Scheduling

COSC-361

Stephen Marz

TICKLE COLLEGE OF ENGINEERING

MIN H. KAO DEPARTMENT OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

1

Scheduling

- The act of picking which process(es) will use the CPU and in which order.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

2

Short Term vs Long Term

- Long term is given to new processes.
  - Much like a "promise" that the scheduler will get to it sometime.
- Short term is given to those "tenured" processes.
  - Has more control and guarantees to already executing processes.

3 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

3

Cooperative Scheduling

- Cooperative scheduling cannot halt a running process.
  - The process runs until it:
    - Yields
    - Waits for a resource
- Can lead to ***starvation***.

4 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

4

Starvation

- The point a which a process has not given enough CPU and the effects are apparent.
- Example: a customer keeps the waiter at their table, and hence you're unable to order your food.

5 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

5

Preemptive Scheduling

- The OS will interrupt a process to give another a share of the CPU.
- Can still lead to starvation for critical processes.

6 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

6

Prioritization

- Some processes may be prioritized above others.
  - They get more time on the CPU than others.
- Starvation can occur if high-priority processes keep getting the CPU over low-priority processes.

7

Arbitration

- What happens if two processes have equal priority?
- Schedulers may incorporate arbitration rules based on:
  - Creation time
  - CPU Runtime
  - # of context switches

8

Batch Scheduling

- Batch scheduling assumes the start and end times are known.
  - Batch processes run until they're done.
    - Cooperative
- FCFS (first-come, first-served)
  - The first process uses the CPU until it's done, then the second, then the third, and so on.
- SJF (shortest job first)
  - The scheduler sorts all processes in their runtime order (shortest first).
  - Most efficient type of batch scheduling algorithm.

9

Interactive Scheduling

- Processes with an indefinite runtime.
  - Most process you work with are interactive.
- Algorithms
  - Round-robin
  - Multi-level
  - Multi-level feedback

10

Real-time Scheduling

- Soft-real-time
  - Process must run at a certain time +/- some "slop".
- Hard-real-time
  - Process run at a certain time +/- a very minimal amount of "slop".

11

Definitions

- Quantum (Q) - The amount of CPU time a process gets per context switch.
- Quantum Multiplier (xQ) - The coefficient given to a Quantum.
  - Ex: 2Q means a process is given 2x the CPU time.

12

Round Robin Scheduling

- Priorities are ignored
- Starts at the top of the queue, goes to the next, then the next, and so forth.
  - At the end of the queue, it goes back to the top.
- Problems
  - Inefficient
  - Can't handle priorities
  - Can lead to starvation for larger process lists.

13 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

13

Multilevel Scheduling

- Sorts processes in a sub-list based on its priority.
  - Requires a queue for each potential priority. Ex: Priorities range from 1 to 20 -- we need 20 queues.
- Round robin is used in each sub-list.
- Higher priority processes can prevent lower priority processes from ever running.
  - As long as one higher-priority process is running, no lower-priority process ever runs.

14 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

14

Multilevel Feedback

- Fixes starvation in ML scheduling by dynamically changing priorities.
- After a process runs, its priority is decremented and it is moved into the next lower queue.
- This prevents higher priority processes from hogging the CPU.

15 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

15

MLF Scheduling

- All processes enter the lowest priority queue N.
  - After process runs, it is put into N-1 queue.
- Queue N is given Q CPU time
- Queue N-1 is given 2Q CPU time
- Queue N-2 is given 3Q CPU time
  - Some MLF algorithms always double Q, but this can lead to unresponsive interactivity amongst processes.
  - Longer time quantum reduces impact of context switching time.
- Priority is inversely correlated with CPU time. (To favor shorter processes).

16 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

16

Completely Fair Scheduler

- CFS is default Linux scheduler
- It attempts to have a level vruntime
  - vruntime = Amount of CPU a process has been given through its lifetime.
- Uses red-black trees keyed on vruntime
  - Lowest vruntime always hits the leaf on the leftmost branch.
  - Very efficient to grab next process for scheduling.

17 13-Feb-19 COSC 361 THE UNIVERSITY OF TENNESSEE KNOXVILLE

17

Scheduling
COSC-361
Stephen Marz

THE UNIVERSITY OF TENNESSEE KNOXVILLE

18