Virtual Memory

Readings: Chapter 8

Hardware and Control Structures

- Two characteristics fundamental to memory management:
  - All memory references are logical addresses that are dynamically translated into physical addresses at run time
  - A process may be broken up into a number of pieces that don’t need to be contiguously located in main memory during execution

- If these two characteristics are present, it is not necessary that all of the pages or segments of a process be in main memory during execution

Execution of a Process

- Operating system brings into main memory a few pieces of the program
  - Resident set
- An interrupt is generated when an address is needed that is not in main memory
- Operating system places the process in a blocking state
- Piece of process that contains the logical address is brought into main memory
  - Operating system issues a disk I/O Read request
  - Another process is dispatched to run while the disk I/O takes place
  - An interrupt is issued when disk I/O is complete, which causes the operating system to place the affected process in the Ready state

Implications

- More processes may be maintained in main memory. Why?
- This leads to more efficient utilization of the processor. Why?
- A process may be larger than all of main memory
  - The OS automatically loads pieces of a process into main memory as required
Real and Virtual Memory

- **Real Memory**: Main memory, the actual RAM
- **Virtual Memory**: Memory on disk

Allows for effective multiprogramming and relieves the user of tight constraints of main memory.

Principle of Locality

- Program and data references within a process tend to cluster
- Only a few pieces of a process will be needed over a short period of time
- Therefore it is possible to make intelligent guesses about which pieces will be needed in the future
- Avoids thrashing

Support Needed for Virtual Memory

- For virtual memory to be practical and effective:
  - Hardware must support paging and segmentation
  - Operating system must include software for managing the movement of pages and/or segments between secondary memory and main memory

Thrashing

A state in which the system spends most of its time swapping process pieces rather than executing instructions. To avoid this, the operating system tries to guess, based on recent history, which pieces are most likely to be used in the near future.