LBSC 690: Week 1

Computers and Networks

Jimmy Lin
College of Information Studies
University of Maryland

Monday, January 29, 2007

Goals

- By the end of this class, you will…
  - Have a basic understanding of computers and networks
  - Know how to think about "space", "time", and "speed"
  - Understand of how computers store data and move data around
  - Be able to evaluate tradeoffs between different technologies

A Very Brief History of Computing

- Computer = "a person who computes" (< 1940's)
- Hardware: all developed for the government
  - Mechanical: essentially a big adding machine
  - Analog: designed for calculus, limited accuracy
  - Digital: early machines filled a room
  - Microchips: designed for missile guidance
- Software: initial applications were military
  - Numeric: computing gun angles
  - Symbolic: code-breaking

Commercial Developments

- Mainframes (1960's)
- Minicomputers (1970's)
- Personal computers (1980's)
- Networked computers (1990's)
- Ubiquitous and embedded computers (2000's)

The Processing Cycle

- Input comes from somewhere
  - Keyboard, mouse, microphone, camera, …
  - Fetch data from memory
- The system does something with it
  - Add, subtract, multiply, etc.
- Output goes somewhere
  - Monitor, speaker, printer, robot controls, …
  - Store data back into memory

Today's Focus

- **Storing** and **moving** around data
  - Within a computer
  - Between computers
- Inside a single computer: connecting the processor with the memory
- Between multiple computers: computer networks

Thinking about Size

- What's a bit?
- How much information can *n* bits represent?
- What's the difference between decimal and binary?
- And octal?
- And hexadecimal?
- 18,446,744,073,709,551,615 grains of wheat

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Units of Size

Unit	Abbreviation	Size (bytes)
bit	b	1/8
byte	B	1
kilobyte	KB	2^{10} = 1024
megabyte	MB	2^{20} = 1,048,576
gigabyte	GB	2^{30} = 1,073,741,824
terabyte	TB	2^{40} = 1,099,511,627,776
petabyte	PB	2^{50} = 1,125,899,906,842,624

How do hard drive manufactures "cheat" you?

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Thinking About Time

- Total "transfer time" is what counts
  - Time for first bit + time between first and last bits
- For long distances, the first factor is important
  - California: 1/80 of a second (by optical fiber)
  - London: 1/4 of a second (by satellite)
- For large files, the second factor dominates
  - Number of bits per second is limited by physics
- **Latency:** the amount of time it takes data to travel from source to destination
- **Bandwidth:** the amount of data that can be transmitted in a fixed amount of time

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Thinking About Speed

- Speed can be expressed two ways:
  - How long to do something once?
    - Memory speed measured as "access time"
  - How many things can you do in one second?
    - Processor speed measured in "clock cycles per second"
    - Bandwidth measured in "bits per second"
- Convenient units are typically used
  - "10 microseconds" rather than "0.00001 seconds"
- When comparing speeds, convert units first!

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Units of Time

Unit	Abbreviation	Duration (seconds)
second	sec/s	1
millisecond	ms	10^{-3} = 1/1,000
microsecond	µs	10^{-6} = 1/1,000,000
nanosecond	ns	10^{-9} = 1/1,000,000,000
picosecond	ps	10^{-12} = 1/1,000,000,000,000
femtosecond	fs	10^{-15} = 1/1,000,000,000,000,000

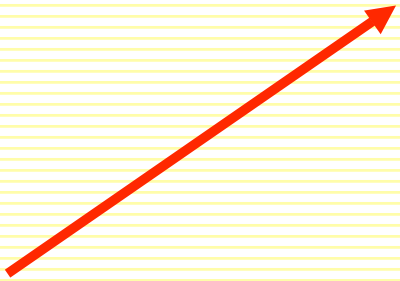
UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

Units of Frequency

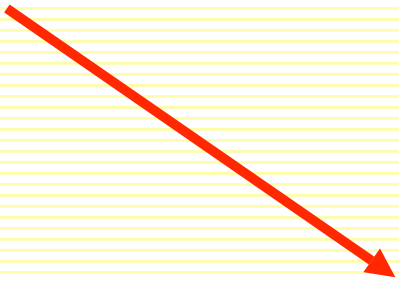
Unit	Abbreviation	Cycles per second
hertz	Hz	1
kilohertz	KHz	10^3 = 1,000
megahertz	MHz	10^6 = 1,000,000
gigahertz	GHz	10^9 = 1,000,000,000

UNIVERSITY OF MARYLAND COLLEGE of INFORMATION STUDIES

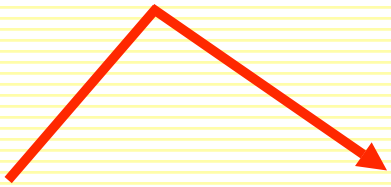
Trends in Technology: #1



Trends in Technology: #2



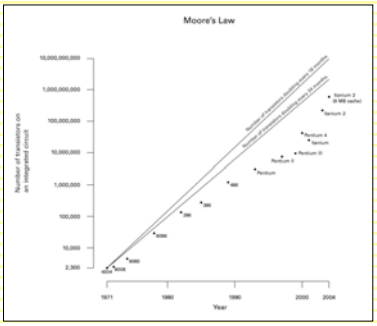
Trends in Technology: #3



Moore's Law

- What is it?
  - Gordon E. Moore, co-founder of Intel: number of components on an integrated circuit will double every 18 months (1965)
- Why is it important?

Illustration of Moore's Law



Aside: The Gigahertz Race

- Intel Pentium 4: 3.80 GHz
- Apple G5: 2.7 GHz
- Intel Core Duo: 2.0 GHz

- What does it mean?
- Which is actually faster?
- Why is this important for consumers?

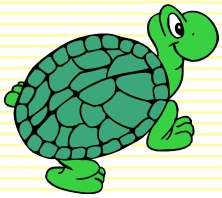
The CPU and the Memory

- CPU (Central Processor Unit) – where actual computation is performed
- Memory – location of data on which computation is performed
- Bus – moves data from memory to and from CPU
- Desiderata for memory:
  - Large
  - Fast
  - Cheap

Large, Fast, and Cheap Memory

- Impossible! (Why?)
- Engineering is all about compromise!

Small, but fast…

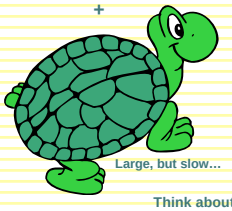


Large, but slow…

Best of Both Worlds?

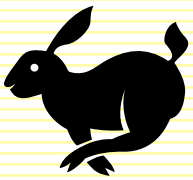
Small, but fast…

+



Large, but slow…

=



Large *and* fast?!

Think about your bookshelf and the library…

Locality

- **Spatial locality:** If the system fetched *x*, it is likely to fetch data located near *x* (Why?)
- **Temporal locality:** If the system fetched *x*, it is likely to fetch *x* again (Why?)
- Insight behind the storage hierarchy: move important data from slow, large memory to fast, small memory
- Cache: a place for concealment and safekeeping, as of valuables. (American Heritage Dict.)
- Caching strategies: what's the most effective strategy for moving data around?

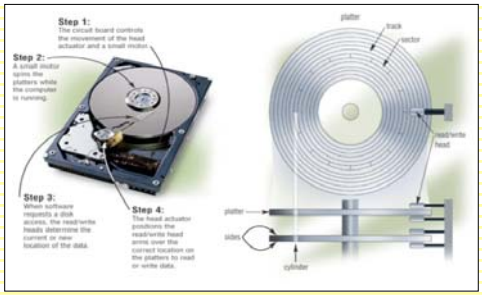
The Storage Hierarchy

Type	Speed	Size	Cost
Registers	< 1 ns	512 bytes	Very expensive
Cache	10 ns	2 MB	Very expensive
RAM	50 ns	1 GB	Cheap
Hard drive	10 ms	100 GB	Very Cheap

Trading Speed for Space

- Hard disk is larger than RAM but much slower
  - 10 ms access time and 100 GB is typical
    - 200,000x slower/100x bigger than RAM!
    - > 10 million times slower than the CPU!
- The initial access is the slow part
  - Subsequent bytes sent at 30 MB/sec (33 ns/byte)
- The importance of caching…
- What's typical cache miss rate?
- What happens if the data doesn't all fit into RAM?

How Hard Drives Work



from Shelly, Cashman, Vermaatt's Discovering Computers 2004

Summary So Far…

- For computation to occur, data must be moved to and from memory
- Different type of memories represent different tradeoffs
- Caching strategies and the storage hierarchy give us the best of both worlds

Why Networking?

- Sharing data
- Sharing hardware
- Sharing software
- Increasing robustness
- Facilitating communications
- Facilitating commerce

Packet vs. Circuit Networks

- Telephone system ("circuit-switched")
  - Fixed connection between caller and called
  - High network load results in busy signals
- Internet ("packet-switched")
  - Each transmission is broken up into pieces and routed separately
  - High network load results in long delays

Packet Switching

- Break long messages into short "packets"
  - Keeps one user from hogging a line
  - Each packet is tagged with where it's going
- Route each packet separately
  - Each packet often takes a different route
  - Packets often arrive out of order
  - Receiver must reconstruct original message
  - How do packet-switched networks deal with continuous data?
- Request retransmission for lost packets
  - Unless the first packet is lost!

Different Networks Types

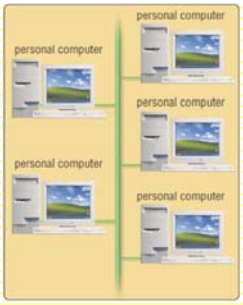
- Local Area Networks (LANs)
  - Connections within a building or a small area
- Metropolitan Area Networks (MANs)
  - Connections across a city or a larger geographic area
- Wide Area Networks (WANs)
  - Connections between multiple LANs/MANs
  - May cover thousands of square miles
- The Internet
  - Collection of WANs across multiple organizations

Local Area Networks

- Usually covering a small area
  - Short-distance lines are **fast** and **cheap**
  - Fast communications makes routing simple
- Ethernet is a common LAN technology
  - All computers are connected to the same cable
  - Ordinary lines can carry 10 Mb/sec
  - Every host broadcasts everything to all others
  - Collisions limit throughput to about 50% utilization
- Network type vs. network typology

Network Typology: Bus

- All attach to the same cable
- Transmit anytime
  - Collision detection
  - Automatic retransmission
- Inexpensive and flexible
  - Easy to add new machines
  - Robust to computer failure
- Practical for short distances
  - Half the bandwidth is wasted



Network Typology: Star

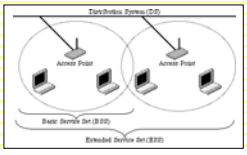
- All attach directly to a hub
- Higher cost
  - Line from hub to each machine
  - Hub must handle every packet
  - Hub requires backup power
- Much higher bandwidth
  - No sharing, no collisions
  - Allows disks to be centralized

Network Typology: Ring

- Unidirectional transmission
  - Used mostly for larger networks
- Very high bandwidth
  - No collisions
  - Simple routing policies
- Complex management
  - Changes must be coordinated

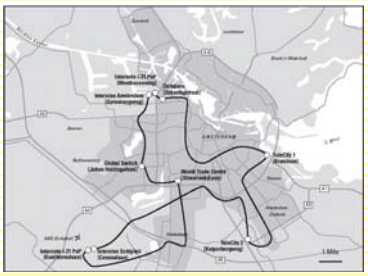
Wireless Networks

- Radio-based Ethernet
  - Effective for a few rooms within buildings
- "Access Point" gateways to wired networks
  - Available throughout most of the Maryland campus
  - Commercial providers offer "hot spots" in airports, etc.
- Available in two speeds
  - IEEE 802.11b: 10Mbps (most common)
  - IEEE 802.11g: 54Mbps (now becoming available)



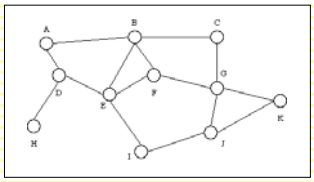
Metropolitan Area Networks

- Span entire cities (often arranged as rings)



Wide Area Networks

- Connect multiple LANs (or MANs) across a large geographic area
- Often employ a “mesh” typology
- The Internet is the largest WAN in existence



The Internet

- Global collection of public networks
  - Private networks are often called “intranets”
  - Each organization maintains its own network
- Use of shared protocols
  - TCP/IP (Transmission Control Protocol/Internet Protocol): basis for communication
  - DNS (Domain Name Service): basis for naming hosts
  - HTTP (HyperText Transfer Protocol): World Wide Web
- Next week: how does all of this work?

A Short History of the Internet

- 1969: Origins in government research
  - Advanced Research Projects Agency (ARPAnet)
- 1983: Design adopted by other agencies
  - Expansion from educational institutions to corporations
- 1991: World Wide Web added point-and-click capabilities

Now You Know About…

- Size, time, and speed
- Different types of memories and their tradeoffs
- The storage hierarchy: large and fast!
- Circuit-switched vs. packet-switched networks
- Networks of different sizes: LANs, WANs, etc.
- Network typologies: bus, star, ring, etc.