Name: Inayat yousuf
Task : Operating System basics.

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.

Functionalities of Operating System:

- **Resource Management:** When parallel accessing happens in the OS means when multiple users are accessing the system the OS works as Resource Manager, Its responsibility is to provide hardware to the user. It decreases the load in the system.
- **Process Management:** It includes various tasks like **scheduling**, **termination** of the process. OS manages various tasks at a time. Here **CPU Scheduling** happens means all the tasks would be done by the many algorithms that use for scheduling.
- **Memory Management:** Refers to the management of primary memory. The operating system has to keep track, how much memory has been used and by whom. It has to decide which process needs memory space and how much. OS also has to allocate and deallocate the memory space.

Every general-purpose computer consists of the hardware, operating system, system programs, and application programs. The hardware consists of **memory, CPU, ALU, and I/O devices, peripheral devices, and storage devices**. System program consists of **compilers, loaders, editors, OS,** etc. The application program consists of **business programs, database programs**.

The Operating system must support the following tasks. The tasks are:

1. Provides the facilities to create, modification of programs and data files using an editor.
2. Access to the compiler for translating the user program from high-level language to machine language.
3. Provide a loader program to move the compiled program code to the computer's memory for execution.
4. Provide routines that handle the details of I/O programming.

Types of OS:

- Batch Operating System- Sequence of jobs in a program on a computer without manual interventions.
- Time-sharing operating System- allows many users to share the computer resources. (Max utilization of the resources).
- Distributed operating System- Manages a group of different computers and makes appear to be a single computer.
- Network operating system- computers running in different operating systems can participate in a common network (It is used for security purposes).
- Real-time operating system – meant applications to fix the deadlines.

Functions of an Operating System :
An operating system has a variety of functions to perform. Some of the prominent functions of an

operating system can be broadly outlined as:

- **Processor Management** This deals with the management of the Central Processing Unit (CPU). The operating system takes care of the allotment of CPU time to different processes. When a process finishes its CPU processing after executing for the allotted time period, this is called scheduling. There is various type of scheduling techniques that are used by the operating systems:
    1. **Shortest Job First(SJF)**: The process which needs the shortest CPU time is scheduled first.
    2. **Round Robin Scheduling:** Each process is assigned a fixed CPU execution time in a cyclic way.
    3. **Priority Based scheduling (**Non-Preemptive**):** In this scheduling, processes are scheduled according to their priorities, i.e., the highest priority process is scheduled first. If priorities of two processes match, then schedule according to arrival time.

- **Device Management:**
  The Operating System communicates with the hardware and the attached devices and maintains a balance between them and the CPU. This is all the more important because the CPU processing speed is much higher than that of I/O devices. In order to optimize the CPU time, the operating system employs two techniques – Buffering and Spooling.

- **Buffering:**
  In this technique, input and output data is temporarily stored in Input Buffer and Output Buffer. Once the signal for input or output is sent to or from the CPU respectively, the operating system through the device controller moves the data from the input device to the input buffer and for the output device to the output buffer. In the case of input, if the buffer is full, the operating system sends a signal to the program which processes the data stored in the buffer. When the buffer becomes empty, the program informs the operating system which reloads the buffer and the input operation continues.

- **Spooling (Simultaneous Peripheral Operation on Line):**
  This is a device management technique used for processing different tasks on the same input/output device. When there are various users on a network sharing the same resource then it can be a possibility that more than one user might give it a command at the same point in time. So, the operating system temporarily stores the data of every user on the hard disk of the computer to which the resource is attached. The individual user need not wait for the execution process to be completed. Instead, the operating system sends the data from the hard disk to the resource one by one.
  Example: printer

  Memory management:
  In a computer, both the CPU and the I/O devices interact with the memory. When a program needs to be executed it is loaded onto the main memory till the execution is completed.

Thereafter that memory space is freed and is available for other programs. The common memory management techniques used by the operating system are Partitioning and Virtual Memory.

- **Partitioning:**
  The total memory is divided into various partitions of the same size or different sizes. This helps to accommodate a number of programs in the memory. The partition can be fixed i.e. remains the same for all the programs in the memory or variable i.e. memory is allocated when a program is loaded onto the memory. The latter approach causes less wastage of memory but in due course of time, it may become fragmented.

- **Virtual Memory:**
  This is a technique used by the operating systems which allows the user can load programs that are larger than the main memory of the computer. In this technique, the program is executed even if the complete program can not be loaded inside the main memory leading to efficient memory utilization.

- **File Management:**
  The operating system manages the files, folders, and directory systems on a computer. Any data on a computer is stored in the form of files and the operating system keeps the information about all of them using File Allocation Table (FAT). The FAT stores general information about files like filename, type (text or binary), size, starting address, and access mode (sequential/indexed sequential/direct/relative). The file manager of the operating system helps to create, edit, copy, allocate memory to the files, and also updates the FAT. The operating system also takes care that files are opened with proper access rights to read or edit them.

Process Schedulers in Operating System

The process scheduling is the activity of the process manager that handles the removal of the running process from the CPU and the selection of another process on the basis of a particular strategy. Process scheduling is an essential part of a Multiprogramming operating systems. Such operating systems allow more than one process to be loaded into the executable memory at a time and the loaded process shares the CPU using time multiplexing.

Process Table and Process Control Block (PCB)

While creating a process the operating system performs several operations. To identify the processes, it assigns a process identification number (PID) to each process. As the operating system supports multi-programming, it needs to keep track of all the processes. For this task, the process control block (PCB) is used to track the process’s execution status. Each block of memory contains information about the process state, program counter, stack pointer, status of opened files,

scheduling algorithms, etc. All these information is required and must be saved when the process is switched from one state to another. When the process makes a transition from one state to another, the operating system must update information in the process's PCB.

What is throughput: it is the number of tasks executed per unit time. Linux provides us highest number of throughputs as compared to windows.

Non-Preemptive.. in this case processes are taken by cpu and not leaves it until it has been done. This process can only leaves the cpu only if the process itself some more resources from input or somehwhere else. Eg incase of multi programming.

Preemptive: in this case the process has given a time slot if the process is completed in this time slot then cpu goes for next process or if not done its work in this time slot then cpu also leaves this process and go for another process waiting in the ram. Time sharing or multi-tasking.

Distributed os A distributed system is a computing environment in which various components are spread across multiple computers (or other computing devices) on a network **.** These multiple computer are connected within a network.

The induividual computers are loosly coupled means they have their own kernal, memory, disk, resources etc

Arrival Time: The point of time at which the process enters the ready queue or ready state.

Burst Time: Time required the process to get executed on CPU.

Completion Time: It is the point of time at which process completes its execution.

Turn Around Time: [Completion time – Arrival Time].

Waiting Time: [Turn Arount Time – Burst Time].

Response Time: [(The time at which process get cpu first time) -(Arrival Time)]

Process Scheduling in OS:

The process scheduling is the activity of the process manager that handles the removal of the running process from the CPU and the selection of another process on the basis of a particular strategy.

Process scheduling is an essential part of a Multiprogramming operating systems. Such operating systems allow more than one process to be loaded into the executable memory at a time and the loaded process shares the CPU using time multiplexing

There are three types of process scheduler.

1. **Long Term or job scheduler :**
   It brings the new process to the 'Ready State'. It controls ***Degree of Multi-programming***, i.e., number of process present in ready state at any point of time. It is important that the long-term scheduler make a careful selection of both IO and CPU bound process.

2 . Short term or CPU scheduler :
It is responsible for selecting one process from ready state for scheduling it on the running state. Note: Short-term scheduler only selects the process to schedule it doesn't load the process on running. Here is when all the scheduling algorithms are used.

3. Medium-term scheduler :
It is responsible for suspending and resuming the process. It mainly does swapping (moving processes from main memory to disk and vice versa).

Process Table and Process Control Block (PCB)

While creating a process the operating system performs several operations. To identify the processes, it assigns a process identification number (PID) to each process. As the operating system supports multi-programming, it needs to keep track of all the processes. For this task, the process control block (PCB) is used to track the process's execution status. Each block of memory contains information about the process state, program counter, stack pointer, status of opened files, scheduling algorithms, etc. All these information is required and must be saved when the process is switched from one state to another. When the process makes a transition from one state to another, the operating system must update information in the process's PCB.

A process control block (PCB) contains information about the process, i.e. registers, quantum, priority, etc. The process table is an array of PCB's, that means logically contains a PCB for all of the current processes in the system.

CPU scheduling

Why do we need scheduling?
A typical process involves both I/O time and CPU time. In a uni programming system like MS-DOS, time spent waiting for I/O is wasted and CPU is free during this time. In multi programming systems, one process can use CPU while another is waiting for I/O. This is possible only with process scheduling.

Different Scheduling Algorithms

First Come First Serve (FCFS): Simplest scheduling algorithm that schedules according to arrival times of processes. First come first serve scheduling algorithm states that the process that requests the CPU first is allocated the CPU first. It is implemented by using the FIFO queue.

Shortest Job First (SJF): Process which have the shortest burst time are scheduled first.If two processes have the same bust time then FCFS is used to break the tie. It is a non-preemptive scheduling algorithm.

Longest Job First (LJF): It is similar to SJF scheduling algorithm. But, in this scheduling algorithm, we give priority to the process having the longest burst time. This is non-preemptive in nature i.e., when any process starts executing, can't be interrupted before complete execution.

Shortest Remaining Time First (SRTF): It is preemptive mode of SJF algorithm in which jobs are schedule according to shortest remaining time.

Longest Remaining Time First (LRTF): It is preemptive mode of LJF algorithm in which we give priority to the process having largest burst time remaining.

Round Robin Scheduling: Each process is assigned a fixed time(Time Quantum/Time Slice) in cyclic way.It is designed especially for the time-sharing system. The ready queue is treated as a circular queue. The CPU scheduler goes around the ready queue, allocating the CPU to each process for a time interval of up to 1-time quantum. To implement Round Robin scheduling, we keep the ready queue as a FIFO queue of processes. New processes are added to the tail of the ready queue. The CPU scheduler picks the first process from the ready queue, sets a timer to interrupt after 1-time quantum, and dispatches the process.

Convoy Effect in Operating Systems

Convoy Effect is phenomenon associated with the First Come First Serve (FCFS) algorithm, in which the whole Operating System slows down due to few slow processes.

FCFS algorithm is non-preemptive in nature, that is, once CPU time has been allocated to a process, other processes can get CPU time only after the current process has finished. This property of FCFS scheduling leads to the situation called Convoy Effect.

To avoid Convoy Effect, preemptive scheduling algorithms like Round Robin Scheduling can be used – as the smaller processes don't have to wait much for CPU time – making their execution faster and leading to less resources sitting idle.

Differences between Deadlock and Starvation in OS :

1. Deadlock occurs when none of the processes in the set is able to move ahead due to occupancy of the required resources by some other process as shown in the figure below, on the other hand Starvation occurs when a process waits for an indefinite period of time to get the resource it requires.
2. Other name of deadlock is **Circular Waiting**. Other name of starvation is **Lived lock**.
3. When deadlock occurs no process can make progress, while in starvation apart from the victim process other processes can progress or proceed.

Solution to Starvation : Aging

Aging is a technique of gradually increasing the priority of processes that wait in the system for a long time.For example, if priority range from 127(low) to 0(high), we could increase the priority of a waiting process by 1 Every 15 minutes. Eventually even a process with an initial priority of 127 would take no more than 32 hours for priority 127 process to age to a priority-0 process.

Process Synchronization: is the task of coordinating the execution of processes in a way that no two processes can have access to the same shared data and resources.

It is specially needed in a multi-process system when multiple processes are running together, and more than one processes try to gain access to the same shared resource or data at the same time.

On the basis of synchronization, processes are categorized as one of the following two types:

- **Independent Process** : Execution of one process does not affects the execution of other processes.
- **Cooperative Process** : Execution of one process affects the execution of other processes.

Process synchronization problem arises in the case of Cooperative process also because resources are shared in Cooperative processes.

Critical Section:
When more than one processes access a same code segment that segment is known as critical section. Critical section contains shared variables or resources which are needed to be synchronized to maintain consistency of data variable.

Inter Process Communication (IPC)

Inter-process communication (IPC) is a mechanism that allows processes to communicate with each other and synchronize their actions. The communication between these processes can be seen as a method of co-operation between them. Processes can communicate with each other through both:

1. Shared Memory
2. Message passing

File Allocation Methods

The allocation methods define how the files are stored in the disk blocks. There are three main disk space or file allocation methods.

- Contiguous Allocation
- Linked Allocation
- Indexed Allocation

File Access Methods

When a file is used, information is read and accessed into computer memory and there are several ways to access this information of the file. Some systems provide only one access method for files. Other systems, such as those of IBM, support many access methods, and choosing the right one for a particular application is a major design problem.

There are three ways to access a file into a computer system:

Sequential-Access,

Direct Access,

Index sequential Method.