HW3 Solution

1. Explain the main differences between a short-term and long-term scheduler.

The primary distinction between the two schedulers lies in the frequency of execution. The short-term scheduler is designed to frequently select a new process for the CPU, at least once every 100 milliseconds. Because of the short time between executions, the short-term scheduler must be fast. The long-term scheduler executes much less frequently; minutes may separate the creation of one new process and the next. The long-term scheduler controls the degree of multiprogramming. Because of the longer interval between executions, the long-term scheduler can afford to take more time to decide which process should be selected for execution.

2. Explain the concept of a context switch.

Whenever the CPU starts executing a new process, the old process's state must be preserved. The context of a process is represented by its process control block. Switching the CPU to another process requires performing a state save of the current process and a state restore of a different process. This task is known as a context switch. When a context switch occurs, the kernel saves the context of the old process in its PCB and loads the saves context of the new process scheduled to run.

3. Explain the terms "at most once" and "exactly once" and indicate how they relate to remote procedure calls.

Because a remote procedure call can fail in any number of ways, it is important to be able to handle such errors in the messaging system. The term "at most once" refers to ensuring that the server processes a particular message sent by the client only once and not multiple times. This is implemented by merely checking the timestamp of the message. The term "exactly once" refers to making sure that the message is executed on the server once and only once so that there is a guarantee that the server received and processed the message.

4. Try to explain why Google Chrome uses multiple processes. (Efficiency? Reliability? Security?)

Each website opens up in a separate tab and is represented with a separate renderer process. If that webpage were to crash, only the process representing that the tab would be affected, all other sites (represented as separate tabs/processes) would be unaffected.

5. Name and describe the different states that a process can exist in at any given time.

The possible states of a process are: new, running, waiting, ready, and terminated. The process is created while in the new state. In the running or waiting state, the process is executing or waiting for an event to occur, respectively. The ready state occurs when the process is ready and waiting to be assigned to a processor and should not be confused with the waiting state mentioned earlier. After the process is finished executing its code, it enters the termination state.