# Accelerated Simulation Method Involving Markovian and Self-Similar Traffic Sources with Non-Fifo Scheduler

Sharifah H. S. Ariffin, John A. Schormans

*Department of Electronic Engineering*
Queen Mary College, University of London, Mile End Road, London, United Kingdom.
sharifah.ariffin@elec.qmul.ac.uk

***Abstract*** **- An efficient accelerated simulation method is essential to obtain fast and accurate results when dealing with self-similar and long range dependent (LRD) traffic sources. In this paper we describe an approach that involves short range dependent (SRD) and LRD traffic sources. The results show accuracy in the approach describe and the original packet-by-packet version. Also, the real time and number of events can be reduced using this method.**

***Keywords - Accelerated simulation, self-similar traffic, non FIFO scheduler***

## I. INTRODUCTION

Traditional stochastic models that exhibits short range dependence (SRD) such as pure Poisson or Markov-Modulated Poisson (MMP) can no longer be apply to capture the burstiness of the modern packet network traffic. Recent studies had led to the conclusion that the Ethernet, ATM traffic, telnet, FTP and variable-bit-rate (VBR) video traffic can be more accurately model by self similar traffic [ Leland et. al.,1994; Paxson and Floyd, 1994; Crovella and Taqqu, 1999; Garret and Willinger, 1994 ]. Some examples of self-similar stochastic process are Fractional Brownian Motion (FBM), Poisson Zeta Process and Pareto process. Network performance analysis, mainly in simulation, of such traffic requires long hours of real clock time due to the long range dependent structure of self similar traffic. More over accurate statistics can only be obtain from a large number of replication. For these reasons, accelerated simulation is needed to simulate events that are rare e.g. buffer overflows.

In Queen Mary, University of London (QMUL) several approaches had been proposed to reduce the amount of time taken to execute a simulation of self-similar and non self-similar traffic, such as a) Hot emulation b) state space reduction method and c) Traffic Aggregation (TA).

Hot emulation technique was a collaborated project of QMUL with Nortel that divided the whole network traffic into background (BG) traffic and foreground (FG) traffic [Schormans et al, 2001]. Since the traffic of interest in the investigation was apart of the traffic, this traffic is classified as FG traffic and the rest as BG traffic. By simulating FG traffic and not the whole detail of the network traffic, the number of events was reduced and proved to save more real time consumed for a simulation experiment.

State space reduction method is a technique to accelerate simulations by aggregating traffic sources [Pitts and Schormans, 2000]. In this technique N number of ON and OFF sources with 2N state space process is reduce to just 2 state space processes. The aggregated process of ON state is when input rate exceeds the output rate while the OFF state is when the input rate is not zero but less than the output rate.

The state space reduction method had proved to be highly accurate for Markovian traffic and it was further developed for self similar traffic called Traffic Aggregation method (TA) [ Ma et al, 2003; Ma and Schormans, 2002a; Ma and Schormans, 2002b; Ma and Schormans 2002c; Ma and Schormans 2001]. This is an accelerated simulation method that uses First-In-First-out (FIFO) buffer scheduler and proved to save 75% of the usual time taken for conventional simulation. It has also proved to reduce time taken with an accurate queuing distribution in the aggregated buffer for the end-to-end model.

Despite of performance network prediction for practical application such as bursty traffic, Differential services e.g. Diffserv., will make network performance design more applicable to the real network. Hence in this paper we describe an accelerated simulation approach based on TA method called Enhanced TA (E_TA). E_TA involves self similar packet-train traffic with a buffer that supports heterogeneous traffic.

The paper is organized as follows: we present the E_TA accelerated simulation approach in section 2. Section 3 shows the simulation experiment results and finally the conclusion and discussion in the section 4.

## II. THE ACCELERATED SIMULATION METHOD

Packet train source means that a burst of arrivals in the E_TA method represents many packet-by-packet events in the TA model. In other words, the ON duration is compressed to 1 unit time instead of >1 unit time, see figure 1 below.

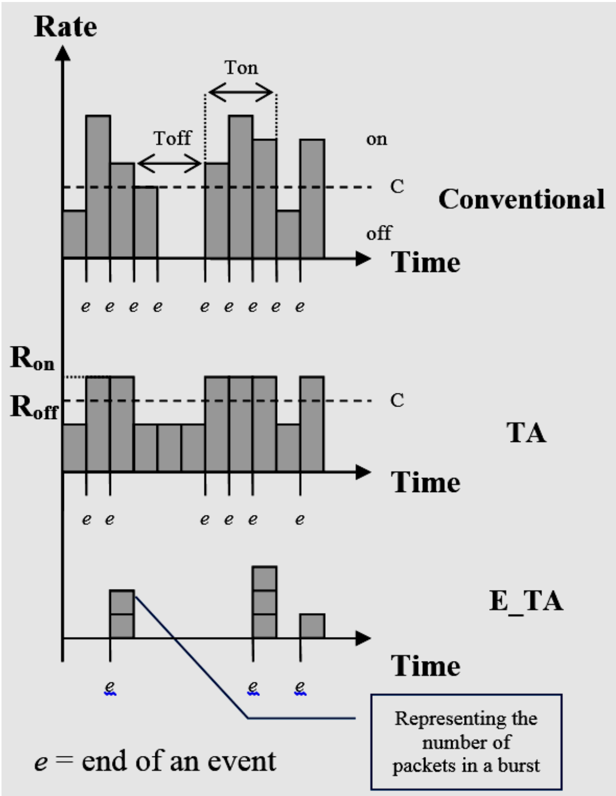


Figure 1. The number of events reduced using E_TA

However in order to get equivalent traffic accuracy in the buffer it requires a few adjustments. In the initial stage of the experiment, a simple FIFO scheduler is use in the buffer. The potentially long active periods in the arriving packets in a burst is filled by the number of packet arrival rate during the ON period. The adjustment was not only done to the ON period but also to the OFF period where the mean OFF time for E_TA is given by:

$$mean_{OFF} = \left[\left(1-\rho \middle/ R_{ON}\right) \times R_{ON} \middle/ \rho\right] \cdot C \quad (1)$$

where $\rho$ is the load, RON is the ON arrival rate, C is the buffer capacity and all packets are assumed to be the same size. This is because in E_TA equivalent traffic distribution in the buffer was only match when the packet arrival rate in the mean ON period was taken into account (i.e. to fill the buffer) followed by a reduction of mean OFF period (i.e. emptying the buffer. Thus, the service rate of each burst in the FIFO queue depends on the arrival packet rate in the mean ON period.

For differential services, a non- FIFO scheduler is use to further developed E_TA method. The original packet-by-packet version of the non FIFO scheduler supports two levels of priority, high priority for real-time applications and low priority for non real-time applications. The buffer with a non FIFO scheduler will have two sub-queues. Sub-queue 1, sq1, holds high priority packets and sub-queue 2, sq2, holds low priority packets. The service time distribution of sq2 equals the busy period distribution of sq1 and this is illustrated in figure 2 below.

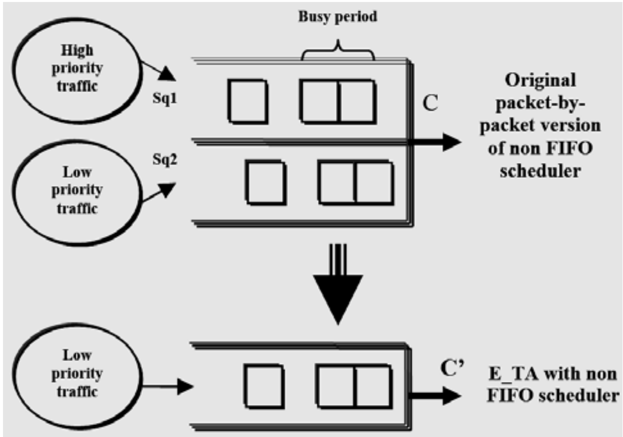


Figure 2. The elimination of sq1 in E_TA accelerated simulation technique

However, the E_TA accelerated simulation model with non FIFO scheduler eliminates sq1 and only simulated sq2 with low priority source. The simulation relies on the prior queue analysis of sq1. The burst of packets in sq2 in E_TA with non FIFO scheduler has to be carefully served as the service rate of each packet equals to the mean busy period distribution of sq1. Hence the service rate of each burst in E_TA with non FIFO scheduler is given by

$$P_{ST} = \sum_{n=1}^{P_{\max}} ST_n \quad (2)$$

where ST is the service time and Pmax is the number of packets in a burst.

The issue of whether or not sq1 that holds high priority packets will starve packets in sq2 is called packet pre-emption. If we say that this non FIFO scheduler gives good performance to high priority packets, in the real network this will significantly degrade the performance of the low priority packets. For example at current time sq1 is empty, the buffer will choose sq2 and serve low priority packets. Halfway through the service, high priority packets arrive. The scheduler will have two choices, either to delay the high priority packets for the low priority packets or to abort the service of the low priority packets in sq2 and immediately switch to sq1. In the former case, the scheduling is called non pre-emptive and the latter case is called pre-emptive. In this experiment, non pre-emptive scheduling is use because the extra delay experienced by the high priority packets is small compared to the delay experienced by the low priority packets in pre-emptive scheduling.

The mean busy period distribution of sq1 can not just be heavy-tailed (because of the self similar traffic) but also exponential. Because exponential traffic is less bursty, the mean busy period distribution of sq1 (whether it is exponentially distributed or heavy-tailed distributed) will not affect the queuing distribution in sq2 if the low priority traffic source is self similar. This is shown in section 3. Hence, if the service rate of sq1 = C1 and service rate of sq2 = C2, the overall service rate in the non FIFO scheduler is C. For E_TA with non FIFO scheduler the overall service rate will be

$$C' = C(1 - C1) \tag{3}$$

## III. EXPERIMENTAL RESULTS

The queue distribution of E_TA with FIFO scheduler compared to the TA model is shown in figure 3 for the following scenario: mean Ton = 4, mean Toff = 10 with rate in the ON period, R = 2 and service rate, C = 1. This graph shows the accuracy of the initial experiment in the accelerated simulation approach compared to the packet-by-packet version.

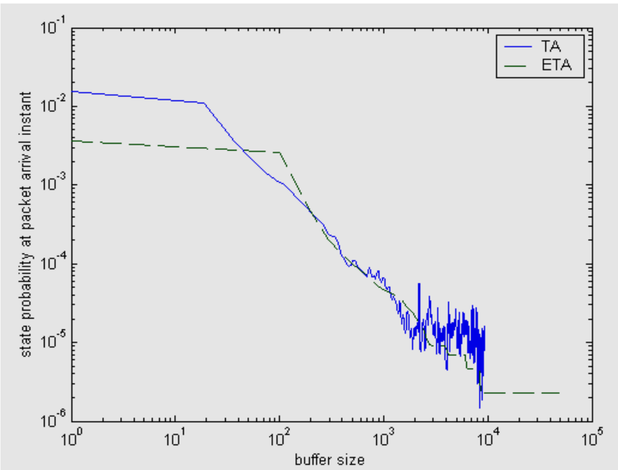


Figure 3: The queuing distribution in the buffer for E_TA with FIFO scheduler (ETA) and Traffic Aggregation model (TA)

Figure 4 shows the queue distribution of E_TA with non FIFO scheduler compared to the original packet-by-packet version of non FIFO scheduler with mean Ton = 8, mean Toff =3 with rate in the ON period, R = 2 and service rate, C = 1. Plotting x on a linear scale and y on the log scale, the result shows accuracy in the queuing distribution of sq2.

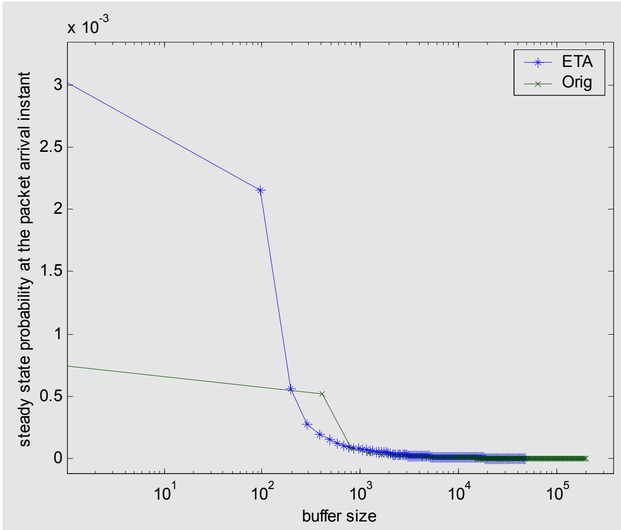


Figure 4: The queuing distribution in sq2 of the non FIFO scheduler for E_TA with non FIFO scheduler (ETA) and original packet-by-packet model (orig)

The essential factors in accelerated simulation method are the number of events reduced and how much real time saved in any simulation experiments. Hence, in figure 5 and 6 experiments of the following scenarios is done with different loads.

i) E_TA with FIFO scheduler compared to TA : mean Ton = 4, mean Toff =10 with rate in the ON period, R = 2 and service rate, C = 1.

ii) E_TA with non FIFO scheduler compared to original packet-by-packet method: mean Ton = 8, mean Toff =3 with rate in the ON period, R = 2 and service rate, C = 1

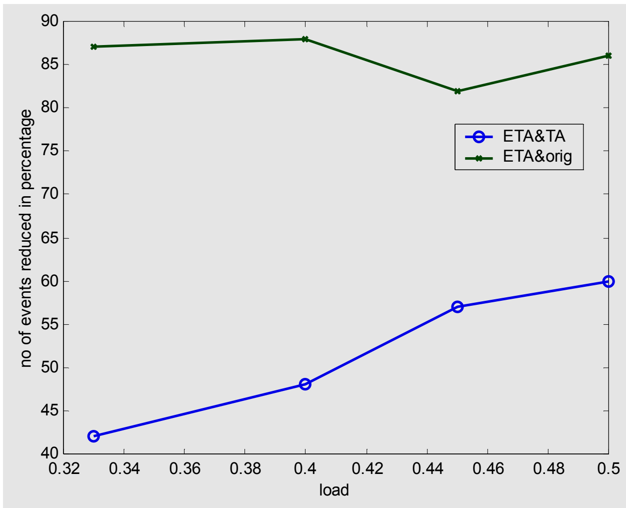


Figure 5: Number of events reduced in percentage for E_TA with FIFO scheduler compared TA and E_TA with non FIFO scheduler compared to original packet-by-packet

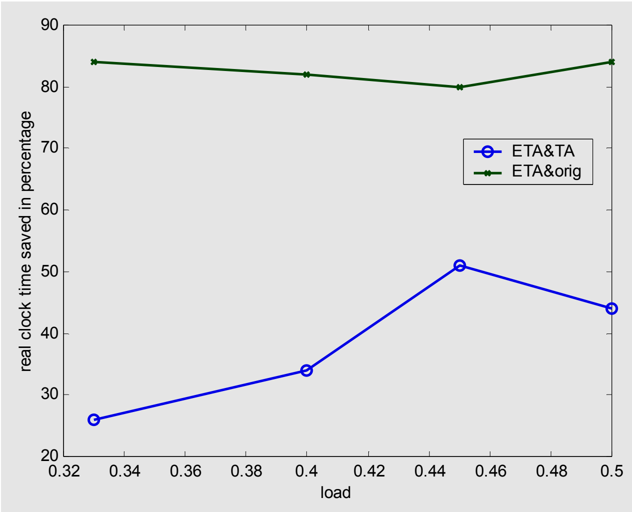


Figure 6: Real simulation clock time saved in percentage for E_TA with FIFO scheduler compared TA and E_TA with non FIFO scheduler compared to original packet-by-packet

The number of events and real clock time saved in percentage had been obtained in both methods. Even though the case i) did not give very much reduction, the number of events and real clock time keep on reduce as the load increases. As for the case ii), the number of events and real time saved are mainly maintained.

The mean busy period distributions of sq1 is shown in figure 7 for exponential distributed and Pareto distributed ON/OFF traffic. The mean On , Ton = 8 units time, mean Off , Toff = 3 units time, ON arrival rate, R = 2 and service rate, C = 1. Using this analysis, both mean busy period distributions are use as the service rate of each packet in sq2 with self-similar low priority traffic and figure 8 is obtained. It shows that even though Markovian traffic is apply at sq1, the power law queuing distribution dominates in sq2.

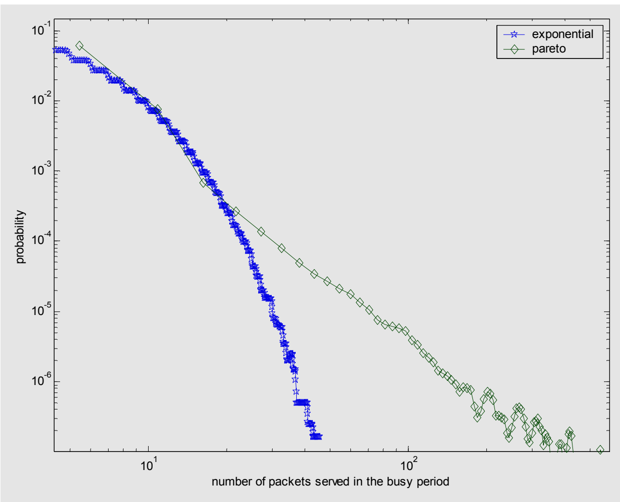


Figure 7: The mean busy period distribution in sq1 for Markovian traffic and self-similar traffic

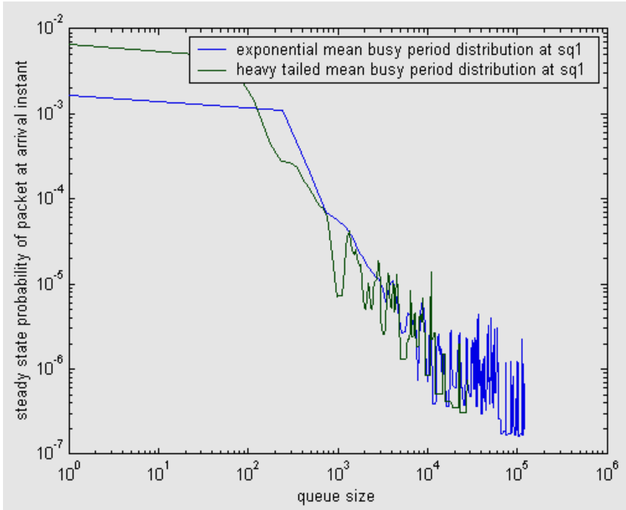


Figure 8: Power law decay in sq2

## IV. CONCLUSIONS AND DISCUSSIONS

From the experiments, it shows that accelerated simulation methods of E_TA with FIFO scheduler and non FIFO scheduler approach is highly accurate compared to the TA and the original packet-by-packet model. It not only reduced the number of events and time spends for a simulation to reach a steady state but this approach can be use for both homogeneous and heterogeneous traffic sources. Moreover from the experiments it was found that high priority traffic can be either exponentially distributed or Pareto distributed ON/OFF traffic because in sq2 heavy tailed distribution dominates the queuing distribution.

## REFERENCES

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

[2] Paxson V. and Floyd S.1994, “Wide-Area Traffic: The Failure of Poisson Modelling”, In the Proceedings of ACM SIGCOMM.

[3] Crovella M. E. and Taqqu M. 1999, “Estimating The Heavy Tail Index From Scaling Properties”, Methodology and Computing in Applied Probability, Vol. 1, No. 1, Available from http://www.cs.bu.edu/faculty/crovella/papers.html.

[4] Garret M. and Willinger W., 1994, “Analysis Modelling and Generation of Self Similar VBR Video Traffic”, In the Proceedings of ACM SIGCOMM.

[5] Schormans J., Liu E., Cuthbert L. and Pitts J.,2001, “A Hybrid Technique for the Accelerated Simulation of ATM Networks and Network Elements”, ACM Transaction on Modelling and Computer (TOMACS) Vol. 11, Issue 2, April.

[6] Pitts J.M. and Schormans J.A.2000, “Introduction to IP and ATM Design and Performance: With Applications and Analysis Software”, Wiley, Chichester.

[7] Ma A. H. I. and Schormans J. A. 2001, "A Fast Simulation Method For Modelling IP Networks", 17th UK Teletraffic Symposium on "Networks: A Service Centred Approach", Dublin, Ireland, May.

[8] Ma A. H. I. and Schormans J. A. 2002a, “Accelerated Simulation Modelling of Power-law Traffic via Aggregation”, ITC2002 Beijing, June.

[9] Ma A. H. I. and Schormans J. A. 2002b, "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, Madrid, Spain.

[10] Ma A. H. I. and Schormans J. A. 2002c, "Hybrid Technique for Analysis of Multiplexed Power-Law Traffic in Broadband Networks", IEE Electronic Letters, Vol.38, No. 6, March.

[11] Ma A. and Schormans J. A. and Cuthbert L. 2003, "Aggregation Technique for Networks with Power Law Traffic and Application to accelerated Simulation", IEE Proc. Comm., Vol. 150, No. 3, June.

*Biography of the First Author*

The author received the BEng degree in Electronics and Communications from the London Metropolitan University, formally known as university of North London) in 1997, and the MEE degree in Telecommunications from University Technology Malaysia, Malaysia, in 2000. She is currently a PhD student in the Department of Electronics Engineering in Queens Mary College, University of London. Here current research interests are traffic queueing, simulation speed and performance evaluation.

.