CSCE 311 - Operating Systems

Scheduling

Qiang Zeng, Ph.D.
Fall 2018
Resource Allocation Graph describing the traffic jam
## Conditions for Deadlock

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mutual Exclusion</strong></td>
<td>A process cannot access a resource that has been allocated to another process.</td>
</tr>
<tr>
<td><strong>Hold-and-Wait</strong></td>
<td>A process may hold allocated resources while awaiting assignment of others.</td>
</tr>
<tr>
<td><strong>No Pre-emption</strong></td>
<td>No resource can be forcibly removed from a process holding it.</td>
</tr>
<tr>
<td><strong>Circular Wait</strong></td>
<td>A closed chain of processes exists, such that each process holds at least one resource needed by the next process in the chain.</td>
</tr>
</tbody>
</table>
Dealing with Deadlock

- Three general approaches exist for dealing with deadlock:

  Prevent Deadlock
  - adopt a **policy** that eliminates one of the conditions
  - E.g., numbering resources, and request from low to high

  Avoid Deadlock
  - make the appropriate **dynamic** choices based on the current state of resource allocation
  - E.g., banker’s algorithm

  Detect Deadlock and Recover
  - attempt to detect the presence of deadlock and take action to recover
Main Points

• Scheduling policy: what to do next
  – when there are multiple threads in the ready queue;
  – more generally, multiple packets to send, or web requests to serve
• Definitions
  – response time, throughput, wait time
• Uniprocessor policies
  – (1) FIFO
  – (2) Shortest Job First
  – (3) Round robin
  – (4) Multilevel feedback queue
• (5) Affinity scheduling for multiprocessor policies

Some of the slides are courtesy of Dr. Thomas Anderson
Example

• You manage a restaurant, and the customers complain that they wait forever and starve. What will you do?

• You manage a web site, that suddenly becomes wildly popular. Do you?
  – Buy more hardware?
  – Turn away some users?
  – Implement a different scheduling policy?
Non-preemptive vs Pre-emptive

• A **preemptive** scheduling means that the scheduler can take resources (e.g., CPU) away from the process
  – Recall timer interrupts and CPU time slice

• A **non-preemptive** scheduling means a process occupies the resources until it voluntarily relinquishes the resources
Metrics to evaluate a scheduler

- **Response time**
  - Time elapsed from the time of submission to the first response
- **Throughput**
  - # of tasks can be done per unit of time?
- **Wait time**
  - Time spent on waiting in the ready queue
- **Turnaround time**
  - $T_{\text{finish}} - T_{\text{start}}$
- **Predictability (low variance)**
  - How consistent is the performance over time?
- **Fairness**
First Come First Serve (FCFS)

- Schedule tasks in the order they arrive
  - Continue running them until they complete or give up the processor
- Easy to implement; very small overhead due to scheduling
- Average waiting time can be large if small jobs wait behind long ones (high turnaround time)
  - You have a basket, but wait behind one with a cart
Example

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>BURST TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>21</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
</tr>
<tr>
<td>P3</td>
<td>6</td>
</tr>
<tr>
<td>P4</td>
<td>2</td>
</tr>
</tbody>
</table>

This is the GANTT chart for the above processes.
Shortest-Process-First (SPF) Scheduling

• Scheduler selects process with smallest time to finish
• Advantages: low average wait time
• Disadvantages:
  – Potentially large variance in wait times
    • Long task can be affected by short tasks once and again
    • Starvation
  – Relies on estimates of time-to-completion, which can be inaccurate or unrealistic
Example

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>BURST TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>21</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
</tr>
<tr>
<td>P3</td>
<td>6</td>
</tr>
<tr>
<td>P4</td>
<td>2</td>
</tr>
</tbody>
</table>

In Shortest Job First Scheduling, the shortest Process is executed first.

Hence the Gantt chart will be following:

<table>
<thead>
<tr>
<th>P4</th>
<th>P2</th>
<th>P3</th>
<th>P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>5</td>
<td>11</td>
</tr>
</tbody>
</table>

Now, the average waiting time will be \( \frac{0 + 2 + 5 + 11}{4} = 4.5 \) ms
Round Robin

• Each task gets resource for a fixed period of time (time quantum, or CPU time slice)
  – If task doesn’t complete, it goes back in line

• Need to pick a time quantum
  – What if time quantum is too long?
    • Becomes FIFO
  – What if time quantum is too short?
    • Large overhead for context switch

• Advantage: fairness (we will further discuss this)

• Disadvantage
  – many context switches bring a large overhead;
  – large wait time
Round Robin = Fairness?

• Is Round Robin always fair?
  – No! See the next slides
CPU bound vs I/O bound

• If a process’s speed is mainly determined by the CPU speed, the process is CPU-bound
  – E.g., multiply matrix

• If a process’s speed is mainly determined by the I/O speed (i.e., most of the time the process blocks for I/O), the process is I/O-bound
  – E.g., emacs
Mixed Workload

Tasks

I/O Bound

CPU Bound

CPU Bound

Issues I/O Request

I/O Completes

Issues I/O Request

I/O Completes

Time
Solution: **Multi-level Feedback Queue**
Multi-level Feedback Queue (MFQ)

- Used in Linux, Windows, MacOS, and Solaris
  - Each system adopts the algorithm with some little modifications
- First developed by Corbato et al; led to Turing Award
MFQ

• Has a set of Round Robin queues
  – Each queue has a separate priority
• High priority queues have short time slices, while low priority queues have long time slices
• Scheduler picks the first thread in highest-priority non-empty queue
• When a process is scheduled out, it is inserted in queues following the two rules
  – If time slice expires, task drops one level
  – If the process relinquishes the slice due to I/O, it is kept in the current priority queue
• Optionally, for a process in the base level queue that becomes I/O bound, it can be promoted to the next-higher queue
<table>
<thead>
<tr>
<th>Priority</th>
<th>Time Slice (ms)</th>
<th>Round Robin Queues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>80</td>
<td></td>
</tr>
</tbody>
</table>

New or I/O Bound Task
Time Slice Expiration
Multiprocessor Scheduling

• What would happen if we used MFQ on a multiprocessor?
  – Contention for scheduler spinlock
  – Poor CPU cache reuse
Per-Processor Affinity Scheduling

- Each processor has its own MFQ
  - Each MFQ is protected by a per-processor spinlock
- When the system puts threads back on the ready list, they are put back where they had most recently run
- But idle processors can steal work from other processors
Per-Processor Multi-level Feedback with Affinity Scheduling
Summary

• Scheduling policy: what to do next, when there are multiple threads ready to run
• Response time, throughput, wait time
• Uniprocessor policies
  – FIFO, Shortest Job First
  – round robin
  – multilevel feedback as approximation of optimal
• Multiprocessor policies