UBINET/SI5: Performance Evaluation of Networks

Homework 6
To be returned on 5 November 2024 at 9 am

Homeworks are a personal effort. Copied solutions will get 0 for a grade.

6.1 A labyrinth

A labyrinth is divided in 3 sectors having each an entry point. There is only one exit point that is in sector 3. Anyone entering a sector spends a time that is exponentially distributed to find a door. In sector $i = 1, 2$, with some probability, the door opens to the entry point of the sector (the person will have to cross the sector all over again) and with the complementary probability it opens to the next sector, sector $i + 1$. In sector 3, the door opens either to the entry point of the sector or to the exit point. Persons wishing to cross this labyrinth arrive according to a Poisson process with rate λ and choose one of the three entry points at random, with equal probability. Persons do not know that only sector 3 has an exit point. Only one person can be inside a sector at any given time. If a sector is not empty, then one must wait at the entry point until the sector empties.

1. Say why this labyrinth can be modeled as the Jackson network of Figure 1.
   Write explicitly the parameters of the Jackson network:
   – exogenous arrival rates at all nodes,
   – routing matrix $\mathbf{P}$,
   – probabilities to leave the network.

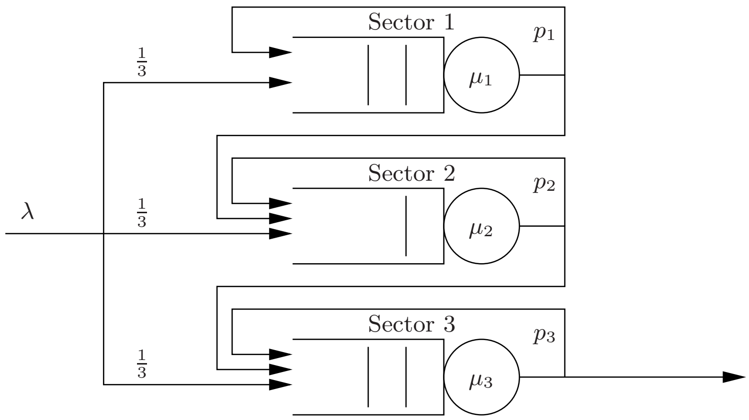

Figure 1: A Jackson network with 3 nodes.

2. Write and solve the traffic equations.
3. What are the irreducibility conditions?
4. Find an upper bound on λ such that the network is stable.
5. What is the stationary probability that only sector 3 is not empty?
6. What time a person needs on average to exit the labyrinth?

6.2 Caching in a Web server

The aim of this problem is to evaluate the impact of a caching mechanism on the performance of a web server. The web server is composed of four components: the processing unit (denoted node A) that processes all requests and manages the cache of the web server, two disks (denoted nodes B and C) and one communication unit (node D) whose role is to send the requested documents to the users. A table storing mappings between documents and their location (disk B or C) is available at node A.

Any document requested by users must be in one and only one of the disks at nodes B or C and possibly in the cache at node A. Let p_A be the probability of this last case. The event of having a document in disk B but not in cache A occurs with probability p_B. p_C is defined similarly, so that $p_A + p_B + p_C = 1$. The web server handles the requests according to the following algorithm:

1. check whether the requested document is in the local cache of the processing unit or not; if yes, send a copy of the document to node D and go to step 4; if not, retrieve its location from a table at node A;
2. access the disk containing the document, according to the information retrieved from the table; make a copy of the requested document and go to node A;
3. update the local cache at the processing unit by inserting in it a copy of the document; send the copy of the document to node D;
4. at the communication unit (node D), transmit the copy of the document to the user that has requested it.

Steps 1 and 3 are handled by node A. The time needed to perform any of these steps is modeled by an exponential random variable with rate μ_A. Step 2 is handled by either node B or node C. The time needed to perform this step is assumed to be exponentially distributed with rate μ_B, when it is node B performing it or rate μ_C, when it is node C performing it. Node D handles step 4 in a time that is exponentially distributed with rate μ_D. Observe that steps 1 and 4 are always performed for any request, whereas steps 2 and 3 are additionally performed only for those requests asking for documents that are only on disks.

We will assume that requests arrive to the web server according to a Poisson process with rate $\lambda > 0$.

1. Explain why this web server can be modeled as a Kelly network with $K = 4$ nodes (A, B, C, D) and $R = 3$ classes $(1, 2, 3)$ with routes r_1, r_2, and r_3 respectively.
   What are the three routes?
   What is the traffic intensity λ_k along route r_k for $k = 1, 2, 3$?
2. Compute $\hat{\lambda}_{i,k}$ the global arrival rate of customers of class $k = 1, 2, 3$ in node $i \in \{A, B, C, D\}$ and $\hat{\lambda}_i$ the global arrival rate in node $i \in \{A, B, C, D\}$.
3. Under which condition is the network stable?
4. Compute the expected response time of the web server (denoted as $\overline{T}$).

5. Numerical applications: assume that $1/\mu_B = 12.5$ ms, $1/\mu_C = 20$ ms, $1/\mu_D = 10$ ms. We will consider two different scenarios:

   - the cache is small: $1/\mu_A = 5$ ms, $p_B = p_C = 0.4$.
   - the cache is large: $1/\mu_A = 25$ ms, $p_B = p_C = 0.1$.

   For each case:
   (i) compute TH, the maximum throughput of the web server (in requests per second),
   (ii) find the bottleneck node, and
   (iii) compute the expected response time when $\lambda = 20$ requests/sec.
   Which configuration is the best?

6. Is it possible to model this web server as a Jackson network? Justify your answer.