UNIVERSITY OF TORONTO

FACULTY OF APPLIED SCIENCE AND ENGINEERING

FINAL EXAMINATION, DECEMBER, 1998

Fourth Year -- Computer Engineering Program

CSC 470F - COMPUTER SYSTEM MODELLING AND ANALYSIS

Examiner: K. C. Sevcik

NAME _____ STUDENT #: _____

Open books and notes. Use of calculators is permitted.

Write all answers in the space provided on the examination paper.

Show your work where appropriate.

Attempt all parts of the test before spending too much time on any single part.

PART	MARK
I.	/10
II.	/25
III.	/18
IV.	/13
V.	/17
VI.	/17
Total	**/100**

PART I. [10 marks]

A. [6 marks]

A system supporting a *single class workload* was measured and found to have the following characteristics:

$N = 6$	(number of users)
$Z = 10$	(think time)
$T = 1000$ *secs.*	(measurement interval)
$C_0 = 200$	(number of completions)
$U_{cpu} = .80$	(cpu utilization)
$U_{disk\,1} = .40$	(disk1 utilization)
$U_{disk\,2} = .60$	(disk2 utilization)

1. What is the *system throughput*, $X(6)$?

2. What is the *service demand* of the *disk* 2, $D_{disk\,2}$?

3. What is the *average response time*, $R(6)$?

4. If the number of customers is increased from 6 to about 100, approximately what will the *utilization of disk* 1, $U_{disk\,1}(100)$, be?

5. If the number of customers is increased to 100, approximately what will *throughput*, $X(100)$, be?

6. If the number of customers is increased to 100, approximately what will *response time*, $R(100)$, be?

B. [4 marks]
Use *Fundamental Laws* to answer the following questions:

1. In a system with a single disk, each transaction does 10 disk I/O's on average and the average service time of the disk is 20 msecs. If the disk utilization is 60%, what is the system *throughput*?

2. A system in which users have an average think time of 15 seconds supports a throughput of 3 transactions per second with an average response time of 5 seconds. How many active online customers are there?

3. Consider a system that supports both a *batch* class with average multiprogramming level 3, and an *interactive* class with 15 active customers and an average think time of 20 seconds. The cpu service demands for the two classes are 1.0 for the interactive class and 10.0 for the batch class, and the total cpu utilization is 90%. If the average response time for interactive users is 10 seconds, then:

   What is the *throughput* of the interactive class?

   What is the *throughput* of the batch class?

PART II. [25 marks]

A. [5 marks]

The *Exact Mean Value Analysis* algorithm for analyzing a **single-class, closed** queueing network is based on five key equations.

1. State the two Exact MVA equations that are based on forms of Little's Law.

2. State the Exact MVA equation that expresses the expected length of a queue at center k seen by an arriving customer to that center.

3. State the Exact MVA equation that reflects the assumption that a customer can hold only one resource at a time.

4. What is assumed in one of the MVA equations relating to the amount of remaining service required by the customer in service at the time another customer arrives?

B. [7 marks]

Some ways in which queueing networks differ are:

- whether or not *homogeneity assumptions* are satisfied (leading to a *product form* solution)
- number of centres (few or many)
- number of classes (few or many)
- total number of customers (few or many)
- required accuracy of model solution (perfect or estimated)

Among the set of solution algorithms,

Global Balance (GB)	ABA (ABA)	BSB (BSB)	Convo-lution (CNV)	Exact MVA (eMVA)	Approx. MVA (aMVA)	Linear-izer (LIN)

(i). Which of these algorithms are applicable only when the homogeneity assumptions are satisfied?

(ii). Which ones become impractical whenever the total number of customers is large?

(iii). Which ones become impractical whenever there are many classes?

(iv). Which ones provide an *exact* solution to the model?

(v). Which one is susceptible to numerical errors (i.e., overflow or underflow)?

(vi). Which ones can be conveniently extended to account for certain characteristics that make models *non–separable* (or non-*product form*)?

(vii). Which ones provide only bounds rather than point estimates of performance measures?

C. [6 marks]

Various exact and approximate algorithms for solving queueing network models use different expressions for the *expected number of customers found in queue* upon arrival at a center. In the table below, place a single check mark in each column to indicate the expression used in the *single class* case by the algorithm associated with that column.

$A_k(N) =$	Ex MVA (open systems)	Ex MVA (closed systems)	Bard's LCP aMVA	Bard/ Schweit. aMVA	Queue Line aMVA	Linear-izer
$Q_k(N)$						
$Q_k(N-1)$						
$(N-1) \times Q_k(N)$						
$\frac{N-1}{N}\ Q_k(N)$						
$\frac{N-1}{N}\ Q_k(N) - (N-1)\delta$						
$\frac{N-2}{N-1}\ Q_k(N) + \frac{1}{N-1}(\frac{D_k}{D+Z})$						

D. [7 marks]

Given the parameters of a single class, closed queueing network:

$$N,\ K,\ D,\ \ (D_1, D_2, \cdots, D_K)$$

(where D is the sum of the K service demands), any of the solution algorithms listed below can be applied. For each one, enter each of N, K, and D in one column of each row to indicate whether the *computational cost of the algorithm* is independent of, linear in, or exponential in the value of that parameter.

	Independent of ...	Linear in ...	Exponential in ...
ABA			
BSB			
Convolution			
Exact MVA			
Approx MVA			
Linearizer			
Global Balance			

PART III. [18 marks]

A. [6 marks]

1. Give two reasons why it is more difficult assure that a *stochastic simulation* program is working "correctly" than it is to assure that, for example, a matrix multiplication program or a sorting program is working "correctly"?

2. Briefly indicate a few tests that can be applied to a *stochastic simulation* program to increase confidence that it is working as intended.

3. In analyzing a queueing network model by simulation (e.g., using Csim), rather than by analysis (e.g., using MAP), what additional information about the workload can be included to increase the accuracy of the results?

4. Why are *confidence intervals* needed to evaluate the validity of results for simulation models, but not for analytic models?

5. Why is a "warm-up" period typically needed in stochastic, dynamic simulations?

6. How can the impact of "end effects" be limited at the end of a simulation?

B. [6 marks]

Consider a situation where simulation is used to estimate the average response time of a particular queueing system. One hundred *independent replications* were carried out, and the means of the response times in each of the one hundred trials are given by:

$$\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots, \vec{x}_{99}, \vec{x}_{100}$$

Then we can calculate the sample mean and sample variance as follows:

$$\text{Sample Mean:} \quad \vec{\vec{x}} = \frac{1}{100} \sum_{j=1}^{100} \vec{x}_j$$

$$\text{Sample Variance:} \quad Var(\vec{x}) = \frac{1}{99} \sum_{j=1}^{100} (\vec{x}_j - \vec{\vec{x}})^2$$

1. If the calculation reveals that:

$$\vec{\vec{x}} = 15.0 \quad \text{and} \quad Var(\vec{x}) = 900.0,$$

then what is the *90% confidence interval* on the *average response time* of the system? **[Show your calculations.]**

Reminder:

$$Prob\left[\vec{\vec{x}} - \sqrt{\frac{Var(\vec{x})}{n}}\, T_\alpha \le \mu \le \vec{\vec{x}} + \sqrt{\frac{Var(\vec{x})}{n}}\, T_\alpha\right] = 1 - \alpha$$

T_α	1.00	1.28	1.65	1.96	2.58
$1 - \alpha$	.68	.80	.90	.95	.99

2. Based on these experimental observations which indicate that $\vec{\vec{x}} = 15.0$ and $Var(\vec{x}) = 900.0$, how likely is it that the *true average response time* is greater than 20.0? **[Again, show your calculations.]**

C. [6 marks]

A simulation experiment is composed of N_{trials} *trials*. Each *trial* consists of a $N_{batches}$ *batches* of $N_{observations}$ *observations* each. Below are six experimental designs, each consisting of 50,000 observations overall.

Design	N_{trials}	$N_{batches}$	$N_{observations}$	*Total Observations*
(a).	1	1	50,000	50,000
(b).	1	50	1,000	50,000
(c).	50	1	1,000	50,000
(d).	1	500	100	50,000
(e).	5	10	1,000	50,000
(f).	50	50	20	50,000

Assume that you would like to do *confidence interval* analysis using an assumption of Normal distributions, and answer the following questions.

(i). Which designs are suitable for use with the *method of independent replications*?

(ii). Which designs are suitable for use with the *method of batch means*?

(iii). Which design or designs will most likely yield the most accurate estimates of the true confidence interval on the results?

PART IV. [13 marks]

A. [5 marks]

Assume that you have a large closed queueing network with many centers and several classes of customers with a large number of customers in each class. If the model is *separable*, then the exact solution to the model can be found by *Exact MVA*, but at high computational cost. Assume that you need to solve this model for a large number of cases that differ *only in the service demands of the classes at one specific center*. Briefly describe how *hierarchical modelling* with a *flow-equivalent service center* can be used to obtain the exact answers to all the cases with a vast reduction in the total amount of computation required.

B. [8 marks]

Consider the *memory-constrained* queueing network model shown below with memory limit M, average think time Z, and service demands $D_c = V_c S_c$ and $D_d = V_d S_d$. Let the *system state* be described by a quadruple, (n_t, n_m, n_c, n_d), where n_x is the number of customers at the terminals, the memory queue, the cpu, and the disk, respectively. For $N = 4$ customers and memory limit $M = 2$, consider finding an exact solution to this non-separable model by solving the set of *global balance equations*.

1. How many *states* are there in the *global state transition diagram*?

2. There are *seven* different non-zero *transition rates* between pairs of states in the *global state transition diagram*. What are they (expressed in terms of system parameters Z, S_c, S_d, and $p = 1 / V_c$)?

3. Write out the full *global state balance equation* for each of the following four states:

   (4,0,0,0):

   (3,0,0,1):

   (2,0,2,0):

   (1,1,1,1):

PART V. [17 marks]

A. [12 marks]

In the three parts of this question, you are to write a fragment of pseudo-code (between one and eight lines long in each case) that generates pseudo-random variates with the specified distribution based on the general method indicated, using a source of Uniform (0, 1) pseudo-random numbers. In your pseudo-code, each reference to U_j will represent a new sample from the Uniform (0, 1) source, and you should leave the variable R_i holding the value of the pseudo-random number.

1. Use the *method of inversion* to determine a way to produce pseudo-random variates with the distribution:
   Probability Density Function:

$$f(x) = \frac{x - 2}{8} \qquad 2 \le x \le 6$$

   Cumulative Distribution Function:

$$F(x) = \frac{(x - 2)^2}{16} \qquad 2 \le x \le 6$$

2. Use the *method of decomposition* to determine a way to produce pseudo-random variates with the distribution:
   Probability Density Function:

$$f(x) = \begin{cases} 1/2 & 2 \le x \le 3 \\ 1/3 & 3 \le x \le 5 \\ 1/6 & 5 \le x \le 6 \\ 0 & \textit{otherwise} \end{cases}$$

3. Use the *method of rejection* to determine a way to produce pseudo-random variates with the distribution:
   Probability Density Function:

$$f(x) = \frac{1}{2\pi}\sqrt{4-(x-4)^2} \quad 2 \le x \le 6$$

B. [5 marks]

Five desirable properties of *Uniform (0,1)* pseudo-random number generators are listed below. In a few words, explain the meaning of each one.

uniformity:

independence:

density:

efficiency:

maximum cycle length:

PART VI. [17 marks]

A. [5 marks]

The equations below are useful in the *approximate mean value analysis* algorithm to approximately account for service time distributions that are non-exponential when the service discipline is *FCFS*. Examine the equations in detail and answer the questions below.

$$R_{c,k}(\vec{N}) = V_{c,k}\left[S_{c,k} + \sum_{j=1}^{C}\left[S_{j,k}[Q_{j,k}(N-\vec{1}_c) - U_{j,k}(N-\vec{1}_c)] + r_{j,k}\,[U_{j,k}(N-\vec{1}_c)]\right]\right]$$

$$= V_{c,k}\left[S_{c,k} + \sum_{j=1}^{C} S_{j,k}\left[Q_{j,k}(N-\vec{1}_c) + \left(\frac{CV_{jk}^2 - 1)}{2}\right) U_{j,k}(N-\vec{1}_c)\right]\right]$$

where $r_{j,k}$ is the expected remaining service time of a class j customer found in service at center k by an arriving class c customer, and CV_{jk}^2 is the squared coefficient of variation of the service times of the class j customers at center k.

1. What part of the equation represents the *service demand* of the arriving customer?

2. What part of each equation represents the likelihood of finding a class j customer in service when a class c customer arrives at center k?

3. Under what conditions will the factor, $[Q_{j,k}(N-\vec{1}_c) - U_{j,k}(N-\vec{1}_c)]$, be negative?

4. For what types of service time distributions will the factor, $\left(\frac{CV_{jk}^2 - 1}{2}\right)$, be equal to zero?

5. For what types of service time distributions will the factor, $\left(\frac{CV_{jk}^2 - 1}{2}\right)$, be negative?

B. [6 marks]

Several disks are connected through a *single* controller to a processor. In this case, a general form of the equation to determine the *service demand* of disk k is:

$$D_k = V_k \left[Seek_k + Latency_k + Contention_k + Transfer_k \right]$$

1. If the basic disk characteristics, and the locality of access properties and block sizes of the workload are known, which term in the sum in the equation is *most difficult* to estimate?

2. What is represented by the term identified in part (1)?

3. In a few words or phrases, explain why *iteration* is required to estimate the unknown term and hence the disk service demand.

C. [6 marks]

Based on measurement data, a single class system consisting of a cpu and a disk with terminals was found to have the following characteristics:

$$N = 3 \qquad Z = 5 \qquad D_{cpu} = 3 \qquad D_{disk} = 2$$

However, further investigation revealed that a software lock had to be held for *one–third* of the cpu time and for *one–half* of the disk time.

1. Performance estimates for this model can be obtained from a two-class model in which one class (U) includes a visit to an artificial *delay* center, $D_{U,delay}$, and the other class (L) in placed in a *domain* with a memory limit, M_L. Fill in the table below with the parameters of such a two class model. Put "?" in two boxes to indicate which parameter value for each class is initially unknown.

class	N_c	Z_c	M_c	$D_{c,cpu}$	$D_{c,disk}$	$D_{c,delay}$
U						
L						

2. Describe briefly an iterative procedure to determine mutually consistent values of the two parameter values that are not known initially.