DATA STORAGE TECHNOLOGIES & NETWORKS (CS C446, CS F446 & IS C446)

LECTURE 08 – STORAGE

Caching - Generic

- Caching as a principle can be applied between any two levels of memory
  - e.g. Buffer Cache (part of RAM)
    - transparent to App,
    - maintained by OS,
    - between main memory and hard disk,
    - $R_{RAM,buffer} = 1$
  - e.g. Disk cache
    - between RAM and hard disk
    - typically part of disk controller
    - typically semiconductor memory
    - may be non-volatile ROM on high end disks to support power breakdowns.
    - transparent to OS and Apps

Caching - Generic

- Other examples:
  - network client cache – transparent to OS and apps.
  - web client (browser cache) - transparent to OS and apps.
  - proxy cache (client proxy, server proxy) – transparent to OS and apps; client proxies are transparent to servers and vice-versa.
  - distribution cache – transparent to OS, apps, and clients or servers or both.
- Exercise: Find out typical R values in this case.

Input Output Devices

- Access to I/O devices is different from access to memory devices (cache, ROM, RAM)
  - I/O devices are usually not semiconductor devices.
    - Need more than electronic decoding/switching
  - They are often bulky and have moving parts
    - So they can not be stored on board.
- Most significantly I/O devices are slow
  - Speed mismatch between Processor and I/O devices
    - Fully synchronous I/O will heavily impair processor utilization

I/O Techniques

- Polling:
  - pseudo-asynchronous
    - Processor inspects (multiple) devices in rotation
  - Even with several devices,
    - processor may still be forced to do useless work or wait or both.
- Interrupts:
  - Processor initiates I/O by requesting an operation with the device.
    - May disconnect if response can't be immediate, which is usually the case
  - When device is ready with a response it interrupts the processor.
    - Processor finishes I/O with the device.
  - Asynchronous but
    - Data transfer between I/O device and memory still requires processor to execute instructions.

I/O Techniques

- Direct Memory Access
  - Processor initiates I/O
  - DMA controller acts as an intermediary:
    - interacts with the device,
    - transfers data to/from memory as appropriate, and
    - interrupts processor to signal completion.
  - From the processor's perspective DMA controller is yet another device
    - But one that works at semiconductor speeds
- I/O Processor
  - More sophisticated version of DMA controller with
    - the ability to execute code: execute I/O routines, interact with the OS etc.

I/O Performance

- I/O may be a small component of a program
  - Why worry about I/O performance?
- Speed-up and Amdahl's Law
  - Say I/O is a fraction f of the total work-load.
  - And you improve the rest of the system by a factor of x
  - Then the effective speedup is:
    - $T_{origin} / (f*T_{origin} + (1-f)*T_{origin}/x) = 1/(f+(1-f)/x)$
- Implications of Amdahl's Law:
  - Example:
    - if 10% of time is spent in I/O and
    - rest of the system performance improves by a factor of 100
    - then the effective speedup is
    - $1/(0.1 + 0.9/100) = 1/0.109 = 9.17$ (approx.)
  - At the limit
    - $Lt_{x \to \infty}$ (Speedup) = $1/(f + 0) = 1/f$

I/O - Performance Parameters

- I/O Rate
  - Number of I/O operations per unit time
- Data Transfer Rate
  - Affects Time Taken for completion of a single I/O operation.
- I/O Performance depends on various factors:
  - The specific I/O device
  - Processing Overhead
    - Architecture and Operating Systems Design
      - E.g. DMA performance, availability of buffer cache

Magnetic Memories

- Non-Volatile
  - Power not needed for retention
- Floppy Disks
  - Flexible, can be carried, limited size

Floppy Disk

- Originally
  - A single surface disk (8 in.)
- Later
  - 2 surfaces; 4X density; (5.25in.)
  - 2 surfaces; High density; (3.5 in)
  - 1.44MB capacity;
  - Access rate 10 ms (complete sector)
- Geometry:
  - 80 tracks per surface;
  - 18 sectors per track (160 x 18 = 2880 sectors)
  - 1 sector = 512 bytes

Magnetic Memories

- Hard Disks
  - Hard cased
  - Originally meant for PCs and mainframes
    - 14 in. diameter for mainframes in 60s
    - 3.5 in. diameter for PCs from 80's
  - Now available in various shapes:
    - Mini disks (2.5 in. dia.) for laptops, gaming consoles and as external pocket storage
      - 2.75 in × 0.275–0.59 in × 3.945 in (69.85 mm × 7–15 mm × 100 mm) = 48.895–104.775 cm^3
      - laptops 9.5mm and desktop 12.5mm
    - Micro disks (1.68 in. or 1.8 in. dia) for iPods / Cameras / other handheld devices
    - 0.85 in. form factor – mobile phones [SD/MMC slot compatible HDD optimized for video storage on 4G handset]

Magnetic Memories