# Exercise 1 – Overview

Questions are taken from Stallings, Operating Systems Internals and Design Principles, fifth edition.

## 1.7

In virtually all systems that include DMA modules, DMA acess to main memory is given higher priority than processor access to main memory. Why?

## 1.8

A DMA module is transferring characters to main memory from an external device transmitting at 9600 bits per second (bps) The processor can fetch instructions at the rate of 1 million instructions per second. By how much will the processor be slowed down due to the DMA activity?

## 1.9

A computer consists of a CPU and an I/O device D connected to main memory M via a shared bus with a data bus width of one word. The CPU can execute a maximum of $10^6$ instructions per second. An average instruction requires five machine cycles, three of which use the memory bus. A memory read or write operation uses one machine cycle. Suppose that the CPU is continuously executing "background" programs that require 95% of its instruction execution rate but not any I/O instructions. Assume that one processor cycle equals one bus cycle. Now suppose that very large blocks of data are to be transferred between M and D.

a) If programmed I/O is used and each one-word I/O transfer requires the CPU to execute two instructions, estimate the maximum I/O data transfer rate, in words per second, possible through D.

b) Estimate the same rate if DMA transfer is used.

## 1.11

Generalize these two equations for caching of memory to n-level memory hierarchies.

$$T_s = H \cdot T_1 + (1 - H) \cdot (T_1 + T_2) = T_1 + (1 - H) \cdot T_2 \tag{1}$$

$T_s$ = average (system) access time

$T_1$ = access time of $M_1$ (e.g., cache, disk cache)

$T_2$ = access time of $M_2$ (e.g., main memory, disk)

$H$ = hit ratio (fraction of time reference is found in $M_1$)

$$C_s = (C1 \cdot S1 + C2 \cdot S2)/(S1 + S2) \tag{2}$$

$C_s$ = average cost per bit for the combined two-level memory

$C_1$ = average cost per bit of upper-level memory $M_1$

$C_2$ = average cost per bit of lower-level memory $M_2$

$S_1$ = size of $M_1$

$S_2$ = size of $M_2$

## 1.13

A computer has a cache, main memory, and a disk used for virtual memory. If a referenced word is in the cache, 20 ns are required to access it. If it is in main memory but not in the cache, 60 ns are needed to load it into the cache (this includes the time to originally check the cache), and then the reference is started again. If the word is not in main memory, 12 ms are required to fetch the word from disk, followed by 60 ns to copy it to the cache, and then the reference is started again. The cache hit ratio is 0.9 and the main-memory hit ratio is 0.6. What is the average time in ns required to access a referenced word on this system?

## 2.1

Suppose that we have a multiprogrammed computer in which each job has identical characteristics. In one computation period, $T$, for a job, half the time is spent in I/O and the other half in processor activity. Each job runs for a total of $N$ periods. Assume that a simple round-robin scheduling is used, and that I/O operations can overlap with processor operation. Define the following quantities:

- Turnaround time = actual time to complete a job.
- Throughput = average number of jobs completed per time period $T$.
- Processor utilization = percentage of time that the processor is active (not waiting).

Compute these quantities for one, two, and four simultaneous jobs, assuming that the period T is distributed in each of the following ways:

a) I/O first half, processor second half

b) I/O first and fourth quarters, processor second and third quarter

## 2.2

An I/O-bound program is one that, if run alone, would spend more time waiting for I/O than using the processor. A processor-bound program is the opposite. Suppose a short-term scheduling algorithm favors those programs that have used little processor time in the recent past. Explain why this algorithm favors I/O-bound programs and yet does not permanently deny processor time to processor-bound programs.