Wavelet and Time-Domain Modeling of Multi-Layer VBR Video Traffic

Min Dai
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843
min@ee.tamu.edu

Dmitri Loguinov*
Department of Computer Science
Texas A&M University
College Station, TX 77843
dmitri@cs.tamu.edu

Abstract— **Modeling of video traffic has always been an important research area for the design, simulation, and performance analysis of packet video networks and Internet streaming. In this paper, we present a frame-level hybrid framework for modeling multi-layer VBR video traffic, in which the base layer is modeled using a combination of wavelet and time-domain methods and the enhancement layer is linearly predicted from the based-layer model. To accurately capture long-range dependent (LRD) and short-range dependent (SRD) properties of VBR base-layer traffic, we use wavelets to model the distribution of I-frame sizes and a simple time-domain model for P/B frame sizes; however, unlike previous studies, we analyze and successfully model both inter-GOP *and* intra-GOP correlation found in many VBR sequences. Through the use of QQ plots and leaky-bucket simulations, we demonstrate that our model not only effectively preserves the temporal burstiness of the original traffic, but also captures important statistical features of the original video such as the autocorrelation function and the frame-size distribution. We also show that our model performs better than the previous methods in single-layer and multi-layer MPEG-4 sequences and demonstrate that certain results obtained in prior work for MPEG-1 video do not hold for general GOP-based coding schemes.**

I. INTRODUCTION

Video traffic modeling is a traditional research area that has captivated researchers for over 15 years. The main goal of the studies in this field is to understand the properties (such as the distribution and correlation) of frame sizes produced by a variety of video encoders and to develop models that can generate synthetic traffic with properties close to those of the actual video sequences. Besides providing an insight into the coding process and structure of video traffic, these models can later be used for many practical purposes including allocation of network resources, design of efficient streaming networks, and delivery of certain Quality of Service (QoS) guarantees to the end users.

A good traffic model should capture the characteristics of video sequences and accurately predict network performance (e.g., buffer overflow probabilities and packet loss). Among the various characteristics of video traffic, there are two major interests: (1) the distribution of frame sizes; and (2) the autocorrelation function (ACF) that captures common dependencies between frame sizes in VBR traffic. In regard to the first issue, several models have been proposed for the frame-size distribution, including the lognormal [7], Gamma [25], and various hybrid distributions (e.g., Gamma/Pareto [15] or Gamma/lognormal [23]). 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 [9], [16]. The co-existence of SRD and LRD indicates that the ACF structure of video traffic is similar to that of SRD processes at small time lags and to that of LRD processes at large time lags [9]. Since both LRD and SRD are embedded in the signal, using either a long-range dependent or a short-range dependent model *alone* does not provide satisfactory results. Thus, many studies are conducted to address this problem, but only a few of them manage to model the complicated LRD/SRD ACF structure of real video traffic (e.g., [15], [16]).

The correlation that most models try to capture is the *inter-GOP* correlation, which is well characterized by the ACF of the I-frames; however, another dimension to video traffic includes the *intra-GOP* correlation[1]. This is an important characteristic useful in computing precise bounds on network packet loss [14], which is rarely addressed in related work. Additionally, most existing traffic models only apply to single-layer VBR video and often overlook the multi-layer aspects of common streaming video traffic in the current Internet [1], [27].

In this paper, we develop a modeling framework that is able to capture the complex LRD/SRD structure of single-layer and multi-layer video traffic, while addressing the issues of both intra-GOP and multi-layer correlation. We model I-frame sizes in the wavelet domain and replace the wavelet coefficients with more tractable approximations, which are later used to construct synthetic I-frame sizes. To preserve the intra-GOP correlation, we generate synthetic P-frame traffic using a linear model of the preceding I-frame in the time domain. We use a similar model to maintain the cross-layer correlation in multi-layer video sequences and show that the complexity (in terms of parameters or computational cost) of the resulting model is no worse than that of prior methods.

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

Although we use MPEG-4 sequences in this paper, our framework applies to general GOP-based multi-layer video traffic. The specifics of the three test sequences discussed in this paper are as following: a single layer `Star Wars` [5] (25 frames/s), a two-layer spatially-scalable `The Silence of the Lambs` [21] (30 frames/s), and a two-layer FGS-coded `Star Wars` [21] (30 frames/s). All three sequences have GOP structure $IBBPBBPBBPBB$.

This paper is organized as follows. In Section II, we overview the related work on traffic modeling and provide the background on wavelet analysis. In Section III, we discuss the intra-GOP correlation and show how to model I, P, and B-frame sizes of general video traffic. Section IV explains how to generate a synthetic enhancement layer, given certain base-layer information. Section V concludes the paper.

II. RELATED WORK AND PRELIMINARIES

In this section, we provide a brief overview of related work and discuss the basics of wavelet analysis.

A. Single-Layer Models

Numerous studies have been conducted in modeling VBR video traffic. According to the dominant stochastic method applied in each model, we group them into five categories: autoregressive (AR) models [6], [7], [10], [15], Markov-modulated models [13], [25], self-similar (fractal) models [9], [11], wavelet-based methods [16], [22], and other approaches [19].

In the first category, we discuss AR models, which are considered a classical approach in the area of traffic modeling. After the first auto-regressive (AR) model was applied to video traffic in 1988 [17], AR processes and their variations remain highly popular in this area of research [15]. For example, Corte *et al.* [3] use a linear combination of two AR(1) processes to model the ACF of the original video traffic, in which one AR(1) model is used for modeling small lags and the other one for large lags. Since using a single AR process is generally preferred, Krunz *et al.* [7] model the deviation of I-frame sizes from their mean in each scene using an AR(2) process. Building upon Krunz' work [7], Liu *et al.* [15] propose a *nested* AR(2) model, which uses a second AR(2) process to model the mean frame-size of each scene. Heyman [10] propose an AR model called GBAR with Gamma-distributed marginal statistics and a geometric autocorrelation. By considering the group-of-picture (GOP) cyclic structure of video traffic, Frey *et al.* [6] extend the GBAR model in [10] 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 [13]). Rose [24] uses nested Markov chains to model GOP sizes. Since synthetic data is generated at the GOP level, this model actually coarsens the time scale and thus is not suitable for high-speed networks. Chen *et al.* [2] use a doubly Markov modulated punctured AR model, in which a nested Markov process describes the transition between the different states and an AR process describes the frame size at each state. The computation complexity of this method is quite high due to the combination of a doubly Markov model and an AR process. Sarkar *et al.* [25] 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 the different states in the time domain due to the dynamic nature of video traffic [16].

We group models based on a self-similar process into the third category. Garrett *et al.* [9] propose a fractional ARIMA (Autoregressive Integrated Moving Average) model to replicate the LRD properties of compressed sequences, but do not provide an explicit model for the SRD structure of video traffic. Using the results of [9], Huang *et al.* [11] present a self-similar fractal traffic model; however, this model does not capture the multi-timescale variations in video traffic [7]. This problem can be overcome using the Transform-Expand-Sample (TES) method [19]. Although this method is accurate in matching the ACF at both small and large lags, it has high computational complexity and often must be used in special software (e.g., *TEStool*) that generates synthetic sequences.

Different from the above time-domain methods, several wavelet models [16], [22] recently emerged due to their ability to accurately capture both LRD and SRD properties of video traffic [16]. We provide more background on wavelets and an initial analysis of approximation coefficients in section II-C.

B. Multi-Layer Models

All models discussed above focus on single-layer video traffic and only a handful of studies analyze multi-layer sequences. For example, Chandra *et al.* [1] use a finite-state Markov chain to model one- and two-layer video traffic of all activity levels. They assume that only one I-frame exists in the whole video sequence and the I-frame size is simply an *i.i.d.* Gaussian random variable. The model clusters P-frame sizes into K states according to the correlation between successive P-frame sizes and uses a first-order AR process to model the frame size in each state. The goal of [1] is to model one or two-layer video traffic with a CBR base layer, while many multi-layer video sequences have *more* than two layers and the base-layer is VBR.

Similarly to the work in [1], Zhao *et al.* [27] build a K-state Markov chain based on frame-size clusters. The clustering feature in [27] is the cross-correlation between the frame size of the base layer and that of the enhancement layer at the same frame index. In each state of the Markov chain, the base and the enhancement-layer frame sizes follow a multivariate normal distribution. However, the computational cost of the hierarchical clustering approach applied in [27] limits its application to short video sequences. Furthermore, in both [1] and [27], there is no general method for choosing the optimal number of states and the parameters are often chosen empirically.

C. Wavelet Models and Preliminaries

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 produces the *approximation* coefficients of the original signal. The wavelet transform strongly reduces the temporal correlation in the input signal, which means that signals with LRD properties produce short-range dependent wavelet coefficients [16].

In order to understand the structure of the wavelet transform, we next examine the relationship between the original signal and the detailed and approximation coefficients. We use the Haar wavelet transform as a typical example since it is often chosen for its simplicity and good performance [16], [22]. Recall that the Haar scaling and wavelet functions are, respectively:

$$\varphi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

$$\psi(t) = \begin{cases} 1 & 0 \le t < 1/2 \\ -1 & 1/2 \le t < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

In the following discussion, we define $\{A_j\}$ to be the random process modeling approximation coefficients A_j^k and $\{D_j\}$ to be the process modeling detailed coefficients D_j^k at the wavelet decomposition level j, where k is the spatial location of A_j^k and D_j^k. We also assume that $j = J$ is the coarsest scale and $j = 0$ is the original signal. Recall that the Haar approximation coefficients A_j^k are obtained via [22]:

$$A_j^k = 2^{-1/2}(A_{j-1}^{2k} + A_{j-1}^{2k+1}). \tag{3}$$

In Fig. 1 (a), we show the autocorrelation of processes $\{A_3\}$ and $\{D_3\}$ computed based on the I-frame sizes in single-layer `Star Wars` using Haar wavelets (labeled as "ACF detailed" and "ACF approx", respectively). As shown in the figure, the ACF of $\{D_3\}$, which is a typical example of detailed coefficients, is almost zero at non-zero lags, which means that it is an *i.i.d.* (uncorrelated) noise. This explains why previous literature commonly models detailed coefficients as zero-mean *i.i.d.* Gaussian variables [16]. Fig. 1 (a) 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.

Recalling that I-frame sizes $\{A_0\}$ follow a Gamma distribution [23], we next examine the relationship between $\{A_0\}$ and the approximation coefficients $\{A_j, j > 0\}$ in various sequences with the help of the following lemma. Notice that $\{A_j\}$ is a random process $A_j = (A_j^1, A_j^2, \cdots, A_j^k, \cdots)$ and A_j^k is a random variable.

Lemma 1: Given that the I-frame sizes follow a Gamma distribution, the approximation coefficients $A_j^k, j \ge 1$ is a linear combination of several Gamma distributions.

Proof: For brevity, we only derive the distribution of A_1^k and note that the derivations for $A_j^k, j \ge 2$ are very similar.
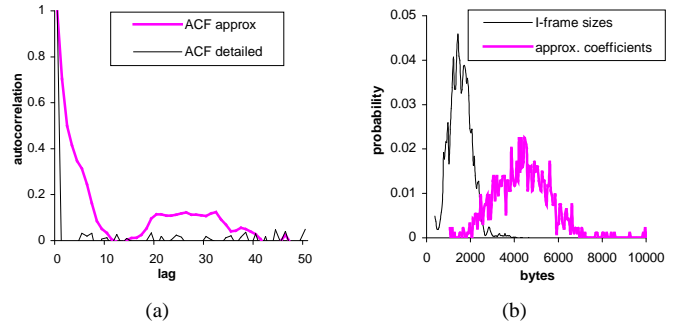

Fig. 1. (a) The ACF structure of coefficients $\{A_3\}$ and $\{D_3\}$ in single-layer `Star Wars`. (b) The histogram of I-frame sizes and that of coefficients A_3.

According to (3), each value of A_1^k is a linear summation of the sizes of two neighboring I-frames, which we denote by X_1^k and X_2^k, respectively. Notice that X_1^k and X_2^k are two correlated Gamma distributed random variables. Then,

$$A_1^k = 2^{-1/2}(X_1^k + X_2^k), \tag{4}$$

where $X_i^k \sim \text{Gamma}(\alpha_i, \lambda_i), i = 1, 2$, and $\lambda_1 = \lambda_2$. We can rewrite X_i^k in the form of the standard Gamma distribution:

$$X_1^k = \lambda_1 Y_1, \tag{5}$$
$$X_2^k = \lambda_2 Y_2, \tag{6}$$

where $Y_i \sim \text{Gamma}(\alpha_i, 1)$ are two standard Gamma random variables.

To catch the correlation between X_1^k and X_2^k, we further decompose Y_1 and Y_2 into a sum of two *independent* standard Gamma random variables using the decomposition properties of standard Gamma distributions [6]:

$$Y_1 = Y_{11} + Y_{12}, \tag{7}$$
$$Y_2 = Y_{12} + Y_{22}, \tag{8}$$

where Y_{11}, Y_{12}, and Y_{22} are independent of each other and follow the standard Gamma distribution with parameters α_{11}, α_{12}, and α_{22}, respectively. Then the correlation between X_1^k and X_2^k becomes:

$$\text{cov}(X_1^k, X_2^k) = \lambda_1 \lambda_2 \text{var}(Y_{12}) = \lambda_1 \lambda_2 \alpha_{12}. \tag{9}$$

Combining (4) and (9), re-write A_1^k as:

$$A_1^k = 2^{-1/2}\left(\lambda_1 Y_{11} + (\lambda_1 + \lambda_2)Y_{12} + \lambda_2 Y_{22}\right). \tag{10}$$

As observed from (10), A_1^k is a linear combination of independent standard Gamma distributions, which leads to the statement of the lemma after detailed statistical analysis. ∎

As a typical example, we illustrate the distribution of the approximation coefficients $\{A_3\}$ and that of $\{A_0\}$ (original I-frame sizes) of single-layer `Star Wars` in Fig. 1 (b). The figure shows that the two distributions have a similar shape, but with different parameters. During extensive experiments, we find a single Gamma distribution is accurate enough to describe the actual histogram of $\{A_j\}$. In the next section, we use this information to efficiently estimate the approximation coefficients.

III. MODELING GENERAL VIDEO TRAFFIC

In this section, we model I-frame sizes in the wavelet domain and P-frame sizes based on the intra-GOP correlation. We address the enhancement layer in the next section. There are two contributions to our framework discussed below: (1) we show a novel method for estimating the coefficients of the wavelet transform, which is both efficient and accurate; and (2) we model the intra-GOP correlation and propose a simple model that accurately generates synthetic P-frame sizes, which is in contrast to much of the previous work that relied on $i.i.d.$ random variables to model the sizes of P/B-frame sizes in each GOP [7], [11], [15], [25].

A. Generating Synthetic I-Frame Sizes

Since the wavelet transform has a great advantage over the time-domain methods in capturing the LRD and SRD properties of video [16], [22], we model the I-frame sizes in the wavelet domain and thus need to estimate both detailed and approximation coefficients, which we already defined as $\{D_j\}$ and $\{A_j\}$, respectively.

Even though previous wavelet-based traffic modeling methods often model $\{D_j\}$ as zero-mean $i.i.d.$ Gaussian variables [16], there is insufficient evidence as to the distribution of the actual $\{D_j\}$ found in GOP-based video traffic. To provide some insight into the structure of detailed coefficients, we compare the histogram of the *actual* coefficients $\{D_1\}$ in Star Wars with those generated by several alternative models in Fig. 2 (note that the y-axis is scaled logarithmically). Fig. 2 (a) displays the histogram of the actual $\{D_1\}$, part (b) shows that the Gaussian fit matches neither the shape, nor the the range of the actual distribution, and part (c) demonstrates that the Generalized Gaussian Distribution (GGD) produces an overly sharp peak at zero (the number of zeros in GGD is almost three times larger than that in the actual $\{D_1\}$) and also does not model the range of the real $\{D_1\}$.

Additional simulations (not shown for brevity) demonstrate that a low-variance Laplacian distribution can describe the high peak of the actual histogram and a high-variance one can cover the large range of the actual data, but none of them can achieve both goals at the same time. However, a *mixture-Laplacian* distribution describes the real data very well:

$$f(x) = p\frac{\lambda_0}{2}e^{-\lambda_0|x|} + (1-p)\frac{\lambda_1}{2}e^{-\lambda_1|x|}, \quad (11)$$

where $f(x)$ is the PDF of the mixture-Laplacian model, p is the probability to obtain a sample from a low-variance Laplacian component, and λ_0 and λ_1 are the shape parameters of the corresponding low- and high-variance Laplacian distributions. Fig. 2 (d) shows that the histogram of the mixture-Laplacian synthetic coefficients $\{D_1\}$ is much closer to the actual one than the other discussed distributions.

We next discuss approximation coefficients $\{A_j\}$. Recall that current methods generate the coarsest approximation coefficients (i.e., $\{A_J\}$) either as independent Gaussian [16] or Beta random variables [22]. However, as mentioned in Section II-C, the approximation coefficients are non-negligibly
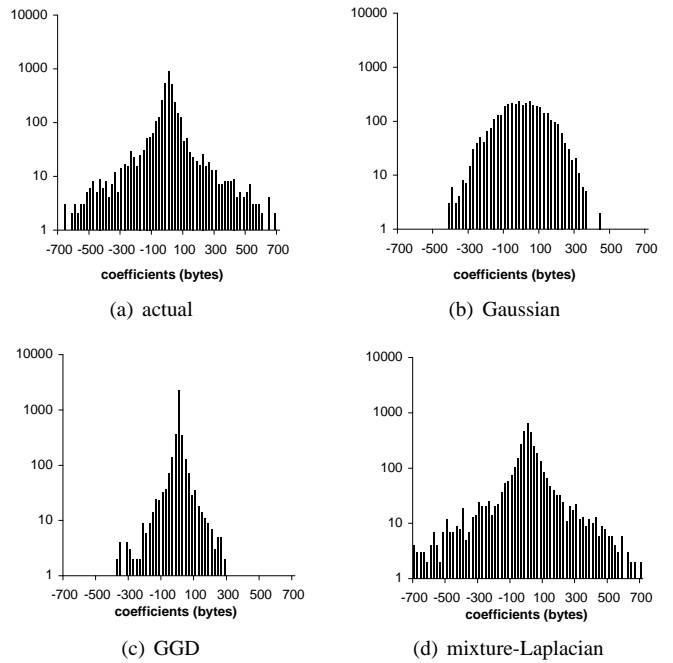

Fig. 2. Histograms of (a) the actual detailed coefficients; (b) the Gaussian model; (c) the GGD model; and (d) the mixture-Laplacian model.

correlated and are not $i.i.d.$ To preserve the correlation of approximation coefficients and achieve the expected distribution in the synthetic coefficients, we assume that the coarsest approximation coefficients $\{A_J\}$ 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 the actual coefficients $\{A_J\}$. The number of preserved correlation lags k is chosen to be a reasonable value (e.g., the average scene length[2]). 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 [4].

Using the estimated approximation and detailed coefficients, we perform the inverse wavelet transform to generate synthetic I-frame sizes. Fig. 3 (a) shows the ACF of the actual I-frame sizes and that of the synthetic traffic. Fig. 3 (b) shows the correlation of the synthetic traffic from the GOP-GBAR model [6] and Gamma_A model [25] in short range. As observed in both figures, our synthetic I-frame sizes capture both the LRD and SRD properties of the original traffic better than the previous models.

---

[2]This is a reasonable choice because there is much less correlation among I-frames of different scenes than among I-frames of the same scene.
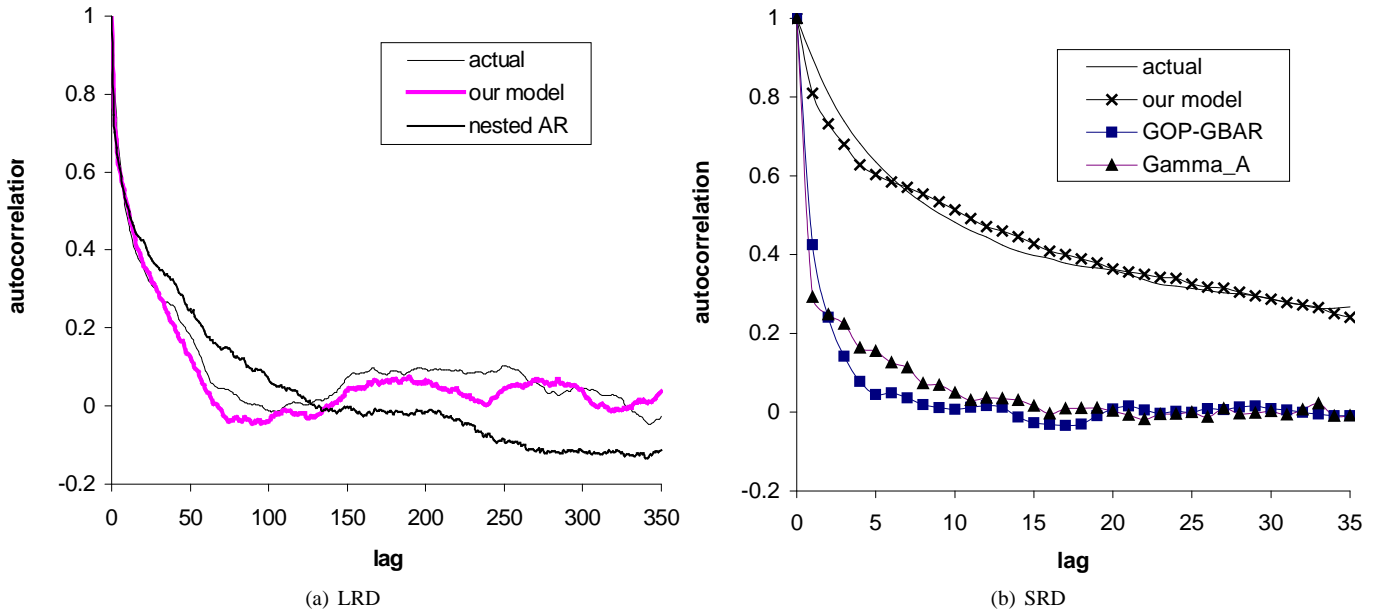
(a) LRD

(b) SRD

Fig. 3. The ACF of the actual I-frame sizes and that of the synthetic traffic in (a) long range and (b) short range.

B. Intra-GOP Correlation

Lombardo *et al*. [13] 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[3] between P/B-frames and the I-frame from the same GOP.

Before further discussion, we define I, P and B-*frame size sequences* as follows. Assuming that $n \geq 1$ represents the GOP 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. 4. 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 our modeling purposes. The correlation between $\{\phi_i^B(n)\}$ and the I-frame sequence $\{\phi^I(n)\}$ also does not change as a function of i, which we show in the following subsection.

Lombardo *et al*. [13] further modeled 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 [13] are MPEG-1 coded; however, we find that this linear estimation does not hold for general video traffic. As shown in Fig. 5 (a), the means of P and B-frames are *not* linear functions of I-frame sizes in MPEG-4 coded `Star Wars`. Therefore, we propose an alternative model for generating P and B-frame sizes, which captures the intra-GOP correlation in general GOP-based VBR video.

---

[3]In traffic modeling literature, the normalized auto-covariance function is often used instead of the autocorrelation function [15].
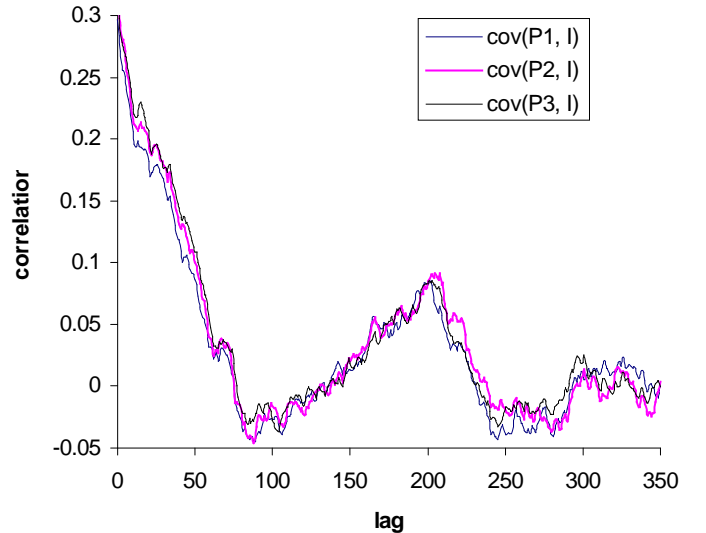

Fig. 4. The correlation between $\{\phi_i^P(n)\}$ and $\{\phi^I(n)\}$, for $i = 1, 2, 3$.

C. 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. Motivated by this observation, we propose a linear model to estimate the size of the i-th P-frame in the n-th GOP:

$$\phi_i^P(n) = a\tilde{\phi}^I(n) + \tilde{v}(n), \qquad (12)$$

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

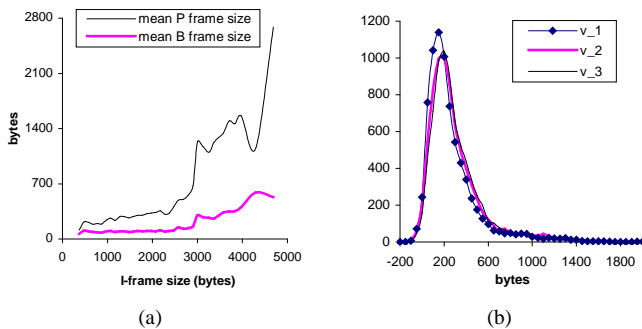*Lemma 2:* To capture the intra-GOP correlation, the value

Fig. 5. (a) The mean sizes of P and B-frames in a GOP given the size of the corresponding I-frame in `Star Wars`. (b) The histogram of $\{v(n)\}$ for $\{\phi_i^P(n)\}$ with $i = 1, 2, 3$.
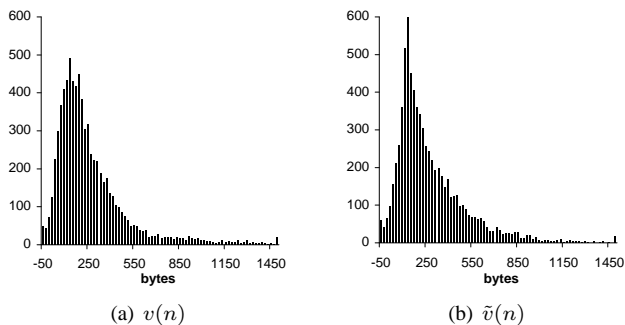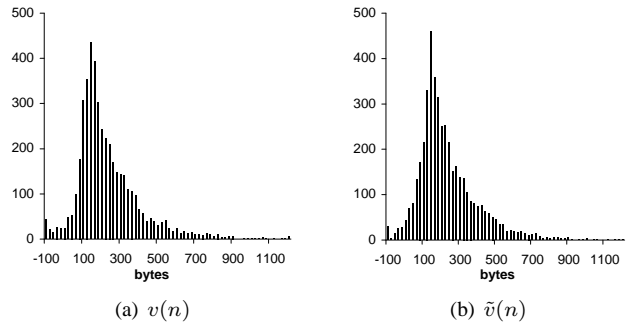

Fig. 7. Histograms of (a) the actual noise $\{v(n)\}$ for $\{\phi_2^P(n)\}$ in the base layer of `The Silence of the Lambs`; and (b) synthetic noise $\{\tilde{v}(n)\}$.

Fig. 6. Histograms of (a) the actual noise $\{v(n)\}$ for $\{\phi_1^P(n)\}$ in `Star Wars`; and (b) synthetic noise $\{\tilde{v}(n)\}$.

of coefficient a in (12) must be equal to:

$$a = \frac{r(0)\sigma_P}{\sigma_I}, \tag{13}$$

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:

$$E[\tilde{\phi}^I(n - k)] = E[\tilde{\phi}^I(n)] = 0. \tag{14}$$

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

$$C(k) = E[(\phi_i^P(n) - E[\phi_i^P])(\tilde{\phi}^I(n - k) - E[\tilde{\phi}^I])]. \tag{15}$$

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

$$
\begin{aligned}
C(k) &= E[(a\tilde{\phi}^I(n) + v(n) - E[\phi_i^P])\tilde{\phi}^I(n - k)] \\
&= aE[\tilde{\phi}^I(n)\tilde{\phi}^I(n - k)] \tag{16}
\end{aligned}
$$

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

$$r(0) = \frac{C(0)}{\sigma_P \sigma_{\tilde{I}}} = \frac{aE[\tilde{\phi}^I(n)^2]}{\sigma_P \sigma_{\tilde{I}}}, \tag{17}$$

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:

$$\frac{a \cdot \sigma_I}{\sigma_P} = r(0), \tag{18}$$

which leads to (13). ∎

To understand how to generate $\{\tilde{v}(n)\}$, 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 the single-layer `Star Wars` in Fig. 5 (b). The figure shows that the residual process $\{v(n)\}$ does not change much as a function of i. We also observe that the histogram of $\{v(n)\}$ is asymmetric and decays fast on both sides. Although a generalized Gamma distribution (including scale parameter, location parameter, and shape parameter) might be able to describe this type of distribution, its parameter estimation is quite complicated [12].

To model the asymmetry and quickly decaying trend of $v(n)$, we use two exponential distributions to estimate its PDF. We first left-shift $\{v(n)\}$ by an offset δ to make the mode (i.e., the peak) appear at zero. We then model the right side using one exponential distribution $1 - e^{\lambda_1}$ and the absolute value of the left side using another exponential distribution $1 - e^{\lambda_2}$. Afterwards, we generate synthetic data $\{\tilde{v}(n)\}$ based on these two exponential distributions and right-shift the result by δ. As shown in Fig. 6 and Fig. 7, the histograms of $\{\tilde{v}(n)\}$ are close to those of $\{v(n)\}$ in both `Star Wars` and the base layer of `The Silence of the Lambs`. Statistical parameters $(r(0), \sigma_P, \sigma_I, \lambda_1, \lambda_2)$ needed for this model are 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., [15], [25]) in Fig. 8 (a). The figure shows that our model indeed preserves the intra-GOP correlation of the original traffic, while the previous methods produce white (uncorrelated) noise.

As we shown earlier in Fig. 5 (a), the sizes of B-frames are relatively small compared to those of P-frames. From Fig. 8 (b), we observe that the correlation between $\{\phi_i^B(n)\}$ and the I-frame sequence $\{\phi^I(n)\}$ are much smaller than that between P-frame sequences and the I-frame sequence. Thus, we can generate the synthetic B-frame traffic simply by an *i.i.d.* lognormal random number generator.
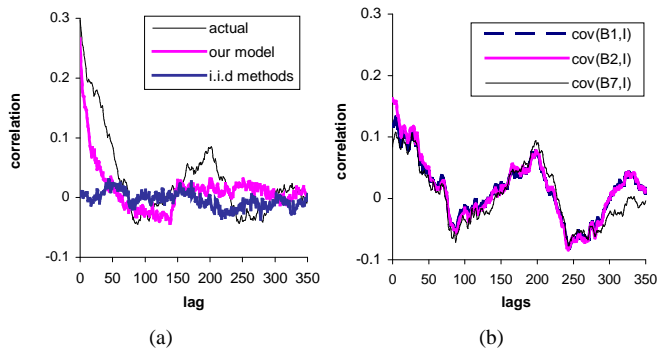
6

Fig. 8. (a) The correlation between $\{\phi_1^P(n)\}$ and $\{\phi^I(n)\}$ in `Star Wars`. (b) The correlation between $\{\phi_i^B(n)\}$ and $\{\phi^I(n)\}$, for $i = 1, 2, 7$.
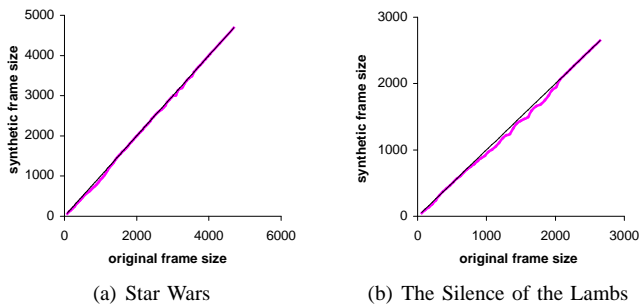

Fig. 9. QQ plots of (a) the synthetic single-layer `Star Wars` traffic and (b) the synthetic `The Silence of the Lambs` base-layer traffic.

D. Model Accuracy Study

There are two popular studies to verify the accuracy of a video traffic model [25]: quantile-quantile (QQ) plots and packet-loss evaluation. The QQ plot is a graphical technique to verify 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. The greater the departure from this reference line, the greater the difference between the two test data sets. Fig. 9 shows QQ plots of the synthetic single-layer `Star Wars` traffic and the synthetic base layer traffic of `The Silence of the Lambs`, both of which are generated by our model. As shown in the figure, the generated frame sizes and the original traffic are almost 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 router buffer with capacity c and drain rate d [25]. 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 as the difference 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:

$$e = \frac{|p - p_{model}|}{p}. \tag{19}$$

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 [6]	2.44%	2.51%	4.01%
	Nested AR [15]	4.02%	2.05%	5.63%
	Gamma_A [25]	5.54%	1.04%	0.99%
	Gamma_B [25]	5.76%	1.81%	1.15%
20ms	Our Model	0.93%	0.61%	1.13%
	GOP-GBAR [6]	3.84%	2.16%	3.77%
	Nested AR [15]	5.81%	2.77%	8.46%
	Gamma_A [25]	5.20%	0.61%	2.57%
	Gamma_B [25]	4.89%	1.93%	2.05%
30ms	Our Model	0.25%	0.33%	0.95%
	GOP-GBAR [6]	4.94%	3.33%	5.68%
	Nested AR [15]	6.94%	4.14%	9.92%
	Gamma_A [25]	4.88%	1.10%	4.48%
	Gamma_B [25]	4.67%	2.17%	4.03%

In Table I, we illustrate the values of e for various buffer capacities c and drain rates d. As shown in the table, the synthetic traffic generated by our model provides a very accurate estimate of the actual data loss probability p and significantly outperforms the other methods. In addition, our synthetic traffic is approximately 30% more accurate than the $i.i.d.$ models of prior work in estimating the loss ratio of P-frames. We should finally note that the complexity of our method is no worse than that of prior work (and sometimes, even lower) and that simulations with additional video sequences have demonstrated results similar to those shown throughout this paper.

IV. MODELING MULTI-LAYER VIDEO TRAFFIC

We next investigate methods to capture cross-layer dependency and model enhancement-layer traffic. Due to its flexibility and high bandwidth utilization, layered video coding is common in video applications. Layered coding is often referred to as "scalable coding," which can be further classified as coarse-granular (e.g., spatial scalability) or fine-granular (e.g., fine granular scalability (FGS)) [26]. The major difference between coarse granularity and fine granularity is that the former provides quality improvements only when a *complete* enhancement layer has been received, while the latter continuously improves video quality with every additionally received codeword of the enhancement bitstream.

In both coarse granular and fine granular coding methods, an enhancement layer is coded with the residual between the original image and the reconstructed image from the base layer. Therefore, the enhancement layer has a strong dependency on the base layer. Zhao *et al.* [27] also indicate that there exists a cross-correlation between the base layer and the enhancement layer; however, this correlation has not been fully addressed in previous studies. In the next subsection, we investigate the cross-correlation between the enhancement layer and the base layer using spatially-scalable `The Silence of the Lambs` sequence and an FGS-coded `Star Wars` sequence as examples. For brevity, we only show the analysis of two-
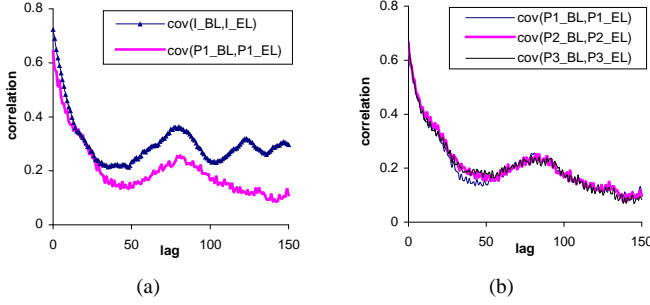
Fig. 10. (a) The correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$, $\{\varepsilon_1^P(n)\}$ and $\{\phi_1^P(n)\}$ in Star Wars. (b) The correlation between $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ in The Silence of the Lambs for $i = 1, 2, 3$.
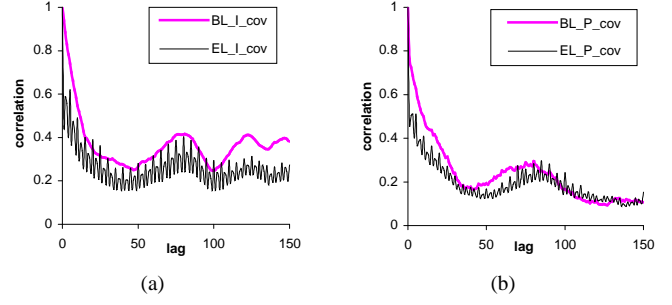

Fig. 11. (a) The ACF of $\{\varepsilon^I(n)\}$ and that of $\{\phi^I(n)\}$ in Star Wars. (b) The ACF of $\{\varepsilon_1^P(n)\}$ and that of $\{\phi_1^P(n)\}$ in The Silence of the Lambs.
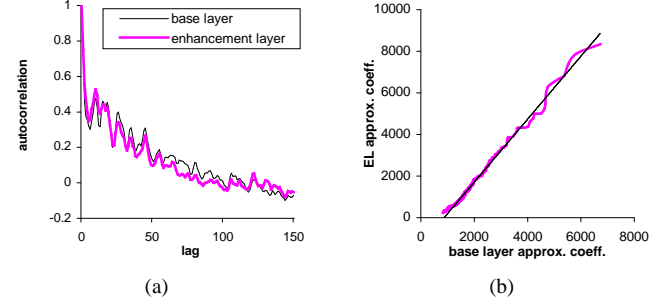

Fig. 12. (a) The ACF of $\{A_3(\varepsilon)\}$ and $\{A_3(\phi)\}$ in The Silence of the Lambs. (b) QQ plot of $\{A_3(\varepsilon)\}$ and $\{A_3(\phi)\}$ in Silence of the Lambs.

layer sequences and note that similar results hold for video streams with more than two layers.

A. Analysis of the Enhancement Layer

For discussion convenience, we define the enhancement layer frame sizes as follows. Similar to the definition in the base layer, we define $\varepsilon^I(n)$ to be the I-frame size of the n-th GOP, $\varepsilon_i^P(n)$ to be the size of the i-th P-frame in GOP n, and $\varepsilon_i^B(n)$ to be the size of the i-th B-frame in GOP n.

Since each frame in the enhancement layer is predicted from the corresponding frame in the base layer, we examine the cross-correlation between the enhancement layer frame sizes and the corresponding base layer frame sizes in various sequences. In Fig. 10 (a), we display the correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ and that between $\{\varepsilon_1^P(n)\}$ and $\{\phi_1^P(n)\}$, which are labeled as cov(I_BL, I_EL) and cov(P1_BL, P1_EL), respectively. As shown in the figure, the correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ is stronger than that between $\{\varepsilon_1^P(n)\}$ and $\{\phi_1^P(n)\}$, especially at large lags. This observation indicates that $\{\varepsilon^I(n)\}$ exhibits LRD properties and we should preserve these properties in the synthetic enhancement layer I-frame sizes.

In Fig. 10 (b), we show the cross-correlation between processes $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ for $i = 1, 2, 3$. The figure demonstrates that the correlation between the enhancement layer and the base layer is quite strong, and the correlation structures between each $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ are very similar to each other. To avoid repetitive description, we do not show the correlation between $\{\varepsilon_i^B(n)\}$ and $\{\phi_i^B(n)\}$, which is similar to that between $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$.

Aside from cross-correlation, we also examine the auto-correlation of each frame sequence in the enhancement layer and that of the corresponding sequence in the base layer. We show the ACF of $\{\varepsilon^I(n)\}$ and that of $\{\phi^I(n)\}$ (labeled as "EL_I_cov" and "BL_I_cov", respectively) in Fig. 11 (a). The figure shows that although the ACF structure of $\{\varepsilon^I(n)\}$ has some oscillation, its trend closely follows that of $\{\phi^I(n)\}$. In addition, we display the ACF of $\{\varepsilon_1^P(n)\}$ and that of $\{\phi_1^P(n)\}$ in Fig. 11 (b). From this figure and other experimental results, one also observes that the ACF structures of processes $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ are similar to each other.

B. Modeling Frame Sizes

Although cross-layer correlation is obvious in multi-layer traffic, previous work neither considered it during modeling [1], nor explicitly addressed the issue of preserving it in the synthetic traffic [27]. In this section, we first describe how we model the enhancement layer frame sizes and then evaluate the performance of our model in capturing the cross-layer correlation.

Recalling that $\{\varepsilon^I(n)\}$ also possesses both SRD and LRD properties, we model it in the wavelet domain as we modeled $\{\phi^I(n)\}$. We define $\{A_j(\varepsilon)\}$ and $\{A_j(\phi)\}$ to be the approximation coefficients of $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ at the wavelet decomposition level j, respectively. To better understand the relationship between $\{A_j(\varepsilon)\}$ and $\{A_j(\phi)\}$, we show the ACF of $\{A_3(\varepsilon)\}$ and $\{A_3(\phi)\}$ using Haar wavelets in Fig. 12 (a). The figure shows that $\{A_3(\varepsilon)\}$ and $\{A_3(\phi)\}$ have a similar ACF structure. We also display a QQ plot of $\{A_3(\phi)\}$ (the x-axis) and $\{A_3(\varepsilon)\}$ (the y-axis) in Fig. 12 (b). The straight (but not diagonal) line in the figure shows that the two sets of coefficients come from the same distribution but with different parameters [18].

As shown in Fig. 12, $\{A_j(\varepsilon)\}$ and $\{A_j(\phi)\}$ exhibit similar ACF structure and come from the same distribution. Thus, we generate $\{A_J(\varepsilon)\}$ by borrowing the ACF structure of $\{A_J(\phi)\}$, which is known from our base-layer model. Using the ACF of $\{A_J(\phi)\}$ in modeling $\{\varepsilon^I(n)\}$ not only saves computational cost, but also preserves the cross-layer correlation.

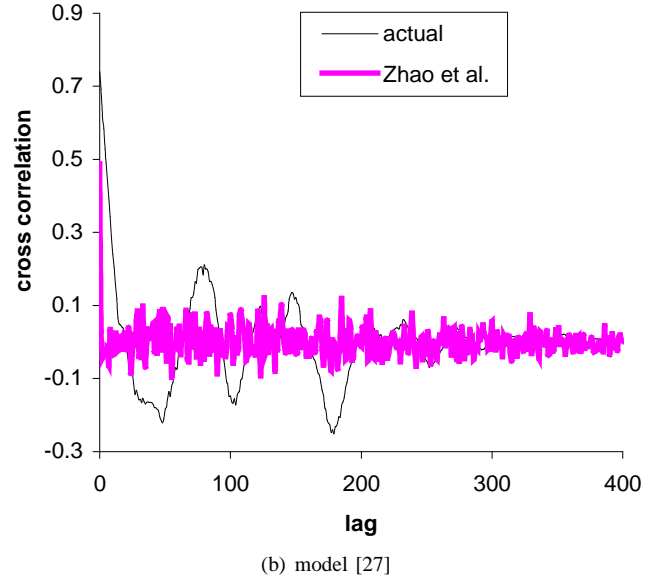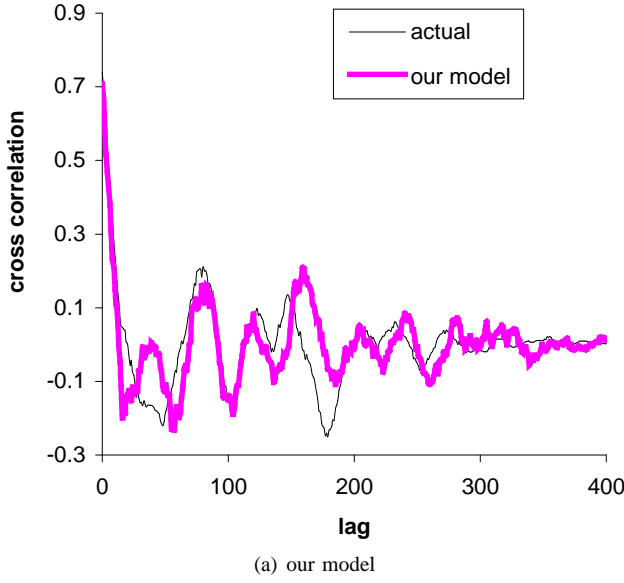(a) our model

(b) model [27]

Fig. 13. The cross-correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ in The Silence of the Lambs and that in the synthetic traffic generated from (a) our model and (b) model [27].
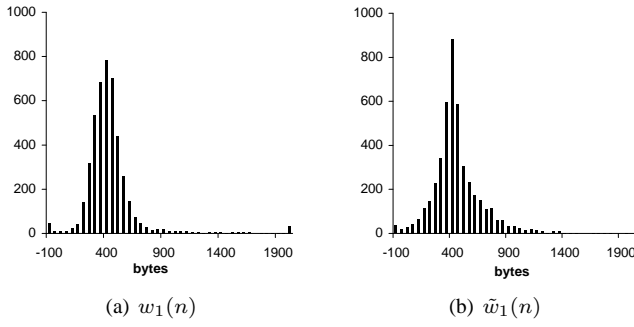

(a) $w_1(n)$

(b) $\tilde{w}_1(n)$

Fig. 14. (a) The histogram of $\{w_1(n)\}$ and (b) that of $\{\tilde{w}_1(n)\}$ when modeling $\{\varepsilon_1^P(n)\}$ in The Silence of the Lambs.
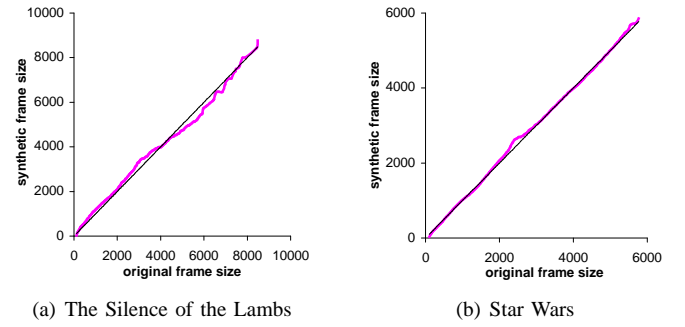

(a) The Silence of the Lambs

(b) Star Wars

Fig. 15. QQ plots of the synthetic enhancement-layer traffic: (a) The Silence of the Lambs and (b) Star Wars.

In Fig. 13, we compare the actual cross-correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ to that between the synthetic $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ generated from our model and Zhao's model [27]. The figure shows that our model significantly outperforms Zhao's model in preserving the cross-layer correlation.

Furthermore, recall that the cross-correlation between $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ and that between $\{\varepsilon_i^B(n)\}$ and $\{\phi_i^B(n)\}$ are also strong, as shown in Fig. 10. We use the linear model from Section III-C to estimate the sizes of the i-th P and B-frames in the n-th GOP:

$$\varepsilon_i^P(n) = a\phi_i^P(n) + \tilde{w}_1(n), \tag{20}$$
$$\varepsilon_i^B(n) = a\phi_i^B(n) + \tilde{w}_2(n), \tag{21}$$

where $a = r(0)\sigma_\varepsilon/\sigma_\phi$, $r(0)$ is the lag-0 cross-correlation coefficient, σ_ε is the standard deviation of the enhancement-layer sequence, and σ_ϕ is the standard deviation of the corresponding base-layer sequence. Processes $\{\tilde{w}_1(n)\}, \{\tilde{w}_2(n)\}$ are independent of $\{\phi_i^P(n)\}$ and $\{\phi_i^B(n)\}$. As described in Section III-C, we also examine $\{w_1(n)\}$ and $\{w_2(n)\}$, and

find that they can be modeled in the same way that we model $\{v(n)\}$, as shown in Fig. 14.

We evaluate the accuracy of the synthetic enhancement layer by using QQ plots and show two examples in Fig. 15, which displays two QQ plots of the synthetic The Silence of the Lambs and Star Wars enhancement-layer traffic. The figure shows that the synthetic frame sizes in both sequences have the same distribution as those in the original traffic.

We next examine the data loss ratio predicted by our synthetic traffic passed through a generic buffer as shown in Section III-D. We are not able to show simulation results for the other multi-layer models given our sample sequences since the model in [27] is suitable only for short sequences and the one in [1] is only applicable to sequences with a CBR base layer. In Fig. 16 and Fig. 17, we show the overflow data loss ratio of the enhancement layers in both The Silence of the Lambs and Star Wars with different drain rates d for buffer capacity $c = 10$ ms and $c = 30$ ms, respectively. The x-axis in the figure represents the ratio of the drain rates to
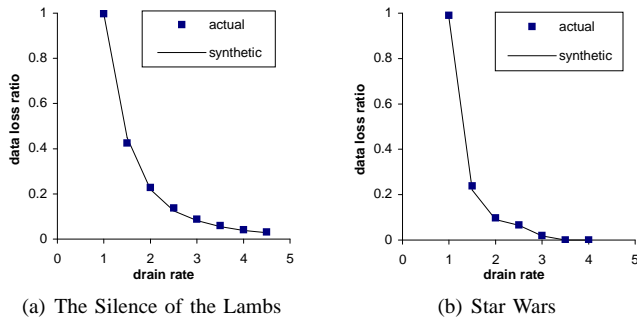
(a) The Silence of the Lambs   (b) Star Wars

Fig. 16. Overflow data loss ratio of the original and synthetic enhancement layer traffic for $c = 10$ ms for (a) `The Silence of the Lambs` and (b) `Star Wars`.

(a) The Silence of the Lambs   (b) Star Wars

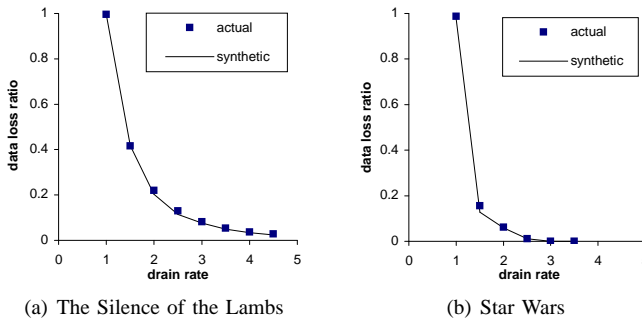Fig. 17. Overflow data loss ratio of the original and synthetic enhancement layer traffic for $c = 30$ ms for (a) `The Silence of the Lambs` and (b) `Star Wars`.

the average traffic rate $\bar{r}$. The figure shows that the synthetic enhancement layer preserves the temporal information of the original traffic very well.

V. CONCLUSION

In this paper, we presented a framework for modeling multi-layer full-length VBR video traffic. This framework incorporated wavelet-domain analysis into time-domain modeling. This work precisely captured the LRD as well as SRD properties of video traffic, accurately described the intra-GOP correlation in compressed VBR sequences, and proposed novel methods to model cross-layer correlation in multi-layer sequences. Since our framework is developed at frame-size level (whereas much of the previous work uses slice-level or even block-level [25]), we can examine the loss ratio for each type of frames and apply other methods (e.g., guaranteeing the transmission of certain frames) to improve the video quality at the receiver.

Furthermore, we derived the sizes of P/B-frames based on their dependency on the I-frame belonging to the *same* GOP, which makes our framework also applicable to the case that I-frames are unexpectedly inserted and GOP structure are changed. In future work, our traffic model will also be helpful in designing a layered peer-to-peer video system.

REFERENCES

[1] K. Chandra and A. R. Reibman, "Modeling One- and Two-Layer Variable Bit Rate Video," *IEEE/ACM Trans. on Networking*, vol. 7, June 1999.
[2] T. P.-C. Chen and T. Chen, "Markov Modulated Punctured Autoregressive Processes for Video Traffic and Wireless Channel Model-ing," *Packet Video*, April 2002.
[3] A. L. Corte, A. Lombardo, S. Palazzo, and S. Zinna, "Modeling Activity in VBR Video Sources," *Signal Processing: Image Commu-nication*, vol. 3, June 1991.
[4] P. Embrechts, F. Lindskog, and A. McNeil, "Correlation and Dependence in Risk Management: Properties and Pitfalls," *Cambridge University Press*, 2002.
[5] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation (Extended Version)," available at *http://www-tkn.ee.tu-berlin.de*.
[6] 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.
[7] M. Krunz and S. K. Tripathi, "On the Characterization of VBR MPEG Streams," *Proc. of ACM SIGMETRICS*, vol. 25, June 1997.
[8] M. Krunz and A. Makowski, "Modeling Video Traffic Using M/G/infinity Input Processes: A Compromise Between Markovian and LRD Models," *IEEE J. Select. Areas Commun.*, vol. 16, June 1998.
[9] M. W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," *Proc. of ACM SIGCOMM*, Aug. 1994.
[10] D. P. Heyman, "The GBAR Source Model for VBR Video Conferences," *IEEE/ACM Trans. on Networking*, vol. 5, Aug. 1997.
[11] 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 ACM SIGCOMM*, Aug. 1995.
[12] T.-Y. Hwang and P.-H. Huang, "On New Moment Estimation of Parameters of the Gamma Distribution Using its Characterization," *Annals of the Institute of Statistical Mathematics*, vol. 54, Issue 4, 2002.
[13] A. Lombardo, G. Morabito, and G. Schembra, "An Accurate and Treatable Markov Model of MPEG-Video Traffic," *Proc. of INFOCOM*, March 1998.
[14] A. Lombardo, G. Morabito, S. Palazzo, and G. Schembra, "A Markov-Based Algorithm for the Generation of MPEG Sequences Matching Intra- and Inter-GoP Correlation," *European Trans. on Telecommuni-cations*, vol. 12, April 2001.
[15] D. Liu, E. I. Sára, and W. Sun, "Nested Auto-Regressive Processes for MPEG-Encoded Video Traffic Modeling," *IEEE Trans. on CSVT*, vol. 11, Feb. 2001.
[16] S. Ma and C. Ji, "Modeling Heterogeneous Network Traffic in Wavelet Domain," *IEEE/ACM Trans. on Networking*, vol. 9, Oct. 2001.
[17] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins, "Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Trans. on Comm.*, vol. 36, July 1988.
[18] "Matlab documents," *http://www.mathworks.com*.
[19] B. Melamed and D. E. Pendarakis, "Modeling Full-Length VBR Video Using Markov-Renewal-Modulated TES Models," *IEEE J. Select. Areas Commun.*, vol. 16, June 1998.
[20] MPEG, "Coding of Moving Pictures and Audio," ISO/IEC JTC1/SC29/WG11 N3908, Jan. 2001.
[21] M. Reisslein, J. Lassetter, S. Ratnam, O. Lotfallah, F. H. P. Fitzek, and S. Panchanathan, "Video Traces for Network Performance Evaluation," available at http://trace.eas.asu.edu.
[22] 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*, March 2000.
[23] O. Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems," *Proc. of the 20th Annual Conference on Local Computer Networks*, Oct. 1995.
[24] O. Rose, "Simple and Efficient Models for Variable Bit Rate Mpeg Video Traffic," *Performance Evaluation*, vol. 30, 1997.
[25] 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.
[26] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Prentice Hall, 2001.
[27] J.-A. Zhao, B. Li, and I. Ahmad, "Traffic Model For Layered Video: An Approach On Markovian Arrival Process," *Packet Video*, April 2003.