Software Categories

• Application software—software written to address specific needs—to solve problems in the real world
  —I.e. Internet Explorer (and other web browsers), Microsoft Word (and other word processors), business accounting software, the applets from “Computer Science Illuminated” text, etc.
• System software—software that manages a computer system at a fundamental level
  —I.e. Microsoft Windows, Unix/Linux, Mac OS

Roles of an Operating System

• Operating system—system software that:
  —Manages computer resources such as memory, CPU processing time and input/output devices
  —Provides an interface through which a human can interact with the computer
  —Allows an application program to interact with these other system resources

Roles of an Operating System

• The various roles of an operating system generally revolve around the idea of “sharing nicely”
• An operating system manages resources, and these resources are often shared in one way or another among programs that want to use them

Resource Management

• Memory management—the process of keeping track of what programs are in memory and where in memory they reside
• Multiprogramming—the technique of keeping multiple programs (or processes) that compete for access to the CPU in main memory at the same time so that they all can execute

Resource Management

• Process—a program (or a thread within a program) in execution
• Process management—the act of carefully tracking the progress of a process and all of its intermediate states
• CPU scheduling—determining which process in memory is executed by CPU at any given point

Batch Processing

• The first operating system was a human operator, who organize various jobs from multiple users into batches of jobs that needed the same resources

Timesharing

• As computer speed increased, the human operator became the bottleneck
• Timesharing system—a system that allows multiple users to interact with a computer at the same time
• Virtual machine—the illusion created by a time-sharing system that each user has his/her own machine

1963 Timesharing: A Solution to Computer Bottlenecks
• http://www.youtube.com/watch?v=Q07PhW5sCEk

Other Factors

• Real-time system—a system in which response time is crucial given the nature of the application
• Response time—the time delay between receiving a stimulus and producing a response
• Device driver—a small program that “knows” the way a particular device expects to receive and deliver information
14 Updating Drivers (the Easy Way)
   - [http://www.youtube.com/watch?v=z70UffBPuk&feature=related](http://www.youtube.com/watch?v=z70UffBPuk&feature=related)

15 Memory Management
   - Operating systems must employ techniques to:
     - Track where and how a program resides in memory
     - Convert logical addresses into actual addresses
   - Logical address—reference to a stored value relative to the program making the reference
   - Physical address—actual address in main memory

17 Single Contiguous Memory Management (Page 1)
   - There are only two programs in memory
     - The operating system
     - The application program
   - This approach is called single contiguous memory management (mm)

18 Single Contiguous Memory Management (Page 2)
   - In concrete terms:
     - A logical address is simply an integer value relative to the starting point of the program
     - A physical address is a logical address added to the starting location of the program in main memory

20 Partition Memory Management (Page 1)
   - Single contiguous MM has only the OS and one other program in memory at one time
   - Partition MM has the OS and any number of other programs in memory at one time
   - There are two schemes for dividing up memory for programs:
     - Fixed partitions—main memory is divided into a fixed number of partitions into which programs can be loaded
     - Dynamic partitions—partitions are created as needed to fit the programs waiting to be loaded

21 Partition Memory Management (Page 2)
   - Memory is divided into a set of partitions, some empty and some allocated to programs
   - Base register—a register that holds the beginning address of the current partition (the one that is running)
   - Bounds register—a register that holds the length of the current partition

23 Partition Selection Algorithms
   - Which partition should we allocate to a new program?
   - First fit—allocate program to the first partition big enough to hold it
   - Best fit—allocate program to the smallest partition big enough to hold it
   - Worst fit—allocate program to the largest partition big enough to hold it

25 Paged Memory Management (Page 1)
   - Paged memory technique—a technique in which processes are divided into fixed-size pages and stored in memory frames when loaded
   - Frame—a fixed-size portion of main memory that holds a process page
   - Page—a fixed-size portion of a process that is stored into a memory frame
   - (Assume that a frame and a page are the same size)

27 Paged Memory Management (Page 2)
   - Integer logical address is mapped into a <page number, offset> logical address
     - Page number—address divided by the page size (say 1024)
     - Offset—remainder of the address divided by page size
   - I.e.
—2566 DIV 1024 = 2
—2566 MOD 1024 = 518  ==> <2, 518>

Paged Memory Management (Page 3)
- This new logical address is mapped to a physical address with the help of a page-map table (PMT)
- Every program has a PMT that shows into which frame each page of the program is stored

Paged Memory Management (Page 4)
- Demand paging—an extension of paged memory management in which pages are brought into memory on demand
- Page swap—the act of bringing in a page from secondary memory, which often causes another page to be written back to secondary memory

Paged Memory Management (Page 5)
Virtual memory—the illusion that there are no restrictions on the size of a program because entire process does not have to be in memory at the same time
Thrashing—inefficient processing caused by constant page swaps

Process Management (Page 1)
- Process management—the act of managing the use of the CPU by individual processes
  —Recall that a process is a program in execution
- Considers which stages a process goes through

Process Management (Page 2)
- Process control block (PCB)—a data structure used by the OS to manage information about a process, including:
  —Current value of the program counter
  —Values of all CPU registers for the process
  —Base and bound register values (or page tables)
  —Accounting information
- Each state is represented by a list of PCBs, one for each process in that state

Process Management (Page 3)
- There is only one CPU and therefore only one set of CPU registers which can contain the values for the currently executing process
- Each time a process is moved to the running state:
  —Register values for the currently running process are stored into its PCB
  —Its PCB is moved to the list of the state into which it goes
  —Register values of the new process moving into the running state are loaded into the CPU
  —Exchange of register information is called a context switch

CPU Scheduling (Page 1)
- CPU Scheduling—the act of determining which process in the ready state should be moved to the running state
- Many processes may be in the ready state, but only one process can be in the running state, making progress at any one time

CPU Scheduling (Page 2)
- Nonpreemptive scheduling—currently executing process gives up the CPU voluntarily
- Preemptive scheduling—operating system decides to favor another process, preempting the currently executing process
- Turnaround time—the amount of time between when a process first arrives in the ready state and when it exits the running state for the last time

CPU Scheduling Algorithms (Page 1)
• First-Come, First-Served—processes are moved to the CPU in the order in which they arrive in the running state
• Shortest Job Next—process with the shortest estimated running time in the ready state is moved into the running state first

CPU Scheduling Algorithms (Page 2)
• Round Robin—each process runs for a specified time slice and moves from the running state to the ready state to await its next turn if not finished
  —Every process is treated the same
  —Time slice (quantum)—the amount of time that each process receives before being preempted and returned to ready state allowing another process its turn