Self-Similar Traffic and Its Implications for ATM Network Design

Changcheng Huang, Michael Devetsikiotis, Ioannis Lambadaris and A. Roger Kaye

Department of Systems & Computer Engineering
Carleton University
Ottawa, Ontario K1S 5B6, Canada

Tel: (613) 788-5740, Fax: (613) 788-5727, email: huang@sce.carleton.ca

ABSTRACT *Self-similar* (or *fractal*) stochastic processes have been proposed as more accurate models of certain categories of traffic (e.g., Ethernet traffic, variable-bit-rate video) which will be transported in ATM networks. Self-similar processes exhibit *long range dependence* structure which is not the case for traditional models. The distinct differences between these two classes of models, have significant implications for performance prediction and network design.

In this paper, we describe simulation results using both synthetic self-similar processes and empirical video traces. Based on these simulation results, we analyze certain existing congestion control schemes and show that, although these schemes may be promising under traditional models, they face serious challenge under self-similar models.

I. INTRODUCTION

Several recent papers [1, 2, 3, 4, 5] have shown that traditional traffic models may be inadequate for modeling real traffic to be carried on ATM networks. Studies have reported that LAN traffic [1], WAN traffic [4] and variable bit rate (VBR) video traffic [2] [3] often display *long range dependence* (LRD) and can be better modeled by self-similar processes. Furthermore, it has been shown in [5] and [3] that, in addition to LRD structure, VBR video traffic also possesses strong *short range dependence* (SRD) structure which suggests an asymptotic self-similar model rather than exact self-similar model. Apart from Poisson distribution, marginal distributions with heavy tails were reported in [3] and [5].

A stochastic process is said to exibit LRD when it has a nonsummable autocorrelation function [6]. Self-similar processes (both exact and asymptotic) are among those LRD processes which are widely used. Traditional traffic models, on the other hand, typically possess some form of Markovian structure and display a SRD structure only.

Due to the distinct differences between these two classes of models, their implications for network design and performance estimation will be significantly different. In [7], a large deviation approach has been successfully applied to a single server queue with FGN as input. It was shown that the queueing tail distribution decays in according to a Weibull function with increasing buffer size rather than exponentially as in traditional models. Similar results were shown in [8] [9]. In [10] and [5], a fast simulation approach which can efficiently simulate the transient queue occupation distribution with input processes possessing both LRD and SRD was developed. Furthermore, this approach can accommodate arbitrary marginal distribution. Simulation results were compared with real traffic results, and a close match was achieved. These results have shown that performance estimates based on traditional models are far more optimistic. Consequently, studies related to the design and performance evaluation of an ATM network based on traditional traffic models must be revisited in the context of these new models.

In this paper, based on the simulation results in [10] and [5], we investigate some serious implications for ATM network design introduced by LRD structure. We show that, the buffering gain of self-similar traffic is much lower than that of traditional models. This means that, for certain applications, peak rate allocation may be necessary for access links. While peak rate allocation may be unavoidable at the access links, we show that peak rate allocation is not necessary within networks due to the significant multiplexing gain of self-similar traffic streams. But this conclusion must be applied cautiously. We show that multiplexing heterogeneous self-similar traffics, the traffic with highest Hurst parameter will dominate the QoS (Quality of Service). This means that, traffic streams with lower Hurst parameter values may suffer the same mean delay as traffic streams with higher Hurst parameter. In its extreme case, starvation problem may be introduced. This kind of problem can not be solved by traditional priority strategies. As pointed out in [4], for self-similar traffic, high priority traffic may block low priority traffic for quite a long time making it enter into starvation again. Based on these observations, we further investigate some congestion control schemes which may seem promising under traditional mod-

0-7803-2916-3

els. We show that, although some of the problems mentioned above can be solved by introducing fair queueing approach, other problems remain to be addressed.

This paper is organized as follows. In section II, we make a brief introduction to self-similar traffic models and a result of modeling VBR video traffic using both SRD and LRD components. In section III, we describe simulation results based on self-similar processes. Based on these simulation results, we investigate congestion control issues posed by self-similar processes in section IV. In section V, we summarize our conclusions.

II. SELF-SIMILAR TRAFFIC MODELS

A. DEFINITION OF THE FGN PROCESS

While there are numerous stochastic models which exhibit the self-similar property, two of them, namely the exactly self-similar *fractional Gaussian noise* (FGN) and the asymptotically self-similar *fractional autoregressive integrated moving-average* (F-ARIMA) process, are the most commonly used. FGN can be viewed as a reasonable first approximation of more complex LRD processes, since it can be derived from a special type of central limit theorem applied to LRD processes.

A fractional Gaussian noise process $\mathbf{X} = \{X_k : k = 1, 2, \ldots\}$ is a stationary Gaussian process with mean $m = E[X_k]$, variance $\sigma^2 = E[(X_k - m)^2]$, and autocorrelation function

$$r(k) = 1/2(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}), \; k = 1, 2, 3, \ldots \tag{1}$$

Therefore, if $1/2 < H < 1$, FGN is exactly second-order self-similar with Hurst parameter H.

Definitions of self-similar processes in a more general sense can be found in [11]. Intuitively, one of the most striking features of such processes is that their aggregated processes possess a nondegenerate correlation structure as $n \to \infty$.

B. SELF-SIMILAR MODELING VBR VIDEO

Earlier efforts in modeling video traffic have been confined to short traces of empirical records or to conference video, due to the difficulties in obtaining empirical data from realistically long sequences.

Recent extensive measurements of real traffic data [2], have led to the conclusion that VBR video traffic cannot be sufficiently represented by traditional models, but instead can be more accurately matched by self-similar models. In [3] the authors presented a detailed statistical analysis of a 2-hour long empirical VBR video trace ("Star Wars"). It was found that, VBR video traces possess both LRD and SRD structures. The authors estimated the Hurst parameter of the empirical stream, modeled the marginal distribution of the video "bandwidth" (i.e., number of bits per video frame or slice) with a combined Gamma/Pareto distribution, and generated synthetic traces by appropriately transforming a fractional ARIMA$(0, d, 0)$ process [12] that provided the LRD behavior. However, explicit modeling of the SRD structure was left for future work.

In [5], we developed a unified approach which, in addition to accurately modeling the marginal distribution of empirical video records, also models directly both the short and the long term empirical autocorrelation structures. In Fig. 1 and Fig. 2, we compare the autocorrelation functions and q-q plots for marginal distributions of empirical data trace and synthetic trace generated by our built model in [5]. Close matches are achieved.

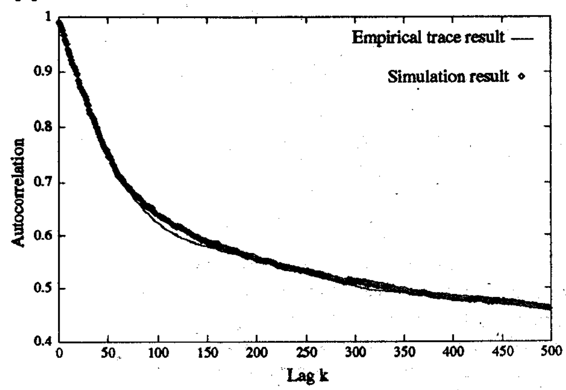

Figure 1: Autocorrelation of the empirical trace and the synthetic process.

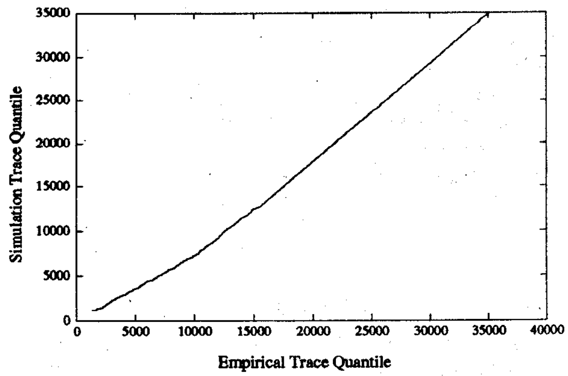

Figure 2: q-q plot for the marginal distributions of the empirical trace and the synthetic process.

III. SIMULATION RESULTS

In this section, we show our simulation results for a FCFS single server queue with deterministic service rate μ. We consider two classes of sources. One is FGN processes representing general self-similar processes, the other is the model we built in [5] and showed in last section representing a specific VBR video model.

A. FGN CASE

In Fig. 3, we show the following results:

(1) one FGN source with Hurst parameter $H = 0.7$; (2) one FGN source with Hurst parameter $H = 0.9$; (3) two FGN sources with $H = 0.7$ and $H = 0.9$ respectively; (4) a large deviation result for (3).

From this figure we can see that, the buffering gain with increasing buffer size is much smaller for self-similar sources than for traditional models. Furthermore when multiplexing two self-similar processes, the one with larger Hurst parameter will dominate the queueing tail distribution although multiplexing gain exists for both sources.

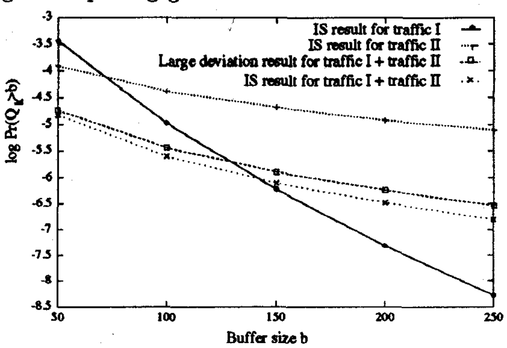

Figure 3: Estimated $\log \Pr(Q_k > b)$ versus the buffer size b (heterogeneous sources, one with $H = 0.7$, the other with $H = 0.9$). Each simulation is based on 1000 iid replications.

In Fig. 4, we show that when multiplexing several sources with the same Hurst parameters, significant multiplexing gain is achieved.

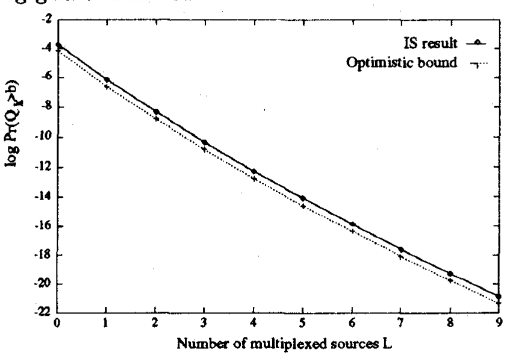

Figure 4: Estimated $\log \Pr(Q_k > b)$ versus the number of multiplexed sources L. The Hurst parameter is $H = 0.9$.

For more about these simulation results, see [10].

B. REALISTIC VIDEO CASE

In Fig. 5, we show the comparison of synthetic video trace and empirical video trace. Consistent with the FGN case, the buffering gain is also very small with increasing buffer size.

For more about this simulation result, see [5].

IV. NETWORK DESIGN ISSUES

One of the major advantages of packet switching networks is to multiplex different traffic streams asynchronously to achieve multiplexing gain. Although significant multiplexing gain can be achieved for self-similar traffics as shown in Fig. 3 and Fig. 4, the burstier one will dominate the queueing tail distribution as shown in Fig. 3. This means that, traffic with lower Hurst parameter may suffer the same mean delay as traffic with higher Hurst parameter. In

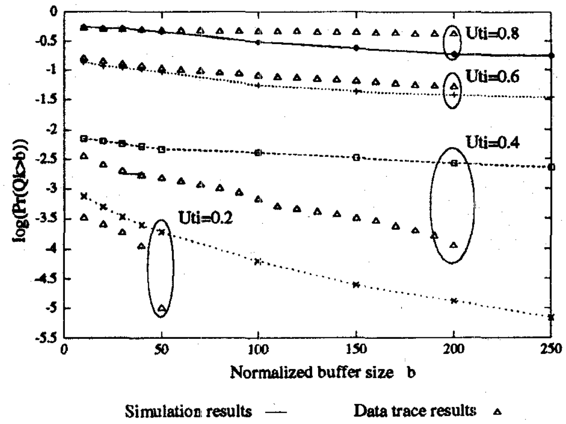

Figure 5: Overflow probability versus buffer size b, for different utilization values, using 1000 replications and $k = 10b$.

its extreme case, a starvation problem may be introduced. This kind of problem can not be solved by traditional priority strategies. As pointed out in [4], for self-similar traffic, high priority traffic may block low priority traffic for quite a long time making it enter into starvation again.

The above mentioned problem exists under traditional models although with less severe probabilities. People have found that it is caused by the interference between different traffics within a FCFS queue. Traditional approach to solving this problem is to introduce fair queueing schemes. Several rate-based service disciplines have been proposed to provide firewall protection among individual traffic flows: Delay-EDD [13], VirtualClock [14], and PGPS [15, 16]. The common features of these approaches are:

(1) Provide bounded end-to-end delay for each individual traffic within the whole network.

(2) Packets from different flows are maximally interleaved to obtain the multiplexing gain.

(3) Maximum throughput is preserved due to their work preserving nature.

Other approaches attempt to solve the propagation of delay jitter existing in the approaches above and further decouple the interference between nodes. Namely they are: Jitter-EDD [17], RCSP [18], Stop-and-Go [19] and Hierarchical Round Robin [20].

The above approaches are very promising under traditional models because they provide a congestion-free and topology-free solution to the high speed packet switch network. Unfortunately, they all require some burstiness constraints [21] at the access node. A well-known such constraint is the so-called (σ, ρ) constraint which can be implemented by a leaky bucket. As shown in Fig. 6, a leaky bucket can be modeled by a virtual queue. Queueing results in Fig. 3 and Fig. 6 can be directly applied to estimate loss probability at the access node regulated by a leaky bucket. The result is that, although the user can be guaranteed loss-free transport within network using the approaches above, the user may suffer heavy loss at the access node. Without raising the bucket rate, the loss may become intolerable.

But increasing the buffer rate results in a loose control or no control at all over the burtiness of the user traffic. This dilemma poses significant challenge to various congestion control schemes.

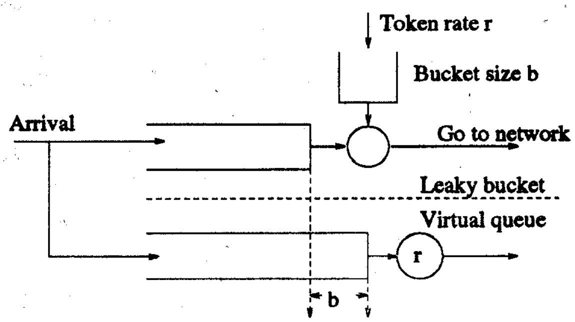

Figure 6: Equivalence of leaky bucket and virtual queue in terms of loss rate.

V. CONCLUSIONS

Self-similar traffic models are regarded as more accurate for modeling certain traffic types. Due to their extreme burstiness, self-similar models pose significant challenge to existing congestion control schemes. In this paper, we have shown that, without certain changes, traditional congestion control algorithms may fail to achieve their goal.

VI. ACKNOWLEDGMENTS

This research was supported by grants from the Telecommunications Research Institute of Ontario.

References

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the Self-Similar Nature of Ethernet Traffic (Extended Version). *ACM/IEEE Transactions on Networking*, 2(1):1–15, Feb. 1994.

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-Range Dependence in Variable-Bit-Rate Video Traffic. To appear on *IEEE Transactions on Communications*, 1994.

[3] M. W. Garrett and W. Willinger. Analysis, Modeling and Generation of Self-Similar VBR Video Traffic. In *Proc. ACM SIGCOMM '94*, London, U. K., Aug. 1994.

[4] Vern Paxson and Sally Floyd. Wide Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[5] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Modeling and Simulation of Self-similar VBR Compressed Video:A Unified Approach. In *Proc. ACM SIGCOMM'95*, Boston, USA, August 1995.

[6] D.R. Cox. Long-Range Dependence: A Review. In H. A. David and H. T. David, editors, *Statistics: An appraisal.* The Iowa State University Press, 1984.

[7] N. G. Duffield and N. O'Connell. Large Deviations and Overflow Probabilities for the General Single-Server queue, with Appplications. *DIAS-STP-93-30*, 1993.

[8] I. Norros. A Storage Model with Self-Similar Input. *Queueing Systems*, 16:387 – 396, 1994.

[9] Nol Rananand and Prakash Narayan. Upper-bounds on the Performance of a Queue with Long Range Dependent Input. *Preprint*, 1994.

[10] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Fast Simulation for Self-similar Traffic in ATM Network. In *Proc. IEEE ICC'95*, Seattle, USA, June 1995.

[11] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437, 1968.

[12] J. R. M. Hosking. Fractional Differencing. *Biometrika*, 68(1):165–176, 1981.

[13] D. Ferrari and D. Verma. A Scheme for Real-Time Channel Establishment in Wide-Area Networks. *IEEE J. Select. Areas Commun.*, 8(4):368–379, April 1990.

[14] L. Zhang. VirtualClock: A New Traffic Control Algorithm for Packet Switching Networks. *ACM Transactions on Computer Systems*, 9(2):101–124, May 1991.

[15] A. K. Parekh and G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks–The Single Node Case. In *Proc. IEEE INFOCOM '92*, pages 915–924, 1992.

[16] A. K. Parekh and G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks–The Multiple Node Case. In *Proc. IEEE INFOCOM '93*, pages 521–530, 1993.

[17] D. Verma, H. Zhang, and D. Ferrari. Delay Jitter Control for Real-Time Communication in a Packet Switch Network. In *Proc. IEEE TriCOM '91*, pages 35–43, 1991.

[18] H. Zhang and D. Ferrari. Rate-Controlled Static-Priority Queueing. In *Proc. IEEE INFOCOM '93*, pages 227–236, 1992.

[19] S. J. Golestani. Duration-Limited Statistical Multiplexing of Delay-Sensitive Traffic in Packet Networks. In *Proc. IEEE INFOCOM '91*, pages 323–332, 1991.

[20] C. Kalmanek, H. Kanakia, and S. Keshav. Rate Controlled Servers for Very High-Speed Networks. In *Proc. IEEE GlobeCom '90*, pages 300.3.1–300.3.9, 1990.

[21] R. L. Cruz. A Calculus for Network Delay, Part I: Network Elements in Isolation. *IEEE Trans. Inform. Theory*, 37(1):114–131, Jan. 1991.