COMP9334 Capacity Planning
Assignment, Session 1, 2017

April 6, 2017

Instructions

(1) There are 3 questions in this assignment. Answer all questions.

(2) The total marks for this assignment is 15 marks.

(3) In answering the questions, it is important for you to show us your intermediate steps and tell us what arguments you have made to obtain the results. You need to note that both the intermediate steps and the arguments carry marks. Please note that we are **not** just interested in whether you can get the final numerical answer right, but we are **more** interested to find out whether you understand the subject matter. We do that by looking at your intermediate steps and the arguments that you have made to obtain the answer. Thus, if you can show us the perfect intermediate steps and the in-between arguments but get the numerical values wrong for some reason, we will still award you marks for having understood the subject matter.

If you use a computer program to perform any part of your work, you are also required to submit the program.

(4) The submission deadline is 11:00pm Sunday, 9 April 2017. Late submission will cap the maximum mark that you receive. Submissions after 11:00pm on 14 April will no longer be accepted.

(5) Your submission should consist of:

  (a) A report describing the solution to the problems. This report can be typewritten or a scan of handwritten pages. This report must be in pdf format and must be named assignment.pdf. The submission system will only accept the name assignment.pdf.

  (b) One or more computer programs if you use them to solve the problems numerically. You should use tar/zip/rar to archive all the computer programs into one file with the name supp.tar, supp.zip or supp.rar. The submission system will only accept these names.

(6) Submission can be made via the course website.

Question 1 (3 marks)

An interactive computer system consists of three devices: a CPU and three disks (denoted by Disk1, Disk2 and Disk3). The system was monitored for 30 minutes and the following measurements were taken:

Number of completed jobs	1,231
Number of CPU accesses	2,789
Number of Disk1 accesses	17,412
Number of Disk2 accesses	13,424
Number of Disk3 accesses	15,978
CPU busy time	1102 seconds
Disk1 busy time	929 seconds
Disk2 busy time	1017 seconds
Disk3 busy time	1265 seconds
~~Think time~~	~~27 seconds~~

In addition, you can assume that the think time per job is 27 seconds.

(a) Determine the service demand at each device of the system.

(b) Use bottleneck analysis to determine the asymptotic bound on the system throughput when there are 40 active terminals.

(c) Using your results in Part (b), compute the minimum possible response time when the number of terminals is 40.

Question 2 (6 marks)

A call centre has 4 operators. Calls arrive at the call centre obey the Poisson distribution with a rate of 20 calls per hour. The service time required by each call is exponentially distributed with mean service time 10 minutes.

(a) Assuming the call centre has no facilities to place an incoming call on hold. This means that if all the operators are busy, an incoming call will be rejected. Compute the probability that an incoming call is rejected.
Note: You do not need to derive the Markov chain for this part. You are allowed to apply standard results from Queueing Theory.

(b) The owner of the call centre would like to decrease the call rejection rate to less than 50% of the value calculated in Part (a). The owner decides to achieve this by introducing a queue which places the incoming calls on hold when all the operators are busy.

The holding queue will consist of M holding slots where M is to be determined. If an incoming call arrives when all operators are busy, it will be placed in the holding queue provided that a vacant holding slot is available. If the call arrives when all operators are busy and all M holding slots are used, then the call is rejected. Assuming that the customers are infinitely patient in the sense that once their call is accepted in the (holding) queue, they will wait until they get to the operator and will only leave the system after they have been served.

(i) Formulate a continuous-time Markov chain for the call centre with 4 operators and M holding slots. Your formulation should include the definition of the states and the transition rates between states.

(ii) Write down the balance equations for the continuous-time Markov chain that you have formulated in Part (b,i).

(iii) Derive expressions for the steady state probabilities of the continuous-time Markov chain that you have formulated.

(iv) Use your answer in Part (b,iii) to determine the smallest value of M required to reduce the call rejection rate to less than 50% of the value calculated in Part (a).

Note: There are many (in fact, infinite) choices of M that can reduce the call rejection rate to less than 50% of the value calculated in Part (a). We are only interested in the smallest value of such M's.

(v) For the value of M that you have calculated in Part (b,iv), determine how long an accepted call will need to wait before it will be served by an operator.

Note: If you use a computer program to derive your numerical answers, you must include your computer program in your submission. Do not forget to show us your steps to obtain your answer.
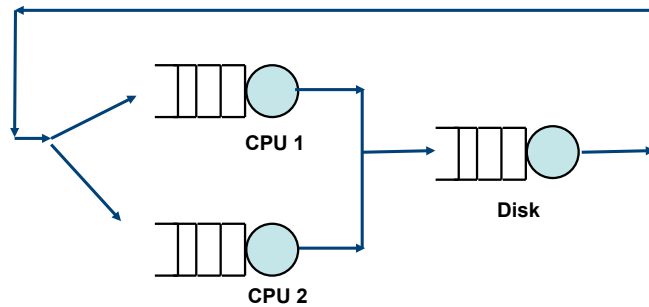
Question 3 (6 marks)

Figure 1: Figure for Question 3.

Consider the computer system shown in Figure 1. The system consists of three devices: a disk and 2 CPUs. Each device is modelled as a server and a queue. The system is at peak load and there are three (3) jobs circulating in the system at all times. During each round that a job circulates the system, the job requires processing from *one* of the CPUs and then followed by the disk. Assuming that:

- The processing time required by each job per visit to the disk is exponentially distributed with mean 50 milli-seconds.

- The two CPUs have different mean processing times. The mean processing times for CPU1 and CPU2 are, respectively, 50 and 100 milli-seconds. Both processing time distributions are assumed to be exponential.

- After a job has left the disk, it will proceed to receive processing at one of the CPUs immediately. In any attempt to utilise the faster CPU (i.e. CPU1), a job will only be sent to CPU2 if ~~it is idle~~ CPU2 is idle and CPU1 is busy. In other words, if CPU2 is busy, the job will be sent to CPU1; and if both CPU1 and CPU2 are idle, the job will be sent to CPU1.

Answer the following questions.

(a) Let the states be the following 3-tuple:

(number of users in the CPU1, number of users in CPU2, number of users in the disk),

formulate a continuous-time Markov chain for this computer system. Your formulation should include (1) a list of states; (2) the transition rates between the states.

(b) Write down the balance equations for the continuous-time Markov chain that you have formu-
lated in Part (a).

(c) What is the steady state probability for each state?

(d) What is the throughput of the system?

(e) What is the mean response time of the CPU1?

(f) How long does a user have to wait, on average, at the disk before it gets served?

*Note: If you use a computer program to solve for the steady state probabilities, you need to show
us your code. Also, do not forget to show us the steps you use to get your answers.*

$- - -- $ End of assignment $- - --$