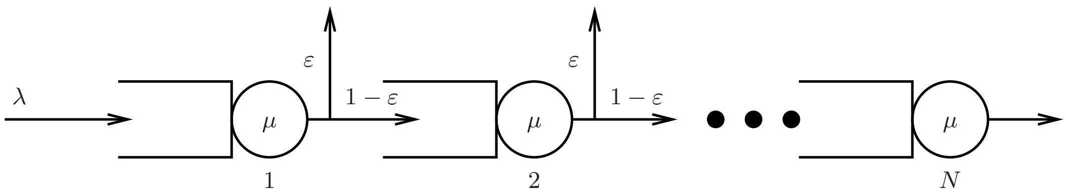
UBINET: Performance Evaluation of Networks

Homework 6
To be returned on 3 November 2015

6.1 A series network

Exercise borrowed from Philippe Nain

Consider the series network in the figure below. It contains N nodes, each being an infinite capacity single-server queue with exponentially distributed service time. The service rate of each node is μ. Packets enter this network at node 1 following a Poisson process with rate λ, such that $\lambda < \mu$. At each node except for the last one, a fraction ε of packets leaves the network after receiving service. Packets leaving the network before the Nth node are considered lost.

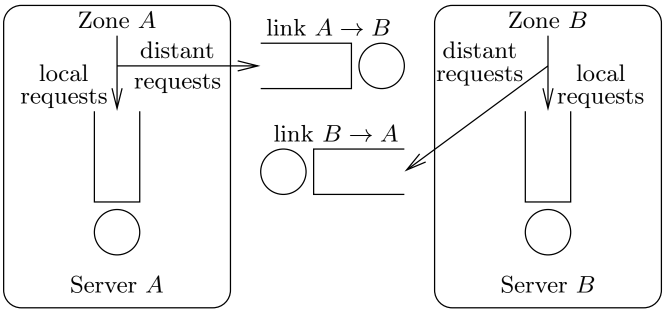


1. Write the routing matrix when $N = 4$.
2. Express the throughput of node i in terms of λ, ε and i, for $i = 1, 2, \ldots, N$.
3. What is the throughput of the entire system (output rate of non-lost packets)?
   Compute then the loss probability p_{loss}.
4. What is the system response time seen by non-lost packets $\overline{T}_{\text{real}}$?
   How does it compare to $\overline{T}$, the expected sojourn time of a packet picked at random?

6.2 Web server mirroring

Exercise borrowed from Alain Jean-Marie and Philippe Nain

Consider a web server A that is located in a geographical zone A. Server B, located in geographical zone B, is a mirror for server A. Zones A and B are inter-connected so users from either zone can access documents stored on either server. Only the web servers and the communication links between the zones incur non-negligible delays. Local communication links need not to be modeled. As a result, the only queues that need to be considered are those shown in the figure below. Web servers and communication links are seen as infinite capacity single-server queues.[1]



Assume that 80% of the users in zone B send their requests to the nearby server whereas only half of the users in zone A favor the geographically close server. Requests to a distant server go through the *outbound queue* to be processed by the *distant server* which sends the requested document through the *inbound queue*. We assume the following: (i) each web server may process 2000 requests per second; (ii) the bandwidth of each communication link is 32 Megabits per second (Mb/s); and (iii) documents and requests have an average size of 1000 bytes.[2]

Requets in zones A and B are generated according to independent Poisson processes with respective rates $\lambda_A > 0$ and $\lambda_B > 0$. We assume that service times at web servers A and B and at communication links $A \to B$ and $B \to A$ are exponentially distributed. The respective service rates are $\mu_A > 0$, $\mu_B > 0$, $\mu_{AB} > 0$ and $\mu_{BA} > 0$. Service times are all mutually independent and independent of both Poisson arrival processes. These assumptions will allow us to model this system as a four-node Kelly network (nodes are denoted A, B, AB and BA).

1. Give μ_A, μ_B, μ_{AB} and μ_{BA} in customers per second (requests and documents are all customers).

2. Specify the routes in this Kelly network and give the arrival rate on each route.
   Find the total arrival rate in each of the four nodes (use notation seen in class).

3. Give the pairs (λ_A, λ_B) for which the system is stable (justify your answer).

4. Find $\overline{T}_A$ and $\overline{T}_B$, the average response time of the users in zones A and B, respectively.
   Find $\overline{T}$, the average response time of an arbitrary user.

5. Numerical application: $\lambda_A = 2000$ requests per second, $\lambda_B = 500$ requests per second. Find $\overline{T}_A$, $\overline{T}_B$ and $\overline{T}$.

6. We now assume that each web server stores permanently copies of all the documents that can be required by the users, so that users are always served locally. As a result, inter-zone communications become useless. However, the number of requests per second that the web servers A and B can handle falls down to 1500.

[1] The service discipline is irrelevant as the system will be modeled as a Kelly network.
[2] Clearly this is not realistic (made for the sake of simplicity).

Give the new formula for $\overline{T}_A$, $\overline{T}_B$ and $\overline{T}$.
By comparing to the numerical values found earlier, say if this system is more or less efficient than the original one as far as $\overline{T}_A$, $\overline{T}_B$ and $\overline{T}$, respectively, are concerned.