Motivation

- Computers “become” slower over time
- Money spent ineffectively
- Power consumption
- E-waste

Outline

- Problem motivation
- How computers work today
- Some history
- What can you do?
- What are computer scientists doing?
- The future

How a Computer System Works

1. User clicks on “paper.doc”
2. Copy file into memory
3. Copy part of file into CPU
4. Present result to user

Show-n-Tell: CPU

1. CPU speeds are in the nano-seconds (billions of operations per second)
2. CPU’s internal speed is ~10-100 times faster than its access to main memory.
3. Typical CPU’s internal memory: 256MB to several GB, multiple caches/cores

Show-n-Tell: CPU Heat Sink

Note: modern faster CPUs require an even larger CPU fan
**Show-n-Tell: Main Memory**

Common name: Dynamic Random Access Memory (DRAM)

Typical memory sizes today: 4-16 GB (Giga Bytes), or billion characters
Typical memory speeds: micro-seconds (millions of operations per second)

**Show-n-Tell: Flash Memory**

Major advantage: persistent (non-volatile) memory

Typical flash memory sizes today: 1GB-1TB (Giga Bytes), or billion characters
Typical flash speeds: between DRAM and hard disks
Lifetime: only thousands/millions of writes
Performance depends on fullness, age, and internal processes (GC, defrag)

**Show-n-Tell: Hard Disk**

Typical disk sizes today: 1-4 TB (Giga Bytes), or billion characters
Typical disk speeds: milli-seconds (of operations per second)

**The Storage Hierarchy Pyramid**

Smaller, faster, more expensive

CPU

100x faster

DRAM Memory

1,000x faster

Hard Disk

Bigger, slower, cheaper

**Moore’s Law**

“Transistor density doubles every 18-24 months” (c. 1965)

**Kryder’s Law**

“Magnetic storage density doubles annually” (c. 1995)
Moore’s vs. Kryder’s Law

- Hard disk size doubling annually
- CPU/Memory doubling “only” every 1.5–2 years
- Hard disk speed lagging far behind!

Moore’s Law: Transistor Density
Kryder’s Law: Hard Disk Density

How Computers REALLY Work (1)

Scenario 1: CPU and memory are large and fast enough.

How Computers REALLY Work (2)

Scenario 2: CPU too small. Memory is large enough.

How Computers REALLY Work (3)

Scenario 3: CPU and memory are BOTH too small.

The Perfect Storm

- Disk sizes getting larger the fastest
- CPU/Memory getting larger/faster
  - But lagging behind disk-size growth rates
- Disk speeds lagging far behind the rest
- Software getting larger
  - Fills disks to capacity
  - Software updates & system service packs
- E.g., anti-virus/spyware scanners
  - How long to scan a 500GB disk?

Buying a New Computer (1)

- Given a fixed budget, spend your $$ on the following, by importance:
  1. More main memory (2–8 GB+)
     - Try to keep a free memory upgrade slot
  2. Larger CPU caches
  3. Faster disks (SSD)

Caveat: may increase power consumption
Buying a New Computer (2)

• Where to cut back on (if fixed budget):
  1. Don’t get the fastest CPU (in MHz)
     • E.g., get a 2.8GHz instead of a 3.2GHz
  2. You may skip on multi-core CPUs
     • Extra “brains” but share same CPU cache
  3. Consider a smaller disk

Recommend: avoid “canned” solutions

Ongoing Comp. Sci. Research

• How to use multi-core CPUs effectively?
  ◦ Trade CPU for reduced I/O?
• How to reduce the need to go to disk?
  ◦ Compressing data (spare CPU cores)
  ◦ Prediction/Al algorithms, prefetch disk data
• Use new storage devices
  ◦ E.g., FLASH memory

Show-n-Tell: Flash Memory

Major advantage: persistent (non-volatile) memory
Typical flash memory sizes today: 128MB–4GB (Giga bytes), or billion characters
Typical flash speeds: between DRAM and hard disks
Lifetime: only millions of writes

The Future of FLASH Memory

• Intermediate cache between main memory and hard disks
• Reliability
  ◦ Distribute writes evenly across device
• Staging device inside hard disks
  ◦ Hybrid disks
• Replace spinning disks
  ◦ Solid State Disks (SSDs)
    • MacBook Air (2008)

Computer Power Use

• Energy savings trails off everything
  ◦ Small percentage improvements annually
• Energy consumption secondary design consideration:
  ◦ Idle laptop: 20–30 Watts
  ◦ Idle workstation: 60–80 Watts
  ◦ Idle server: 200+ Watts
“Green” Research

• Turn off components automatically
  ◦ Without annoying users
• Better hardware components
  ◦ Without slowing down the system
• Use FLASH to store data
  ◦ Lower costs, improve reliability
• Lightweight software
  ◦ Without reduced functionality

Mommy, Daddy, Why is My New Laptop So Dog Slow?!
or Recent Trends in Systems Research

Q&A

Erez Zadok
Stony Brook University
Computer Science Department
http://www.cs.sunysb.edu/~ezk

Today’s Storage Hierarchy Pyramid

Smaller, faster, more expensive

Bigger, slower, cheaper