Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified Approach

Changcheng Huang Michael Devetsikiotis Ioannis Lambadaris A. Roger Kaye

Department of Systems and Computer Engineering
Carleton University
1125 Colonel By Drive
Ottawa, Canada K1S 5B6

Abstract

Variable bit rate (VBR) compressed video is expected to become one of the major loading factors in high-speed packet networks such as ATM-based B-ISDN. However, recent measurements based on long empirical traces (complete movies) revealed that VBR video traffic possesses *self-similar* (or *fractal*) characteristics, meaning that the dependence in the traffic stream lasts much longer than traditional models can capture.

In this paper, we present 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. We also present simulation results using synthetic data and compare with results based on empirical video traces.

Furthermore, we extend the application of efficient estimation techniques based on *importance sampling* that we had used before only for simple fractal processes. We use importance sampling techniques to efficiently estimate low probabilities of packet losses that occur when a multiplexer is fed with synthetic traffic from our self-similar VBR video model.

1 Introduction

An important advantage of packet switched networks (e.g., ATM-based B-ISDN networks), is that such networks support variable bit rate (VBR) connections, thus allowing efficient statistical multiplexing of bursty traffic. Video sources (coders) generate inherently VBR traffic, however, in order to transmit video information in circuit-switched networks, the variable content of moving pictures has to be coded in constant bit rate (CBR) form, resulting in inefficient bandwidth utilization and variable picture quality.

Due to the advantages of VBR video transmission and the packet-switched nature of ATM, and given the development of highly-sophisticated compression techniques for video sources, VBR compressed video traffic is expected to become one of the main loading components in future B-ISDN networks. However, the high bandwidth and burstiness of VBR video traffic, can make network design and management difficult to perform. Effective design and performance analysis depend on accurate modeling of the various traffic types. Among bursty traffic types, VBR video sources are arguably among the most important and demanding to model, due to their bandwidth fluctuation and autocorrelation, as well as their complex generation scheme (coding algorithm). Numerous studies have been conducted on issues of video coding, transmission over packet networks, and related modeling and performance analysis topics, see for example [10, 24, 27, 30, 32, 28] and references within.

Traditional models based on Markovian structures (e.g., MMPP, IBP, etc.) have been widely used to statistically approximate VBR video traffic. All these models have in common an asymptotically exponential decay of the autocorrelation function and a rapidly decaying marginal distribution tail. Furthermore they lack a systematic way of simultaneously fitting both the empirical marginal distribution and the autocorrelation function.

In a series of papers (see [22] and references within), B. Melamed and colleagues at NEC USA, Inc., developed the TES (Transform-Expand-Sample) modeling technique which can capture both the marginal distribution and the autocorrelation structure of empirical records. The TES approach was used to model transmission of VBR video traffic over high-speed networks also in [15, 29]. A composite TES-based model of the "Star Wars" sequence was presented in [21].

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 (weeks of computer processing time are required at this time to generate statistics from fully compressed, full-length movies).

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* (*fractal*) models [19, 18]. The crucial feature of self-similar processes is that they exhibit *long range dependence* (LRD), that is, their autocorrelation function decays less than exponentially fast and is non-summable. This is in contrast to traditional stochastic models, all of which exhibit *short range dependence* (SRD), i.e., have an autocorrelation function that decays exponentially or faster. The serious implication for ATM network design is that, conclusions based on traditional models may not be applicable under self-similar traffic. Recent studies on self-similar traffic have shown that the LRD structure may have a significant impact on queue-

Presented at ACM SIGCOMM '95, *Cambridge, MA, August, 1995*

ing performance [6, 23, 1, 13].

In [7] the authors presented a detailed statistical analysis of a 2-hour long empirical VBR video trace ("Star Wars"). 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 [11] that provided the LRD behavior. However, explicit modeling of the SRD structure was left for future work.

Although an ARIMA(p,d,q) model [12] can be used to model both LRD and SRD at the same time, it may be difficult to obtain accurate estimates of the p and q parameters required for the generation of traces with arbitrary marginals. This fact motivated us to develop modeling techniques that may capture the autocorrelation structure directly. In this paper, we extend the work in [7] and present a unified approach which, in addition to modeling the marginal distribution of empirical records, also models directly both the SRD and LRD empirical autocorrelation structures. While here we utilize MPEG-1 compressed VBR video, the approach itself can be readily applied to other VBR video compression schemes (e.g., JPEG, MPEG-2, H.261). Briefly, we generate a background self-similar Gaussian process with both LRD and SRD explicitly incorporated, we use a histogram-based inversion technique to generate a foreground process with the marginal distribution of the empirical data, and systematically iterate until the SRD part of the foreground process matches that of the empirical stream. We also prove that the value of the Hurst parameter H is not affected by a large family of transformations, a result we use above to synthesize self-similar traffic with arbitrary marginal distributions.

Furthermore, we extend the application of efficient estimation techniques based on *importance sampling* that were used in [13] for simple fractal Gaussian noise (FGN) processes. Here we use importance sampling techniques to efficiently estimate the probability of rare packet losses that occur when a multiplexer is fed with synthetic traffic from our self-similar VBR video model.

This paper is organized as follows: In Section 2 we provide a brief introduction to self-similar traffic models. We describe our unified approach to modeling VBR video traffic using both SRD and LRD components in Section 3. In Section 4 we present simulation results on the performance of a queueing system and compare with simulations based on the empirical trace. Conclusions are provided in Section 5. The proof of the invariance of the Hurst parameter under certain transformations is given in Appendix A. In Appendix B we outline the importance sampling technique we use in order to obtain efficient estimates of low cell loss probabilities under self-similar traffic.

2 Self-Similar Traffic Models and Trace Generation

Extensive measurements of real traffic data [18], have led to the conclusion that Ethernet traffic cannot be sufficiently represented by traditional models, but instead can be more accurately matched by *self-similar* (*fractal*) models [19, 20]. More recently, variable-bit-rate (VBR) video traffic was also found to exhibit self-similar characteristics, similarly to LAN traffic [2].

The crucial feature of self-similar processes is that they exhibit *long range dependence* (LRD), that is, their autocorrelation function $r(k)$ decays less than exponentially fast, and is non-summable, i.e., $r(k) \sim k^{-\beta}$, as $k \to \infty$, for $0 < \beta \leq 1$ (the quantity $H = 1 - \frac{\beta}{2}$ is called the *Hurst* parameter). This is in contrast to traditional stochastic models, all of which exhibit *short range dependence* (SRD), i.e., have an autocorrelation function that decays exponentially or faster. For formal definitions of *self-similarity*, *second-order self-similarity*, and *asymptotical second-order self-similarity* the interested reader can see [18] and references therein. The serious implication for ATM network design is that conclusions based on traditional models may not be applicable under self-similar traffic models.

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. The advantage of F-ARIMA models is that they can model both long time dependence and short time dependence at the same time [11].

Clearly, generation of long synthetic traces from self-similar processes poses significant difficulties, due to their long range dependence. In the following paragraphs we briefly describe Hosking's procedure [12] for generating traces from self-similar Gaussian processes. For such a process $\mathbf{X}$ with mean $m = 0$, the conditional mean and variance of X_k, given the past values $x_{k-1}, x_{k-2}, \ldots, x_0$, may be written as [26]:

$$m_k = \mathrm{E}(X_k | x_{k-1}, x_{k-2}, \ldots, x_0) = \sum_{j=1}^{k} \phi_{kj} x_{k-j} \tag{1}$$

$$v_k = \mathrm{var}(X_k | x_{k-1}, x_{k-2}, \ldots, x_0) = \sigma^2 \prod_{j=1}^{k} (1 - \phi_{jj}^2) \tag{2}$$

Here ϕ_{jj} is the jth partial correlation coefficient of $\{X_k\}$ and the ϕ_{kj} are partial linear regression coefficients. For simulating a sample $\{x_0, x_1, \ldots, x_{n-1}\}$ of size n from a self-similar Gaussian process with correlation $r(k)$, we have the following algorithm:

1. Generate a starting value x_0 from a Gaussian distribution $N(0, v_0)$. Set $N_0 = 0$, $D_0 = 1$.

2. For $k = 1, \ldots, n-1$, calculate ϕ_{kj}, $j = 1, \ldots, k$, recursively via the equations

$$N_k = r(k) - \sum_{j=1}^{k-1} \phi_{k-1,j} r(k) \tag{3}$$

$$D_k = D_{k-1} - N_{k-1}^2 / D_{k-1} \tag{4}$$

$$\phi_{kk} = N_k / D_k \tag{5}$$

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} \quad j = 1, \ldots, k-1 \tag{6}$$

Calculate $m_k = \sum_{j=1}^{k} \phi_{kj} x_{k-j}$ and $v_k = (1 - \phi_{kk}^2) v_{k-1}$. Generate x_k from the Gaussian distribution $N(m_k, v_k)$.

3 Modeling VBR Video

As pointed out by Hosking [12], the above method is applicable to any causal Gaussian process as long as the correlation function $r(k)$ is known. While Beran *et al.* [2] have shown that VBR video traffic has self-similar characteristics, there is strong evidence that VBR video possesses significant SRD components also. Furthermore, in [7] it was shown that the marginal distribution of VBR video possesses a long tail which is far from the Gaussian distribution.

Coder	MPEG-1
Duration	2 hours, 12 minutes, 36 seconds
Number of frames	238,626
Frame dimensions	320x240 pixels
Resolution	8 bits/pixel (3-band color)
Format	YUV colorspace, CCIR 601-2
Frame rate	30 per second
Slice rate	15 per frame

Table 1: Parameters of compressed empirical video sequence.

In this paper, our goal is to generate a process with marginal distribution and autocorrelation function that closely match the corresponding functions of empirical traces. In the following sections, we will outline our approach that is based on experiments involving a real video trace.

Most often in the past, long video traffic traces have been taken from the "Star Wars" movie. In this paper, we use approximately two hours of video from the movie "Last Action Hero". The movie was initially encoded using the MPEG-1 algorithm [17, 16], with a hardware *intraframe* MPEG-1 encoder on a Sun SPARC 20 computer [31]. The movie was then decompressed and re-encoded with both intraframe and *interframe* coding, using the PVRG-MPEG 1.1 software codec [25]. A summary of the parameters of the empirical trace is given in Table 1.

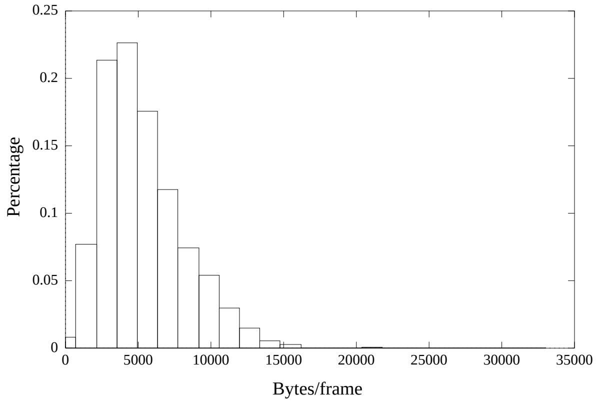

Figure 1: Empirical distribution function for "Last Action Hero".

3.1 Generation of a Self-Similar Process with an Arbitrary Marginal Distribution

Let X_k be a Gaussian process with zero mean, unity variance and autocorrelation function $r(k)$. Let $F_X(x)$ be its marginal cumulative probability function. Let $F_Y(y)$ be a marginal cumulative probability function corresponding to a process Y_k. Then we can generate the process Y_k with the desirable marginal cumulative probability function $F_Y(y)$ from the process X_k by using the following transformation [22, 7]:

$$Y_k = h(X_k) = F_Y^{-1}(F_X(X_k)) \quad k = 1, 2, \ldots \tag{7}$$

In real modeling procedures, $F_Y(y)$ can be obtained either by modeling an empirical distribution using parametric mathematical functions or, as we do in our approach, by inverting the empirical distribution directly. An important issue however, regarding this approach is that if the process $\mathbf{X}$ is a self-similar Gaussian process with Hurst parameter H, then the nature of the process $\mathbf{Y}$ may not be known. In the appendix at the end of the paper we show that, under general conditions, the process $\mathbf{Y}$ will be a self-similar process having the same Hurst parameter with process $\mathbf{X}$.

The empirical distribution function and the transform $h(.)$ corresponding to the trace of "Last Action Hero" is shown in Fig. 1 and 2.

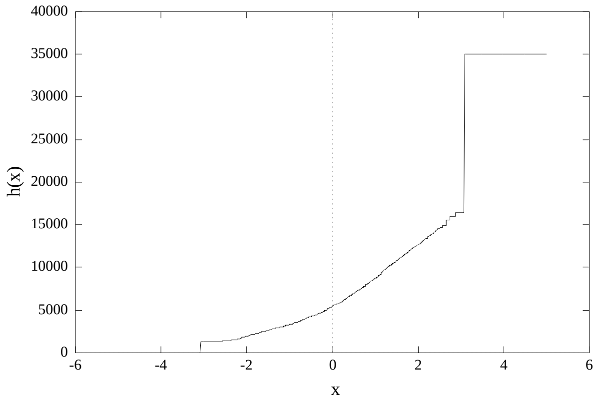

Figure 2: Transform function $h(X)$ that converts a normal distribution to the marginal distribution of the "Last Action Hero" trace.

3.2 Generation of a Process with both LRD and SRD

In the following parts, we describe our approach to modeling both VBR video with both LRD and SRD in four steps:

Step 1: Estimation of the Hurst parameter H:

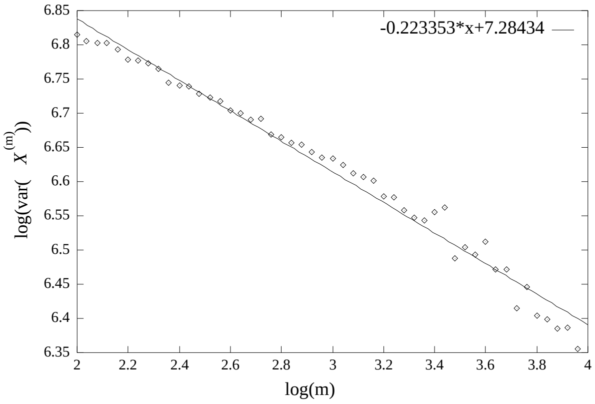

Figure 3: Variance-time plot for "Last Action Hero".

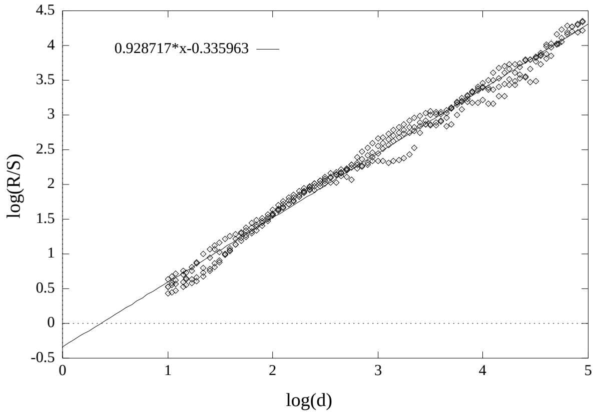

Figure 4: Pox diagram of R/S for "Last Action hero".

From the different approaches recommended for estimating the Hurst parameter in [18], we use here the methods of *variance-time plots* and *R/S analysis*.

For a self-similar process $\mathbf{X}$ (denoting in the following the number of bytes per video frame), the variance of the aggregated processes $X^{(m)}$, defined by

$$X_k^{(m)} = 1/m(X_{km-m+1}+\ldots+X_{km}), \;\; k \geq 1, \;\; m = 1, 2, 3, \ldots$$

decreases linearly (for large m) in log-log plots against m. The *variance-time plots* are obtained by plotting the function $\log(\mathrm{var}(\mathrm{X}^{(m)}))$ against $\log(m)$ and by fitting a simple least squares line through the resulting points in the plane, ignoring the small values for m. An estimate $\hat{\beta}$ of β will be the absolute value of the asymptotic slope of the line fit. Values of the estimate $\hat{\beta}$ between 0 and 1 suggest self-similarity, and an estimate for the degree of self-similarity is $\hat{H} = 1 - \hat{\beta}/2$. Fig.3 shows the variance-time plot for the empirical trace of "Last Action Hero". Logarithms are taken to base 10. An estimate for the corresponding Hurst parameter is $\hat{H} = 0.89$.

The R/S analysis is based on the *Hurst effect.* Briefly, for a given set of observations $(X_k : k = 1, 2, \ldots, n)$ with sample mean $\bar{X}(n)$ and sample variance $S^2(n)$, the rescaled adjusted range or the "R/S statistic" is given by

$$\begin{aligned} R(n)/S(n) &= 1/S(n)[\max(0, W_1, \ldots, W_n) \\ &- \min(0, W_1, \ldots, W_n)] \end{aligned} \tag{8}$$

with $W_k = (X_1 + X_2 + \ldots + X_k) - k\bar{X}(n), k = 1, 2, \ldots, n$. For self-similar processes, we have the following relation [14]:

$$E[R(n)/S(n)] \sim cn^H, \;\; as\ n \longrightarrow \infty \tag{9}$$

Given a sample of N observations $(X_k : k = 1, 2, 3, \ldots, N)$, one subdivides the whole sample into K non-overlapping blocks and computes the rescaled adjusted range $R(t_i, n)/S(t_i, n)$ for each of the new "starting points" $t_1 = 1, t_2 = N/K + 1, \ldots$ which satisfy $(t_i - 1) + n \leq N$. Here, the R/S statistic $R(t_i, n)/S(t_i, n)$ is defined as in (8) with W_k replaced by $W_{t_i+k} - W_{t_i}$ and $S^2(t_i, n)$ being the sample variance of $X_{t_i+1}, X_{t_i+2}, \ldots, X_{t_i+n}$. Next, we plot $\log(R(t_i, n)/S(t_i, n))$ versus $\log(n)$. This plot is the rescaled adjusted range plot. An estimate $\hat{H}$ is given by a least squares fit. For the empirical trace of "Last Action Hero", the corresponding plot is shown in Fig. 4. An $\hat{H} = 0.92$ was determined.

Combining the results of above two approaches, we decided to set $\hat{H} = 0.9$ and $\hat{\beta} = 0.2$ for the empirical trace of "Last Action Hero".

Step 2: Modeling the autocorrelation function:

The autocorrelation resulting from the actual empirical trace of "Last Action Hero" movie is shown in Fig. 5. Upon inspection of the plot it is evident that a "knee" around lag 60 to 80 exists. For lags less than the "knee" we observe that the autocorrelation decreases relatively fast thus indicating a short term dependence. When the lag is larger than the "knee", we can observe a slowly decreasing autocorrelation indicating long range dependence. The rapidly decaying part of the autocorrelation can be approximated by superimposing a number of decreasing exponentials of the form $\exp(-\lambda_i k)$ with different rates λ_i. Furthermore the part corresponding to long range dependence can be approximated by $Lk^{-\beta}$, where L is a constant. We can now write the following:

$$\begin{aligned} R(k) &= Lk^{-\beta} I(k \geq K_t) \\ &+ \sum_{i=1}^{j} w_i \exp(-\lambda_i k) I(k < K_t), k = 1, 2, \ldots \end{aligned} \tag{10}$$

$$\sum_{i=1}^{j} w_i = 1 \tag{11}$$

$$LK_t^{-\beta} = \sum_{i=1}^{j} w_i \exp(-\lambda_i K_t) \tag{12}$$

where K_t is the lag value corresponding to the "knee", $I(.)$ is the indicator function. Then we can use a least square fitting to determine the coefficients for the SRD and LRD parts of the autocorrelation. Such a fit is shown in Fig. 6. In our case we used one exponential for modeling the SRD and finally obtained the following expression for the autocorrelation:

$$\hat{r}(k) = \exp\left(-0.00565k\right) I(k < K_t) + 1.59k^{-0.2} I(k \geq K_t) \tag{13}$$

As will be illustrated later on by simulation experiments, the exponential component was necessary since the polynomial component decays too fast in the early lags.

Step 3: Measurement of the "attenuation" factor:

Let $r_h(k)$ and $r(k)$ be the autocorrelation functions of the processes $\mathbf{Y}$ and $\mathbf{X}$ respectively. By using Hosking's technique employing the autocorrelation $\hat{r}(k)$ we generate the

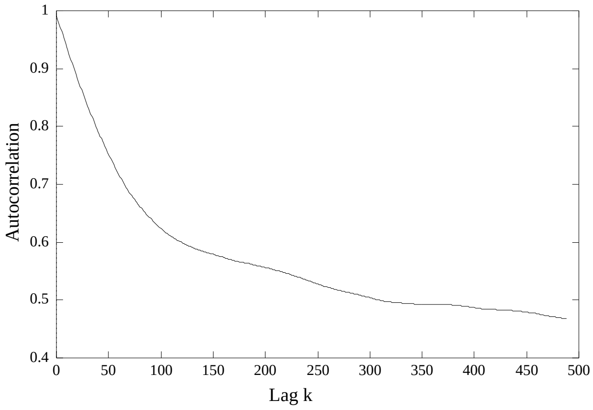

Figure 5: The estimated autocorrelation function of "Last Action Hero".

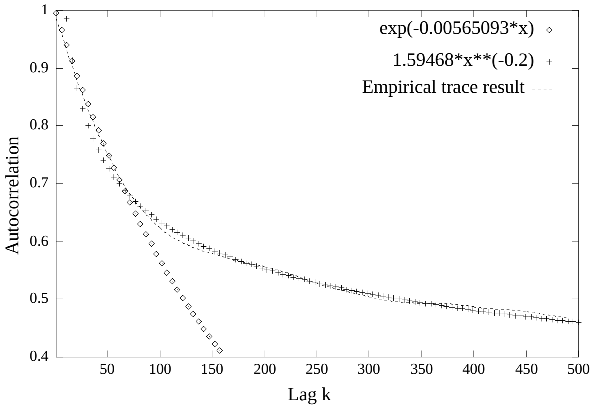

Figure 6: Autocorrelation fitting result.

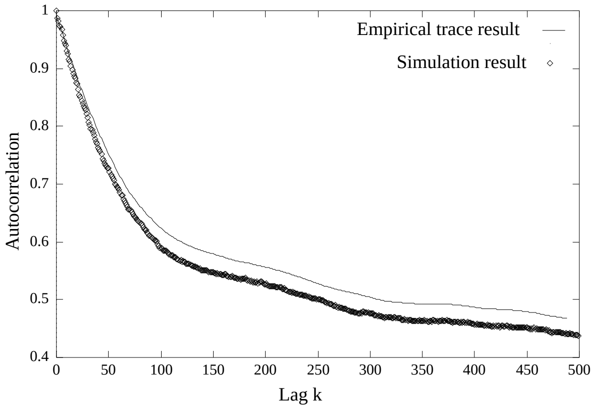

Figure 7: The autocorrelation functions of $\mathbf{X}$ and $\mathbf{Y}$, illustrating the attenuation factor, a.

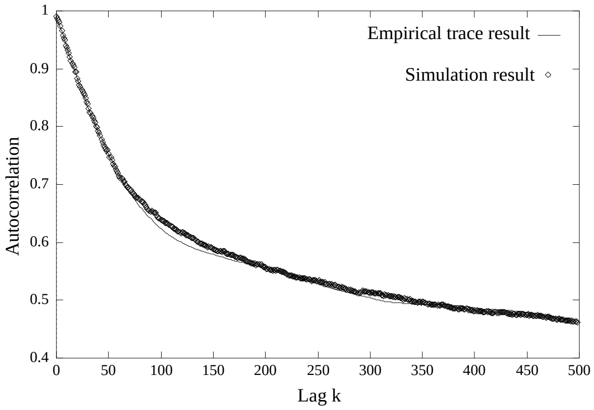

Figure 8: Autocorrelation of the empirical trace and the final simulated process.

process $\mathbf{X}$. The process $\mathbf{Y}$ is then generated by equation (7). It can be shown (refer to Appendix A) that $r_h(k) = ar(k)$, as $k \longrightarrow \infty$, where a is a constant in $(0, 1]$. We call a the "attenuation" factor. In Fig. 7 we show the autocorrelations of the process $\mathbf{X}$ and the process $\mathbf{Y}$. By measuring the ratio $r_h(k)/r(k)$ at a large lag we found $a = 0.94$.

Step 4: Generation a process with the desired autocorrelation:

Let $r(k) = \hat{r}(k)/a$, for $k \geq K_t$. Then for the short term part, we solve the following equation to get the rate λ:

$$\exp\left(-\lambda K_t\right) = \hat{r}(K_t)/a \qquad (14)$$

and we let $r(k) = \exp(-\lambda k) \quad for\ k < K_t$. We decided to set $K_t = 60$ based on the intersection point of the two fitting curves. We then generate process $\mathbf{X}$ using Hosking's method and the process $\mathbf{Y}$ using equation (7). The final autocorrelation result of process $\mathbf{Y}$ is shown together with the empirical autocorrelation of Fig. 8, indicating a satisfactory match.

3.3 Modeling VBR Video with Interframe Compression

In this section, we generalize our approach to the modeling of VBR video with both intraframe and interframe compression. The codec we used is the PVRG-MPEG 1.1 software codec based on the Santa Clara 1991 draft of MPEG-1 [25].

The MPEG-1 coder [16] consists of five stages: a motion compensation stage, a transformation stage, a lossy quantization stage, and two lossless coding stages. The motion compensation stage subtracts the current image from shifted view of the previous image if they are both alike. The transform concentrates the information energy into the first few transform coefficients, the quantizer causes a controlled loss of information, and the two coding stages further compress the data closer to symbol entropy.

A typical MPEG-1 sequence consists of three separate parts: a series of intraframes (I frames), which are image frames coded individually without any temporal prediction; a series of forward predicted frames (P frames), interspersed between these I frames; and bidirectionally predicted frames (B frames) interspersed between the forward predicted frames and the intraframes. A typical frame sequence in a GOP (group of pictures) is as follows:

I B B P B B P B B P B B I ...

Our approach to modeling interframe-encoded MPEG-1 VBR video is to generate a single stationary background process $\mathbf{X}$ with both SRD and LRD structures and then generate the foreground process using three different transforms $h_I(X)$, $h_B(X)$ and $h_P(X)$ based on the histograms of I, B and P frames, respectively, according to above frame sequence structure. The PVRG-MPEG 1.1 software codec used in our experiments produced video traffic in which I frames appear periodically once every 12 frames.

We model the composite I-B-P video traffic as follows:

Step 1: Isolate I frames only and model the I-frames process according to the previous sections;

Step 2: Rescale the estimated autocorrelation of the I frames :

$$r(k) = r_I(k/K_I) \tag{15}$$

where, $r_I(k)$ is the autocorrelation and $K_I = 12$ is the period of I frames.

The similarity between the synthetic and real data trace is evaluated by means of the corresponding estimates of autocorrelation functions and marginal distribution histograms. Figures 9, 10, and 11 show the foreground autocorrelation of the synthetic trace in comparison to the autocorrelation of the original empirical trace from "Last Action Hero". Figures 12 and 13 compare the marginal distributions of the model process versus the empirical data trace, using a histogram and a Q-Q plot, respectively. The agreement shown in the figures above support the use of our approach for modeling complex traffic streams.

The possibility of establishing an automatic search for the best background autocorrelation structure is currently under investigation.

4 Cell Loss Studies in an ATM Environment

In this section, we focus on the simulation of single buffer queue with a single arrival process and a single deterministic server. We consider a slotted-time single server queue with deterministic service rate μ and a stationary self-similar arrival process $\mathbf{Y}$, with Y_k representing the number of arriving cells within the kth time slot. Here, without loss of generality, we assume Y_k can take any non-negative real value. Let Q_k denote the size of the queue at time $k = 0, 1, \ldots$. We are interested in calculating the probability (transient and steady-state) that Q_k will exceed a given value b.

Assuming $Q_0 = 0$, we have the following Lindley equa-

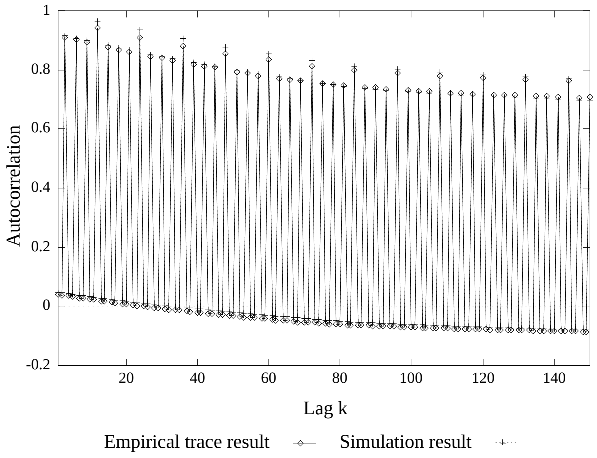

Figure 9: Comparison of autocorrelations of simulation process and empirical trace (lags 1 to 150).

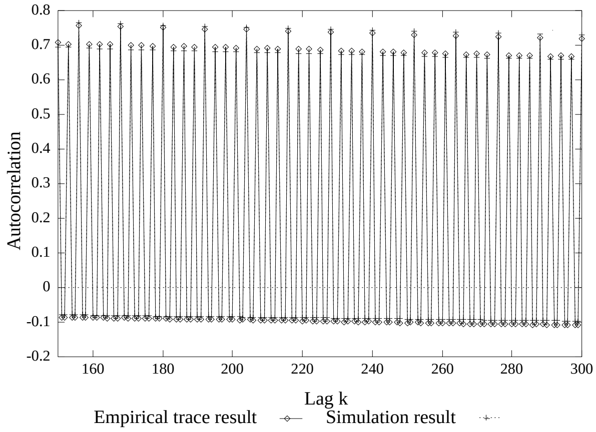

Figure 10: Comparison of autocorrelations of simulation process and empirical trace (lags 151 to 300).

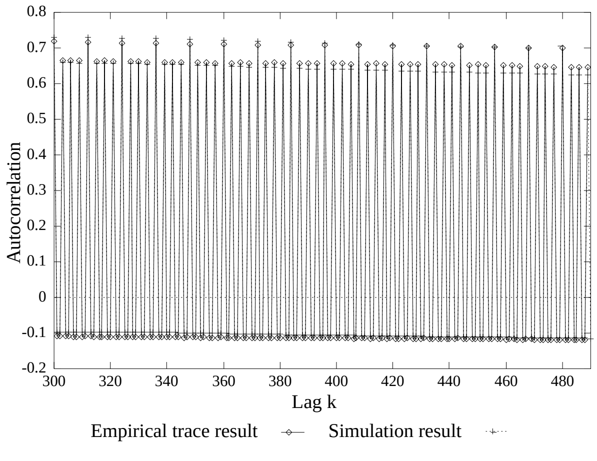

Figure 11: Comparison of autocorrelations of simulation process and empirical trace (lags 301 to 490).

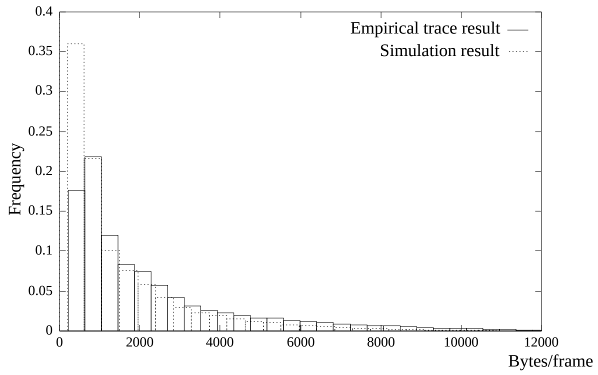

Figure 12: Comparison of histograms of simulation process and empirical trace.

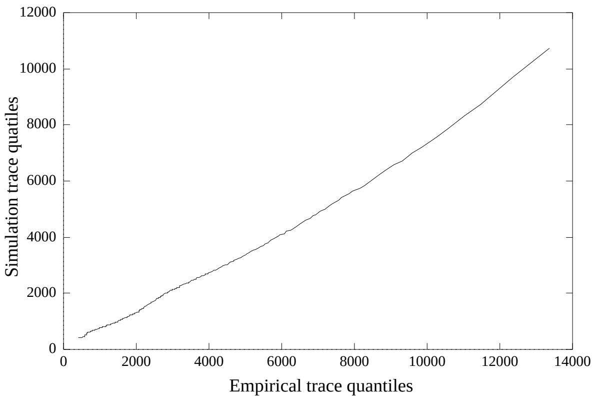

Figure 13: Q-Q plot comparing the marginal distributions of the simulation process and the empirical trace.

tion [4]:

$$Q_k = \langle Q_{k-1} + Y_k - \mu \rangle^+ = \langle Q_{k-1} + Z_k \rangle^+, \text{ for } k = 1, 2, \dots \tag{16}$$

where we define the process $\mathbf{Z} = \{Z_k : Z_k = Y_k - \mu, k = 1, \dots\}$ as the *work load* process. Now define the *total work load process* $\mathbf{W}$ as $\{W_k : W_k = \sum_{i=1}^{k} Z_i, \; k = 1, 2, \dots\}$. Then, assuming that $\mathrm{E}[\mathrm{Y_k}] < \mu$, $\mathbf{W}$ is a stationary increment process, and since $\mathbf{Y}$ is a stationary process, we have

$$\Pr(Q_k > b) = \Pr(\sup_{0 \le i \le k} W_i > b), \text{ for } k = 0, 1, 2, \dots \tag{17}$$

Our goal is to efficiently simulate the single buffer queue and estimate $P(Q_k > b)$ which may be viewed as the probability of buffer overflow. As the buffer size b increases it is clear that the frequency of the event of buffer overflow will drop, necessitating longer simulations. This observation along with the fact that the generation of self similar traffic using Hosking's method is computationally quite demanding, motivated us to use the method of importance sampling in order to built efficient fast simulations. For more details on importance sampling and fast simulations the reader is referred to Appendix B.

The probability $\Pr(Q_k > b)$ can be estimated by observing N iid replications of the realization $w_1^{(n)}, \dots, w_k^{(n)}$ of $\mathbf{W}$, for $n = 1, \dots, N$. Let $L^{(n)}$, $n = 1, \dots, N$, denote the corresponding likelihood ratio for each replication. Then, the following simulation procedure can be used for estimating $\Pr(Q_k > b)$ (see also Appendix B):

1. Initialize $i = 1, n = 1$;

2. Generate a sample point x_i by Hosking's method described in Section 3; generate the corresponding twisted sample point $x_i' = x_i + m^*$;

3. Generate a sample point y_i' from equation (7): $y_i' = h(x_i')$;

4. Generate a sample point w_i by replacing the process $\mathbf{X}$ with the process $\mathbf{X}'$ in the definition of total work load process;

5. If $w_i \le b$ and $i < k$, then repeat from step 2 with $i = i + 1$; otherwise continue with step 6;

6. If $w_i \le b$ and $i = k$, set $I_n = 0$ and go to step 8; otherwise continue with step 7;

7. Set $I_n = 1$ and calculate $L^{(n)} = L(i)$ via equations (42) to (48);

8. If $n = N$ estimate $\Pr(Q_k > b)$ by $\hat{P} = \frac{1}{N} \sum_{n=1}^{N} I_n L^{(n)}$; otherwise set $n = n + 1$, $i = 1$ and goto step 2.

Based on the above description, we can apply IS by suitably modifying (twisting) the mean of the arrival process. However, an efficient method to obtain a favorable (or near-optimal) twisted mean m^* remains to be devised. Analytical approaches to optimizing the form and amount of twisting for SRD models have been investigated in [8, 3, 9]. For the case of FGN processes, analytical arguments for optimizing the twisting process were given in [13]. However, after the transformation in (7), a closed-form optimization becomes intractable, therefore we resort here to the heuristic search approach, which is based on the fact that the IS estimator of $\Pr(Q_k > b)$ is always *unbiased*, while the sample path properties as well as the variance of the IS estimator are dramatically affected by the choice of twisting parameter values. Typically, if we observe estimates of $\Pr(Q_k > b)$ and its normalized variance as the twisting parameters change, the normalized variance exhibits a clear "valley" around the most favorable parameter values, which can be thus, approximately identified. This approach has been successfully

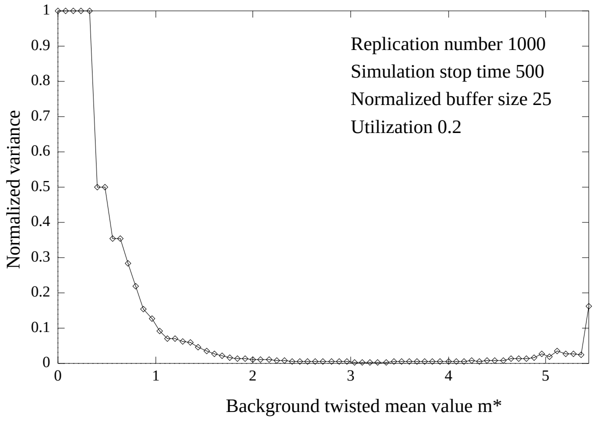

Figure 14: Plot of the estimated normalized variance of the estimator versus the mean value of background process twisting, m^*. Results correspond to a stopping time $k = 500$, utilization 0.2, buffer size $b = 25$, and 1000 replications.

applied to traditional (SRD) models (see [5] and references within) and to FGN processes in [13].

An optimal selection of the (twisted) mean will result in a greatly reduced variance of the estimator for $P(Q_k > b)$. A favorable (near-optimal) background (twisted) mean value can be found from plots such as the one shown in Fig. 14. For our experiments, we found the value 3.2 to be a near-optimal twisted mean value, resulting in a variance reduction of approximately 1000 (conversely, the required number of replications for the same accuracy is reduced by 1000). In the figures that follow, when we refer to buffer size we will essentially mean the normalized buffer size, i.e., the ratio of true buffer size to mean arrival rate.

All the simulations that we have described thus far have a transient nature in the sense that they provide an estimate of the probability of buffer overflow at a given time slot k given the initial conditions. It is of particular interest to decide on a simulation run length in order to achieve steady state results, i.e. for the case when $k \to \infty$. Fig. 15 shows the transient buffer overflow probability for a given buffer size b, corresponding to two initial buffer occupancy conditions, namely empty and full buffer. From this figure we can see that the transient time in a simulation may be reduced if the initial conditions are chosen properly. Since the generation of the background process $\mathbf{X}'$ may be computationally demanding, a small transient period may be highly desirable.

Fig. 16 shows approximately steady state results ($k = 2000$) for several service rates (corresponding to different system utilization values). Simulations were run using the empirical video trace as well. We should note however that all simulations involving synthetic traffic were based on 1000 independent replications for each different utilization and buffer size. Since only one empirical trace was available, it was impossible to perform independent replications for each simulation involving real data. Even if the real data were split into batches we would expect significant correlations between batches due to the self similar nature of the traffic. Therefore, simulations involving the empirical trace were based only on one (long) replication. Furthermore, the same empirical trace was used for simulating all different buffer sizes! Therefore, we may expect a slight disagreement between the results obtained using synthetic and empirical

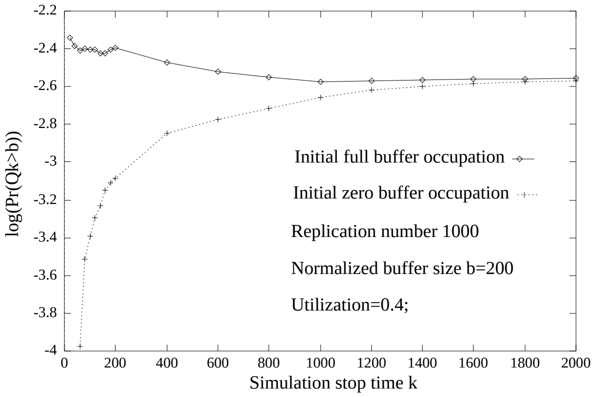

Figure 15: Transient buffer overflow probability, using 1000 replications, $b = 200$, and a utilization of 0.4.

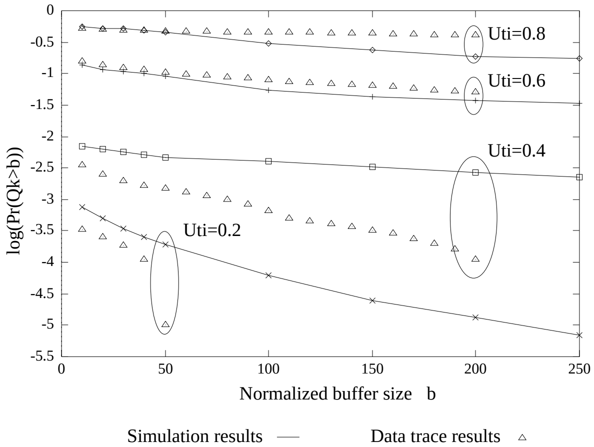

Figure 16: Overflow probability versus buffer size b, for different utilization values, using 1000 replications and $k = 10b$.

data. This is evident in Fig. 16 for the case with utilizations 0.8 and 0.6. For lower utilizations and larger buffer sizes however this disagreement is expected to be more profound as shown for the cases with utilizations 0.4 and 0.2 since on top of the abovementioned reasons the real data may not be long enough to provide accurate estimates.

Traditional models typically focus on the short range dependence structure. As shown in [6] and [13], the loss probability of a queueing system driven by an FGN process decays less than exponentially fast with respect to buffer size. In Fig. 17, we compare three models. The first video model possesses only the SRD and includes only the exponentially decaying part of the autocorrelation as it was derived in Section 3. The second model is the one exhibiting both the SRD and LRD. The third model captures only the LRD structure, based on a single FGN background process (i.e., there is no short-term exponential component). It is easy to see that for small buffer size the difference in the probability of buffer overflow is not significant, but as the buffer size increases the estimate based on the SRD model decays much faster than the one based on the model characterized by both LRD and SRD. Finally, as expected, although the third model exhibits the appropriate asymptotic behavior, the corresponding loss probability decays too fast for small buffer sizes.

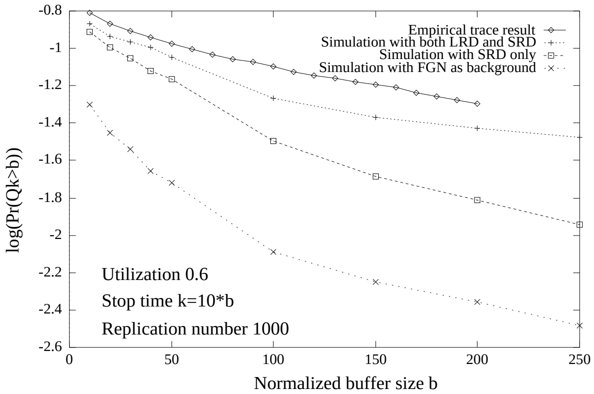

Figure 17: Overflow probability versus buffer size b for four cases: using the empirical trace, using the simulated model with both LRD and SRD, using a simulated model *without* LRD, and using a model with *only* LRD.

5 Conclusions

Effective design and performance analysis in high-speed networks depend on the accurate modeling of the various traffic types. Recent measurements based on long empirical traces (complete movies) revealed that VBR compressed video traffic possesses self-similar characteristics, meaning that the dependence in the traffic stream lasts much longer than traditional models can capture.

In this paper, we extended previous modeling approaches and presented a unified approach which, in addition to accurately modeling the marginal distribution of empirical records, also models directly *both* the short and the long-term empirical autocorrelation structures. Simulation results suggest that our approach is promising, and work is in progress for its further refinement.

Finally, we extended the application of efficient estimation techniques based on importance sampling, in order to efficiently estimate the probability of rare packet losses that occur when a multiplexer is fed with synthetic traffic from our VBR video model.

6 Acknowledgements

This research was supported by grants from the Telecommunications Research Institute of Ontario and the National Science and Engineering Research Council of Canada.

References

[1] R. Addie, M. Zukerman, and T. Neame. Performance of a Single Server Queue with Self Similar Input. In *Proc. IEEE ICC '95*, Seattle, June 1995.

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-Range Dependence in Variable-Bit-Rate Video Traffic. *IEEE Trans. Commun.*, 43(2/3/4):1566–1579, 1995.

[3] J. A. Bucklew. *Large deviation Techniques in Decision, Simulation, and Estimation.* John Wiley and Sons, 1990.

[4] J. W. Cohen. *The Single Server Queue.* North-Holland, 1982.

[5] M. Devetsikiotis and J. K. Townsend. Statistical Optimization of Dynamic Importance Sampling Parameters for Efficient Simulation of Communication Networks. *IEEE/ACM Trans. Networking*, 1(3), June 1993.

[6] N. G. Duffield and N. O'Connell. Large Deviations and Overflow Probabilities for the General Single-Server Queue, with Applications. Technical Report DIAS-STP-93-30, Dublin Institute for Advanced Studies, 1993.

[7] 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.

[8] P. W. Glynn and D. L. Iglehart. Importance Sampling for Stochastic Simulations. *Management Science*, 35(11):1367–1392, Nov. 1989.

[9] P. Heidelberger. Fast Simulation of Rare Events in Queueing and Reliability Models. In *Proc. of Performance '93*, Rome, Italy, October 1993.

[10] D. Heyman, T. V. Lakshman, A. Tabatabai, and H. Heeke. Modeling Teleconference Traffic from VBR Video Coders. In *Proc. IEEE ICC '94*, New Orleans, 1994.

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

[12] J. R. M. Hosking. Modeling Persistence in Hydrological Time Series Using Fractional Differencing. *Water Resources Research*, 20(12):1898–1908, 1984.

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

[14] H. E. Hurst. Long-Term Storage Capacity of Reservoirs. *Trans. of the Am. Soc. of Civil Eng.*, 116:770–799, 1951.

[15] M. R. Ismail, I. Lambadaris, M. Devetsikiotis, and A. R. Kaye. Modeling Prioritized MPEG Video Using TES and a Frame Spreading Strategy for Transmission in ATM Networks. In *Proc. IEEE INFOCOM '95*, Boston, April 1995.

[16] ISO. *MPEG-1 Specification.* CD 11172.

[17] D. LeGall. MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM*, 34(4), Apr. 1991.

[18] 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.

[19] B. B. Mandelbrot. *The Fractal Geometry of Nature.* Freeman, San Fransisco, 1983.

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

[21] B. Melamed and D. Pendarakis. A TES-Based Model for Compressed "Star Wars" Video. In *Proc. Comm. Theory Mini-Conf., IEEE Globecom '94*, San Fransisco, November 1994.

[22] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski. TES-Based Video Source Modeling For Performance Evaluation of Integrated Networks. *IEEE Trans. Commun.*, 42(10), Oct. 1994.

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

[24] P. Pancha and M. El Zarki. Bandwidth Allocation Schemes for Variable Bit Rate MPEG Sources in ATM Networks". *IEEE Trans. Circ. Syst. Video Tech.*, Vol. 3(3), June 1993.

[25] Portable Video Research Group, Stanford University. *PVRG-MPEG Codec 1.1*, June 1993.

[26] F. L. Ramsey. Characterization of the Partial Autocorrelation Function. *The Annals of Statistics*, 2(6):1296–1301, 1974.

[27] A. R. Reibman and B. G. Haskell. Constraints on Variable Bit Rate Video for ATM Networks". *IEEE Trans. Circ. Syst. Video Tech.*, Vol. 2(4), Dec. 1992.

[28] D. Reininger, D. Raychaudhuri, B. Melamed, B. Sengupta, and J. Hill. Statistical Multiplexing of VBR MPEG Compressed Video on ATM Networks. In *Proc. IEEE INFOCOM '93*, San Fransisco, Mar. 1993.

[29] C. M. Sharon, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Rate Control of VBR H.261 Video on Frame Relay Networks. In *Proc. IEEE ICC '95*, Seattle, June 1995.

[30] P. Skelly, M. Schwartz, and S. Dixit. A Histogram-Based Model for Video Traffic Behavior in an ATM Multiplexer. *IEEE/ACM Trans. on Networking*, 1(4), Aug. 1993.

[31] Sun Microsystems Computer Corporation. *SunVideo 1.0 User's Guide*, Oct. 1993.

[32] F. Yegenoglu, B. Jabbari, and Ya-Qin Zhang. Motion-Classified Autoregressive Modeling of Variable Bit Rate Video. *IEEE Trans. Circ. Syst. Video Tech.*, Vol. 3(1), Feb. 1993.

A Proof of the Invariance of Parameter H

We now prove the following theorem that was used in Section 3.2:

Theorem: Let $\mathbf{X} = \{X_i, i = 0, 1, \ldots\}$ be a zero mean, unity variance self-similar Gaussian process defined on a probability space $(\Omega, \mathcal{F}, P)$ with Hurst parameter H and $h : \mathbf{R} \longmapsto \mathbf{R}$ be a function measurable to Borel field. If $h^2(X)$ is integrable with respect to P and $E(h(X)X) \neq 0$, then the process $\mathbf{Y} = \{Y_i = h(X_i), i = 0, 1, \ldots\}$ is an asymptotically self-similar process with the same Hurst parameter H.

Proof: Without loss of generality, we assume $E(h(X)) = 0$. Let $r(k), k = 1, 2, \ldots$ be the autocorrelation function of $\mathbf{X}$. Let $\psi(k) = h(x_i)h(x_{i+k})\frac{1}{2\pi\sqrt{1-r^2(k)}}$. Since $h^2(X)$ is integrable with respect to P, it follows by a straightforward application of the Schwartz inequality that the autocovariance of $\mathbf{Y}$ is finite or

$$\psi(k)\exp\{-\frac{x_i^2 - 2r(k)x_i x_{i+k} + x_{i+k}^2}{2(1-r^2(k))}\} \tag{18}$$

is integrable with respect to Lebesgue measure.

Due to the symmetry of the zero mean Gaussian probability distribution, and the fact that $h^2(-X)$ is also integrable with respect to P we can easily conclude that the function

$$\psi(k)\exp\{-\frac{x_i^2 + 2r(k)x_i x_{i+k} + x_{i+k}^2}{2(1-r^2(k))}\} \tag{19}$$

is integrable with respect to Lebesgue measure.

Furthermore, since for any real numbers x and y we have $\exp(|xy|) < \exp(xy) + \exp(-xy)$ the function

$$\psi(k)\exp\{-\frac{x_i^2 - |2r(k)x_i x_{i+k}| + x_{i+k}^2}{2(1-r^2(k))}\} \tag{20}$$

is integrable with respect to Lebesgue measure. By the expansion of exponential function, we have

$$\exp\{|\frac{r(k)}{1-r^2(k)}x_i x_{i+k}|\} = \sum_{n=0}^{\infty}[|\kappa(k,n)|]^n/n! \tag{21}$$

where

$$\kappa(k,n) = [\frac{r(k)}{1-r^2(k)}x_i x_{i+k}]^n/n! \tag{22}$$

Therefore, the function

$$\sum_{n=0}^{\infty}\psi(k)|\kappa(k,n)|\exp\{-\frac{x_i^2 + x_{i+k}^2}{2(1-r^2(k))}\} \tag{23}$$

is integrable with respect to Lebesgue measure and furthermore,

$$\int\int\{\sum_{n=m}^{\infty}[|g(k,n)|]\exp\{-\frac{x_i^2 + x_{i+k}^2}{2(1-r^2(k))}\}\}dx_i dx_{i+k}$$
$$\longrightarrow 0 \quad \text{as} \quad m \longrightarrow \infty \tag{24}$$

where

$$g(k,n) = \psi(k)\kappa(k,n) \tag{25}$$

Therefore, the function

$$g(k,n)\exp\{-\frac{x_i^2 + x_{i+k}^2}{2(1-r^2(k))}\} \tag{26}$$

is integrable with respect to Lebesgue measure and its integral converges to zero as $n \longrightarrow \infty$. We now calculate the autocovariance of $\mathbf{Y}$ as follows:

$$\begin{aligned} \mathrm{cov}(\mathrm{Y_i}, \mathrm{Y_{i+k}}) &= \\ &= \int\int\psi(k)\exp\{-\frac{x_i^2 - 2r(k)x_i x_{i+k} + x_{i+k}^2}{2(1-r^2(k))}\}dx_i dx_{i+k} \\ &= \int\int\psi(k)\{\sum_{n=0}^{\infty}\kappa(k,n)\}\exp\{-\frac{x_i^2 + x_{i+k}^2}{2(1-r^2(k))}\}dx_i dx_{i+k} \end{aligned} \tag{27}$$

By use of the dominated convergence theorem and equation (24), we can write

$$\mathrm{cov}(\mathrm{Y_i}, \mathrm{Y_{i+k}}) =$$

$$= \sum_{n=0}^{\infty} \int\int g(k,n) \exp\{-\frac{x_i^2 + x_{i+k}^2}{2(1-r^2(k))}\} dx_i dx_{i+k}$$

$$= \gamma(k)\{\int h(x_i) x_i \exp\{-\frac{x_i^2}{2(1-r^2(k))}\} dx_i\}^2 + o(r(k))$$

$$\text{as} \quad k \longrightarrow \infty \tag{28}$$

where

$$\gamma(k) = \frac{r(k)}{2\pi[1-r^2(k)]^{3/2}} \tag{29}$$

If $r_h(k), k = 1, 2, \ldots$ denotes the autocorrelation function of $\mathbf{Y} = h(\mathbf{X})$, then as $k \longrightarrow \infty$, we have

$$\begin{aligned}
\frac{r_h(k)}{r(k)} &= \frac{\mathrm{cov}(\mathrm{Y_i}, \mathrm{Y_{i+k}})}{\mathrm{r(k)var(Y_i)}} \\
&= \frac{1}{2\pi} \frac{\{\int h(x_i) x_i \exp\{-\frac{x_i^2}{2}\} dx_i\}^2}{E(h^2(X_i))} \\
&= \frac{[E(h(X_i)X_i)]^2}{E(h^2(X_i))} \\
&= a
\end{aligned} \tag{30}$$

where $a = \frac{[E(h(X_i)X_i)]^2}{E(h^2(X_i))}$. Finally by using the Schwarz inequality it follows that

$$a \leq 1 \tag{31}$$

B Importance Sampling

B.1 Preliminaries

Let U be a random variable that has a probability density function $p(u)$ and consider estimating the probability P that U is in some set A, then

$$P = \int_{-\infty}^{\infty} I_A(t) p(t) dt = \mathrm{E}_p[I_A(U)] \tag{32}$$

where $I_A(\cdot)$ is the indicator function of event A. Assume that $p'(u)$ is another density function. Assuming that $p(u) = 0$ whenever $p'(u) = 0$ (*absolute continuity* condition), we have

$$\begin{aligned}
P &= \int_{-\infty}^{\infty} I_A(t) \frac{p(t)}{p'(t)} p'(t) dx \\
&= \mathrm{E}_{p'}[I_A(U)\frac{p(U)}{p'(U)}] = \mathrm{E}_{p'}[I_A(U)L(U)]
\end{aligned} \tag{33}$$

where $L(u) = p(u)/p'(u)$ is a *likelihood ratio* (*weight function*) and the notation p' denotes sampling from the density $p'(u)$. This equation suggests the following variance reduction estimation scheme which is called *importance sampling* (IS) (see [8] and references within): Draw N samples $u_1, \ldots, u_N$ using the density p'. Then, by equation (33), an unbiased estimate of P is given by

$$\hat{P}_N = \frac{1}{N} \sum_{n=1}^{N} I_A(u_n) L(u_n) \tag{34}$$

i.e., P can be estimated by simulating a random variable with a different density and then unbiasing the output $I_A(u_n)$ by multiplying with the likelihood ratio. We call $p'(u)$ the *twisted density*. Since any density can be used as the twisted density, the question arising is which is the *optimal* twisted density, i.e., which is the density that minimize the variance of $\hat{P}$. Although the unconstrained optimal density is easy to describe, implementing it is not practically feasible because it represents a tautology (i.e., requires knowledge of P).

Typically, the search for $p'(u)$ focuses on constrained or parametric solutions. A general rule for choosing a favorable twisted density is to make the likelihood ratio small on the set A. When A is a rare event under density $p(u)$, what we have to do is choose a density to make the event A more likely to occur. In doing this, we reduce the variance of the estimate $\hat{P}$. For more about the IS technique, see [8] and [3]. Importance sampling has been successfully applied to the simulation of various SRD processes. A variety of approaches, namely analytical, large deviation-based, and statistical have been proposed in order to choose $p'(u)$ ([8, 3, 5, 9] and references within).

B.2 Twisted Process and Likelihood Ratio

In order to apply importance sampling to efficiently simulate rare cell losses in an ATM multiplexer under VBR video traffic , we need to construct an appropriate twisted traffic stream, calculate the corresponding likelihood ratio, and choose optimal (or simply favorable) twisting parameter values. In [13] the twisted process and likelihood ratio were described for simulating FGN processes. Here, we extend those results for the case of a self-similar Gaussian process that serves as the *background* process for the generation of realistic VBR video traffic.

Let $\mathbf{X}$ be the background self-similar Gaussian process as defined in Section 3.2, with mean value $m = 0$. Define a new process $\mathbf{X}' = \{X'(k) : X'(k) = X(k) + m^*, k = 1, \ldots\}$. It is easy to see that process $\mathbf{X}'$, which we call the *twisted background process*, is a Gaussian process with mean m^*, and that its variance and correlation function are the same as for $\mathbf{X}$. Given a realization $(x'_1, \ldots, x'_{k-1})$ of process $\mathbf{X}'$, the corresponding realization of process $\mathbf{X}$ satisfies $x_j = x'_j - m^*$, for $j = 1, 2, \ldots, k-1$. From equations (1)–(2),

$$\begin{aligned}
\mathrm{E}_{X'}(X'_k | x'_{k-1}, \ldots, x'_1) &= m^* \\
&+ \mathrm{E}_X(X_k | x'_{k-1} - m^*, \ldots, x'_1 - m^*) \\
&= m^* + \mathrm{E}_X(X_k | x_{k-1}, \ldots, x_1) \\
&= m^* + \sum_{j=2}^{k} \phi_{kj}(x_{k-j}) \\
&= m^* + \sum_{j=2}^{k} \phi_{kj}(x'_{k-j} - m^*) \\
&= m^* + m_{k,X'} \quad \text{for } k = 2, 3, \ldots
\end{aligned} \tag{35}$$

where

$$m_{k,X'} \triangleq \sum_{j=2}^{k} \phi_{kj}(x'_{k-j} - m^*) \tag{36}$$

Also from equations (1)–(2)

$$\mathrm{var}_{X'}(X'_k | x'_{k-1}, \ldots, x'_1) = \mathrm{var}_X(X_k | x_{k-1}, \ldots, x_1) \tag{37}$$

In IS simulation, we simulate a twisted *foreground* arrival process $\mathbf{Y}'$ instead of the arrival process $\mathbf{Y}$, where $\mathbf{Y}$ is defined in equation (7) and $Y'_k = h(X'_k) = F_Y^{-1}(F_X(X'_k))$. It is straightforward to observe that, during the simulation we need only calculate the likelihood ratio of the *background* processes, $\mathbf{X}$ and $\mathbf{X}'$.

The likelihood ratio of the corresponding background processes, $\mathbf{X}$ and $\mathbf{X}'$ respectively, is calculated as follows: Let $(x_1', \ldots, x_{k-1}')$ be also taken as a realization of the work load process $\mathbf{X}$. Then,

$$\begin{aligned} \mathrm{E}_X(X_k | x_{k-1}', \ldots, x_1') &= \sum_{j=2}^{k} \phi_{kj}(x_{k-j}') \\ &= m_{k,X} \text{ for } k = 2, 3, \ldots \end{aligned} \tag{38, 39}$$

where

$$m_{k,X} \triangleq \sum_{j=2}^{k} \phi_{kj}(x_{k-j}') \tag{40}$$

We also have

$$\mathrm{var}_X(X_k | x_{k-1}', \ldots, x_1') = \mathrm{var}_{X'}(X_k' | x_{k-1}', \ldots, x_1') \tag{41}$$

The likelihood ratio of the background processes up to time k is

$$\begin{aligned} L(k) &= \frac{f_X(x_1', \ldots, x_k')}{f_{X'}(x_1', \ldots, x_k')} \\ &= \frac{f_X(x_1') f_X(x_2'|x_1') \cdots f_X(x_k'|x_{k-1}', \ldots, x_1')}{f_{X'}(x_1') f_{X'}(x_2'|x_1') \cdots f_{X'}(x_k'|x_{k-1}', \ldots, x_1')} \\ &= \prod_{i=1}^{k} L_i \end{aligned} \tag{42}$$

where

$$L_1 = \frac{f_X(x_1')}{f_{X'}(x_1')} \tag{43}$$

$$L_i = \frac{f_X(x_i'|x_{i-1}', \ldots, x_1')}{f_{X'}(x_i'|x_{i-1}', \ldots, x_1')} \quad \text{for} \quad i = 2, 3, \ldots, k \tag{44}$$

Then, from equations (35) to (39), we have

$$L_i = \frac{e^{\theta_i x_i}}{M_i} \quad for\ i = 2, 3, \ldots \tag{45}$$

where

$$\theta_i = -\frac{-m_{i,X} + m^* + m_{i,X'}}{\sigma^2 \prod_{j=2}^{i}(1 - \phi_{jj}^2)} \tag{46}$$

$$M_i = e^{-\theta_i/2(-m_{i,X} - m^* - m_{i,X'})} \tag{47}$$

and

$$L_1 = e^{-\frac{2(m^*)x_0 + (m^*)^2}{2\sigma^2}} \tag{48}$$