HW19

UNIX "Fork" Command

The fork command is the **only** way to create a new process in UNIX operating systems. Once called, it duplicates the calling process, copying everything (excluding the process ID) from the original (parent) process, including its' address space, registers, files in use/open file descriptors, variables, as well as all variable values. The child process only copies the calling thread, as in the thread that invoked the fork. A parent-child relationship is established during a fork via the command's return value, returning the child's process ID to the parent, 0 to the child, and -1 if an error occurred during the fork operation. Since the child process is identical to the parent process, it will continue to operate **exactly** like the parent (executing onwards from the point of the fork) which demands that logic exists within the program's code which uses the fork return value to differentiate the child from the parent. It is also possible to execute a different program in the child process by running the exec/execve command, passing the desired program as a parameter, which will replace all of the contents of the child process with the desired program, in effect "creating" a completely new process.

Upon a successful fork, both the parent and child processes will be in the READY state, waiting for execution by the processor. Since a processor can only run a single process at a time, a scheduling algorithm will need to decide which process to run, which process to suspend, and and which process run next, treating both the newly created child process and original parent process as separate processes competing for execution time.

Works Cited

- A. S. Tanenbaum and H. Bos, Modern operating systems. Boston Massachusetts: Pearson, 2014.
- fork(2) Linux manual page. [Online]. Available: https://man7.org/linux/man-pages/man2/fork.2.html. [Accessed: 03-Jun-2021].
- fork. [Online]. Available: https://pubs.opengroup.org/onlinepubs/7908799/xsh/fork.html. [Accessed: 03-Jun-2021].