2001 Paper 8 Question 14

Computer Systems Modelling

Two servers operate with different performance characteristics at mean rates μ_1 and μ_2. You wish to combine them into a single system by associating each server with a separate FIFO queue and dispatching incoming work items to the first queue with probability p_1 and to the other queue with probability p_2. Incoming items arrive at a rate λ and none are discarded from the system.

You may assume that the inter-arrival-time distribution and both service-time distributions are exponential, that there is no limit on the queue lengths and that the population size is infinite.

(a) Using Kendall notation, describe the first server and its queue. Construct a Markov-chain model for this part of the system. [2 marks]

(b) Let $q_{k,i}$ denote the probability that there are exactly i items of work in server k and its queue. By using detailed flow balance equations or otherwise express $q_{k,i}$ in terms of λ, p_k and μ_k. [6 marks]

(c) Hence derive T_k, the mean response time of work items served at k. [6 marks]

(d) Suppose that the system administrator wishes to ensure that work items receive the same mean response time irrespective of which server they visit. Express p_1 in terms of λ, μ_1 and μ_2. Qualitatively, when is it reasonable to consider dispatching work to both servers to maintain an equal mean response time? How will the system behave at other times? [6 marks]