Study Session 1: Overview of Operating Systems

Introduction

Have you ever wondered or thought about how easy it is for you to communicate with the computer without knowing how to speak the computer's "language"? Operating System is the special software that makes this possible in a computer system. Without an operating system, a computer is useless.

In this study session, you will learn the basic knowledge on how an Operating System works. The purpose of an operating system is to provide an environment in which a user can execute computer programs in a convenient and efficient manner.

Learning Outcomes for Study Session 1

On completion of this study session, you should be able to:

1.1 Define and explain the meaning of Operating Systems

1.2 Highlight the functions of an Operating System

1.3 Discuss the types of Operating System

1.4 Explain the Operating System Services

1.1 Definition of Operating System

An Operating System (OS) is an intermediary between computer users and the computer hardware. It provides an environment in which a user can execute programs conveniently and efficiently using natural languages.

Box 1.1: Operating System

An Operating System is a program that acts as an interface between the user and the computer hardware and controls the execution of all kinds of programs.

What is an Operating System?

In technical terms, it is software which manages hardware. An operating System controls the allocation of resources and services such as memory, processors, devices and information. Users usually interact with operating system using operating system commands while programs invoke the service of operating system using Operating System Call.

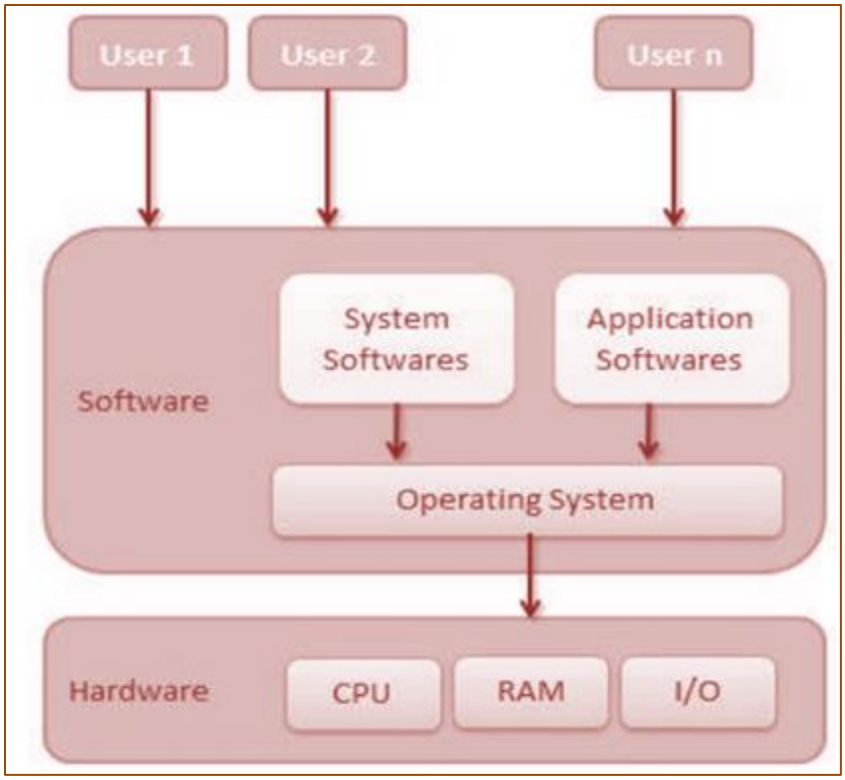

Figure 1.1: Relationship between User, Operating system, Application Software and Hardware

Pilot Question 1.1

Explain Operating System as discussed in this study session

1.2 Functions of Operating System

The following are some of the important functions of an operating System:

i. Memory Management
ii. Processor Management
iii. Device Management
iv. File Management
v. Security
vi. Control over system performance
vii. Job Accounting
viii. Error detecting aids

ix. Coordination between other software and users

1.2.1 Memory Management

Memory management refers to management of Primary Memory or Main Memory. Main memory is a large array of words or bytes where each word or byte has its own address.

What is memory management?

Main memory provides a fast storage that can be accessed directly by the CPU. So for a program to be executed, it must be in the main memory. Operating System does the following activities for memory management:

i. Keeps track of primary memory i.e. what part of it are in use by whom, what part are not in use.
ii. In multiprogramming, OS decides which process will get memory when and how much.
iii. Allocates the memory when the process requests it to do so.
iv. De-allocates the memory when the process no longer needs it or has been terminated.

1.2.2 Processor Management

In multiprogramming environment, Operating System decides which process gets the services of the processor, when and how much time. This function is called process scheduling. Operating System does the following activities for processor management:

i. Keeps track of processor and status of process. Program responsible for this task is known as traffic controller.
ii. Allocates the processor (CPU) to a process.
iii. De-allocates processor when processor is no longer required.

1.2.3 Device Management

OS manages device communication via their respective drivers. Operating System does the following activities for device management.

i. Keeps track of all devices. Program responsible for this task is known as the I/O controller.
ii. Decides which process gets the device when and for how much time.
iii. Allocates the device in the efficient way.
iv. De-allocates devices

1.2.4 File Management

A file system is normally organized into directories for easy navigation and usage. These directories may contain files and other directions. Operating System does the following activities for file management.

i. Keeps track of information, location, uses, status etc. The collective facilities are often known as file system.
ii. Decides who gets the resources.
iii. Allocates the resources.
iv. De-allocates the resources.

1.2.5 Other Important Activities/Functions

The following are some of the important activities that Operating System does:

i. Security -- By means of password and similar other techniques, preventing unauthorized access to programs and data.
ii. Control over system performance -- Recording delays between request for a service and response from the system.
iii. Job accounting -- Keeping track of time and resources used by various jobs and users.
iv. Error detecting aids -- Production of dumps, traces, error messages and other debugging and error detecting aids.
v. Coordination between other software and users -- Coordination and assignment of compilers, interpreters, assemblers and other software to the various users of the computer systems.

Pilot Questions 1.2

i. State the functions of Operating System
ii. Define main memory
iii. Differentiate between memory management and processor management functions of an Operating System

1.3 Types of Operating Systems

Operating systems are there from the very first computer generation. Operating systems keep evolving over the period of time. Following are few of the important types of operating system which are most commonly used.

1.3.1 Batch Operating System

The users of batch operating system do not interact with the computer directly. Each user prepares his job on an off-line device like punch cards and submits it to the computer operator.

To speed up processing, jobs with similar needs are batched together and run as a group. Thus, the programmers left their programs with the operator. The operator then sorts programs into batches with similar requirements.

The problems with Batch Systems are the following:

i. Lack of interaction between the user and job.
ii. CPU is often idle, because the speeds of the mechanical I/O devices are slower than CPU.
iii. Difficult to provide the desired priority.

1.3.2 Time-sharing Operating Systems

Time sharing is a technique which enables many people, located at various terminals, to use a particular computer system at the same time. Time-sharing or multitasking is a logical extension of multiprogramming. Processor's time which is shared among multiple users simultaneously is termed as time-sharing.

The main difference between Multiprogrammed Batch Systems and Time-Sharing Systems is that in case of multi-programmed batch systems, objective is to maximize processor use, whereas in Time-Sharing Systems objective is to minimize response time.

Multiple jobs are executed by the CPU by switching between them, but the switches occur so frequently. Thus, the user can receive an immediate response. For example, in a transaction processing, processor execute each user program in a short burst or quantum of computation.

That is if n users are present, each user can get time quantum. When the user submits the command, the response time is in few seconds at most. OS uses CPU scheduling and multiprogramming to provide each user with a small portion of time. Computer systems that were designed primarily as batch systems have been modified to time-sharing systems.

Advantages of Timesharing operating systems are following:

i. Provide advantage of quick response.
ii. Avoids duplication of software.
iii. Reduces CPU idle time.

Disadvantages of Timesharing operating systems are following.

i. Problem of reliability.
ii. Question of security and integrity of user programs and data.
iii. Problem of data communication.

1.3.3 Distributed Operating System

Distributed systems use multiple central processors to serve multiple real time application and multiple users. Data processing jobs are distributed among the processors accordingly to which one can perform each job most efficiently.

The processors communicate with one another through various communication lines (such as high-speed buses or telephone lines). These are referred to as loosely coupled systems or distributed systems. Processors in a distributed system may vary in size and function. These processors are referred as sites, nodes, and computers and so on.

The advantages of distributed systems are as follows:

i. With resource sharing facility, user at one site may be able to use the resources available at another.

ii. Speedup the exchange of data with one another via electronic mail.
iii. If one site fails in a distributed system, the remaining sites can potentially continue operating.
iv. Better service to the customers.
v. Reduction of the load on the host computer.
vi. Reduction of delays in data processing

The disadvantages are;

i. Communication Protocol Overhead.
ii. Insufficient Simplicity.
iii. High dependence on their education associated with fault tolerance.
iv. Lack of global state information
v. Atomic Transactions.
vi. Process and Data Migration.

1.3.4 Network Operating System

Network Operating System runs on a server and provides server the capability to manage data, users, groups, security, applications, and other networking functions. The primary purpose of the network operating system is to allow shared file and printer access among multiple computers in a network, typically a local area network (LAN), a private network or to other networks.

Examples of network operating systems are Microsoft Windows Server 2003/2008/2010, UNIX, Linux, Mac OS X, Novell NetWare, and BSD.

The advantages of network operating systems are as follows:

i. Centralized servers are highly stable.
ii. Security is server managed.
iii. Upgrades to new technologies and hardware can be easily integrated into the system.
iv. Remote access to servers is possible from different locations and types of systems.

The disadvantages of network operating systems are as follows:

i. High cost of buying and running a server.
ii. Dependency on a central location for most operations.
iii. Regular maintenance and updates are required.

1.3.5 Real Time Operating System

Real time system is defined as a data processing system in which the time interval required to process and respond to inputs is so small that it controls the environment. Real time processing is always on line whereas on line system needs not be real time. The time taken by the system to respond to an input and display of required updated information is termed as response time. So in this method, response time is very less as compared to the online processing.

Define real time operating system

Real-time systems are used when there are rigid time requirements on the operation of a processor or the flow of data and real-time systems can be used as a control device in a dedicated application. Real-time operating system has well-defined, fixed time constraints otherwise system will fail. For example Scientific experiments, medical imaging systems, industrial control systems, weapon systems, robots, and home-appliance controllers, Air traffic control system etc.

Box 1.3: Real Time System

Real time system is defined as a data processing system in which the time interval required to process and respond to inputs is so small that it controls the environment.

There are two types of real-time operating systems.

I. Hard Real-Time Systems

Hard real-time systems guarantee that critical tasks complete on time. In hard real-time systems, secondary storage is limited or missing with data stored in ROM. In these systems, virtual memory is almost never found.

II. Soft Real-Time Systems

Soft real time systems are less restrictive. Critical real-time task gets priority over other tasks and retains the priority until it completes. Soft real-time systems have limited utility than hard real-time systems.

For example, In Multimedia, virtual reality, Advanced Scientific Projects like undersea exploration and planetary rovers, etc.

1.3.6 Multiprogramming Operating System

To overcome the problem of underutilization of CPU and main memory, the multiprogramming was introduced. The multiprogramming is interleaved execution of multiple jobs by the same computer.

In multiprogramming system, when one program is waiting for I/O transfer; there is another program ready to utilize the CPU. So, it is possible for several jobs to share the time of the CPU. But it is important to note that multiprogramming is not defined to be the execution of jobs at the same instance of time. Rather, it does mean that there are a number of jobs available to the CPU (placed in main memory) and a portion of one is executed then a segment of another and so on. A simple process of multiprogramming is shown in figure 1.2 below.

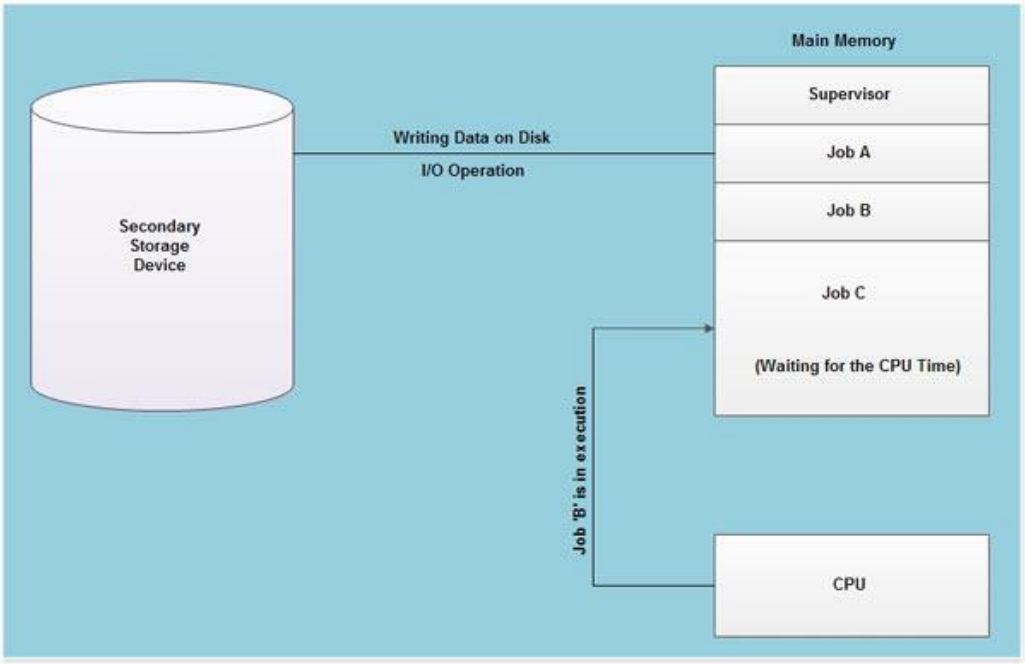

Figure 1.2: Multiprogramming Operating System

As shown in figure 1.2, at the particular situation, job' A' is not utilizing the CPU time because it is busy in I/O operations. Hence the CPU becomes busy to execute the job 'B'. Another job C is waiting for the CPU for getting its execution time. So in this state the CPU will never be idle and utilizes maximum of its time.

A program in execution is called a "Process", "Job" or a "Task". The concurrent execution of programs improves the utilization of system resources and enhances the system throughput as compared to batch and serial processing. In this system, when a process requests some I/O to allocate; meanwhile the CPU time is assigned to another ready process. So, here when a process is switched to an I/O operation, the CPU is not set idle.

Multiprogramming is a common approach to resource management. The essential components of a single-user operating system include a command processor, an input/ output control system, a file system, and a transient area. A multiprogramming operating system builds on this base, subdividing the transient area to hold several independent programs and adding resource management routines to the operating system's basic functions.

Pilot Question 1.3

Differentiate between Distributed Operating System and Network Operating System

1.4 Operating System Services

An Operating System provides services to both the users and to the programs. It provides programs, an environment to execute. Also, it provides users, services to execute the programs in a convenient manner.

The following are few general services provided by operating systems.

1. Program execution
2. I/O operations
3. File System manipulation
4. Communication
5. Error Detection
6. Resource Allocation
7. Protection

1.4.1 Program Execution

Operating system handles many kinds of activities from user programs to system programs like printer spooler, name servers, file server etc. Each of these activities is encapsulated as a process. A process includes the complete execution context (code to execute, data to manipulate, registers, OS resources in use). The following are the major activities of an operating system with respect to program management.

1. Loads a program into memory.
2. Executes the program.
3. Handles program's execution.
4. Provides a mechanism for process synchronization.
5. Provides a mechanism for process communication.
6. Provides a mechanism for deadlock handling.

1.4.2 I/O Operation

I/O subsystem comprised of I/O devices and their corresponding driver software. Drivers hides the peculiarities of specific hardware devices from the user as the device driver knows the peculiarities of the specific device.

Operating System manages the communication between user and device drivers. The following are the major activities of an operating system with respect to I/O Operation.

1. I/O operation means read or write operation with any file or any specific I/O device.
2. Program may require any I/O device while running.
3. Operating system provides the access to the required I/O device when required

Pilot Question 1.4

How does an Operating System enhance execution of programs?

Summary of Study Session 1: Overview of Operating System

In this Study Session 1, you have learnt that:

1. An operating System (OS) is an intermediary between users and computer hardware. It provides users an environment in which a user can execute programs conveniently and efficiently.It is software which manages hardware. OS controls the allocation of resources and services such as memory, processors, devices and information
2. Functions of OS include;
   Memory Management, Processor Management, Device Management, File Management, Security, Control over system performance, Job accounting, Error detecting aids, Coordination between other software and users
3. Types of OS are;
   Batch operating system, Time-sharing operating systems, Distributed operating System, Network operating System, and Real Time operating System
4. OS Services include;
   Program Execution, I/O operations, File System manipulation, Communication, Error Detection, Resource Allocation and Protection

Glossary of Terms

Hard Real-Time Systems: Hard real-time systems guarantee that critical tasks complete on time. In hard real-time systems secondary storage is limited or missing with data stored in ROM. In these systems virtual memory is almost never found.

Main memory: It is a large array of words or bytes where each word or byte has its own address.

Soft Real-Time Systems: Soft real time systems are less restrictive. Critical real-time task gets priority over other tasks and retains the priority until it completes. Soft real-time systems have limited utility than hard real-time systems.

Pilot Answers

Pilot Answer 1.1

i. An operating System (OS) is an intermediary between users and computer hardware. It provides users an environment in which a user can execute programs conveniently and efficiently. Pilot Answer 1.2

   i. The functions of Operating System are;

      a. Memory Management
      b. Processor Management
      c. Device Management
      d. File Management
      e. Security
      f. Control over system performance
      g. Job accounting
      h. Error detecting aids
      i. Coordination between other software and users

ii. Main memory is a large array of words or bytes where each word or byte has its own address.

iii. In Memory Management, Operating System decides which process will get memory when and how much. It allocates the memory when the process requests it to do so while in Processor Management, Operating System decides which process gets the processor when and how much time

Pilot Answer 1.3

i. Distributed systems use multiple central processors to serve multiple real time application and multiple users while Network Operating System runs on a server and provides server the capability to manage data, users, groups, security, applications, and other networking functions

Pilot Answer 1.4

i. An Operating System aids execution of programs through the following: Loads a program into memory, Executes the program, Handles program's execution, Provides mechanism for process synchronization, Provides mechanism for process communication, Provides mechanism for deadlock handling

Self-Assessment Questions (SAQs) for Study Session 1

Now that you have completed this study session, you can assess how well you have achieved its Learning Outcomes by answering these questions. Write your answers in your Study Diary and discuss them with your Tutor at the next Study Support Meeting. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

1. Explain Operating System as discussed in this study session
2. State the functions of Operating System
3. Describe the disadvantages of the following types of Operating System

i. Batch Operating System

ii. Time Sharing Operating System

4. Highlight the common services provided by operating systems