CSE 538 – Fall 2015 Final
4 Problems – 80 points total

Your Name:

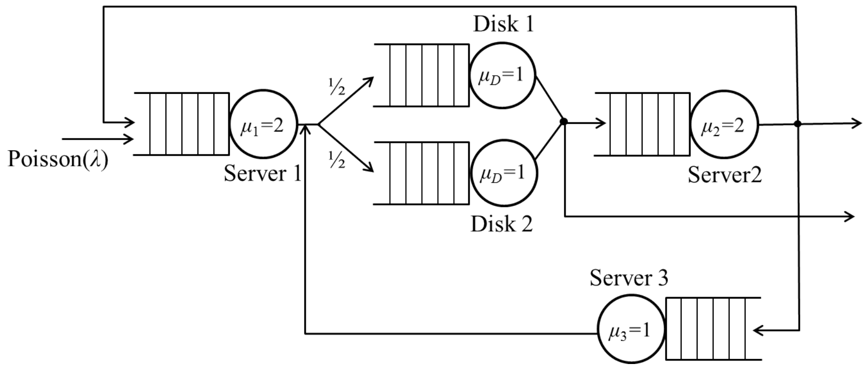

Figure 1: Transactions processing system.

Problem 1 [15 points] Consider the transactions processing system of Fig. 1, which consists of three servers, servers 1 to 3, and a Disk sub-system consisting of two disks, Disks 1 and 2. Transactions arriving at the Disk sub-system are load-balanced equally between the two disks. Transactions enter the system at server 1, but the path they take through the system depends on the transaction type. There are four (4) types of transactions, a to d, that the system handles, each with the following path through the system:

1. T_a: Server 1–Disk–Server 2–Out
2. T_b: Server 1–Disk–Server 2–Server 1-Disk–Out
3. T_c: Server 1–Disk–Server 2–Server 3-Disk–Out
4. T_d: Server 1–Disk–Server 2–Server 3–Disk–Server 2–Out

Transactions arrive to the system according to a Poisson process of rate λ. A new transaction is of type k with probability $p_k = \frac{1}{4}, k = a, b, c, d$. Server i has a service rate Exp(μ_i), $i = 1, 2, 3$, as shown on the figure, *i.e.,* $\mu_1 = \mu_2 = 2, \mu_3 = 1$ and the two Disks in the Disk sub-system each have a service rate Exp(μ_D), where $\mu_D = 1$.

1. **[5 points]** What is the maximum total transactions arrival rate λ^* below which the system is stable?

Write your answer in the space below

2. **[5 points]** What is the average response time through the system $E[T_d]$ for transactions of type d, assuming that the total transactions arrival rate is $\lambda = 1$?

Write your answer in the space below

3. **[5 points (3+2)]**
(i) **[3 points]** What is the probability that Server 2 is busy serving a transaction of type d?

Write your answer in the space below

(ii) **[2 points]** What is the probability that there are exactly one transaction of type c and one transaction of type d at Server 3?

Write your answer in the space below

Problem 2 [30 points] Consider an M/G/1 system with vacations, *i.e.,* arrivals are Poisson with rate λ, and service times follow a general distribution with p.d.f. $f_S(t), t \geq 0$, and whenever the server becomes idle it goes on "vacation" for a period of duration V, where V follows a general distribution with p.d.f. $f_V(t), t \geq 0$. The one difference with the vacation system we have studied is that the server goes on vacation only once. When it returns from vacation, it either starts serving jobs that are present, if any, or waits until the first arrival to start work again, *i.e.,* it does not go on another vacation if it returns from vacation to an empty system.

1. **[15 points]** Under these assumptions, derive an expression for $E[B_{V_1}]$, the average duration of a busy period, *i.e.,* the server is busy serving.
   Hint 1: Express the probability f_0 of 0 arrival in a vacation as a function of the LST $\widetilde{V}(s)$ of $f_V(t)$.
   Hint 2: Use this result to first obtain an expression for the duration of an average idle period $E[I]$, and then use this to find $E[B_{V_1}]$ by leveraging the relationship between the system utilization ρ, $E[I]$, and $E[B_{V_1}]$.

Write your answer in the space below

2. **[15 points]** Derive an expression for the average waiting time $E[T_Q^{V_1}]$ of an arrival to the system.
   Hint: Follow the same derivation as that on p. 397 of the book, but accounting for the fact that the "unfinished work at the server" now has two components; one when the server is busy, as before, as well as one when the server is not busy. In this latter case, the "unfinished work" is the excess time of the vacation **if** the server is still on vacation, and 0 otherwise. You also need to make sure you correctly "weigh" the different components of the unfinished work, *i.e.,* by their respective probabilities.

Write your answer in the space below

Problem 3 [25 points] Consider the two servers system of Fig. 2, where the first server is configured to operate according to a FCFS policy, while the second server operates according to a PS policy. Both servers have the same unit service rate. We will be exploring different job assignment policies to the two servers.

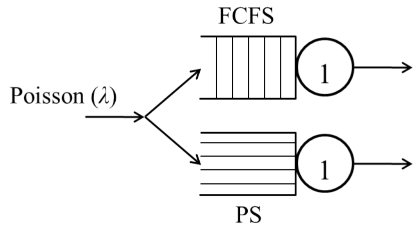

Figure 2: FCFS and PS servers system.

1. **[5 points]** Assume that incoming jobs have a general service time distribution with p.d.f. $f_S(t), t \geq 0$ and first and second moments $E[S] = 1$ and $E[S^2] = 10$, respectively. Our first policy assigns incoming jobs to the FCFS server with probability p and to the PS server with probability $(1 - p)$. obtain an expression, function of λ, for the overall average response time $E[T_p]$ of this assignment policy. Explicitly identify constraints that p and λ must satisfy for an assignment to be feasible.

Write your answer in the space below

2. **[5 points]** Assuming that $\lambda = 1$, identify an expression for the value p^* of p that minimizes $E[T_p]$, and use it to numerically compute the value of p^*.

Write your answer in the space below

3. **[5 points]** Next, we consider a policy that distinguishes between small jobs and large jobs. Jobs still arrive according to a Poisson process of rate λ (assume that λ is small enough to ensure stability) but now come in two sizes, namely, $S_1 = 1$ and $S_2 = 99$, with 99% of the jobs being small and 1% of the jobs being large. Small and large jobs are sent to a different server.
   Should small jobs be sent to the FCFS server or the PS server to minimize their average response time? Rigorously justify your answer. No points will be awarded without quantitative reasoning behind it.

Write your answer in the space below

4. **[5 points]** Assume the job size distribution of the previous question, and that small jobs are sent to the FCFS server and large jobs to the PS server. Is the average response time of small jobs under this configuration always smaller than the average response time they would have experienced if both job types had been sent to a single FCFS server that is twice as fast? Assume again that λ is small enough to ensure stability.

Write your answer in the space below

5. **[5 points]** Repeat the previous question but now comparing sending small jobs to a dedicated FCFS server to a configuration where all jobs, small and large, share a single PS server that is again twice as fast. How does the average response time of small jobs in this latter scenario compare to when they have a dedicated FCFS server?

Write your answer in the space below

Problem 4 [10 points] Consider a single server system where arrivals are according to a Poisson process of rate λ, and where service times follow a two-phase Coxian distribution with a first phase of average duration $E[S_1] = \frac{1}{\mu}$ and a second phase of average duration $E[S_2] = \frac{10}{\mu}$. Jobs transition from the first phase to the second phase with probability $p = 0.5$. The server uses a processor sharing (PS) service discipline. Assume $\lambda E[S] < 1$.

1. **[5 points]** What are, as a function of μ and λ, the average response times of jobs that only require the first phase of the Coxian service distribution, and of jobs that require both phases of the Coxian distribution?

Write your answer in the space below

2. **[5 points]** How much faster should the processor be in order to halve the average response time of jobs that only require the first phase of the Coxian distribution? By how much would such a faster processor improve the overall average response time over all job types? Rigorously justify your answer. No points will be awarded without quantitative reasoning behind it.

Write your answer in the space below