Unit 7

SCHEDULING ALGORITHM

A Process Scheduler

A Process Scheduler schedules different processes to be assigned to the CPU based on particular scheduling algorithms.

There are six popular process scheduling algorithms which we are going to discuss

1. First-Come, First-Served (FCFS) Scheduling
2. Shortest-Job-Next (SJN) Scheduling
3. Priority Scheduling
4. Shortest Remaining Time
5. Round Robin(RR) Scheduling
6. Multiple-Level Queues Scheduling

Algorithms

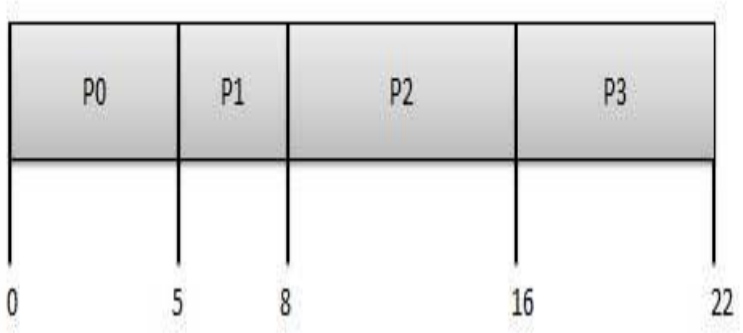
These algorithms are either **non-preemptive or preemptive**. Non-preemptive algorithms are designed so that once a process enters the running state, it cannot be preempted until it completes its allotted time, whereas the preemptive scheduling is based on priority where a scheduler may preempt a low priority running process anytime when a high priority process enters into a ready state.

First Come First Serve (FCFS)

1. Jobs are executed on first come, first serve basis.
2. It is a non-preemptive, pre-emptive scheduling algorithm.
3. Easy to understand and implement.
4. Its implementation is based on FIFO queue.
5. Poor in performance as average wait time is high.

First Come First Serve (FCFS)

Process	Arrival Time	Execute Time	Service Time
P0	0	5	0
P1	1	3	5
P2	2	8	8
P3	3	6	16



Wait time of each process is as follows −

Process	Wait Time : Service Time - Arrival Time
P0	0 - 0 = 0
P1	5 - 1 = 4
P2	8 - 2 = 6
P3	16 - 3 = 13

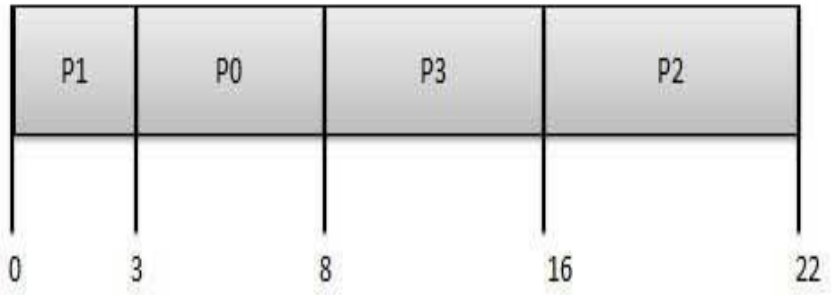
Average Wait Time: (0+4+6+13) / 4 = 5.75

Shortest Job Next (SJN)

1. This is also known as **shortest job first**, or SJF
2. This is a non-preemptive, pre-emptive scheduling algorithm.
3. Best approach to minimize waiting time.
4. Easy to implement in Batch systems where required CPU time is known in advance.
5. Impossible to implement in interactive systems where required CPU time is not known.
6. The processer should know in advance how much time process will take.

Shortest Job Next (SJN)

Process	Arrival Time	Execute Time	Service Time
P0	0	5	3
P1	1	3	0
P2	2	8	16
P3	3	6	8



Wait time of each process is as follows –

Process	Wait Time : Service Time - Arrival Time
P0	3 - 0 = 3
P1	0 - 0 = 0
P2	16 - 2 = 14
P3	8 - 3 = 5

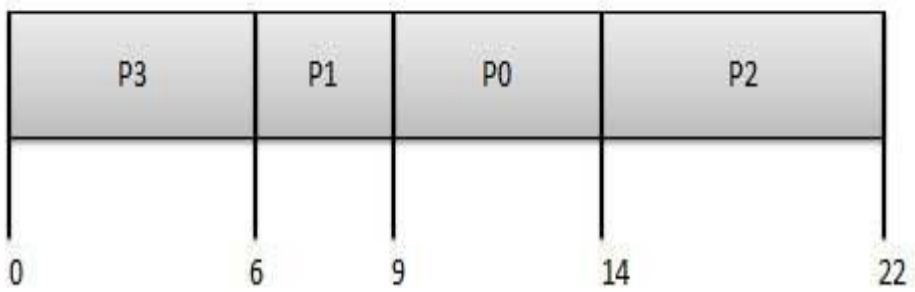
Average Wait Time: (3+0+14+5) / 4 = 5.50

Priority Based Scheduling

1. Priority scheduling is a non-preemptive algorithm and one of the most common scheduling algorithms in batch systems.
2. Each process is assigned a priority. Process with highest priority is to be executed first and so on.
3. Processes with same priority are executed on first come first served basis.
4. Priority can be decided based on memory requirements, time requirements or any other resource requirement.

Priority Based Scheduling

Process	Arrival Time	Execute Time	Priority	Service Time
P0	0	5	1	9
P1	1	3	2	6
P2	2	8	1	14
P3	3	6	3	0



Wait time of each process is as follows −

Process	Wait Time : Service Time - Arrival Time
P0	9 - 0 = 9
P1	6 - 1 = 5
P2	14 - 2 = 12
P3	0 - 0 = 0

Average Wait Time: (9+5+12+0) / 4 = 6.5

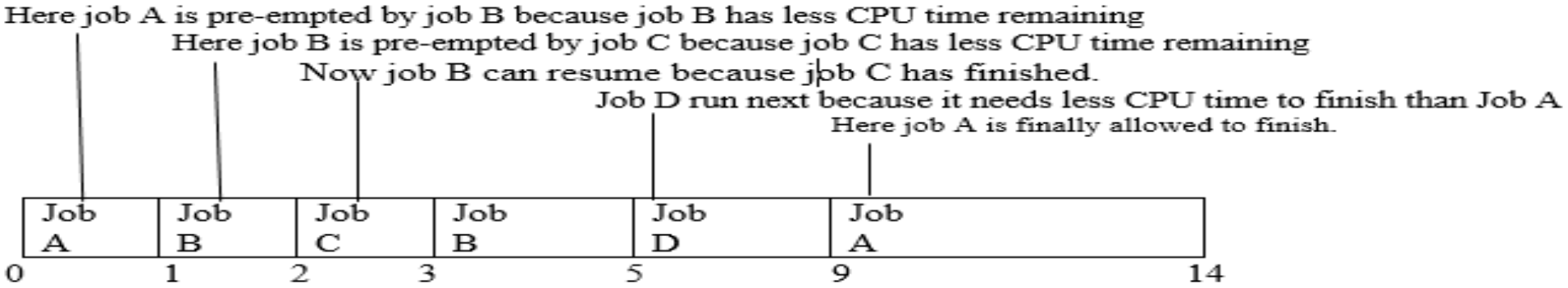
Shortest Remaining Time

1. Shortest remaining time (SRT) is the preemptive version of the SJN algorithm.
2. The processor is allocated to the job closest to completion but it can be preempted by a newer ready job with shorter time to completion.
3. Impossible to implement in interactive systems where required CPU time is not known.
4. It is often used in batch environments where short jobs need to give preference.

Example

Four jobs arrived in quick succession (1 CPU cycle a part) as shown below. Use SRT algorithm.

Arrival time:	0	1	5	6
Job:	A	B	C	D
CPU cycle:	6	3	1	4



In this case the **turnaround time** is the **completion time** of each job minus its **arrival time.**

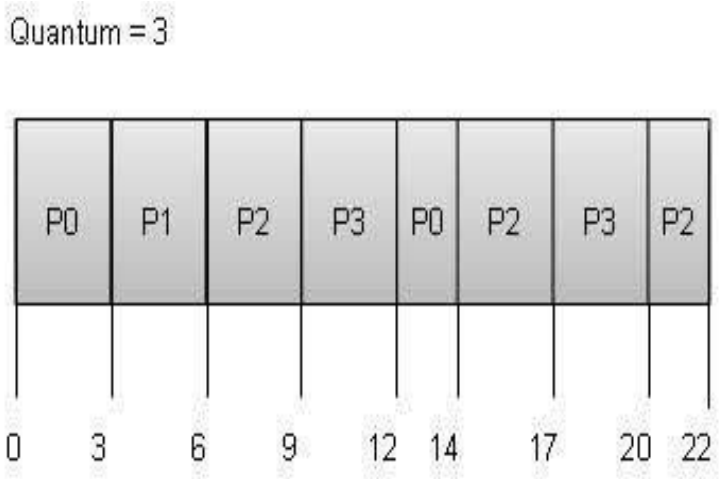
Job:	A	B	C	D
Turnaround:	14	4	1	6

Average turnaround time $= \frac{14 + 4 + 1 + 6}{4} = 6.25$

Round Robin Scheduling

1. Round Robin is the preemptive process scheduling algorithm.
2. Each process is provided a fix time to execute, it is called a **quantum**.
3. Once a process is executed for a given time period, it is preempted and other process executes for a given time period.
4. Context switching is used to save states of preempted processes

Round Robin Scheduling



Wait time of each process is as follows –

Process	Wait Time : Service Time - Arrival Time
P0	(0 - 0) + (12 - 3) = 9
P1	(3 - 1) = 2
P2	(6 - 2) + (14 - 9) + (20 - 17) = 12
P3	(9 - 3) + (17 - 12) = 11

Average Wait Time: (9+2+12+11) / 4 = 8.5

Multiple-Level Queues Scheduling

Multiple-level queues are not an independent scheduling algorithm. They make use of other existing algorithms to group and schedule jobs with common characteristics.

1. Multiple queues are maintained for processes with common characteristics.
2. Each queue can have its own scheduling algorithms.
3. Priorities are assigned to each queue.

For example, CPU-bound jobs can be scheduled in one queue and all I/O-bound jobs in another queue. The Process Scheduler then alternately selects jobs from each queue and assigns them to the CPU based on the algorithm assigned to the queue.