

Midterm for CpE/EE 345 – Modeling and Simulation Stevens Institute of Technology Spring 2004	Name:
Pledge:	

The midterm is open book/open notes. Total value is 100 points (30% of course grade). All questions are equally weighted. Do any 10 of the 12 question. Do more than 10 for extra credit. Some question can be answered in more than one way. Only one answer is required, but extra credit will be given for identifying and explaining alternate answers. Some questions ask for N answers. Extra credit will be given for more than N answers.

- (1) In the late 1970's LANs didn't exist. It was not uncommon for a large company to use a voice PBX with modem pools for in-building data communications, much like ISPs provide dial-up services today. The BLT Sandwich Company had a bank of 250 300 bps Bell 103 modems to serve their employee's data networking needs and was planning to upgrade to Bell 212 300/1200 bps dual speed modems, since most employees were getting the faster 212 modems. Proposal A was to leave the old modems in one PBX hunt group (a set of phone lines that all share the same phone number) and add a second hunt group for the new modems. The 250 computer data connections would be moved from the old modems to the new ones as new modems were installed. Users who had 300 bps modems would call the low-speed hunt group, while users with 1200 bps modems would call the high-speed hunt group. Proposal B was to combine the old and new modems into one hunt group – users would call a single PBX exchange and would be connected to the first available modem, whether it was 300 or 300/1200 bps. Discuss the advantages and disadvantages of Proposal A versus Proposal B with respect to issues we have covered so far in this course. Note: ISPs faced this same issue when they started upgrading their dial-in modem pools to 56 kb/s modems.

[7 pts] This is the classic server-splitting problem. Generally, creating two separate server groups will increase blocking – the probability that all the resources in the server group are used, while the other server group has idle devices.

If you consider hunt groups to be M/M/c queues, Proposal A calls for two individual queues, M/M/n and M/M/(250-n) where n is the number of modems in one of the pools (it doesn't matter which way we assign n – as the number of 300 bps modems or the number of 1200 bps modems). Proposal B has a single M/M/250 queue. With fewer servers in Proposal A's separate server groups, users would encounter a higher blocking probability, forcing them to dial the other server group.

[3 pts] On the other hand, with a single server group, while the blocking probability would be lower, allowing users a greater chance to complete their call on a single attempt, 1200 bps users might get connected to a 300 bps modem, slowing down their access. Having separate server groups would ensure the 300 bps users would only use 300 bps modems, leaving the 1200 bps modems for 1200 bps users. Assuming that the 300 bps users didn't try to sneak into the 1200 bps server group...

[5 pts EC]: This problem could be analyzed if the holding times and calling rates of the users were known, as well as the mix of user modems and the number of upgraded modems in the modem pool(s). We could assume users have a fixed amount of traffic to send and would encounter $1/4^{\text{th}}$ the holding time with a higher speed connection.

p.s. – this describes a real problem that existed at Bell Telephone Labs in the late 1970s.

- (2) The BLT Sandwich Company's original modem pool from problem (1), has 250 modems. If the length of a data call was exponentially distributed with a mean length of 20 minutes, and each user made an average of 4 data calls in an 8 hour shift to read their email, is it likely that the number of data ports was sufficient if BLT had 2000 employees?

For an M/G/ ∞ queue, with $\lambda = (4/8)*2000$ per hour. $\mu = 3/\text{hour}$, so $L = \lambda/\mu = 333$. This means that the expected number of users, if there were an infinite number of phone lines is 333, which is greater than the 250 lines provided. The number of data ports is not sufficient.

- (3) The men on the Snevets University co-ed Bocce Ball team have heights that are uniformly distributed between 65 and 77 inches, while women on the team have heights that are uniformly distributed between 62 and 70 inches. If 30% of the members of the team are women, find the mean and variance of the distribution of heights of members of the team.

There are two ways to approach this problem. The simplest is to consider that there are two separate groups of team members. We can calculate their average heights separately and then combine the results. In this manner,

$E(\text{team height}) = E(\text{men's heights}) \cdot \text{fraction_men} + E(\text{women's heights}) \cdot \text{fraction_women} = 71 \cdot .7 + 66 \cdot .3 = 69.5$. Unfortunately, calculating the variance is not as straightforward.

The other approach, which does allow calculation of the variance, is as follows:

First, find the probability density function:

$$w_l := 62$$

$$w_u := 70$$

$$p_w := .3$$

$$m_l := 65$$

$$p_m := 1 - p_w$$

$$m_u := 77$$

$$\text{norm1} := \int_{w_l}^{w_u} p_w dx + \int_{m_l}^{m_u} p_m dx$$

$$\text{norm1} = 10.8$$

We have to include the interval size, since a wider interval reduces the maximum value of the pdf.

$$f(x) := \text{if} \left[x < w_l, 0, \text{if} \left[x > w_u, 0, \frac{p_w (w_u - w_l)^{-1}}{\text{norm1}} \right] \right] + \text{if} \left[x < m_l, 0, \text{if} \left[x > m_u, 0, \frac{p_m (m_u - m_l)^{-1}}{\text{norm1}} \right] \right]$$

We have a way of weighting the individual contributions, but this is not enough to ensure $f(x)$ is a pdf, since it must integrate to 1.

$$\text{norm2} := \int_{\min(w_l, m_l)}^{\max(w_u, m_u)} f(x) dx$$

$$f2(x) := \frac{f(x)}{\text{norm2}}$$

Now, $f_2(x)$ is a pdf, so we can calculate the mean and variance in the standard way. We'll only integrate over the interval where $f_2(x)$ is nonzero.

$$m := \int_{\min(w_l, m_l)}^{\max(w_u, m_u)} x \cdot f_2(x) \, dx$$

$m = 69.5$

So, we get the same value for the mean as the other method and we can use this to calculate the variance:

$$v := \int_{\min(w_l, m_l)}^{\max(w_u, m_u)} (x - m)^2 \cdot f_2(x) \, dx$$

$v = 15.25$

- (4) Students arrive independently at Major Disaster's Snack Bar with exponentially distributed interarrival times, with a mean arrival rate of 15 students per hour. What is the probability that 30 students will arrive between 1 pm and 3 pm? If no students have arrived by 1:30pm, what is the probability that 10 will arrive between 2 pm and 3 pm?

[3 pts] The interarrival process is exponential, so the number of students arriving in a given interval is Poisson.

$$P[N(t) = n] = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad \text{where } n=30, \lambda=15, t=2.$$

[2 pts] Here, the probability is 0.073.

[3 pts] The fact that no students arrived by 1:30 has no influence on the number of arrivals between 2 and 3.

[2 pts] Here we can apply the same formula with $n=10, \lambda=15, t=1$. Here, the probability is 0.049.

- (5) A single queue system has exponentially distributed interarrival times with a mean arrival rate of 3 customers per minute. The service time is uniformly distributed with a mean of 2 per minute and a variance of 1 per minute². If the queue has the following parameters at $t=t_0$: $L_Q(t_0)=10$, next customer arrival at $(t_0+8$ seconds), next service completion at $(t_0+30$ seconds). What is the system state at t_0 ?

The system state is described by the number of customers in the queue and the busy/idle state of the server. Here, there are 10 customers in the queue. Since there are customers waiting, the server must be busy, so the system state is (busy, 10). The other information about the queue is irrelevant.

- (6) Able and Baker have decided to split their carhop business, since they are rarely in the building at the same time. Their lawyer wants to meet with them at times when both are present simultaneously, but Able is there 50% of the time, while Baker is there 40% of the time, independently. What is the probability that both Able and Baker will both be present three times out of the five times the lawyer comes to visit?

First, since Able and Baker arrive independently, the joint probability that both will be present at any given time is the product of the probabilities that each is there or $.5 \cdot .4 = .2$. The probability that both will be present 3 out of 5 times is a Bernoulli trial with a sample size of 5 and $p = .2$, so

$$P(X \geq 3) = \sum_{x=3}^5 \binom{5}{x} (.2)^x (.8)^{5-x}$$

where

$$\binom{5}{x} = \frac{5!}{x!(5-x)!}$$

$$P(X \geq 3) = \sum_{x=3}^5 \binom{5}{x} (.2)^x (.8)^{5-x} =$$

$$\left(\frac{5!}{3!(5-3)!} (.2)^3 (.8)^{5-3} \right) + \left(\frac{5!}{4!(5-4)!} (.2)^4 (.8)^{5-4} \right) + \left(\frac{5!}{5!(5-5)!} (.2)^5 (.8)^{5-5} \right) =$$

$$\left(\frac{5 \cdot 4}{2} (.2)^3 (.8)^2 \right) + \left(\frac{5}{1} (.2)^4 (.8)^1 \right) + (.2)^5 =$$

$$\left(\frac{5 \cdot 4}{2} (.2)^3 (.8)^2 \right) + \left(\frac{5}{1} (.2)^4 (.8)^1 \right) + (.2)^5 =$$

$$.0512 + .0064 + .00032 =$$

$$.0579$$

- (7) A Beowulf Linux cluster is made up of thirty-two 2.5 GHz PCs, each capable of processing jobs with exponentially distributed service times at an average rate of $\mu = .2$ jobs/second. Jobs arrive independently at the cluster with exponential interarrival time distributions at the rate $\lambda = 10$ jobs/second and are assigned randomly to the first available processor. What observations can you make about the average waiting time for a job? How might you change the system configuration to get the best performance for the least cost?

[3 pts] The overall server utilization, $\rho = \lambda / (c\mu) = 10 / (.2 \cdot 32) > 1$.

[4 pts] The average waiting time grows without bound.

[3 pts] More processors are needed to reduce the average utilization to less than 1, so that the jobs can all be completed. At least 50 processors would be needed to reach this level of service rate.

- (8) Packets arrive randomly with exponentially distributed interarrival rates at a router in the Hard Knocks University computer center. On average, 10 packets/minute arrive at the router. Packets are forwarded to the Internet via one of three routes – 30% of the packets are printed on slips of paper and sent via trained pigeons that congregate in the park at the south end of the campus, 60% are printed on punch cards and sent via pneumatic tube, and 10% are sent via a 2.4 kb/s dial-up connection. What is the probability that 20 pigeons will be dispatched in the next 10 minutes?

[4 pts] First, recognize that this is the Poisson splitting property applies here. The pigeons must process $.3 \cdot 10 = 3$ packets/minute.

[3 pts] The probability of $n=20$ events in time $t=10$ minutes, with $\lambda=3$ /minute is

$$P[N(t) = n] = \frac{e^{-\lambda t} (\lambda t)^n}{n!}$$

[3 pts] The probability is 0.013

- (9) You observe a system simulation that goes through the following states:

Time	System State
<0	(0,0)
0	(1,0)
5	(1,1)
6	(1,2)
8	(1,1)
10	(1,0)
11	(0,0)
15	(1,0)
18	(1,1)
20	(1,0)
25	(0,0)

Describe the meaning of the system state and the events that are happening.

[1 pt] The first parameter in the system state is the server status. 1 means busy, 0 means idle.

[1 pt] The second parameter is the number of customers in the system.

[1 pt for each event correctly identified] The events are:

Arrivals at $t=0, 5, 6, 15,$ and 18

Departures at $t=8, 10, 11, 20$ and 25 .

- (10) Create the FEL for the system described in problem (9) at each point in time shown above. Assume that the system simulation will end at $t=26$

Time	System State	FEL
<0	(0,0)	(A,0),(E,26)
0	(1,0)	(A,5),(D,8),(E,26)
5	(1,1)	(A,6),(D,8),(E,26)
6	(1,2)	(D,8),(A,15),(E,26)
8	(1,1)	(D,10),(A,15),(E,26)
10	(1,0)	(D,11),(A,15),(E,26)
11	(0,0)	(A,15),(E,26)
15	(1,0)	(A,18),(D,20),(E,26)
18	(1,1)	(D,20),(E,26)
20	(1,0)	(D,25),(E,26)
25	(0,0)	(E,26)

- (11) System A has a single queue with a service time that is exponentially distributed, having a mean service rate of .5 jobs/minute. System B has two tandem queues through which all customers must pass, both queues have exponentially distributed service times with a mean service time of 1 minute. In which system would you expect to encounter a longer waiting time?

[10 pts] Both systems have the same average service time: 2 minutes. However, while the variance of the first system is 4 minute^2 , the second system has a variance that is the sum of the variances of the two service time variances: $1 \text{ minute}^2 + 1 \text{ minute}^2 = 2 \text{ minute}^2$. Thus, the coefficient of variation is higher for System A, creating a longer average queue length and, by Little's formula, a longer average waiting time.

- (12) You have been asked to write a simulation of various disk access algorithms for the disk controller for a computer system. You are considering several single server data block queuing disciplines to decide which is optimum for writing data blocks to a hard disk with multiple platters, a fixed rotational speed, and a known track to track radial seek time. The queuing disciplines you are considering are: First-in-first-out (FIFO), Last-in-first-out (LIFO), Service-in-random-order (SIRO), Priority, and Shortest-Processing-Time (SPT). Pick at least two of these queuing disciplines and describe how you would expect that queuing discipline to perform. Include information on you might decide how to decide which data block will be written next.

[5 pts for each]:

FIFO: This would be the simplest queuing discipline to manage and the simplest to analyze, since it is the typical queuing discipline. However, this might not be the optimum discipline to use. For instance, the first block in the queue might need to be written to a position on the disk that will require the longest time to get to – the disk needs to rotate to the correct position and the write head needs to seek to the right track. The next block to be written would always be the first in the queue, even if the second block in the queue was to be written to the current write head position.

LIFO: Generally, this would not be an appropriate queuing discipline for a disk controller. It offers no advantage over the FIFO, since it is insensitive to the data and the current position of the disk and the write head. Since we would expect the disk write queuing delay to be rather short, on average, and since data blocks would not be removed from the queue if their processing time were too long, there is no incentive to use LIFO. To implement LIFO, data blocks could be pushed onto a stack, much like the stack of dishes in a restaurant. The next block to be processed would always be the one on the top of the stack.

SIRO: This queuing discipline would generally not be a desirable queuing discipline for a disk controller. Like FIFO and LIFO, it does not examine the state of the disk (rotation and head position) in relationship to where the data needs to be written. Like LIFO, data could stagnate while waiting to be written. Blocks could be written into random access memory and could be retrieved in any order.

Priority: This queuing discipline might be one worth seriously considering for a disk controller. If there were some means to assign a priority to data blocks, the most important would be written first. While this would not reduce the average waiting time before a block was written to disk, at least it would ensure that the most important blocks were written. Priority could be assigned based on block length, with short, blocks that could be written in the shortest number of operations getting the highest priority. Data would probably be stored in random access memory with a table of priorities and pointers to the data stored

in a list.

SPT: This queuing discipline would very likely be the best performing disk scheduling algorithm, if processing time included a measure of the length of the data block as well as how far the disk rotation and write head were away from the place where the data was to be written. By always writing the data that required the least head movement and ordering the data writes to allow multiple blocks of data to be written as the disk continued to rotate, performance could be maximized. Data blocks would probably be stored in random access memory with markers that indicated where the data was destined for on the disk. The disk controller would need to keep track of the status of the disk rotation and write head and would need to quickly find the next block to be written, based on a cost measure.