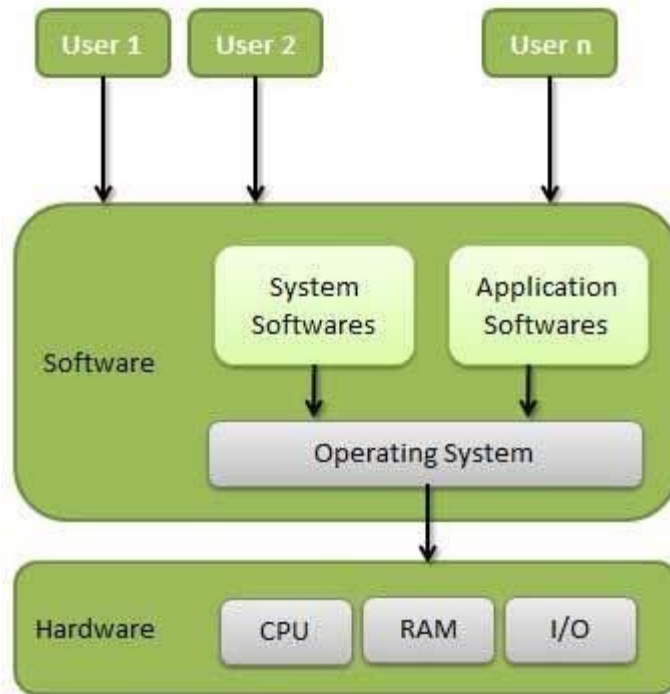


OPERATING SYSTEM

- An operating system is a program which manages all the computer hardwares.
- It provides the base for application program and acts as an intermediary between a user and the computer hardware.
- The operating system has two objectives such as: Firstly, an operating system controls the computer's hardware. The second objective is to provide an interactive interface to the user and interpret commands so that it can communicate with the hardware.
- The operating system is very important part of almost every computer system.
- The prime objective of operating system is to manage & control the various hardware resources of a computer system. These hardware resources include processer, memory, and disk space and so on.
- The output result was display in monitor. In addition to communicating with the hardware the operating system provides on error handling procedure and display an error notification.
- The operating system organizes application so that users can easily access, use and store them.
- It provides a stable and consistent way for applications to deal with the hardware without the user having known details of the hardware.
- If the program is not functioning properly, application and displays the appropriate error message
- Computer system components are divided into 5 parts
 1. Computer hardware
 2. operating system
 3. utilities
 4. Application programs
 5. End user



Types of Operating System

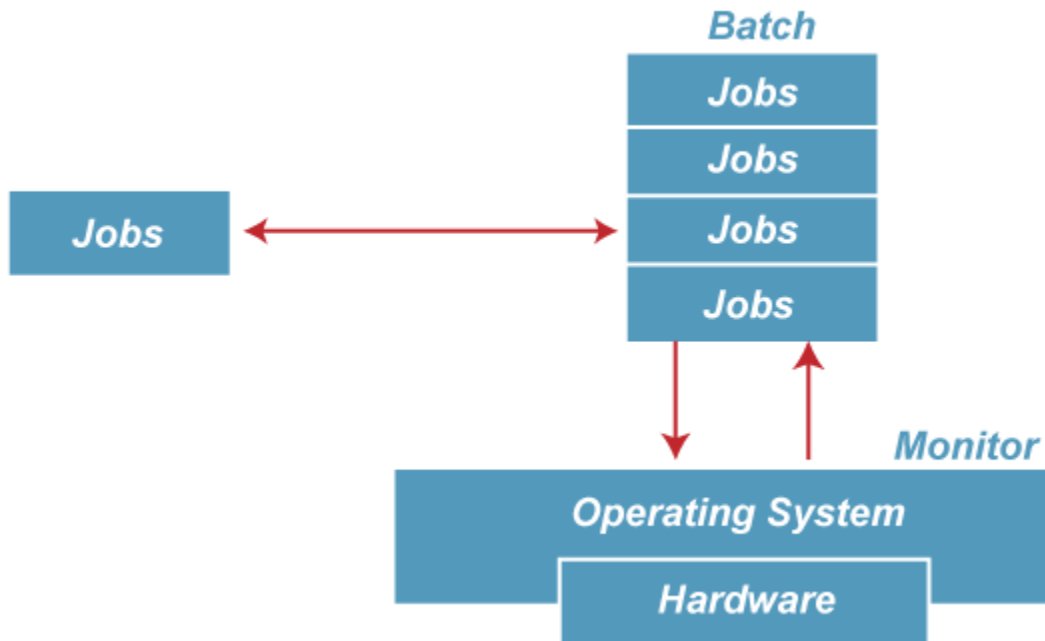
1. Mainframe System:

It is the system where the first computer used to handle many commercial scientific applications. The growth of mainframe systems traced from simple batch system where the computer runs one and only one application to time shared systems which allowed for user interaction with the computer system a.

2. Batch /Early System:

Early computers were physically large machine. The common input devices were card readers, tape drivers. The common output devices were line printers, tape drivers and card punches. In these systems the user did not interact directly with the computer system. Instead the user preparing a job which consists of programming data and some control information and then submitted it to the computer operator after some time the output is appeared. The output in these early computer was fairly simple is main task was to transfer control automatically from one job to next. The operating system always resides in the memory. To speed up processing operators batched the jobs with similar needs and ran then together as a group. The

disadvantages of batch system are that in this execution environment the CPU is often idle because the speed up of I/O devices is much slower than the CPU.



3. Multiprogrammed System:

Multiprogramming concept increases CPU utilization by organization jobs so that the CPU always has one job to execute the idea behind multiprogramming concept. The operating system keeps several jobs in memory simultaneously as shown in below figure. This set of job is subset of the jobs kept in the job pool. The operating system picks and beginning to execute one of the jobs in the memory. In this environment the operating system simply switches and executes another job. When a job needs to wait the CPU is simply switched to another job and so on. The multiprogramming operating system is sophisticated because the operating system makes decisions for the user. This is known as scheduling. If several jobs are ready to run at the same time the system choose one among them. This is known as CPU scheduling.

The disadvantages of the multiprogrammed system are -

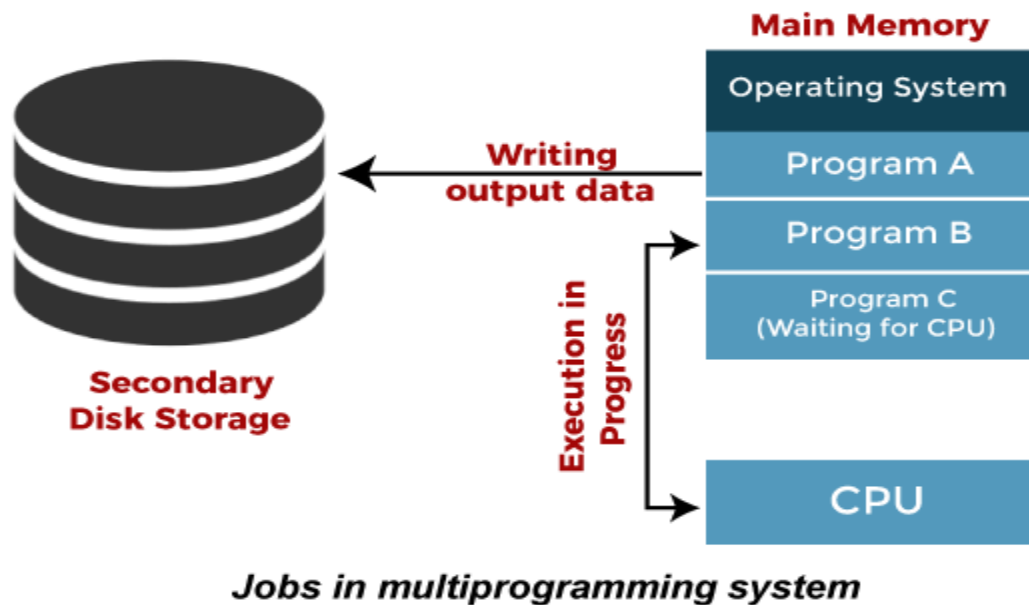
It does not provide user interaction with the computer system during the program execution.

The introduction of disk technology solved these problems rather than reading the cards from card reader into disk.

This form of processing is known as spooling. SPOOL stands for simultaneous peripheral operations online.

It uses the disk as a huge buffer for reading from input devices and for storing output data until the output devices accept them. It is also use for processing data at remote sides.

- The remote processing is done and its own speed with no CPU intervention. Spooling overlaps the input, output one job with computation of other jobs.
- Spooling has a beneficial effect on the performance of the systems by keeping both CPU and I/O devices working at much higher time.

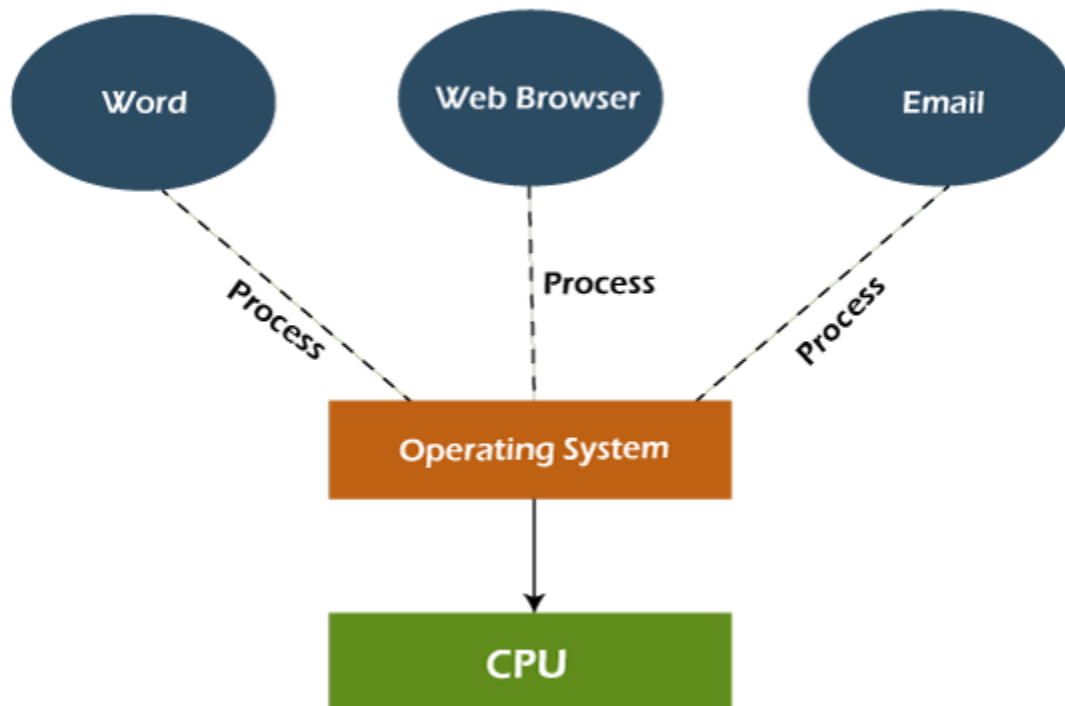


4. Time Sharing System:

The time sharing system is also known as multi user systems. The CPU executes multiple jobs by switching among them but the switches occurs so frequently that the user can interact with each program while it is running. An interactive computer system provides direct communication between a user and system. The user gives instruction to the operating systems or to a program directly using keyboard or mouse and wait for immediate results. So the response time will be short. The time sharing system allows many users to share the computer simultaneously. Since each action in this system is short, only a little CPU time is needed for each user. The system switches rapidly from one user to the next so each user feels as if the entire computer system is dedicated to his use, even though it is being shared by many users.

The disadvantages of time sharing system are:

- It is more complex than multiprogrammed operating system .
- The system must have memory management & protection, since several jobs are kept in memory at the same time.
- Time sharing system must also provide a file system, so disk management is required.
- It provides mechanism for concurrent execution which requires complex CPU scheduling schemes.



5. Personal Computer System/Desktop System:

Personal computer appeared in 1970's. They are microcomputers that are smaller & less expensive than mainframe systems. Instead of maximizing CPU & peripheral utilization, the systems opt for maximizing user convenience & responsiveness. At first file protection was not necessary on a personal machine. But when other computers 2nd other users can access the files on a pc file protection becomes necessary. The lack of protection made it easy for malicious programs to destroy data on such systems. These programs may be self replicating & they spread via worm or virus mechanisms. They can disrupt entire companies or even world wide networks. E.g : windows 98, windows 2000, Linux.

Microprocessor Systems/ Parallel Systems/ Tightly coupled Systems:

These Systems have more than one processor in close communications which share the computer bus, clock, memory & peripheral devices. Ex: UNIX, LINUX.

Multiprocessor Systems have 3 main advantages.

- a. Increased throughput: No. of processes computed per unit time. By increasing the no. of processors more work can be done in less time. The speed up ratio with N processors is not N, but it is less than N. Because a certain amount of overhead is incurred in keeping all the parts working correctly.
- b. Increased Reliability: If functions can be properly distributed among several processors, then the failure of one processor will not halt the system, but slow it down. This ability to continue to operate in spite of failure makes the system fault tolerant.
- c. Economic scale: Multiprocessor systems can save money as they can share peripherals, storage & power supplies.

6. Distributed System/Loosely Coupled Systems:

In contrast to tightly coupled systems, the processors do not share memory or a clock. Instead, each processor has its own local memory. The processors communicate with each other by various communication lines such as high speed buses or telephone lines. Distributed systems depend on networking for their functionalities. By being able to communicate distributed systems are able to share computational tasks and provide a rich set of features to the users. Networks vary by the protocols used, the distances between the nodes and transport media. TCP/IP is the most common network protocol. The processor in a distributed system varies in size and function. It may be microprocessors, work stations, minicomputer, and large general purpose computers. Network types are based on the distance between the nodes such as LAN (within a room, floor or building) and WAN (between buildings, cities or countries). The advantages of distributed system are resource sharing, computation speed up, reliability, communication.

7. Real time Systems:

Real time system is used when there are rigid time requirements on the operation of a processor or flow of data. Sensors bring data to the computers. The computer analyzes data and adjusts controls to modify the sensors inputs. Systems that control scientific experiments, medical imaging systems and some display systems are real

time systems. The disadvantages of real time system are: a. A real time system is considered to function correctly only if it returns the correct result within the time constraints. b. Secondary storage is limited or missing instead data is usually stored in short term memory or ROM. c. Advanced OS features are absent. Real time system is of two types such as:

- **Hard real time systems:**

It guarantees that the critical task has been completed on time. The sudden task is takes place at a sudden instant of time.

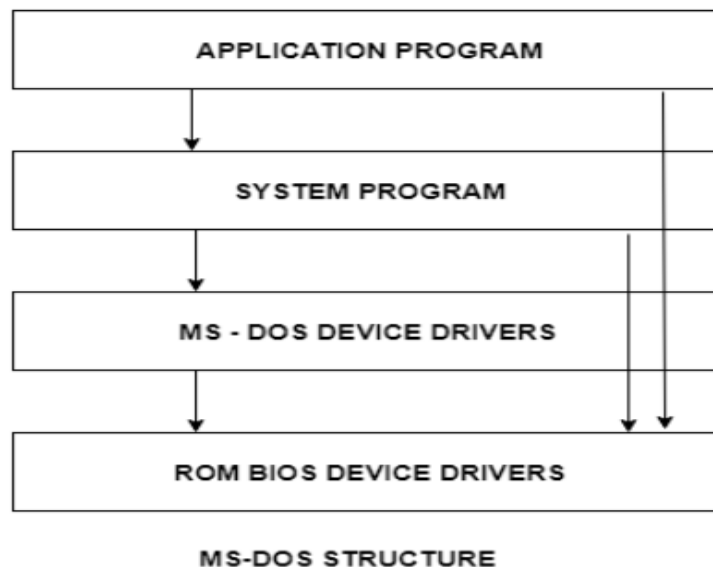
- **Soft real time systems:**

It is a less restrictive type of real time system where a critical task gets priority over other tasks and retains that priority until it computes. These have more limited utility than hard real time systems. Missing an occasional deadline is acceptable e.g. QNX, VX works. Digital audio or multimedia is included in this category. It is a special purpose OS in which there are rigid time requirements on the operation of a processor. A real time OS has well defined fixed time constraints. Processing must be done within the time constraint or the system will fail. A real time system is said to function correctly only if it returns the correct result within the time constraint. These systems are characterized by having time as a key parameter.

Operating System Structure

1. Simple structure:

There are several commercial system that don't have a well- defined structure such operating systems begins as small, simple & limited systems and then grow beyond their original scope. MS-DOS is an example of such system. It was not divided into modules carefully. Another example of limited structuring is the UNIX operating system.



2. Layered approach:

In the layered approach, the OS is broken into a number of layers (levels) each built on top of lower layers. The bottom layer (layer 0) is the hardware & top most layer (layer N) is the user interface.

The main advantage of the layered approach is modularity.

- The layers are selected such that each users functions (or operations) & services of only lower layer. This approach simplifies debugging & system verification, i.e. the first layer can be debugged without concerning the rest of the system. Once the first layer is debugged, its correct functioning is assumed while the 2nd layer is debugged & so on.
- If an error is found during the debugging of a particular layer, the error must be on that layer because the layers below it are already debugged. Thus the design & implementation of the system are simplified when the system is broken down into layers.
- Each layer is implemented using only operations provided by lower layers. A layer doesn't need to know how these operations are implemented; it only needs to know what these operations do.
- The layer approach was first used in the operating system. It was defined in six layers.

Layers	Functions
5	User Program
4	I/O Management
3	Operator Process Communication
2	Memory Management
1	CPU Scheduling
0	Hardware

The main disadvantage of the layered approach is:

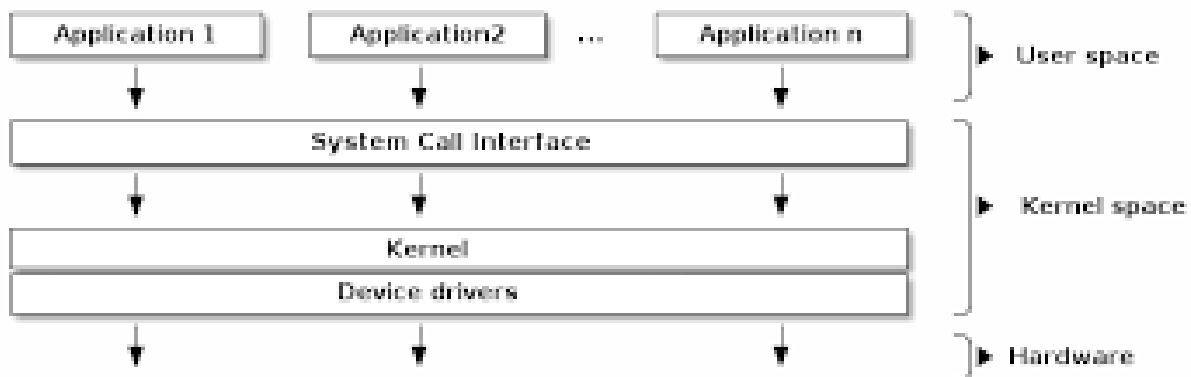
- The main difficulty with this approach involves the careful definition of the layers, because a layer can use only those layers below it. For example, the device driver for the disk space used by virtual memory algorithm must be at a level lower than that of the memory management routines, because memory management requires the ability to use the disk space.
- It is less efficient than a non layered system (Each layer adds overhead to the system call & the net result is a system call that take longer time than on a non layered system).

3. Kernel Structure

Kernel Structure is a control module of an os. It is one structure which loads first and remain in memory. It lies between system program and hardware. This kernel uses system call to perform all its functions like CPU scheduling ,memory management etc. It is also called command line interpreter.

Main Function Of Kernel –

- It provide mechanism for creation and deletion of processes.
- It provide CPU scheduling, memoery management.
- It provide mechanism for inter process communication.



Types of Kernel :

1. Monolithic Kernel –

It is one of types of kernel where all operating system services operate in kernel space. It has dependencies between systems components. It has huge lines of code which is complex.

Example

Unix, Linux, Open VMS, XTS-400 etc.

- Advantage
It has good performance.

- Disadvantage
It has dependencies between system component and lines of code in millions.

2. Micro Kernel –

It is kernel types which has minimalist approach. It has virtual memory and thread scheduling. It is more stable with less services in kernel space. It puts rest in user space.

Example

Mach, L4, AmigaOS, Minix, K42 etc.

- Advantage
It is more stable.
- Disadvantage
There are lots of system calls and context switches.

3. HybridKernel

It is the combination of both monolithic kernel and microkernel. It has speed and design of monolithic kernel and modularity and stability of microkernel.

Example

Windows NT, Netware, BeOS etc.

- Advantage
It combines both monolithic kernel and microkernel.
- Disadvantage
It is still similar to monolithic kernel.

COMPONENTS OF OPERATING SYSTEM AND SERVICES

The various functions of operating system are as follows:

1. Process Management:

- A program does nothing unless their instructions are executed by a CPU. A process is a program in execution. A time shared user program such as a compiler is a

process. A word processing program being run by an individual user on a pc is a process.

- A system task such as sending output to a printer is also a process. A process needs certain resources including CPU time, memory files & I/O devices to accomplish its task.
- These resources are either given to the process when it is created or allocated to it while it is running. The OS is responsible for the following activities of process management.
- Creating & deleting both user & system processes.
- Suspending & resuming processes.
- Providing mechanism for process synchronization.
- Providing mechanism for process communication.
- Providing mechanism for deadlock handling.

2. Main Memory Management:

The main memory is central to the operation of a modern computer system. Main memory is a large array of words or bytes ranging in size from hundreds of thousand to billions. Main memory stores the quickly accessible data shared by the CPU & I/O device. The central processor reads instruction from main memory during instruction fetch cycle & it both reads & writes data from main memory during the data fetch cycle. The main memory is generally the only large storage device that the CPU is able to address & access directly. For example, for the CPU to process data from disk. Those data must first be transferred to main memory by CPU generated E/O calls. Instruction must be in memory for the CPU to execute them.

The OS is responsible for the following activities in connection with memory management.

- Keeping track of which parts of memory are currently being used & by whom.
- Deciding which processes are to be loaded into memory when memory space becomes available.

- Allocating & deallocating memory space as needed.

3. File Management:

File management is one of the most important components of an OS computer can store information on several different types of physical media magnetic tape, magnetic disk & optical disk are the most common media. Each medium is controlled by a device such as disk drive or tape drive those has unique characteristics. These characteristics include access speed, capacity, data transfer rate & access method (sequential or random). For convenient use of computer system the OS provides a uniform logical view of information storage. The OS abstracts from the physical properties of its storage devices to define a logical storage unit the file. A file is collection of related information defined by its creator.

The OS is responsible for the following activities of file management.

- Creating & deleting files.
- Creating & deleting directories.
- Supporting primitives for manipulating files & directories
- Mapping files into secondary storage.
- Backing up files on non-volatile media.

4. I/O System Management:

One of the purposes of an OS is to hide the peculiarities of specific hardware devices from the user. For example, in UNIX the peculiarities of I/O devices are hidden from the bulk of the OS itself by the I/O subsystem.

The I/O subsystem consists of:

- A memory management component that includes buffering, catching & spooling.
- A general device- driver interfaces drivers for specific hardware devices. Only the device driver knows the peculiarities of the specific device to which it is assigned.

5. Secondary Storage Management:

The main purpose of computer system is to execute programs. These programs with the data they access must be in main memory during execution. As the main memory is too small to accommodate all data & programs & because the data that it holds are lost when power is lost. The computer system must provide secondary storage to back-up main memory. Most modern computer systems are disks as the storage medium to store data & program. The operating system is responsible for the following activities of disk management.

- Free space management.
- Storage allocation.
- Disk scheduling

Because secondary storage is used frequently it must be used efficiently.

6. Networking:

A distributed system is a collection of processors that don't share memory peripheral devices or a clock. Each processor has its own local memory & clock and the processor communicate with one another through various communication lines such as high speed buses or networks. The processors in the system are connected through communication networks which are configured in a number of different ways. The communication network design must consider message routing & connection strategies are the problems of connection & security.

7. Protection or security:

If a computer system has multi users & allow the concurrent execution of multiple processes then the various processes must be protected from one another's activities. For that purpose, mechanisms ensure that files, memory segments, CPU & other resources can be operated on by only those processes that have gained proper authorization from the OS.

8. Command interpretation:

One of the most important functions of the OS is command interpretation where it acts as the interface between the user & the OS.

Operating System Services

An operating system provides an environment for the execution of the program. It provides some services to the programs. The various services provided by an operating system are as follows:

- **Program Execution:** The system must be able to load a program into memory and to run that program. The program must be able to terminate this execution either normally or abnormally.
- **I/O Operation:** A running program may require I/O. This I/O may involve a file or a I/O device for specific device. Some special function can be desired. Therefore the operating system must provide a means to do I/O.
- **File System Manipulation:** The programs need to create and delete files by name and read and write files. Therefore the operating system must maintain each and every files correctly.
- **Communication:** The communication is implemented via shared memory or by the technique of message passing in which packets of information are moved between the processes by the operating system.
- **Error detection:** The operating system should take the appropriate actions for the occurrences of any type like arithmetic overflow, access to the illegal memory location and too large user CPU time.
- **Resource Allocation:** When multiple users are logged on to the system the resources must be allocated to each of them. For current distribution of the resource among the various processes the operating system uses the CPU scheduling run times which determine which process will be allocated with the resource.
- **Accounting:** The operating system keep track of which users use how many and which kind of computer resources.
- **Protection:** The operating system is responsible for both hardware as well as software protection. The operating system protects the information stored in a multiuser computer system.

SYSTEM CALLS

System calls provide the interface between a process & the OS. These are usually available in the form of assembly language instruction. Some systems allow system calls to be made directly from a high level language program like C, BCPL and PERL etc. systems calls occur in different ways depending on the computer in use. System calls can be roughly grouped into 5 major categories.

1. Process Control:

- End, abort: A running program needs to be able to has its execution either normally (end) or abnormally (abort).
- Load, execute: A process or job executing one program may want to load and executes another program.
- Create Process, terminate process: There is a system call specifying for the purpose of creating a new process or job (create process or submit job). We may want to terminate a job or process that we created (terminates process, if we find that it is incorrect or no longer needed).
- Get process attributes, set process attributes: If we create a new job or process we should able to control its execution. This control requires the ability to determine & reset the attributes of a job or processes (get process attributes, set process attributes).
- Wait time: After creating new jobs or processes, we may need to wait for them to finish their execution (wait time).
- Wait event, signal event: We may wait for a specific event to occur (wait event). The jobs or processes then signal when that event has occurred (signal event).

2. File Manipulation:

- Create file, delete file: We first need to be able to create & delete files. Both the system calls require the name of the file & some of its attributes.
- Open file, close file: Once the file is created, we need to open it & use it. We close the file when we are no longer using it.
- Read, write, reposition file: After opening, we may also read, write or reposition the file (rewind or skip to the end of the file).

- Get file attributes, set file attributes: For either files or directories, we need to be able to determine the values of various attributes & reset them if necessary. Two system calls get file attribute & set file attributes are required for their purpose.

3. Device Management:

- Request device, release device: If there are multiple users of the system, we first request the device. After we finished with the device, we must release it.
- Read, write, reposition: Once the device has been requested & allocated to us, we can read, write & reposition the device.

4. Information maintenance:

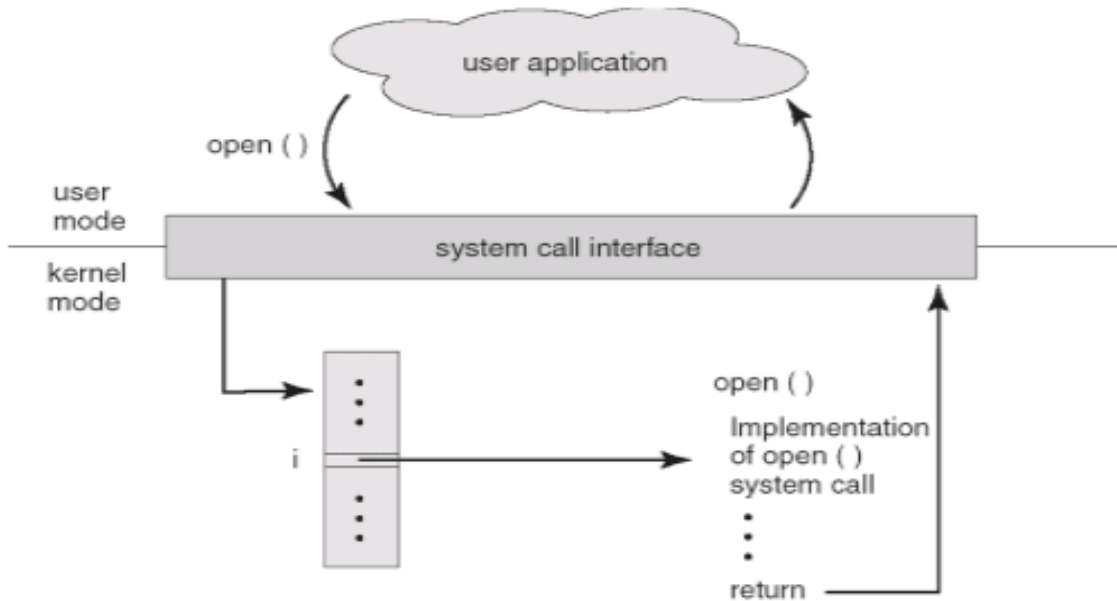
- Get time or date, set time or date: Most systems have a system call to return the current date & time or set the current date & time.
- Get system data, set system data: Other system calls may return information about the system like number of current users, version number of OS, amount of free memory etc.
- Get process attributes, set process attributes: The OS keeps information about all its processes & there are system calls to access this information.

5. Communication:

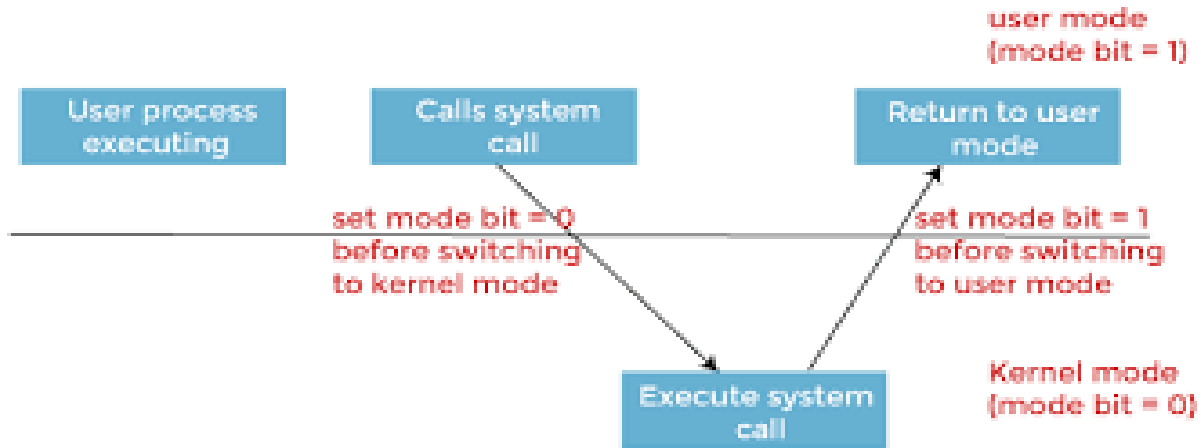
There are two modes of communication such as:

- Message passing model: Information is exchanged through an inter process communication facility provided by operating system. Each computer in a network has a name by which it is known. Similarly, each process has a process name which is translated to an equivalent identifier by which the OS can refer to it. The get hostid and get processed systems calls to do this translation. These identifiers are then passed to the general purpose open & close calls provided by the file system or to specific open connection system call. The recipient process must give its permission for communication to take place with an accept connection call. The source of the communication known as client & receiver known as server exchange messages by read message & write message system calls. The close connection call terminates the connection.

- Shared memory model: processes use map memory system calls to access regions of memory owned by other processes. They exchange information by reading & writing data in the shared areas. The processes ensure that they are not writing to the same location simultaneously.



DUAL OPERATION IN OS



When the computer system executes on behalf of a user application, the system is in user mode. However, when a user application requests a service from the operating system via a system call, it must transition from user to kernel mode to fulfill the request. As we can say, this architectural enhancement is useful for many other aspects of system operation. At system boot time, the hardware starts in kernel mode. The operating system is then loaded and starts user applications in user mode. Whenever a trap or interrupt occurs, the hardware switches from user mode to kernel mode, changing the mode bit's state to 0. Thus, whenever the operating system gains control of the computer, it is in kernel mode. The system always switches to user mode by setting the mode bit to 1 before passing control to a user program.

Need for Dual-Mode Operation

Certain types of processes are to be made hidden from the user, and certain tasks that do not require any type of hardware support. Using the *dual mode* of the OS, these tasks can be dealt with separately.

Also, the Operating System needs to function in the *dual mode* because the Kernel Level programs perform all the bottom level functions of the OS like process management, Memory management, etc. If the user alters these, then this can cause an entire system

failure. So, for specifying the access to the users only to the tasks of their use, Dual Mode is necessary for an Operating system.

SYSTEM PROGRAM

System programs provide a convenient environment for program development & execution. They are divided into the following categories.

- File manipulation: These programs create, delete, copy, rename, print & manipulate files and directories.
- Status information: Some programs ask the system for date, time & amount of available memory or disk space, no. of users or similar status information.
- File modification: Several text editors are available to create and modify the contents of file stored on disk.
- Programming loading and execution: Once a program is assembled or compiled, it must be loaded into memory to be executed.
- Communications: These programs provide the mechanism for creating virtual connections among processes users 2nd different computer systems.
- Application programs: Most OS are supplied with programs that are useful to solve common problems or perform common operations. Ex: web browsers, word processors & text formatters etc.

VIRTUAL MACHINES

By using CPU scheduling & virtual memory techniques an operating system can create the illusion of multiple processes, each executing on its own processors & own virtual memory. Each processor is provided a virtual copy of the underlying computer. The resources of the computer are shared to create the virtual machines. CPU scheduling can be used to create the appearance that users have their own processor.

Implementation: Although the virtual machine concept is useful, it is difficult to implement since much effort is required to provide an exact duplicate of the underlying machine. The CPU is being multiprogrammed among several virtual machines, which slows down the virtual machines in various ways.

Difficulty: A major difficulty with this approach is regarding the disk system. The solution is to provide virtual disks, which are identical in all respects except size. These are known as mini disks in IBM's VM OS. The sum of sizes of all mini disks should be less than the actual amount of physical disk space available.

Advantages:

1. There are no protection problems because each virtual machine is completely isolated from all other virtual machines.
2. Virtual machine can provide an instruction set architecture that differs from real computers.
3. Easy maintenance, availability and convenient recovery.

Disadvantages:

1. When multiple virtual machines are simultaneously running on a host computer, one virtual machine can be affected by other running virtual machines, depending on the workload.
2. Virtual machines are not as efficient as a real one when accessing the hardware.

