A HYBRID WAVELET FRAMEWORK FOR MODELING VBR VIDEO TRAFFIC

Min Dai, Dmitri Loguinov

Texas A&M University
College Station, TX 77843
min@ee.tamu.edu, dmitri@cs.tamu.edu

Hayder Radha

Michigan State University
East Lansing, MI 48824
radha@egr.msu.edu

ABSTRACT

Traffic models play an important role in network simulation and performance analysis. This paper presents a frame-level hybrid framework for modeling variable bitrate (VBR) video traffic. To accurately capture long-range dependent (LRD) and short-range dependent (SRD) properties of video traffic, we incorporate elements of wavelet-domain analysis into classical time-domain modeling found in prior work. However, unlike previous studies, we analyze and successfully model both inter-GOP *and* intra-GOP correlation. Through the use of QQ plots and leaky-bucket simulations, we evaluate the accuracy of our approach and demonstrate that the autocorrelation function and the frame-size distribution of synthetic traffic match those of the original traffic very well. The leaky-bucket simulation also demonstrates that our model effectively preserves the temporal burstiness of the original video and can be used to predict buffer overflow probabilities and network packet loss.

1. INTRODUCTION

Accurate modeling of VBR video traffic is important for properly allocating network resources, effectively designing networks, and providing certain Quality of Service (QoS) to video applications. A good traffic model should capture the characteristics of video sequences and accurately predict network performance (e.g., buffer overflow probability and packet loss). Of all characteristics of video traffic, two major concerns are the distribution of frame sizes and their autocorrelation function (ACF).

Among the proposed models for the frame-size distribution are the lognormal distribution [5], the Gamma distribution [15], and several hybrid distributions such as Gamma/Pareto [10]. From numerous video sequences, Rose [14] concludes that the Gamma distribution is a good approximation for the PDF of I and P-frame sizes, and the lognormal distribution matches that of B-frame sizes.

Compared to the task of fitting a model to the frame-size distribution, capturing the ACF structure of VBR video traffic is more challenging due to the fact that VBR video exhibits both LRD and SRD properties [11], [6]. Since LRD and SRD are embedded in video traffic, using either a long-range dependent or a short-range dependent model *alone* cannot give satisfactory results. Thus, many studies have been conducted to address this problem, but only a few of them manage to model the ACF structure of video traffic (e.g., [11], [10]).

The correlation that most models try to capture is the *inter*-GOP (i.e., group of pictures) correlation, which exhibits both LRD and SRD trends and is well characterized by the correlation function of the I-frames. In contrast, even fewer studies analyze or model the *intra-GOP*[1] correlation, which is also an important characteristic of video traffic, especially in precisely evaluating the data loss of video transmission over the network [9].

In this paper, we develop a modeling framework that can capture both inter and intra-GOP correlation of video traffic. Given the excellent performance of wavelet analysis in matching the properties of bursty network traffic [11], we model I-frame sizes in the wavelet domain. The wavelet coefficients in the wavelet domain are replaced with more tractable approximations, which are later used to construct synthetic I-frame sizes. Furthermore, to preserve intra-GOP correlation, we generate synthetic P-frame traffic using a linear model of the preceding I-frame in the time domain. Although for demonstration purposes, we use the MPEG-4 coded `StarWars` sequence [3], our framework applies to general GOP-based video traffic. The sequence is in QCIF format, coded at 25 frames/s with GOP structure $IBBPBBPBBPBB$ and quantization parameters 10-14-18 for I-P-B frames, respectively.

This paper is organized as follows. In Section 2, we briefly overview the related work and provide a background on wavelet analysis. In Section 3, we give a short discussion of intra-GOP correlation and explain how to generate synthetic I, P, and B-frame sizes. We also evaluate the accuracy of our modeling framework in this section. Section 4 concludes the paper.

2. RELATED WORK AND PRELIMINARIES

In this section, we provide a brief overview of related work and background knowledge on wavelet analysis.

2.1. Related Work

Numerous studies have been conducted in modeling VBR video traffic, which can be grouped into five categories according to the dominant stochastic method applied in each model: Autoregressive (AR) models [4], [5], [7], [10], Markov-modulated models [9], [15], self-similar (fractional) models [6], [8], wavelet-based methods [11], [13], and other methods [12].

In the first category, we discuss AR models, which are considered a classical approach in the area of traffic modeling. Krunz *et al*. [5] model the deviation of I-frame sizes from their mean (in each scene) using an AR(2) process. Building upon Krunz' work [5], Liu *et al*. [10] utilize a *nested* AR(2) model, which uses a second AR(2) process to model the mean frame-size of each scene. Heyman [7] propose an AR model called GBAR with Gamma-distributed marginal statistics and a geometric autocorrelation. By

This work was supported in part by NSF grant ANI-0312461.

[1] The correlation between P/B-frames and the I-frame in the same GOP.

considering the GOP cyclic structure of video traffic, Frey *et al*. [4] extend GBAR in [7] to the GOP-GBAR model.

The second category consists of Markov-modulated models, which employ Markov chains to create other processes (e.g., the Bernoulli process [9]). Sarkar *et al*. [15] propose two Markov-modulated Gamma-based algorithms. At each state of the Markov chain, the sizes of I, P, and B-frames are generated as Gamma-distributed random variables with different sets of parameters. Although Markov-modulated models can capture the LRD of video traffic, it is difficult to accurately define and segment video sources into different states in the time domain due to the dynamic nature of video traffic [11].

We group models based on a self-similar process in the third category. Garrett *et al*. [6] propose a fractional ARIMA (Autoregressive Integrated Moving Average) model to capture the LRD of video traffic, but explicit modeling of the SRD structure has not been provided. Using the results of [6], Huang *et al*. [8] present a self-similar fractal traffic model. However, this model does not capture the multi-timescale variations in video traffic [5]. By contrast, Transform-Expand-Sample (TES) methods (e.g., [12]) are accurate in matching the ACF at both small and large lags. However, they have high computational complexity and often must use special software (e.g., *TEStool*) to generate synthetic sequences.

Different from the above time-domain methods, several powerful wavelet models [11], [13] recently emerged due to their ability to accurately capture both LRD and SRD properties of video traffic [11]. Since background of wavelets is required to better understand these methods, we elaborate on the basics of these models next.

2.2. Wavelet Analysis

Wavelet analysis is typically based on the decomposition of a signal using an orthonormal family of basis functions, which includes a high-pass wavelet function and a low-pass scaling filter. The former generates the *detailed* coefficients while the latter generates the *approximation* coefficients of the original signal.

Wavelet transform can strongly reduce the temporal correlation of an input signal, which indicates that even though a signal has LRD properties, its corresponding wavelet coefficients are short-range dependent [11]. In Fig. 1 (left), we show the autocorrelation of the detailed and approximation coefficients at decomposition level three using Haar wavelets (labeled as "ACF detailed" and "ACF approx", respectively). As shown in the figure, the ACF of detailed coefficients is almost zero at non-zero lags, which means that they are *i.i.d.* (uncorrelated) random variables. This explains why previous literature commonly models detailed coefficients as zero-mean *i.i.d.* Gaussian variables [11]. Fig. 1 (left) also shows that the approximation coefficients have a slower decaying ACF compared to that of the detailed coefficients, which implies that they *cannot* be modeled as *i.i.d.* random variables.

We next examine the relationship between the original signal and the approximation coefficients in various sequences and observe that the approximation coefficients follow a Gamma distribution, which is also a very good approximation of the histogram of I-frame sizes [9], [14]. To understand this scenario better, in Fig. 1 (right), we illustrate the distribution of the original signal and that of the third-level approximation coefficients using Haar wavelets. The figure shows that the two distributions have a similar shape, but with different parameters.

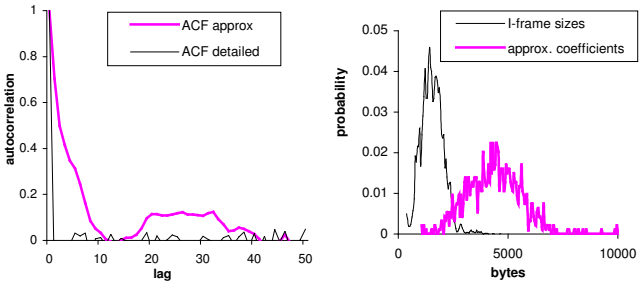

Fig. 1. The ACF structures of the third-level approximation coefficients and a typical set of detailed coefficients (left). The distribution of I-frame sizes and the third-level approximation coefficients (right).

3. MODELING VIDEO TRAFFIC

In this section, we model I-frame sizes in the wavelet domain and P-frame sizes based on the intra-GOP correlation.

3.1. Generating Synthetic I-Frame Sizes

In the following discussion, we define A_j to be the *approximation* coefficients and D_j the *detailed* coefficients at the wavelet decomposition level j. We also assume that $j = J$ is the coarsest scale and $j = 0$ is the original signal. As mentioned earlier, previous wavelet-based methods often model D_j as zero-mean *i.i.d.* Gaussian variables [11]. We examined the statistical properties of D_j and found that a mixture-Laplacian distribution [1] matches the PDF of D_j better than the Gaussian distribution. In Fig. 2 (left), we show the distributions of the actual and the estimated coefficients D_1 in `StarWars`. The figure shows that a mixture-Laplacian estimation outperforms the Gaussian model.

Recall that current methods generate the coarsest approximation coefficients A_J either as independent Gaussian [11] or Beta random variables [13]. However, as shown in Fig. 1 (left), these approximation coefficients are not *i.i.d.* distributed. To preserve the correlation in A_J and achieve the expected distribution of the synthetic coefficients, we assume that the coarsest approximate coefficients are dependent random variables with marginal Gamma distributions. We first generate N dependent Gaussian variables x_i using a $k \times k$ correlation matrix, where N is the length of A_J and the correlation matrix is obtained from A_J. The number of preserved correlation lags k is chosen to be a reasonable value (e.g., the average scene length). By applying the Gaussian CDF $F_G(x)$ directly to x_i, we convert them into a uniformly distributed set of variables $F_G(x_i)$. It is well known that if F is a continuous distribution with inverse F^{-1} and u is a uniform random number, then $F^{-1}(u)$ has the distribution F. Based on this insight, we pass the result from the last step through the inverse Gamma CDF to generate (still dependent) Gamma random variables [2].

With the estimated approximation and detailed coefficients, inverse wavelet transform is performed and the synthetic I-frame sizes are generated. Fig. 2 (right) shows the ACF of the actual I-frame sizes and that of the synthetic traffic in long range. Fig. 3 (left) shows the correlation of the synthetic traffic from the GOP-GBAR model [4] and Gamma_A model [15] in short range. As observed in both figures, our synthetic I-frame sizes capture both the LRD and SRD properties of the original traffic and outperforms the previous models.

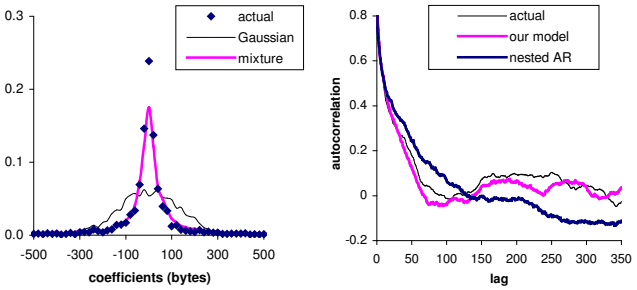

Fig. 2. The PDF of actual D_1 with Gaussian and mixture-Laplacian estimations in `StarWars` (left). The ACF of the actual I-frame sizes and that of the synthetic traffic in long range (right).

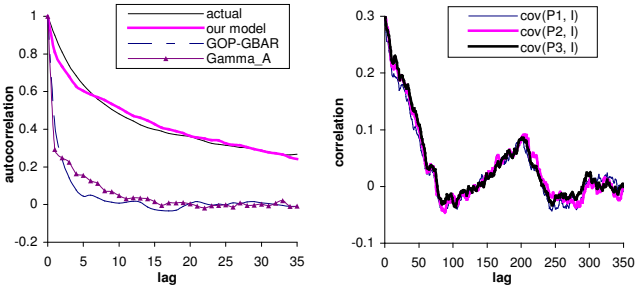

Fig. 3. The ACF of the actual I-frame sizes and that of the synthetic traffic in short range (left). The correlation between $\phi_i^P(n)$ and $\phi^I(n)$, with $i = 1, 2, 3$ (right).

3.2. Intra-GOP Correlation

Lombardo *et al*. [9] noticed that there is a strong correlation between the P/B-frames and the I-frame belonging to the same GOP. Motivated by their results, we investigate the correlation[2] between P/B-frame sizes and the I-frame sizes from the same GOP.

Before further discussion, we define I, P and B-*frame size sequences* as follows. Assuming that n represents the GOP sequence number, we define $\phi^I(n)$ to be the I-frame size of the n-th GOP, $\phi_i^P(n)$ to be the size of the i-th P-frame in GOP n, and $\phi_i^B(n)$ to be the size of the i-th B-frame in GOP n. For example, $\phi_3^P(10)$ represents the size of the third P-frame in the 10-th GOP.

We display the correlation between processes $\{\phi^I(n)\}$ and $\{\phi_i^P(n)\}$ in Fig. 3 (right). As shown in the figure, the correlation is almost identical between the different P-frame sequences and the I-frame sequence, which is rather convenient for modeling P-frame sizes. Due to limited space, we do not show the correlation between $\{\phi_i^B(n)\}$ and the I-frame sequence $\{\phi^I(n)\}$, which also does not change as a function of i.

Lombardo *et al*. [9] further model the sizes of P and B-frames as Gamma distributed random variables, with mean and variance estimated by a *linear* function of I-frame sizes. The sample video sequences in [9] are MPEG-1 coded; however, we find that this linear estimation does not hold for general video traffic. As shown in Fig. 4 (left), the means of P and B-frames are not linear functions of I-frame sizes in MPEG4-coded `StarWars`. Therefore, we propose an alternative modeling framework in the following section that captures the intra-GOP correlation for general GOP-based video.

3.3. Modeling P and B-Frame Sizes

The above discussion shows that there is a similar correlation between $\{\phi_i^P(n)\}$ and $\{\phi^I(n)\}$ with respect to different i. Thus, we propose the following model to estimate the size of the i-th P-frame in the n-th GOP:

$$\phi_i^P(n) = a\tilde{\phi}^I(n) + \tilde{v}(n), \tag{1}$$

where $\tilde{\phi}^I(n) = \phi^I(n) - E[\phi^I(n)]$ and $\tilde{v}(n)$ is a process that is independent of $\tilde{\phi}^I(n)$.

Lemma 1. *To capture the intra-GOP correlation, the value of coefficient a in (1) must be equal to:*

$$a = \frac{r(0)\sigma_P}{\sigma_I}, \tag{2}$$

where σ_P is the standard deviation of $\{\phi_i^P(n)\}$, σ_I is the standard deviation of $\{\phi^I(n)\}$, and $r(0)$ is their normalized correlation coefficient at lag zero.

Proof. Without loss of generality, we assume that both $\tilde{\phi}^I(n)$ and $\phi_i^P(n)$ are wide-sense stationary processes. Thus, $E[\phi_i^P(n)]$ is constant and:

$$E[\tilde{\phi}^I(n-k)] = E[\tilde{\phi}^I(n)] = 0. \tag{3}$$

Denote by $C(k)$ the covariance between $\phi_i^P(n)$ and $\tilde{\phi}^I(n)$ at lag k:

$$C(k) = E[(\phi_i^P(n) - E[\phi_i^P])(\tilde{\phi}^I(n-k) - E[\tilde{\phi}^I])]. \tag{4}$$

Recall that $\tilde{v}(n)$ and $\tilde{\phi}^I(n)$ are independent of each other and thus $E[\tilde{v}(n)\cdot\tilde{\phi}^I(n)] = E[\tilde{v}(n)]\cdot E[\tilde{\phi}^I(n)] = 0$. Then $C(k)$ becomes:

$$\begin{aligned} C(k) &= E[(a\tilde{\phi}^I(n) + \tilde{v}(n) - E[\phi_i^P])\tilde{\phi}^I(n-k)] \\ &= aE[\tilde{\phi}^I(n)\tilde{\phi}^I(n-k)] \end{aligned} \tag{5}$$

Next, observe that the normalized correlation coefficient r at lag zero is:

$$r(0) = \frac{C(0)}{\sigma_P\sigma_{\tilde{I}}} = \frac{aE[\tilde{\phi}^I(n)^2]}{\sigma_P\sigma_{\tilde{I}}}, \tag{6}$$

where $\sigma_{\tilde{I}}$ is the standard deviation of $\tilde{\phi}^I(n)$.

Recalling that $E[\tilde{\phi}^I(n)] = 0$, we have $E[\tilde{\phi}^I(n)^2] = \sigma_{\tilde{I}}^2 = \sigma_I^2$ and:

$$\frac{a\cdot\sigma_I}{\sigma_P} = r(0), \tag{7}$$

which leads to (2). □

To better understand the distribution of $\{\tilde{v}(n)\}$, we next examine the *actual* residual process $v(n) = \phi_i^P(n) - a\tilde{\phi}^I(n)$ for each i. We show the histograms of $\{v(n)\}$ for P-frame sequences $i = 1, 2, 3$ in `StarWars` in Fig. 4 (right). The figures shows that the residual process $\{v(n)\}$ possesses very similar statistical properties when modeling different P-frame sequences. We also observe that $\{v(n)\}$ follows a shifted Gamma distribution with parameters μ, α, β. The statistical parameters ($r(0), \sigma_P, \sigma_I, \mu, \alpha, \beta$) needed for this model can be easily estimated from the original sequences.

We illustrate the difference between our model and a typical *i.i.d* method of prior work (e.g., [10], [15]) in Fig. 5 (left). The figure shows that our model indeed preserves the intra-GOP correlation of the original traffic much better than the common methods of related work.

Since the sizes of B-frames are relatively small compared to those of P-frames (as shown in Fig. 4 (left)), we generate the synthetic B-frame traffic simply using an *i.i.d.* lognormal random number generator.

[2] In traffic modeling literature, the normalized auto-covariance function is often used instead of the autocorrelation function.

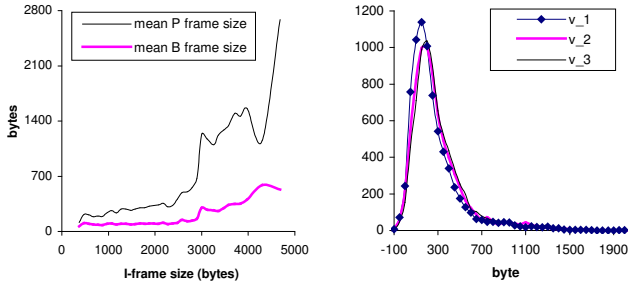

Fig. 4. The mean sizes of P and B-frames given the sizes of I-frames (left). The histograms of $\{v(n)\}$ when modeling $\{\phi_i^P(n)\}$, with $i = 1, 2, 3$ (right).

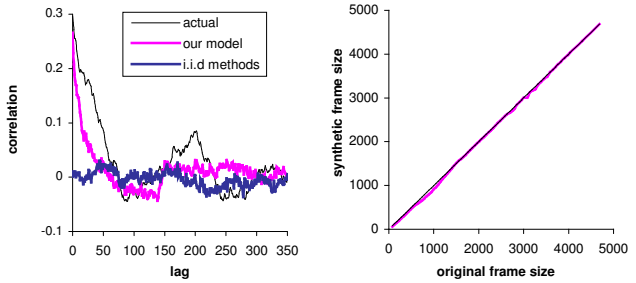

Fig. 5. The correlation between $\phi_1^P(n)$ and $\phi^I(n)$ in `StarWars` (left). The QQ plot for the synthetic `StarWars` traffic (right).

3.4. Model Accuracy Study

There are two popular studies to verify the accuracy of a video traffic model [15]: quantile-quantile (QQ) plots and packet-loss buffer evaluation. The QQ plot is a graphical technique to verify the distribution similarity between two test data sets. If the two sets have the same distribution, the points should fall along the 45 degree reference line. Fig. 5 (right) shows the QQ plot of the synthetic traffic generated by our model and demonstrates that the generated frame sizes and the original traffic are statistically identical.

Besides the distribution, we also examine how well our approach preserves the temporal information of the original traffic. A common test for this is to pass the synthetic traffic through a generic buffer with capacity c and drain rate d [15]. The drain rate is the number of bytes drained per second and is simulated as different multiples of the average traffic rate $\bar{r}$. To understand the performance difference between the various models, we define the relative error e between the *actual* packet loss p observed in the buffer fed with the original traffic and that observed using the synthetic traffic generated by each of the models:

$$e = \frac{|p - p_{model}|}{p}. \quad (8)$$

In Table 1, we illustrate the values of e for various buffer capacities and drain rates d. As shown in the table, the synthetic traffic generated by our model provides a very accurate estimate of the actual packet-loss probability p and significantly outperforms the other methods. We should note that additional simulations with other video sequences have demonstrated results similar to those shown throughout this paper.

4. CONCLUSION

In this paper, we presented a framework for modeling full-length VBR video traffic. This framework incorporated wavelet-domain analysis into time-domain modeling. This work precisely captures the LRD as well as SRD properties of video traffic and accurately describes its intra-GOP correlation. Furthermore, since our framework is developed at the frame-size level (much of previous work uses slice-level or even block-level [15]), future applications can predict the loss ratio for each type of frames and apply various methods (e.g., guarantee the transmission of certain frames) to improve the video quality at the receiver.

Buffer capacity	Traffic type	Drain rate		
		$2\bar{r}$	$4\bar{r}$	$5\bar{r}$
10ms	Our Model	1.80%	0.93%	0.50%
	GOP-GBAR [4]	2.44%	2.51%	4.01%
	Nested AR [10]	4.02%	2.05%	5.63%
	Gamma_A [15]	5.54%	1.04%	0.99%
	Gamma_B [15]	5.76%	1.81%	1.15%
20ms	Our Model	0.93%	0.61%	1.13%
	GOP-GBAR [4]	3.84%	2.16%	3.77%
	Nested AR [10]	5.81%	2.77%	8.46%
	Gamma_A [15]	5.20%	0.61%	2.57%
	Gamma_B [15]	4.89%	1.93%	2.05%
30ms	Our Model	0.25%	0.33%	0.95%
	GOP-GBAR [4]	4.94%	3.33%	5.68%
	Gamma_A [15]	4.88%	1.10%	4.48%
	Gamma_B [15]	4.67%	2.17%	4.03%

Table 1. Relative packet-loss error e using the `StarWars` sequence.

5. REFERENCES

[1] M. Dai, D. Loguinov, and H. Radha, "Statistical Analysis and Distortion Modeling of MPEG-4 FGS," *IEEE ICIP*, Sept. 2003.

[2] P. Embrechts, F. Lindskog, and A. McNeil, "Correlation and Dependence in Risk Management: Properties and Pitfalls," *Cambridge University Press*, 2002.

[3] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation," aviailable at http://www-tkn.ee.tu-berlin.de.

[4] M. Frey and S. Nguyen-Quang, "A Gamma-Based Framework for Modeling Variable-Rate MPEG Video Sources: the GOP GBAR Model," *IEEE/ACM Trans. on Networking*, vol. 8, Dec. 2000.

[5] M. Krunz and S. K. Tripathi, "On the Characterization of VBR MPEG Streams," *ACM SIGMETRICS*, vol. 25, 1997.

[6] M. W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," *Proc. of SIGCOM*, 1994.

[7] D. P. Heyman, "The GBAR Source Model for VBR Video Conferences," *IEEE/ACM Trans. on Networking*, vol. 5, Aug. 1997.

[8] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye, "Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified Approach," *Proc. of SIGCOM*, 1995.

[9] A. Lombardo, G. Morabito, and G. Schembra, "An Accurate and Treatable Markov Model of MPEG-Video Traffic," *Proc. of INFOCOM*, 1998.

[10] D. Liu, E. I. Sára, and W. Sun, "Nested Auto-Regressive Processes for MPEG-Encoded Video Traffic Modeling," *IEEE Trans. on CSVT*, vol. 11, 2001.

[11] S. Ma and C. Ji, "Modeling Heterogeneous Network Traffic in Wavelet Domain," *IEEE/ACM Trans. on Networking*, vol. 9, Oct. 2001.

[12] B. Melamed and D. E. Pendarakis, "Modeling Full-Length VBR Video Using Markov-Renewal-Modulated TES Models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 600-611, June 1998.

[13] V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk, "Multiscale Queuing Analysis of Long-Range-Dependent Network Traffic," *Proc. of INFOCOM*, vol. 2, 2000.

[14] O. Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems," *Technical Report*, Univ. of Wurzburg, Institute of Computer Science, Feb. 1995.

[15] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, "Modeling Full-Length Video Using Markov-Modulated Gamma-Based Framework," *IEEE/ACM Trans. on Networking*, vol. 11, Aug. 2003.