CSC 224 Principles of OS

3. OS Processes Mngt Scheduling

Process Management/Scheduling

- In a multiprogramming system, multiple processes exist concurrently in main memory.
- Each process alternates between using a processor and waiting for some event to occur such as the completion of an I/O operation.
- The key to multiprogramming is Scheduling.

Process scheduling

- 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.

Scheduling queues

- Scheduling queues refers to queues of processes or devices.

Scheduling queues

- **Job queue**
- Any process that enters into the system is put in a job queue, that consists of all processes in the system.
- **Ready queue**
- Set of all processes residing in main memory, ready and waiting to execute
- **Device queue**
- List of devices waiting for a particular I/O device

Two state process model

- Refers to running and non-running states.

- **Running**
- when new process is created by OS, that process enters into the system in the running state.

- **Non-Running**
- Processes that are not running are kept in queue, waiting for their turn to execute.
- When a process is interrupted, that process is transferred in the waiting queue.
- If the process has completed or aborted, the process is discarded. In either case, the dispatcher then selects a process from the queue to execute.

Schedulers

- Schedulers are special system software which handles process scheduling in various ways.
- Their main task is to select the jobs to be submitted into the system and to decide which process to run

Types of Schedulers

- Schedulers are of three types:
- Long term
- Short term
- Medium term

Long term (Job) schedulers

- Determines which programs are admitted to the system for processing. Job scheduler selects processes from the queue and loads them into memory for execution.
- The primary objective of the job scheduler is to provide a balanced mix of jobs, such as I/O bound and processor bound.

- It also controls the degree of multiprogramming. If the degree of multiprogramming is stable, then the average rate of process creation must be equal to the average departure rate of processes leaving the system.

Short term (CPU) scheduler

- Main objective is increasing system performance in accordance with the chosen set of criteria.
- It is the change of ready state to running state of the process.
- CPU scheduler selects a process among the processes that are ready to execute and allocates CPU to one of them.

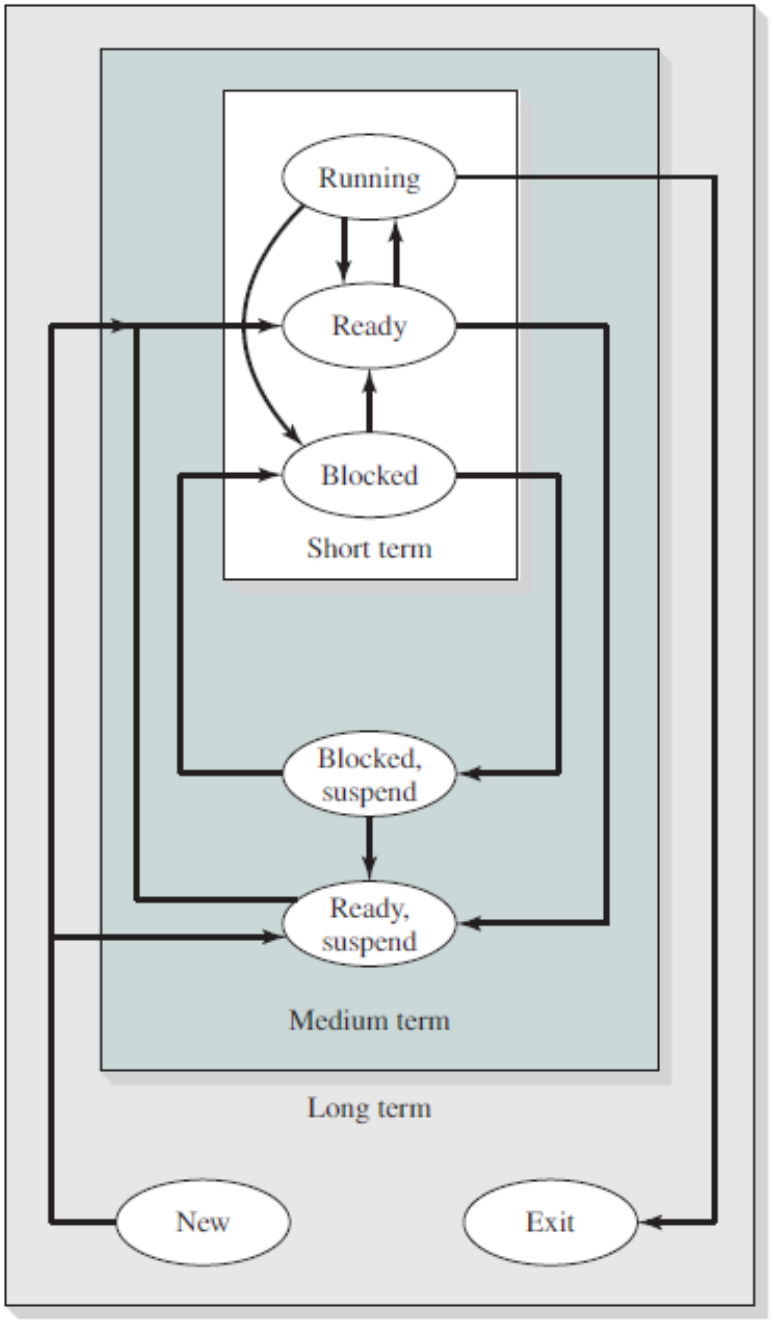
- Short term scheduler also known as dispatcher, execute most frequently and makes the fine grained decision of which process to execute next.
- Short term scheduler is faster than long term scheduler

Medium term scheduler

- Medium term scheduling is part of the swapping. It removes the processes from the memory.
- It reduces the degree of multiprogramming.
- The medium term scheduler is in-charge of handling the swapped out-processes.

Long term	Short Term	Medium Term
Job scheduler	CPU scheduler	Process swapping Scheduler
Speed lesser than short term	Speed is fastest	Speed in between
Controls degree of multiprogramming	Lesser control of multiprogramming	Reduces degree of multiprogramming
Minimal in time sharing systems	Minimal in time sharing systems	Part of time sharing systems
Selects processes from pool & loads them to memory	Selects processes ready to execute	Can reintroduce process into memory & execution can be

Levels of scheduling



Context switch

- Is the mechanism to store and restore the state or context of a CPU in Process Control block so that a process execution can be resumed from the same point at a later time.
- Using this technique a context switcher enables multiple processes to share a single CPU. Context switching is an essential part of a multitasking operating system features.

Process Scheduling Algorithms

- First Come First Served (FCFS) (FIFO)
- Shortest Job First (SJF) Scheduling
- Priority Scheduling
- Round Robin (RR) Scheduling
- Multilevel Queue Scheduling

1. FCFS

- Jobs are executed on first come, first serve basis.
- Easy to understand and implement.
- Poor in performance as average wait time is high

Process	Arrival Time	Execute Time	Service Time
P 0	0	5	0
P 1	1	3	5
P 2	2	8	8
P 3	3	6	16

Execution

P0	P1	P2	P3
0	5	8	16	22

Wait time

- 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 = 5.55

2. Shortest Job First

- Best approach to minimize waiting time.
- Impossible to implement
- Processer should know in advance how much time process will take.

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

Execution

P1	P0	P3	P2
0	3	8	14	22

- Wait time : Service time – Arrival time

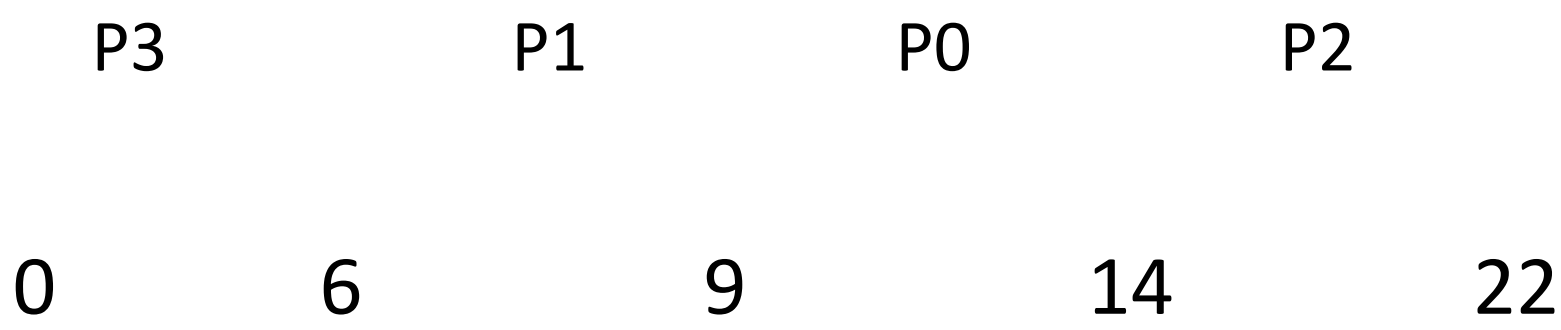
- P0 3 – 0 = 3
- P1 0 – 0 = 0
- P2 16 – 2 = 14
- P3 8 - 3 = 5

- Average wait time: (3+0+14+5)/4 = 5.50

3. Priority Based Scheduling

- 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	0
P1	1	3	2	3
P2	2	8	1	8
P3	3	6	3	16


P3
P1
P0
P2
0
6
9
14
22

- Wait time : Service time – Arrival time

- P0 0 – 0 = 0
- P1 3 – 1 = 2
- P2 8 – 2 = 6
- P3 16 - 3 = 13

- Average wait time: (0+2+6+13)/4 = 5.25

4. Round Robin Scheduling

- 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.

- P0 P1 P2 P3 P0 P2 P3 P2

0 3 6 9 12 15 18 21 24

- Wait time : Service time – Arrival time

- P0 (0 – 0) + (12 – 3) = 9
- P1 (3 – 1) = 2
- P2 (6 – 2) + (15 – 9) = 10
- P3 (9 – 3) + (18 – 12) = 12

- Average wait time: (9+2+10+12)/4 = 8 .25

Multi Queue Scheduling

- Multiple queues are maintained for processes.
- Each queue can have its own scheduling algorithms.
- Priorities are assigned to each queue.