Queuing Consequences of Self-similar Traffic

P.J.B. King Hadi A. Larijani*

December 11, 1998

Abstract

The discovery of the self similarity of Ethernet traffic at Bellcore has had a profound effect on the world of performance modelling.

This paper examines the effects of self similarity of input traffic on the performance of a queue. We use a traffic characterisation due to Robert and Le Boudec which allows self similar traffic with an arbitrary mean and Hurst parameter to be generated. The technique gives a Markovian environment in which a queue can be analysed using matrix geometric techniques. The effect of increasing the Hurst parameter on the queuing behaviour is investigated.

1 Introduction

In the early 1990s the world of network performance analysis was fundamentally changed by traffic measurements undertaken at Bellcore. Using a full trace of traffic on an Ethernet, Leland and colleagues[1], established that the traffic showed patterns that were independent of the time scale used to measure them. For example, the proportion of time periods that were idle (say) was independent of the actual length of the periods involved. The traffic pattern had all the self similar characteristics of a fractal curve or a chaotic map. This self similarity is not evident in any of the standard traffic models. In particular, Poisson assumptions not lead to self similar traffic. Since then, a number of attempts have been made to adapt conventional queueing models to this new traffic pattern, or to develop new queueing models. This paper follows the route of adapting conventional models and attempts to show how

*Dept. of Computing and Electrical Engineering, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS. Email: pjbk@cee.hw.ac.uk, hal@cee.hw.ac.uk

self similar traffic will change the expected queue length. A self similar traffic model due to Robert and Le Boudec[5] is used to specify the input process to a queue. The matrix geometric techniques of Neuts[2, 3] are used to evaluate the steady state probabilities and calculate mean queue lengths.

2 Self Similarity

For a time series, such as a set of interarrival times, *self-similarity* is present if there are invariants which are independent of the scale in which the measurement is made. For example, if we consider a Poisson process, as longer and longer intervals are considered, the probability that an interval will have no arrivals will become significantly smaller. As the traffic is aggregated it becomes less bursty. If one were to treat this graphically, a graph of the variable, assuming it is self similar, can be rescaled without changing its overall "shape". If $X(t)$ is a random variable, then changing the time scale over which the graph is drawn will only change the amplitude of the function, not its other characteristics. In particular, if the functions $X(t) - X(t_0)$ and $(X(rt) - X(rt_0))/r^H$ are statistically indistinguishable, then $X(t)$ is said to be self similar with Hurst parameter H.

A Hurst parameter between 0 and 1 corresponds to fractional Brownian motion. Normal Brownian motion has $H = 0.5$. Measurements on Ethernet traffic have shown Hurst parameters ranging from 0.7 to 0.99[1]. Robert and Le Boudec[5] have developed a discrete time Markov chain which can represent self similar traffic. This model requires only three parameters and can describe a wide range of self similar behaviours. They assume that the input process can be in one of n states, in discrete time. When in state 1, a single arrival occurs with probability 1. In all other states, no arrivals occur. The Markov chain governing the changes between states of the input process is such that transitions are made from state 1 to all states, and also from all states to state 1, but not between arbitrary states. The transition matrix of the Markov chain is:

$$P = \begin{pmatrix} 1 - 1/a - 1/a^2 \cdots - 1/a^{n-1} & 1/a & 1/a^2 & \cdots & 1/a^{n-1} \\ q/a & 1 - q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1 - (q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (q/a)^{n-1} & 0 & 0 & \cdots & 1 - (q/a)^{n-1} \end{pmatrix}$$

The expected number of arrivals in a slot is given by

$$E[X] = \frac{1 - 1/q}{1 - (1/q)^n}$$

For a particular mean arrival rate, one can determine the parameter q. The parameter a does not affect the mean, but it has an effect on the Hurst parameter. There is no simple relationship, but increases in a increase the Hurst parameter of the traffic.

3 Queueing Model

Using this input process in a queueing model is possible in a number of ways. First, a discrete time queue could be constructed, and the process used as the input to that queue. Since at most 1 customer arrives per slot, and one customer can be served, this would not be very interesting. A more interesting approach would be to use several of the self similar processes as input, so that queues might have a chance to build up. This would involve construction of a Markov chain with each state representing the states of the individual self similar processes. Construction of such a chain is straightforward, and then the analysis of a discrete time queue with that input process could be conducted.

A second approach, which is followed here, is to (incorrectly) assume that the process is continuous, and to assume that when the input process is in state 1 arrivals form a Poisson process with rate 1, and that the server gives exponentially distributed service times with mean 1, whatever the state of the input. This means that we are analysing an $M/M/1$ queue in a Markovian environment, which has been the subject of many studies[2, 4].

The state of the system can be denoted by a pair of integers, (I, J), with I representing the number of customers in the system, and $J, 1 \leq J \leq n$, representing the state of the input process. When the input process is in state 1, jobs arrive in a Poisson stream at rate 1, and in all other states of the input process there are no arrivals. Whatever the state of the input process, service takes place at rate 1.

The steady state probabilities can be denoted as $\pi_{ij} = \Pr(I = i \wedge J = j)$ can be related using the balance equations, and if the vector $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{in})$ is defined, then the balance equations can be expressed as:

$$A_2 \boldsymbol{\pi}_{i+1} + A_1 \boldsymbol{\pi}_i + A_0 \boldsymbol{\pi}_{i-1} = 0$$

The matrices A_2 and A_0 are diagonal matrices, with entries consisting of the arrival and service rates, respectively, in the corresponding state of the input process. A_1 is $P - I - A_2 - A_0$, where P is the transition matrix of the input process. Neuts[2] shows that

$$\boldsymbol{\pi}_i = R^i \boldsymbol{\pi}_0$$

n	q	$E[N]$
2	2.00	2.05
3	2.73	1.97
4	2.91	1.98
8	2.99	1.98
10	3.00	1.98

Table 1: Effect of increasing n, fixed $a = 10$

where R is the unique solution of

$$A_2 R^2 + A_1 R + A_0 = 0$$

and

$$\boldsymbol{\pi}_0 = (1 - R)\boldsymbol{\alpha}$$

where $\boldsymbol{\alpha}$ is the steady state distribution of the Markov chain representing the input process.

The mean queue length can be found using

$$\mathbf{n} = R(1 - R)^{-2}\boldsymbol{\pi}_0$$

since R is known to have spectral radius less than 1.

4 Numerical Results

The behaviour of the single server queue when fed with input generated using the Robert and Le Boudec process was investigated. In all cases, the mean arrival rate was set to 0.666667, so that a simple $M/M/1$ queue would have an expected queue length of 2. the effect of increasing the number of states of the input process was investigated first.

The results are shown in Table 1, where q is the value of the parameter needed to keep the mean arrival rate at 0.6666667, and $E[N]$ is the mean queue length.

The effect of increasing the a parameter can be seen in Table 2. It is not easy to calculate the Hurst parameter of the process directly from the parameters, but increases in a correspond to increases in the Hurst parameter.

5 Conclusions

It appears that the Robert and Le Boudec model can effectively capture self similarity of input traffic. For a small computational cost, the effect of self

a	$E[N]$
5	1.41
10	1.98
15	2.43
20	2.81
25	3.15
30	3.46
35	3.74
40	4.01

Table 2: Effect of increasing a, fixed $n = 8$ and $q = 2.99$

similarity on queue lengths can be seen. Increasing the Hurst parameter of the input process increases the mean queue length, and hence the mean delay encountered by messages. There does not seem to be an inherent limit on the queue length that may be encountered.

Further work is in progress to develop discrete time models which will capture the characteristics of ATM traffic more precisely. Software is being developed to evaluate the parameters of the Robert and Le Boudec model from a trace of packet traffic. When this is complete, the model can be evaluated against the delays encountered by the real life traffic.

References

[1] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994.

[2] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, volume 2 of *Johns Hopkins Series in the Mathematical Sciences*. The Johns Hopkins University Press, Baltimore, Maryland, 1981.

[3] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, New York, NY, 1989.

[4] P. Purdue. The M/M/1 queue in a Markovian environment. *Oper. Res.*, 22(3):562–569, 1974.

[5] S. Robert and J.-Y. Le Boudec. New models for pseudo self-similar traffic. *Performance Evaluation*, 30(1):57–68, 1997.