CSE 538 – Fall 2016 Final

3 Problems – 75 points total

Your Name:

Problem 1 [30 points] Consider a single unit rate server system with an infinite storage capacity, where jobs arrive according to a Poisson process of rate λ. Jobs sizes follow an exponential distribution with parameter μ. Assume $\frac{\lambda}{\mu} < 1$.

1. **[10 points]** Assume that the system is configured so that jobs of size $< \Delta$ are queued in FCFS order *ahead* of all the jobs of size $\geq \Delta$ that are also queued in FCFS order. In other words, small jobs (of size less than Δ) are served first ahead of jobs of size greater than or equal to Δ. In both categories jobs are served in order of their arrival, and once in service jobs cannot be preempted.

   Provide and expression, function of λ, μ, and Δ for the system's overall average response time $E[T]$.

Write your answer in the space below

2. **[5 points]** Assuming $\lambda = 0.5, \mu = 1$, what is the value Δ^* that minimizes the system response time $E[T]$ and how much smaller is $E[T]$ than if all jobs were served in FCFS order, irrespective of their size?
Note: Either derive an explicit expression for Δ^* and solve for its value when $\lambda = 0.5, \mu = 1$, or write a small numerical search procedure to compute its values (the latter may be faster).

Write your answer in the space below

3. **[10 points]** Consider now a configuration where we split the single unit rate server into two servers. Jobs of size less than Δ are assigned to the first server, while jobs of size greater than or equal to Δ are sent to the second server. The unit capacity of the original server is split between the two servers so that both have the same load that, therefore, remains equal to $\rho = \frac{\lambda}{\mu}$.

   Provide and expression, function of λ, μ, and Δ for the system's overall average response time $E[\widehat{T}]$ in this two-server configuration.

Write your answer in the space below

4. **[5 points]** Assuming as in question 2 that $\lambda = 0.5, \mu = 1$, what is the value $\widehat{\Delta}^*$ that minimizes the system response time $E[\widehat{T}]$, and how does it compare to the value obtained in question 2 for the configuration where small jobs were served ahead of big jobs? Can you explain the difference?
Note: Again, either derive an explicit expression for $\widehat{\Delta}^*$ and solve for its value when $\lambda = 0.5, \mu = 1$, or write a small numerical search procedure to compute its values.

Write your answer in the space below

Problem 2 [30 points] Consider an M/G/1 system with a unit rate server where jobs arrive according to a Poisson process of rate λ, but where when the system empties out, service resumes only after k jobs have arrived (as opposed to as soon as the first job arrives). Once service resumes, it proceeds as in a regular M/G/1 system until the system is again empty. Assume also that $\lambda \cdot E[S] < 1$. Under those assumptions, show that

1. **[5 points - 1, 2, 2 points]** In steady-state, we have

$$\begin{aligned} P\{\text{system is non-empty and serving}\} &= \rho \\ P\{\text{system is non-empty and waiting}\} &= \frac{(k-1)(1-\rho)}{k} \\ P\{\text{system is empty}\} &= \frac{1-\rho}{k} \end{aligned}$$

Write your answer in the space below

2. **[5 points]** The average length of a "busy" period is given by

$$E[B_{(k)}] = \frac{\rho + k - 1}{\lambda(1-\rho)}$$

where a busy period is defined as a period of time during which the system is not-empty.

Write your answer in the space below

3. **[15 points]** Assuming that a busy period is divided into busy/waiting and busy/serving periods, show that the average number of jobs in the system during a busy/waiting period is $\frac{k}{2}$ and the average number of jobs in the system during a busy/serving period is

$$E[N_{\text{busy/serving}}] = \frac{E[N_{\text{M/G/1}}]}{\rho} + \frac{k-1}{2}$$

where $E[N_{\text{M/G/1}}]$ is the average number of customers in a regular M/G/1 system.
Hint: Relate the busy/serving period to k independent busy periods of a regular M/G/1 system.

Write your answer in the space below

4. **[5 points]** Finally, show that the average number of jobs in the system, $E[N_{(k)}]$ is given by

$$E[N_{(k)}] = E[N_{\text{M/G/1}}] + \frac{k-1}{2}$$

where $E[N_{\text{M/G/1}}]$ is again the average number of customers in a regular M/G/1 system.

Write your answer in the space below

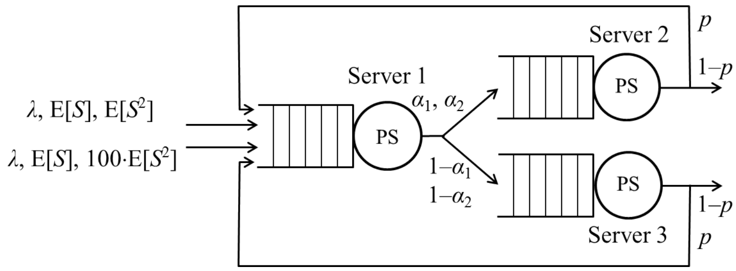

Figure 1: Open network of PS servers with two types of jobs.

Problem 3 [15 points] Consider the open network of Fig. 1 consisting of three unit rate PS servers with infinite storage capacity. Two types of jobs arrive to the network at server 1. Upon leaving server 1, jobs of type $i, i = 1, 2$, can be directed to either server 2 or server 3 with probabilities α_i and $1 - \alpha_i$, respectively, and when leaving either server 2 or 3, jobs of both types depart the system with probability $1 - p$ or return to server 1 with probability p. Both types of jobs arrive according to a Poisson process of the same rate λ, and have general service time distributions. We also know that $E[S_1] = E[S_2] = E[S]$, but that $E[S_1^2] = 100 \cdot E[S_2^2]$.

1. **[5 points - 2, 3 points]** What relationship should λ satisfy to ensure that the system is stable, and what is the arrival rate of type $i, i = 1, 2$, jobs at server 2?

Write your answer in the space below

2. **[10 points]** Identify a pair of values (α_1^*, α_2^*) that minimizes the overall system response time $E[T]$. Justify your answer.

Write your answer in the space below