FALL 2006
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN-MADISON
PH. D. QUALIFYING EXAMINATION
Modeling and Analysis
Monday, September 18, 2006
3:00-7:00 PM

GENERAL INSTRUCTIONS:

1. Answer each question in a separate book.

2. Indicate on the cover of *each* book the area of the exam, your code number, and the question answered in that book. On one of your books list the numbers of all the questions answered. *Do not write your name on any answer book.*

3. Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

Answer ALL four questions.

POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area to resolve problems during the *first hour* of the exam. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

MA Ph.D. Qualifying Exam

Fall 2006

Directions:

Use careful reasoning to develop the answers to each of the following questions. The correctness of your derivations and the soundness of your explanations count more than the completeness of your answers. Be sure to state any assumptions you make in your solutions. Note that it's fine not to simplify numerical answers.

1. Provide short answers to each of the following.

   a. By the BCMP result, the solution for the mean residence time in an open single-class single-server queue is as follows:

   $$R = S\,(1+Q)$$

   where Q is the mean queue length (including the customer in service) and S is the mean service time. List at least three scheduling disciplines that this result holds for, and the restrictions on service time distribution, if any, for each scheduling discipline.

   b. The P-K formula for mean residence time in the single-class single-serve M/G/1 queue is as follows:

   $$R = S + \frac{US(1+C_S^2)}{2(1-U)} = S(1+Q-U) + U\frac{S}{2}(1+C_S^2)$$

   where U is the server utilization, C_S is the coefficient of variation in service time, and the other symbols are as defined in part a. Explain which scheduling disciplines this result holds for.

   c. In the case of the M/G/c queue, the following approximation for mean residence time has been shown to be reasonably accurate:

   $$R = S + \frac{U^{\sqrt{2(c+1)}}(1+C_S^2)}{2\lambda(1-U)}$$

   where $U = \lambda S/c$ and the remaining symbols are as defined in parts a and b.

   Suppose that $c = 2$, $U = 0.5$, $S = 4$, and $C_S^2 = 3$. What is the value of R?

   For the same workload, assume the two servers are replaced by one server that is twice as fast. Explain the value of each of the parameters in the mean residence time equation and compute the new value for R.

2. Consider a FCFS queue with Poisson arrivals at rate λ and exponential service times with mean S. Assume that at each arrival event, exactly two customers arrive.

   a. What relationship between λ and S is required for the mean residence time in the queue to be finite?

   b. Assume the relationship in a holds and derive the mean residence time in terms of λ and S.

3. Consider a new transport protocol on host A that transmits a file to host B in 1500-byte packets at average rate λ packets per second, with approximately fixed packet spacing. For each packet received, host B sends an acknowledgement packet back to host A.

   Assume the packets only wait for service at the bottleneck link in the path from A to B. That is, the waiting occurs at the link that has the least available bandwidth for the flow, and all other points in the path have negligible packet waiting time.

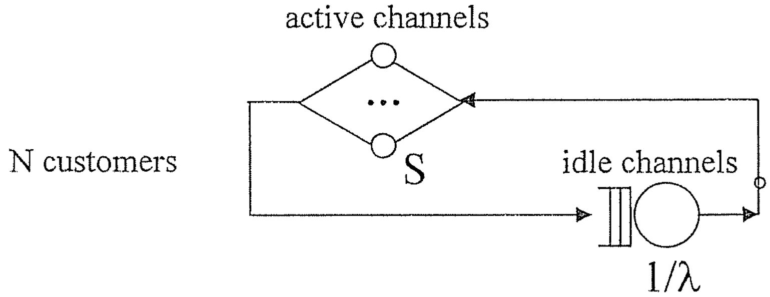
   While the protocol is transmitting the data, the protocol also uses the acknowledgment packets to measure: (1) the average round trip time, RTT, and (2) the minimum round trip time, RTT_{min}. The protocol also sends a few closely spaced packet pairs at the beginning of the flow and uses the spacing between the acknowledgments for each packet pair to measure (3) the capacity or transmission rate on the bottleneck link, C (in bits/second).

   Assume that the measures of RTT, RTT_{min}, and C are all accurate and that the value of these parameters does not change during the time of the flow. Develop formulas for the protocol to compute each of the following from these three measures, and explain why each formula holds.

   a. The average number of packets from the flow that are in the bottleneck link buffer.

   b. At the instants that the packets from the flow arrive to the bottleneck link buffer, the average total amount of data queued in the buffer, in units of 1500-byte packets.

4. For a video-on-demand server with clients who will not wait for service, the Sigmetrics 2002 paper by Tan et al. uses a two-center queueing network (shown in the following figure) to compute the client balking rate (or fraction of client requests that are lost because they arrive when all of the server channels are busy serving other clients) as a function of total server bandwidth.



   In this model, the number of customers in the network, N, is equal to the total amount of server bandwidth measured in units of the streaming rate. (Each unit of bandwidth equal to the streaming rate is called a "channel". That is, if N=100, then the server has bandwidth to serve 100 streams simultaneously.

   Customers at the FCFS queue represent idle channels waiting to serve new client requests that arrive at rate λ. Assume that the time between client requests (modeled as a service time at the FCFS queue) is exponentially distributed with mean $1/\lambda$.

The customers at the delay center represent channels that are streaming data to clients. The average duration of a stream is denoted by S.

a. The outputs of the customized MVA solution – for the case that the video server uses a scalable multicast streaming protocol – include: (1) the system throughput, X, (2) the mean residence time in the FCFS queue, R, (3) the mean number of customers in the FCFS queue (including the customer in service), Q, and (4) the utilization of the server in the FCFS queue, U. Explain how to compute the client balking rate from these model outputs.

b. If the server uses unicast streaming and each client views an entire requested video of duration S, is a customized MVA solution required, or can we use a standard MVA solution to obtain X, R, Q, and U? Explain briefly.

c. Next consider one of the output ports at an Internet router. Assume the output port has a maximum buffer size of B packets.

   Suppose you are asked to use a similar idea to the Tan et al. model – i.e., a two-center queueing network with B customers – to predict the fraction of packets that are dropped because they arrive to the link when the buffer is full. The customers at one of the nodes represent empty slots in the buffer waiting for packets to arrive. At the other node, packets in the buffer are transmitted on the outgoing link.

   Assume an average packet arrival rate of λ, and a coefficient of variation in packet inter-arrival times of C_a. Also assume a link speed equal to r bits/second, and a fixed packet size of X bits.

   Draw the queueing network and develop the customized MVA equations that compute the packet loss rate for the buffer.