OPERATING SYSTEM SCHEDULING ALGORITHMS

The Process Scheduler schedule different processes to be assigned to the CPU based on particular scheduling algorithm. There are six popular process scheduling algorithms which we are going to discuss in the following section:

- First-Come, First-Served (FCFS) Scheduling

- Shortest-Job-Next (SJN) Scheduling

- Priority Scheduling

- Shortest Remaining Time

- Round Robin(RR) Scheduling

- Multiple-Level Queues Scheduling

These algorithms are either **nonpreemptive** 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 where as 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)

- Jobs are executed on first come, first serve basis.

- It is a nonpreemptive scheduling algorithm.

- Easy to understand and implement.

- Its implementation is based on FIFO queue.

- Poor in performance as average wait time is high.

Process	Arrival Time	Execute Time	Service Time
P0	0	5	0
P1	1	3	5
P2	2	8	8
P3	3	6	16
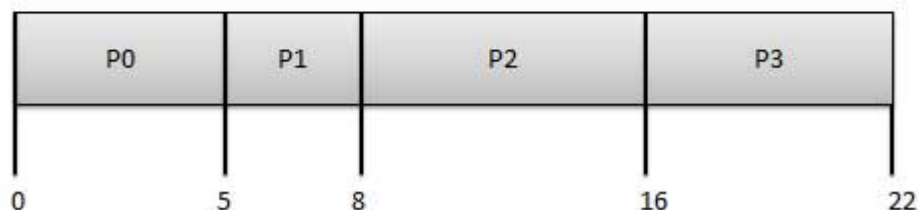
Wait time of each process is following

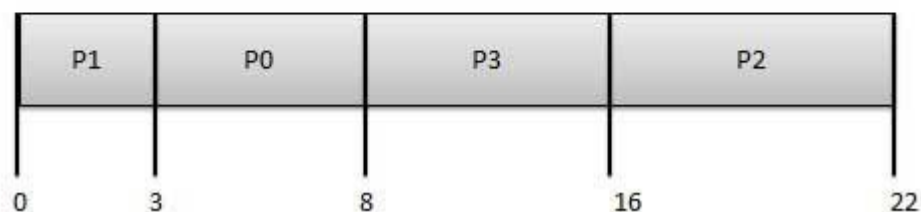| Process	Wait Time : Service Time - Arrival Time
P0	0 - 0 = 0
P1	5 - 1 = 4

P2	8 - 2 = 6
P3	16 - 3 = 13

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

Shortest Job Next (SJN)

- This is also known as shortest job first, or SJF

- This is a nonpreemptive scheduling algorithm.

- Best approach to minimize waiting time.

- Easy to implement in Batch systems where required CPU time is known in advance.

- Impossible to implement in interactive systems where required CPU time is not known.

- Processer should know in advance how much time process will take.

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

P1	P0	P3	P2

```
0       3              8              16              22
```

Wait time of each process is following

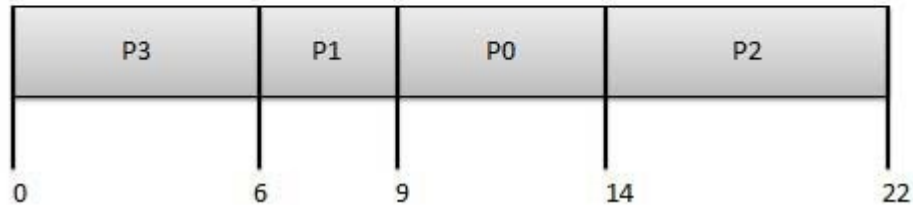| Process	Wait Time : Service Time - Arrival Time
P0	3 - 0 = 3
P1	0 - 0 = 0
P2	16 - 2 = 14
P3	8 - 3 = 5

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

Priority Based Scheduling

- Priority scheduling is a nonpreemptive algorithm and one of the most common sched- uling algorithms in batch systems.

- Each process is assigned a priority. Process with highest priority is to be executed first and so on.

- Processes with same priority are executed on first come first serve basis.

- Priority can be decided based on memory requirements, time requirements or any other resource requirement.

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

P3	P1	P0	P2

0　　　　6　　　9　　　14　　　　22

Wait time of each process is following

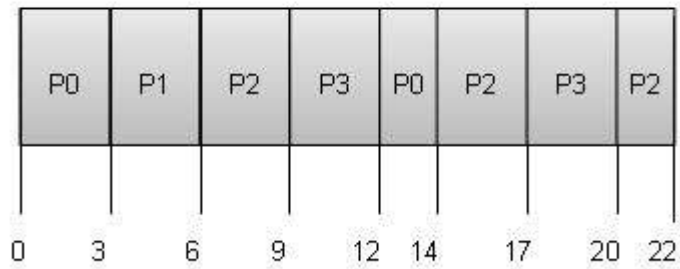| Process	Wait Time : Service Time - Arrival Time
P0	9 - 0 = 9
P1	6 - 1 = 5
P2	14 - 2 = 12
P3	0 - 0 = 0

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

Shortest Remaining Time

- Shortest remaining time (SRT) is the preemptive version of the SJN algorithm.

- 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.

- Impossible to implement in interactive systems where required CPU time is not known.

- It is often used in batch environments where short jobs need to give preference.

Round Robin Scheduling

- Round Robin is the preemptive process scheduling algorithm.

- Each process is provided a fix time to execute called quantum.

- Once a process is executed for given time period. Process is preempted and other process executes for given time period.

- Context switching is used to save states of preempted processes.

Quantum = 3

Wait time of each process is following

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 is not an independent scheduling algorithm but it makes use of other existing algorithms to group and schedule jobs with common characteristic.

- Multiple queues are maintained for processes with common characteristic.

- Each queue can have its own scheduling algorithms.

- 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 assign them to the CPU based on the algorithm assigned to the queue.