CS252
Graduate Computer Architecture

Lecture 9:
Network 2: Protocols, Routing, Wireless

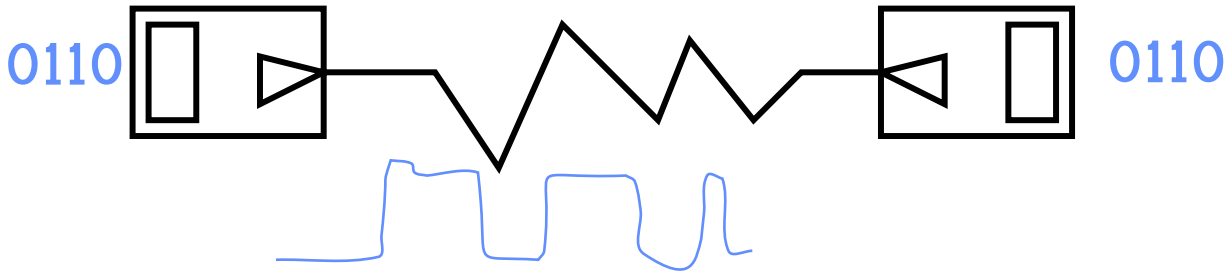
February 14, 2001

Prof. David A. Patterson

Computer Science 252

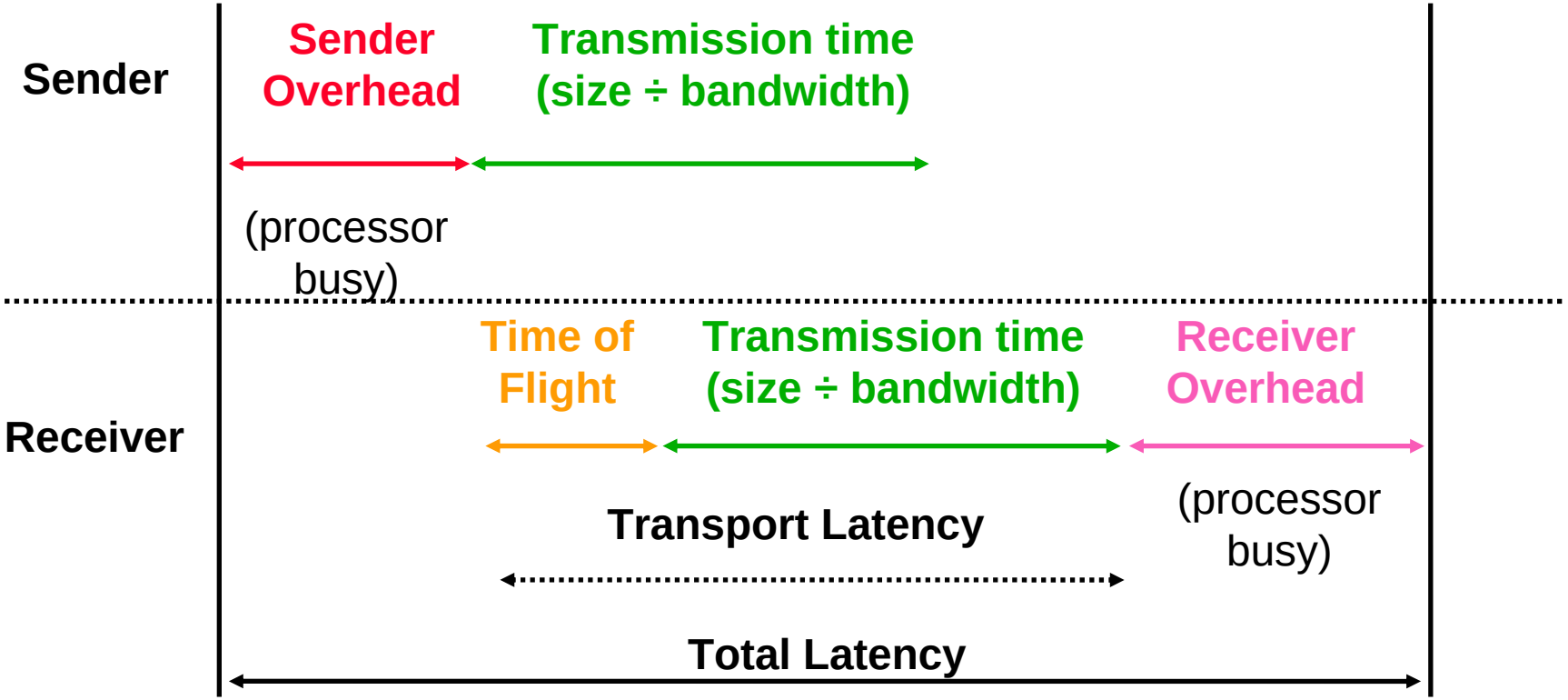
Spring 2001

Review: Network Basics



- Link made of some physical media
  - wire, fiber, air
- with a transmitter (tx) on one end
  - converts digital symbols to analog signals and drives them down the link
- and a receiver (rx) on the other
  - captures analog signals and converts them back to digital signals
- tx+rx called a transceiver

Review: Performance Metrics



Total Latency = Sender Overhead + Time of Flight +
Message Size ÷ BW + Receiver Overhead

Includes header/trailer in BW calculation?

Review: Interconnections

- Communication between computers
- Packets for standards, protocols to cover normal and abnormal events
- Performance issues: HW & SW overhead, interconnect latency, bisection BW
- Media sets cost, distance

Compare Media

- Assume 40 2.5" disks @ 25 GB (1 TB), Move 1 km
- Compare Cat 5 (100 Mbit/s), Multimode fiber (1000 Mbit/s), single mode (5000 Mbit/s), and car
- Cat 5: 1000 x 1024 x 8 Mb / 100 Mb/s = 23 hrs
- MM: 1000 x 1024 x 8 Mb / 1000 Mb/s = 2.3 hrs
- SM: 1000 x 1024 x 8 Mb / 5000 Mb/s = 0.5 hrs
- Car: 5 min + 1 km / 50 kph + 10 min = 0.25 hrs
- Car of disks = high BW media

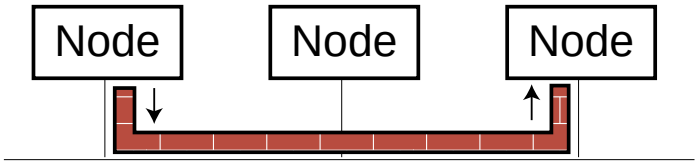
Interconnect Issues

- Performance Measures
- Network Media
- Connecting Multiple Computers

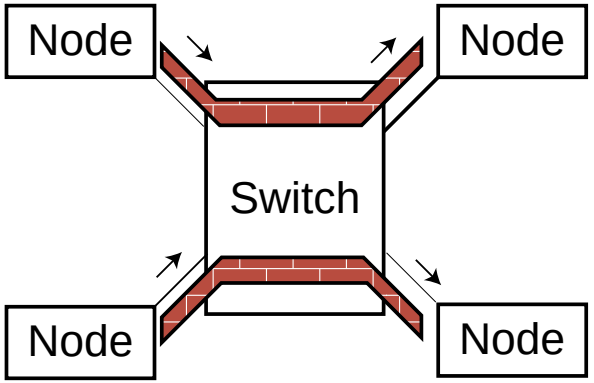
Connecting Multiple Computers

- Shared Media vs. Switched: pairs communicate at same time: "point-to-point" connections
- Aggregate BW in switched network is many times shared
  - point-to-point faster since no arbitration, simpler interface
- Arbitration in Shared network?
  - Central arbiter for LAN?
  - Listen to check if being used ("Carrier Sensing")
  - Listen to check if collision ("Collision Detection")
  - Random resend to avoid repeated collisions; not fair arbitration;
  - OK if low utilization

Shared Media (Ethernet)



Switched Media (CM-5,ATM)



(A. K. A. data switching interchanges, multistage interconnection networks, interface message processors)

Connection-Based vs. Connectionless

- Telephone: operator sets up connection between the caller and the receiver
  - Once the connection is established, conversation can continue for hours
- Share transmission lines over long distances by using switches to multiplex several conversations on the same lines
  - "Time division multiplexing" divide B/W transmission line into a fixed number of slots, with each slot assigned to a conversation
- Problem: lines busy based on number of conversations, not amount of information sent
- Advantage: reserved bandwidth

Connection-Based vs. Connectionless

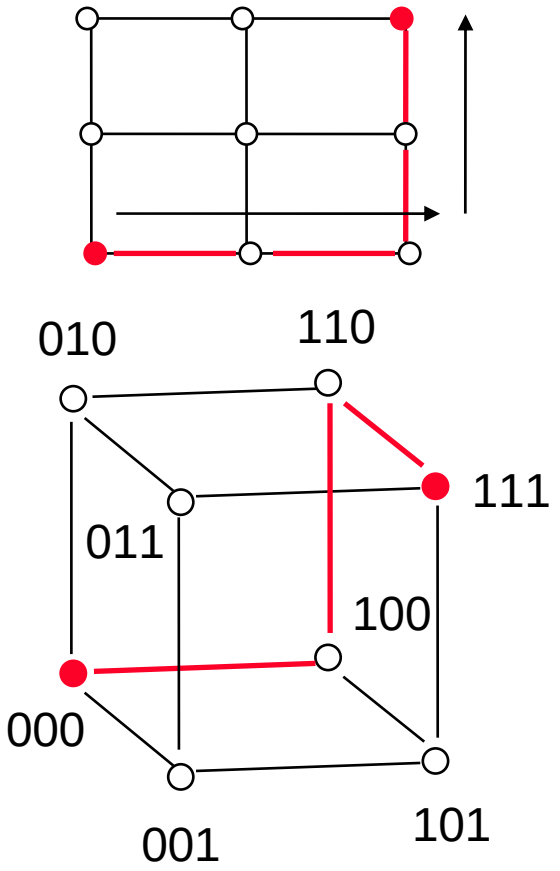
- Connectionless: every package of information must have an address => packets
  - Each package is routed to its destination by looking at its address
  - Analogy, the postal system (sending a letter)
  - also called "Statistical multiplexing"
  - Note: "Split phase buses" are sending packets

Routing Messages

- Shared Media
  - Broadcast to everyone
- Switched Media needs real routing. Options:
  - Source-based routing: message specifies path to the destination (changes of direction)
  - Virtual Circuit: circuit established from source to destination, message picks the circuit to follow
  - Destination-based routing: message specifies destination, switch must pick the path
    - » deterministic: always follow same path
    - » adaptive: pick different paths to avoid congestion, failures
    - » Randomized routing: pick between several good paths to balance network load

Deterministic Routing Examples

- mesh: dimension-order routing
  - (x_1, y_1) -> (x_2, y_2)
  - first $\Delta x = x_2 - x_1$,
  - then $\Delta y = y_2 - y_1$,
- hypercube: edge-cube routing
  - $X = x_o x_1 x_2 \ldots x_n$ -> $Y = y_o y_1 y_2 \ldots y_n$
  - R = X xor Y
  - Traverse dimensions of differing address in order
- tree: common ancestor
- Deadlock free?



Store and Forward vs. Cut-Through

- Store-and-forward policy: each switch waits for the full packet to arrive in switch before sending to the next switch (good for WAN)
- Cut-through routing or worm hole routing: switch examines the header, decides where to send the message, and then starts forwarding it immediately
  - In worm hole routing, when head of message is blocked, message stays strung out over the network, potentially blocking other messages (needs only buffer the piece of the packet that is sent between switches).
  - Cut through routing lets the tail continue when head is blocked, accordioning the whole message into a single switch. (Requires a buffer large enough to hold the largest packet).

Cut-Through vs. Store and Forward

- Advantage
  - Latency reduces from function of:

    number of intermediate switches X by the size of the packet

    to

    time for 1st part of the packet to negotiate the switches
    + the packet size ÷ interconnect BW

Congestion Control

- Packet switched networks do not reserve bandwidth; this leads to *contention* (connection based limits input)
- Solution: prevent packets from entering until contention is reduced (e.g., freeway on-ramp metering lights)
- Options:
  - Packet discarding: If packet arrives at switch and no room in buffer, packet is discarded (e.g., UDP)
  - Flow control: between pairs of receivers and senders; use feedback to tell sender when allowed to send next packet
    - » Back-pressure: separate wires to tell to stop
    - » Window: give original sender right to send N packets before getting permission to send more; overlapslatency of interconnection with overhead to send & receive packet (e.g., TCP), adjustable window
  - Choke packets: aka "rate-based"; Each packet received by busy switch in warning state sent back to the source via choke packet. Source reduces traffic to that destination by a fixed % (e.g., ATM)

Protocols: HW/SW Interface

- Internetworking: allows computers on independent and incompatible networks to communicate reliably and efficiently;
  - Enabling technologies: SW standards that allow reliable communications without reliable networks
  - Hierarchy of SW layers, giving each layer responsibility for portion of overall communications task, called protocol families or protocol suites
- Transmission Control Protocol/Internet Protocol (TCP/IP)
  - This protocol family is the basis of the Internet
  - IP makes best effort to deliver; TCP guarantees delivery
  - TCP/IP used even when communicating locally: NFS uses IP even though communicating across homogeneous LAN

CS 252 Administrivia

- Select partner, project?
- Read Amdahl's Law paper

Network/Routers Berkeley/Stanford

2. gig10-cnr1.EECS.Berkeley.EDU (169.229.3.65)

| full-duplex 1000baseSX

3. gigE5-0-0.inr-210-cory.Berkeley.EDU (169.229.1.45)
[cisco 7513/RSP4]

| full-duplex 100baseFX (1 of 2)

4. fast4-0-0.inr-002-eva.Berkeley.EDU (128.32.0.34)
[cisco 7507/RSP4]

| OC-3 PoS (1 of 2; 132 Mbit/sec)

5. pos0-2.inr-000-eva.Berkeley.EDU (128.32.0.73)
[cisco 12008 (GSR)]

| OC-12 PoS (628 Mbit/sec)

6. pos3-0.c2-berk-gsr.Berkeley.EDU (128.32.0.90)
[cisco 12012 (GSR)]

Network/Routers Berkeley/Stanford II

6. pos3-0.c2-berk-gsr.Berkeley.EDU (128.32.0.90)
   [cisco 12012 (GSR)]
   | OC-12 PoS (628 Mbit/sec)
7. SUNV--BERK.POS.calren2.net (198.32.249.14)
   [cisco 12008 (GSR)]
   | OC-12 PoS (628 Mbit/sec)
8. STAN--SUNV.POS.calren2.net (198.32.249.74)
   [cisco 12008 (GSR)]
   | OC-12 PoS (628 Mbit/sec)
9. i2-gateway.Stanford.EDU (171.64.1.214)
   [cisco 120xx (GSR)]
10. Core4-gateway.Stanford.EDU (171.64.1.226)
11. 171.64.3.89 (171.64.3.89)
12. CS.Stanford.EDU (171.64.64.64)

TraceRoute Berkeley to Stanford, I (round trip times for 3 probes)

1 fast1-1.snr1.CS.Berkeley.EDU (128.32.131.1)
1.12 ms 0.593 ms 0.546 ms

2 gig10-cnr1.EECS.Berkeley.EDU (169.229.3.65)
0.695 ms 0.615 ms 0.662 ms

3 gigE5-0-0.inr-210-cory.Berkeley.EDU (169.229.1.45)
0.783 ms 0.741 ms 0.708 ms

4 fast4-0-0.inr-002-eva.Berkeley.EDU (128.32.0.34)
1.89 ms 1.3 ms 1.24 ms

5 pos0-2.inr-000-eva.Berkeley.EDU (128.32.0.73)
1.34 ms 1.99 ms 1.51 ms

6 pos3-0.c2-berk-gsr.Berkeley.EDU (128.32.0.90)
1.82 ms 1.65 ms 2.18 ms

7 SUNV--BERK.POS.calren2.net (198.32.249.14)
2.34 ms 2.78 ms 3.18 ms

TraceRoute Berkeley to Stanford, II

7 SUNV--BERK.POS.calren2.net (198.32.249.14)
2.34 ms 2.78 ms 3.18 ms

8 STAN--SUNV.POS.calren2.net (198.32.249.74)
3.36 ms 3.36 ms 2.91 ms

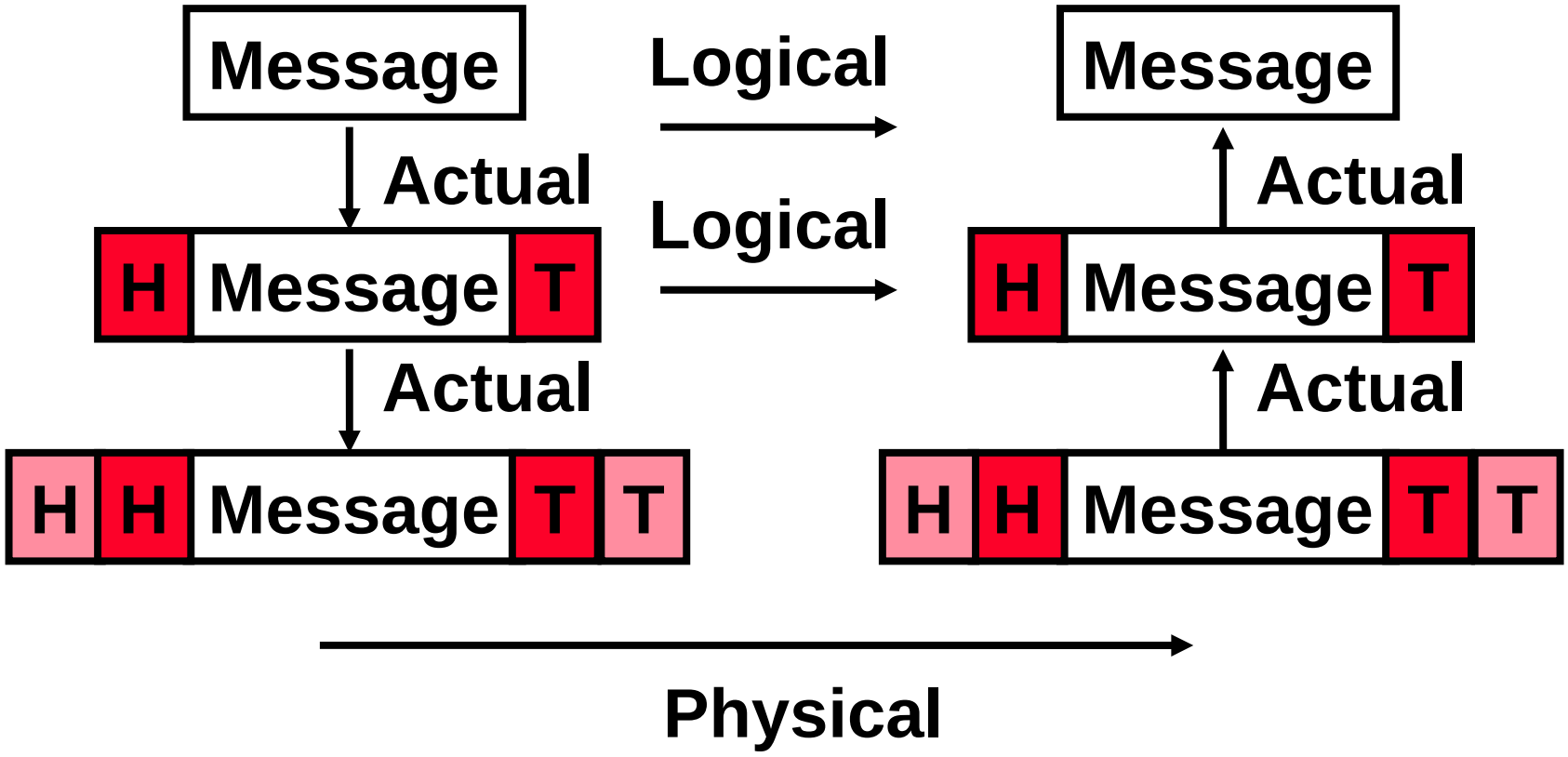
9 i2-gateway.Stanford.EDU (171.64.1.214)
3.73 ms 3.50 ms 2.98 ms

10 Core4-gateway.Stanford.EDU (171.64.1.226)
3.52 ms 3.69 ms 3.34 ms

11 171.64.3.89 (171.64.3.89)
5.46 ms 4.38 ms 4.13 ms

12 CS.Stanford.EDU (171.64.64.64)
4.23 ms * ms 4.37 ms

Protocol Family Concept

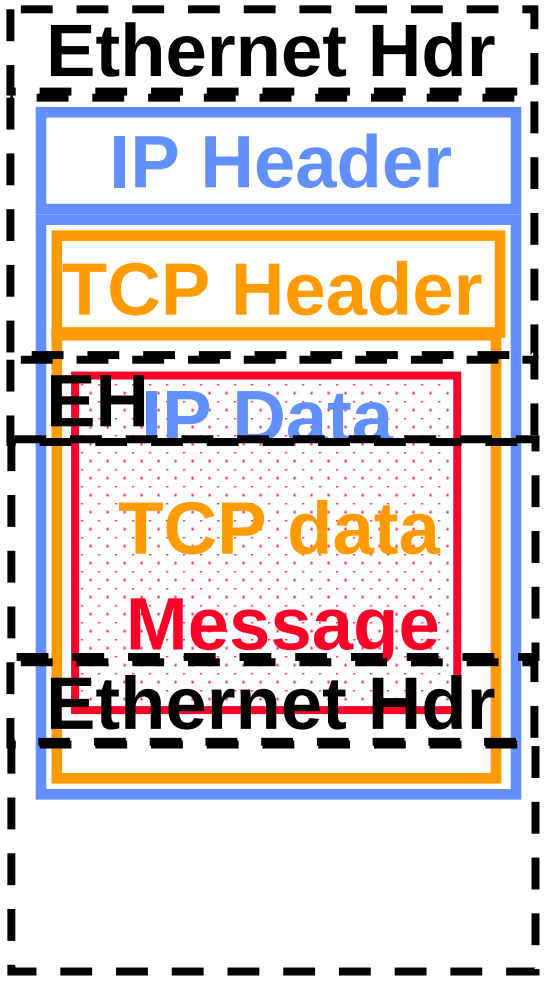


Protocol Family Concept

- Key to protocol families is that communication occurs logically at the same level of the protocol, called peer-to-peer,
- but is implemented via services at the next lower level
- Encapsulation: carry higher level information within lower level "envelope"
- Fragmentation: break packet into multiple smaller packets and reassemble
- Danger is each level increases latency if implemented as hierarchy (e.g., multiple check sums)

TCP/IP packet, Ethernet packet, protocols

- Application sends message
- TCP breaks into 64KB segments, adds 20B header
- IP adds 20B header, sends to network
- If Ethernet, broken into 1500B packets with headers, trailers (24B)
- All Headers, trailers have length field, destination, ...



Example Networks

- Ethernet: shared media 10 Mbit/s proposed in 1978, carrier sensing with expotential backoff on collision detection
- 15 years with no improvement; higher BW?
- Multiple Ethernets with devices to allow Ehternets to operate in parallel!
- 10 Mbit Ethernet successors?
  - FDDI: shared media (too late)
  - ATM (too late?)
  - Switched Ethernet
  - 100 Mbit Ethernet (Fast Ethernet)
  - Gigabit Ethernet
  - 10 Gigabit Ethernet in 2002?

Connecting Networks

- Bridges: connect LANs together, passing traffic from one side to another depending on the addresses in the packet.
  - operate at the Ethernet protocol level
  - usually simpler and cheaper than routers
- Routers or Gateways: these devices connect LANs to WANs or WANs to WANs and resolve incompatible addressing.
  - Generally slower than bridges, they operate at the internetworking protocol (IP) level
  - Routers divide the interconnect into separate smaller subnets, which simplifies manageability and improves security
- Cisco is major supplier; basically special purpose computers

Comparing Networks

	SAN		LAN			WAN
	FC-AL	Infini-band	10 Mb Ethernet	100 Mb Ethernet	1000 Mb Ethernet	ATM
Length (meters)	30/1000	17/100	500/2500	200	100	
Data lines	2	1, 4, 12	1	1	4/1	1
Clock (MHz)	1000	2500	10	100	1000	155/ 622
Switch?	Opt.	Yes	Optional	Opt.	Yes	Yes
Nodes	<=127	~1000	<=254	<=254	<=254	~10000
Material	Copper / fiber	Copper /fiber	Copper	Copper	Copper /fiber	Copper /fiber

Comparing Networks

	SAN		LAN			WAN
	FC-AL	Infini-band	10 Mb Ethernet	100 Mb Ethernet	1000 Mb Ethernet	ATM
Switch?	Opt.	Yes	Optional	Opt.	Yes	Yes
Bisection BW (Mbits /sec)	800 shared or 800 x switch ports	(2000 - 24000) x switch ports	10 shared or 10 x switch ports	100 shared or 100 x switch ports	1000 x switch ports	155 x switch ports
Peak link BW(Mbits /sec)	800	2000, 8000, 24000	10	100	1000	155/ 622
Topology	Ring or Star	Star	Line or Star	Line or Star	Star	Star

Comparing Networks

	SAN		LAN			WAN
	FC-AL	Infiniband	10 Mb Ethernet	100 Mb Ethernet	1000 Mb Ethernet	ATM
Connec-tionless?	Yes	Yes	Yes	Yes	Yes	No
Store & forward?	No	No	No	No	No	Yes
Conges-tion control	Credit-based	Back-pressure	Carrier sense	Carrier sense	Carrier sense	Credit based
Standard	ANSI Task Group X3T11	Infiniband Trade Associa-tion	IEEE 802.3	IEEE 802.3	IEEE 802.3 ab-1999	ATM Forum

Packet Formats

- See Fig 7.20 on page 634

Wireless Networks

- Media can be air as well as glass or copper
- Radio wave is electromagnetic wave propagated by an antenna
- Radio waves are modulated: sound signal superimposed on stronger radio wave which carries sound signal, called *carrier signal*
- Radio waves have a wavelength or frequency: measure either length of wave or number of waves per second (MHz): long waves => low frequencies, short waves => high frequencies
- Tuning to different frequencies => radio receiver pick up a signal.
  - FM radio stations transmit on band of 88 MHz to 108 MHz using frequency modulations (FM) to record the sound signal

Issues in Wireless

- Wireless often => mobile => network must rearrange itself dynamically
- Subject to jamming and eavesdropping
  - No physical tape
  - Cannot detect interception
- Power
  - devices tend to be battery powered
  - antennas radiate power to communicate and little of it reaches the receiver
- As a result, raw bit error rates are typically a thousand to a million times higher than copper wire

Reliability of Wires Transmission

- *bit error rate* (*BER*) of wireless link determined by received signal power, noise due to interference caused by the receiver hardware, interference from other sources, and characteristics of the channel
  - Path loss: power to overcome interference
  - Shadow fading: blocked by objects (walls, buildings)
  - Multipath fading: interference between multiple version of signals arriving different times
  - Interference: reuse of frequency or from adjacent channels

2 Wireless Architectures

- *Base-station* architectures
  - Connected by land lines for longer distance communication, and the mobile units communicate only with a single local base station
  - More reliable since 1-hop from land lines
  - Example: cell phones
- *Peer-to-peer* architectures
  - Allow mobile units to communicate with each other, and messages hop from one unit to the next until delivered to the desired unit
  - More reconfigurable

Cellular Telephony

- Exploit exponential path loss to reuse same frequency at spatially separated locations, thereby greatly increasing customers served
- Divide region into nonoverlaping hexagonal cells (2-10 mi. diameter) which use different frequencies if nearby, reusing a frequency when cells far apart so that mutual interference OK
- Intersection of three hexagonal cells is a base station with transmitters and antennas
- Handset selects a cell based on signal strength and then picks an unused radio channel
- To properly bill for cellular calls, each cellular phone handset has an electronic serial number

Cellular Telephony II

- Orginal analog design frequencies set for each direction: pair called a channel
  - 869.04 to 893.97 MHz, called the *forward path*
  - 824.04 MHz to 848.97 MHz, called the *reverse path*
  - Cells might have had between 4 and 80 channels
- Several digital successors:
  - *Code division multiple access* (*CDMA*) uses a wider radio frequency band
  - *time division multiple access* (*TDMA*)
  - *global system for mobile communication* (*GSM*)
  - *International Mobile Telephony* 2000 (IMT-2000) which is based primarily on two competing versions of CDMA and one TDMA, called Third Generation (3G)

Practical Issues for Inteconnection Networks

- Connectivity: max number of machines affects complexity of network and protocols since protocols must target largest size
- Connection Network Interface to computer
  - Where in bus hierarchy? Memory bus? Fast I/O bus? Slow I/O bus? (Ethernet to Fast I/O bus, Inifiband to Memory bus since it is the Fast I/O bus)
  - SW Interface: does software need to flush caches for consistency of sends or receives?
  - Programmed I/O vs. DMA? Is NIC in uncachable address space?

Practical Issues for Inteconnection Networks

- Standardization advantages:
  - low cost (components used repeatedly)
  - stability (many suppliers to chose from)
- Standardization disadvantages:
  - Time for committees to agree
  - When to standardize?
    - » Before anything built? => Committee does design?
    - » Too early suppresses innovation
- Reliability (vs. availability) of interconnect

Practical Issues

Interconnection	SAN	LAN	WAN
Example	Inifiband	Ethernet	ATM
Standard	Yes	Yes	Yes
Fault Tolerance?	Yes	Yes	Yes
Hot Insert?	Yes	Yes	Yes

- Standards: required for WAN, LAN, and likely SAN!
- Fault Tolerance: Can nodes fail and still deliver messages to other nodes?
- Hot Insert: If the interconnection can survive a failure, can it also continue operation while a new node is added to the interconnection?

Cross-Cutting Issues for Networking

- Efficient Interface to Memory Hierarchy vs. to Network
  - SPEC ratings => fast to memory hierarchy
  - Writes go via write buffer, reads via L1 and L2 caches
- Example: 40 MHz SPARCStation(SS)-2 vs 50 MHz SS-20, no L2$ vs 50 MHz SS-20 with L2$ I/O bus latency; different generations
- SS-2: combined memory, I/O bus => 200 ns
- SS-20, no L2$: 2 busses +300ns => 500ns
- SS-20, w L2$: cache miss+500ns => 1000ns

Crosscutting: Smart Switch vs. Smart Network Interface Card

	Less Intelligent	More Intelligent
Switch	Small Ethernet Myrinet Inifiband	Large Ethernet
NIC	Ethernet Infiniband Target Channel Adapter	Myrinet Inifiband Host Channel Adapter

- Inexpensive NIC => Ethernet standard in all computers
- Inexpensive switch => Ethernet used in home networks

Summary: Networking

- Protocols allow hetereogeneous networking
  - Protocols allow operation in the presense of failures
  - Internetworking protocols used as LAN protocols
    => large overhead for LAN
- Integrated circuit revolutionizing networks as well as processors
  - Switch is a specialized computer
  - Faster networks and slow overheads violate of Amdahl's Law