Simulation and Modelling
Problem Sheet 1

1. Assume a system with a single resource (server) has been observed for 1 minute. If there were 20 completions, the server was busy for 40 seconds, and the average number of jobs was 2,
   (a) What was the service requirement of each job?
   (b) What was the throughput of the system?
   (c) What was the utilisation of the resource?
   (d) What was the average time each job spent in the system?
   (e) How long on average did each job wait before being served, assuming a FCFS queueing discipline?

2. For the closed network in Figure 1, assume initially that $\mu_1 = \mu_2$.

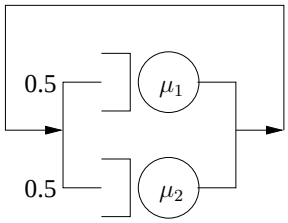

Figure 1: Closed queueing network

   (a) What can you say about the utilisations of the two servers and mean populations of the two queues?
   (b) What is the throughput of the system under i. low load (e.g. $N = 1$), ii. high load (large N)?
   (c) Under high load (large population, N) what would you expect to happen to the utilisations and mean queue populations if μ_1 were to increase by 10%, say?
   (d) Under the same high-load assumption what will happen to the throughput and mean response time if μ_1 were to increase by the same 10%. What about 50% or a factor of 20?

3. (Exam question 2a, 2016) In an interactive system a number of clients submit jobs to a server that comprises a database engine and a processor. 200 user job completions were observed in an observation period of 140 seconds during which time there was a fixed population of 20 clients. There were 1260 completions at the processor and 810 at the database engine and the mean service times at the two devices were 50ms (0.05s) and 130ms (0.13s) respectively. The average client think time was 5 seconds and there is a single class of jobs.

(a) Compute the utilizations of the two devices over the observation period and determine the bottleneck device.

(b) In order to cope with an increasing number of users it has been proposed that the bottleneck device be duplicated so that there are two identical devices sharing the load that is currently being imposed on the original. If both duplicates experience half the average service demand of the original device, what will be the maximum throughput that can be achieved for large N? Show your working.

4. Consider the interactive system shown in Figure 2 for which the following measurements have been taken:

   - $N = 25$
   - $Z = 20$
   - $X = 0.5$
   - $N_{pre} = 12$
   - $V_{cpu} = 3$
   - $S_{cpu} = 0.2$

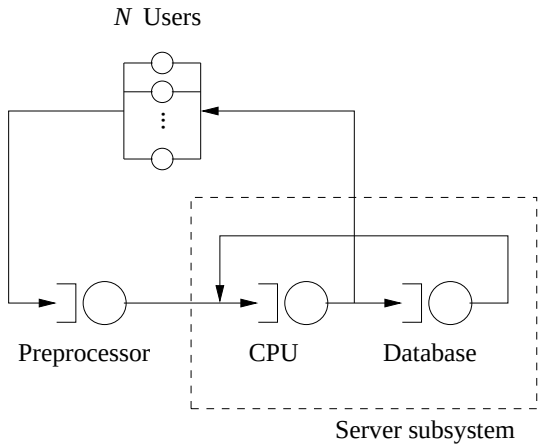

Figure 2: Interactive System

(a) What is the average response time of the preprocessor?

(b) What is the CPU utilisation?

(c) What is the average response time of the main server?