A SIMPLE DERIVATION OF QUEUEING DELAY IN A TREE NETWORK OF DISCRETE-TIME QUEUES WITH DETERMINISTIC SERVICE TIME

Eytan Modiano[1] and Jeffrey E. Wieselthier
Information Technology Division
Naval Research Laboratory
Washington, DC 20375

Anthony Ephremides
Elec. Eng. Dept. & Inst. of Systems Research
University of Maryland
College Park, MD 20742

In this paper we consider a network of discrete-time queues in which the service time is deterministic and the same at each queue. Such a network of queues arises in a data communication network model, where data is formatted into fixed-length packets. An important performance index in such networks is queueing delay. The model generally used for the analysis of delay in a large network is based on ***Kleinrock's Independence Assumption,*** which assumes that the queues at each link behave as independent queues regardless of the interaction of traffic between the different links [1]. This model is reasonably good for systems involving exponential arrivals, a densely connected network, and uniform loading among source-destination pairs. Otherwise the model becomes very inaccurate in predicting delay.

The exact computation of delay in a network of queues with constant service time is complicated by the fact that the outputs of intermediate queues in the network (which serve as inputs to other queues) are correlated. This correlation, in an N-node network, gives rise to an N-dimensional infinite Markov chain whose analysis is very difficult. There has been some effort in the literature to develop a delay model for a packet-switched network with fixed-length packets. It is clear that in a discrete-time network with fixed-length packets of one slot duration, only one packet can depart from a link during any slot. In [2] an attempt is made to take advantage of this regular nature of departures by modeling the departures from a link in the network as an independent Bernoulli process with rate equal to the utilization of that link. In [3] a direct calculation of the moment generating function for the steady-state queue sizes for a tandem of queues is obtained. However, the direct nature of the approach makes it difficult to extend to more than a simple tandem because of the rapidly increasing complexity of the calculations.

In this paper we offer an exact solution to a tree network of discrete-time queues with identical service times. Our solution is based on considering an equivalent network, where priority is given to customers in transit (over exogenous customers entering the system). The solution to the equivalent model is easily computed, from which the solution to the original model (without priorities) can be obtained. Thus, the priority scheme is devised only as a means to simplify the analysis of the system, but is not proposed for actual use in a network.

We begin by considering the two discrete-time queues in tandem shown in Figure 1. The input to the first queue is a Poisson process of rate λ_1, and its output is fed into the second queue. Additionally, the second queue has an independent exogenous Poisson arrival stream of rate λ_2. Both queues have the same, constant, service time. The first queue is then a simple discrete-time M/D/1 queue with arrival rate λ_1. The average delay for such a system is well known. The analysis of delay at the second queue is complicated by the fact that the inputs to that queue are a mixture of the outputs of the first queue and an exogenous Poisson stream. For our analysis we assume that in-route traffic has priority over exogenous traffic. That is, customers departing from the first queue are immediately served at

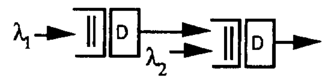

Figure 1. Two discrete-time queues in tandem.

the second queue without experiencing any delay. This assumption is not required of the actual system, and is only a hypothetical one made for the purpose of simplifying the analysis. In fact, we note that the overall average delay at the second queue is the same for either the system with priority or the actual system without priority.

Now considering the tandem with priority, we observe that customers departing from the first queue will experience no delay at the second queue. Exogenous traffic arriving at the second queue will be served only when there are no customers departing from the first queue. Since a customer departs the first queue exactly one time unit after it begins receiving service, the entire two-stage tandem system can be thought of as a single server system with two exogenous Poisson streams where one stream receives priority over the other. With this observation we can show that the average queue size at the second queue is equal to

$$\overline{Q}_2 = \frac{(\lambda_1 + \lambda_2)^2}{2(1-\lambda_1-\lambda_2)} - \frac{\lambda_1^2}{2(1-\lambda_1)}$$

and the average delay is

$$D_2 = \frac{\overline{Q}_2}{\lambda_1 + \lambda_2}$$

This result is consistent with that obtained in [3] when the arrival processes are independent Poisson processes. Using a similar priority scheme this result can now be extended to obtain a solution to the queueing delay in a tree-network of queues.

References

[1] L. Kleinrock, Communication Nets: Stochastic Message Flow and Delay, McGraw-Hill, New York, 1964.

[2] E. Modiano, J. E. Wieselthier and A. Ephremides, "An Approach for the Analysis of Packet Delay in an Integrated Mobile Radio Network," Proc. of the 1993 Conference on Information Sciences and Systems, p. 138, Baltimore, MD, March 1993.

[3] J. A. Morrison, "Two Discrete-Time Queues in Tandem," IEEE Transactions on Communications, COM-27, pp. 563-573, March 1979.

[1] Eytan Modiano was a NRC post-doctoral research associate at NRL when this work was performed. He is presently with the Communications Division of MIT Lincoln Laboratory, Lexington, MA 02173.