Operating System

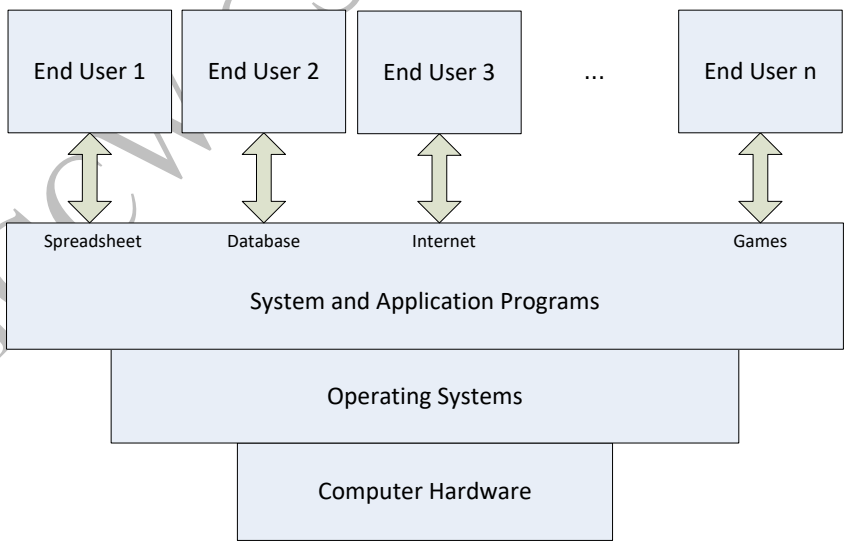
Introduction

An operating system is a program that controls the execution of the application programs and act as an interface between the user of the computer and the computer hardware

A computer is build with a set of hardware components properly linked together to interact in such a way that the total system is reliable and accurate. The operating system can be defined as the set of programs that controls the hardware and provide services for application software. For example, when a user requests the computer to perform an operation namely printing a document or executing a program, the operating system takes the task of controlling the software and hardware to give the preferred output.

Computer system is generally serialized into hardware, system software, application software, and user.

The following essentially gives the different layers of computer system.



Layers of Computer System

The *hardware* components provide the basic computing resources for the system. The *system software* creates environment for programmers to create software applications that suit their own needs. It acts as an intermediate between the computer hardware and the application software. The *application software* specifies the ways in which these resources are used to solve user's computing problems. The operating system controls the hardware and coordinates its use among the application programs for the different users. The operating system hides the features of hardware components.

The primary goal of an operating system is to make the system convenient to the user. A secondary level is the efficient operation of a computer system. Thus operating system acts upon the following three main functions:

- **Ease of use:** An operating system allows computer for convenient usage of resources.
- **Effectiveness**: An operating system makes resources to be used in an effectual manner.
- **Ability to progress**: An operating system should be built in such a way so that it supports hardware upgrades with new services and fault fixing.

Operating system can be viewed as a resource allocator. System resources are namely CPU time, memory space, file storage space, input and output devices etc.

Functionalities of Operating System:

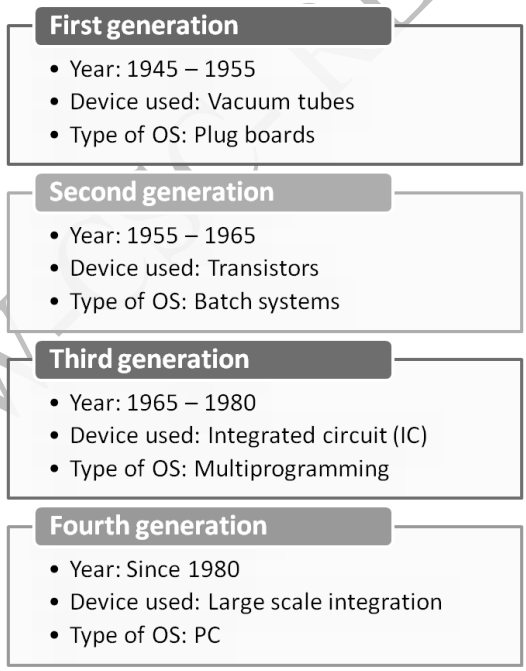
In order to build up a totally well performed computer system, it is one of most important responsibilities of operating system designers to control work at the contiguous levels. The responsibilities of the operating system are listed below:

- When the computer is turned on, operating system ensure for initial setup of the computer system.
- Whenever a user request for any resource or I/O, operating system permits system resources to be shared among users and protects from unauthorized intervention.

- Operating system schedules the demand on system resources when multiple-user applications operating system requested.
- Operating system records the sequence of events that lead to error condition.
- Operating system provides access to compilers to translate programs from high-level languages to machine language.
- Operating system ensures long term storage of user information in the form of files.
- If multiple programs run at same time, the operating system selects which applications should run first and how much time should be allocated to each application before giving access to another application.

History of Operating System

The advent of operating system has been evolving through the years. The following *fig 1.9* depicts the four generations with usage of devices and types of operating system employed.



History of Operating System

Mainframe Systems

Operating system manages a computer system's internal memory, processors, devices, and file system. Mainframe systems are principally designed for optimize utilization of hardware that tackles many commercial and scientific applications. The execution of a process is done serially

or concurrently by the operating system. The resources are allocated to a single program until its completion. They may also be reallocated dynamically to a group of active programs in various stages of execution. Mainframe systems have several divergences of both serial and multi-programmed operating system.

Mainframe systems have the following distinct features:

- The main frame computers are used to deal with different applications and found today in corporate data centre
- Main focus is on input and output bound business data applications.

The three main functions of mainframe Systems are:

1. Batch Processing
2. Transaction Processing
3. Time Sharing

- **Batch Processing**

In batch processing systems, a series of programs are executed in a queue without user interaction. Set of data files are taken as input and processed to produce the output data files. All input files are processed in batches by the program.

Memory is divided into two areas namely *System*, and *User space.* Scheduling is simple by which the jobs are processed in the order of FCFS (First Come First Serve) manner. After execution, the memory is released and output is copied into output spool. Spooling is a special area in memory to place jobs in a buffer for easy access of device. It utilizes the disk as a large buffer for printing data through printers and other devices. Spooling reduces idle (unused) time and overlapped I/O and CPU.

- **Transaction Processing**

In transaction processing, the data is collected, stored, modified and retrieved by the system. The system reacts immediately to the user request. Each request is considered as a *transaction*.

Every transaction must be either executed successfully or failed completely; it cannot remain in an unstable state.

Errors in transaction system are handled in safe and consistent manner. There are certain errors which cannot be avoided (e.g. Network errors or Database deadlocks).

- **Time Sharing**

In time sharing, resources are shared among many users. The computer's time is scheduled so that they are shared across multiple tasks and multiple users. The system is used to support interactive users for CPU scheduling and multiprogramming. A common form of multitasking is time sharing. But multiprogramming is not a time sharing system.

High priority is given to the interactive process. For example, consider an I/O request. If a key is pressed, the CPU control is given to the appropriate process to perform the task. The process switches rapidly form one user to the next user.

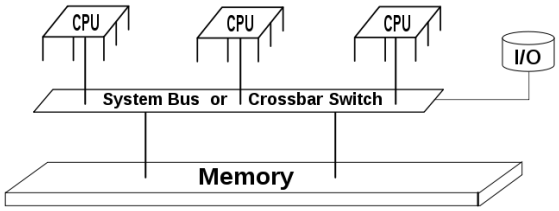
Multiprocessor Systems

There is more than one processor in multiprocessor system that can communicate with each other to allocate task between them. In multiprocessor systems, there exists two or more CPU in a single computer system. The computer bus, system clock and the input output devices are shared in multiprocessor system. Even sometimes, the memory is shared among processes. It is also possible for two or more processes to run in parallel.

Multiprocessor systems are of two types:

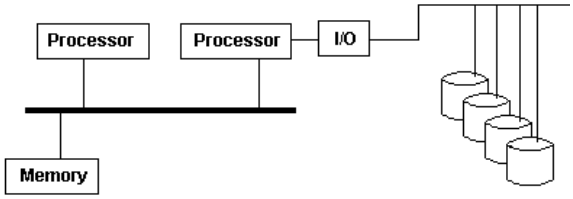
- Symmetric multiprocessing or Symmetric multiprocessor (SMP)
- Asymmetric multiprocessing or Asymmetric multiprocessor (ASMP)

Symmetric multiprocessing refers to an architecture in which one or more processors access the same shared memory. Every identical processor communicates with one another provided that each process in the system is not in execution at the same time. Nowadays many multiprocessor systems use SMP architecture.



SMP System

In asymmetric multiprocessing, the processor are different from each other with different physical memory space. Each processor is assigned with a specific task. For example, one processor may be allocated to disk operations, another processor to video operations, and the remaining processors to standard processor tasks. This system does not provide the flexibility to allocate processes to the least-loaded CPU, unlike an SMP system.



ASMP System

Because of the difficulty and unique nature of this architecture, it was not implemented by many vendors or programmers during 1970 – 1980

Features of multiprocessor systems

- Multiple processors share the common memory. This reduces the cost and resources.
- If one processor in the system fails, then other processors should retrieve the interrupted process state so that execution of the process can continue. The system automatically reschedules the failed job.
- The processors always maintain efficient context switching operation.
- Multiprocessor system retains large physical address space and large virtual address space.
- The IPC mechanism is implemented in hardware which is very effective and easy to use.

- If the number of processors is increased, the task will be completed in less time by increasing the throughput.

Types of Systems

The following are the types of systems in operating system.

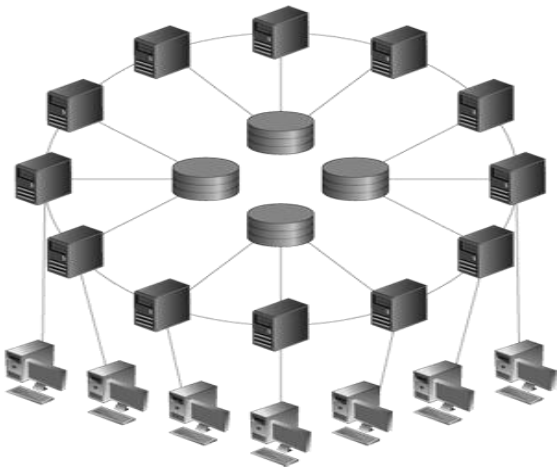
Distributed Systems

A distributed operating system is one that looks to its users like an ordinary operating system but runs on multiple, independent CPU. A distributed system is a set of interconnected, autonomous computers that cooperatively solve large, single problems or facilitate parallel execution of separate, but possibly related task.

Distributed systems refers a structure with a set of computers connected together using physical or wireless links. Links are considered to be fast enough to facilitate efficient information interchange especially between concurrently running processes.

The most vital goal is to use the power of many computers within the distributed system to run concurrent parts of a complex and time-consuming process such as weather forecasting, in order to provide the required information in time.

Several computers and personal computers are connected together, but they are not being used all the time. The processing power can be used to solve large and complex problems. It is not only the processing power than can be used towards solving a common problem, but rather most of the resources can be used, for this purpose. The main memory of all computers may collectively be used as a common **Distributed Shared Memory** (DSM).



Distributed Systems

Examples of distributed systems are given below:

- Amoeba
- Chorus
- Mach
- v-system
- SUN Network File System (NFS)
- CODA
- Plan 9
- XFS
- SFS

Advantages of Distributed Systems

- Distributed systems provides various file system services which depends on specification and interface that may include storage, access, structure, protection of files on multiple machines.
- They allow multiple heterogeneous file to share a local file system.
- The resources namely software libraries, database and hardware resources such as hard disks, printers and CD-ROM can be shared among all computers and the users.

- Hard disk availability is increased by having multiple disks located at different sites. Thus if one hard disk fails or is unavailable the program can use some other hard disk.
- The system power and functionality can be increased by simply adding the additional resources to the system.

Disadvantages of Distributed Systems

- It is very difficult to design the distributed system than the centralized system
- The resources are physically kept in separate location and there exists no common clock between multiple processors
- The messages are delivered late which may even be lost

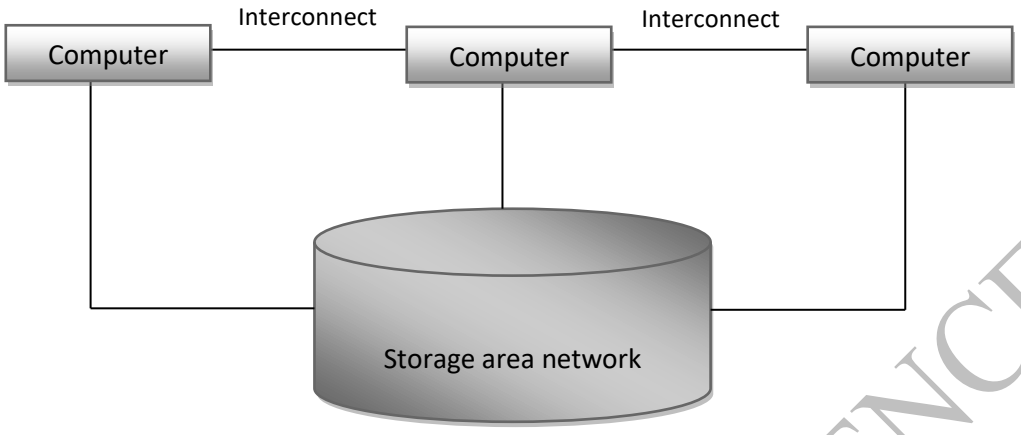
Clustered Systems

A computer cluster is a group of linked computers, working together closely

A user friendly interface is provided by clustered systems among the users, the applications and the hardware. The cluster components are built and connected through fast networks. This network is a local area network which connects all the computers within a room, a floor, or a building.

Clusters are typically arranged for better performance and availability than the single computer. When compared to single computer, the clustered systems are considered to be much more cost-effective with increased speed and high availability service. Even if one or more systems in the cluster fail, the service will continue to fulfil the system requirements.

The clustered systems work with multiple CPUs to accomplish a given task. They are made up of two or more individual systems that are coupled together. The storage area is shared by two or more systems. A layer of cluster software runs on cluster nodes. Every node will monitor one or more nodes through the LAN. If any one system in the network fails, then the monitoring system can take the ownership and restart the application that was failed during execution.



Clustered System

Symmetric Clustering: This is the form of clustering, two or more systems run the applications. They also monitor each other at the same time. It becomes more effective as it uses all available machines. If there are multiple applications which are to be executed, then symmetric clustering can be used.

Asymmetric Clustering: In this form of clustering, one system will be in standby mode and other system is running the application. This standby machine performs actually nothing. It only monitors the server. It becomes the active server if the server fails.

Examples of clustered systems are given below:

- Alpha Kernel
- GLUnix
- Guide
- Hurricane
- MOSIX
- QNX
- RHODOS
- Sumo
- Solaris M
- VAXclusters

- UnixWare Non Stop Cluster

Advantages of Clustered Systems

- Clustering allows two or more systems to share storage
- Provides high reliability
- Scalability
- Lower cost of ownership

Disadvantages of Clustered Systems

- Increased conflict at the clustered resource
- Interfering among the flow of the clustered processors

Real-time Systems

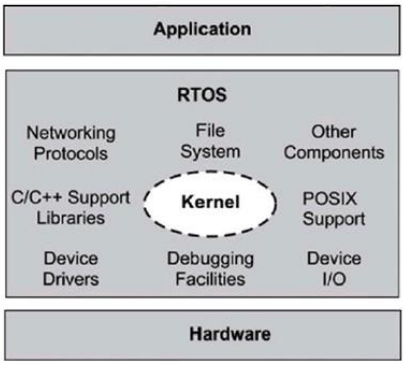
Real time systems are used to control systems such as satellites, robots, etc. They are very useful for measuring and controlling applications. They quickly respond to the input given by the user and produce the output. It has time constraints and those constraints are well defined and fixed. The data processing is done within the defined constraints or the system will fail. It is contrast to time-sharing systems, where there is no quick response to the input given. It is also compared to batch systems where there is no well-defined and fixed time constraint.

Real time systems are categorized into hard real time systems and soft real time systems. In h*ard real time systems*, maximum time is generated for the operations to complete execution. All critical task is completed on time.

In *soft real time systems*, there exists only specific time for the operations to complete its execution. All critical tasks gets priority over the other tasks and the priority is retained until the task is completed.

Deterministic scheduling algorithm is used in real time systems. Determinism is a part of application that runs in a hard real time operating system where events and interrupts are handled

in a defined time. Consider the following fig, the heart of any real time operating system is the kernel. "Kernel" acts as a bridge between the application software and the hardware of a computer.



Real time operating system

Scheduler is the internal part of kernel. It is basically a set of algorithms which manage the task running order. Multitasking enables the kernel to control multiple tasks that must run within time deadlines. Multitasking makes multiple threads to run concurrently where the processor runs task by task, following scheduled order.

Examples of real time systems are given below:

- Automatic Teller Machines (ATM)
- Scientific experiments
- Medical imaging systems
- Industrial control systems
- Display systems
- Weapon systems
- Home appliances controller systems

Advantages of real time systems

- Task scheduling
- Multitasking
- Deterministic behavior

- Inter-task communication
- System management

Disadvantages of real time systems

- Cost of the operating system
- Complicated implementation

System components

A system can be created with large and complex problems that can be partitioned into smaller parts. Each part is well defined with proper input, functions and output. One important point to note is that all systems are not designed with same structure. The system components are:

- Process Management
- Main-Memory Management
- File Management
- I/O System Management
- Secondary-Storage Management
- Networking
- Protection System
- Command Interpreter System
- Graphical User Interface

Operating System Services

Operating systems provides various services to the user of the program. Some specific services differ from one operating system to other operating system. The following *fig* demonstrates the view of operating system services.

The following are the five services provided by an operating system

- Program Execution
- I/O Operations
- File System

- Communications
- Error Management
- Resource Allocation
- Auditing
- Security

The operating system provides the types of system calls that can be grouped into following six major categories:

- Process control
- File management
- Device management
- Information maintenance
- Communications
- Protection

The system programs can be classified into following groups:

- **File management**: The system can create, copy, rename, print, dump, list and delete files and directories.
- **Program loading and execution**: After the compilation, the program must be loaded into the memory for execution. System programs loads and resolves links for the program to execute.
- **Status information**: Status of the system can be date, time, amount of memory or disk space, number of users.
- **Programming language support** : Operating systems provide compilers, interpreters, assemblers, debuggers to the user for better support.
- **File modifications** : Text editor may be used to modify the contents of the file stored in disk or other storage devices.
- **Communications** : System programs provide environment for communication between different users, processes and different computer system.

References

1. "Operating Systems" by Achyut S Godbole, Achyut, Tata McGraw-Hill Education – 2005
2. "Operating Systems Concepts" by Abraham Silberschatz, Peter B. Galvin, Greg Gagne, Eighth Edition, John Wiley & Sons, Inc. – 2010
3. "Modern Operating Systems", International Version by Andrew S. Tanenbaum, Pearson Education, 2008
4. "Operating Systems" by I.A.Dhotre, Technical Publications; Eight Revised Edition – 2009
5. "Operating Systems" by Deitel Deitel Choffnes, Third Edition, Pearson Education, Inc. – 2004
6. "Operating Systems: Internals and Design Principles", by William Stallings, Pearson Education, Inc. – 2009
7. http://nptel.iitm.ac.in
8. http://courses.cs.vt.edu
9. http://www.iu.hio.no