1. Introduction to Operating system

An operating system acts as an intermediary between the user of a computer and the computer hardware. The purpose of an operating system is to provide an environment in which a user can execute programs in a convenient and efficient manner. An operating system is software that manages the computer. It must provide appropriate mechanisms to ensure the correct operation of the computer system and to prevent user programs from interfering with the proper operation of the system.

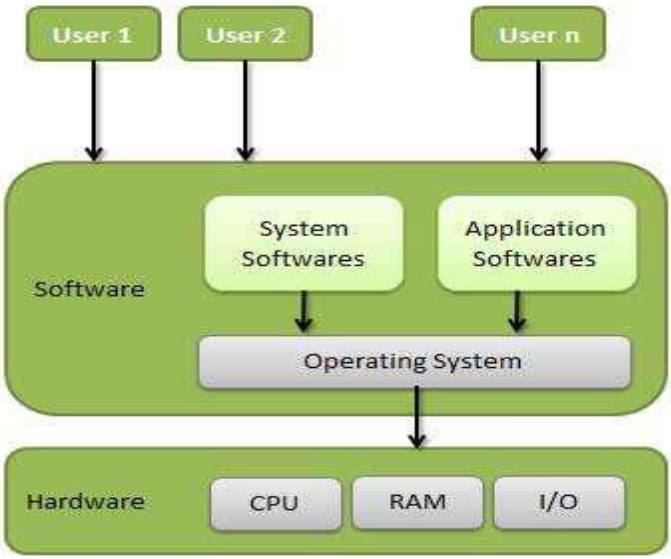
Definition

- An Operating system is a program that controls the execution of application programs and acts as an interface between the user of a computer and the computer hardware.

An Operating system is concerned with the allocation of resources and services, such as memory, processors, devices and information. The Operating System correspondingly includes programs to manage these resources, such as a traffic controller, a scheduler, memory management module, I/O programs, and a file system.

Overview

Every general purpose computer consists of the hardware, operating system, system programs, and application programs. The hardware consists of memory, CPU, I/O devices, peripheral device and storage device. System program consists of compilers, loaders, editors, OS etc. The application program consists of business program and database program. Below is a conceptual view of a computer system.



Every computer must have an operating system to run other programs. The operating system coordinates the use of the hardware among the various system programs and application program for various users. It simply provides an environment within which other programs can do useful work.

The operating system is a set of special programs that run on a computer system that allow it to work properly. It performs basic tasks such as recognizing input from the keyboard, keeping track of files and directories on the disk, sending output to the display screen and controlling peripheral devices.

OS is designed to serve **two basic purposes/objectives of Operating System**:

1. Allocation of resources.

It controls the allocation and use of the computing system's resources among the various user and tasks, i.e. controls how processes may access the resources.

2. Hide details of hardware by creating abstractions.

It provides an interface between the computer hardware and the programmer thus simplifies and makes it feasible for coding, creation, and debugging of application programs.

Functions of an operating system

There are a number of important functions of an operating system. They include;

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.

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

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

2. **Process Management**

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

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

3. Device Management

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

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

4. File Management

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

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

5. Provision of user interface

User interface controls how user enter data and instructions and how information is displayed on the screen. Operating system provides user interface through three (3) ways;

Command line- where users type commands using keyboards

Menu driven/based –users select commands as means of entering commands. Menus with listings of commands are provided, from which users select.

Graphical user interface- users use icons, buttons and other graphical objects to issue commands.

Other Important Activities

Following are some of the important activities that Operating System does.

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

History of Operating System

Operating systems have been evolving through years in generations.

1st Generation – 1940s

The earliest electronic digital computers had no operating system. They were very large and used vacuum tubes. Programs were entered manually one bit at a time and on machine/plug boards.

2nd Generation – 1950s

With the implementations of transistors, operations improved with introduction of punch cards and the operating systems used to run one job at a time and are referred to as single stream batch processing system because data and programs were submitted in forms of batches.

3rd Generation- 1960s

Batch processing systems were also used but with the use of Integrated Circuits (IC) the systems were able to take advantage of computer resources by running several jobs at once. The concept of multiprogramming was developed where several jobs were loaded into memory and the processor could switch from one job to another. Also timesharing technique, a variant of multiprogramming, where each user was directly connected to the system through a terminal, was stimulated.

4th Generation- 1970s

Operating systems of this generation were developed as a result of Very Large Scale Integration (VLSI) circuits. Microprocessor technology evolved to the point that desktop computers as powerful as mainframes were built. Operating systems were developed which dominated the personal computer scene.

Types of operating systems

There are a few important types of operating systems. They include;

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 following.

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

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 Multi-programmed 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. Operating system use 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

- Provide advantage of quick response.
- Avoids duplication of software.
- Reduces CPU idle time.

Disadvantages of Timesharing operating systems are following.

- Problem of reliability.
- Question of security and integrity of user programs and data.
- Problem of data communication.

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 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 following.

- With resource sharing facility user at one site may be able to use the resources available at another.
- Speedup the exchange of data with one another via electronic mail.
- If one site fails in a distributed system, the remaining sites can potentially continue operating.
- Better service to the customers.
- Reduction of the load on the host computer.
- Reduction of delays in data processing.

4. Network operating System

Network Operating System runs on a server and provides server with 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 other

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

The advantages of network operating systems are following.

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

The disadvantages of network operating systems are following.

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

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 need 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 little as compared to the online processing.

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.

Services of an Operating System

An Operating System provides services to both the users and the programs.

- It provides programs, an environment to execute.
- It provides users, services to execute the programs in a convenient manner.

Common services include;

1. Program execution

Operating system with respect to program management does the following; loads a program into memory, executes the program, handles program execution, provides a mechanism for process

synchronization, provides a mechanism for process communication, and provides a mechanism for deadlock handling.

2. I/O Operation

An operating system with respect to I/O Operation does the following; I/O operation, means read or write operation with any file or any specific I/O, issuance of I/O device to a program while running, and provides the access to the required I/O device when required.

3. File system manipulation

Major activities of OS in file management include; Program needs to read a file or write a file and thus the operating system gives the permission to the program for operation on file- Permission varies from read-only, read-write, denied and so on. It provides an interface to the user to create/delete files. It provides an interface to the user to create/delete directories, and it provides an interface to create the backup of file system.

4. Communication

Two processes often require data to be transferred between them and both processes can be on the one computer or on different computers but are connected through computer network. Thus OS may implement communication by two methods, either by Shared Memory or by Message Passing.

5. Error handling

OS constantly remains aware of possible errors and takes the appropriate action to ensure correct and consistent computing.

6. Resource Management

OS manages all kind of resources using schedulers and CPU scheduling algorithms are used for better utilization of CPU.

7. Protection

OS ensures that all access to system resources is controlled, that external I/O devices are protected from invalid access attempts and provides authentication feature for each user by means of a password.