

LEC-12: Intro to Process Scheduling | FCFS | Convoy Effect



1. Process Scheduling

- a. Basis of Multi-programming OS.
- b. By switching the CPU among processes, the OS can make the computer more productive.
- c. Many processes are kept in memory at a time, when a process must wait or time quantum expires, the OS takes the CPU away from that process & gives the CPU to another process & this pattern continues.

2. CPU Scheduler

- a. Whenever the CPU become ideal, OS must select one process from the ready queue to be executed.
- b. Done by STS.

3. Non-Preemptive scheduling

- a. Once CPU has been allocated to a process, the process keeps the CPU until it releases CPU either by terminating or by switching to wait-state.
- b. Starvation, as a process with long burst time may starve less burst time process.
- c. Low CPU utilization.

4. Preemptive scheduling

- a. CPU is taken away from a process after time quantum expires along with terminating or switching to wait-state.
- b. Less Starvation
- c. High CPU utilization.

5. Goals of CPU scheduling

- a. Maximum CPU utilization
- b. Minimum Turnaround time (TAT).
- c. Min. Wait-time
- d. Min. response time.
- e. Max. throughput of system.

6. Throughput: No. of processes completed per unit time.

7. Arrival time (AT): Time when process is arrived at the ready queue.

8. Burst time (BT): The time required by the process for its execution.

9. Turnaround time (TAT): Time taken from first time process enters ready state till it terminates. (CT - AT)

10. Wait time (WT): Time process spends waiting for CPU. (WT = TAT - BT)

11. Response time: Time duration between process getting into ready queue and process getting CPU for the first time.

12. Completion Time (CT): Time taken till process gets terminated.

13. FCFS (First come-first serve):

- a. Whichever process comes first in the ready queue will be given CPU first.
- b. In this, if one process has longer BT. It will have major effect on average WT of diff processes, called **Convoy effect**.
- c. Convoy Effect is a situation where many processes, who need to use a resource for a short time, are blocked by one process holding that resource for a long time.
 - i. This cause poor resource management.