15-857/47-774 Homework 12: Scheduling

Special due date: Homework is due on **Tuesday, Dec 2, at 5 p.m.**. Please turn these in at Mor's office: GHC 7207. Solutions will go out at Wednesday's class.

These problems are from your textbook, *Performance Modeling and Design of Computer Systems*. Starred problems are either not in your textbook, or have some modifications, given below.

<u>Exercises</u>: **28.1* ; 29.2 ; 30.3* ; 30.5(a,c) ; 30.7 ; 31.2 ; 32.2* ; 32.3***

Exercise 28.1*: [From Transforms to Tails]
Let T_Q denote the Laplace transform of waiting time in an M/D/1 queue (Deterministic serivice time). Assume that the job size is $S = 1$, and the arrival rate is $\lambda = 0.5$. Your goal is to figure out $\mathbf{P}\{T_Q > t\}$, where we will assume that $t = 2$.

(a) Argue that for any non-negative r.v. X, $\mathbf{P}\{X \geq a\} \leq \frac{\mathbf{E}[X]}{a}$, for any $a > 0$. In your proof, assume that X is continuous.

(b) Argue that, for any $s > 0$,

$$\mathbf{P}\{T_Q > t\} = \mathbf{P}\left\{e^{sT_Q} > e^{st}\right\} \leq \frac{\widetilde{T_Q}(-s)}{e^{st}}$$

(c) Let $t = 2$. Evaluate the upper bound of $\mathbf{P}\{T_Q > t\}$ obtained in (b), and plot this as a function of $s > 0$. Eyeball the minimizing s and report the upper bound on $\mathbf{P}\{T_Q > 2\}$ that you obtain.

(d) Before you get too excited about this way of obtaining an upper bound on the tail, try it for the case where $t = 0.5$. What happens?

Exercise 30.3*: [FB versus PS under Exponential Workloads]
This problem is in your book. Hint: Use a CTMC.

Exercise 31.2*: [Can adding variability reduce response time?]
Consider an M/M/1 queue with average arrival rate λ and average service rate μ. In an effort to reduce average response time, Ying suggests adding variability to the service rate, by serving some jobs at rate $k\mu$ and serving other jobs at rate μ/k, for some $k > 1$. Specifically, Ying argues for changing the job size distribution, S, to the following:

$$S = \begin{cases} \text{Exp}(k\mu) & \text{w/prob } \frac{k}{k+1} \\ \text{Exp}(\mu/k) & \text{w/prob } \frac{1}{k+1} \end{cases}$$

and for giving jobs with shorter expected size non-preemptive priority over jobs with longer expected size. Observe that $\mathbf{E}[S]$ is still $\frac{1}{\mu}$ as before, but job sizes have more variability, for higher k.

Derive mean response time for Ying's altered system and compare with the original M/M/1. It will help to use a program like Mathematica or Matlab to simplify computations. Consider the case where $\mu = 1$ and $\lambda = 0.98$. Is there a value of k for which Ying's system is superior? Is there a value of k for which the original M/M/1 is superior? Show your formulas.

Exercise 32.2*: [The $c\mu$-Rule]
This problem is in your book, however the last line of this problem in part (d) has a typo. The μ should be μ_i.

Exercise 32.3*: [Preemptive Priority Transform]
Consider an M/G/1 with two priority classes: L (low priority) and H (high priority). Low priority jobs arrive with rate λ_L and have size S_L. High priority jobs arrive with rate λ_H and have size S_H. Let T_L denote the response time of the low priority jobs. Derive $\widetilde{T_L}(s)$.

[Hint: This will be much easier if you think about the response time as composed of a waiting time plus a residence time, then figure out each of those separately. Your answer will involve busy periods all over the place. Be careful to define these busy periods in words. Some of these will be busy periods started by certain jobs, or started by certain work. It may be helpful to use notation like, B_H, to express the duration of a busy period made up of only high-priority jobs.]