

PROCESS SCHEDULING

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 Scheduling Queues

The OS maintains all PCBs in Process Scheduling Queues. The OS maintains a separate queue for each of the process states and PCBs of all processes in the same execution state are placed in the same queue. When the state of a process is changed, its PCB is unlinked from its current queue and moved to its new state queue.

The Operating System maintains the following important process scheduling queues –

Job queue – This queue keeps all the processes in the system.

Ready queue – This queue keeps a set of all processes residing in main memory, ready and waiting to execute. A new process is always put in this queue.

Device queues – The processes which are blocked due to unavailability of an I/O device constitute this queue.

Why do we need Scheduling?

In Multiprogramming, if the long term scheduler picks more I/O bound processes then most of the time, the CPU remains idle. The task of Operating system is to optimize the utilization of resources.

If most of the running processes change their state from running to waiting then there may always be a possibility of deadlock in the system. Hence to reduce this overhead, the OS needs to schedule the jobs to get the optimal utilization of CPU and to avoid the possibility to deadlock.

Objectives of Process Scheduling Algorithm

Max CPU utilization [Keep CPU as busy as possible]

Fair allocation of CPU.

Max throughput [Number of processes that complete their execution per time unit]

Min turnaround time [Time taken by a process to finish execution]

Min waiting time [Time a process waits in ready queue]

Min response time [Time when a process produces first response]

WHAT IS SCHEDULER

A scheduler is a type of system software that allows you to handle process scheduling.

Three types of the scheduler are

- 1) Long term
- 2) Short term
- 3) Medium-term

Long Term Scheduler

Long term scheduler is also known as a job scheduler. This scheduler regulates the program and select process from the queue and loads them into memory for execution. It also regulates the degree of multi-programing. However, the main goal of this type of scheduler is to offer a balanced mix of jobs, like Processor, I/O jobs., that allows managing multiprogramming.

Medium Term Scheduler

Medium-term scheduling is an important part of swapping. It enables you to handle the swapped out-processes. In this scheduler, a running process can become suspended, which makes an I/O request. A running process can become suspended if it makes an I/O request. A suspended processes can't make any progress towards completion. In order to remove the process from memory and make space for other processes, the suspended process should be moved to secondary storage.

Short Term Scheduler

Short term scheduling is also known as CPU scheduler. The main goal of this scheduler is to boost the system performance according to set criteria. This helps you to select from a group of processes that are ready to execute and allocates CPU to one of them. The dispatcher gives control of the CPU to the process selected by the short term scheduler.

Scheduling Algorithms

There are various algorithms which are used by the Operating System to schedule the processes on the processor in an efficient way.

1. First Come First Serve
2. Shortest Job First

3. Shortest Remaining Time First
4. Round Robin Scheduling
5. Priority Scheduling
6. Multilevel Queue Scheduling
7. Multilevel Feedback Queue Scheduling
8. Highest Response Ratio Next