Performance Analysis of a Single Server Queue Loaded by Long Range Dependent Input Traffic

R.G. Garroppo, S. Giordano, M. Pagano, G. Procissi, F. Russo
Department of Information Engineering - University of Pisa
Via Diotisalvi 2 I-56126 Pisa - Italy
tel.+39 50 568539 fax +39 50 568522
e-mail: {garroppo@, giordano@, pagano@, procissi@edison., russo@}iet.unipi.it

Abstract

Recent studies on broadband networks show that real traffics exhibit Long Range Dependence (LRD). Because of its dramatic consequences on queueing performances this property has to be taken into account in network design and management. To these aims we analysed the behaviour of a single server queueing system loaded by an LRD arrival process modelled by fractional Gaussian noise. Our goal was to obtain an analytical estimation of the complementary probability, that represents an upper bound for Cell Loss Ratio. The approach is based on the Large Deviations Theory and we successfully validated the results by means of simulation studies.

Keywords

Long Range Dependence, Large Deviations Theory, Queueing Performances

1 INTRODUCTION

The LRD feature represents a relevant statistical memory property of traffic flows in LANs, MANs, WANs, D-channel signaling in ISDN networks and VBR coded video sequences (Willinger, 1996). This characterising behaviour implies an hyperbolic decay of the traffic data autocorrelation that cannot be parsimoniously captured using traditional Markov processes.

The LRD correlation structure is related to the high burstiness presented by multimedia traffic. When investigating the temporal behaviour of traffic patterns, the eyeballing evidence of LRD is the presence at each time scale of clusters of arrivals, called *bursts*. Hence indexes of burstiness based on the *maximum burst length* concept have to be replaced by new metrics able to characterise the traffic behaviour at the different time scales. In this framework a promising approach is the introduction of the Hurst parameter H, which expresses the degree of self-similarity and represents a measure of LRD and therefore of the related burstiness: the greater is the burstiness of the traffic, the higher is its Hurst parameter.

In the future telecommunications scenario a major role will be played by broadband networks, based on ATM architecture. A key characteristic of these networks is the provision of Quality of Service (QoS) requirements. As a consequence it is particularly important to obtain some quantitative results on network performances in terms of cell loss ratio, maximum cell delay, cell delay variation, delay jitter etc.. The estimation of these parameters involves the study of elementary queueing systems, which represent simplified models of single network elements. The mentioned long memory property has a deep impact on queueing performances because of the bursty nature of the arrival process. Moreover, traffic models capturing LRD features makes unfeasible the analytical study of even very simple queueing systems.

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35398-2_19

G. de Marchis et al. (eds.), *Optical Networks: Design and Modelling*

In this paper we analysed an infinite buffer single server queue to estimate the complementary probability of buffer occupancy, that gives many useful insights on QoS parameters. In fact it is widely used as an upper bound for the estimation of the overflow probability for finite buffer systems. Typical loss probability requirements for broadband networks are around 10^{-6}-10^{-8}: to estimate these target values a simulation based approach becomes impassable, since it requires unacceptable long running times. To overcome these difficulties, two different techniques have been considered: rare event simulation procedures (Giordano, 1998) and mathematical approaches based on Large Deviations Theory (LDT).
In this work we have considered the latter philosophy, which leads to analytical results very useful for the design of connection admission control schemes. The paper is organised as follows: in section 2 we present an overview on LDT introducing basic concepts such as the rate function and the Large Deviations Principle. In section 3 we show some theoretical results achieved by the application of LDT to fractal queueing networks, while in section 4 we validate the analytical results concerning the overflow probability by means of discrete event simulations. In the Conclusions Section we highlight the efficiency of our approach and further research directions.

2 NOTES ON LARGE DEVIATIONS THEORY

In this section we introduce some concepts of LDT which are necessary in the paper (for a more complete analysis see (Dembo, 1993), (Shwartz, 1995) and (Bucklew, 1990)). The function $I: R \to (-\infty,+\infty]$ is assumed to be a *rate function* if it satisfies the following conditions:

- I is non-negative;
- I is lower semi-continuos;
- its *level sets* are compact (*good rate function).*

Moreover, the pair $\left\{\frac{W_t}{a_t}, v_t\right\}$ satisfies a *Large Deviations Principle (LDP)* with rate function I if the following conditions hold:

- For every closed subset $F \subset R$:

$$\limsup_{t\to\infty} \frac{1}{v_t} \log P\left(\frac{W_t}{a_t} \in F\right) \le -\inf_{x\in F} I(x) \tag{1}$$

- For every open subset $G \subset R$:

$$\liminf_{t\to\infty} \frac{1}{v_t} \log P\left(\frac{W_t}{a_t} \in G\right) \ge -\inf_{x\in G} I(x). \tag{2}$$

Roughly speaking, if the pair $\left\{\frac{W_t}{a_t}, v_t\right\}$ satisfies an LDP then

$$P\left(\frac{W_t}{a_t} > x\right) \approx e^{-v_t I(x)}.$$

The previous expression justifies the name *rate function* given to *I*: in fact it establishes the decay rate for the tail of the mentioned random variable distribution.

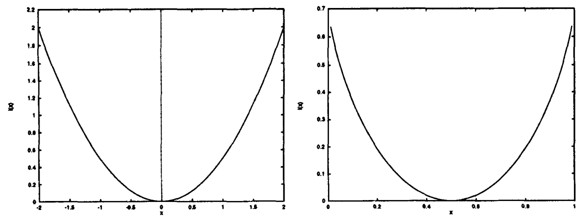

Figure 2. Rate functions for Gaussian standard (left) and Bernoulli (right) random variables.

For example, the sequence of empirical means $\{M_n\}$

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

of i.i.d. random variables X_n satisfies a LDP (see *Cramér Theorem* in reference (Dembo, 1993)) with good rate function depending on the X_n distribution. In figures 1 e 2 we show the plots of *I*-functions for Gaussian, Bernoulli, Poisson and Exponential random variables; as can be seen *I* is usually convex.

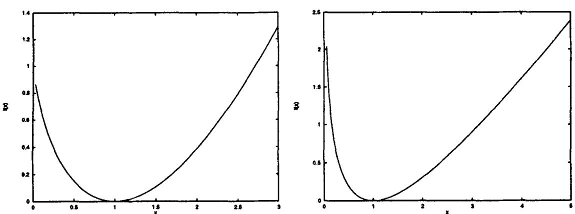

Figure 3. Rate functions for Poisson (left) and Exponential (right) random variables.

3 QUEUEING PERFORMANCES WITH LDT

A widely used queue model in ATM environment is represented by a single server queue with deterministic service rate C (because of the fixed-length of information units, i.e. cells).

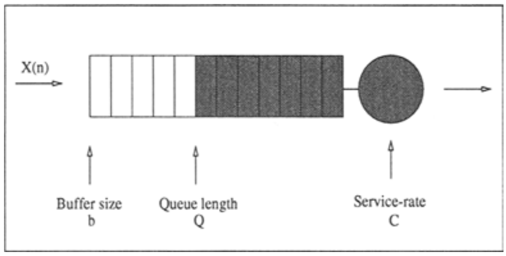

Figure 4. Reference queueing model.

In figure 3, we show the relevant parameters of the actual queueing model. In order to obtain an estimation of an upper bound for the overflow probability when the queue is fed by LRD traffic processes, we applied LDT to a simplified version of that model, characterised by an infinite buffer.
In the analysis we considered a slotted time *fractal queueing system*, whose input process X_t represents the number of arrivals in the time interval $((t-1)T_u, tT_u]$, where T_u is the time unit (assumed equal to one) and $t \in Z$

The arrival process is modelled as follows:

$$X_t = m + G_t$$

where G_t is the so-called *fractional Gaussian Noise (fGn)* and m represents the mean arrival rate. The *fGn* is a discrete-time, stationary, Gaussian, exactly second-order self-similar (and so LRD) process with zero mean, Hurst parameter H and autocorrelation function:

$$r(k) = \frac{1}{2}\left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right).$$

The associated workload process, representing the difference between the total amount of work offered to the queue in the interval *(0, t]* and the maximum number of served arrivals in the same time period, is given by:

$$W_t = B_t - (C - m) \cdot t$$

where B_t is the sample of a *fractional Brownian motion* process at time tT_w, which is related to G_t by relation $G_t = B_t - B_{t-1}$.
Denoting with Q the asymptotic queue length and considering the assumed weakly stationarity of arrival and service processes, from the Lindley equation (Duffield, 1994) it follows:

$$Q = \sup_t W_t$$

Substituting a_t and v_t with t and t^2 / σ_t^2 (where σ_t^2 is the variance of the workload process) respectively in relations (1) and (2), it was proved (Duffield, 1994) that the pair $\left\{\frac{W_t}{t}, \frac{t^2}{\sigma_t^2}\right\}$ satisfies a LDP with good rate function given by the *Legendre-Fenchel* transform

$$I(x) = \sup_\theta \{\theta x - \lambda(\theta)\}$$

of the *scaled cumulant generating function*:

$$\lambda(\theta) = \lim_{t\to\infty} \frac{\sigma_t^2}{t^2} \log E e^{\frac{t^2}{\sigma_t^2}\theta\frac{W_t}{t}}.$$

Moreover in (Duffield, 1994) the following fundamental asymptotic property is proved:

$$\lim_{b\to+\infty} b^{-2(1-H)} \log P(Q > b) = -\delta,$$

where δ is given by:

$$\delta = \frac{1}{2\sigma^2}\left(\frac{C-m}{H}\right)^{2H} \frac{1}{(1-H)^{2(1-H)}}$$

Then we can say that, for large b values, the complementary probability presents a weibullian behaviour:

$$\log P(Q > b) \approx -\delta b^{2-2H}$$

The above relation, named *Norros's Bound* (Norros, 1995), represents a conservative approximation for the overflow probability when we deal with large values of buffer size, b. Figures 4 and 5 show that this bound is quite conservative and so we have to introduce a suitable additive constant which allows to achieve a tighter estimation.
We set this constant value equal to the probability of having a non empty buffer queue:

$$-\mu = \log P(Q > 0).$$

Then the estimation of the overflow probability with a buffer size equal to b is given by:

$$\log P(Q > b) \approx -\delta b^{2-2H} - \mu . \qquad (3)$$

Once obtained the analytical estimation of the overflow probability, in the following section we validate it by means of discrete event simulations.

4 SIMULATION RESULTS

We arranged a set of simulations to verify the effectiveness of the modified bound (3) for different traffic loads and burstiness levels (i.e. H values). To this aim we used Random Midpoint Displacement algorithm (Lau, 1995) to generate several data sets representing sample paths of fractional Gaussian noise with different H values. We show the simulated queueing behaviour for two values of the normalised offered load which are very close to the desired working conditions of the network. For each simulation we assumed a mean arrival rate equal to 300 cells per Time unit and a standard deviation of 100.

Figures 4-7 show the behaviour of the complementary probability $P(Q > b)$ in logarithmic scale against b (so called *watermark plot*): we represent simulation results with dots and error bars (95% confidence interval) and the two analytical approximations with dashed lines (theoretical expression (3) and Norros's bound).

In each case we can see that, for values of b large enough, bound (3) is conservative and represents a tighter approximation for overflow probability than the well-known Norros's bound for different H and normalised load values. When low buffer sizes are considered, the asymptotic nature of bound (3) comes out; the simulation results show that the gap between the proposed relation and the simulated results grows as H increases.

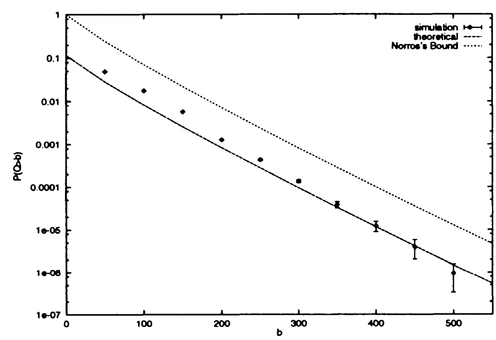

Figure 5. Complementary probability for H=0.55 and offered load 0.7.

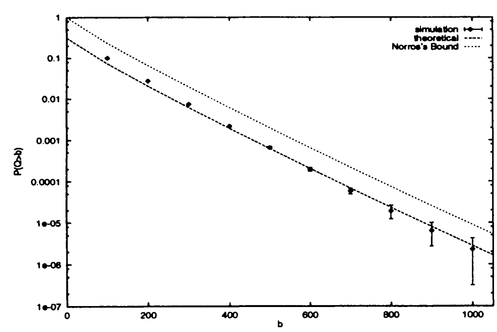

Figure 6. Complementary probability for H=0.55 and offered load 0.8.

Figure 7 highlights the impact of LRD on queueing performances. Moreover, by comparing the results with those shown in figures 4 and 6, we can see that an overflow probability around 10^{-3} is achieved with b=150, b=250, b=500 and b=3000 cells respectively for H=0.55, 0.65, 0.75 and 0.85 with the same offered load equal to 0.7. Thus the higher is H (i. e. the LRD) the more critical is the queueing behaviour.

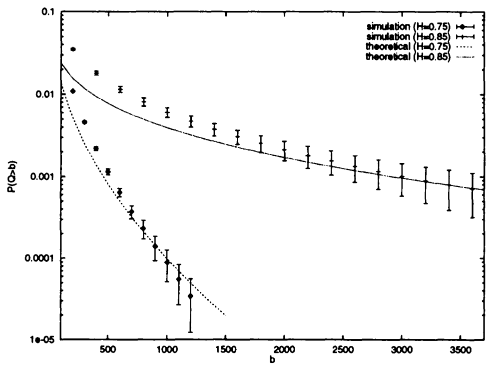

Figure 7. Complementary probability for H=0.65 and offered load 0.7 and 0.8.

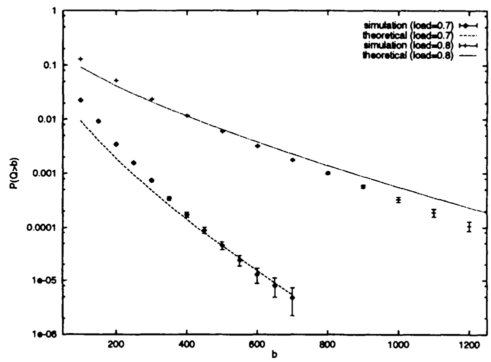

Figure 8. Complementary probability for H=0.75, H=0.85 and offered load 0.7.

5 CONCLUSIONS

In this paper we investigated the performances of an infinite buffer single server queue, with deterministic service rate, loaded by fractional Gaussian noise, a well-known LRD process. In particular we have tested the analytical estimation of the complementary probability determined using LDT. The main contribution of this paper is the introduction of a proper biasing in the results presented in (Duffield, 1994) that permits us to give a tight approximation for buffer overflow probability. This estimator could be useful in the context of preventive resource allocation techniques with particular interest for *call admission control* with stringent QoS requirements.

REFERENCES

Bucklew J. A. (1990), Large Deviations Techinques in Decision, Simulation, and Estimation, John Wiley & Sons, New York.

Dembo A., and Zeitouni O. (1993), Large Deviations Techniques and Applications, Jones and Bartlett Boston, London.

Duffield N. G., and Connell N. O. (1994), Large deviations and overflow probabilities for the general single-server queue, with applications, DIAS Dublin.

Giordano S., Pagano M., Russo F., and Tartarelli S. (1998), Efficiency of Importance Sampling Technique for Queueing Performance Evaluation with LRD Traffic, to be published in AEÜ Int. J. Electron. Commun.

Lau W. C., Erramilli A., Wang J. L., and Willinger W. (1995), Self-similar traffic generation: the random midpoint displacement algorithm and its properties, in Proceedings of the ICC '95, pp. 466-472, Seattle, WA.

Norros I. (1995), On the Use of Fractional Brownian Motion in The Theory of Connenctionless Networks, IEEE-JSAC, Vol. 13, N. 6, pp. 953-962.

Shwartz A., and Weiss A. (1995), Large Deviations for Performance Analysis. Queues, Communications and Computing, Chapman & Hall, London.

Willinger W., Taqqu M. S., and Erramilli A. (1996), A bibliographical guide to self-similar traffic and performance modeling for modern high speed networks, in *Stochastic Networks: Theory and Applications in Telecommunication Networks* (F. P. Kelly, S. Zachary and I. Ziedins, eds.), Vol. 4 of Royal Statistical Society Lecture Notes Series, pp. 91-104, Oxford University Press, Oxford.

Rosario Giuseppe Garroppo was born in Trebisacce, Italy 1969. In 1995 he received the Italian laurea degree in Telecommunications Engineering from the University of Pisa, Italy. Currently he is a PhD student at the Department of Information Engineering of the University of Pisa and his research activity deals with realistic traffic modelling and queueing performances evaluation.

Stefano Giordano was born in Pisa, Italy in 1963. He received the Italian laurea degree in Electronics Engineering from the University of Pisa in 1990 and the Dottorato di Ricerca (PhD) degree in Telecommunication Engineering in 1995.
Since the end of 1993 he is assistant professor at the Department of Information Engineering of the University of Pisa (Telecommunication Networks Group).
His main research and professional areas of interest are Broadband Communications, Telecommunication Networks Analysis, Design and Simulation, Multimedia Communications and Source Coding.

Michele Pagano was born in Lerici, Italy 1968. In 1994 he received the Italian laurea degree in Electronical Engineering from the University of Pisa, Italy, and in the same year he began his PhD studies. Since 1997 he is assistant professor at the Department of Information Engineering of the University of Pisa. His research activity deals with self similar modelling and wavelet analysis of multimedia traffic in broadband networks.

Gregorio Procissi was born in Montecatini Terme, Italy 1972. In 1997 he received the Italian laurea degree in Telecommunications Engineering from the University of Pisa, Italy, and currently he is studying at BRIMS Institute in Bristol. His research activity deals with Large Deviation Theory and self similar processes.

Franco Russo was born in Nola, Italy 1937 and he received his Italian laurea degree in Electronic Engineering in 1962 from the University of Pisa. From 1963 to 1981 he was researcher in the Italian National Research Council (CNR) and since 1981 he is full professor in "Digital Signal Processing". He is member of the Italian National Research Council (CNR) Group in "Telecommunication and Information Theory" and of the executive board of National University Consortium for Telecommunications (CNIT). Currently his main research area covers digital signal processing and telecommunication networks.