Jeremy Holloway
CPSC-3220-001
6/3/2019

Pre-class Assignment #12

1. How frequently does each of scheduler listed below run? (see slides)
  short-term scheduler (a.k.a. dispatcher) – runs after every interrupt.
  medium-term scheduler (a.k.a. swapper) – runs every few seconds.
  long-term scheduler (a.k.a. batch job initiator) – minutes.

2. What are three performance problems identified by the textbook when the thread scheduler uses a single ready list for a multiprocessor?
    1. Contention for the MFQ lock.
    2. Cache coherence overhead.
    3. Limited cache reuse.

3. When should threads be moved from one per-processor ready list to another one?
    When another processor is idle and can accomplish the work instead of it sitting in its current processes queue waiting to run.

4. Define the following terms:
  affinity scheduling – A scheduling policy where tasks are preferentially scheduled onto the same processor they had previously been assigned, to improve cache reuse.
  oblivious scheduling – A scheduling policy where the operating system assigns threads to processors without knowledge of the intent of the parallel application.
  priority donation (a.k.a. priority inheritance) – When a thread waits for a lock held by a lower priority thread, the lock holders priority is temporarily increased to the waiter's priority until the lock is released.

5. Define the following terms: (see slides)
  priority aging – threads priority to decrease as it runs and increase as it waits.
  priority boosting – threads priority increases/decreases when signaled.

6. What is Little's Law and under what condition is it applicable?
    Little's law is a theorem provided by John Little in 1961 that applies to any stable system where the arrival rate matches the departure rate.

7. What are two performance problems identified by the textbook when server utilization is high?
    1. Response time
    2. Utilization

8. Why are bursty arrivals of tasks more of a problem than a steady arrival rate?
    This is because queues tend to be full during busy periods and empty during the idle periods, so few requests enjoy high priority and many suffer long queue waits.

9. What are two possible responses identified by the textbook to manage overload?
    1. Reject requests
    2. Sophisticated scheduling

10. Identify and explain at least one reason why a system might have have to do more work per request as load increases.

This is because the rest of the system is also running at a high capacity so response times overall will be slower, which results in all individual responses to be slower.