Accelerated Simulation Method for Power-law Traffic and Non-FIFO Scheduling

Authors:

Sharifah H. S. Ariffin and John A. Schormans

Department of Electronic Engineering,

Queen Mary, University of London,

Mile End Road,

London, E1 4NS

Accelerated Simulation Method for Power-law Traffic and Non-FIFO Scheduling

Sharifah H. S. Ariffin and John A. Schormans

Department of Electronic Engineering, Queen Mary, University of London,

Abstract: The conventional method of simulating networks, i.e. packet-by-packet, always consumes many hours and often days of 'real' time just to simulate a few hours of 'simulation' time. Accelerated simulation techniques reduce the number of events simulated and decrease the time needed for an experiment. In this paper we describe our approach to accelerate simulation of a power-law aggregated ON/OFF traffic through a FIFO queue, and then extend the same idea to a non-FIFO queue.

1. Introduction

Recent research has discovered self-similarity and long-range dependence (LRD) in a variety types of packet traffic, which includes the Ethernet, ATM traffic, Telnet and FTP [1-3]. Packet switched network traffic such as IP data are more accurately represented by this type of model rather than a Poisson or renewal based process. Hence, network performance analysis become more complex and simulation consumes more time [4]. Much research has been done to reduce the amount of time taken to execute a simulation of self-similar traffic with a First-In-First-Out (FIFO) buffer scheduler [4 -6]. The technique of Traffic Aggregation (TA) was used to speed up the simulation time of several ON/OFF input sources and it saved 75% of the usual time taken for a conventional simulation. The TA technique has been extended in Enhanced TA (E_TA), and instead of simulating packet-by-packet arrivals, an event is simulated at the end of an ON period [8]. Furthermore, E_TA places the Imbedded Markov Chain (IMC) such that it significantly reduced the number of events needed to reach the steady state. However, to get an accurate acceleration technique for power-law ON/OFF traffic, a few adjustments have to be made, and there are discussed in section 2.

In the real network, customers' satisfaction is the main priority and to cope with the demands non-FIFO queue scheduling, e.g. DiffServ, is used [9]. This type of scheduler discipline simultaneously satisfies the requirement of the real-time and the non-real-time applications. Hence, it separates different type of services to different queues such as in the Weighted Fair Queuing (WFQ) scheduler. However since events in the E_TA model are simulated at the end of ON periods, modifications are needed to the service times of packets in the buffer. In section 2 the approach of E_TA with a FIFO queue are presented. Section 3 gives the fundamental idea of the E_TA model with a non- FIFO queue, while section 4 presents the results and lastly the conclusions are in section 5.

2. E_TA with FIFO queue

To represent Power-law traffic in the model we have adopted, the ON and OFF periods of the sources are assumed to be Pareto distributed. In TA model, the Imbedded Markov Chain (IMC) is located such that the simulated stochastic process evolves from packet arrival to packet arrival:

$$s_A(k) = s_A(k-1) + 1 - d_A(k) \qquad (1)$$

where:

$s_A(k)$ = the state of the buffer after the kth packet arrival

$d_A(k)$ = the number of packet departures between contiguous packet arrivals 'k-1' and 'k'

This forms an acceleration technique because the packet arrival process driving up the level of packets in the buffer is a concatenation of the N original ON/OFF processes, and it has already been shown elsewhere [4-7] that this allows the same buffer statistics to be derived at a smaller cost in terms of absolute number of simulated events. In E_TA this idea is taken further where the IMC is located such that the simulated stochastic process evolves from a 'burst' of packet arrivals to the next such burst, where each burst constitutes the increase in packet length due to the associated ON period:

$$s_O(k) = s_O(k-1) + a_O(k) - d_O(k) \qquad (2)$$

where:
$s_O(k)$= the state of the buffer after the k th packet arrival

$d_O(k)$= the number of packet departures between contiguous ON periods 'k-1' and 'k'

$a_O(k)$= the number of packet arrivals in an ON period that act to increase the buffer length (the number of Excess Rate arrivals, [7]).

This means that a burst of arrivals in the E_TA method represents many packet-by-packet events in the TA model; this is illustrated in figure 1. To get equivalent traffic accuracy in the FIFO buffer, the input load must match. During the ON periods in the E_TA model, the packet arrival rate is equal to 1 unit time but since E_TA simulates at the end of an ON period, a few things need to be considered: a) the potentially long active periods of arriving packets in a burst, b) service rate in the buffer. In [6], it is reported that mean OFF times have little impact on the simulation. This is incorrect in E_TA because when mean ON time in E_TA is increased from 10 to 30 the equivalent traffic failed to match. However the increment of mean ON time (filling the buffer) followed by a reduction of mean OFF time (emptying the buffer) proved to provide an accurate queue distribution in the buffer, this is justified by the results in section 4. The mean OFF time for E_TA is given by

$$mean_{OFF} = \left[\left(1 - r/R_{ON} \right) \times R_{ON}/r \right] \cdot C \qquad (3)$$

where r is the load, R_{ON} is the ON arrival rate of the aggregated traffic, C is the buffer capacity and all packets are assumed to be the same size.

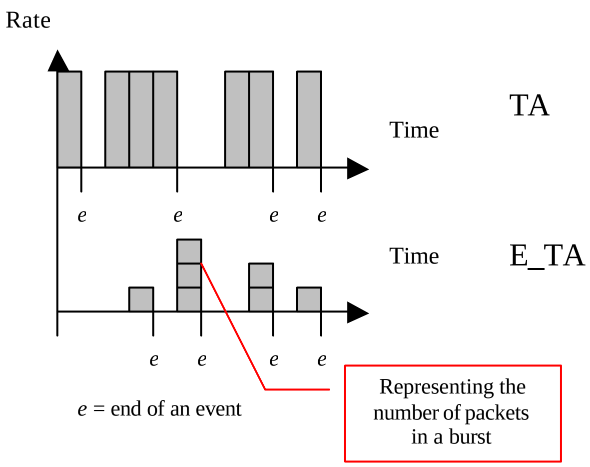

Figure 1: E_TA reduces the number of events

In terms of timeslots, the ON duration in E_TA is 1 unit time but the service time of a burst of packets depends on the number of packets contained in that burst. Hence, instead of deterministic service times, adjustment has to be made to ensure the queue distribution is equivalent. E_TA accelerates simulation by not scheduling events associated specifically with packet-by-packet traffic, instead by incorporating the time that would have been taken in the buffer by a full packets-by-packet simulation into the service process of the traffic we are interested in, the burst-of-packets traffic. It is easy to state this as an operational principle, but more difficult to design algorithms that allow us to implement this idea in a practical fashion, in working simulators. In order for our resulting simulators to be effective and accurate we must fulfil the criteria that they accurately reproduce the queuing behaviour of the original queuing system, i.e. the one that explicitly models the packet-by-packet traffic. This means that these two queuing systems have to produce the same results for certain measures of interest. The measure we choose (because it critically affects both information loss and delay) is the state probabilities as seen by arriving packets. Figure 2 illustrates the idea where, if burst 1 has 2 packets, service time 1, ST_1 equals 2 timeslot, if burst 2 contains 1 packet; ST_2 equals 1 and so on.

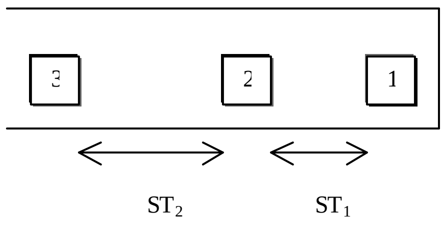

Figure 2: Service time in FIFO queue depends on the number of packets in the burst

3. E_TA with a non-FIFO scheduler

The point of having a non-FIFO queue is to provide differential quality of service (QoS) to the customers. For example real-time applications such as VoIP require very low jitter and a one-way mean delay of 100 milliseconds. Non-real-time applications such as FTP (file transfer protocol) do not suffer from jitter, but are sensitive to packet loss. The non-FIFO queue for the E_TA model is adopted from the Static Priority Full Service (SPFS) scheduler [10], with certain modifications because we are not simulating individual events. Instead, we are applying a hybrid technique on the service time which relies on prior knowledge of applicable queuing analysis.

There are two levels of priority, high priority for real-time applications and low priority for non real-time applications. The buffer will have two sub-queues, where sub-queue 1, sq1, holds high priority packets and sub-queue 2, sq2, holds low priority packets. The busy period distribution of the sq1 is added to the service times of the sq2 traffic, and this is illustrated in figure 3.

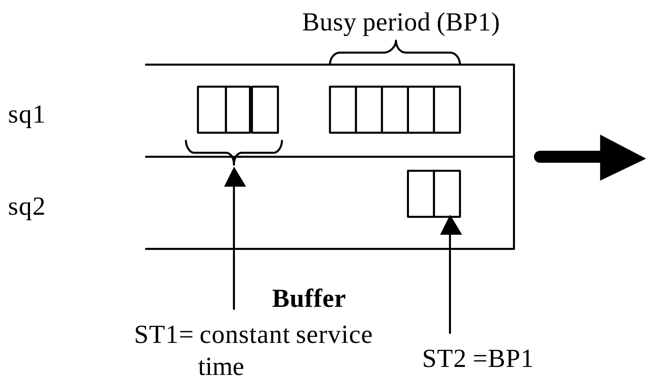

Figure 3: The non-FIFO scheduler

In the non- FIFO queue, some consequences need to be considered a) the fact that the input traffic are in terms of burst b) the adjustment of the service time for both sub queues.

4. Results

The E_TA model with FIFO queue buffer is compare with a TA model for the following scenarios;
Scenario 1:
Mean ON time = 4 unit time
Mean OFF time = 10 unit time
ON arrival rate = 2 unit time
Service rate =1 unit time

Scenario 2:
Mean ON time = 6 unit time
Mean OFF time = 10 unit time
ON arrival rate = 2 unit time
Service rate =1 unit time

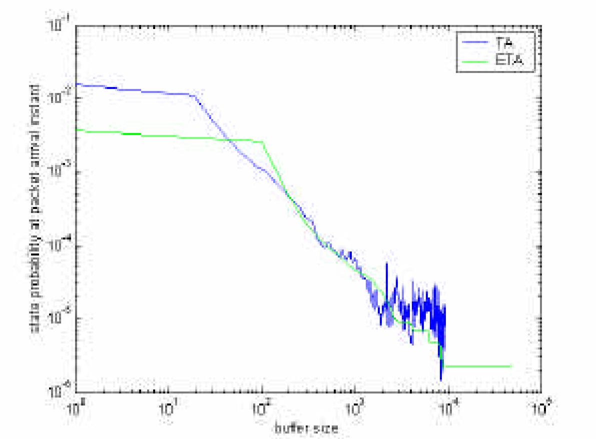

Figure 4: Scenario 1

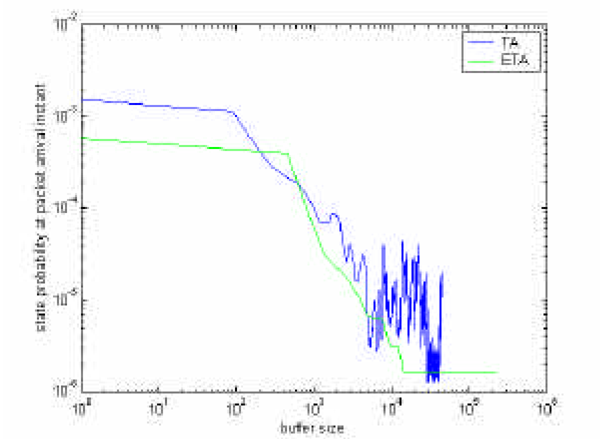

Figure 5: Scenario 2

It was found that over 3 billion timeslots the E_TA method simulates 8 397 736 events, while TA simulated 12 526 707 events. Hence E_TA simulates 32% fewer events than simulated by TA in the traffic example. In general this speedup is a function of the burstiness of the traffic being simulated.

5. Conclusion

The results show that E_TA with a FIFO queue has equivalent traffic accuracy to the TA model and the speed up of simulation in terms of the number of events was significant. FIFO queue is a simple buffer to model but practically a real network supports many types of services that have different priorities. The service rates of each type of service will also varies. Further work we will apply E_TA model to accelerate the simulation of non-FIFO queue, and we anticipate that even more significant speedup will be obtained in these cases.

References

[1] W. Leland, M. Taqqu, W. Willinger and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended version)", IEEE/ ACM Transaction on Networking, February 1994.

[2] V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modelling", Proc. of ACM SIGCOMM, 1994.

[3] M.E. Crovella and M. Taqqu, "Estimating The Heavy Tail Index From Scaling Properties", Methodology and Computing in Applied Probability, Vol. 1, No. 1, 1999.

[4] A. H. I. MA and J. A. Schormans, "A Fast Simulation Method for Modelling IP Networks", 17th UK Teletraffic Symposium on "Networks: A service centred approach", Dublin, Ireland, May 2001

[5] A. H. I. MA and J. A. Schormans.: "Efficient Simulation Modelling for Long Range Dependent Traffic", Fourth Workshop on Rare Event Simulation and related Combinatorial Optimization Problems (RESIM / COP'02), April 11-12, 2002, Madrid, Spain.

[6] A. H. I. MA and J. A. Schormans: "Accelerated simulation modelling of Power-law traffic via aggregation", ITC2002 Beijing, June 2002.

[7] J.M. Pitts and J.A. Schormans, "Introduction to IP and ATM Design and Performance", Wiley, Chischester, 2000.

[8] S. Syed Ariffin and J. A. Schormans, "ON/OFF Pareto Distributed Source At Burst Level: A Study Of Fast Simulation Methods For Packet Network", PGNET2002.

[9] G. Zhang and H. T. Mouftah, "End-To-End Qos Guarantees Over Diffserv Networks", Computers and Communications, 2001. Proceedings. Sixth IEEE Symposium, 2001

[10] A. Narayanan and V.G. Kulkarni, "First Passage Times In Fluid Model With Application To Two Priority Fluid Systems", Proc. of IPDS 96', IEEE Computer Society Press, 1996.