A Unified Traffic Model for MPEG-4 and H.264 Video Traces

Min Dai, Yueping Zhang, and Dmitri Loguinov

***Abstract*—This paper presents a frame-level hybrid framework for modeling MPEG-4 and H.264 multi-layer variable bit rate (VBR) video traffic. To accurately capture long-range dependent and short-range dependent properties of VBR sequences, 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 (Group of Pictures) *and* intra-GOP correlation in VBR video and build an enhancement-layer model using cross-layer correlation. 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 possesses lower complexity and has better performance than the previous methods in both single- and multi-layer sequences.**

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 [24].

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 traffic in the current Internet [2], [35]. In addition, traffic modeling research is falling behind the rapid advances in video techniques since very limited research has been done to model H.264 video sequences [5]. Therefore, the goal of this work is to better understand the statistical properties of various video sequences and to develop a unified model for single and multi-layer MPEG-4 and H.264 video traffic.

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 video. In regard to the first issue, several models have been proposed for the frame-size distribution, including the lognormal [10], Gamma [31], and various hybrid distributions (e.g., Gamma/Pareto [20] or Gamma/lognormal [29]).

Furthermore, 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 traces exhibit both long-range dependent (LRD) and short-range dependent (SRD) properties [12], [21]. 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 [12]. Thus, using either an LRD or SRD 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., [20], [21]).

Besides the complex autocorrelation properties, video traffic also exhibits *inter*- and *intra*-GOP[1] correlation due to the GOP-based coding structure of many popular standards. While the former is well characterized by the ACF of the sizes of I-frames of each GOP, the latter refers to the correlation between the sizes of P/B-frames and the I-frame size in the same GOP. Whereas most models try to capture the inter-GOP correlation, the intra-GOP correlation has been rarely addressed in related work even though it is an important characteristic useful in computing precise bounds on network packet loss [19].

In this paper, we develop a modeling framework that is able to capture the complex LRD/SRD structure of single/multi-layer video traffic, while addressing the issues of both inter/intra-GOP and cross-layer correlation. We model I-frame sizes in the wavelet domain using *estimated* wavelet coefficients, which are more mathematically tractable than the *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 that of prior methods.

The specifics of the sample sequences used in this paper are shown in Table I. As see in the table, single-layer traffic includes both MPEG-4 and H.264 video sequences and multi-layer traffic includes Fine Granular Scalability (FGS) [25], temporal and spatial scalability [32], and Multiple Description Coded (MDC) [33] sequences. Most non-MDC sequences are one-hour long with GOP structure $(12, 2)$, except that `Star Wars IV` [27] is half an hour long with GOP structure

A shorter version of this paper appeared in IEEE INFOCOM 2005.

Min Dai is with Qualcomm Inc., San Diego, CA 92121 USA (e-mail: mdai@qualcomm.com)

Yueping Zhang is with NEC Laboratories America, Inc., Princeton, NJ 08540 USA (e-mail: yueping@nec-labs.com).

Dmitri Loguinov is with Texas A&M University, College Station, TX 77843 USA (e-mail: dmitri@cs.tamu.edu).

[1] A GOP includes one I-frame and several P- and B-frames.

TABLE I
CHARACTERISTICS OF SAMPLE SEQUENCES

Sequence name	Scalability	Rate (fps)	Standard
Starship Troopers [27]	None	25	H.264
Star Wars IV [27]	None	30	H.264
Star Wars IV-A [8]	None	25	MPEG-4
Jurassic Park I [8]	None	25	MPEG-4
Starship Troopers [8]	None	25	MPEG-4
Star Trek - First Contact [8]	None	25	MPEG-4
The Silence of the Lambs-A [8]	None	25	MPEG-4
Bridge [27]	MDC	25	MPEG-4
Star Wars IV-B [27]	FGS	30	MPEG-4
Clip CIF [27]	FGS	30	MPEG-4
Citizen Kane [27]	Temporal	30	MPEG-4
The Silence of the Lambs-B [27]	Spatial	30	MPEG-4

$(16, 1)$ and `Clip CIF` is 30 seconds long with GOP structure $(12, 0)$. Note that GOP structure (N, M) means that there are N frames in a GOP and M B-frames between every two non-B frames, e.g., GOP $(12, 2)$ stands for *IBBPBBPBBPBB* and GOP $(12, 0)$ refers to *IPPPPPPPPPPP*. The length of the sample MDC-coded sequence is varying since temporal subsampling is applied. More details about MDC coding is given in Section VI. Further note that sequences coded from the same video but with different quantization steps are not repetitively listed. For example, we only show `Jurassic Park I` once in Table I, while this paper uses several single-layer `Jurassic Park I` sequences that are coded with different quantization steps.

The outline of this paper is as follows. Section II overviews the related work on traffic modeling. In Section III, we provide the technical background on wavelet analysis and statistical properties of wavelet coefficients. In Section IV, we show how to model single-layer and base-layer video traffic by generating synthetic I-traces in the wavelet domain and P/B traces with a linear I-trace model. Sections V and VI analyze and model the cross-layer correlation in layer-coded and MDC-coded video traces, respectively. In Section VII, we evaluate the accuracy of our model using both single-layer and multi-layer video traffic. Section VIII concludes the paper.

II. RELATED WORK

The topic of VBR traffic modeling has been extensively studied 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 auto-regressive (AR) models, since they are classical approaches in the area of traffic modeling. After the first AR model was applied to video traffic in 1988 [22], AR processes and their variations remain highly popular in this area of research [20]. For example, Corte *et al*. [4] 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*. [10] model the deviation of I-frame sizes from their mean in each scene using an AR(2) process. Building upon Krunz' work [10], Liu *et al*. [20] 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. In [14], Heyman *et al.* propose a discrete autoregressive (DAR) model to model videoconferencing data. Since the DAR model is not effective for single-source video traffic, Heyman [15] later develops a GBAR model, which has Gamma-distributed marginal statistics and a geometric auto-correlation function. By considering the GOP cyclic structure of video traffic, Frey *et al*. [9] extend the GBAR model in [15] 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 [18], AR process [3]). Rose [30] uses nested Markov chains to model GOP sizes. Since synthetic data are generated at the GOP level, this model actually coarsens the time scale and thus is not suitable for high-speed networks. Ramamurthy and Sengupta [26] propose a hierarchical video traffic model, which uses a Markov chain to capture scene change and two AR processes to match the autocorrelation function in short and long range, respectively. Extending the above work, Chen *et al*. [3] 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*. [31] 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 [21].

We classify self-similar processes and fractal models as the third category. Garrett *et al*. [12] 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 [12], Huang *et al*. [16] present a self-similar fractal traffic model; however, this model does not capture the multi-timescale variations in video traffic [10].

Other approaches include the $M/G/\infty$ process [11] and Transform-Expand-Sample (TES) based models [23]. The former creates SRD traffic [20] and the latter has high computational complexity and often requires special software (e.g., *TEStool*) to generate synthetic sequences. Different from the above time-domain methods, several wavelet models [21], [28] recently emerged due to their ability to accurately capture both LRD and SRD properties of video traffic [21].

B. Multi-Layer Models

Most traffic modeling studies focus on single-layer video traffic and much less work has been done to model multi-layer sequences. Ismail *et al*. [17] use a TES-based method to model VBR MPEG video that has two levels of priority, which might be considered the first multi-layer traffic model. Later, Chandra *et al*. [2] use a finite-state Markov chain to model one- and two-layer scalable video traffic. 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 [2] 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 [2], Zhao *et al*. [35] build a K-state Markov chain based on frame-size clusters. The clustering feature in [35] is the cross-layer 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 in [35] is high and only suitable for video sequences with few scene changes. Furthermore, even though methods [1] exist for choosing the optimal number of states in a Markov chain, [2] and [35] do not examine their performance and instead select the necessary parameters heuristically.

III. Wavelet Analysis

The wavelet transform has become a powerful technique in the area of traffic modeling [21]. Wavelet analysis is typically based on a decomposition of the signal using a 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.

In order to better understand the structure of wavelet coefficients, we investigate statistical properties of both detailed and approximation coefficients in this section.

A. Detailed Coefficients

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. We also assume that $j = J$ is the coarsest scale and $j = 0$ is the original signal.

As we show next, one big advantage of the wavelet transform is its ability to provide short-range-dependent detailed coefficients for long-range-dependent processes.

Theorem 1: The detailed coefficients of an LRD process possess SRD properties.

Proof: Assume that $\{X(t)\}$ is an LRD process with spectral density function $\Gamma(\nu) \sim c|\nu|^{-\alpha}$, where $c > 0$ and $0 < \alpha < 1$.

The covariance function of any two detailed coefficients D_j^k and $D_{j'}^{k'}$ of $\{X(t)\}$ is:

$$cov[D_j^k, D_{j'}^{k'}] = E[D_j^k D_{j'}^{k'}] - E[D_j^k]E[D_{j'}^{k'}], \qquad (1)$$

where j, j' are the decomposition levels and k, k' show the sample locations. Note that the high-pass nature of wavelet function leads to $E[D_j^k] = E[D_{j'}^{k'}] = 0$.

Furthermore, Wornell [34] shows that $E[D_j^k D_{j'}^{k'}]$ decreases hyperbolically with the distance between the two wavelet coefficients of $\{X(t)\}$ as $|2^j k - 2^{j'} k'|^{2H-2N}$, where Hurst parameter $H \in (0.5, 1)$, N is the *vanishing moment*[2] of mother wavelet ψ_0, and term $|2^j k - 2^{j'} k'|$ is the shortest distance between two detailed coefficients. Note that $|2^j k - 2^{j'} k'|$ is always greater than 1.

Thus, we write (1) as:

$$cov[D_j^k, D_{j'}^{k'}] = E[D_j^k D_{j'}^{k'}] = \tau^{-(2N-1-\alpha)}, \qquad (2)$$

where $\tau = |2^j k - 2^{j'} k'|$ and $\alpha = 2H - 1$ is the parameter in the spectral density function of $\{X(t)\}$. Since $\alpha \in (0, 1)$ and $N \geq 1$, we have $(2N - 1 - \alpha) > 0$.

Due to the fact that $\tau > 1$ and $(2N - 1 - \alpha) > 0$, $\tau^{-(2N-1-\alpha)}$ converges to zero and $\sum_{\tau=-\infty}^{\infty} \tau^{-(2N-1-\alpha)}$ converges to a constant. Recall that a process is short-range dependent if the sum of its autocorrelation[3] function $\gamma(k)$ is summable, i.e., $\sum_{k=-\infty}^{\infty} \gamma(k)$ is finite. Thus, the detailed coefficients are short-range dependent. ■

B. Approximation Coefficients

After analyzing the detailed coefficients, we next examine the autocorrelation function and the distribution of the approximation coefficients. We use the Haar wavelet transform as a typical example since it is often chosen for its simplicity and good performance [21], [28]. Recall that the Haar scaling and wavelet functions are, respectively:

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

Then the Haar approximation coefficients A_j^k are obtained via [28]:

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

where j is the decomposition level and k is the index of a process. Since video traffic possesses strong self-similarity [24], we have the following result.

Theorem 2: The Haar approximation coefficients of a self-similar process preserve the correlation structure of the original signal.

Proof: From (3), we observe that the process of Haar approximation coefficients $\{A_j\}$ is generated by calculating the mean values of each two neighboring samples in a process

[2]A wavelet has vanishing moments of order m if $\int_{-\infty}^{\infty} t^p \psi(t) dt = 0$, where $p = 0, \cdots, m-1$ [13].

[3]In the traffic modeling literature, the normalized auto-covariance function is often used instead of the autocorrelation function [20].

$\{X(t)\}$. Recall that the *aggregated process* $\{X^{(m)}\}$ of a self-similar process $\{X(t)\}$ at aggregation level m is:

$$X^{(m)}(i) = \frac{1}{m} \sum_{t=m(i-1)+1}^{mi} X(t). \tag{4}$$

Comparing (4) to (3), one can observe that $\{A_j\}$ is a weighted aggregated process $\{X^{(2^j)}\}$ of the original signal $\{X(t)\}$. Since that a self-similar process and its aggregated process have the same autocorrelation structure [24], the theorem follows. ∎

We apply the wavelet transform to the sizes of I-frames in sample sequences and examine the statistical properties of the detailed and approximation coefficients. In Fig. 1 (a), we show the ACF of processes $\{A_3\}$ and $\{D_3\}$ computed based on the I-frame sizes in single-layer `Star Wars IV-A` 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 $\{D_j\}$ as zero-mean $i.i.d.$ Gaussian variables [21]. Fig. 1 (a) also shows that the approximation coefficients $\{A_j\}$ 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.

In Fig. 1 (b), 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-A`, as a typical example. Fig. 1 (b) shows that the *symmetric* Gaussian distribution does not describe the heavy tail of the actual PDF of $\{A_3\}$, even though it is a popular distribution model for the approximation coefficients based on the *Central Limit Theorem* [21]. Next, we examine the relationship between I-frame size $\{A_0\}$ and its approximation coefficients $\{A_j, j > 0\}$ with the help of the following theorem.

Theorem 3: Given that the I-frame sizes follow a Gamma distribution, the approximation coefficients $A_j^k, j \geq 1$ follow a linear combination of Gamma distributions.

Proof: Note that $\{A_j\}$ is a random process with $A_j = (A_j^1, A_j^2, \cdots, A_j^k, \cdots)$ and A_j^k is a random variable.

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

$$A_1^k = 2^{-1/2}(X_1^k + X_2^k), \tag{5}$$

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

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

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

Next, to capture 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 [9]:

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

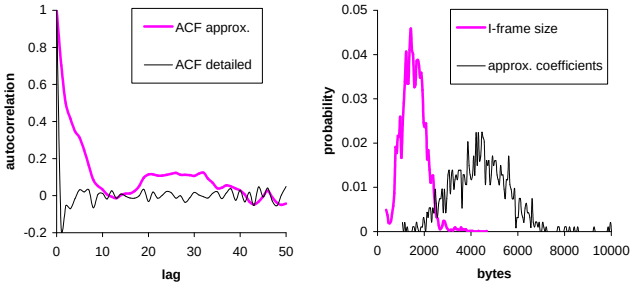
where Y_{11}, Y_{12}, and Y_{22} are independent of each other and follow standard Gamma distributions with parameters α_{11}, α_{12}, and α_{22}, respectively. Then the correlation between X_1^k and X_2^k becomes:

$$\text{cov}(X_1^k, X_2^k) = \lambda_1\lambda_2 \text{var}(Y_{12}) = \lambda_1\lambda_2\alpha_{22}. \tag{8}$$

Combining (5) and (8), we rewrite A_1^k as:

$$A_1^k = 2^{-1/2}\left(\lambda_1 Y_{11} + (\lambda_1 + \lambda_2)Y_{12} + \lambda_2 Y_{22}\right). \tag{9}$$

As can be observed from (9), A_1^k is a linear combination of independent standard Gamma distributions, which leads to the statement of the theorem. ∎



(a) (b)

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

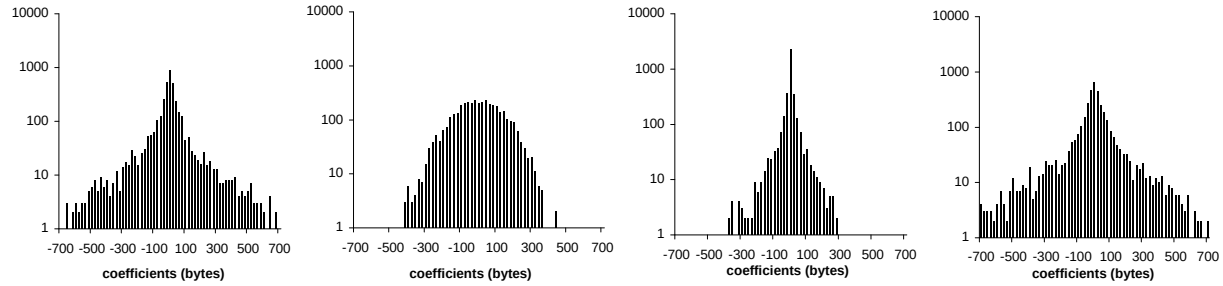
Fig. 1 (b) shows that the distribution of $\{A_j\}$ has a similar Gamma shape as that of I-frame sizes, but with different parameters. Extensive experimental results also demonstrate that a single Gamma distribution is accurate enough to describe the actual histogram of $\{A_j\}$. In the next section, we use this information to efficiently estimate the approximation coefficients.

IV. Modeling Single/Base-Layer

In this section, we discuss the issue of modeling the single-layer traffic and the base-layer of layer-coded traces, since the latter can be considered as the former from video coding perspective. We generate synthetic I-frame sizes in the wavelet domain and then model P/B-frame sizes in the time domain based on the intra-GOP correlation.

A. Modeling I-Frame Sizes

Wavelet-based algorithms have an advantage over the time-domain methods in capturing the LRD and SRD properties of video [21], [28]. Furthermore, wavelet methods do not need to specifically model scene-change lengths since wavelets are good at detecting discontinuities in video traffic, which are most often generated by scene changes. Due to these characteristics of wavelet transform, we model the I-frame sizes in the wavelet domain using the estimated approximation



(a) actual (b) Gaussian (c) GGD (d) mixture-Laplacian

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

and detailed coefficients, which are represented by $\{D_j\}$ and $\{A_j\}$, respectively.

Previous wavelet-based traffic modeling methods often model $\{D_j\}$ as zero-mean $i.i.d.$ Gaussian variables [21], without thorough investigation to $\{D_j\}$ś actual distribution. To provide some insight into the structure of detailed coefficients, we compare the histogram of the *actual* coefficients $\{D_1\}$ in `Star Wars IV-A` 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 over-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 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:

$$f(x) = p\frac{\lambda_0}{2}e^{-\lambda_0|x|} + (1-p)\frac{\lambda_1}{2}e^{-\lambda_1|x|}, \qquad (10)$$

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 $\{A_j\}$. Current methods generate the coarsest approximation coefficients (i.e., $\{A_J\}$) either as independent Gaussian [21] or Beta random variables [28]. However, as mentioned in Section III-B, 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 model the coarsest approximation coefficients $\{A_J\}$ as *dependent* random variables with marginal Gamma distributions[4] according to Theorem 3. The procedure is as follows:

- Generate N dependent Gaussian variables x_i using a $k\times k$ correlation matrix, where N is the length of $\{A_J\}$ and the number of preserved correlation lags k is chosen to be a reasonable value (e.g., the average scene length[5]). The correlation matrix is obtained from the actual coefficients $\{A_J\}$.
- Apply the Gaussian CDF $F_G(x)$ directly to x_i to convert them into a uniformly distributed set of variables $F_G(x_i)$.
- Pass the result from the last step through the inverse Gamma CDF to generate (still dependent) Gamma random variables [6], [7].

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 [9] and Gamma_A model [31] 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.

B. Intra-GOP Correlation Analysis

We next provide a detailed analysis of intra-GOP correlation for various video sequences and model P/B-frame sizes in the time domain based on intra-GOP correlation. Before further discussion, we define I and P/B-traces as follows. Assuming that $n \geq 1$ represents the GOP number,

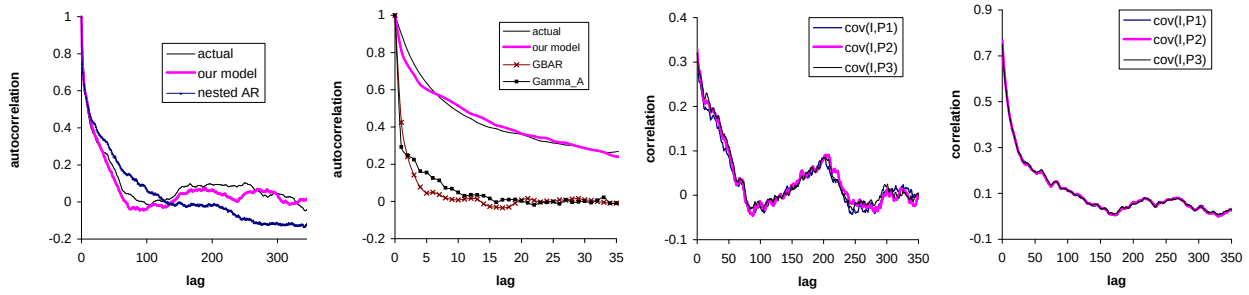
- $\phi^I(n)$ is the I-frame size of the n-th GOP;
- $\phi_i^P(n)$ is the size of the i-th P-frame in GOP n;
- $\phi_i^B(n)$ is the size of the i-th B-frame in GOP n.

For example, $\phi_3^P(10)$ represents the size of the third P-frame in the 10-th GOP.

Although previous work model the P/B-frame sizes as $i.i.d.$ random variables [10], [20], [31], Lombardo *et al.* [18] noticed that there is a strong correlation 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

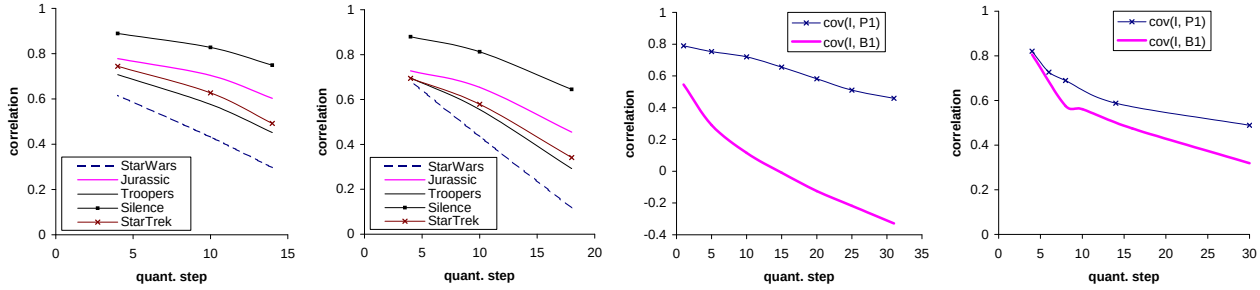
[4]More details about how to construct dependent random variables are available in [6], [7].

[5]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) (b) (c) (d)

Fig. 3. The ACF of the actual I-frame sizes and that of the synthetic traffic in (a) long range and (b) short range; The correlation between $\{\phi_i^P(n)\}$ and $\{\phi^I(n)\}$ in (c) `Star Wars IV-A` and (d) `Jurassic Park I`, for $i = 1, 2, 3$.



(a) (b) (c) (d)

Fig. 4. (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 $Q = 4, 10, 18$; The correlation between $\{\phi^I(n)\}$ and $\{\phi_1^P(n)\}$ and that between $\{\phi^I(n)\}$ and $\{\phi_1^B(n)\}$ in (c) H.264 `Starship Troopers` and (d) the base layer of the spatially scalable `The Silence of the Lambs-B` coded at different Q.

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 investigate the correlation between $\{\phi^I(n)\}$ and $\{\phi_i^P(n)\}$ and that between $\{\phi^I(n)\}$ and $\{\phi_i^B(n)\}$ for different i in various sequences. Fig. 3(c) shows the intra-GOP correlation in single-layer `Star Wars IV-A`, which is coded with quantization step $Q = 10, 14, 18$ for I/P/B frames, respectively. Fig. 3(d) shows the same correlation in `Jurassic Park I` that is coded at $Q = 4$ for all frames[6]. 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 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. 4(a)-(b). We also show the same correlation in H.264 coded `Starship Troopers` [27] in Fig. 4(c) and in the base layer of the spatially scalable `The Silence of the Lambs-B` in Fig. 4(d).

As observed from Fig. 4, the intra-GOP correlation decreases as the quantization step increases. This is due to the fact that sequences coded with smaller Q share more source information among the different frames in one GOP and thus have stronger intra-GOP correlation than sequences coded with larger Q. This observation is very useful for users to decide whether to preserve intra-GOP correlation at the expense of an increase in model complexity.

[6]If a sequence is coded with the *same* quantization step c for all frames, we say this sequence is coded at $Q = c$. Otherwise, we describe the quantization step for each type of frames in this sequence.

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:

$$\phi_i^P(n) = a\tilde{\phi}^I(n) + \tilde{v}(n), \tag{11}$$

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 $\tilde{\phi}^I(n)$.

Theorem 4: To capture the intra-GOP correlation, the value of coefficient a in (11) must be equal to:

$$a = \frac{r(0)\sigma_P}{\sigma_I}, \tag{12}$$

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:

$$E[\tilde{\phi}^I(n-k)] = E[\tilde{\phi}^I(n)] = 0. \tag{13}$$

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

$$C(k) = E[(\phi_i^P(n) - E[\phi_i^P])(\tilde{\phi}^I(n-k) - E[\tilde{\phi}^I])]. \quad (14)$$

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:

$$\begin{aligned} C(k) &= E[(a\tilde{\phi}^I(n) + v(n) - E[\phi_i^P])\tilde{\phi}^I(n-k)] \\ &= aE[\tilde{\phi}^I(n)\tilde{\phi}^I(n-k)] \end{aligned} \quad (15)$$

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

$$r(0) = \frac{C(0)}{\sigma_P \sigma_{\tilde{I}}} = \frac{aE[\tilde{\phi}^I(n)^2]}{\sigma_P \sigma_{\tilde{I}}}, \quad (16)$$

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:

$$\frac{a \cdot \sigma_I}{\sigma_P} = r(0), \quad (17)$$

which leads to (12). ■

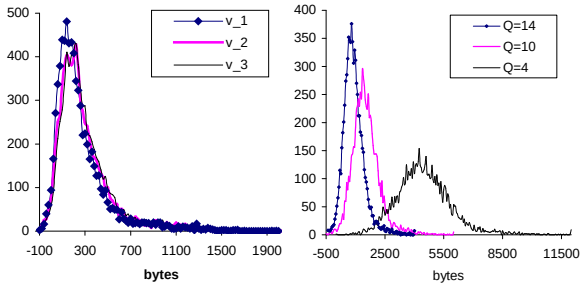
To understand how to generate $\{\tilde{v}(n)\}$, we next examine the *actual* residual process $v(n) = \phi_i^P(n) - a\tilde{\phi}^I(n)$ for each i and for video sequences coded at various Q. Fig. 5(a) shows the histograms of $\{v(n)\}$ for P-traces with different i in single-layer `Star Wars IV-A` and Fig. 5(b) shows the histograms of $\{v(n)\}$ for sequences coded at different Q. The figures show that the residual process $\{v(n)\}$ does not change much as a function of i but its histogram becomes more bell-shaped 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 $\{v(n)\}$.

From Fig. 4(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-A` coded at $Q = 18$) or as large as 0.9 (e.g., in `The Silence of the Lambs-A` 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 (11) when the correlation is large. The linear model has the following form:

$$\phi_i^B(n) = a\tilde{\phi}^I(n) + \tilde{v}_B(n), \quad (18)$$

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., [20], [31]) in Fig. 6(a)-(b). 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.



(a) (b)

Fig. 5. (a) Histograms of $\{v(n)\}$ for $\{\phi_i^P(n)\}$ with $i = 1, 2, 3$ in `Star Wars IV-A` coded at $Q = 14$; (b) Histograms of $\{v(n)\}$ for $\{\phi_1^P(n)\}$ in `Jurassic Park I` coded at $Q = 4, 10, 14$.

D. Further Discussion and Algorithm Summary

As shown in Fig. 4, the intra-GOP correlation is small in video sequences coded with a large quantization step. Furthermore, the intra-GOP correlation decreases as the GOP size increases, since the P/B-frames are far away from the I-frame in the same GOP due to the large GOP size. Under these circumstances, we can model $\{\phi_i^P(n)\}$ or $\{\phi_i^B(n)\}$ using the correlation between them and their reference frame due to the fact that P/B-frames are predictively coded.

This discussion benefits the modeling of temporally scalable traffic. Note that in temporally scalable video, 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 [27]. In other words, we model the enhancement layer of a temporally scalable-coded sequence as modeling the B-frames of a single-layer video traffic.

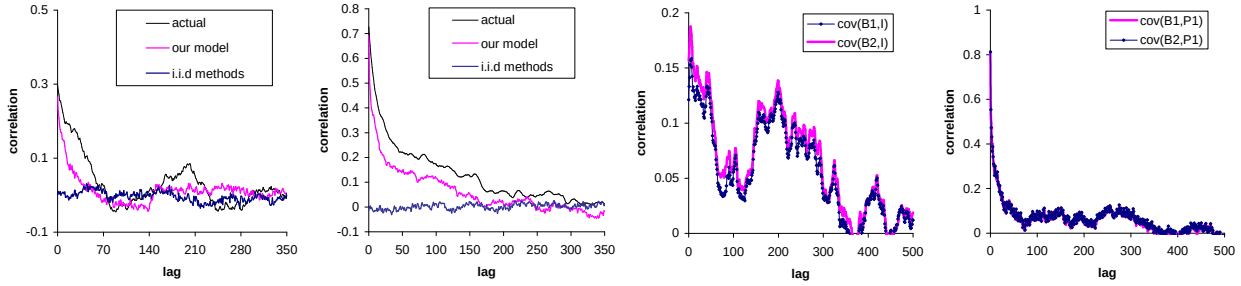
To better understand the correlation between the neighboring frames, we examine the correlation between the I-trace and P/B-traces and that between two neighboring P/B-traces in various sequences. In Fig. 6(c)-(d), we show the correlation between $\{\phi^I(n)\}$ and $\{\phi_i^B(n)\}$ and that between $\{\phi_1^P(n)\}$ and $\{\phi_i^B(n)\}$, for $i = 1, 2$, in temporally scalable `Citizen Kane` coded with quantization step $Q = 30$.

As Fig. 6(c) shows, the correlation between the base layer I-frames and the enhancement layer is not large enough to apply the linear model (18) in this sequence. However, Fig. 6(d) shows that the enhancement layer and its neighboring base layer P-frames are highly correlated. Therefore, we rewrite linear model (18) to:

$$\phi_i^B(n) = a\tilde{\phi}^P(n) + \tilde{v}_B(n), \quad (19)$$

where parameter $a = r(0)\sigma_B/\sigma_P$, $r(0)$ is the lag-0 correlation between the neighboring P and B-frame sizes, and σ_B, σ_P are the standard deviations of these two P/B-traces.

Before we finish this section, we summarize the procedures of our algorithm in Fig. 7 and discuss its complexity. Assume there is a video trace of length N, which includes M I-frames and $N - M$ P/B-frames. The required operations for I-frame size modeling is $O(M)$ since the computational complexity of DWT is in the order of signal length [21]. Note that P/B-traces are generated in a batch, which has computational cost of $O(N-M)$. Therefore, the computational complexity of our algorithm to generate a video trace of length N is $O(N)$.



(a) (b) (c) (d)

Fig. 6. (a) The correlation between $\{\phi_1^P(n)\}$ and $\{\phi^I(n)\}$ in `Star Wars IV-A`; (b) The correlation between $\{\phi_1^B(n)\}$ and $\{\phi^I(n)\}$ in `Jurassic Park I`; (c) The correlation between $\{\phi_i^B(n)\}$ and $\{\phi^I(n)\}$ and (d) that between $\{\phi_i^B(n)\}$ and $\{\phi_1^P(n)\}$ in temporally scalable-coded `Citizen Kane`, for $i = 1, 2$.

1) Generate the I-trace:
- Perform J levels of wavelet decomposition on the original I-trace
- For $i = 1$ to J do:
  - Estimate mixture-laplacian distribution parameters from original detailed coefficients;
  - Generate synthetic detailed coefficients using the estimated parameters.
- At level J:
  - Estimate Gamma distribution parameters from the original approximation coefficients;
  - Use copula to generate correlated synthetic approximation coefficients.

2) Generate P-traces:
- Estimate parameters of the generalized Gamma distribution from the original residual process;
- Generate synthetic P-traces using (11) based on synthetic I-trace.

3) Generate B-traces: repeat step 2) using B frames.

Fig. 7. Summary of the proposed algorithm.

V. Modeling Layer-Coded (Scalable) Traffic

In this section, we provide brief background knowledge of multi-layer video, investigate methods to capture cross-layer dependency, and model the enhancement-layer traffic.

Due to its flexibility and high bandwidth utilization, scalability [25], [32] is common in video applications. Layered coding can be further classified as coarse-granular (e.g., spatial scalability) or fine-granular (e.g., FGS) [32]. 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.* [35] also indicate that there exists a cross-layer 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-layer correlation between the enhancement layer and the base layer using spatially scalable `The Silence of the Lambs-B`, FGS-coded `Star Wars IV-B`, and three-layer FGS-coded `Clip CIF` as examples. We only show the analysis of these sequences for brevity and note that similar results hold for video streams with more layers.

A. Analysis of the Enhancement Layer

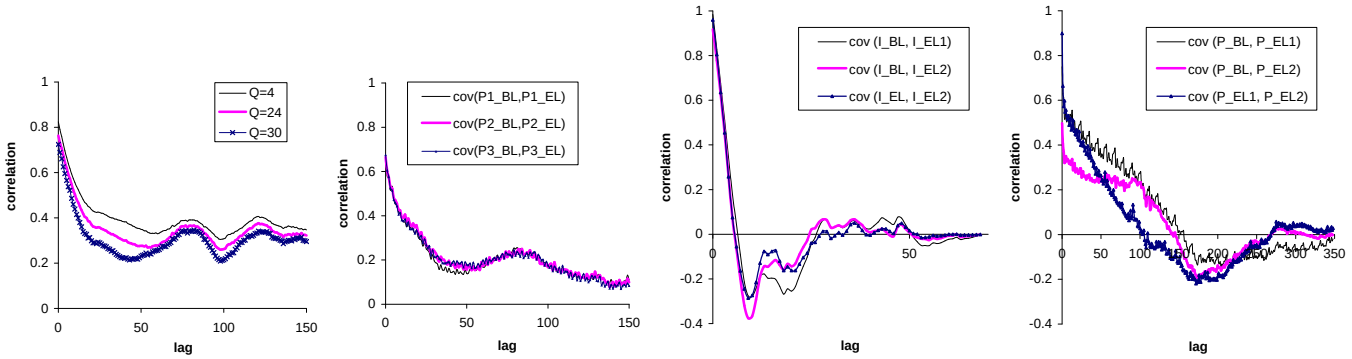
For discussion convenience, we define the enhancement layer frame sizes as follows. Similar to the definition in the base layer, we define $\varepsilon^I(n)$ to be the I-frame size of the n-th GOP, $\varepsilon_i^P(n)$ to be the size of the i-th P-frame in GOP n, and $\varepsilon_i^B(n)$ to be the size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

In Fig. 8(a), we display the correlation between the enhancement layer $\{\varepsilon^I(n)\}$ and the base layer $\{\phi^I(n)\}$ 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-layer correlation curves is not as obvious as that in intra-GOP correlation. We also observe that cross-layer 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.

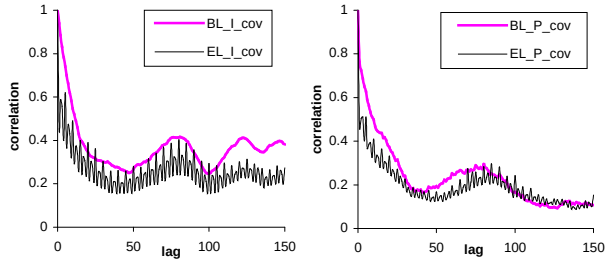
In Fig 8(b), we show the cross-layer correlation between processes $\{\varepsilon_i^P(n)\}$ and $\{\phi_i^P(n)\}$ for $i = 1, 2, 3$. The figure demonstrates that the correlation between the enhancement layer and the base layer is quite strong, and the correlation structures between each $\{\varepsilon_i^P(n)\}$ 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 $\{\phi_i^P(n)\}$.

We further evaluate the cross-layer correlation between the base layer and different enhancement layers as well as that between neighboring enhancement layers, using a three-layer FGS coded sequence `Clip CIF`. In Fig 8(c), we demonstrate that cross-layer correlation is strong between $\{\phi^I(n)\}$ and $\{\varepsilon_i^I(n)\}$, for $i = 1, 2$. Fig 8(d) shows that while P-frame cross-correlation between $\{\varepsilon_2^P(n)\}$ and $\{\phi^P(n)\}$ is strong, it is somewhat smaller than that between $\{\varepsilon_1^P(n)\}$ and $\{\phi^P(n)\}$, which is to be expected. In both figures, the cross correlation



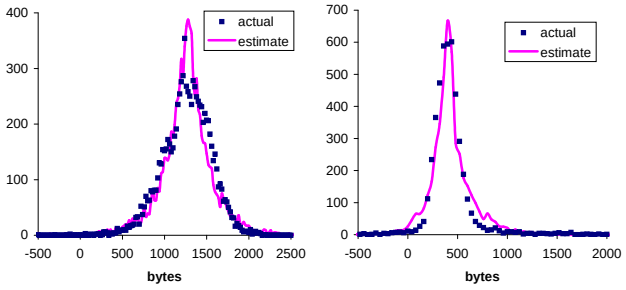
(a) (b) (c) (d)

Fig. 8. (a) The correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ in `The Silence of the Lambs-B` 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-B` coded at $Q = 30$, for $i = 1, 2, 3$; (c) The correlation between $\{\varepsilon_i^I(n)\}$ and $\{\phi_I(n)\}$ in `Clip CIF`, where $i = 1, 2$; (d) The correlation between $\{\varepsilon_i^P(n)\}$ and $\{\phi^P(n)\}$ in `Clip CIF`, where $i = 1, 2$.



(a) (b)

Fig. 9. (a) The ACF of $\{\varepsilon^I(n)\}$ and that of $\{\phi^I(n)\}$ in `Star Wars IV-B`; (b) The ACF of $\{\varepsilon_1^P(n)\}$ and that of $\{\phi_1^P(n)\}$ in `The Silence of the Lambs-B`.



(a) Star Wars IV-B (b) The Silence of the Lambs

Fig. 11. Histograms of $\{w_1(n)\}$ and $\{\tilde{w}_1(n)\}$ for $\{\varepsilon_1^P(n)\}$ in (a) `Star Wars IV-B` and (b) `The Silence of the Lambs-B` coded at $Q = 30$.

between the base layer and the first enhancement layer is similar to that between two enhancement layers.

Aside from cross-layer correlation, we also examine the autocorrelation 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. 9(a); and display the ACF of $\{\varepsilon_1^P(n)\}$ and that of $\{\phi_1^P(n)\}$ in Fig. 9(b). The figure shows that although the ACF structure of $\{\varepsilon^I(n)\}$ has some oscillation, its trend closely follows that 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 [2], nor explicitly addressed the issue of its modeling [35]. 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 (as shown in Fig. 9(a)), 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. 10(a)-(b).

As shown in Fig. 10(a)-(b), $\{A_j(\varepsilon)\}$ 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. 10(c)-(d), we compare the actual cross-layer 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 [35]. The figure demonstrates 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-layer 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. 8(a)-(b). We use the linear model from Section IV-C to estimate the sizes of the i-th P and B-frames in the n-th GOP:

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

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

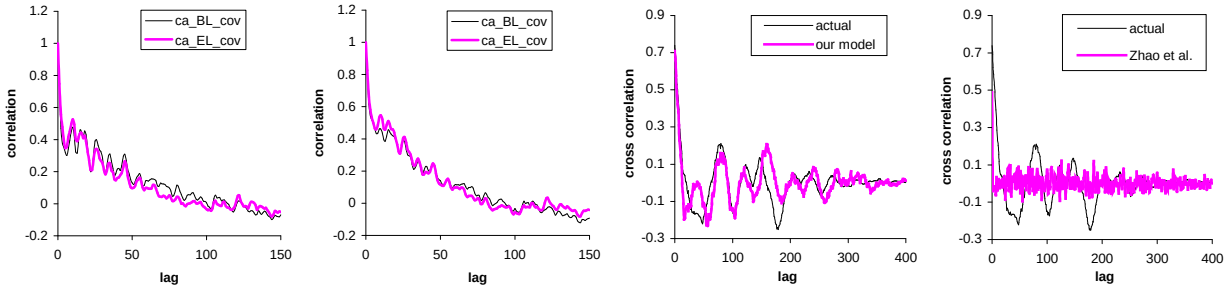


Fig. 10. The ACF of $\{A_3(\varepsilon)\}$ and $\{A_3(\phi)\}$ in `The Silence of the Lambs-B` coded at (a) $Q = 30$ and (b) $Q = 4$; The cross-layer correlation between $\{\varepsilon^I(n)\}$ and $\{\phi^I(n)\}$ in `The Silence of the Lambs` and that in the synthetic traffic generated from (c) our model and (d) model [35].

where $a = r(0)\sigma_\varepsilon/\sigma_\phi$, $r(0)$ is the lag-0 cross-layer 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 $\{w_1(n)\}$ in Fig. 11 for two video sequences and find that their histograms are asymmetric but decay fast on both sides. To capture this asymmetry, we use two exponential distributions to estimate the PDF of $\{w_1(n)\}$:

- We left-shift $\{w_1(n)\}$ by an offset δ to place the peak at $x = 0$;
- We then model the right side using one exponential distribution $\exp(\lambda_1)$;
- We then model the mirrored value of the left side using another exponential distribution $\exp(\lambda_2)$.
- We finally generate synthetic data $\{\tilde{w}_1(n)\}$ based on these two exponential distributions and right-shift the result by δ.

As shown in Fig. 11, the histograms of $\{\tilde{w}_1(n)\}$ are close to those of the actual data in both `Star Wars IV-B` and `The Silence of the Lambs-B`. We generate $\{\tilde{w}_2(n)\}$ in the same way and find its histogram is also close to that of $\{w_2(n)\}$. We do not show the histograms here for brevity.

We finish the section by summarizing the procedures of our algorithm to generate layered traffic:

1) Generate the base-layer traffic;
2) Produce the enhancement-layer I-trace in the wavelet domain using the base-layer algorithm in Fig. 7;
3) Generate P/B-traces using linear model (20) and (21).

VI. Analysis and Modeling of MDC Traffic

While layered coding techniques use a hierarchical structure, multiple description coding (MDC) is a non-hierarchical coding scheme that generates *equal-importance* layers. In MDC-coded sequences, each layer (i.e., description) alone can provide acceptable quality and several layers together lead to higher quality. Each description can be *individually* coded with other layered coding techniques [33].

For instance, the sample MDC sequences used in this paper are coded with both spatial scalability and MDC. After the original video stream is split into L descriptions (i.e., L-D), the l-th description ($l \in [1, L]$) contains frames $m, m+L, m+2L, \ldots, m = l$, of the original video sequence. Afterwards, description l is further encoded into one base layer and one enhancement layer using spatially scalable coding techniques. The GOP pattern of each description is $(12, 0)$. Since spatial scalability is applied to each individual description, no extra dependency or correlation is introduced between the distinct descriptions.

A. Analysis of MDC Traffic

Since each description is able to independently provide acceptable quality to users, all descriptions must share fundamental source information and thus they are highly correlated between each other. This cross-layer correlation enables the receiver/decoder to estimate the missing information of one description from another received one.

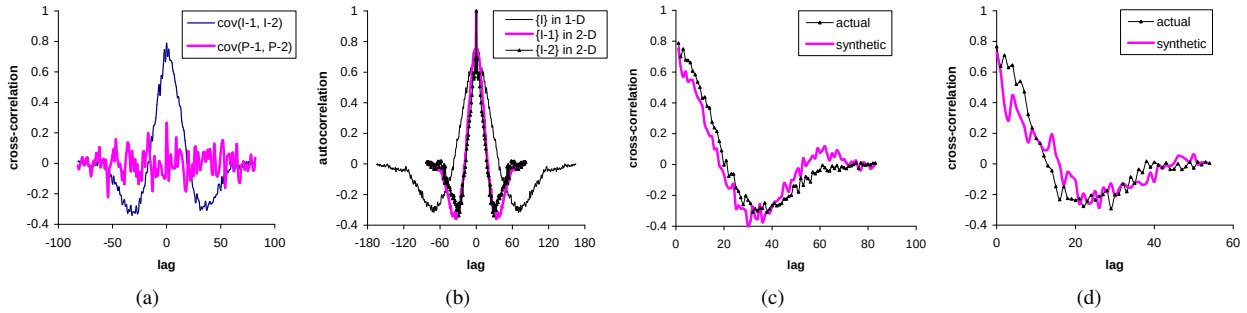
To better understand the correlation in MDC traffic, we analyze the correlation structures of MDC-coded sequences for further modeling purposes. We give the following definition for demonstration purposes. Assuming that n represents the GOP number in the i-th description of an L-D sequence, we use $\{\phi^I_{L-i}(n)\}$, $\{\phi^{P_t}_{L-i}(n)\}$, and $\{\phi^{B_t}_{L-i}(n)\}$ to represent the I-trace, the t-th P- and B-trace, respectively.

In Fig. 12(a), we show the cross-layer correlation between the I-traces and P-traces of two descriptions of the 2-D `Bridge` sequence. As the figure shows, the cross-layer correlation between I-traces is much stronger than that between P-traces, which is because I-frames contain more fundamental source information than P-frames. We also investigate the ACF structure of the original sequence and that of $\{\phi^I_{2-1}(n)\}$ and $\{\phi^I_{2-2}(n)\}$ in Fig. 12(b). One observes that the autocorrelation of $\{\phi^I_{2-1}(n)\}$ and that of $\{\phi^I_{2-2}(n)\}$ are almost identical, both of which are a shifted and scaled version of the ACF of the original sequence.

B. MDC Traffic Model

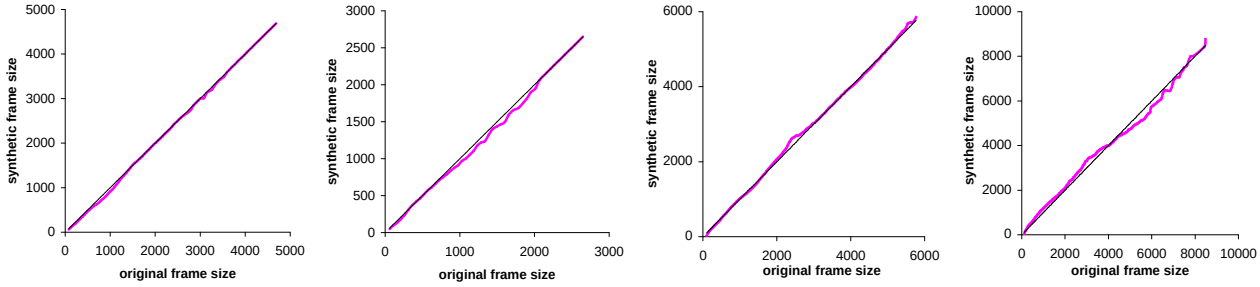
Due to the strong correlation between I-traces of different descriptions, we generate the synthetic $\{\phi^I_{L-i}(n)\}$ *simultaneously* to preserve this cross-layer correlation. Modeling I-frame sizes is also conducted in the wavelet domain to preserve their LRD and SRD properties.

Recall that the approximation coefficients preserve the correlation structure of the original signal. Also note that the ACF of different descriptions of an L-D sequence are very similar to each other (as shown in Fig. 12(b)). Therefore, after generating



(a) (b) (c) (d)

Fig. 12. (a) The cross-layer correlation between $\{\phi^I_{2-1}(n)\}$ and $\{\phi^I_{2-2}(n)\}$ and that between $\{\phi^{P_1}_{2-1}(n)\}$ and $\{\phi^{P_1}_{1-i}(n)\}$ 2-D `Bridge` traffic; (b) The autocorrelation of the original sequence and that of $\{\phi^I_{2-1}(n)\}$ and $\{\phi^I_{2-2}(n)\}$ in 2-D `Bridge`; The actual and synthetic cross-layer correlation between $\{\phi^I_{L-1}(n)\}$ and $\{\phi^I_{L-2}(n)\}$ in `Bridge`, with (c) $L = 2$ and (d) $L = 3$.



(a) (b) (c) (d)

Fig. 13. QQ plots for the synthetic (a) single-layer `Star Wars IV-A` traffic and (b) `The Silence of the Lambs-B` base-layer traffic; QQ plots for the synthetic enhancement-layer traffic: (c) `Star Wars IV-B` and (d) `The Silence of the Lambs-B`.

the first description $\{\phi^I_{L-1}(n)\}$, we borrow the ACF structure of $\{\phi^I_{L-1}(n)\}$ to construct the approximation coefficients of other descriptions. Detailed coefficients $\{D_j\}$ are estimated as *i.i.d.* mixture-Laplacian distributed random variables.

We evaluate the model performance in preserving cross-layer correlation using various MDC-coded sequences and show the actual and synthetic correlation between $\{\phi^I_{L-1}(n)\}$ and $\{\phi^I_{L-2}(n)\}$ of 2-D and 3-D `Bridge` traffic in Fig. 12(c)-(d). As the figure shows, our model indeed captures the cross-layer correlation between the different descriptions in L-D MDC traffic. To reduce model complexity, P/B-traces are modeled in the same way as single-layer P/B-traces. More performance evaluation is conducted in the following section.

VII. Model Performance Evaluation

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. In this regard, there are three popular studies to verify the accuracy of a video traffic model [31]: 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 and the second-order moment of the synthetic traffic match those of the original one, the overflow loss simulation examines the effectiveness of a traffic model to capture the temporal burstiness of original traffic.

In the following subsections, we evaluate the accuracy of our model in both single-layer and multi-layer traffic using the above three measures.

A. QQ Plots

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.

We show QQ plots of the synthetic single-layer `Star Wars IV-A` and the synthetic base layer of `The Silence of the Lambs-B` that are generated by our model in Fig. 13(a) and (b), respectively. As shown in the figure, the generated frame sizes and the original traffic are almost identical.

We also evaluate the accuracy of the synthetic enhancement layer by using QQ plots and show two examples in Fig. 13(c)-(d), which confirms the accuracy of synthetic `The Silence of the Lambs-B` and `Star Wars IV-B` enhancement-layer traffic. The figure shows that the synthetic frame sizes in both sequences have the same distribution as those in the original traffic. More simulation results for MDC and temporarily scalable traffic are shown in Fig. 14(a)-(b), which includes QQ plots for the temporarily scalable `Citizen Kane` and synthetic 1-st description in 3-D `Bridge`. As shown in Fig. 14(a)-(b), the synthetic and the original traffic have almost identical distributions.

B. Second-Order Descriptor

Different from QQ plots, the variance of traffic during various time intervals is a second-order descriptor, which

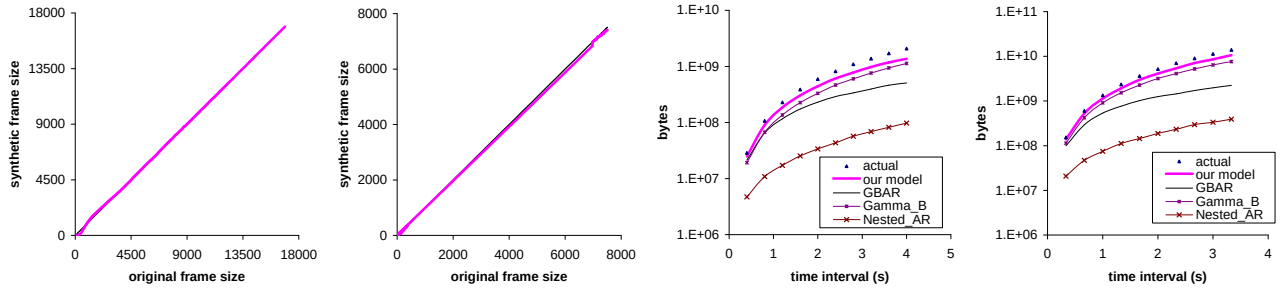


Fig. 14. QQ plots for the synthetic traffic of (a) temporarily scalable `Citizen Kane` and (b) the first description in 3-D `Bridge`; Comparison of the variance between synthetic and original traffic in single-layer (c) `Star Wars IV-A` and (d) `The Silence of the Lambs-B` base layer.

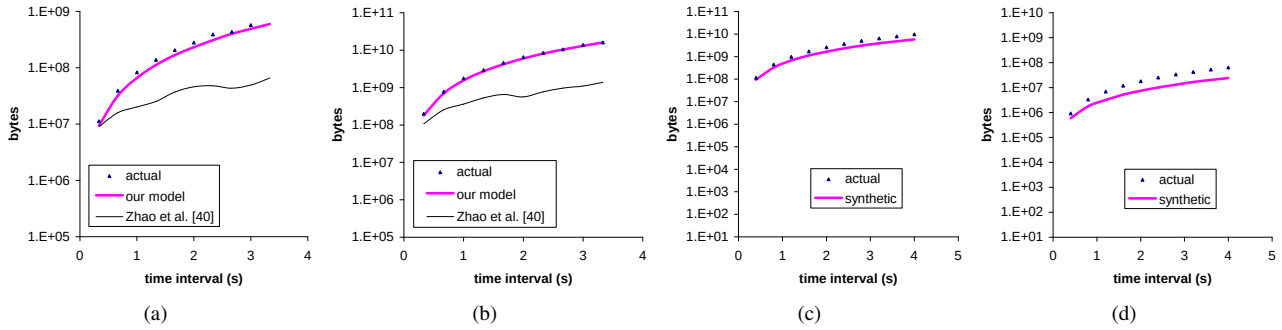


Fig. 15. Comparison of the variance between the synthetic and original enhancement layer traffic in (a) `Star Wars IV-B` and (b) `The Silence of the Lambs-B`; Comparison of the variance between the original and synthetic temporarily scalable `Citizen Kane` coded at (c) $Q = 4$ and (d) $Q = 30$.

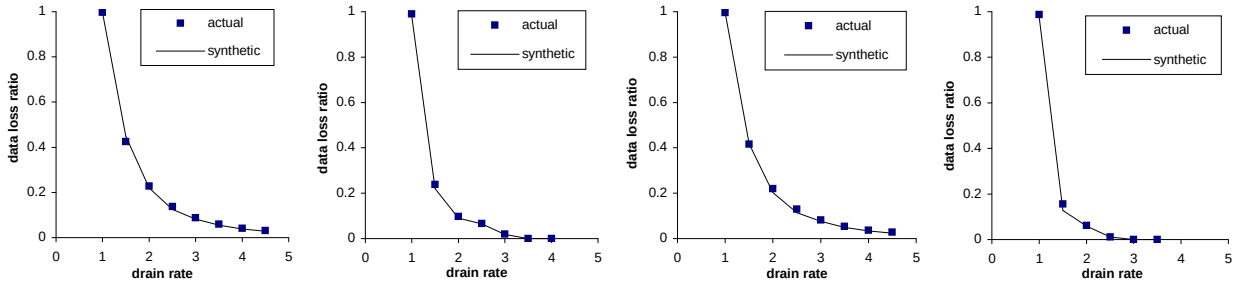
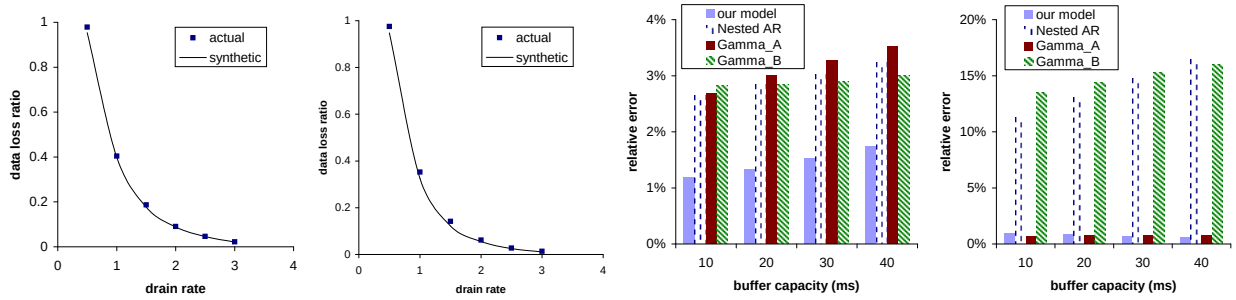


Fig. 16. Overflow data loss ratio of the original and synthetic enhancement layer traffic for (a) `The Silence of the Lambs-B` ($c = 10$ ms), (b) `Star Wars IV-B` ($c = 10$ ms), (c) `The Silence of the Lambs-B` ($c = 30$ ms), and (d) `Star Wars IV-B` ($c = 30$ ms).

shows whether the model captures the burstiness properties of the arrival processes [2]. This metric is computed as follows. Assume that the length of a video sequence is l and there are m frames in a given time interval t. We segment the one-dimensional data into n non-overlapping block of size m, where $n = l/m$. After summarizing all the data in each block, we obtain a data sequence of length n and then calculate its variance. Given a set of time intervals (i.e., various values of m), we can obtain a set of variances.

In Fig. 14(c)-(d), we give a comparison between the variance of the original traffic and that of the synthetic traffic generated from different models using different time intervals. The figure shows that the second-order moments of our synthetic traffic are in good agreement with those of the original sequences. Note that we display the y-axis on logarithmic scale to clearly show the difference among the performance of the various models.

We also compare the variance of the original enhancement layer traffic and that of the synthetic traffic in Fig. 15(a)-(b). Due to the computational complexity of Zhao's model [35] in calculating long sequences, we only take the first 5000 frames of `Star Wars IV-B` and `The Silence of the Lambs-B`. As observed from the figure, our model preserves the second-order moment of the original traffic well. In Fig. 15(c) and (d), we display the variance of the original and synthetic traffic in two temporally scalable `Citizen Kane` sequences coded at $Q = 4$ and $Q = 30$, respectively. The figure demonstrates that the synthetic traffic captures the burstiness of the original traffic very well. Note that model [35] works slightly better in short sequences (e.g., 30 seconds) with very few scene changes; however, our model still outperforms [35] in such scenarios (not shown for brevity).



(a) (b) (c) (d)

Fig. 17. Overflow data loss ratio of the original and synthetic temporarily scalable `Citizen Kane` for (a) $c = 10$ ms and (b) $c = 40$ ms; Given $d = \bar{r}$, the error e of various synthetic traffic in H.264 `Starship Troopers` for (c) $Q = 1$ and (d) $Q = 31$.

TABLE II
RELATIVE DATA LOSS ERROR IN `Star Wars IV-A` GOP $(12, 2)$

Buffer capacity	Traffic type	Drain rate $2\bar{r}$	Drain rate $4\bar{r}$	Drain rate $5\bar{r}$	Buffer capacity	Traffic type	Drain rate $2\bar{r}$	Drain rate $4\bar{r}$	Drain rate $5\bar{r}$
20 ms	Our model	0.93%	0.61%	1.13%	30 ms	Our model	0.25%	0.33%	0.95%
	GOP-GBAR [9]	3.84%	2.16%	3.77%		GOP-GBAR [9]	4.94%	3.33%	5.68%
	Nested AR [20]	5.81%	2.77%	8.46%		Nested AR [20]	6.94%	4.14%	9.92%
	Gamma_A [31]	5.20%	0.61%	2.57%		Gamma_A [31]	4.88%	1.10%	4.48%
	Gamma_B [31]	4.89%	1.93%	2.05%		Gamma_B [31]	4.67%	2.17%	4.03%
	Wavelet model [21]	12.6%	48.4%	57.7%		Wavelet model [21]	21.4%	50.0%	57.1%

TABLE III
RELATIVE DATA LOSS ERROR IN `Bridge`

Buffer capacity	Traffic type	Drain rate $1\bar{r}$	Drain rate $1.5\bar{r}$	Drain rate $2\bar{r}$	Drain rate $2.5\bar{r}$	Buffer capacity	Traffic type	Drain rate $1\bar{r}$	Drain rate $1.5\bar{r}$	Drain rate $2\bar{r}$	Drain rate $2.5\bar{r}$
10 ms	Descrip. 1	1.64%	1.38%	0.49%	0%	30 ms	Descrip. 1	3.03%	0.93%	0.25%	0.65%
	Descrip. 2	0.32%	2.36%	0.72%	0.30%		Descrip. 2	0.22%	0.18%	0.49%	0%
20 ms	Descrip. 1	2.28%	0.53%	0.49%	1.22%	40 ms	Descrip. 1	3.38%	0.96%	1.07%	0.73%
	Descrip. 2	0.75%	0.35%	0.48%	0%		Descrip. 2	1.85%	1.69%	0%	0.31%

C. Buffer Overflows

Besides the distribution and burstiness, we are also concerned how well our approach preserves the temporal information of the original traffic. A common test for this purpose is a leaky-bucket simulation, which is to pass the original or the synthetic traffic through a generic buffer with capacity c and drain rate d [31]. The drain rate is the number of bytes drained per second and is simulated as different multiples of the average traffic rate $\bar{r}$.

We first examine the model accuracy using the data loss ratio. In Fig. 16, we show the overflow in the enhancement layers of both `The Silence of the Lambs-B` $(54,000$ frames) and `Star Wars IV-B` $(108,000$ 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 $\bar{r}$. In Fig. 17(a)-(b), we illustrate the loss rate of the original and synthetic temporarily scalable `Citizen Kane` coded with $Q = 4$ with buffer capacity $c = 10, 40$ ms. All simulation results show that the synthetic traffic preserves the temporal information of the original traffic very well.

To better evaluate the model performance, we also compare the accuracy of our model with that of several other traffic models using the relative error as the main metric. We define error e as the relative difference between the *actual* packet loss p in a leaky-bucket simulation and the *synthetic* packet loss p_{model} obtained under the same simulation constraints using synthetic traffic generated by each of the models: $e = |p - p_{model}|/p$.

In Table II and Table IV, we illustrate the values of e for various buffer capacities and drain rates d, for two sequences of different GOP structures. As shown in the tables, 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. Fig. 17(c)-(d) shows the relative error e of synthetic traffic generated from different models in H.264 `Starship Troopers` coded at $Q = 1$ and $Q = 31$. Since the GOP-GBAR model [9] is specifically developed for MPEG traffic, we do not apply it to H.264 sequences. The figure shows that our model outperforms the other three models in `Starship Troopers` coded at small Q and performs as well as Gamma_A [31] at large Q (relative error e of both models is less than 1% in Fig. 17(d)).

We also compare the computation complexity of different methods using Matlab under the same computing environment. The processing time of our model is 4.8 sec, which is less than that of the pure wavelet model [21] (11.65 sec), that of Gamma_A/B [31] (5.88 sec) and that of Nested AR [20]

TABLE IV
RELATIVE DATA LOSS ERROR IN `Star Wars IV` GOP $(16, 1)$

Buffer capacity	Traffic type	Drain rate $\bar{r}$	Drain rate $3\bar{r}$	Drain rate $4\bar{r}$
20 ms	Our model	2.20%	3.61%	3.57%
	GOP-GBAR [9]	2.46%	8.21%	22.62%
	Nested AR [20]	2.42%	10.47%	4.59%
	Gamma_A [31]	3.09%	16.37%	34.18%
	Gamma_B [31]	10.82%	27.89%	38.78%
	Wavelet model [21]	5.58%	51.75%	44.44%

(9.90 sec), but is higher than that of GOP-GBAR [9] (1 sec). However, given the higher accuracy, our model seems to offer a good tradeoff between complexity and fidelity.

We are not able to show results for previous multi-layer models given the nature of our sample sequences since the model in [2] is only applicable to sequences with a CBR base layer and the one in [35] is suitable only for short sequences. Therefore, we only illustrate the relative error e of the synthetic 2-D MDC-coded `Bridge` generated by our model in Table III. As shown in the table, our method accurately preserves the temporal information of MDC traffic.

VIII. CONCLUSION

In this paper, we presented a framework for modeling MPEG-4 and H.264 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 [31], 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. The linear computation complexity of our model is no worse than that of [9], [20], [31] and significantly lower than the $O(N^2 \log N)$ complexity of [35].

This framework also applies to adaptive GOP structure cases, but requires small modifications, e.g., with large GOP size, we prefer using the neighboring correlation (19) rather than Lag-0 Intra-GOP correlation (11) to model P/B frames. The main limitation of our model is that it does not apply to video traffic generated by codecs without a concept of GOP structure (e.g., motion JPEG 2000).

Future work involves understanding traffic prediction using the proposed framework and modeling of non-stationary VBR sources.

REFERENCES

[1] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics, 2005.

[2] K. Chandra and A. R. Reibman, "Modeling One- and Two-Layer Variable Bit Rate Video," *IEEE/ACM Trans. Netw.*, vol. 7, June 1999.

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

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

[5] M. Dai and D. Loguinov, "Analysis and Modeling of MPEG-4 and H.264 Multi-Layer Video Traffic," *IEEE INFOCOM*, Mar. 2005.

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

[7] P. Embrechts, F. Lindskog, and A. J. McNeil, "Modelling Dependence with Copulas and Applications to Risk Management," *Handbook of Heavy Tailed Distributions in Finance*, Elsevier/North-Holland, Amsterdam, 2003.

[8] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation (Extended Version)," Available from http://www-tkn.ee.tu-berlin.de.

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

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

[11] 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.

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

[13] J. C. Goswami and A. K. Chan, *Fundamentals of Wavelets: Theory, Algortihms, and Applications*, John Wiley & Sons, 1999.

[14] 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.

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

[16] 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.

[17] M. R. Ismail, I. E. Lambadaris, M. Devetsikiotis, and A. R. Kaye, "Modeling Prioritized MPEG Video Using TES and a Frame Spreading Strategy for Transmission in ATM Networks," *Proc. of INFOCOM*, Apr. 1995.

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

[19] 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, Apr. 2001.

[20] 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.

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

[22] 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.

[23] 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.

[24] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*, John Wiley & Sons, 2000.

[25] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 Fine-grained Scalable Video Coding Method for Multimedia Streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, Mar. 2001, pp. 53–68.

[26] G. Ramamurthy and B. Sengupta, "Modeling and Analysis of a Variable Bit Rate Multiplexer," *IEEE INFOCOM*, Florence, Italy, 1992.

[27] M. Reisslein, J. Lassetter, S. Ratnam, O. Lotfallah, F. H. P. Fitzek, and S. Panchanathan, "Video Traces for Network Performance Evaluation," Available from http://trace.eas.asu.edu.

[28] 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*, Mar. 2000.

[29] 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.

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

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

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

[33] Y. Wang, A. R. Reibman, and S. Lin, "Multiple Description Coding for Video Delivery," *Proceedings of the IEEE*, vol. 93, No. 1, Jan. 2005.

[34] G.W. Wornell, *Signal Processing with Fractals: A wavelet based approach*, Prentice Hall, 1996.

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