

Operating Systems

Lecture No. 1

Reading Material

- Operating Systems Concepts, Chapter 1
- PowerPoint Slides for Lecture 1

Summary

- Introduction and purpose of the course
- Organization of a computer system
- Purpose of a computer system
- Requirements for achieving the purpose – Setting the stage for OS concepts and principles
- Outline of topics to be discussed
- What is an Operating System?

Organization of a Computer System

As shown in Figure 1.1, the major high-level components of a computer system are:

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

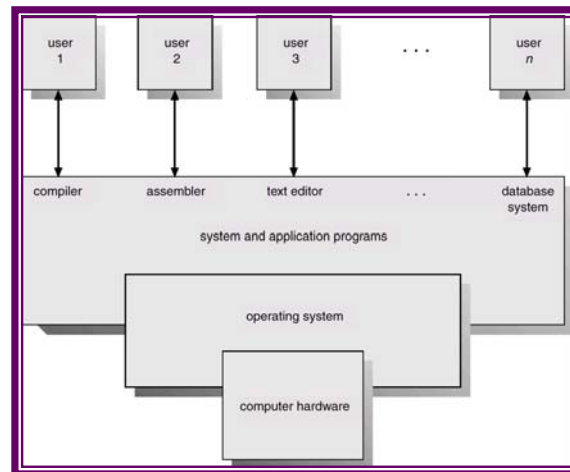


Figure 1.1. High-level components of a computer system

Purpose of a Computer—Setting the Stage for OS Concepts and Principles

Computer systems consist of software and hardware that are combined to provide a tool to implement solutions for specific problems in an efficient manner and to execute programs. Figure 1.2 shows the general organization of a contemporary computer system and how various system components are interconnected.

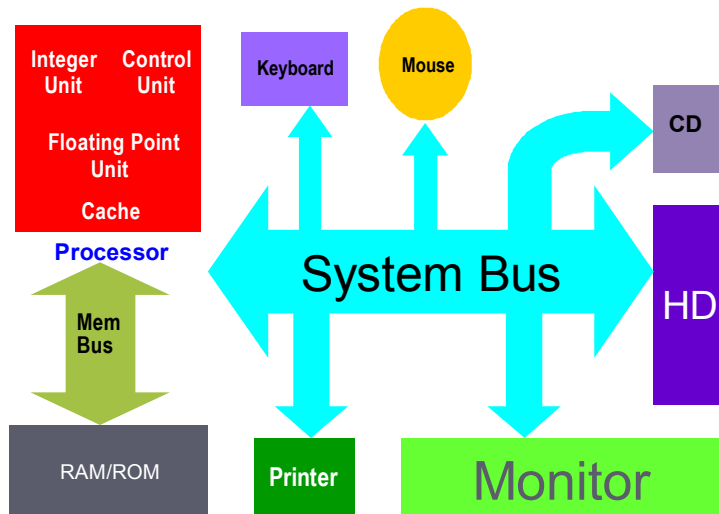


Figure 1.2. Organization of a Computer System

Viewing things closely will reveal that the primary purpose of a computer system is to generate executable programs and execute them. The following are some of the main issues involved in performing these tasks.

1. Storing an executable on a secondary storage device such as hard disk
2. Loading executable from disk into the main memory
3. Setting the CPU state appropriately so that program execution could begin
4. Creating multiple cooperating processes, synchronizing their access to shared data, and allowing them to communicate with each other

The above issues require the operating system to provide the following services and much more:

- Manage secondary storage devices
 - Allocate appropriate amount of disk space when files are created
 - Deallocate space when files are removing
 - Insure that a new file does not overwrite an existing file
 - Schedule disk requests
- Manage primary storage
 - Allocate appropriate amount of memory space when programs are to be loaded into the memory for executing
 - Deallocate space when processes terminate
 - Insure that a new process is not loaded on top of an existing process
 - Insure that a process does not access memory space that does not belong to it
 - Minimize the amount of unused memory space
 - Allow execution of programs larger in size than the available main memory
- Manage processes

- Allow simultaneous execution of processes by scheduling the CPU(s)
- Prevent deadlocks between processes
- Insure integrity of shared data
- Synchronize executions of cooperating processes
- Allow a user to manage his/her files and directories properly
 - User view of directory structure
 - Provide a mechanism that allows users to protect their files and directories

In this course, we will discuss in detail these operating system services (and more), with a particular emphasis on the UNIX and Linux operating systems. See the course outline for details of topics and lecture schedule.

What is an Operating System?

There are two views about this. The top-down view is that it is a program that acts as an intermediary between a user of a computer and the computer hardware, and makes the computer system convenient to use. It is because of the operating system that users of a computer system don't have to deal with computer's hardware to get their work done. Users can use simple commands to perform various tasks and let the operating system do the difficult work of interacting with computer hardware. Thus, you can use a command like `copy file1 file2` to copy 'file1' to 'file2' and let the operating system communicate with the controller(s) of the disk that contain(s) the two files.

A computer system has many hardware and software resources that may be required to solve a problem: CPU time, memory space, file storage space, I/O devices etc. The operating system acts as the manager of these resources, facing numerous and possibly conflicting requests for resources, the operating system must decide how (and when) to allocate (and deallocate) them to specific programs and users so that it can operate the computer system efficiently, fairly, and securely. So, the bottom-up view is that operating system is a resource manager who manages the hardware and software resources in the computer system.

A slightly different view of an operating system emphasizes the need to control the various I/O devices and programs. An operating system is a control program that manages the execution of user programs to prevent errors and improper use of a computer.