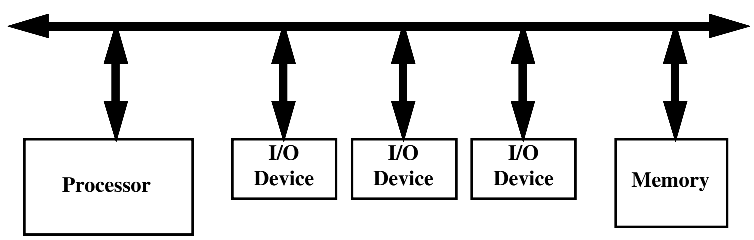
CS152
Computer Architecture and Engineering
Lecture 21: Interconnection Networks

April 12, 1995

Dave Patterson (patterson@cs) and
Shing Kong (shing.kong@eng.sun.com)

Slides available on http://http.cs.berkeley.edu/~patterson

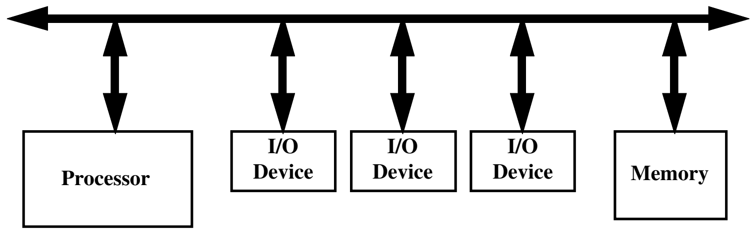
Recap: Advantages of Buses



° Versatility:
  - New devices can be added easily
  - Peripherals can be moved between computer systems that use the same bus standard

° Low Cost:
  - A single set of wires is shared in multiple ways

Recap: Disadvantages of Buses



° It creates a communication bottleneck
  • The bandwidth of that bus can limit the maximum I/O throughput

° The maximum bus speed is largely limited by:
  • The length of the bus
  • The number of devices on the bus
  • The need to support a range of devices with:
    - Widely varying latencies
    - Widely varying data transfer rates

Recap: Types of Buses

° Processor-Memory Bus (design specific)
  • Short and high speed
  • Only need to match the memory system
    - Maximize memory-to-processor bandwidth
  • Connects directly to the processor

° I/O Bus (industry standard)
  • Usually is lengthy and slower
  • Need to match a wide range of I/O devices
  • Connects to the processor-memory bus or backplane bus

° Backplane Bus (industry standard)
  • Backplane: an interconnection structure within the chassis
  • Allow processors, memory, and I/O devices to coexist
  • Cost advantage: one single bus for all components

Recap: Increasing the Bus Bandwidth

° Separate versus multiplexed address and data lines:
  • Address and data can be transmitted in one bus cycle if separate address and data lines are available
  • Cost: (a) more bus lines, (b) increased complexity

° Data bus width:
  • By increasing the width of the data bus, transfers of multiple words require fewer bus cycles
  • Example: SPARCstation 20's memory bus is 128 bit wide
  • Cost: more bus lines

° Block transfers:
  • Allow the bus to transfer multiple words in back-to-back bus cycles
  • Only one address needs to be sent at the beginning
  • The bus is not released until the last word is transferred
  • Cost: (a) increased complexity
    (b) decreased response time for request

Bus Summary:

° Bus arbitration schemes:
  • Daisy chain arbitration: it cannot assure fairness
  • Centralized parallel arbitration: requires a central arbiter

° I/O device notifying the operating system:
  • Polling: it can waste a lot of processor time
  • I/O interrupt: similar to exception except it is asynchronous

° Delegating I/O responsibility from the CPU
  • Direct memory access (DMA)
  • I/O processor (IOP)

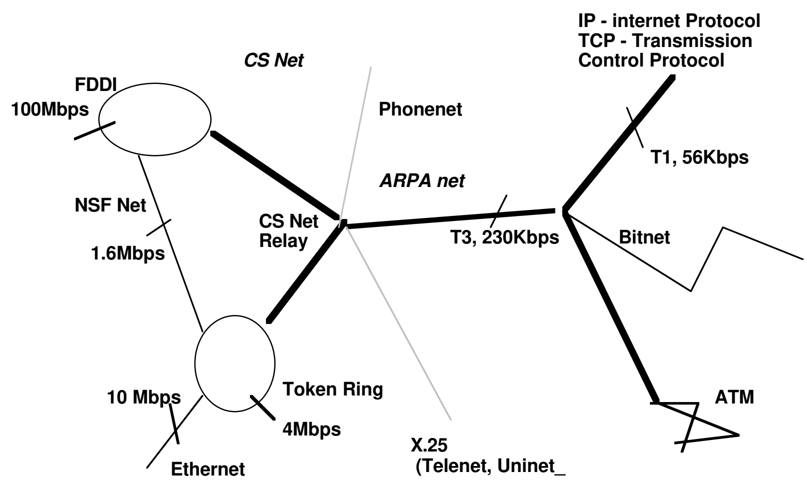
Outline of Today's Lecture

- Recap and Introduction (5 minutes)
- Introduction to Buses (15 minutes)
- Questions and Administrative Matters (5 minutes)
- Bus Types and Bus Operation (10 minutes)
- Bus Arbitration and How to Design a Bus Arbiter (15 minutes)
- Break (5 minutes)
- Operating System's Role (15 minutes)
- Delegating I/O Responsibility from the CPU (5 minutes)
- Summary (5 minutes)

Networks

- Goal: Communication between computers
- Eventual Goal: treat collection of computers as if one big computer
- Theme: Different computers must agree on many things
  => Overriding importance of standards
- Warning: Buzzword rich environment

Current Major Networks



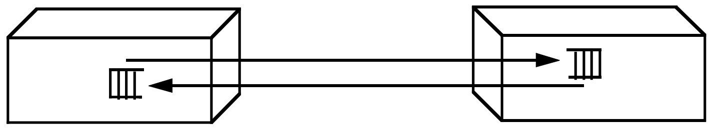
Networks

° Facets people talk a lot about
  - direct vs indirect
  - topology
  - routing algorithm
  - switching
  - wiring

° What matters
  - latency
  - bandwidth
  - cost
  - reliability

ABCs of Networks

° Starting Point: Send bits between 2 computers



° FIFO Queue on each end

° Can send both ways ("Full Duplex")

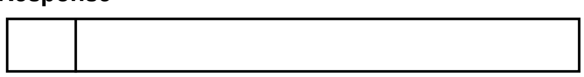
° Rules for communication? "protocol"
  • Inside a computer?
  • Loads/Stores: Request(Address) & Response (Data)
  • Need Request & Response
  • Name for standard group of bits sent: Packet

A Simple Example

° What is format of packet?
  • Fixed? Number bytes?

Request/
Response

Address/Data



1 bit

32 bits

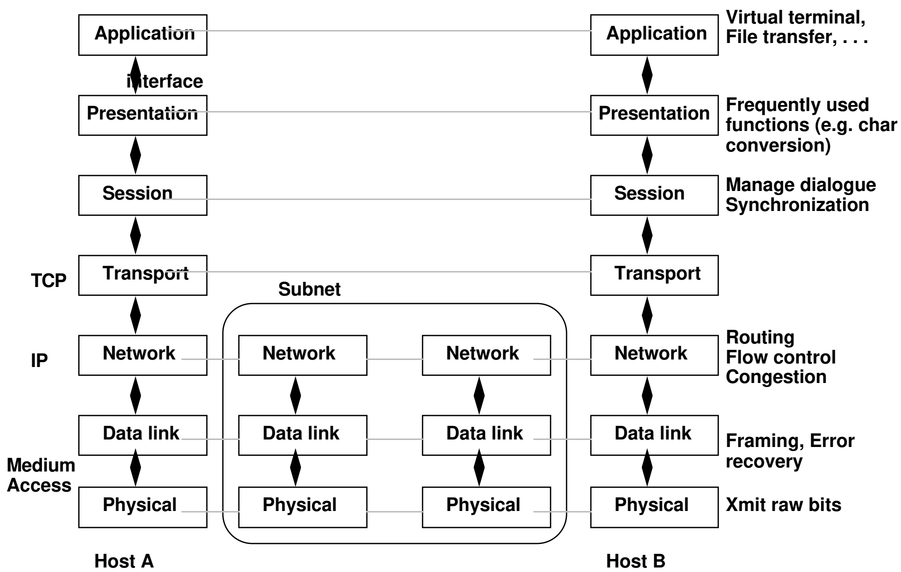
0: Please send data from Address
1: Data corresponding to request

Questions about Simple Example

- What if more than 2 computers want to communicate?
  - Need computer address field in packet?
- What if packet is garbled in transit?
  - Add error detection field in packet?
- What if packet is lost?
  - More elaborate protocols to detect loss?
- What if multiple processes/machine?
  - Queue per process?
- Questions such as these lead to more complex protocols and packet formats

Protocol Stacks

OSI Reference Model



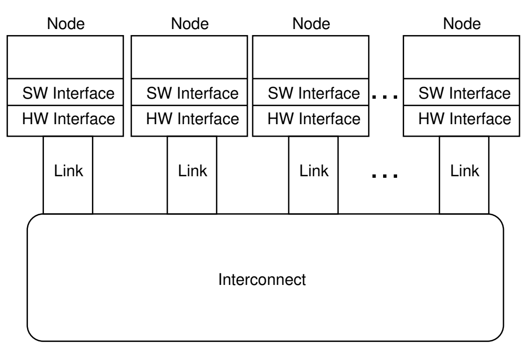
Questions and Administrative Matters (5 Minutes)

- **Home Stretch**
- **Last assignment today (cache), including project description**
- **Remaining schedule (Guests live in 306 Soda)**

12-Apr	Wed	Networks and MP
14-Apr	Fri	Superscalar
19-Apr	Wed	Midterm II 5 PM in Bechtel/Pizza
21-Apr	Fri	SGI T5 Microarchitecture
26-Apr	Wed	DEC Alpha Microarchitecture
28-Apr	Fri	UltraSPARC Microarchitecture
3-May	Wed	Project Demo Meetings
5-May	Fri	Class Summary
8-May	Mon	Projects Due 4PM in 634

Interconnection Networks

- **Examples**
  - **MPP networks (CM-5): 1000s nodes; ≤ 25 meters per link**
  - **Local Area Networks (Ethernet): 100s nodes; ≤ 1000 meters**
  - **Wide Area Network (ATM): 1000s nodes; ≤ 5,000,000 meters**



Interconnection Network Issues

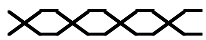
- Implementation Issues
- Performance Measures
- Architectural Issues
- Practical Issues

Implementation Issues

Interconnect	*MPP*	*LAN*	*WAN*
Example	CM-5	Ethernet	ATM
Maximum length between nodes	25 m	500 m; ≤5 repeaters	copper: 100 m optical: 1000 m
Number data lines	4	1	1
Clock Rate	40 MHz	10 MHz	≥ 155.5 MHz
Shared vs. Switch	Switch	Shared	Switch
Maximum number of nodes	2048	254	> 10,000
Media Material	Copper	Twisted pair copper wire or Coaxial cable	Twisted pair copper wire or optical fiber

Media

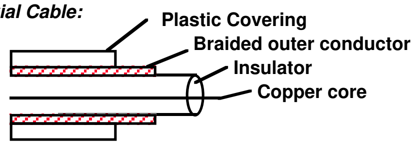
Twisted Pair:



Several Mb/s up to km
- more with shielded twisted pair
- category 5: 4 wires

Why twisted?

Coaxial Cable:

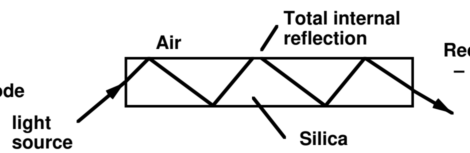


10Mbps at 1km
- more at shorter length

Tap with T-junction or vampire

Fiber Optics

Transmitter
- L.E.D
- Laser Diode



Receiver
- Photodiode

Multimode: many rays bouncing at different angles
Single mode: diameter of fiber less than one wavelength
- acts like a wave guide

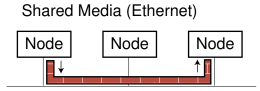
Gb/s at 1 km

Line of sight (microwave) 2-40 GHz

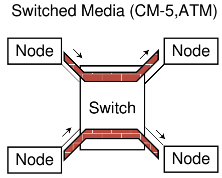
Implementation Issues

- Advantages of Serial vs. Parallel lines:
  - No synchronizing signals
  - Higher clock rate and longer distance than parallel lines. (e.g., 60 MHz x 256 bits x 0.5 m vs. 155 MHz x 1 bit x 100 m)
    - Imperfections in the copper wires or integrated circuit pad drivers can cause skew in the arrival of signals, limiting the clock rate, and the length and number of the parallel lines.
- Switched vs. Shared Media: pairs communicate at same time: "point-to-point" connections

Shared Media (Ethernet)

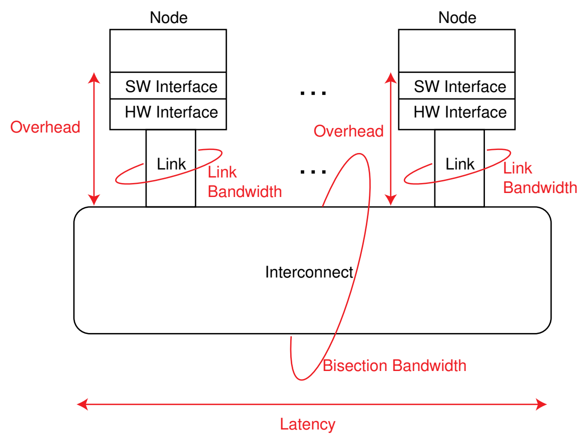


Switched Media (CM-5,ATM)



Network Performance Measures

° Overhead: latency of interface vs. Latency: network



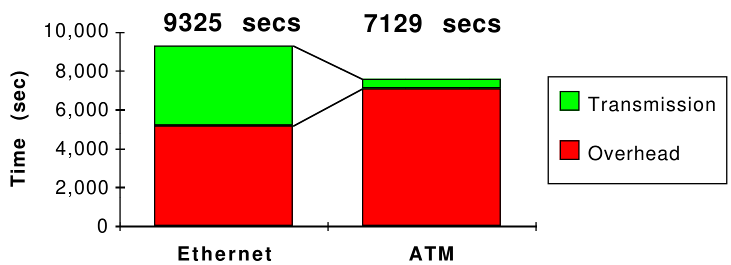
Example Performance Measures

Interconnect	*MPP*	*LAN*	*WAN*
Example	CM-5	Ethernet	ATM
Bisection BW	Nx 5MB/s	1.125 MB/s	N x 10 MB/s
Int./Link BW	20 MB/s	1.125 MB/s	10 MB/s
Latency	5 µsec	15 µsec	50 to 10,000 µs
HW Overhead to/from	0.5/0.5 µs	6/6 µs	6/6 µs
SW Overhead to/from	1.6/12.4 µs	200/241 µs	207/360 µs

(TCP/IP on LAN/WAN)

Importance of Overhead (+ Latency)

- Ethernet / SS10: 9 Mb/s BW, 900 µsecs ovhd
- ATM Synoptics: 78 Mb/s BW, 1,250 µsecs ovhd.
- NFS trace over 1 week: 95% msgs < 200 bytes



• Link Bandwidth as misleading as MIPS

Break (5 Minutes)

Example Performance Measures

Interconnect	*MPP*	*LAN*	*WAN*
Example	CM-5	Ethernet	ATM
Topology	“Fat” tree	Line	Variable, constructed from multistage switches
Connection based?	No	No	Yes
Data Transfer Size	Variable: 4 to 20B	Variable: 0 to 1500B	Fixed: 48B

Topology

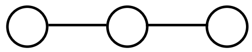
° Structure of the interconnect

° Determines
- degree: number of links from a node
- diameter: max number of links crossed between nodes
- average distance: number of hops to random destination
- bisection
  - minimum number of links that separate the network into two halves

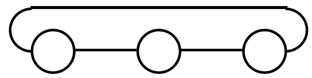
° Warning: these three-dimensional drawings must be mapped onto chips and boards which are essentially two-dimensional media
- elegant when sketched on the blackboard may look awkward when constructed from chips, cables, boards, and boxes

Important Topologies

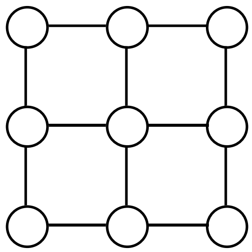
1D mesh



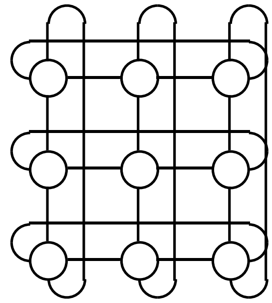
Ring



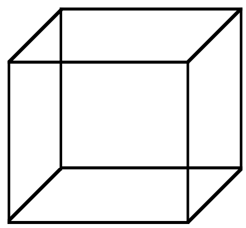
2D mesh



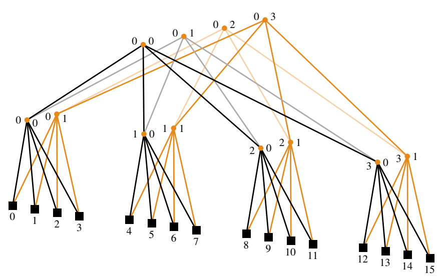
2D torus



Hypercube



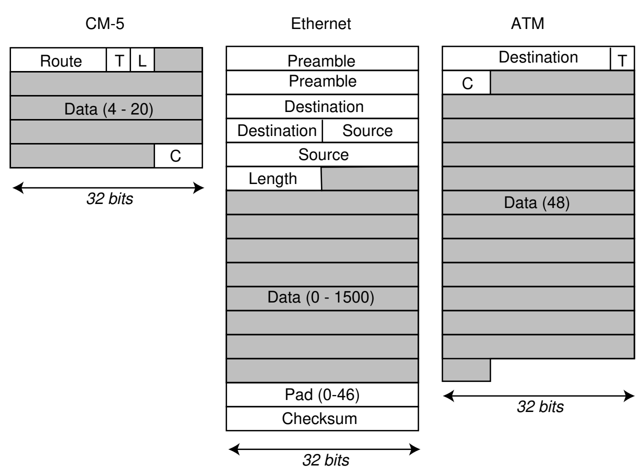
Fat Tree



Connection based vs. Connectionless

- Telephone: operator sets up connection between the caller and the receiver
  - once the connection was established, conversation could continue for hours
- Share transmission lines over long distances by using switches to multiplex several conversations on the same lines
  - " time division multiplexing" divide BW 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
- connectionless: every package of information must have an address => packets
  - Each package is routed to the destination by looking at its address e.g., the postal system
  - Split phase buses send packets

Packet formats



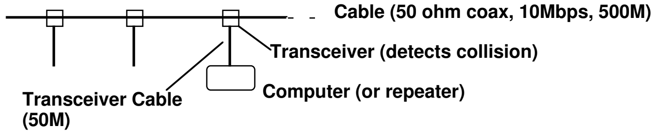
- Fields: Destination, Checksum(C), Length(L), Type(T)
- Data/Header Sizes in bytes: (4 to 20)/4, (0 to 1500)/26, 48/5

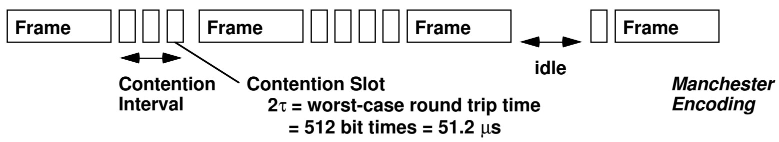
Example: Ethernet (IEEE 802.3)

- **Essentially 10Kb/s 1 wire bus with no central control**

Collision based protocol
– 1-persistent carrier sense multiple access

Listen. If nobody is taking, go ahead and talk.
f you hear anybody else talking, yell SORRY and try again later.

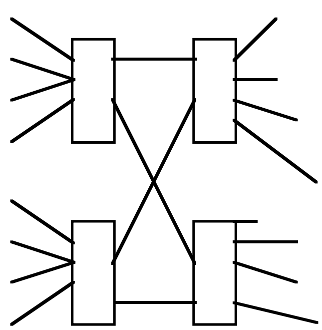




binary exponential backoff
After 1st collision, wait for 0 or 1 slots, at random
After 2nd collison, choose between 0,1,2,3
etc up to 1023 slots.
After 16 collisions fail

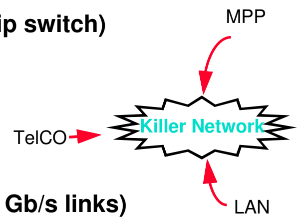
Example: ATM (Asynchronous Transfer Mode)

- **Asynchronous Transfer Mode (155Mb/s, 622 in the future)**
- **Point-to-point, dedicated, switched**
- **5+48 byte fixed sized cells**
- **Connection Oriented using Virtual Channels**
- **Bandwidth guarantees**



Towards the Killer Network

- High bandwidth, scalable (switched) LANs
  - Repackaged MPP backplane (single chip switch)
    - TMC, Intel, . . .
    - IBM SP-2
    - Myrinet (Seitz & Cohen)
  - Research ATM efforts
    - DEC AN2 (ATM switch capable of Gb/s links)
  - Commercial ATM products
    - “off the curve,” but catching up
  - Ethernet successors
    - 100 Mbit/s: Fast Ethernet (Sun et al) 100 VGA (HP et al)
    - Switched Ethernet
    - Switched 100 Mbit/sec Ethernet



Summary: Interconnections

- Communication between computers
- Packets for standards, protocols to cover normal and abnormal events
- Implementation issues: length, width, media
- Performance issues: overhead, latency, bisection BW
- Topologies: many to chose from, but (SW) overheads make them look the alike; cost issues in topologies