What is an Operating System?

- A program that acts as an intermediary between a user of a computer and the computer hardware.
- Operating system goals:
  - Execute user program.
  - Make solving user problems easier. Make the computer system convenient to use.
  - Use the computer hardware in an efficient manner.

Computer System Components

1. Hardware – provides basic computing resources (CPU, memory, I/O devices).
2. Operating system – controls and coordinates the use of the hardware among the various application programs for the various users.
3. Applications programs – define the ways in which the system resources are used to solve the computing problems of the users (compilers, database systems, video games, business programs).
4. Users (people, machines, other computers).

Abstract View of System Components
Operating System Definitions

- Control program – controls the execution of users programs.
- I/O manager – manages and allocates I/O devices.
  → Kernel – the one program ready to run at all times (all else being application programs).

Early Systems

- Structure
  - Large machines run from console.
  - Single user system.
  - Programmer/User as operator.
  - Paper Tape or Punched cards.
  - No hard disk in computer.
- Early Software
  - Assemblers, Compilers, Libraries of common subroutines, Device Drivers, Linkers.
- Secure
- Low CPU utilization.
- Significant amount of setup time.

Multiprogramming Systems

Several jobs are kept in main memory at the same time, and the CPU is multiplexed among them.

The Kernel

- program1
- program2
- program3
- kernel
- I/O
- scheduler
- I/O request
- device driver
- scheduler
- Time slice exceeded
- scheduler
- Interrupt
- device driver
OS Features Needed for Multiprogramming

- I/O routine (device driver) supplied by the system.
- Memory management – the system must allocate the memory to several jobs.
- CPU scheduling – the system must choose among several jobs ready to run.
- Allocation of devices.

Spooling

- Overlap an execution of one job with a slow I/O of another job.
- While executing jobs, special daemons:
  - Outputs data of previous job from disk to a slow output device (e.g. printer).
  - Reads next job from slow input device (e.g. communication line) into a storage area on the disk.
- The storage area on the disk is called Spool

Time-Sharing Systems–Interactive Computing

- A synonymous definition for multiprogramming system.
- The CPU is multiplexed among several jobs that are kept in memory and on disk (the CPU is allocated to a job only if the job is in memory).
- A job swapped in and out of memory to the disk.
- On-line communication between the user and the system is provided; when the operating system finishes the execution of one command, it seeks the next “control statement” from the user’s keyboard.

Desktop Systems

- Personal computers (PC) – computer system dedicated to a single user.
- User convenience and responsiveness.
- Can adopt technology developed for larger operating system. Often individuals have sole use of computer and do not need advanced CPU utilization of protection features.
Parallel Systems

- Multiprocessor systems with more than one CPU in close communication.
- *Tightly coupled system* – processors share memory and a clock; communication usually takes place through the shared memory.
- Advantages of parallel systems:
  - Increased *throughput*.
  - Economical.
  - Increased reliability, fault-tolerant systems.

Parallel Systems (Cont.)

- *Symmetric multiprocessing (SMP)*
  - Each processor runs an identical copy of the operating system.
  - Many processes can run at once without performance deterioration.
  - Most modern operating systems support SMP
- *Asymmetric multiprocessing*
  - Each processor is assigned a specific task; master processor schedules and allocates work to slave processors.
  - More common in extremely large systems

Symmetric multiprocessing (SMP)

Distributed Systems

- Distribute the computation among several physical processors.
- *Loosely coupled system* – each processor has its own local memory; processors communicate with one another through various communications lines, such as high-speed buses or telephone lines.
- Advantages of distributed systems:
  - Resources sharing
  - Computation speed up – load sharing
  - Reliability
Real-Time Systems

• Hard real-time:
  – guaranteed worst-case response times.
  – Secondary storage limited or absent, data stored in short term memory, or read-only memory (ROM)
  – Conflicts with time-sharing systems, not supported by general-purpose operating systems.

• Soft real-time
  – A soft real-time system is one in which its users are happier when the system responds optimally, but which is not considered to have "failed" when the system doesn't meet each and every desired response time.
  – Useful in applications (multimedia, virtual reality) requiring advanced operating-system features.

Handheld Systems

• Personal Digital Assistants (Palms)
• Cellular telephones

• Features:
  – Limited memory
  – Slow processors
  – Small display screens.