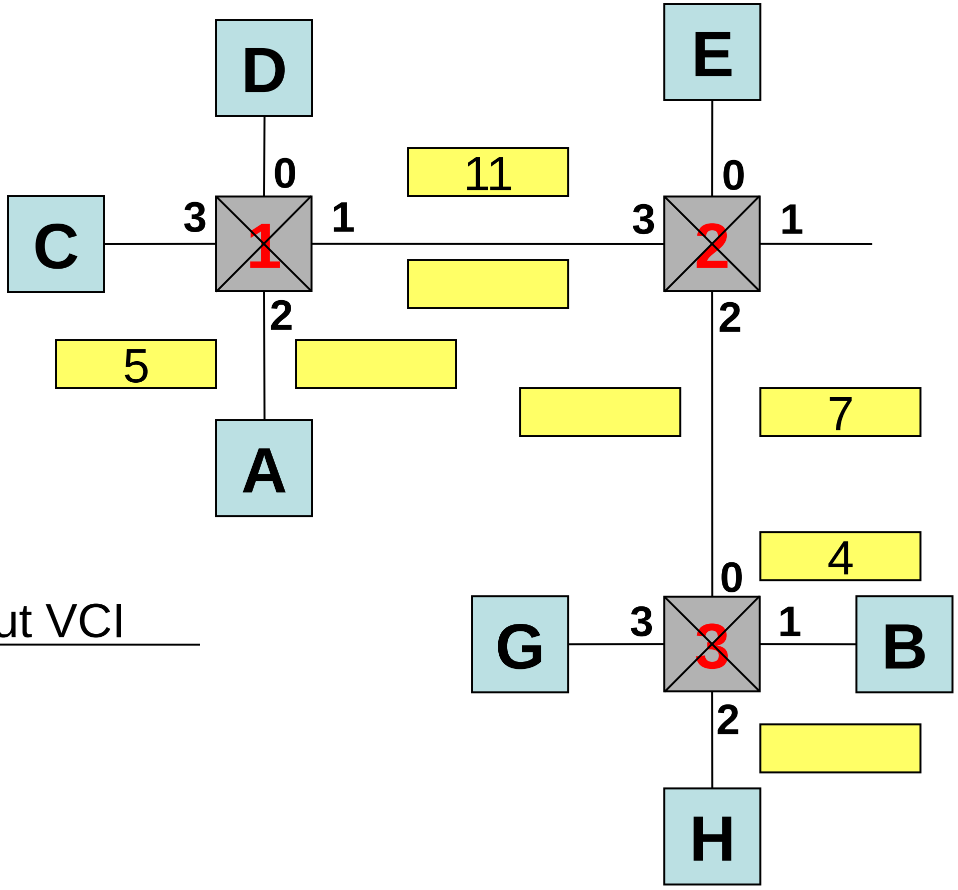
- Set up the connection before data transfer
  – Allocate resources on circuits
  – Set up forwarding tables

- Benefits of virtual circuit approach
  – Performance: per-packet switching cost is low
  – Reliability: predictable latency and throughput
- Drawbacks
  – Setup time is long
    • At least one RTT – why?
  – Fault tolerance
    • What if the circuit fails during the transmission?

Virtual Circuit Switching

- VCI = Virtual Circuit Identifier
- Incoming port + VCI uniquely identify virtual circuit
- Setup phase constructs circuit table entries at each switch

A wants to send to B



Switch	In Port	In VCI	Out port	Out VCI
1				
2				
3				

Conclusion

- Ethernet
- Fast Ethernet
  - 100 Mbps
  - 1 Gbps
- Switching
  - Layer 2 (Link)
  - Layer 3 (Network)
    - Datagram
    - Virtual Circuit