CS 798 – Winter 2014
Assignment 1
Due: January 29, 2014

Problem 1

(a) The activities in a system are observed for a time period of length L (from 0 to L). Let

n = number of jobs arrived in $(0, L)$ = number of jobs processed in $(0, L)$

a_j = probability that an arriving job sees j jobs in the system

d_j = probability that a departing job leaves behind j jobs in the system

There are no group arrivals (i.e., jobs arrive one after another). Show that $a_j = d_j$ for all j.

(b) Suppose arrivals occur in groups of two (i.e., we always see two jobs arriving at the same time), and we assume that both arriving jobs see the same number of jobs in the system. Is the relationship $a_j = d_j$ for all j still true for this case? Explain your answer.

Problem 2

Suppose there are n arrivals in a time period of length L (from 0 to L), and the n^{th} arrival (or last arrival) occurs at time L. Let t_1 be the time between 0 and the time of the first arrival, and t_i be the interarrival time between the $(i\text{-}1)^{st}$ and i^{th} arrivals, $i = 2, 3, \ldots, n$.

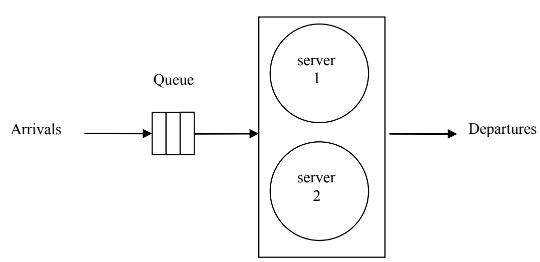
(a) Let X be the time until the next arrival. Plot X as a function of time t for $0 \le t \le L$.

(b) Let $E[Y] = \frac{1}{n}\sum_{i=1}^{n} t_i$ and $E[Y^2] = \frac{1}{n}\sum_{i=1}^{n} t_i^2$ be the mean and second moment of the interarrival time, respectively. Using the plot in part (a), obtain an expression for $E[X]$, the mean of X, as a function of $E[Y]$ and $E[Y^2]$.

Problem 3

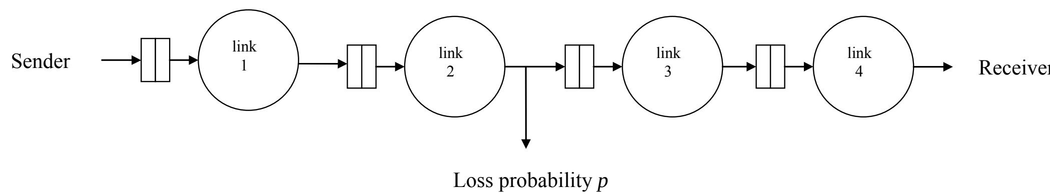
(a) Consider a single server queue model with two classes of jobs. Let n_j = number of class j jobs arrived in $(0, L)$ = number of class j jobs processed in $(0, L)$, $j = 1, 2$. Also let x_{ij} be the service time of the i^{th} class j job. Obtain analytic results for U_j', the utilization of the server by class j jobs, $j = 1, 2$.

(b) Consider a queueing model with a single queue and two parallel servers (see Figure below). Let n = number of jobs arrived in $(0, L)$ = number of jobs processed in $(0, L)$. Also let x_i be the service time of the i^{th} job. Is it possible to obtain analytic results for U_i, the utilization of server i, $i = 1, 2$? Explain your answer.



Problem 4

Consider the tandem queue model shown in the Figure below. Packets sent by the sender are transmitted along a path consisting of links 1, 2, 3 and 4, and then delivered to the receiver. Each link is modeled by a server.



Suppose link 3 is under heavy load and the loss probability at link 3 is p (the loss probabilities at the other three links are zero). Let

Y be the system throughput = the rate at which packets are delivered to the receiver

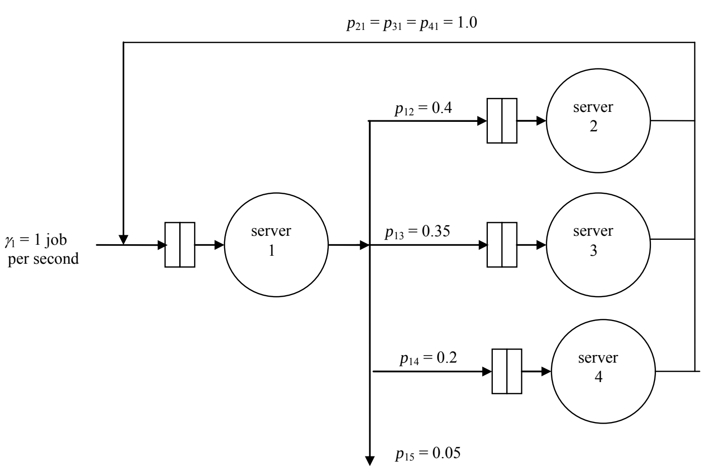
R be the mean end-to-end delay from sender to receiver

Q be the mean number of packets in the network

It has been suggested that according to Little's Law, we can write $YR = Q$. Do you agree? Explain your answer.

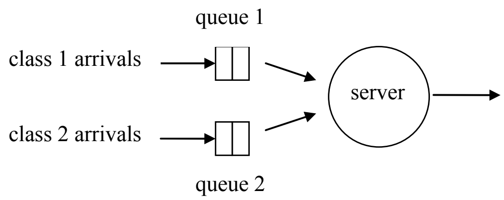
Problem 5

Consider the following open queueing network model. Suppose the mean response times at the four servers, denoted by R_1, R_2, R_3, and R_4, are 0.02, 0.03, 0.05, and 0.02 seconds, respectively. For this model, the end-to-end delay is the elapsed time from when a job arrives from outside the network to when this job departs from the network. Determine R, the mean end-to-end delay.



Problem 6

(a) Consider a queueing model with a single server, two classes of jobs, and cyclic service discipline. This is illustrated in the diagram below.



Under cyclic service, class 1 job arrivals join queue 1 and class 2 job arrivals join queue 2. Service is provided in cycles. A cycle starts with the server checking queue 1. If queue 1 is non-empty, the first job at queue 1 is processed. The server then checks queue 2 and processes the first job at queue 2 if queue 2 is non-empty. The cycle ends when processing of this job is complete. The cycle time is defined to be the sum of:

- Service time of first job at queue 1 (if queue 1 is non-empty), and zero otherwise
- Service time of first job at queue 2 (if queue 2 is non-empty), and zero otherwise
- Time spent in moving from queue 1 to queue 2 and back to queue 1

Suppose there are K cycles in a time period of length L (from 0 to L). Let

- m = number of class 1 jobs arrived in $(0, L)$ = number of class 1 jobs processed in $(0, L)$
- n = number of class 2 jobs arrived in $(0, L)$ = number of class 2 jobs processed in $(0, L)$
- s_i = service time of the i^{th} class 1 job, $i = 1, 2, \ldots, m$
- x_j = service time of the j^{th} class 2 job, $j = 1, 2, \ldots, n$
- H = time spent in moving from queue 1 to queue 2 and back to queue 1; H is a constant regardless of whether queue 1 is empty or not, or whether queue 2 is empty or not.

Also let

- S_1' = mean service time of class 1 jobs
- S_2' = mean service time of class 2 jobs
- C = mean cycle time
- λ_1' = arrival rate of class 1 jobs
- λ_2' = arrival rate of class 2 jobs

Derive an analytic expression for C. You must express your answer as a function of H, S_1', S_2', λ_1' and λ_2'only.

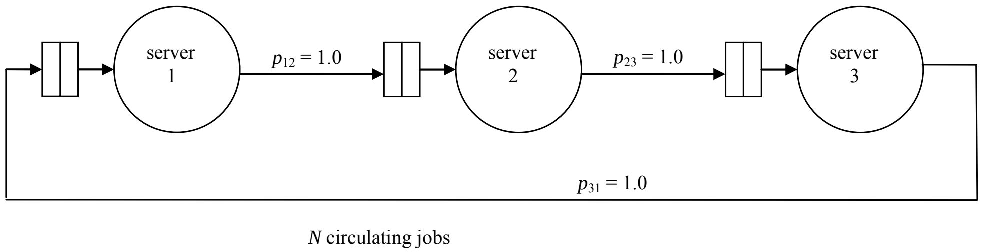
(b) Repeat part (a) for the case of a cyclic service discipline with exhaustive service. Under exhaustive service, each time queue j is checked during a cycle ($j = 1, 2$), jobs at queue j are processed until queue j is empty. For this case, the cycle time is the sum of:

- Total service time of jobs at queue 1 that have been processed (if queue 1 is non-empty), and zero otherwise
- Total service time of jobs at queue 2 that have been processed (if queue 2 is non-empty), and zero otherwise
- Time spent in moving from queue 1 to queue 2 and back to queue 1

Derive an analytic expression for C, the mean cycle time. You must express your answer as a function of H, S_1', S_2', λ_1' and λ_2'only.

Problem 7

Consider the following closed queueing network with N circulating jobs.



Let S_i be the mean service time at server i, $i = 1, 2$, and 3. Consider the case where $S_1 = 0.3$ second, $S_2 = 0.6$ second, and $S_3 = 0.4$ second.

(a) When $N = 1$, what are the values of λ_i, R_i, and Q_i, the arrival rate to server i, the mean response time at server i, and the mean number of jobs at server i, respectively, for $i = 1, 2$, and 3?

(b) Give the upper bound utilization for each of the three servers.

Problem 8

Consider the web application model described in the lecture notes entitled "Queueing Models – Analytic Results". The transition probabilities are $p_{10} = 0.2$, $p_{12} = 0.4$, $p_{01} = p_{21} = 1$; all others have value zero. The mean think time is 5 seconds, and the mean service times per visit to the web/application server and the database server are 0.05 and 0.1 seconds, respectively.

(a) Derive a lower bound for the mean response time R as a function of N, the number of user workstations.

(b) Suppose the web/application server is deployed on a machine that is twice as fast (i.e., the mean service time at the web/application server is 0.025 second instead of 0.05 second). The values of the other parameters remain unchanged. Derive a lower bound for the mean response time R as a function of N.

(c) Suppose the database server is deployed on a machine that is twice as fast (i.e., the mean service time at the database server is 0.05 second instead of 0.1 second). The values of the other parameters remain unchanged. Derive a lower bound for the mean response time R as a function of N.

(d) Explain the similarity (or difference) of the three lower bounds obtained in parts (a), (b), and (c).