Exam #1 for Performance Evaluation (CIS 6930) Fall 2004

>>> SOLUTIONS <<<

Welcome to Exam #1 in Computer Networks II (CIS 6930). You have 75 minutes. Read each problem carefully. There are seven required problems (each problem is worth 14 points - you get 2 points for correctly following instructions). You may have with you a calculator, pencils, eraser, blank paper, lucky rabbit's foot, and one 8.5 x 11 inch "formula sheet". On this formula sheet you may have anything you want (definitions, formulas, etc.) *handwritten by you*. You may use both sides. Computer generated text, photocopies, and scans are not allowed on this sheet. Please submit your formula sheet with your exam. Please start each numbered problem on a new sheet of paper and do not write on the back of the sheets (I do not care about saving paper!). Submit everything in problem order. No sharing of calculators. Good luck and be sure to show your work. A Z-transform table is included as part of this exam.

Problem #1

Answer the following short-answer questions in no more than one to three sentences each:

a) What is performance?

A quantitative measure of a system.

b) What are the three causes of delay in a computer network? How can each of these delays be reduced?

Transmission, propagation, and queueing. Transmission and queueing can be reduced with higher data rate links. Propagation can only be reduced by bringing the two nodes closer together (e.g., by caching of server data near to a client).

c) Briefly describe capacity planning.

A process used to design large IT systems with inputs of workload evolution, system parameters, desired service and outputs of saturation point, cost-effective alternatives.

d) There are two reasons (or goals) for studying time series, what are they?

1) To be able to identify underlying phenomena causing the time series
2) To be able to forecast future values

e) In order of "ugliness", what are the ways to model a system according to Kleinrock?

1) Mathematical analysis leading to a closed-form expression
2) Mathematical analysis leading to a numerical procedure
3) Simulation study
4) Experimental study (build the system)

f) What is operational analysis?

Fundamental relationships that hold independent of distributions (of interarrivals and service times). Relationships that can be measured.

Problem #2

Here are some simple probability questions. Answer them.

a) Assume that you have a system with a packet rate of R pkts/sec and a probability of packet loss of p (i.e., $p = \Pr[\text{packet is lost}]$). Assume that p is iid. Assume that R and p are such that when you increase R by N you also increase p by N. One means of reducing information loss is to send each packet duplicated N times, or as an N-tuple of duplicated packets. For example, for a rate R the packets sent are: 1 2 3 4 5 ... and for a rate 2R the packets sent are: 1 1 2 2 3 3 4 4 5 5 ... Define goodput as the rate of original packets received. For example for packets 1 to 5 for a packet stream of rate 2R we might receive: 1 X 2 X 3 3 4 4 XX where X is a lost packet. The goodput is 0.80*R (packet 5 was never delivered). What is the expression for goodput as a function of N, R, and p?

   For any packet tuple what is the probability that at least one packet gets through (i.e., at least one packet is not lost)? This is simply 1 minus the probability that all packets in the tuple were lost, which is: $\Pr[\text{At least one packet is not lost in a tuple of } N \text{ packets}] = 1-(N\cdot p)^N$. Note that the $N \cdot p$ is a result of the problem statement where the probability of a packet loss increases by N as the overall rate also increases by N (obviously, $N \cdot p < 1$ must hold for this to be interesting). Thus, the goodput is simply: $\text{goodput} = \left[1-(N\cdot p)^N\right]\cdot R$.

b) Assume a server has a failure probability of 0.25 during an overnight operation (i.e., the probability of any given server having failed by morning is 0.25). For two servers, plot the pmf and CDF of the random variable X where $\Pr[X = x] = \Pr[x \text{ servers have failed}]$. Find the means and variance of X. You may assume that failures are independent and identically distributed..

   This is a binomial distribution, but we can do this problem “by hand” quite easily:

$$\Pr[X=0] = (1-0.25)(1-0.25) = 0.5625$$
$$\Pr[X=1] = (1-0.25)(0.25)+(0.25)(1-0.25) = 0.3750$$
$$\Pr[X=2] = (0.25)(0.25) = 0.0625$$

$$Mean = 0\cdot 0.5625 + 1\cdot 0.3750 + 2\cdot 0.0625 = 0.50$$
$$Var = \left(0^2\cdot 0.5625 + 1^2\cdot 0.3750 + 2^2\cdot 0.0625\right) - 0.50^2 = 0.3750$$

```
| pmf                          | CDF
|                              |
|                              |
|                              |
|                              |         +--- 1.0
|                              |     +---+ 0.9375
| 0.5625                       |     |
|   |                          | +---+  0.5625
|   | 0.3750                   | |
|   |   |   . 0.0625           | |
+---+---+---+---               +---+---+---+---
    0   1   2  x                   0   1   2  x
```

Problem #3

Let U be a random variable that is the sum of N Poisson distributed random variables. What is the distribution of U? Show your work.

A sum of random variables is a convolution of pdf's which is a multiplication of the *z* transforms of the pdf's. Multiplying exponentials is done by adding their exponents. So, a product of multiple F(z)'s for a Poisson process results is a Poisson process. For example, two Poisson distributions with the same rate parameter, λ, sum to $F(z) = e^{-2\lambda(1-z)}$, <u>which is the same transform and hence the same distribution</u> (recall that each distribution has a unique transform). This sum of rv's results in the same distribution is a unique property of the Poisson process.

Problem #4

Answer the following short questions on measurement:

a) Given a series of time interval values, T_i, and corresponding rate values, M_i, for $i = 1,2,\ldots N$. Show how to compute the mean time interval value and the mean rate value.

We use arithmetic mean for time intervals and harmonic mean for rates. So,

$$E[T] = \frac{1}{N}\sum_{i=1}^{N} T_i$$

$$E[M] = \frac{N}{\sum_{i=1}^{N} \frac{1}{M_i}}$$

b) Briefly describe SNMP

The Simple Network Management Procotol is a request/response protocol between a server and an agent in a network device. The network device collects measurement data (e.g., bytes in and out, packet errors, etc.) and stores them in a Management Information Base. The Agent communicates with the manager to GET, SET, and TRAP on these MIB values.

c) Briefly describe the MRTG tool.

A tool (written in Perl) that automatically GETs a select MIB variable (using SNMP) and plots it. The plot is stored on a web server for global access. MRTG consolidates graphs by blocking allow it to present data by minute, hour, day, week, and so on.

Problem #5

Describe the experiment you would conduct to determine how much longer the multiplication of two double variables (say, `x = a*b;`) takes than the addition of two variables (say, `x = a + b;`) in the C programming language. Sketch-out the necessary code and describe how you would use it.

The following code will measure the per iteration time of the loop (for) and the statement for addition or multiplication. Key to this solution is the start and stop of timing and use of a loop to repeat the experiment (which will have time interval much less than a clock tick or the overhead to start and stop the timing) many times. Running this experiment for each of the two statements and then subtracting the recorded elapsed times will give the difference in time between an addition and multiplication. The experiment should be run many times (say, many tens of times) and the means used.

```c
#include <stdio.h>          // Needed for printf()
#include <sys\timeb.h>      // Needed for ftime() and timeb structure

#define NUM 2000000000

void main(void)
{
  double          x, a, b;    // Variables
  struct timeb start, stop;   // Start and stop times structures
  double          elapsed;    // Elapsed time in seconds
  int                   i;    // Loop counter
  // Start timing
  ftime(&start);

  for (i=0; i<NUM; i++)
  {
    x = a * b;
    // x = a + b;
  }

  // Stop timing and compute elapsed time
  ftime(&stop);
  elapsed=((double) stop.time + ((double) stop.millitm * 0.001)) -
          ((double) start.time + ((double) start.millitm * 0.001));

  // Output elapsed time
  printf("  Time per iteration = %f microsec \n", (1.0e6 * elapsed / NUM));
}
```

Problem #6

For an M/D/1 queue $L = \rho + \rho^2 / [2(1-\rho)]$. Find L_q, W, and W_q,. Here L is the mean number in the systems, L_q the mean number in the queue, W the mean time in the system, and W_q the mean time in the queue.

For $\rho = \frac{\lambda}{\mu}$ we have:

$$L_q = L - \rho = \frac{\rho^2}{2(1-\rho)}$$

$$W = \frac{L}{\lambda} = \frac{1}{\mu} + \frac{\rho}{2\mu(1-\rho)}$$

$$W_q = W - \frac{1}{\mu} = \frac{\rho}{2\mu(1-\rho)}$$

Problem #7

Consider an ATM switch with a cell switching rate of 10,000 cells per second. Cells are a fixed length. Assume that the ATM switch has a buffer of capacity 1000 cells. Using fluid-flow assumptions, plot the queue length (buffer length) as a function of time for the following input (assume that at time 0 the queue contains zero cells):

- From time 0 to 2 seconds the input rate is 5,000 cells per second
- From time 2 to 4 seconds the input rate is 12,000 cells per second
- From time 4 to 9 seconds the input rate is 9,000 cells per second

From time 0 to 9 seconds, about how many cells are lost due to buffer overflow? Show your work!

From time 0 to 2 seconds the arrival rate is less than the service rate, so the queue length remains 0 cells. From time 2 to 4 seconds the queue is increasing in length at a rate 2,000 cells per second. After 0.5 second (at time = 2.5 seconds) the queue is full. Thus, from time 2.5 to 4 seconds 3,000 cells are lost. From time 4 to 9 seconds the queue is decreasing at 1,000 cells per second and thus empties out at time 5 (and from time 5 to 9 the queue is empty). The plot is: