SNS COLLEGE OF TECHNOLOGY

Coimbatore-35

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

23CST202- OPERATING SYSTEMS

II YEAR AIML B IV SEM

UNIT 2 – PROCESS SCHEDULING AND SYNCHRONIZATION

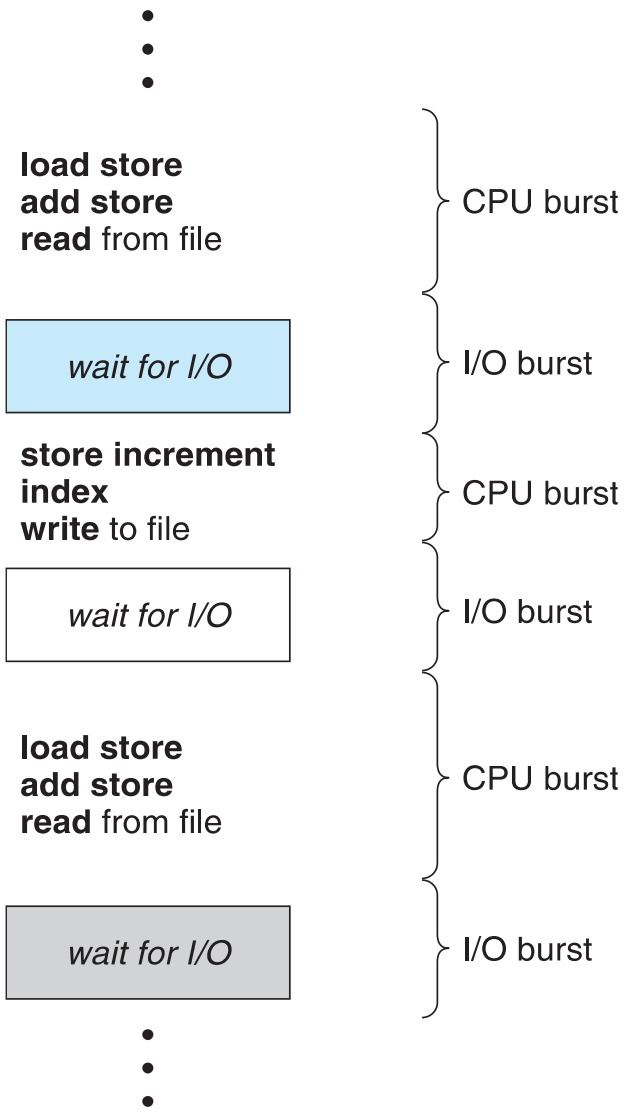
TOPIC – CPU SCHEDULING-SCHEDULING CRITERIA

Objectives

- To introduce CPU scheduling, which is the basis for multiprogrammed operating systems
- To describe various CPU-scheduling algorithms
- To discuss evaluation criteria for selecting a CPU-scheduling algorithm for a particular system

Basic Concepts

- Scheduling is a fundamental operating system function.
- Maximum CPU utilization obtained with multiprogramming.
- The algorithm, the scheduler uses is called scheduling algorithm.
- Several processes are kept in memory at one time.
- CPU–I/O Burst Cycle – Process execution consists of a **cycle** of CPU execution and I/O wait
- **CPU burst** followed by **I/O burst**
- CPU burst distribution is of main concern



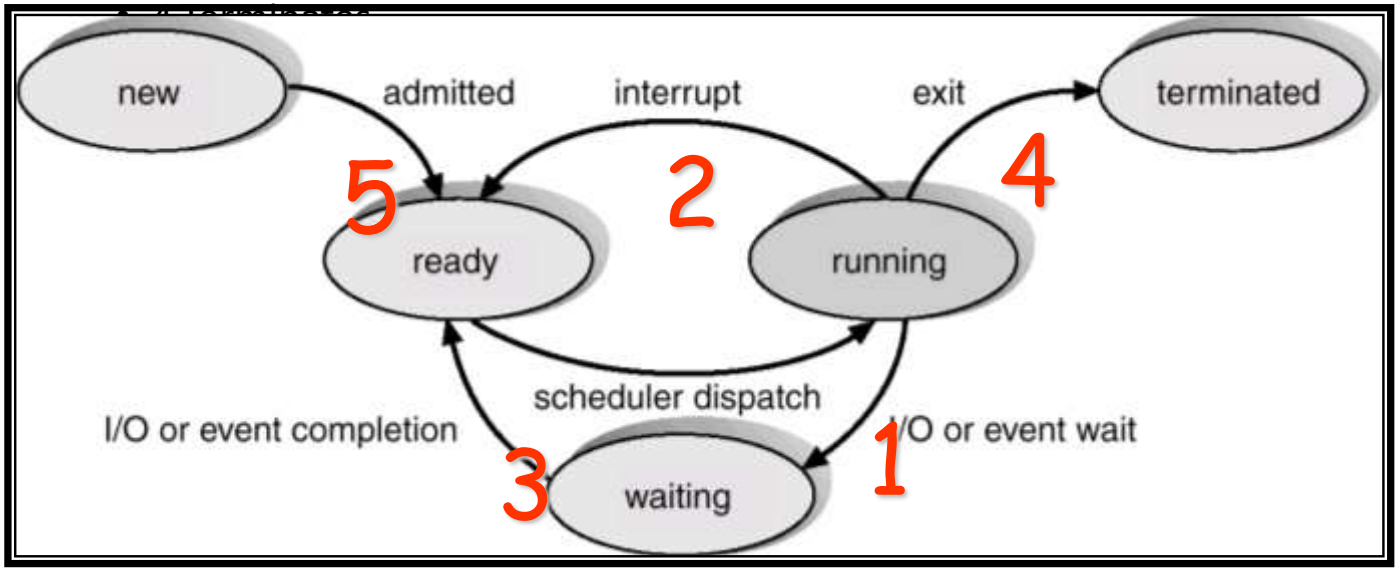
CPU-Scheduling (with multiprograming)

- In a single processor system, only one process can run at a time. Others must wait until CPU is free and can be rescheduled.
- Whenever the CPU becomes idle, the operating system must select one of the processes in the ready queue to be executed (the records in the queues are generally PCBs of the pocesses).
- The short term scheduler (or CPU scheduler- a part of OS) selects a process to get the processor (CPU) among the processes which are already in memory. (i.e. from the ready queue).
- Processor scheduling algorithms try to answer the following crucial question.
  - Which process in the ready queue will get the processor?
- In order to answer this, one should consider the relative importance of several **performance criteria**.

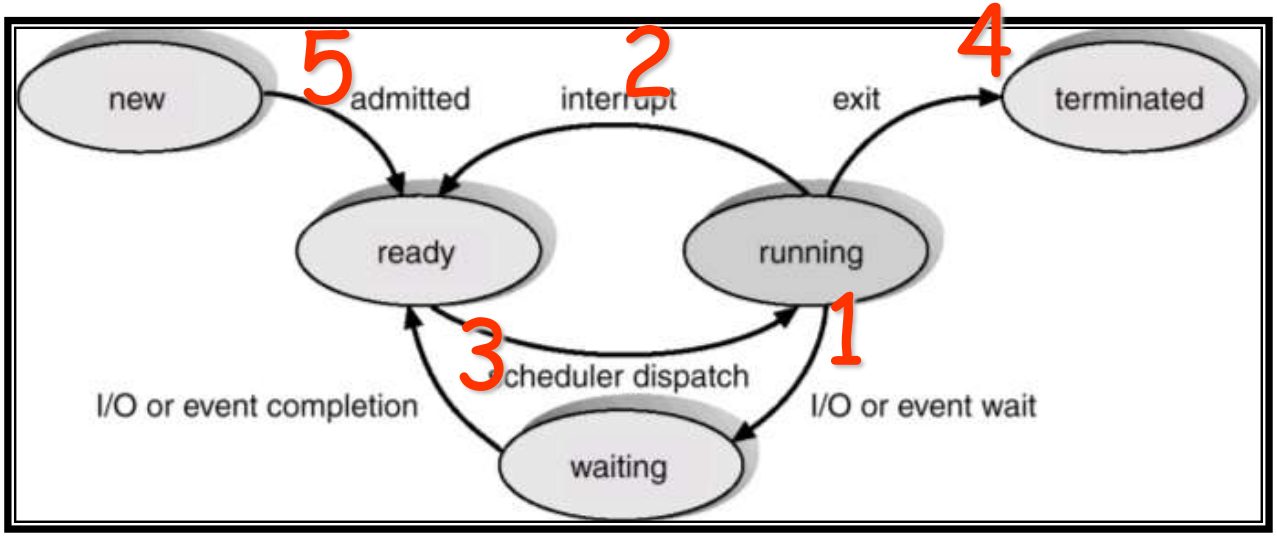
CPU Scheduler

- CPU scheduler selects from the processes in memory that are ready to execute, and allocates the CPU to one of them.

- CPU scheduling decisions may take place when a process:
  - 1.Switches from running to waiting state.
  - 2.Switches from running to ready state.
  - 3.Switches from waiting to ready.



Preemptive – Nonpreemptive Scheduling



Interrupt: Scheduler picks another process
Scheduler dispatch: Scheduler picks this process to execute

- In (1) and (4), a new process must be selected from the ready queue.
- In (2), (3) and (5), previously running process or a new process may be selected.
- Scheduling algorithms that act only in circumstances (1) and (4) are called **nonpreemptive(or cooperative).** Once CPU has been allocated to a process, that process keeps the CPU until it releases the CPU (either by termination or by requesting I/O). This scheduling method was used by Microsoft Windows 3.x.
- Otherwise it is **preemptive.** Windows 95 and all subsequent versions of Windows operating systems have used preemptive scheduling.

Preemptive – Nonpreemptive Scheduling

- A scheduling algorithm which acts on all circumstances is called preemptive. (i.e. such an algorithm can select a new process in circumstances 2, 3 and 5).
- The CPU is allocated to the highest-priority process among all ready processes. The scheduler is called each time a process enters the ready state.
- Advantages of non-preemptive scheduling algorithms:
  - They cannot lead the system to a race condition.
  - They are simple.
- Disadvantage of non-preemptive scheduling algorithms:
  - They do not allow real multiprogramming.
- Advantage of preemptive scheduling algorithms:
  - They allow real multiprogramming.
- Disadvantages of preemptive scheduling algorithms:
  - They can lead the system to a race condition.

Dispatcher

- Dispatcher module (invoked during every process switch) gives control of the CPU to the process selected by the short-term scheduler; this involves:
  - switching context
  - switching to user mode
  - jumping to the proper location in the user program to restart that program

- **Dispatch latency** – time it takes for the dispatcher to stop one process and start another running.

Scheduling Criteria

- CPU utilization - keep the CPU as busy as possible. It is defined as:

  (processor busy time) / (processor busy time + processor idle time)

- Throughput – # of processes that complete their execution per time unit.

- Turnaround time – The interval from the time of submission to the time of completion of a process.

- Waiting time – amount of time a process has been waiting in the ready queue (waiting for I/O device is not counted)

- Response time – amount of time it takes from when a request was submitted until the first response is produced, **not** output.

Performance Criteria

- It is desirable that we,
  - Maximize CPU utilization
  - Maximize throughput
  - Minimize turnaround time
  - Minimize waiting time
  - Minimize response time