EP2200 Queuing Theory and Teletraffic Systems

Monday, June 3rd, 2013.

Available teacher: Vitkoria Fodor

Allowed help: Calculator, Beta mathematical handbook or similar, provided sheets of queuing theory formulas, Laplace transforms and Erlang tables.

1. You decide to model a stochastic system with a three state continuous time Markov chain, with states S_1, S_2 and S_3. The system can move from S_1 to S_2 or to S_3, from S_2 to S_3 and from S_3 to S_1.

a) Does this system has stationary solution? Motivate your answer. (1p)

b) The transition rates are $q_{12} = 1$, $q_{13} = 1$ $q_{23} = 1$ $q_{31} = 1$. Give the Q matrix and the matrix equation of the stationary solution. Calculate the stationary state distribution. (3p)

c) Assume, the system is in S_1 at time zero. Give the distribution of time the system stays in S_1, before moving to another state. Prove or motivate your answer. Give the probability that the system does not move to another state in the first 2 time units. (3p)

d) The system is actually an ice cream machine. In S_1 it is in perfect state and produces 100 ice creams per time unit. In S_2 it is half clogged, but still produces 40 ice creams per time unit. Finally, in S_3 it needs to be stopped and cleaned. Considering the parameters in part b, what is the average time of cleaning? How many ice creams are produced per time unit in average? (3p)

2. You would like to move to the cloud computing business with low prices (and on small scale). For that you need to dimension your servers very well, such that you can provide fast response with the delay users still accept. For that you model the cloud computing service with a queuing model. You consider exponential service time, with an average of 12sec, and you assume that in average one request will arrive per second, according to a Poisson process. You decide to install $N = 15$ servers, so your cloud service can serve N requests at the same time. If all servers are busy, requests wait in a buffer that is not limited. Waiting requests are served in FIFO order.

a) Define the queuing model of the system, and draw its Markov chain. (2p)

b) What is the offered load in your system? What is the utilization of a server? Is the system stable? Motivate your answer. (2p)

c) Give the probability, that an arriving request can not be served immediately and the average time the users need to wait for answer (that is, waiting plus service time). (2p)

d) Consider a request that arrives when all the servers are busy and there are 2 requests waiting. Give the distribution of the waiting time of this arriving request in transform form and in time domain. (4p).

3. Consider the STEX office with only one secretary helping the students. Students arrive according to a Poisson process with one student per 10 minutes in average. The service time of the students is exponentially distributed with a mean of 2 minutes for one third of the students, and with a mean of 5 minutes for the others. Students arriving when the secretary is busy need to wait in a queue.

a) Give the Kendall notation of the system, and calculate the probability that the secretary is busy. (2p)

b) Calculate the average number of students waiting in the queue. (2p)

c) Consider a student that arrives when the secretary is busy but noone is waiting. Calculate the expected time this student spends in the STEX office (waiting time + service time). Calculate the probability that the student needs to wait for more than 5 minutes.(3p)

d) Assume now that students do not wait but leave the office directly, if upon arriving find the secretary busy. What is the probability that an arbitrary arriving student leaves without being served? (3p)

4. Consider resident permit applications received and processed by a case officer in the Swedish Migration Board. Applications arrive according to a Poisson process with 2 applications per working hour in average. Applications are processed with preemptive resume priority. One fourth of the applications have high priority, whereas the others have low priority. The processing time for high priority applications is exponentially distributed with a mean of 1/4 working hour, whereas for low priority applications it is constant 1/2 working hour.

a) Calculate the mean system time (queuing time + processing time) for both classes of applications, and the mean system time for an arbitrary application. (3p)

b) Calculate the average number of high priority applications that wait to be processed at the case officer. What is the probability that the case officer has more than two high priority applications waiting? (2p)

c) Consider a low priority application that arrives where there is one high priority and no low priority application at the officer. Calculate the probability that the processing of this application starts immediately after the processing of the current application. Calculate also the probability that the processing of this application will be interrupted. (3p)

d) To simplify calculations, you decide to model the system as an M/G/1 queuing system without considering priorities. Do you underestimate or overestimate the average system time of the applications this way? (2p)

5. Answer the following short questions.

a) Consider an M/M/1 queue. Prove that the departure process is Poisson and give the departure intensity. (2p)

b) Consider an office with two copy machines, shared by 5 employees. Each of the employees uses the copy machine once in a while and goes back to work in between. Assume that printing a document takes exponentially distributed time with a mean of 2 minutes, and between copying employees work for 20 minutes in average, again according to an exponential distribution. If both of the machines are occupied, employees go back to work immediately. What is the probability that both printers are busy and what is the probability that an employee has to turn back without printing? (2p)

c) Consider a queuing network of two M/M/1 queues. Requests arrive to the first queue. After service, a request goes to the second queue with probability 2/3 and goes back to the first queue otherwise. After service at the second queue, the request leaves the system with probability 2/3, and goes back to the first queue otherwise. The average service time is 1 time unit at both of the queues. What is the maximum arrival rate to the queuing network, such that it remains stable? Assume, that the arrival rate is 1/4. What is the average time that a request spends in the queuing network? (3p)

d) You need to add garbage collection in your computer system. You find two versions. Both of them start to run every time when there are no jobs in the system, and arriving jobs need to wait until the garbage collection process completes. The first version needs exactly 1.5 ms to complete. The second version runs for exponentially distributed time with an average of 1 ms. Which version should you choose to minimize the average waiting time of the jobs? (3p)