Toby Chappell

CPSC 380 Operating Systems

M.W, 4:00-5:15pm

Homework #1

Part I – Overview

1. What are the three main purposes of an operating system? (6 pts)
   * Execute user programs and make solving user problems easier
   * Make the computer system convenient to use
   * Use the computer hardware in an efficient manner
2. Give two reasons why caches are useful. What problems do they solve? What problems do they cause? If a cache can be made as large as the device for which it is caching (for instance, a cache as large as a disk), why not make it that large and eliminate the device? (10 pts)

Caches are useful since they temporarily allow information to be accessed from faster storage. In addition, internal programmable registers can be used to provide high-speed cache for main memory. A problem with cache is that they are limited in size and as such cache management becomes an important design problem (determine cache replacement policy and size to increase performance). The reason why caches are not made large enough to replace the device in its entirety is because they are more expensive per byte compared to disks. Additionally, they consume more power which would cause the machine to need a larger supply of power in order to operate.

Caches are useful when two or more components need to exchange data, and the components perform transfers at differing speeds. Caches solve the transfer problem by providing a buffer of intermediate speed between the components. If the fast device finds the data it needs in the cache, it need not wait for the slower device.  
The data in the cache must be kept consistent with the data in the components.  
If a component has a data value change, and the datum is also in the cache, the cache must also be updated. This is especially a problem on multiprocessor systems where more than one process may be accessing a datum. A component may be eliminated by an equal-sized cache, but only if: (a) the cache and the component have equivalent state-saving capacity (that is, if the component retains its data when electricity is removed, the cache must retain data as well), and (b) the cache is affordable, because faster storage tends to be more expensive.

1. In a multiprogramming and time-sharing environment, several users share the system simultaneously. This situation can result in various security problems. (4 pts)
   1. What are two such problems?

Privacy: One user can read the private data of another user

Integrity: One user can corrupt the private data of another user

* 1. Can we ensure the same degree of security in a time-shared machine as in a dedicated machine?

Probably not, since any protection scheme devised by human scan inevitably be broken by a human, and the more complex the scheme, the more difficult it is to feel confident of its correct implementation.

1. What is the purpose of interrupts? What are the differences between a trap and an interrupt? (4 pts)

An interrupt signals the occurrence of an event (either prompted by the software or hardware). When the CPU is interrupted, it will immediately stop what it is doing and transfer execution to a fixed location. The difference between a trap (or exception) and an interrupt is that a trap is software driven (software error, request for OS service, etc.) and an interrupt is hardware driven.

1. Direct memory access is used for high-speed I/O devices in order to avoid increasing the CPU’s execution load. (6 pts)
   1. How does the CPU interface with the device to coordinate the transfer?

The CPU first sets up the DMA registers (including defining pointers to the source/destination for the transfer and a counter of the number of bytes to be transferred) and then the DMA controller puts the addresses on the bus to carry out the transfers.

* 1. How does the CPU know when the memory operations are complete?

The DMA controller interrupts the CPU.

* 1. The CPU is allowed to execute other programs while the DMA controller is transferring data. Does this process interfere with the execution of the user programs?

This would only become a problem if the CPU and the DMA controller want to access the same memory at the same time. To avoid this problem, the CPU should be prevented from accessing main memory when the DMA controller takes control of the memory bus.

1. Describe some of the challenges of designing operating systems for mobile devices compared with designing operating systems for traditional PCs. (3 pts)

In comparison with desktop/laptop, mobile systems have a decreased memory capacity at around 64 GB in comparison to 1 TB on a desktop. Additionally, mobile devices generally use processors that are smaller, slower, and have fewer processing cores in favor of power consumption.

1. What is the purpose of system calls in an operating system? (2 pts)

System calls allow for user access into the operating system without causing file corruption.

1. What is the purpose of the command interpreter? Why is it usually separate from the kernel? (3 pts)

The command interpreter allows users to directly enter commands to be performed by the operating system. Its main function is to get and execute the next user-specified command. It is normally separate from the kernel since the command interpreter is subject to changes.

1. What are the two models of inter-process communication? (2 pts)

Message-passing model: the communicating processes exchange messages with one another to transfer information.

Shared-memory model: processes use shared memory create() and shared memory attach() system calls to create and gain access to regions of memory owned by other processes.

1. List five services provided by an operating system. (5 pts)

* Program execution
* I/O operations
* File-system manipulation
* Communications
* Error detection