

Operating Systems

COMS W4118

Lecture 23

Alexander Roth

2015 – 04 – 21

1 Scheduling

1.1 Dispatcher versus Scheduler

- Dispatcher is a low-level mechanism.
- It's responsibility is for context-switching.
- Scheduler picks which process to run.
- Scheduler has a scheduling policy, that is a low-level algorithm that decides which process to run at a given point in time.
- The scheduler makes a decision when a process switches from running to waiting state, when it switches from running to ready, or when it switches from waiting to ready, or it terminates.

1.2 Scheduling Algorithm

- There are computation heavy processes, that is, they are CPU bound
- There are a lot of processes that are I/O heavy; that is, they are I/O bound.
- I/O bound processes should have a higher priority when they have something to do, because they involve user interaction.
- However, they are only used temporarily.
- It is better to do calculations in a batch so we can avoid the overhead of context-switching. That is why we use batch systems when there is a long wait time in I/O bound processes.
- There are possible specializations in systems, such as real-time systems. For example, we have embedded systems, such as medical devices.

1.3 Scheduling Performance Metrics

- The shorter the waiting time the better the system will handle.
- A good scheduling algorithm should be wasting CPU cycles as little as possible.
- Waiting time considers the whole entire time until the process completes.

1.4 FCFS Algorithm

- First-come first-serve
- The scheduling time depends on the arrival order.
- So a long process can occur before a short process
- It is more optimal to run the shorter jobs first.

1.5 Shortest Job First

- Schedule the process with the shortest time.
- Advantages - minimizes average wait time.
- Disadvantage - difficult to predict burst time

1.6 Round-Robin

- Context-switch heavy process.
- Uses a time slice method to determine what's running
- If the time slice is too long, then it basically becomes first-come first-serve.
- If the time slice is too short, then it becomes weighted down by context-switches.
- Advantages are there is good interactivity, provided a reasonable time slice.
- Lower average waiting time when job lengths vary widely.
- Fair allocation of CPU across processes.
- Disadvantages are poor average waiting time when jobs have similar lengths.
- Performance depends on the length of the time slice.

1.7 Priorities

- Scheduling systems have the notion of priorities.
- A priority is associated with each process.
- If there is a high-priority process is running, it will not be interrupted by the low-priority functions.
- Round-robin between processes that are of equal priority.
- There is a binary tree structure that separates processes based on priorities.
- Highest priority processes will always run before the lower priorities.
- There is an issue with starvation of the lower priority processes.