

Analysis and Modeling of MPEG-4 and H.264 Multi-Layer Video Traffic

Min Dai

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

Dmitri Loguinov*

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

Abstract— This paper presents a frame-level hybrid framework for modeling H.264 and MPEG-4 multi-layer variable bit rate (VBR) video traffic. In this work, the base layer is modeled using a combination of wavelet and time-domain methods and the enhancement layer is linearly predicted from the base layer using the cross-layer correlation. Unlike previous studies, we analyze and successfully model both inter-GOP and intra-GOP correlation in VBR sequences. To accurately capture long-range dependent (LRD) and short-range dependent (SRD) properties of VBR traffic, we use wavelets to model the distribution of I-frame sizes and a simple time-domain model for P/B frame sizes. Simulation results demonstrate that our model effectively preserves the temporal burstiness and captures important statistical features (e.g., the autocorrelation function and the frame-size distribution) of original traffic. We also show that our model has better performance than the previous methods in both single and multi-layer sequences.

Index Terms— Wavelet transform, Multi-layer, Video traffic, Enhancement layer

I. INTRODUCTION

Video traffic modeling plays an important role in the characterization and analysis of network traffic. Besides providing an insight into the coding process and structure of video sequences, traffic models can be used for many practical purposes including allocation of network resources, design of efficient networks for streaming services, and delivery of certain Quality of Service (QoS) guarantees to end users.

Although many studies have been conducted in this area, most existing traffic models only apply to single-layer VBR video and often overlook the multi-layer aspects of streaming video traffic in the current Internet [1], [28]. In addition, traffic modeling research is falling behind the rapid advances in video techniques, which include standards MPEG-4 [21] and H.264 [14]. Therefore, the goal of this work is to better understand the statistical properties of various video sequences and to develop a model that can generate synthetic traffic with the properties close to those of original single/multi-layer MPEG-4 and H.264 video sequences.

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 the 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 [8], Gamma [26], and various hybrid distributions (e.g., Gamma/Pareto [17] or Gamma/lognormal [24]).

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 long-range dependent (LRD) and short-range dependent (SRD) properties [10], [18]. 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 [10]. Thus, using either a long-range dependent or a short-range dependent model *alone* does not provide satisfactory results. Many studies have been conducted to address this problem, but only a few of them have managed to model the complicated LRD/SRD ACF structure of real video traffic (e.g., [17], [18]). Furthermore, the correlation that most models try to capture is the *inter-GOP* (Group of Pictures) correlation, which is well characterized by the ACF of the I-frames. However, another dimension of video traffic, the *intra-GOP* correlation¹, is rarely addressed in related work, even though it is an important characteristic useful in computing precise bounds on network packet loss [16].

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 using estimated wavelet coefficients, which are more mathematically tractable than actual coefficients. After a thorough analysis of intra-GOP correlation, we generate synthetic P-frame traffic using a time-domain linear model of the preceding I-frame to preserve the intra-GOP correlation. We use a similar model to preserve the cross-layer correlation in multi-layer video sequences and show that the performance of the resulting model is better than

*Supported by NSF grants CCR-0306246, ANI-0312461, and CNS-0434940.

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

that of prior methods.

The specifics of four one-hour video sequences discussed in this paper are as following: a single layer MPEG-4 Star Wars IV [6] (25 frames/s), a single layer H.264 Starship Troopers [22] (25 frames/s), a two-layer MPEG-4 spatially-scalable The Silence of the Lambs [22] (30 frames/s), and a two-layer MPEG-4 FGS-coded Star Wars IV [22] (30 frames/s). All four sequences have GOP structure *IBBPBBPBBPBB*.

This paper is organized as follows. In Section II, we overview the related work on traffic modeling. In Section III, we provide the background on wavelet analysis and show how to generate synthetic I-frame sizes in the wavelet domain. In Section IV, we discuss the intra-GOP correlation in various sequences and present a linear model for P and B-frame sizes. Section V analyzes the cross-correlation between the base layer and the enhancement layer, and explains how to generate a synthetic enhancement layer based on this information. In Section VI, we evaluate the accuracy of our model using both single-layer and multi-layer video traffic. Section VII concludes the paper.

II. RELATED WORK

Numerous studies have been conducted in modeling VBR video traffic and a variety of models have been proposed in the literature. In this section, we briefly overview related work on single-layer and multi-layer traffic models.

A. Single-Layer Models

According to the dominant stochastic method applied in each model, we group existing single-layer models into several categories and present the main results of each group below.

We first discuss AR models, since they are classical approaches in the area of traffic modeling. After the first autoregressive (AR) model was applied to video traffic in 1988 [19], AR processes and their variations remain highly popular in this area of research [17]. 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) process 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.* [8] model the deviation of I-frame sizes from their mean in each scene using an AR(2) process. Building upon Krunz' work [8], Liu *et al.* [17] propose a *nested* AR(2) model, which uses a second AR(2) process to model the mean frame-size of each scene. In both cases, scene changes are detected and scene length is modeled as a geometrically distributed random variable.

To model videoconferencing data, Heyman *et al.* propose a discrete autoregressive (DAR) model in [11] and a GBAR model in [12], the latter of which has Gamma-distributed marginal statistics and a geometric autocorrelation. By considering the group-of-picture (GOP) cyclic structure of video traffic, Frey *et al.* [7] extend the GBAR model in [12] 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 [15]). Rose [25] 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.* [26] 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 usually difficult to accurately define and segment video sources into the different states in the time domain due to the dynamic nature of video traffic [18].

The third category consists of self-similar processes and fractal models. Garrett *et al.* [10] 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 [10], Huang *et al.* [13] present a self-similar fractal traffic model; however, this model does not capture the multi-timescale variations in video traffic [8].

Other approaches include the $M/G/\infty$ process [9] and Transform-Expand-Sample (TES) based models [20]. The former creates SRD traffic [17] and the latter has high computational complexity and often requires special software (e.g., *TESool*) to generate synthetic sequences. Different from the above time-domain methods, several wavelet models [18], [23] recently emerged due to their ability to accurately capture both LRD and SRD properties of video traffic [18]. We provide more background on wavelets and an initial analysis of approximation coefficients in section III-A.

B. Multi-Layer Models

Most traffic modeling studies focus on single-layer video traffic and rarely model multi-layer sequences. Among several multi-layer studies, 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 a 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.* [28] build a K -state Markov chain based on frame-size clusters. The clustering feature in [28] 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 [28] limits its application only to short video sequences. Furthermore, in both [1] and [28], there is no general method for choosing the optimal number of states and the necessary parameters are often selected empirically.

III. MODELING I-FRAME SIZES

In this section, we address the problem of modeling I-frame sizes and show a novel method for estimating the coefficients of the wavelet transform. Using the estimated wavelet coefficients, we later generate synthetic I-frame sizes, which preserve the LRD and SRD properties of the original traffic.

A. 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 [18].

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. Assume that $j = J$ is the coarsest scale and $j = 0$ is the original signal. For discussion convenience, 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 . Notice that $\{A_j\}$ is a random process $A_j = (A_j^1, A_j^2, \dots, A_j^k, \dots)$ and A_j^k is a random variable.

In the following discussion, we use the Haar wavelet transform as a typical example since it is often chosen for its simplicity and good performance [18], [23]. Recall that the Haar approximation coefficients A_j^k are obtained via [23]:

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

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 IV 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 *i.i.d.* (uncorrelated) noise. This explains why previous literature commonly models detailed coefficients as zero-mean *i.i.d.* Gaussian variables [18]. 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.

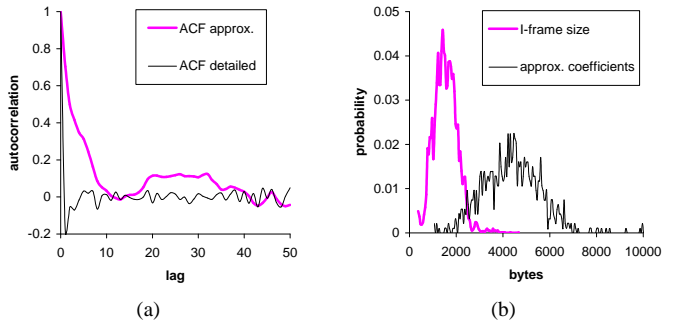


Fig. 1. (a) The ACF structure of coefficients $\{A_3\}$ and $\{D_3\}$ in single-layer Star Wars IV. (b) The histogram of I-frame sizes and that of approximation coefficients $\{A_3\}$.

Recalling that I-frame sizes $\{A_0\}$ follow a Gamma distribution [24], 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.

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

Proof: See [4]. ■

Extensive experimental results show that a single Gamma distribution is accurate enough to describe the actual histogram of $\{A_j\}$. 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 IV in Fig. 1 (b). The figure shows that the two distributions have a similar Gamma shape, but with different parameters. In the next section, we use this information to efficiently estimate the approximation coefficients.

B. 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 [18], [23], 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 [18], 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 IV 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 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 $\{D_1\}$.

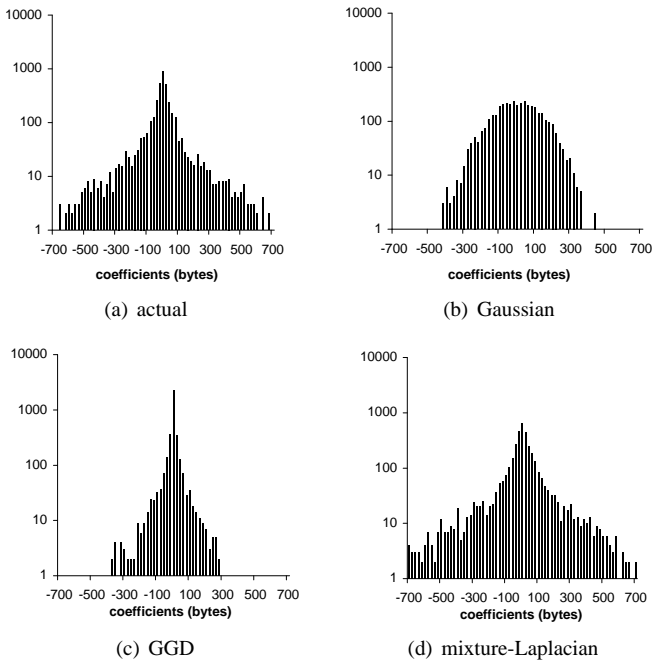


Fig. 2. Histograms of (a) the actual detailed coefficients; (b) the Gaussian model; (c) the GGD model; and (d) the mixture-Laplacian model.

Additional simulations (not shown for brevity) demonstrate that a single Laplacian distribution is not able to describe the fast decay and large data range of the actual histogram; however, a *mixture*-Laplacian distribution follows the real data very well:

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

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 [18] or Beta random variables [23]. However, as mentioned in Section III-A, the approximation coefficients are non-negligibly 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

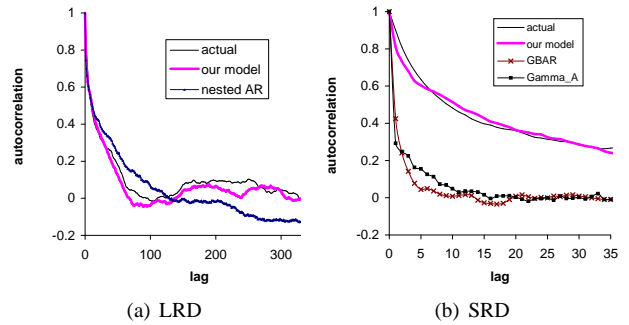


Fig. 3. The ACF of the actual I-frame sizes and that of the synthetic traffic in (a) long range and (b) short range.

reasonable value (e.g., the average scene length²). 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 [5].

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 in long range. Fig. 3 (b) shows the correlation of the synthetic traffic from the GOP-GBAR model [7] and Gamma_A model [26] 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.

IV. MODELING P AND B-FRAME SIZES

We next model P-frame sizes in the time domain based on intra-GOP correlation. The framework in this section has two contributions: (1) provide a detailed analysis of intra-GOP correlation for various video sequences, and (2) model intra-GOP correlation and propose a simple model that accurately generates synthetic P/B-frame sizes based on intra-GOP correlation, which is in contrast to much of the previous work that relied on *i.i.d.* random variables to model the P/B-frame sizes in each GOP [8], [13], [17], [26].

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.

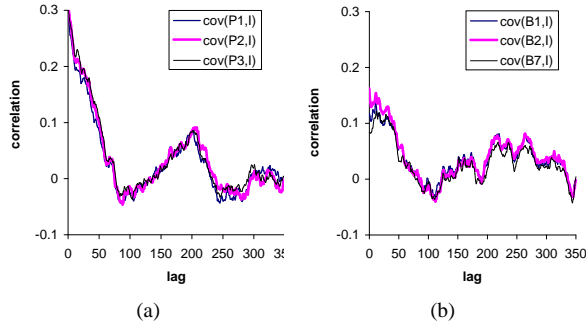


Fig. 4. (a) The correlation between $\{\phi_i^P(n)\}$ and $\{\phi^I(n)\}$ in Star Wars IV, for $i = 1, 2, 3$. (b) The correlation between $\{\phi_i^B(n)\}$ and $\{\phi^I(n)\}$ in Star Wars IV, for $i = 1, 2, 7$.

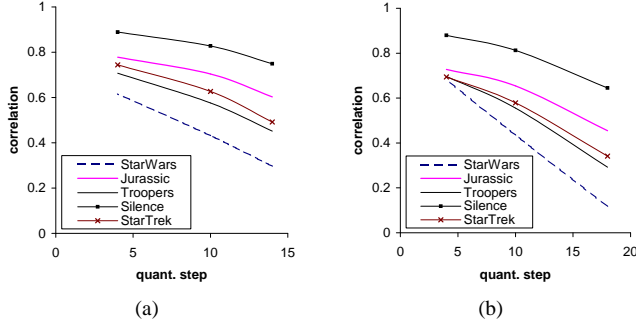


Fig. 5. (a) The correlation between $\{\phi^I(n)\}$ and $\{\phi_1^P(n)\}$ in MPEG-4 sequences coded at $Q = 4, 10, 14$. (b) The correlation between $\{\phi^I(n)\}$ and $\{\phi_1^B(n)\}$ in MPEG-4 sequences coded at $Q = 4, 10, 18$.

A. Intra-GOP Correlation

Lombardo *et al.* [15] noticed that there is a strong correlation³ between the P/B-frame sizes and the I-frame size belonging to the same GOP, which is also called intra-GOP correlation. Motivated by their results, we conduct the analysis of the intra-GOP correlation between $\{\phi^I(n)\}$ and $\{\phi_i^P(n)\}$ or $\{\phi_i^B(n)\}$ in two situations: (a) the intra-GOP correlation for different i in a specific video sequence with fixed quantization step Q ; and (b) the intra-GOP correlation for the same i in various sequences coded at different steps Q .

For the first part of our analysis, we display the correlation between $\{\phi^I(n)\}$ and $\{\phi_i^P(n)\}$ and that between $\{\phi^I(n)\}$ and $\{\phi_i^B(n)\}$ in single-layer Star Wars IV for $i = 1, 2, 3$ in Fig. 4. As shown in the figure, the correlation is almost identical for different i , which is rather convenient for our modeling purposes. For the second part of our analysis, we examine the various video sequences coded at different quantization steps to understand the relationship between intra-GOP correlation and quantization steps. We show the correlation between $\{\phi^I(n)\}$ and $\{\phi_1^P(n)\}$ and that between $\{\phi^I(n)\}$ and $\{\phi_1^B(n)\}$ in five MPEG-4 coded video sequences in Fig. 5.

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

³In traffic modeling literature, the normalized auto-covariance function is often used instead of the autocorrelation function [17].

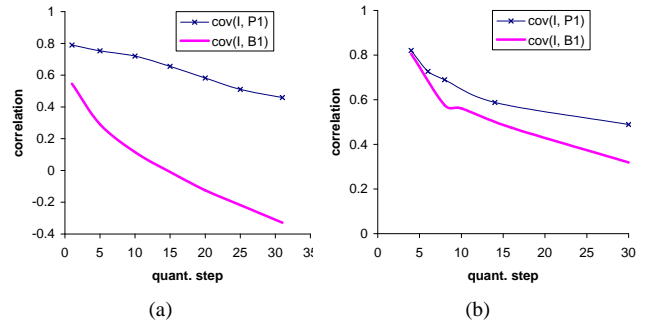


Fig. 6. The correlation between $\{\phi^I(n)\}$ and $\{\phi_1^P(n)\}$ and that between $\{\phi^I(n)\}$ and $\{\phi_1^B(n)\}$ in (a) H.264 Starship Troopers and (b) the base layer of the spatially scalable The Silence of the Lambs coded at different Q .

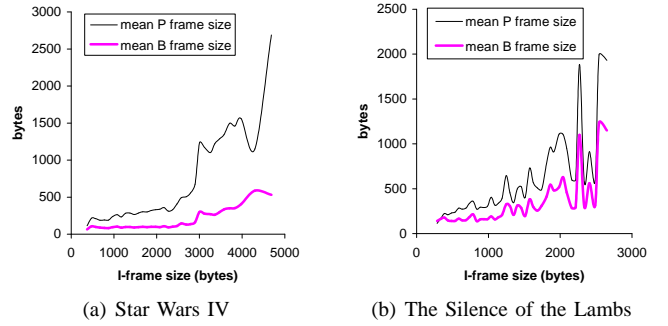


Fig. 7. The mean sizes of P and B-frames of each GOP given the size of the corresponding I-frame in (a) the single-layer Star Wars IV and (b) the base layer of the spatially scalable The Silence of the Lambs.

We also show the same correlation in H.264 coded Starship Troopers [22] and in the base layer of the spatially scalable The Silence of the Lambs in Fig. 6 (a) and (b), respectively. As observed from Fig. 5 and Fig. 6, the intra-GOP correlation decreases as the quantization step increases.

To better model P and B-frame sizes, we also investigate the relationship between P/B-frame sizes and the size of I-frame from the same GOP. Lombardo *et al.* [15] modeled the sizes of MPEG-1 coded P/B-frames as Gamma distributed random variables, with mean and variance estimated by a linear function of $\{\phi^I(n)\}$. However, we find that this linear estimation does not hold for general video traffic. As shown in Fig. 7, the means of P and B-frames are *not* linear functions of I-frame sizes in MPEG-4 coded Star Wars IV and The Silence of the Lambs. Therefore, in the next section, we propose an alternative model for generating P and B-frame sizes, which captures the intra-GOP correlation in general GOP-based VBR video.

B. 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

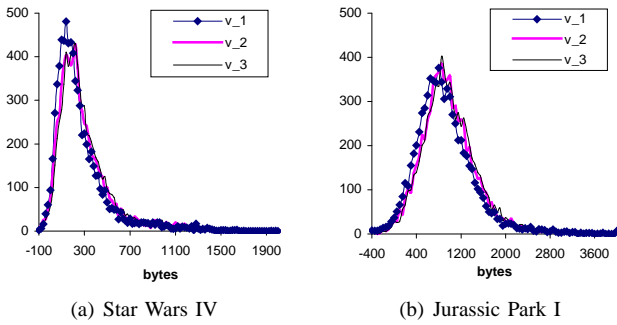


Fig. 8. Histograms of $\{v(n)\}$ for $\{\phi_i^P(n)\}$ with $i = 1, 2, 3$ in (a) Star Wars IV and (b) Jurassic Park I. Both sequences are coded at $Q = 14$.

to estimate the size of the i -th P-frame in the n -th GOP:

$$\phi_i^P(n) = a\tilde{\phi}^I(n) + \tilde{v}(n), \quad (3)$$

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 of coefficient a in (3) must be equal to:

$$a = \frac{r(0)\sigma_P}{\sigma_I}, \quad (4)$$

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:

$$E[\tilde{\phi}^I(n-k)] = E[\tilde{\phi}^I(n)] = 0. \quad (5)$$

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

$$C(k) = E[(\phi_i^P(n) - E[\phi_i^P(n)])(\tilde{\phi}^I(n-k) - E[\tilde{\phi}^I(n)])]. \quad (6)$$

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:

$$\begin{aligned} C(k) &= E[(a\tilde{\phi}^I(n) + v(n) - E[\phi_i^P(n)])\tilde{\phi}^I(n-k)] \\ &= aE[\tilde{\phi}^I(n)\tilde{\phi}^I(n-k)] \end{aligned} \quad (7)$$

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

$$r(0) = \frac{C(0)}{\sigma_P\sigma_I} = \frac{aE[\tilde{\phi}^I(n)^2]}{\sigma_P\sigma_I}, \quad (8)$$

where σ_I is the standard deviation of $\tilde{\phi}^I(n)$. Recalling that $E[\tilde{\phi}^I(n)] = 0$, we have $E[\tilde{\phi}^I(n)^2] = \sigma_I^2 = \sigma_I^2$ and:

$$\frac{a \cdot \sigma_I}{\sigma_P} = r(0), \quad (9)$$

which leads to (4). ■

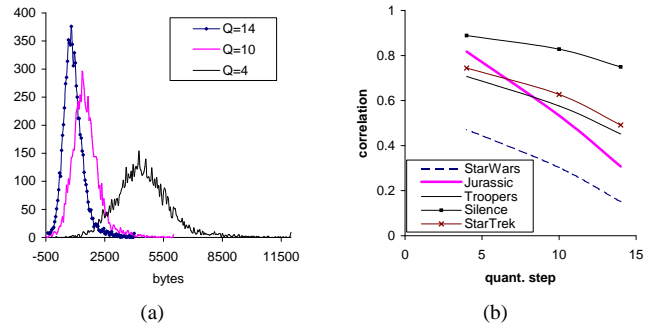


Fig. 9. (a) Histograms of $\{v(n)\}$ for $\{\phi_1^P(n)\}$ in Jurassic Park I coded at $Q = 4, 10, 14$. (b) Linear parameter a for modeling $\{\phi_i^P(n)\}$ in various sequences coded at different Q .

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 IV and Jurassic Park I in Fig. 8. The figures shows that the residual process $\{v(n)\}$ does not change much as a function of i .

In Fig. 9 (a), we show the histograms of $\{v(n)\}$ for sequences coded at different Q . The figure shows that the histogram becomes more Gaussian-like when Q increases. Due to the diversity of the histogram of $\{v(n)\}$, we use a generalized Gamma distribution $Gamma(\gamma, \alpha, \beta)$ to estimate $\{v(n)\}$. Fig. 9 (b) shows that the smaller the quantization step Q , the larger the value of parameter a in (4), which is helpful for modeling sequences coded from the same video content but at different quantization steps.

From Fig. 5 (b), we observe that the correlation between $\{\phi_i^B(n)\}$ and $\{\phi^I(n)\}$ could be as small as 0.1 (e.g., in Star Wars IV coded at $Q = 18$) or as large as 0.9 (e.g., in The Silence of the Lambs coded at $Q = 4$). Thus, we can generate the synthetic B-frame traffic simply by an *i.i.d.* lognormal random number generator when the correlation between $\{\phi_i^B(n)\}$ and $\{\phi^I(n)\}$ is small, or by a linear model similar to (3) when the correlation is large. The linear model has the following form:

$$\phi_i^B(n) = a\tilde{\phi}^I(n) + \tilde{v}_B(n), \quad (10)$$

where $a = r(0)\sigma_B/\sigma_I$, $r(0)$ is the lag-0 correlation between $\{\phi^I(n)\}$ and $\{\phi_i^B(n)\}$, σ_B and σ_I are the standard deviation of $\{\phi_i^B(n)\}$ and $\{\phi^I(n)\}$, respectively. Process $\tilde{v}_B(n)$ is independent of $\tilde{\phi}^I(n)$.

We illustrate the difference between our model and a typical *i.i.d.* method of prior work (e.g., [17], [26]) in Fig. 10. The figure shows that our model indeed preserves the intra-GOP correlation of the original traffic, while the previous methods produce white (uncorrelated) noise. Statistical parameters ($r(0), \sigma_P, \sigma_I, \gamma, \alpha, \beta$) needed for this model are easily estimated from the original sequences.

V. MODELING THE ENHANCEMENT LAYER

In this section, we provide brief background knowledge of multi-layer video, investigate methods to capture cross-layer

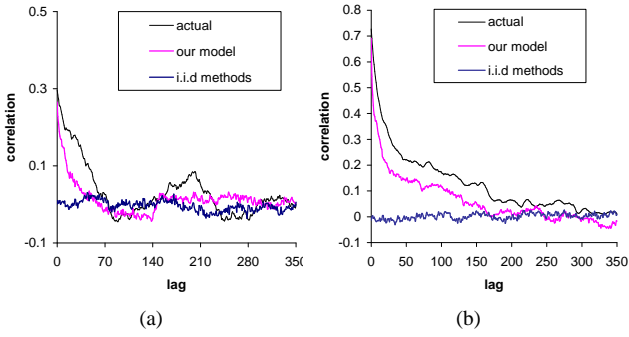


Fig. 10. (a) The correlation between $\{\phi_1^P(n)\}$ and $\{\phi^I(n)\}$ in Star Wars IV. (b) The correlation between $\{\phi_1^B(n)\}$ and $\{\phi^I(n)\}$ in Jurassic Park I.

dependency, and model the enhancement-layer traffic.

Due to its flexibility and high bandwidth utilization, layered video coding has become 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)) [27]. 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 layer 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.* [28] 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 IV sequence as examples. We only show the analysis of two-layer sequences for brevity since similar results hold for video streams with more than two layers.

A. Analysis of the Enhancement Layer⁴

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. 11 (a), we display the correlation

⁴We do not consider temporally scalable coded sequences, in which the base layer and the enhancement layer are approximately equivalent to extracting I/P-frames and B-frames out of a single-layer sequence, respectively [22].

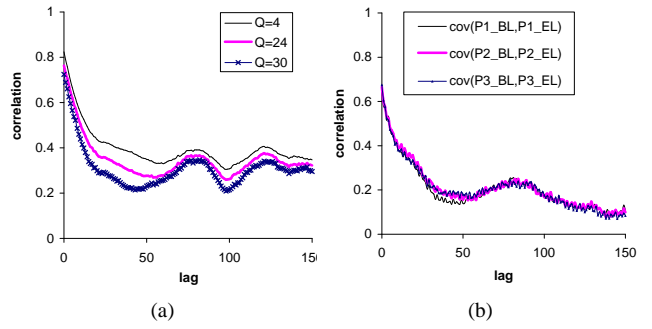


Fig. 11. (a) The correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ in The Silence of the Lambs coded at $Q = 4, 24, 30$. (b) The correlation between $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ in The Silence of the Lambs coded at $Q = 30$, for $i = 1, 2, 3$.

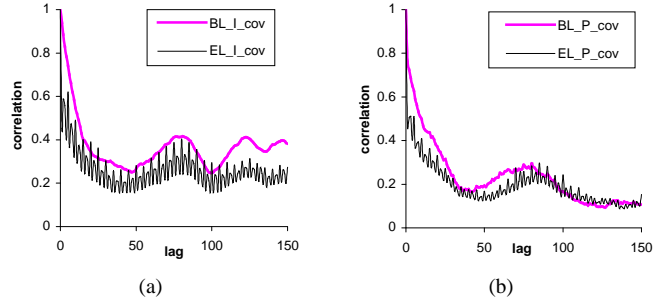


Fig. 12. (a) The ACF of $\{\varepsilon^I(n)\}$ and that of $\{\phi^I(n)\}$ in Star Wars IV. (b) The ACF of $\{\varepsilon_1^P(n)\}$ and that of $\{\phi_1^P(n)\}$ in The Silence of the Lambs.

between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ in The Silence of the Lambs coded at different Q . As observed from the figure, the correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ is stronger when the quantization step Q is smaller. However, the difference among these cross-correlation curves is not as obvious as that in intra-GOP correlation. We also observe that the cross-correlation is still strong even at large lags, which indicates that $\{\varepsilon^I(n)\}$ exhibits LRD properties and we should preserve these properties in the synthetic enhancement layer I-frame sizes.

In Fig 11 (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. 12 (a); and display the ACF of $\{\varepsilon_1^P(n)\}$ and that of $\{\phi_1^P(n)\}$ in Fig. 12 (b). The figure shows that although the ACF structure of $\{\varepsilon^I(n)\}$ has some oscillation, its trend closely follows that

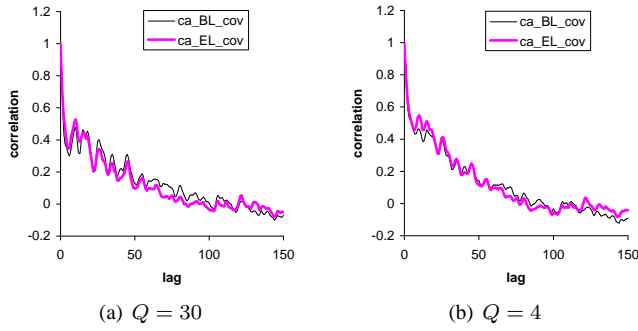


Fig. 13. The ACF of $\{A_3(\varepsilon)\}$ and $\{A_3(\phi)\}$ in The Silence of the Lambs coded at (a) $Q = 30$ and (b) $Q = 4$.

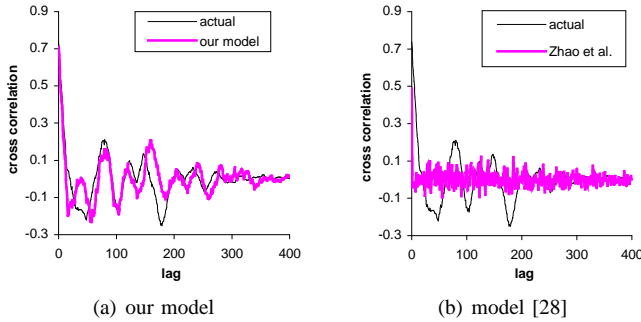


Fig. 14. The cross-correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ in The Silence of the Lambs and that in the synthetic traffic generated from (a) our model and (b) model [28].

of $\{\phi^I(n)\}$. One also observes from the figures that the ACF structures of processes $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ are similar to each other.

B. Modeling the Enhancement Layer I-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 its modeling [28]. In this section, we first describe how we model the enhancement layer I-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 (labeled as “ca_EL_cov” and “ca_BL_cov”, respectively) in Fig. 13.

As shown in Fig. 13, $\{A_j(\varepsilon)\}$ and $\{A_j(\phi)\}$ exhibit similar ACF structure. 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. In Fig. 14, 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

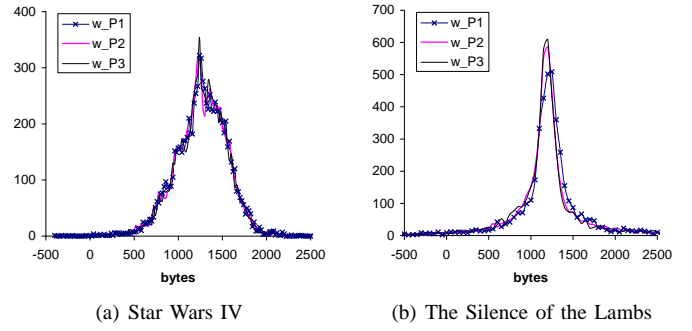


Fig. 15. Histograms of $\{w_1(n)\}$ in (a) Star Wars IV and (b) The Silence of the Lambs ($Q = 24$), with $i = 1, 2, 3$.

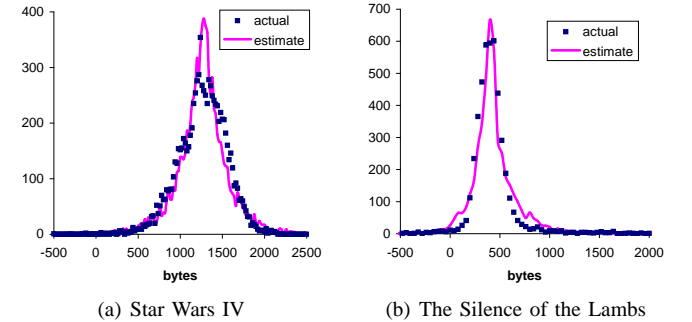


Fig. 16. Histograms of $\{w_1(n)\}$ and $\{\tilde{w}_1(n)\}$ for $\{\varepsilon_1^P(n)\}$ in (a) Star Wars IV and (b) The Silence of the Lambs ($Q = 30$).

Zhao’s model [28]. The figure shows that our model significantly outperforms Zhao’s model in preserving the cross-layer correlation.

C. Modeling P and B-Frame Sizes

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. 11. We use the linear model from Section IV-B to estimate the sizes of the i -th P and B-frames in the n -th GOP:

$$\varepsilon_i^P(n) = a\phi_i^P(n) + \tilde{w}_1(n), \quad (11)$$

$$\varepsilon_i^B(n) = a\phi_i^B(n) + \tilde{w}_2(n), \quad (12)$$

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 $\{\phi_i^B(n)\}$. We examine $\{w_1(n)\}$ and $\{w_2(n)\}$ and find they exhibit similar properties. We show two examples of $\{w_1(n)\}$ in Fig. 15.

As observed from Fig. 15, the histogram of $\{w_1(n)\}$ is asymmetric and decays fast on both sides. Therefore, we use two exponential distributions to estimate its PDF. We first left-shift $\{w_1(n)\}$ 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 $\exp(-\lambda_1)$ and the absolute value of the left side using another exponential distribution $\exp(-\lambda_2)$.

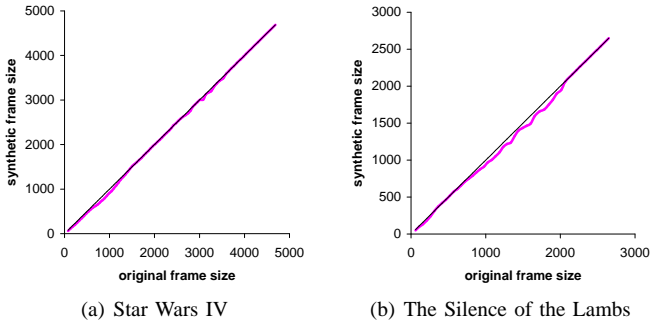


Fig. 17. QQ plots for the synthetic (a) single-layer *Star Wars IV* and (b) *The Silence of the Lambs* base-layer traffic.

Afterwards, we generate synthetic data $\{\tilde{w}_1(n)\}$ based on these two exponential distributions and right-shift the result by δ . As shown in Fig. 16, the histograms of $\{\tilde{w}_1(n)\}$ are close to those of the actual data in both *Star Wars IV* and *The Silence of the Lambs*. We generate $\{\tilde{w}_2(n)\}$ in the same way and find its histogram is also close to that of $\{w_2(n)\}$.

VI. MODEL ACCURACY STUDY

As we stated earlier, a good traffic model should capture the statistical properties of the original traffic and be able to accurately predict network performance. There are three popular studies to verify the accuracy of a video traffic model [26]: quantile-quantile (QQ) plots, the variance of traffic during various time intervals, and buffer overflow loss evaluation. While the first two measures visually evaluate how well the distribution of the synthetic traffic and that of the original one matches, the overflow loss simulation examines the effectiveness of a traffic model to capture the temporal burstiness of original traffic.

The QQ plot is a graphical technique to verify the distribution similarity between two test data sets. If the two data 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.

Different from the QQ plot, the variance of traffic during various time intervals shows whether the second-order moment of the synthetic traffic fits that of the original one. This second-order descriptor is used to capture burstiness properties of arrival processes [1]. This measure operates as follows. Assume that the length of a video sequence is l and there are m frames at a given time interval. We segment the one-dimensional data into a $m \times n$ matrix, where $n = l/m$. After summing all the data in each column, we obtain a sequence of length n and then calculate its variance. Thus, we can obtain a set of variances given a set of time intervals.

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 [26]. The drain rate is the number of bytes drained per second

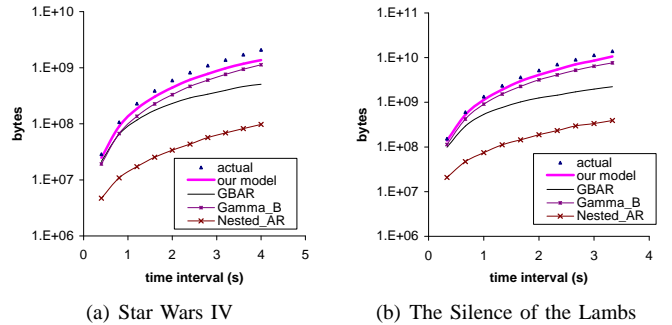


Fig. 18. Comparison of variance between synthetic and original traffic in (a) single-layer *Star Wars IV* and (b) *The Silence of the Lambs* base layer.

and is simulated as different multiples of the average traffic rate \bar{r} .

In the following two sections, we evaluate the accuracy of our model in both single-layer and multi-layer traffic using the above three measures. We should note that simulations with additional video sequences have demonstrated results similar to those shown throughout this paper.

A. Single-layer and the Base layer Traffic

We first show QQ plots of the synthetic single-layer *Star Wars IV* and the synthetic base layer of *The Silence of the Lambs* that are generated by our model in Fig. 17 (a) and (b), respectively. As shown in the figure, the generated frame sizes and the original traffic are almost identical.

In Fig. 18, we give a comparison between the variance of the original traffic and that of the synthetic traffic generated from different models at various time intervals. The figure shows that the second-order moment of our synthetic traffic is in a good agreement with that of the original one.

We also compare the accuracy of several models using a leaky-bucket simulation. To understand the performance differences between 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:

$$e = \frac{|p - p_{model}|}{p}. \quad (13)$$

In Table I, 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 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.

In Fig. 19, we show the relative error e of synthetic traffic generated from different models in H.264 *Starship Troopers* coded at $Q = 1, 31$, given $d = \bar{r}$. Since GOP-GBAR model [7] is specifically developed for MPEG traffic, we do not apply it to H.264 sequences. The figure shows that

TABLE I
RELATIVE DATA LOSS ERROR e IN Star Wars IV

Buffer capacity	Traffic type	Drain rate		
		$2\bar{r}$	$4\bar{r}$	$5\bar{r}$
20ms	Our Model	0.93%	0.61%	1.13%
	GOP-GBAR [7]	3.84%	2.16%	3.77%
	Nested AR [17]	5.81%	2.77%	8.46%
	Gamma_A [26]	5.20%	0.61%	2.57%
	Gamma_B [26]	4.89%	1.93%	2.05%
30ms	Our Model	0.25%	0.33%	0.95%
	GOP-GBAR [7]	4.94%	3.33%	5.68%
	Nested AR [17]	6.94%	4.14%	9.92%
	Gamma_A [26]	4.88%	1.10%	4.48%
	Gamma_B [26]	4.67%	2.17%	4.03%

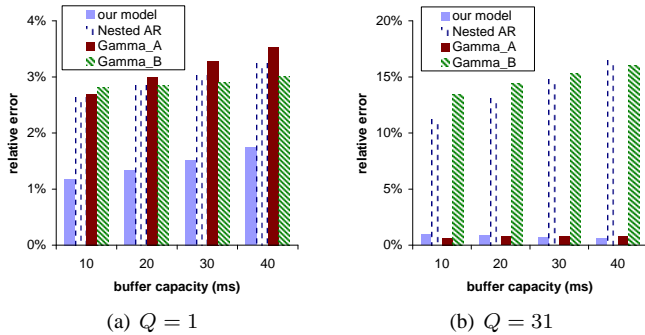


Fig. 19. Given $d = \bar{r}$, the error e of various synthetic traffic in H.264 Starship Troopers coded at (a) $Q = 1$ and (b) $Q = 31$.

our model outperforms the other three models in Starship Troopers coded with small Q and performs as good as model Gamma_A [26] with large Q (the relative error e of both models is less than 1% in Fig. 19 (b)).

B. Enhancement Layer Traffic

We evaluate the accuracy of the synthetic enhancement layer by using QQ plots and show two examples in Fig. 20, which displays two QQ plots for the synthetic The Silence of the Lambs and Star Wars IV 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 also compare the variance of the original traffic and that of the synthetic traffic in Fig. 21. Due to the computational complexity of model [28] in calculating long sequences, we only take the first 5000 frames of Star Wars IV and The Silence of the Lambs. As observed from the figure, our model preserves the second-order moment of the original traffic well.

We next examine the data loss ratio predicted by our synthetic traffic passed through a generic buffer as shown in the previous section. Recall that the model in [1] is only applicable to sequences with a CBR base layer and the one in [28] is suitable only for short sequences. Therefore, we are not able to show results using leaky-bucket simulations for these multi-layer models given the nature of our sample sequences. In Fig. 22 and Fig. 23, we show the overflow data loss ratio

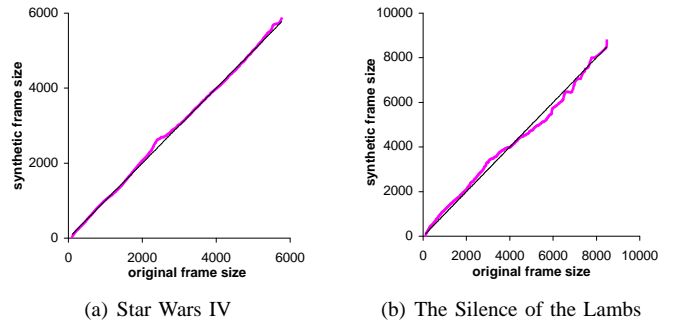


Fig. 20. QQ plots for the synthetic enhancement-layer traffic: (a) Star Wars IV and (b) The Silence of the Lambs.

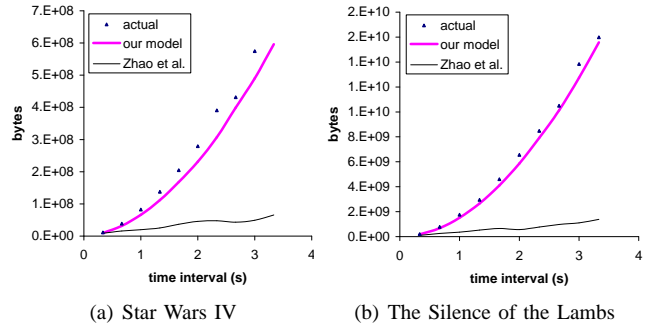


Fig. 21. Comparison of variance between the synthetic and original enhancement layer traffic in (a) Star Wars IV and (b) The Silence of the Lambs.

of the enhancement layers in both The Silence of the Lambs (54,000 frames) and Star Wars IV (108,000 frames) 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 the average traffic rate \bar{r} . The figure shows that the synthetic enhancement layer preserves the temporal information of the original traffic very well.

VII. CONCLUSION

In this paper, we presented a framework for modeling H.264 and MPEG-4 multi-layer full-length VBR video traffic. This work precisely captured the inter- and intra-GOP correlation in compressed VBR sequences, by incorporating wavelet-domain analysis into time-domain modeling. Whereas many previous traffic models are developed at slice-level or even block-level [26], our framework uses frame-size level, which allows us to examine the loss ratio for each type of frames and apply other methods to improve the video quality at the receiver. We also proposed novel methods to model cross-layer correlation in multi-layer sequences. In future work, we plan to apply our traffic model and optimize network delivery of VBR video to design layered peer-to-peer video systems.

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.

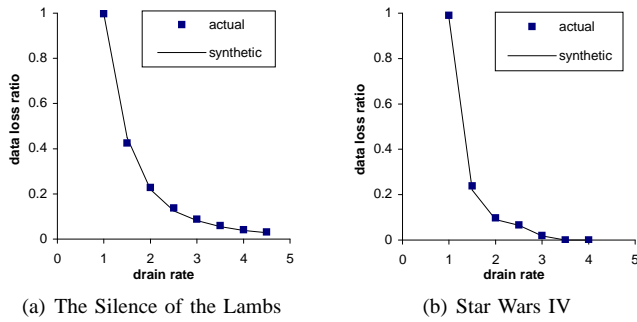


Fig. 22. Overflow data loss ratio of the original and synthetic enhancement layer traffic for $c = 10$ ms for (a) The Silence of the Lambs and (b) Star Wars IV.

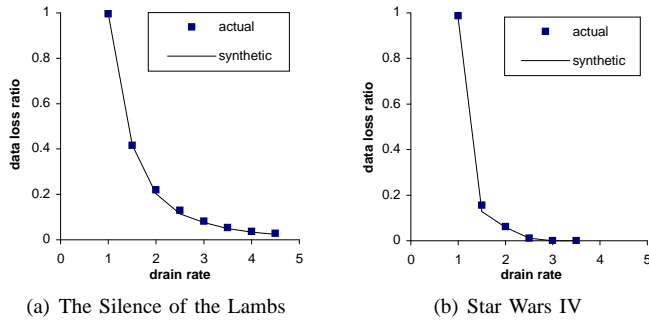


Fig. 23. Overflow data loss ratio of the original and synthetic enhancement layer traffic for $c = 30$ ms for (a) The Silence of the Lambs and (b) Star Wars IV.

[2] T. P.-C. Chen and T. Chen, "Markov Modulated Punctured Autoregressive Processes for Video Traffic and Wireless Channel Modeling," *Packet Video*, April 2002.

[3] A. L. Corte, A. Lombardo, S. Palazzo, and S. Zinna, "Modeling Activity in VBR Video Sources," *Signal Processing: Image Communication*, vol. 3, June 1991.

[4] M. Dai and D. Loguinov, "Analysis and Modeling of MPEG-4 and H.264 Multi-Layer Video Traffic (extended version)," *Texas A&M Technical Report*, July 2004.

[5] P. Embrechts, F. Lindskog, and A. McNeil, "Correlation and dependence in risk management: Properties and pitfalls," available at <http://www.ccfz.ch>, Aug. 1999.

[6] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation (Extended Version)," available at <http://www-tnk.ee.tu-berlin.de>.

[7] 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.

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

[9] 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.

[10] M. W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," *Proc. of ACM SIGCOMM*, Aug. 1994.

[11] D. P. Heyman, A. Tabatabai, and T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks," *IEEE Trans. on CSVT*, vol. 2, Mar. 1992.

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

[13] 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.

[14] Joint Video Team (JVT) of ISO/IEC MPEG, ITU-T VCEG (ISO/IEC JTC1/SC29/WG11, and ITU-T SG16 Q.6, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T rec. H.264/ISO/IEC 14 496-10 AVC)," 7th Meeting, Pattaya, Thailand, Mar. 2003.

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

[16] 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 Telecommunications*, vol. 12, April 2001.

[17] 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.

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

[19] 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.

[20] 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.

[21] MPEG-4: ISO/IEC JTC1/SC29/WG11, "ISO/IEC 14 496:2000-2: Information on technology/coding of audio/visual objects/Part 2: Visual," ISO/IEC, Genf, Switzerland, Dec. 2000.

[22] 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>.

[23] 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.

[24] 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.

[25] O. Rose, "Simple and Efficient Models for Variable Bit Rate Mpeg Video Traffic," *Performance Evaluation*, vol. 30, 1997.

[26] 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.

[27] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Prentice Hall, 2001.

[28] J.-A. Zhao, B. Li, and I. Ahmad, "Traffic Model For Layered Video: An Approach On Markovian Arrival Process," *Packet Video*, April 2003.