Chapter 7 Scheduling

- Scheduling policies are high level rules the OS employs.

7.1 Workload Assumptions

- Workload is the set of processes running on a system.
  - Determining workload is a critical part of building policies.
- Make the following assumptions about processes on a running system for the following discussion.
  - Each job runs for the same amount of time.
  - All jobs arrive at the same time
  - Once started each job runs to completion
  - All jobs only use the CPU
  - The run-time of each job is known
- The assumptions are unrealistic.

7.2 Scheduling Metrics

- Scheduling metrics enable us to compare different scheduling policies.
  - Start by using a single metric: turnaround time.
    - Turnaround time is the time a job completes minus the time the job arrived on the system.
    - Turnaround time is not a performance metric.
  - A second metric is fairness.
    - Performance and fairness are often at odds.

7.3 First In, First Out (FIFO)

- FIFO scheduling is the most basic algorithm to implement.
  - Also called First Come First Serve.
- Positive Properties of FIFO:
  - simple
  - easy to implement
  - works pretty well given the previous assumptions.
- Problems with FIFO
  - Referred to as the convoy effect
  - Several short jobs get bottlenecked behind a very long job.

7.4 Shortest Job First (SJF)

- Solves the confoy effect.
- Problem by relaxing one assumption.
  - Assume that not all jobs arrive at the same time.
    - Shorter jobs arriving after a longer job are still bottlenecked.

7.5 Shortest Time to Completion First (STCF)

- Preempting stops one job to run another job.
  - SJF is a non-preempting scheduler.
- STCF adds preemption to the scheduling to solve the problem of SJF.

7.6 Metric: Response Time

- Response time is the time when a job arrives to the first time it is scheduled.

7.7 Round Robin

- Round Robin address response time.
- RR runs jobs for a time slice then swtiches to the next job in the queue.
    - Sometime RR is referred to as time slicing.
- The shorter the time slice the better the response time metric quality.
    - If the time slice is too Short then the cost of context switching creates too much of a performance drain.
- RR is a bad policy for turnaround time.
    - Any policy that treats all processes the same (fair) does poorly on turnaround time.
    - The trade off is that if you run fairly response time is better and turnaround is poor if you run unfairly, i.e shortest first turnaround is good but response is poor.

7.8 Incorporating I/O

- Relax assumption four that all jobs will only use the CPU.
- When a process is waiting for I/O it is blocked.
    - Won't be using the CPU.
- We can switch to the non-blocked process to execute with the scheduler.

7.9 The Last Assumption

- The OS usually knows very little about the length of a process.