CS415 Module 6 Part A - Scheduling Overview

Athens State University

Outline

Contents

1 Need for Scheduling

Processor Scheduling

- Requirement: assign processes to processor(s) in a way that meets system objectives
  - Optimization of one or more of response time, throughput, processor efficiency, and energy use
- Three separate functions:

Long term
Scheduling

Medium term
Scheduling

Short term
Scheduling

Long-term vs. Short-term Scheduling

Long-term Scheduler Select threads to be brought into the ready queue

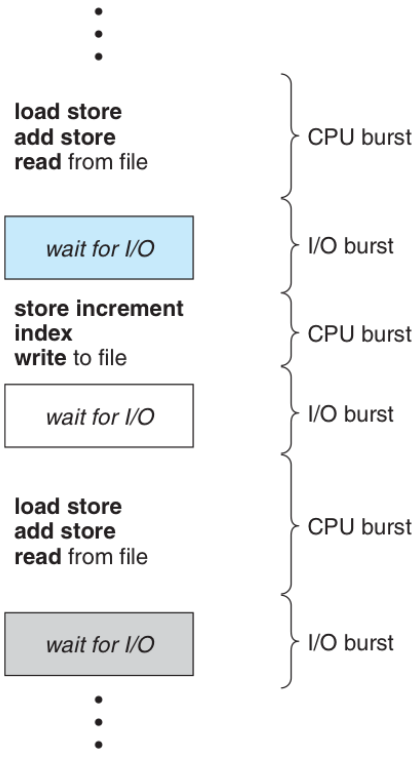
Medium-term Scheduler The decision to add to the number of processes that are partially or fully in main memory

- This is tied to virtual memory management and will be considered later

Short-term Scheduler The decision as to which available thread will be executed by a processor

The short-term scheduler is invoked by the kernel very frequently (usually with a time tick measured in milliseconds) while the long term scheduler is invoked infrequently (seconds to minutes).

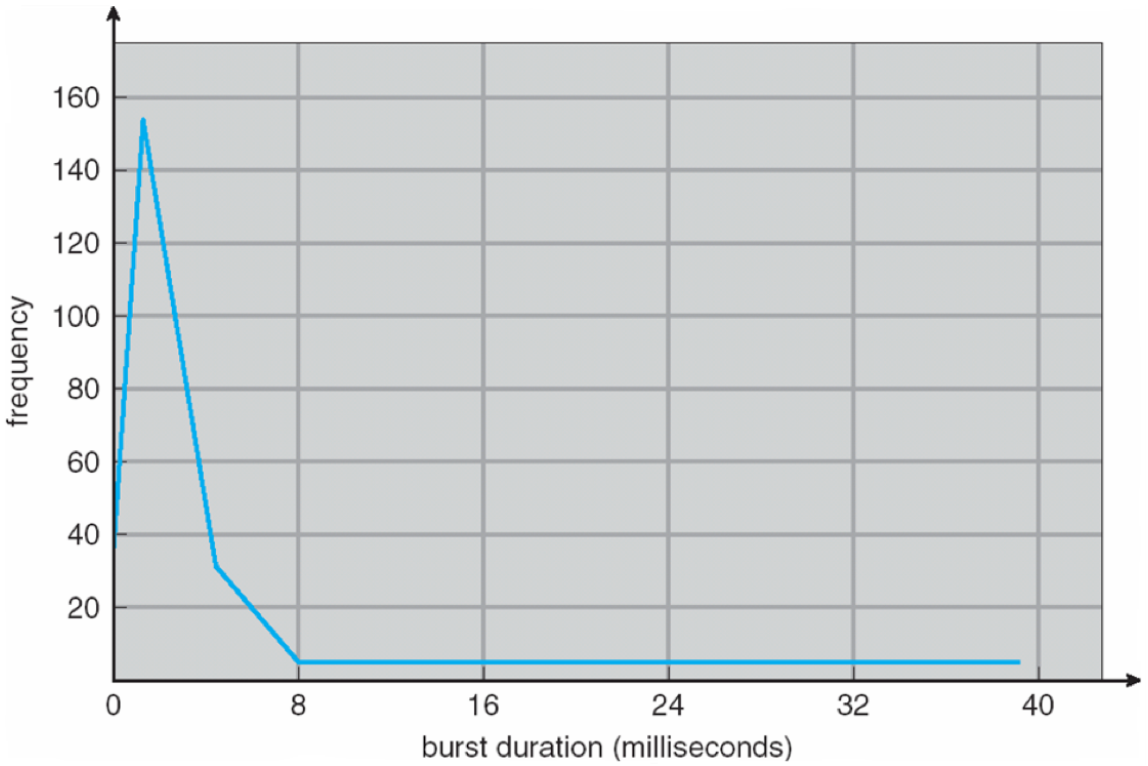
CPU Bound vs. I/O Bound



Burst cycle

- Execution occurs in a cycle of CPU execution and I/O wait
- Scheduling must be concerned with distribution of CPU bursts

CPU Burst Times



Scheduling and Process State Transitions

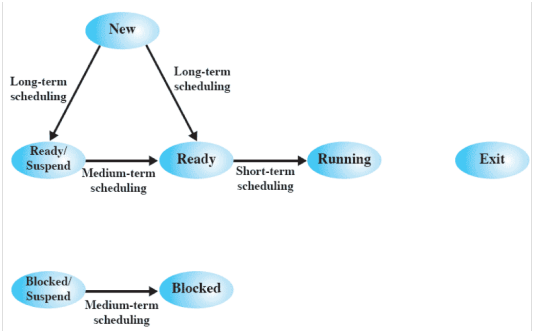
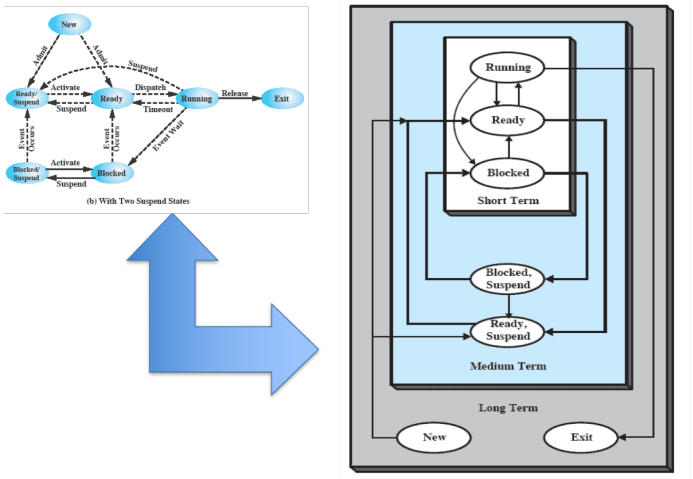


Figure 9.1 Scheduling and Process State Transitions

The Shor-Term Scheduler

- Known as the *dispatcher*
- Makes the fine-grained decision about which thread to execute next
- Invoked when an event occurs that may block the the current process or provide an opportunity to preempt a currently running process in favor of another process
  - Clock interrupts
  - I/O interrupts
  - System calls

– Signals



Scheduling and Process Queues

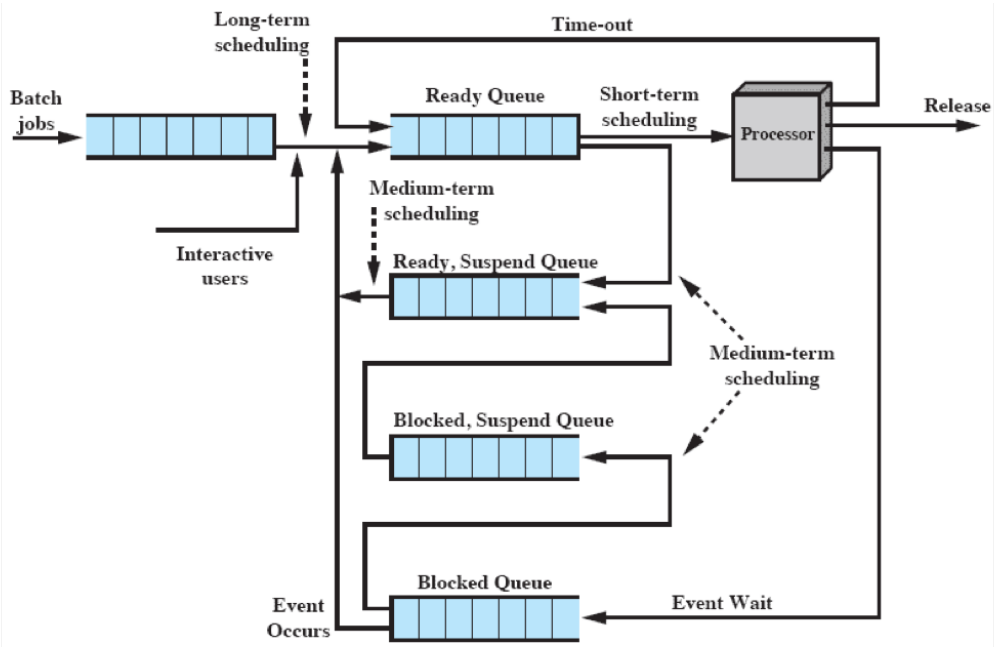


Figure 9.3 Queuing Diagram for Scheduling

2 Scheduling Criteria

Short-Term Scheduling Criteria

A set of criteria is needed to evaluate scheduling policy

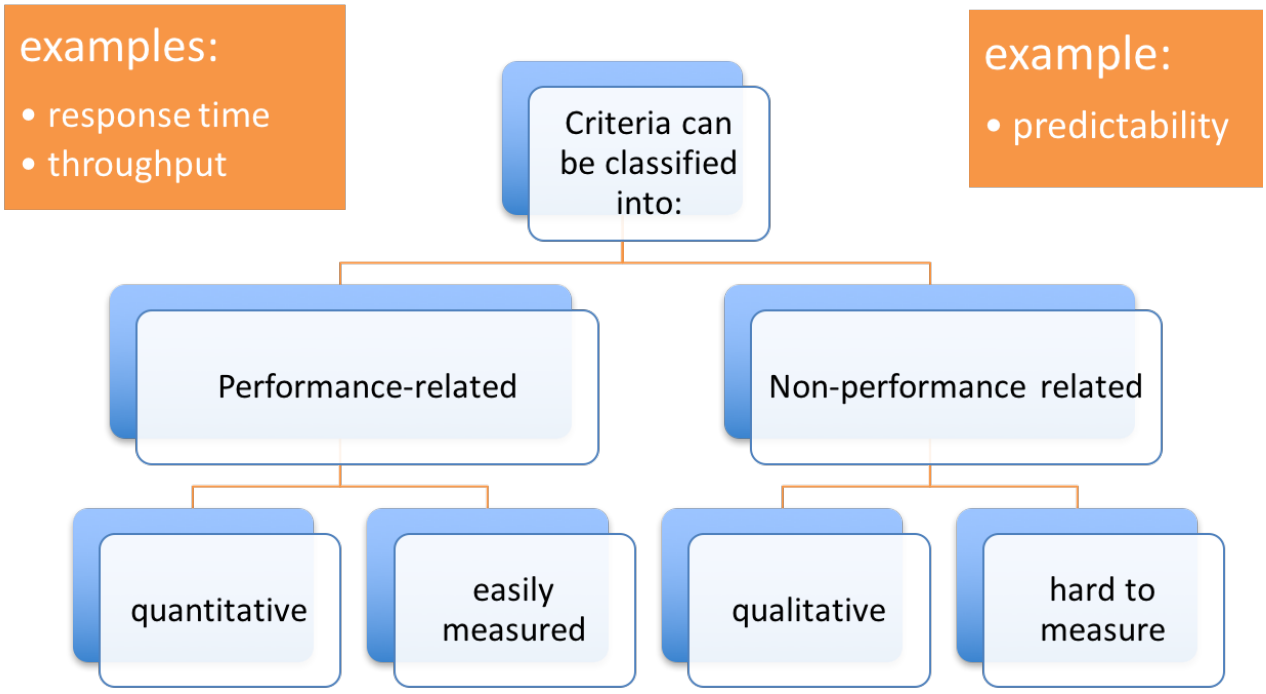
User-oriented criteria

- Relate to the how the user or process perceives the behavior of the system
- Example: Response time in an interactive system

System-oriented criteria

- Focus on effective and efficient utilization of the process
- Example: Rate at which processes complete

Short-Term Scheduling Criteria: Performance



Priority Queuing

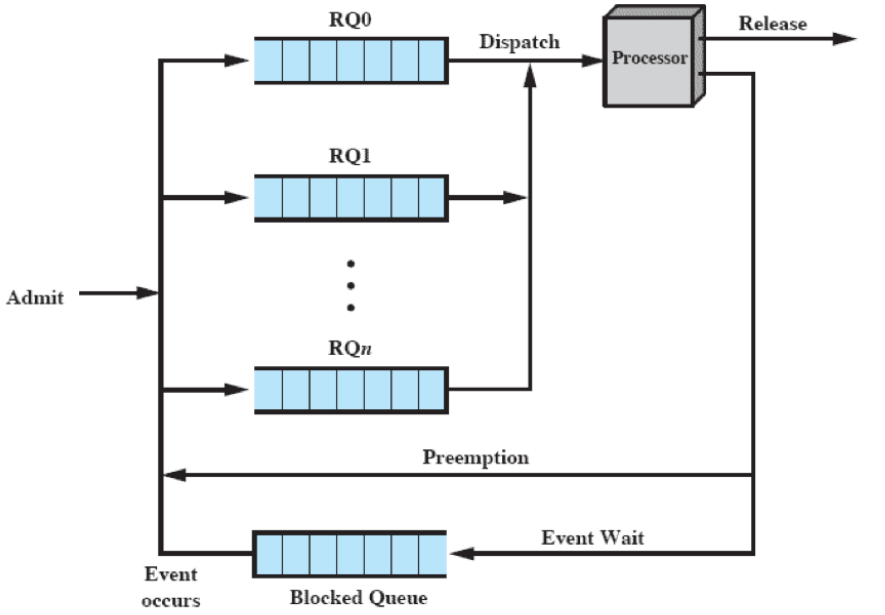


Figure 9.4 Priority Queuing

Selection Function

- Determines which ready thread is selected next for execution
- May be based on priority, resource requirements, or the execution characteristics of the process and thread
- Important execution characteristics
  - Time spent waiting in the system
  - Time spent in execution so far
  - Estimated total service time required by the process

Non-preemptive vs. Preemptive Scheduler

Nonpremptive

- Once a process is in the running state, it continues running until it terminates or must block itself to wait on I/O

Premptive

- Currently running process may be interrupted and moved to the ready state by the OS
- Preemption may occur when a new process arrives, on an interrupt, or periodically

3 Key Points

Key Points

- Types of scheduling
- Scheduling of criteria
- Non-preemptive vs. preemptive