

*I encourage you to discuss these problems with others, but you need to write up the actual solutions alone. At the top of your homework sheet, please list all the people with whom you discussed. Crediting help from other classmates will not take away any credit from you. Also, please limit your use of the web as it is difficult to develop good problems, and so enough search will probably lead you to find the solutions online.*

## 1 Warmup: Practice with Poisson processes [15 points]

- (a) Consider a stream of packets arriving according to a Poisson process with rate  $\lambda$  per second. Suppose each packet is of type “Green” with probability 5% and of type “Yellow” with probability 95%. Given that 100 Green packets arrived during the previous second, (i) what is the expected number of Yellow packets which arrived during the previous second? and (ii) what is the probability that 2000 Yellow packets arrived during the previous second?
- (b) Red packets arrive according to a Poisson Process with rate  $\lambda_1 = 30$  packets/sec. Black packets arrive according to a Poisson Process with rate  $\lambda_2 = 10$  packets/sec. Assume the streams are statistically multiplexed into one stream. Suppose we are told that 60 packets arrived during the second. What’s the probability that exactly 40 of those were red? You can verify your answer by reasoning out this problem in 2 different ways.
- (c) Suppose packets arrive according to a Poisson Process with rate  $\lambda$  and you are told that by in 30 seconds 100 packets have arrived. What is the probability that 20 packets arrived during the first 10 seconds? Explain your answer intuitively.

## 2 Warmup: DTMC to CTMC review [10 points]

Consider the CTMC shown in Figure 1 below.

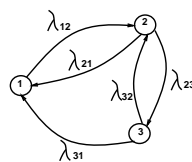


Figure 1: A simple CTMC.

In lecture we saw how to model any CTMC as a DTMC where the timestep in the DTMC is of length  $\delta$ . Show the corresponding DTMC for the CTMC in Figure 1. Then, write out the DTMC balance equations and take the limit as  $\delta \rightarrow 0$  and show the resulting CTMC balance equations.

## 3 More about slowdown [10 points]

Recall in class we saw that for an M/M/1 queue when you increase both the arrival rate and service rate by a factor of  $k$ , the result is a drop in mean response time,  $E[T]$  by a factor of  $k$ . What happens to the mean

slowdown,  $E[S]$ ? Give intuition to explain your result.

#### 4 The distribution of response time [15 points]

In class, we derived the mean response time for an M/M/1 queue, but we often care about the whole distribution of response time – not just its mean. In this problem you will derive the Laplace transform of the response time.

1. Let  $L_X(s) = E[e^{-sX}]$  be the Laplace transform of a random variable  $X$ . Further, for continuous random variables  $Y, A, B$ , define  $Y = \begin{cases} A, & \text{with probability } p; \\ B, & \text{with probability } 1 - p. \end{cases}$   
Prove that  $L_Y(s) = pL_A(s) + (1 - p)L_B(s)$ .
2. Use the above, plus PASTA to derive the Laplace transform of the response time in an M/M/1 queue,  $L_T(s)$ . Do you recognize the resulting Laplace transform as a particular distribution? Hint: Start by conditioning on the number of jobs seen by an arrival.

#### 5 A finite buffer queue [25 points]

Consider an M/M/1/ $n$  queue. That is, consider a queue with Poisson arrivals, Exponential job sizes, a single server that is scheduled according to FCFS, and a finite buffer size of  $n$  (that is  $n - 1$  slots in the queue and 1 slot at the server). If a job arrives when the buffer is full, it is rejected (lost).

Think of the M/M/1/ $n$  as a simple model of a router. Your goal is to reduce the loss probability of the router. You have some money to spend, and it can either (i) double the buffer size or (ii) double the processor speed. Which should you do?

Here's the plan:

- (a) Set up the CTMC.
- (b) Derive the stationary probabilities.
- (c) Derive the utilization of the system (i.e., the probability the system is busy).
- (d) Derive the loss probability (the fraction of jobs that are rejected) – don't forget to use "PASTA" in your justification.
- (e) What is the rate at which jobs are rejected?
- (f) Derive  $E[N]$ , the mean number of jobs in the system.
- (g) Derive  $E[T]$  for the jobs that enter the system.
- (h) Suppose  $n = 5$  and  $\lambda/\mu = 0.4$ , is it better to double the buffer size or the processing speed?
- (i) Suppose  $n = 5$  and  $\lambda/\mu = 0.8$ , is it better to double the buffer size or the processing speed?
- (j) Explain why the answers to the parts above are either the same or different?

## 6 To balance or not to balance [15 points]

Consider the distributed server shown below in Figure 2. The arrival stream is a Poisson process with rate  $\lambda$ . Each job is sent to Host 1 with probability  $p$  and Host 2 with probability  $1 - p$ . The service rate at Host 1 is Exponential( $\mu_1$ ) and at Host 2 is Exponential( $\mu_2$ ).

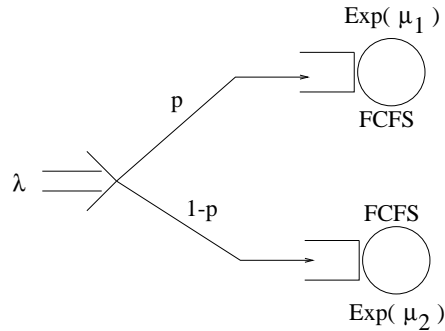


Figure 2: A simple 2-queue load balancing system.

- Assume  $\mu_1 = \mu_2$ . Prove or disprove that  $E[T]$  is minimized when  $p$  is chosen to balance the load (utilization).
- Assume  $\mu_1 \neq \mu_2$ . Prove or disprove that  $E[T]$  is minimized when  $p$  is chosen to balance the load.
- Assume  $\mu_1 \neq \mu_2$  and adjust the system so that it is closed, i.e., there is a fixed population of  $N$  jobs and upon completion each job immediately rejoins the system. Prove or disprove that  $E[T]$  is minimized when  $p$  is chosen to balance the load.

## 7 Open versus Closed [10 points]

Consider the closed and open systems in Figure 3 below. The relationship between closed and open systems is generally very poorly understood, and is something I'm very interested in understanding better.

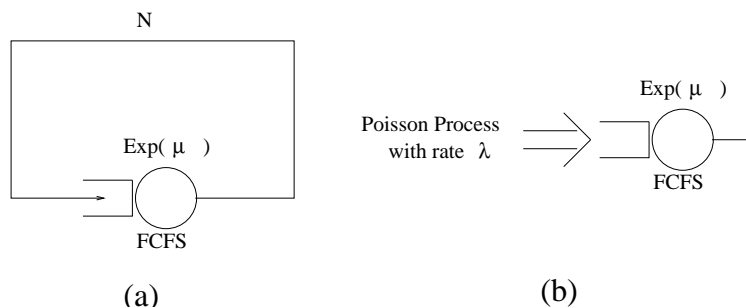


Figure 3: A single server FCFS queue that is closed (a) and open (b).

In this problem, I want you to figure out how  $E[T]$  relates between the two models. In particular, what is a criterion that guarantees that the closed system has larger  $E[T]$  than the open system. Express the criterion in terms of  $\rho = \lambda/\mu$  and  $N$  only.