The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities

Bong K. Ryu
Center for Telecommunications Research, Columbia University, New York, NY 10027
ryu@ctr.columbia.edu

Anwar Elwalid
Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974
anwar@research.att.com

Abstract

There has been a growing concern about the potential impact of long-term correlations (second-order statistic) in variable-bit-rate (VBR) video traffic on ATM buffer dimensioning. Previous studies have shown that video traffic exhibits long-range dependence (LRD) (Hurst parameter large than 0.5). We investigate the practical implications of LRD in the context of *realistic* ATM traffic engineering by studying ATM multiplexers of VBR video sources over a range of desirable cell loss rates and buffer sizes (maximum delays). Using results based on large deviations theory, we introduce the notion of *Critical Time Scale* (CTS). For a given buffer size, link capacity, and the marginal distribution of frame size, the CTS of a VBR video source is defined as the number of frame correlations that contribute to the cell loss rate. In other words, second-order behavior at the time scale beyond the CTS does not significantly affect the network performance. We show that whether the video source model is Markov or has LRD, its CTS is finite, attains a small value for small buffer, and is a non-decreasing function of buffer size. Numerical results show that (i) *even in the presence of LRD*, long-term correlations do not have significant impact on the cell loss rate; and (ii) short-term correlations have dominant effect on cell loss rate, and therefore, well-designed Markov traffic models are effective for predicting Quality of Service (QOS) of LRD VBR video traffic. Therefore, we conclude that it is unnecessary to capture the long-term correlations of a real-time VBR video source under *realistic* ATM buffer dimensioning scenarios as far as the cell loss rates and maximum buffer delays are concerned.

1 Introduction

There has been a growing concern about the potential impact of long-term correlations of variable-bit-rate (VBR) video traffic on ATM traffic engineering. This concern is based on two recent findings: (i) VBR video traffic exhibits long-range dependence (Hurst parameter larger than 0.5) [2]; (ii) the buffer overflow probability (BOP)[1] of long-range dependent (LRD) video traffic decays *hyperbolically* [14, 18] or in a *Weibull* fashion [5, 17]. A direct implication of (ii) is that the notion of effective bandwidth based on Markov models of which the BOP decays *exponentially* does not apply to LRD traffic (see Section 4 for detail). This has led many researchers and practitioners to believe that the LRD property of VBR video traffic will have critical impact on ATM traffic engineering in general.

Different, and perhaps opposing viewpoints about the impact of the LRD property also exist, however. Heyman and Lakshman [8] and Elwalid *et al* [6] analyzed several traces of VBR videoconferencing traffic and modeled them using the first-order Discrete Autoregressive processes [DAR(1)] (see Section 3.1 for detail) by matching the marginal distribution of frame size and first-lag correlation. Despite the fact that these traces were shown to possess long-range dependence by Beran *et al* [2], Elwalid *et al* showed that for practical ranges of buffer size and cell loss rate (CLR)[1], the DAR(1) model, which is a short-range dependent (SRD) process, provides an accurate prediction of CLRs for the purpose of real-time connection acceptance control. Later, we fully explain this result, which appears to contradict the relevant results that appear in the literature (see Section 4), by considering the realities that (i) the autocorrelations of these traces drop quickly for small lags; and (ii) the buffer size required for multiplexing many VBR video sources is typically small due to the delay constraint imposed on real-time applications.

At least two claims have been made in the literature regarding the importance of the LRD property in ATM traffic

[1] Throughout this paper, we use *buffer overflow probability* (BOP) to represent the probability that buffer content exceeds some amount in an infinite buffer system, and *cell loss rate* (CLR) to represent the fraction of lost cells in a finite buffer system.

engineering:

1. While long-term correlations of LRD processes are individually small, their cumulative effect on the CLR is non-negligible [2].

2. The buffer behavior of LRD VBR video traffic cannot be accurately predicted by simple, parsimonious Markov-based (or SRD in general) models [2, 10, 23].

The focus of this study is two-fold: in a narrow sense, we check the validity of the above two claims under realistic scenarios of ATM buffer dimensioning; and in a broad sense, we explore a general relationship between buffer size and video frame correlations (with other system parameters held fixed). We consider ATM multiplexers of real-time VBR video sources over the following ranges: buffer size (maximum delay) less than 20 – 30 msec, and CLR less than 10^{-6}. In other words, we choose a link capacity (or utilization) such that the system operating point lies within the above ranges of buffer size and CLR. We envision that the total end-to-end delay (propagation + processing + network queueing + others) allowed for typical real-time VBR video applications is about 200 msec or so, or even smaller. Since such applications are most likely to pass along several network nodes, we believe that the delay at each node should be kept within 20 – 30 msec at maximum. Using results from large deviations theory, we show that: (i) long-term correlations do not have significant impact on CLR; and (ii) (ii) short-term correlations have dominant effect on cell loss rate, and therefore, well-designed Markov traffic models are effective for predicting Quality of Service (QOS) of LRD VBR video traffic. These results, verified by simulation, disprove the above claims under *practical ranges of buffer size and CLR*.

Since the key issue of this study is to investigate the effect of different types of autocorrelation structures (short-range dependence or long-range dependence) on the multiplexer performance, it is essential that we have models of which the frame correlations can be easily and flexibly adjusted. For this reason, we use stochastic VBR video models, unlike the previous studies in which real video traces were used. In particular, we construct LRD video models such that they resemble real traces exhibiting LRD; see [8] for examples of such traces. Further, in order to exclude the effect of the marginal distribution (of frame size) on buffer behavior, we set the marginal distributions of all models to be identical so that the first-order statistics do not contribute to the difference in queueing behavior. We choose Gaussian frame size distribution to simplify our analysis. It is noted that the Gaussian frame size distribution may not be a universal characterization of video traces; some traces have been found to possess heavier tails than that of the Gaussian distribution [7]. As discussed in Section 6, however, we believe that our main results are unlikely to be affected by this assumption.

Based on the results from large deviations theory, we introduce the notion of *Critical Time Scale* (CTS) which describes a relationship between buffer size and frame correlations (with other system parameters such as link capacity and marginal distribution of frame size held fixed). For given buffer size, the CTS of a VBR video source is defined as the number of frame correlations that contribute to the CLR. For example, for VBR video sources with Gaussian marginal distribution of frame size, knowledge of CTS implies *exactly* how many frame correlations affect the cell loss rate (CLR); correlations after such lag (CTS) do not contribute to CLR. Later we show that whether the model is Markov or has the LRD property, its CTS is finite, attains a small value for small buffer, and is a non-decreasing function of the buffer size. In other words, under realistic scenarios of ATM buffer dimensioning, the number of frame correlations which affect BOP is finite and small *even in the presence of the LRD property*. As a result, it is unnecessary to capture the long-term correlations of a real-time VBR video source as far as the cell loss rates and maximum buffer delays are concerned as QOS metrics.

The rest of this paper is organized as follows. Section 2 provides two definitions of an LRD process used in this paper. Section 3 presents the details of two stochastic processes: Discrete-AutoRegressive process of order p [DAR(p)] [9] and Fractal-Binomial-Noise-Driven Poisson point process (FBNDP) [19]. Four VBR traffic source models, V^v, Z^a, S, and L, are constructed using the above processes and their marginal distributions and autocorrelations are discussed. Section 4 provides an overview of results from large deviations theory on the BOP of ATM multiplexers. Using these analytical tools, we examine the two claims by comparing the CTSs and BOPs of the four video models. Section 5.5 provides simulation results which verify the analytical results of Section 4. Finally, conclusions and discussions are presented in Section 6.

2 Long-Range Dependence: Definitions

Let $X = \{X_k : k \geq 1\}$ be a wide-sense stationary (WSS) process in the discrete-time domain with $\mu \equiv \mathrm{E}[X_k]$, variance $\sigma^2 \equiv \mathrm{E}[(X_k - \mu)^2]$, and autocorrelation function

$$r(k; T_s) \equiv \mathrm{E}[(X_n - \mu)(X_{n+k} - \mu)]/\sigma^2.$$

X_k represents the size of the k-th frame in cells with frame duration of T_s sec. We call X an *asymptotic* LRD process if the tail of its autocorrelation function $r(k; T_s)$ is given by[2] [10]:

$$r(k; T_s) \sim k^{-(2-2H)}, \qquad \text{as } k \to \infty. \tag{1}$$

The quantity H is called the *Hurst parameter* and completely characterizes the relation (1). *Fractional autoregressive integrated moving average* models , *F-ARIMA(p,d,q)*, are an example of asymptotic LRD processes [10].

Likewise, we call X an *exact* LRD process if

$$r(k; T_s) = \frac{1}{2} g(T_s) \nabla^2 (k^{2H}) \qquad \text{for } k > 0 \tag{2}$$

[2] The symbol $\sim$ denotes an asymptotic relation.

with $0.5 < H < 1$, $0 \leq g(\cdot) \leq 1$ independent of k, and $\nabla^2(h(k)) \equiv h(k+1) - 2h(k) + h(k-1)$, the second central difference operator. Due to the asymptotic equivalence of differencing and differentiation, (2) can be approximated to

$$r(k;T_s) \approx g(T_s)H(2H-1)k^{-(2-2H)} \tag{3}$$

for large k. The discrete-time fractional Gaussian Noise (FGN) process has an autocorrelation function of the form (2) with $g(T_s) = 1$ [21], and therefore is an exact LRD process. A family of fractal stochastic point processes (FSPPs) introduced in [19, 20] also yield exact LRD processes in the sense of (2) with $g(T_s) = T_s^{2H-1}/(T_s^{2H-1} + T_0^{2H-1})$. T_0 is called fractal onset time which is used to control the variance of the frame size; see [20] for detail. We distinguish (1) and (2) because we feel that the first claim seems to stem from (1) which completely ignores the short-term correlations of X.

3 LRD and SRD VBR Video Models

We check the validity of the two claims given in the introduction by studying two relevant issues, one for each claim, respectively:

I: Effect of short- and long-term (frame) correlations on CLRs.

II: Efficacy of Markov models in predicting the loss characteristic of LRD VBR video traffic.

The objective of studying the first issue is to examine whether the long-term correlations have significant impact on CLRs in the presence of the LRD property. For this purpose, we use two *asymptotic* LRD models: V^v and Z^a. The long-term correlations of V^v are controlled by v, while its short-term correlations (at least the first-lag correlation) are unchanged. On the other hand, the short-term correlations of Z^a are controlled by a, while its long-term correlations (Hurst parameter) are unchanged; see Fig. 1. We use the

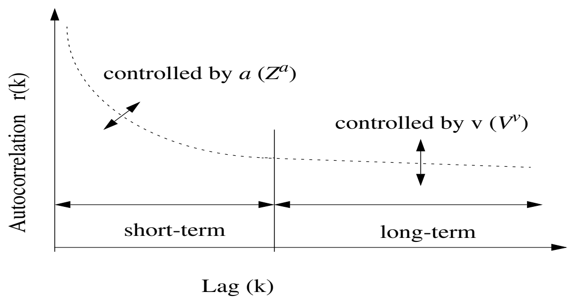


Figure 1: Effect of the change of a and v on the autocorrelation function of the models Z^a and V^v.

superposition of the Fractal-Binomial-Noise-Driven Poisson point process (FBNDP) [20] and DAR(1) process [9] for constructing V^v and Z^a. Details on how we construct these two models are given in Section 3.3. For the first claim to be valid, we must observe, under practical scenarios of ATM buffer dimensioning, that: (i) the CLRs of V^v for different values of v are significantly different; and (ii) the CLRs of Z^a for different values of a are close to each other. Later, we show that what we observe is the opposite, indicating that *even in the presence of the LRD property*, it is the short-term correlations that have a dominant impact on CLRs under *realistic scenarios of ATM buffer dimensioning*, not the long-term correlations.

The objective of studying the second issue is to investigate whether a simple, well designed Markov model can provide a good prediction of the CLRs of an LRD VBR video source. For this purpose, we use two additional models: a short-range dependent (SRD) model S, and an exact LRD model L. For a given asymptotic LRD model Z^a, we construct S such that the first p correlations of Z^a are exactly matched. Also, we use an exact LRD model L to match only the long-term correlations of Z^a. We use DAR(p) process for S and the FBNDP process for L. In summary, Z^a serves the role of a real trace that exhibits the LRD property; S [DAR(p)] captures only the short-term correlations of Z^a; and L only the long-term correlations of Z^a; see [9] and [20, chapter 6] on how DAR(p) processes capture the first p correlations of Z^a. For the second claim to be valid, we must observe that L outperforms S in predicting the CLR of Z^a over practical ranges of buffer sizes and CLRs. Later, we show that when a system operating point is chosen within practical ranges of buffer sizes and CLRs, DAR(p) models perform superior to L even for small p ($p = 1, 2, 3$).

The crucial feature of these four models (V^v, Z^a, S, and L) is that their marginal distributions of frame size are identical, implying that the first-order statistics do not contribute to the difference in buffer behavior. As discussed in the introduction, we choose the Gaussian distribution since the marginal distribution is not the concern of this study.

3.1 DAR(p)

The DAR(p) process (discrete autoregressive process of order p), constructed and analyzed by Jacobs and Lewis [9], is a p-th order Markov chain with the physical and correlation structure of a p-th order autoregressive process; S_n depends explicitly on $S_{n-1}, \ldots, S_{n-p}$. The process is specified by the stationary marginal distribution of $\{S_n\}$ and several other chosen parameters which, independently of the marginal distribution, determine the correlation structure.

This DAR(p) process is defined as follows. Let $\{\epsilon_n\}$ be a sequence of i.i.d. random variables taking values in Z, the set of integers, with distribution π. Let $\{V_n\}$ be a sequence of Bernoulli random variables with $\mathrm{P}(V_n = 1) = 1 - \mathrm{P}(V_n = 0) = \rho$ for $0 \leq \rho < 1$. For the DAR(1) process, ρ represents the first-lag autocorrelation. Let $\{A_n\}$ be a sequence of i.i.d. random variables taking values in $\{1, 2, \ldots, p\}$ with $\mathrm{P}(A_n = i) = a_i \geq 0, \quad i = 1, 2, \ldots, p$ with $\sum_{i=1}^{p} a_i = 1$. Let

$$S_n = V_n S_{n-A_n} + (1 - V_n)\epsilon_n \tag{4}$$

for $n = 1, 2, \ldots$. Then, the process $S = \{S_n\}$ is called the DAR(p) process. Note that this process has p degrees of freedom, and therefore has the capability to match up to the first p autocorrelations. a

3.2 Fractal-Binomial-Noise-Driven Poisson Process (FBNDP)

The FBNDP model is described in [20] in detail. Hence, only a brief description and relevant statistics of the model is provided here. First, we obtain a fractal ON/OFF process with ON/OFF periods independently and identically distributed by the same heavy-tailed density function as

$$p(t) = \begin{cases} \gamma A^{-1} e^{-\gamma t/A} & \text{for } t \le A, \\ \gamma e^{-\gamma} A^{\gamma} t^{-(\gamma+1)} & \text{for } t > A, \end{cases}$$

with $\gamma = 2 - \alpha$ $(1 < \gamma < 2)$. Then, M independent and identical fractal ON/OFF processes are added, yielding a fractal binomial process (FBN). This serves as a stationary stochastic rate function for a Poisson process; the FBNDP results [20].

This model has four parameters: α, A, M, and R (the rate of a fractal ON/OFF process when it is ON) [20]. They determine the relevant statistics of the FBNDP as

$$\begin{aligned} H &= (\alpha+1)/2 \\ \lambda &= RM/2 \\ T_0 &= \left\{\alpha(\alpha+1)(2-\alpha)^{-1}\left[(1-\alpha)e^{2-\alpha}+1\right] R^{-1}A^{\alpha-1}\right\}^{1/\alpha} \end{aligned}$$

where H is the Hurst parameter, λ the mean arrival rate (cells/sec), and T_0 the fractal onset time (sec).

Denote by $N(t)$ the counting process (i.e., number of arrivals up to time t) constructed from the FBNDP. If an exact LRD process $L = \{L_n\}$ is constructed from the FBNDP as

$$L_n \equiv N[nT_s] - N[(n-1)T_s],$$

then we have

$$\begin{array}{lclcl} \mu_L & \equiv & \mathrm{E}[L_n] & = & \lambda T_s \\ \sigma_L^2 & \equiv & \mathrm{Var}[L_n] & = & \left[1 + (T_s/T_0)^{\alpha}\right] \lambda T_s \\ r_L(k) & \equiv & \mathrm{Cov}(L_n, L_{n+k})/\sigma_L^2 & = & \frac{T_s^{\alpha}}{T_s^{\alpha}+T_0^{\alpha}} \cdot \frac{1}{2}\nabla^2(k^{\alpha+1}) \end{array}$$

where T_s is the frame duration and ∇^2 the second central difference operator. Since M is an extra parameter, it is used to control the marginal distribution of $\{L_n\}$. In fact, for large M and μ_L, the marginal distribution of $\{L_n\}$ approaches a Gaussian form with mean μ_L and variance σ_L^2 by the central limit theorem [15].

3.3 $\{V^v, Z^a\}$ = FBNDP + DAR(1)

Many analyses of VBR real traces show that their short-term correlations are characterized by geometric decay [8]. When those traces exhibit the LRD property, mathematically it is represented as power-law decaying long-term correlations; see (1). In order to make V^v and Z^a represent a wide range of real LRD VBR traces in terms of frame correlations, we use the superposition of the FBNDP and DAR(1) processes for constructing Z^a and V^v for which the DAR(1) process contributes to geometric decay for small lags and the FBNDP to power-law decay for large lags; see Figs. 3 (a) and (b).

Let the process $X = \{X_n\}$ be constructed from the FBNDP with a Gaussian marginal distribution with (μ_X, σ_X^2) and $Y = \{Y_n\}$ from DAR(1) also with a Gaussian marginal distribution with (μ_Y, σ_Y^2), and assume X and Y are independent. Then, both the Z^a and V^v also have a Gaussian marginal distribution with

$$\begin{aligned} \mu &= \mu_X + \mu_Y \\ \sigma^2 &= \sigma_X^2 + \sigma_Y^2 \\ r(k) &= \frac{v}{v+1} r_X(k) + \frac{1}{v+1} r_Y(k) \\ &= \frac{v}{v+1} \cdot \frac{T_s^{\alpha}}{(T_s^{\alpha}+T_0^{\alpha})} \cdot \frac{1}{2}\nabla^2(k^{\alpha+1}) + \frac{1}{v+1} a^k, \quad (5) \end{aligned}$$

with a the first-lag correlation of the DAR(1) component and $v \equiv \sigma_X^2/\sigma_Y^2$. Note that $r(k)$ is expressed as a weighted sum of $r_X(k)$ and $r_Y(k)$. The fact that there exists $k_0 < \infty$ such that $r_X(k) > r_Y(k)$ for $k > k_0$, regardless of the choice of α, a, and v (> 0), makes Z^a and V^v asymptotic LRD processes; see Fig. 1.

Indeed, a and v in (5) are the same as in Z^a and V^v, respectively. For V^v, we fix H (or α) and change a such that for different values of v, the first-lag correlation is identical and the next several correlations are very close to each other; see Table 1 and Fig. 3-(a) for the resulting parameters and autocorrelations. For Z^a, we again fix H and change a to obtain diverse behavior of short-term correlations. In this case, we set $v = 1$ (and thus $\sigma_X^2 = \sigma_Y^2$) and $\mu_X = \mu_Y$ so that both the FBNDP and DAR(1) processes contribute equally to the mean and variance of Z^a.

4 Large Deviations Theory and Critical Time Scale

4.1 Brief history on large buffer asymptotics for LRD processes

The first result on queueing analysis of self-similar traffic seems to appear in Norros [17] in which the popular Weibull (lower) bound on the BOP has been established using fractional Brownian Motion. Also, using the techniques of large deviations theory, Duffield *et al* [5] have shown that under mild conditions, the BOP of LRD traffic, whether exact or asymptotic, is again approximated by Weibull behavior *as the buffer size goes to infinity.* While the above results exhibit Weibull decays, Likhanov *et al* [14] and Parulekar and Makowski [18] show that for the $M/G/\infty$-type model of Cox [4], the tail asymptotic of the BOP decays at most hyperbolically. We provide another Weibull approximation for statistically identical N Gaussian exact LRD processes, which is given by

$$\mathrm{P}(W > B) \approx \exp\left[-J(N,b,c) - \frac{1}{2}\log\left\{4\pi J(N,b,c)\right\}\right] \quad (6)$$

with

$$J(N,b,c) \equiv N^{2H-1}\frac{(c-\mu)^{2H}}{2g(T_s)\sigma^2\kappa(H)^2}B^{2-2H}$$

where $\kappa(H) \equiv H^H(1-H)^{1-H}$, $g(T_s)$ is from (2), N is the number of homogeneous sources multiplexed, μ is the mean frame size per source (cells/frame), σ^2 is the variance of frame size, and c is the bandwidth per source (cells/frame). The proof is provided in the Appendix. [3] Note that for $H = 1/2$ and large N, (6) is reduced to a familiar log-linear behavior as predicted by the well-known technique of effective bandwidth based on Markov models. Indeed, one may think of (6) as the *result* of the first claim and the *cause* of the second claim. Also, it serves as a basis for the argument that the popular effective bandwidth scheme based on simple Markovian traffic models would consequently provide a poor estimate of CLRs of LRD traffic.

4.2 Large buffer asymptotics and the Critical Time Scale for Gaussian sources

Large deviations theory has served as a powerful tool for analyzing the performance of communications networks [22, references therein]. We use the *Bahadur-Rao (B-R) Asymptotic* of buffer overflow probability of an ATM multiplexer with N sources [16].

We use the following notations [3, 16]. Denote by C the constant service rate during the frame duration measured in cells/frame and by $\{X_n\}$ the input process (i.e., the size of n-th frame from N identical sources) in cells. The workload at the start of n-th frame interval is denoted by W_n and

$$W_{n+1} = (\min\{(W_n + X_n - C), B\})^+$$

where B is a buffer size (cells) and $x^+ \equiv \max(x,0)$. Let b and c denote respectively the amounts of buffer space and bandwidth per source, so that $B = Nb$ and $C = Nc$. Suppose $\{X_n\}$ is the superposition of N identical sources with Gaussian marginal, each distributed as $\{Y_n\}$ with mean μ, variance σ^2, and autocorrelation function $r(k)$. Then, the BOP $\Psi(c,b,N)$ predicted by the B-R asymptotic takes the following form [16]

$$\Psi(c,b,N) \approx \exp\left(-NI(c,b) + g_1(c,b,N)\right) \tag{7}$$

with

$$I(c,b) \equiv \inf_{m\geq 1}\frac{[b+m(c-\mu)]^2}{2V(m)}, \tag{8}$$

$$g_1(c,b,N) \equiv -1/2\log[4\pi N I(c,b)] \tag{9}$$

and

$$V(m) \equiv \mathrm{Var}\left(\sum_{i=1}^{m} Y_i\right) = \sigma^2\left[m + 2\sum_{i=1}^{m}(m-i)r(i)\right]. \tag{10}$$

[3] We note that similar expression (without the second term in the exponent) has been obtained by Parulekar and Makowski [18] for the discrete-time fractional Gaussian Noise (FGN) process (and thus $g(T_s) = 1$) using the results of Duffield *et al* [5].

Note that ignoring $g_1(c,b,N)$ yields the *Large N Asymptotic* given by Courcoubetis and Weber [3].

Let's examine (8) and (10) with a closer look. Let $f(c,b,m) \equiv [b + m(c-\mu)]^2$ and denote by m_b^* the value of m at which the infimum of $f(c,b,m)/2V(m)$ is obtained, i.e., $m_b^* \equiv \mathrm{arginf}_{m\geq 1} f(c,b,m)/2V(m)$. Assuming that m_b^* is unique [16], for fixed b, c, and μ, m_b^* *represents the number of frame correlations that completely determine* $\Psi(c,b,N)$. More importantly, it implies that *the first* m_b^* *frame correlations are only meaningful in evaluating* $\Psi(c,b,N)$ *through* $V(m_b^*)$. For this reason, we call m_b^* *Critical Time Scale* (CTS). It follows that the LRD property described by (1) and (2) must force m_b^* to be (at least) very large even for small b, in order to support the first claim.

We first show that m_b^* is finite in most practical cases. Assuming that the autocorrelation function $r(k)$ is monotonically decreasing, we note that both $f(c,b,m)$ and $V(m)$ are increasing functions of m. Also, the increasing rate of $V(m)$ is slower than m^2, i.e.,

$$V(m) \approx \begin{cases} \sigma^2[m + O(m^2)] & \text{for SRD processes [8]} \\ \sigma^2 g(T_s) m^{2H} & \text{for exact LRD processes} \end{cases}$$

for large m [4, 8]. It follows that for fixed b,

$$\lim_{m\to\infty} f(c,b,m)/2V(m) = \infty,$$

yielding $m_b^* < \infty$. This result immediately indicates that for finite buffer size, the cumulative effect of autocorrelations is bounded regardless of the presence of the LRD property, disproving the first claim.

Second, our investigation further indicates that m_b^* is actually an order of b, i.e., $m_b^*/b \approx K$ for some constant K. For example, for a relatively large m_b^*, it has been found that $K = 1/(c-\mu)$ for a Gaussian AR(1) process [3] and $K = \frac{H}{(1-H)(c-\mu)}$ for a Gaussian exact LRD process (see Appendix). [4] Moreover, m_0^* is always equal to 1, indicating that correlations do not affect the CLR at all when buffer size is zero. [3, 6]. From this, we stipulate that for small b, m_b^* would be also small.

The following section provides numerical examples which show that whether the model is SRD (Markov) or LRD, its CTS (m_b^*) is finite, attains a small value for small buffer, and is a non-decreasing function of buffer size b. These results are converged to that (i) the Weibull behavior in (6) is not significant for realistic buffer size; and (ii) DAR(p) processes capturing short-term correlations of Z^a provide a satisfactory approximation of CLRs of Z^a.

5 Numerical Results

5.1 Parameter specification

We choose the model parameters of V^v, Z^a, S, and L as follows.

[4] We note that the same result has been independently obtained by Addie *et al* [1] using the superposition of two Gaussian AR(1) processes.

1. The frame rate is chosen to be 25 frames/sec, yielding frame duration $T_s = 0.04$ (sec).

2. The marginal distribution of all the models is Gaussian with mean (μ) 500 (cells/frame) and variance (σ^2) 5000 (cells/frame)^2. The choice of $M = 15$ for Z^a and V^v and $M = 30$ for L for the underlying process FBNDP, respectively, are found to provide good approximations of the Gaussian marginal distribution for simulation purpose.

3. For V^v, three values of v are used: 0.67, 1, and 1.5. Corresponding values of a are determined such that all the $V^{0.67}$, V^1, and $V^{1.5}$ have the same first-lag correlation; see Fig. 3-(a).

4. For Z^a, four values of a are used: $0.7, 0.9, 0.975, 0.99$. Note that once α, λ, T_0, and M are fixed, the marginal distribution of Z^a is not affected by a.

5. For each a of Z^a, we determine the parameters of the DAR(p) model (S) such that it exactly matches the first p correlations of Z^a using the steps given in [20, chapter 6].

6. The value of α for the FBNDP component of Z^a was chosen as 0.8, yielding the Hurst parameter $H = 0.9$.

7. The value of α for L was obtained such that the tail of $r_L(k; T_s)$ provides the best fit to that of $r_{Z^a}(k; T_s)$. Note that because of the $v/(v+1)$ factor in (5), one cannot use the same α of Z^a for L in order to make their long-term correlations identical. The choice of $\alpha = 0.72$ ($H = 0.86$) was found to be sufficiently accurate for the purpose of this study; see Fig. 3-(b).

8. Finally, the values of T_0 for Z^a and L are determined from the given mean, variance, and α of each model.

Table 1 summarizes the values of all the parameters of each model. We note that the different choice of key parameters such as H yields the qualitatively same result as in Table 1. Fig. 2 shows sample paths of $Z^{0.7}$ and its corresponding DAR(1) model which matches the first-lag correlation of $Z^{0.7}$. Strong low-frequency behavior (thus long-term correlations) is clearly visible in $Z^{0.7}$. In particular, the burst-within-burst structure of self-similar traffic, i.e., traffic spikes riding on longer-term ripples, that in turn ride on still longer term swells [10], is easily observed. On the other hand, such characteristic is not observed in the DAR(1) model, as it is a short-range dependent model (Hurst parameter is 0.5). Instead, the DAR(1) process captures the fast-time scale behavior of $Z^{0.7}$ quite closely.

5.2 Autocorrelations

With the parameter specification given in the previous section, we examine the autocorrelations of V^v, Z^a and L for a wide range of time scales. Fig. 3-(a) shows that the short-term correlations of V^v are very close to each other (up to about five lags), and in particular, their first-lag correlation is identical. Fig. 3-(b) shows that (i) the larger value of a of Z^a yields stronger short-term correlations; and (ii) the long-term correlations of both Z^a and L are very close up to at least 1,000 lags. We note that several traces of VBR video-conferencing traffic exhibit the correlation structure similar to Z^a (or V^v) [6, 8]. Figs. 3-(c) and (d) show that indeed the DAR(p) processes match the first p correlations of Z^a.

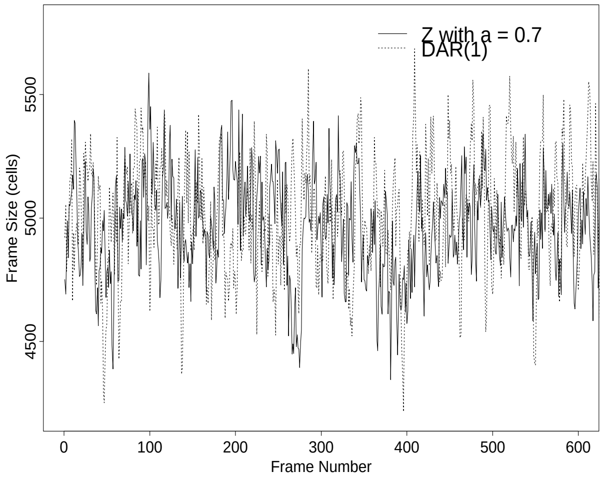


Figure 2: Sample path comparison of $Z^{0.7}$ and matched DAR(1) for 10 sources multiplexed ($N = 10$).

5.3 Critical time scales

Let us examine the effect of different correlation structures (SRD or LRD) on the CTS (m_b^*). Fig. 4-(a) compares m_b^* of V^v for three different values of v. Recall that the short-term correlations of V^v are very close to each other; see Fig. 3-(a). As a result, the values of their m_b^* are much the same for small buffer. On the other hand, m_b^* of Z^a for different values of a shows significant difference, as many as 15 even at $B = 2$ (msec). These results clearly indicate that *it is the short-term correlations which have dominant impact on the CTS, not the long-term correlations.* Both figures also illustrate that m_b^* attains a small value for small b (or B) and non-decreasing function of b (or B). Moreover, we observe that the process with stronger short-term correlations (higher a) yield larger m_b^* for the same size of buffer in spite of the identical long-term correlations.

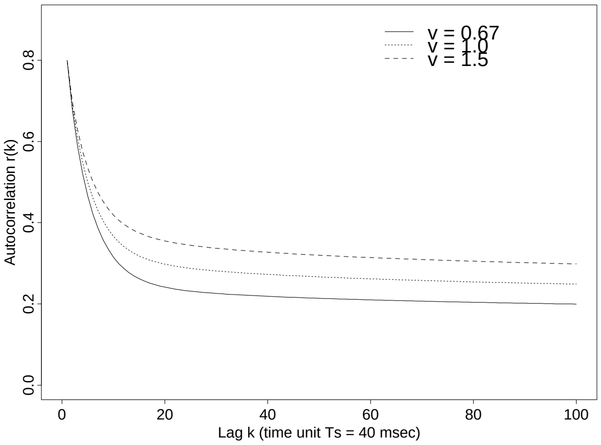
5.4 Large buffer asymptotics

Figs. 5-(a) and (b) compare the buffer overflow probabilities (BOPs) of V^v and Z^a for $N = 30$ based on the B-R asymptotic (7). As predicted by Fig. 4, these figures clearly show that short-term correlations are dominant in determining BOPs (or CLRs) for small buffer. Close behavior in short-term correlations yield close loss probabilities [Fig. 5-(a)], while drastically different short-term correlations yield

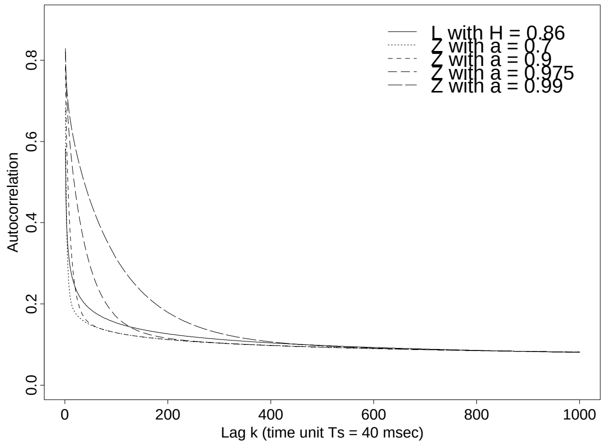
	v	α	a	λ (cells/sec)	T_0 (msec)	M
V^v	0.67	0.9	0.799761	5000	3.48	15
	1	0.9	0.8	6250	3.48	15
	1.5	0.9	0.800362	7500	3.48	15
Z^a	1	0.8	0.7, 0.9, 0.975, 0.99	6250	2.57	15
L	—	0.72	—	12500	1.83	30

S	a	0.7	0.975
	DAR(1)	$\rho = 0.82$, $a_1 = 1$	$\rho = 0.68$, $a_1 = 1$
	DAR(2)	$\rho = 0.87$, $a_1 = 0.70$, $a_2 = 0.3$	$\rho = 0.72$, $a_1 = 0.84$, $a_2 = 0.16$
	DAR(3)	$\rho = 0.89$, $a_1 = 0.63$, $a_2 = 0.18$, $a_3 = 0.19$	$\rho = 0.73$, $a_1 = 0.82$, $a_2 = 0.10$, $a_3 = 0.08$

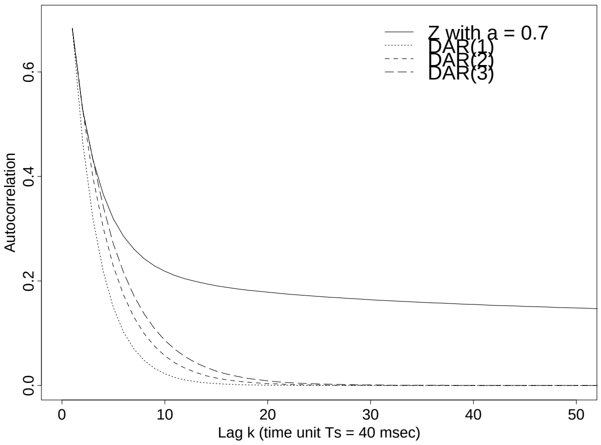
Table 1: Specification of model parameters of V^v, Z^a, S, and L. For S, two cases of Z^a are considered: $Z^{0.7}$ and $Z^{0.975}$.



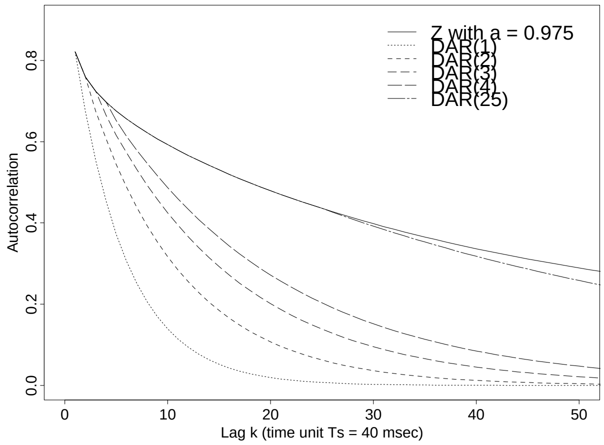
(a) Correlations of V^v for different values of v



(b) Correlations of Z^a for different values of a



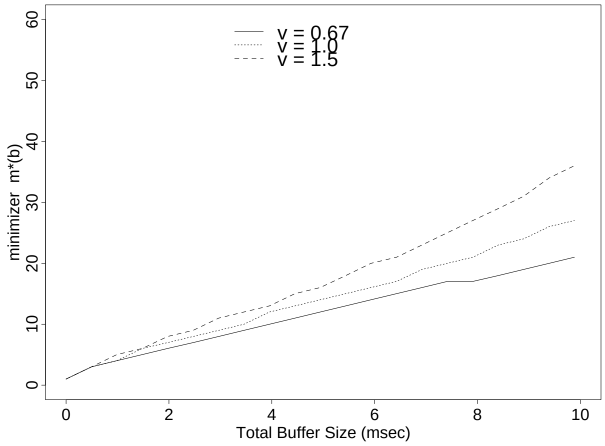
(c) $Z^{0.7}$ and corresponding S (DAR(p))



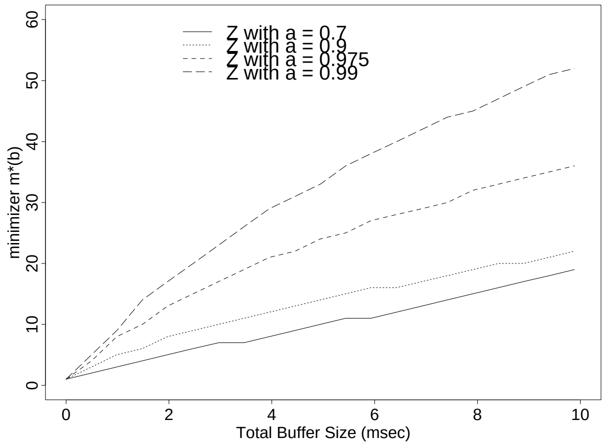
(d) $Z^{0.975}$ and corresponding S (DAR(p))

Figure 3: Analytic autocorrelation functions of V^v, Z^a, S, and L

significant discrepancy despite of having identical long-term correlations [Fig. 5-(b)]. Also, stronger correlations of an arrival process yield slower decaying BOP for the range of buffer size of interest. This illustrates why the first claim is not valid under practical ranges of buffer size and CLR. We note that eventually the curves in Fig. 5-(b) will exhibit the same decaying characteristic when buffer size becomes larger and larger, for they all have the same Hurst parameter [5]; however, such range of the buffer size is beyond the practical consideration; see Fig. 7.



(a) V^v

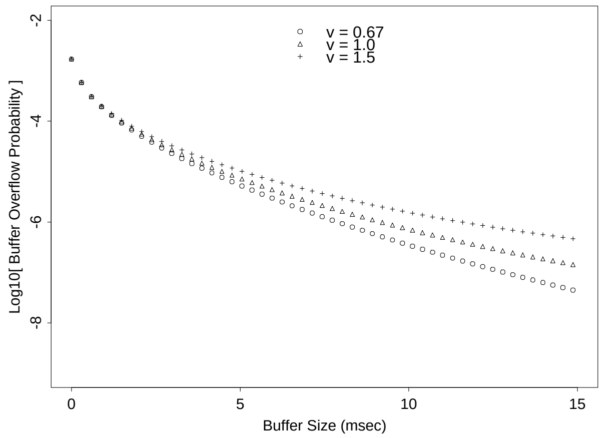


(b) Z^a

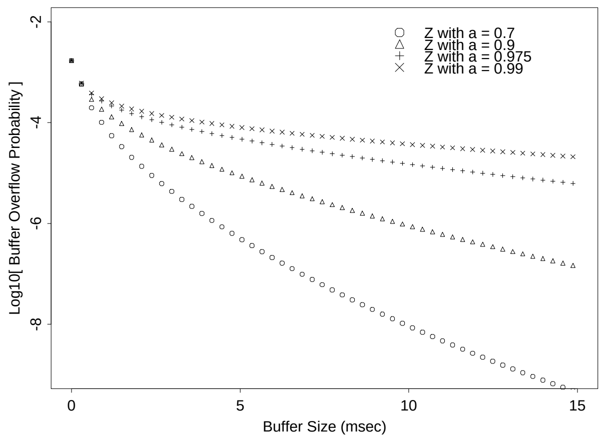
Figure 4: Minimizer m_b^* *vs* total buffer size B. (a) effect of same short-term correlations on m_b^*, (b) effect of same long-term correlations on m_b^*. In both cases, c and μ are fixed as $c = 526$ and $\mu = 500$, and $N = 100$.

Fig. 6-(a) compares the BOPs of $Z^{0.975}$, DAR(p), $p = 1, 2, 3$, and L again based on the B-R asymptotic (7). It is easy to observe that even the DAR(1) model outperforms L for a wide range of buffer size of interest. Moreover, notice that the curve of each DAR(p) model is getting closer to that of $Z^{0.975}$ as p increases, implying that capturing more and more short-term correlations yields more accurate prediction of CLR. This clearly breaks the second claim; simple Markov models are indeed capable of providing good approximations of BOP (or CLRs) of an LRD process for the practical ranges of buffer size and CLR.

Fig. 6-(b) compares the B-R asymptotics of $Z^{0.7}$ and DAR(p), $p = 1, 2, 3$. Recall that all of their autocorrelation functions drop very quickly at small lags (see Figs. 3-(b) and (c)), yielding close loss behavior as we have seen from V^v. This type of autocorrelations has been observed in several VBR videoconferencing traces studied in [6, 8]. We further observe that the difference between all the curves at the loss rate 10^{-6} is only within the order of one. This difference becomes negligible when the loss rate is translated to the number of admissible VBR video connections, which is why the DAR(1) model provides accurate prediction of the number of admissible connections for LRD traces [6].



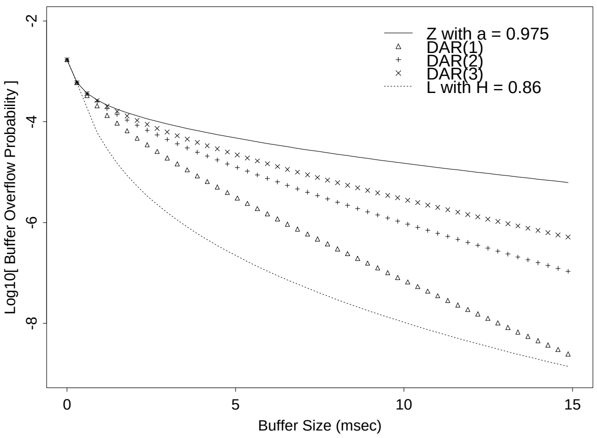
(a) V^v



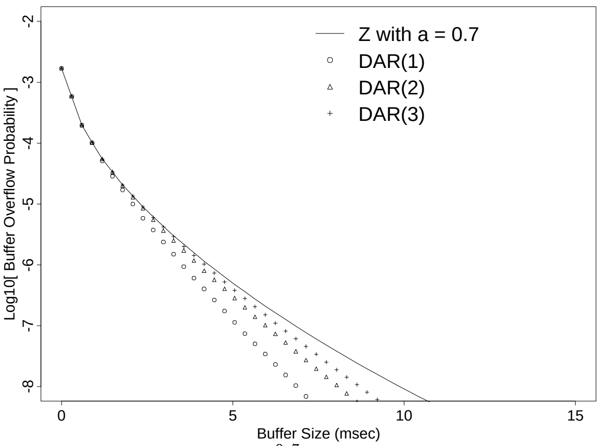
(b) Z^a

Figure 5: Comparison of BOPs of V^v and Z^a based on the Bahadur-Rao asymptotic with $N = 30$ and $c = 538$ (cells/frame): effect of short- and long-term correlations on the cell loss.

We provide one more evidence illustrating the impact of correlation structure on loss behavior. Figs. 7-(a) and (b) repeat Figs. 6-(a) and (b) over much wider range of buffer size, even for *unrealistically large buffer sizes*. The box in the upper left corner of each figure corresponds to the ranges of buffer size and CLR over which Figs. 6-(a) and (b) are drawn. Indeed, these two figures show where the two claims are coming from; L eventually outperforms Markov models over the range of buffer size which is *beyond practical consideration*. Observe that the decaying rates of Z^a follow that of L from about $B = 40$ msec, which serves as another pictorial proof that Z^a and L have the same long-term correlations. All of these arguments converge to the point that future traffic analysis should focus more on finding appropriate time scale at which traffic behavior is to be captured, rather than on providing accurate traffic models. The following section provides simulation results, verifying the results of this section.



(a) $Z^{0.975}$, DAR(p), L

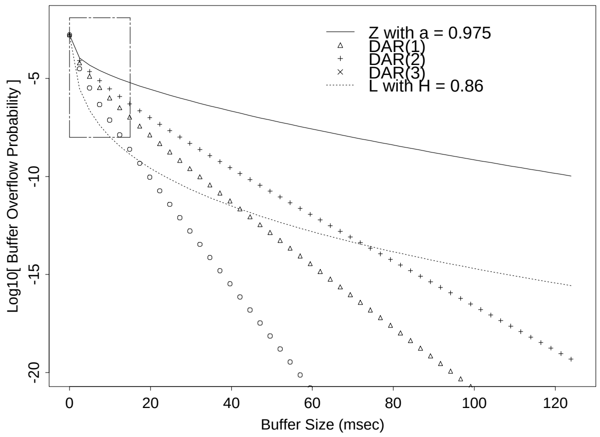


(b) $Z^{0.7}$, DAR(p)

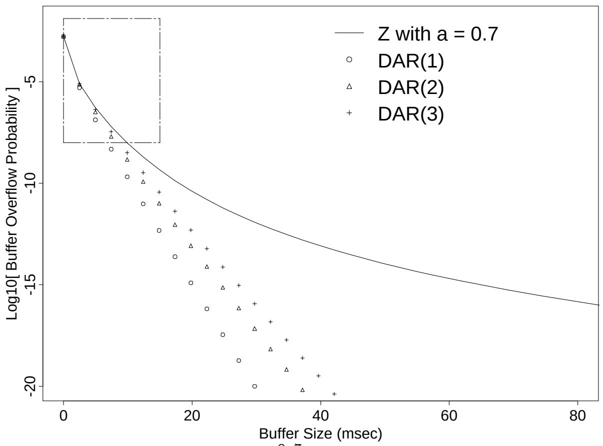
Figure 6: Comparison of BOPs of Z^a, DAR(p), and L with $N = 30$ and $c = 538$: efficacy of simple Markov models.

5.5 Simulation

For each of the four models we run 60 replications, each of which generates half a million frames. This extensive amount of computation ensures accurate and numerically confident estimations which may not be otherwise obtained due to the heavy-tailed ON/OFF times of the FBNDP model. We assume that the beginning of frame of each source is same and that cells are equispaced over the frame duration (deterministic smoothing). For a given buffer size, we estimate the fraction of lost cells (finite buffer system).



(a) $Z^{0.975}$, DAR(p), L

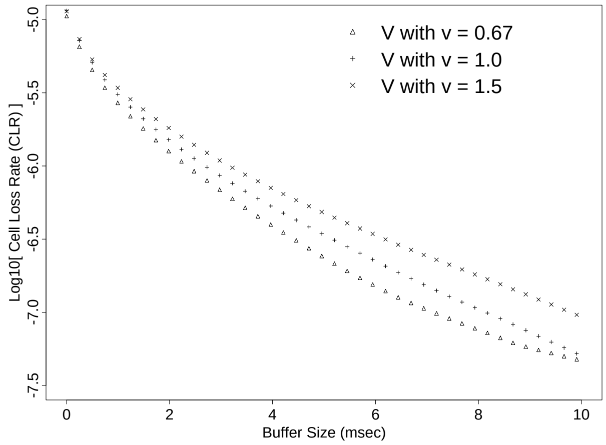


(b) $Z^{0.7}$, DAR(p)

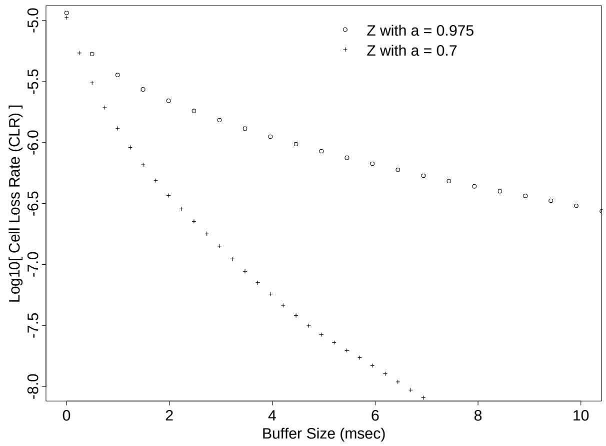
Figure 7: Comparison of BOPs of Z^a, L, and corresponding DAR(p) over a wide range of buffer size with $N = 30$ and $c = 538$.

Most importantly, both Figs. 8 and 9 confirm the results predicted by Figs. 5 and 6, respectively. We note that all the CLR curves begin around the same value at zero buffer (slightly larger than 10^{-5}) which confirms that the parameters given in Table 1 produce the same Gaussian marginal distribution of frame size for all four models. Other results with different choices of N and c, which are omitted here, show qualitatively the same results as Figs. 8 and 9, further supporting our conclusions.

Fig. 10 compares the two large buffer asymptotics (B-R and large N asymptotics) with simulation results for DAR(1) model matched to $Z^{0.975}$. First, we observe that all the curves are parallel under the ranges of buffer size and CLR of interest, indicating that both asymptotics capture the loss behavior of a given model in a qualitative sense. Second, the B-R asymptotic provides tighter upper bound for the CLR than the large N asymptotic, with about one order of magnitude improvement in this case. However, the difference between the B-R asymptotic and the estimated CLR is still about 2 orders of magnitude, posing an open question of the applicability of large buffer asymptotics (infinite buffer) for approximating CLRs (finite buffer).



(a) V^v

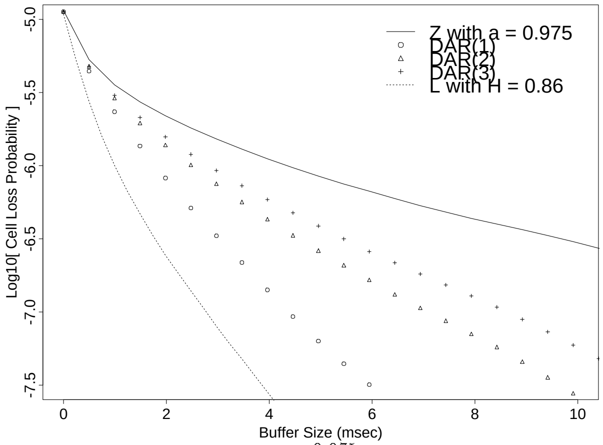


(b) Z^a

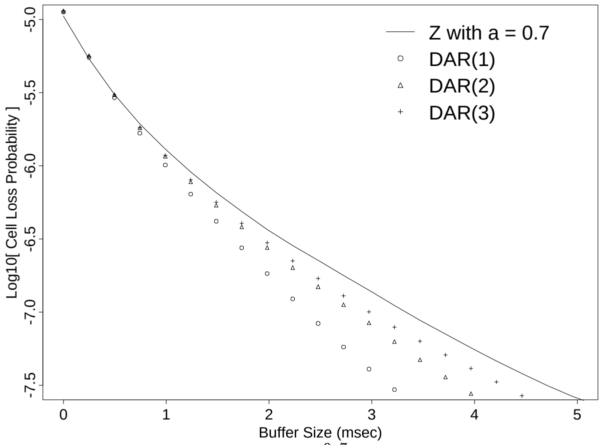
Figure 8: Cell loss rates (CLRs) of V^v and Z^a from simulation (finite buffer). $N = 30$ and $c = 538$ (cells/frame): effect of short- and long-term correlations on the cell loss (corresponding to Fig. 5).

6 Concluding Remarks

This work has looked at the impact of the autocorrelations on the loss behavior of real-time VBR video traffic with particular emphasis on the practical implications of LRD. There exists a growing concern about the potential impact of LRD on the network performance due to the following reasons: (i) previous studies on network performance analysis such as admission control have based on the assumption of Markov traffic models almost exclusively, whereas real video traces exhibit LRD; (ii) traditional Markovian models lack the capability of capturing this property in an economical way. This study reveals that such property is not practically important in determining cell loss rates (CLRs) under the realistic scenarios of ATM buffer dimensioning; short-term correlations (high-frequency behavior) have dominant impact on network performance. By analyzing and comparing the *critical time scales* (CTSs) of carefully designed VBR video models, we have been able to show that under realistic ranges of buffer size and CLR, the number of frame correlations which affect the cell loss is finite and small even in the presence of LRD. Simulation results verify that the two claims on the potential impact of the LRD property are not valid (*myths*) under the practical ranges of buffer size and CLR.



(a) $Z^{0.975}$



(b) $Z^{0.7}$

Figure 9: Comparison of the CLRs of Z^a, L, and matched DAR(p) from simulation. $N = 30$, $c = 538$ (cells/frame): efficacy of simple Markov models (corresponding to Fig. 6).

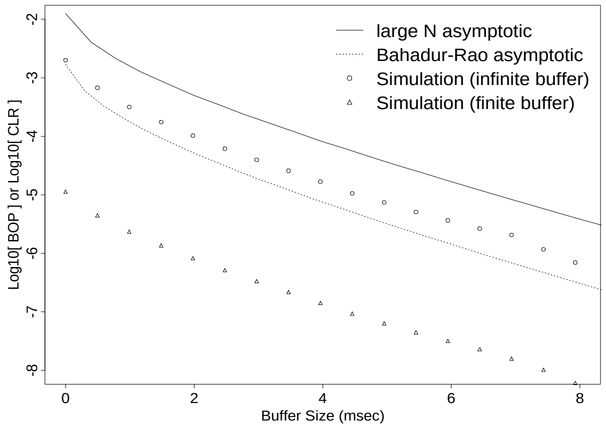


Figure 10: Accuracy of two large buffer asymptotics. The video model is DAR(1) matching $Z^{0.975}$. $N = 30$ and $c = 538$.

We conclude this work with brief discussions on the following relevant issues.

6.1 Effect of other marginal distributions

Strictly speaking, our results are based on the Gaussian marginal distribution, which is considered as having the "lightest" tail. Unfortunately, other distributions such as heavy-tailed frame sizes pose difficulty in mathematical analysis of queue performance. For the negative binomial marginal distribution, the same conclusion was obtained by Heyman and Lakshman using a slightly different approach [8]. In general, the required bandwidth for distributions with heavier tails will be larger than for Gaussian distribution to keep the operating point unchanged (within the practical ranges of buffer size and CLR). Once the bandwidth is properly chosen, and all the video models have the corresponding same marginal distribution of frame size, then the difference in buffer behavior will be again caused by the difference in their higher-order statistics, mainly the autocorrelations [13]. Therefore, we expect that our conclusions are unlikely to be significantly affected by other marginal distributions, though future work might be needed to support this.

6.2 Multiple-time-scale based traffic analysis

There is a significant amount of interest in capturing the time scale at which key statistics of traffic to network performance are to be evaluated [11, 12, 13, 16]. Our conclusions clearly indicate that traffic behavior after certain time scale (i.e., CTS) is not relevant to network performance such as CLR. As discussed in [16], the CTS is closely related with the cutoff frequency ω_c introduced in [11, 12, 13]. Note that a practical buffer size is about one frame duration (30 msec or so), whereas the time scale at which the LRD property of VBR video traffic begins to appear is an order of tens, hundreds, or even thousands frames. Further work is currently under way on finding CTS of various types of traffic sources including MPEG-coded video, and on applying it for traffic management in broadband networks.

Acknowledgement

The authors would like to thank anonymous reviewers for their constructive comments which significantly improved the presentation of this paper. They are also very grateful to Debasis Mitra for helpful discussions and insights he provided during the course of this work. This work was completed while the first author was on leave with Bell Laboratories, Murray Hill, NJ.

Appendix: B-R Asymptotic for N Gaussian LRD sources

In this appendix, we derive (6) using the Bahadur-Rao (B-R) asymptotic [16]. Recall that for an exact LRD process $L = \{L_1, L_2, \ldots\}$ with mean μ and variance σ^2, its autocorrelation function (ACF) is given by

$$\begin{aligned} r(k;T_s) &= \frac{1}{2} g(T_s)\nabla^2(k^{2H}) \qquad \text{for } k > 0 \\ &\approx H(2H-1)g(T_s)k^{2H-2} \end{aligned}$$

due to the equivalence of differencing and differentiation. Then for large m, we approximate $V(m)$ defined by (10) as

$$\begin{aligned} V(m) &\approx \sigma^2 \left[m + 2H(2H-1)g(T_s) \sum_{k=1}^{m} (m-k)k^{2H-2} \right] \\ &\approx \sigma^2 \left[m + 2H(2H-1)g(T_s) \int_1^m (m-x)x^{2H-2} \right] \\ &\approx \sigma^2 g(T_s) m^{2H}. \end{aligned} \tag{11}$$

The above approximation has been found to be accurate even for small m [4]. Substituting $V(m)$ in the rate function (8) with (11) yields,

$$I(c,b) \approx \inf_{m \geq 1} \frac{[b + m(c-\mu)]^2}{2\sigma^2 g(T_s) m^{2H}}.$$

For fixed b and c, let

$$h(x) \equiv \frac{[b + (c-\mu)x]^2}{\sigma^2 g(T_s) x^{2H}}.$$

Differentiating $h(x)$ and setting the result to 0 gives its root x^* at

$$x^* = \frac{H}{(1-H)(c-\mu)} b \approx m_b^*.$$

Then, $I(c,b)$ is given by

$$I(c,b) \approx \frac{(c-\mu)^{2H}}{2g(T_s)\sigma^2 \kappa(H)^2} b^{2-2H}.$$

Employing $B = Nb$ and substituting the result to (7) yields (6).

References

[1] R. G. Addie, M. Zukerman, and T. Neame. Fractal traffic: Measurements, modeling, and performance evaluation. In *Proc. IEEE INFOCOM*, Boston, MA, 1995.

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Comm.*, 43:1566–1579, 1995.

[3] C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a buffer handling many traffic sources. To appear in *J. App. Prob.*, 1995.

[4] D. R. Cox. Long-range dependence: A review. In H. A. David and H. T. Davis, editors, *Statistics: An Appraisal*, pages 55–74. The Iowa State University Press, Ames, Iowa, 1984.

[5] N. G. Duffield, J. T. Lewis, and N. O'Connell. Predicting Quality of Service for traffic with long-range fluctuations. In *Proc. ICC*, Seattle, WA, 1995.

[6] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra, and A. Weiss. Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing. *IEEE JSAC*, 13:1004–1016, 1995.

[7] M. W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar VBR video traffic. In *Proc. ACM SIGCOMM*, London, England, 1994.

[8] D. P. Heyman and T. V. Lakshman. What are the implications of long-range dependence for VBR-video traffic engineering? To appear in *IEEE/ACM Trans. Net.*, 1996.

[9] P. Jacobs and P. Lewis. Discrete time series generated by mixtures III: Autoregressive processes (DAR(p)). Technical Report NPS55-78-022, Naval Postgraduate School, 1978.

[10] W. E. Leland et al. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Net.*, 2:1–15, 1994.

[11] S. Q. Li, S. Chong, and C. L. Hwang. Link capacity allocation and network control by filtered input rate in high-speed networks. *IEEE/ACM Trans. Net.*, 3:678–692, 1995.

[12] S. Q. Li and C. L. Hwang. Queue response to input correlation functions: Continuous spectral analysis. *IEEE/ACM Trans. Net*, 1:678–692, 1993.

[13] S. Q. Li and C. L. Hwang. Queue response to input correlation functions: Discrete spectral analysis. *IEEE/ACM Trans. Net.*, 1:522–533, 1993.

[14] N. Likhanov, B. Tsybakov, and N. D. Georganas. Analysis of an ATM buffer with self-similar ("fractal") input traffic. In *Proc. IEEE INFOCOM*, Boston, MA, 1995.

[15] S. B. Lowen and M. C. Teich. Estimation and simulation of fractal stochastic point processes. *Fractals*, 3:183–210, 1995.

[16] M. Montgomery and G. De Veciana. On the relevance of time scales in performance oriented traffic characterizations. In *Proc. IEEE INFOCOM*, San Francisco, CA, 1996.

[17] I. Norros. A storage model with self-similar input. *Queueing Systems*, 16:387–396, 1994.

[18] M. Parulekar and A. M. Makowski. Buffer overflow probabilities for a multiplexer with self-similar input. In *Proc. IEEE INFOCOM*, San Francisco, CA, 1996.

[19] B. K. Ryu and S. B. Lowen. Point process approaches to the modeling and analysis of self-similar traffic: Part I - Model construction. In *Proc. IEEE INFOCOM*, San Francisco, CA, 1996.

[20] Bong K. Ryu. *Fractal Network Traffic: From Understanding to Implications*. PhD thesis, Columbia University, 1996. CTR Technical Report, CU/CTR/TR 448-96-14, URL: http://www.ctr.columbia.edu/~ryu.

[21] M. S. Taqqu. Self-similar processes. In S. Kotz and N. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 8. Wiley, New York, 1987.

[22] A. Weiss. An introduction to Large Deviations for communication networks. *IEEE JSAC*, 13:938–952, 1995.

[23] W. Willinger. Traffic modeling for high-speed networks: theory versus practice. In *Stochastic Networks*, IMA Volumes in Mathematics and Its Applications. Springer-Verlag, New York, 1994.