Traffic Models for H.264 Video Using Hierarchical Prediction Structures

Akshay Pulipaka*, Patrick Seeling† and Martin Reisslein*
*School of Electrical, Computer, and Energy Engineering
Arizona State University, Goldwater Center MC 5706, Tempe, AZ 85287, USA
http://trace.eas.asu.edu/
Email: {akshay.pulipaka, reisslein}@asu.edu
†Central Michigan University, Mount Pleasant, MI 48859, USA

***Abstract*—We present different video traffic models for H.264 variable bit rate (VBR) videos. We propose our models on top of the recent unified traffic model developed by Dai et al. [1], which presents a frame-level hybrid framework for modeling MPEG-4 and H.264 multi-layer VBR video traffic. We exploit the hierarchical predication structure inherent in H.264 for intra-GoP (group of pictures) analysis. We model the children frames by considering various combinations of the correlation between the parent frames in the prediction structure. Our simulations show that modeling using the hierarchical prediction structure indeed improves capturing the statistical features of the videos and prediction of network performance, without an increase in the complexity as compared to the unified traffic model by Dai et al. [1], which was shown earlier to be better than previous traffic models.**

***Index Terms*—Hierarchical prediction structures, H.264 SVC, intra-GoP correlation, video traffic modeling,**

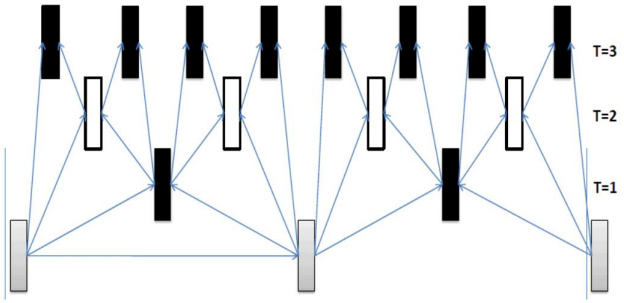


Fig. 1. Hierarchical prediction structure in a G16B7 H.264 video sequence showing 3 temporal levels for the B frames. The arrows indicate the direction of prediction.

I. Introduction

Video traffic modeling plays a major part in network traffic analysis. It is imperative to network design and simulation, providing Quality of Service (QoS) to network applications, besides providing insights into the coding process and structure of video sequences. Many models have been proposed in the past for MPEG video traffic. But very few studies have considered H.264 [2] video traffic modeling and the multi-layer aspects of video traffic streamed over the Internet. Prominent among those is the recent unified traffic model developed by Dai et al. [1] which presented a framework by incorporating wavelet domain analysis into time-domain modeling.

Video traffic modeling has been an active research area for quite many years now. A plethora of models have been developed over time for various applications meeting various demands. There are different ways to classify these models, a few of which will be described in the section II, based on our research interest.

We briefly present our major motivation for this study. We wish to make the hierarchical B-frames structure (shown in Figure 1), which is inherent in H.264, a part of the modeling process by treating correlation between each different pair of hierarchical B-frames separately. This is in contrast to the H.264 video traffic model developed by Dai et al. [1], where the hierarchical B-frames were not considered. Instead, the correlation between any pair of B-frames was considered to be the same. Our approach improves the overall traffic modeling of H.264 single layer video streams, without adding much complexity, as it considers the correlation between the B-frames of different hierarchical levels instead of treating all B-frames alike as suggested by [1]. We have proposed this change through a linear model which estimates the B-frames from the lower hierarchial level B-Frames. Specifically, we propose and evaluate three B-frame prediction models, which are detailed in Section IV. We do not change the estimation of I-frames and the I-B frames correlation presented in [1].

A *Group of Pictures* (GoP) of an encoded video stream is defined as one I-frame and all subsequent P and B-frames before the next I-frame in the stream. One GoP of a G16B7 structure is shown in Figure 1, which means there are 7 B-frames in between any I and P-frame and an I-frame repeats every 16 frames i.e IBBBBBBBPBBBBBBBI... Also note that the 7 B-frames of the first-half of the GoP in Figure 1, are divided into 3 hierarchical levels represented by T = 1 to 3. The I/P-frames belong to the coarsest temporal level T = 0. The arrows in the figure indicate the direction of prediction. Two arrows culminating at a particular frame indicate that the child frame has two parent frames. As illustrated in Figure 1 the pattern of prediction for every B-frame is quite different from another B-frame, hence these individual correlations can be exploited to improve the modeling process. We have additionally validated our approach by using a G16B15 GoP

structure, where the middle P-frame is replaced by another B-frame.

The rest of the paper is organized as follows. In Section II, we classify the varied video traffic models based on different metrics and point the reader to research articles about these traffic models. In Section III, we review in detail the traffic model developed by Dai et al. [1] including the models for estimating I, B and P frames. We propose the changes introduced to [1] and all our models in the following Section IV. The results are shown in Section V and we compare our models to the model proposed in [1]. We summarize our conclusions in Section VI and propose ideas for future/ongoing research.

II. Related Works: Video Traffic Model Classification

A good traffic model is expected to capture the statistical characteristics of video sequences (such as frame size distribution and autocorrelation function (ACF)) and predict the network performance (such as buffer overflow probabilities and packet loss [1]). Several popular models have captured the distribution of frame sizes of video traffic, quite accurately. A few of the most popular distribution models that have been studied are the lognormal model [3], normal model [4], Gamma model [5] and other hybrid models such as Gamma/Pareto [6] and Gamma/Lognormal [7]. Modeling the ACF structure of VBR traffic is much more challenging, because these models need to accurately consider and predict for both Long-Range Dependence (LRD) and Short-Range dependence (SRD). VBR video traffic can be a complex mixture of both SRD and LRD. SRD traffic is generally exhibited by real-time application traffic, such as Voice over IP (VoIP) [8]. LRD traffic is observed for non-real-time applications such as web-request traffic [9] and Ethernet data traffic [10]. The presence of both SRD and LRD properties in video traffic indicates that the ACF structure is similar to SRD processes at small time lags and LRD processes at large time lags [11]. Hence we would require a model which can model both SRD/LRD ACF structures.

Some variants of Markov processes [12]–[14] model SRD processes reasonably well. On the other hand, models such as Fractional Gaussian Noise (FGN) processes capture LRD, but not SRD [10]. In the recent past, a few models have emerged, which were finally able to model both SRD and LRD processes. This complicated problem has been studied by many, but only a very few were successful such as the Nested Auto-regressive (AR) model [6], wavelet model [15], and unified MPEG-4/H.264 model [1].

Single-layer traffic models are the most common type of video traffic models. Much less work has been done though to model multi-layer video traffic. A transform-expand-sample (TES)-based model [16], uses two levels of priority for modeling MPEG video. The finite-state Markov chain model developed by Chandra et al. [17] was used to model single and two layer scalable video traffic. Later, Zhao et al. [18] developed a K-state Markov chain based model on frame-size clusters. Dai et al. [1] have developed a unified traffic model, in which they have tried to fully exploit the cross-layer correlation between base-layer and multiple enhancement layers. Fiems et al. [19] have proposed a multivariate Markovian traffic model to characterize H.264/SVC scalable video traces. More recently multiview video traffic modeling has been explored [20].

Dai et al. [1] have a good discussion on different traffic models. Wavelet analysis is being used increasingly due to their advantages in capturing both LRD and SRD properties of video. In the same reference, the authors have used wavelet analysis to estimate I frames, which we included into our modeling as well. Please refer to [1] for a detailed discussion on wavelet modeling.

III. Unified Traffic Model Proposed by [1]

We base our work on prior research presented by Dai et al. in [1], where both single-layer and scalable-layer traffic was considered based on the publicly available video traces from [21]. We are interested in the presented model's intra-GoP correlation analysis, which is the correlation between I and P/B-frame sizes within a GoP as described earlier in Section I. The wavelet-based estimation of I-frames is described in Section III-A. We present a brief summary of the intra-GoP correlation analysis performed by Dai et al. [1] here, which motivated us to consider an extension for hierarchical prediction structures. Note that the P and B-frame estimation is similar as presented in [1], hence they are used interchangeably.

- Dai et al. [1] show the correlation between I and different P-frames belonging to different time instances i.e $I-P1, I-P2, I-P3$,... where i = 1,2,3,... represent different time instances of P Frames within a GoP. Similar terminology is used to describe B-frames. Please note that these time instances are different than the hierarchical B-Frames, as a B-frame belonging to time instance of 3, may not necessarily refer to the 3rd Hierarchical B-frame structure of Figure 1. Dai et al. have not shown any results for correlation between B-frames of different hierarchical levels. The authors concluded that the "time instance" or "i" value of P or B-frames does not matter as they got similar correlation values for $I-P1, I-P2, I-P3$, etc.
- The results provided by Dai et al. in [1] have been mainly for the MPEG-4 standard and CIF resolution for single-layer traffic analysis.
- Dai et al. suggest that the B-frames can be modeled from the I-frames in two ways:
  a. If the correlation between I and B-frames is not strong, as is the case with a few video sequences, then the B-frames can be represented by an i.i.d lognormal random number generator.
  b. If the correlation between I and B-frames is strong, a linear model and a synthetic generalized gamma distribution can model the B-frames, as

$$\phi_i^B(n) = a\tilde{\phi}^I(n) + \tilde{\upsilon}_B(n), \qquad (1)$$

where $\phi_i^B(n)$ is the size of the ith B-frame in GOP n. $\tilde{\phi}^I(n)$ is defined as

$$\tilde{\phi}^I(n) = \phi^I(n) - \mathbb{E}[\phi^I(n)], \quad (2)$$

where $\phi^I(n)$ is the size of the I-frame in the nth GoP and the estimation of which is discussed in detail in Section III-A. $\tilde{v}_B(n)$ is a synthetic process, independent of $\tilde{\phi}^I(n)$ and is described in more detail in the Section IV. a is defined as

$$a = r(0)\sigma_B/\sigma_I, \quad (3)$$

where $r(0)$ is the lag-0 correlation between $\tilde{\phi}^I(n)$ and $\phi_i^B(n)$, σ_B and σ_I are the standard deviation of $\phi_i^B(n)$ and $\tilde{\phi}^I(n)$, respectively.

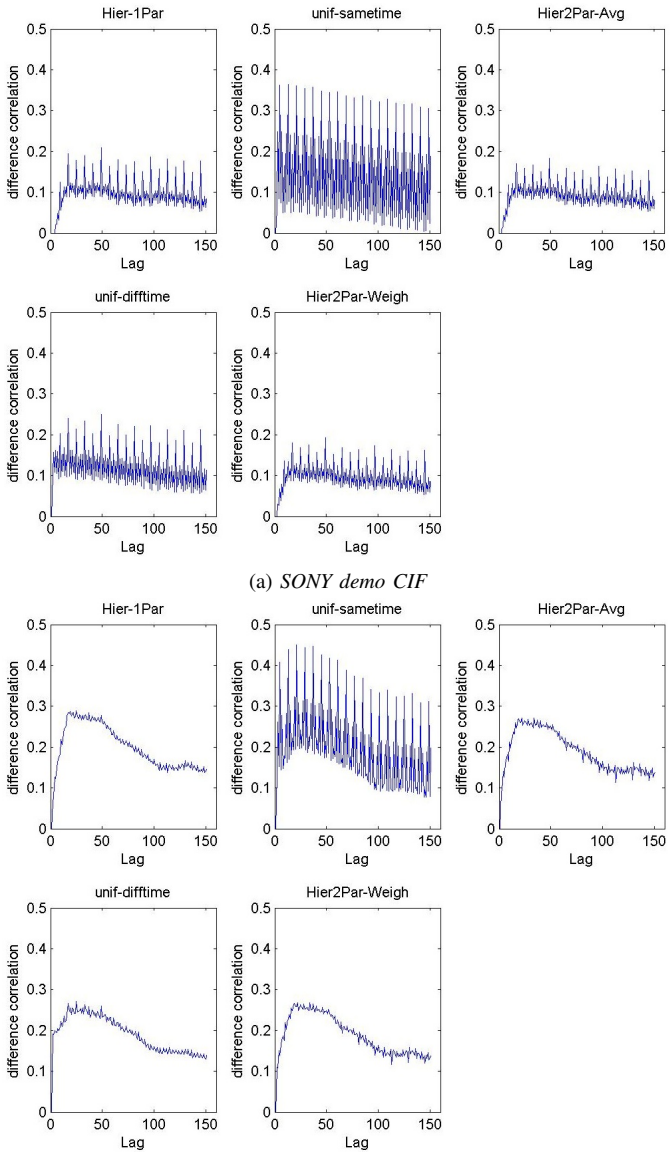
A. Wavelet Model for Estimating I-Frames [1]

In the following, we highlight the main properties of the wavelet model proposed by Dai et al. in [1] for predicting the size of I-frames. The model is a four-step process which can be summarized as follows.

- The I-frame sizes are modeled in the wavelet domain using the estimated approximation $\{A_k\}$ and detailed $\{D_k\}$ coefficients, where k represents the decomposition level. The required signal is typically decomposed using a family of basis functions, which includes a high-pass wavelet function and a low-pass scaling filter which generates the detailed coefficients, and approximation coefficients of the original signal respectively [1]. Haar wavelet function has been used in the model.
- The next step is the modeling of the detailed and approximation coefficients. In this step the detailed coefficients $\{D_k\}$, are modeled by a mixture-Laplacian distribution, whose pdf is given by

$$f(x) = p\frac{\lambda_0}{2}e^{-\lambda_0|x|} + (1-p)\frac{\lambda_1}{2}e^{-\lambda_1|x|} \quad , \quad (4)$$

  where 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. In [1] it is shown that the histogram of the mixture-Laplacian synthetic coefficients $\{D_1\}$ is much closer to the actual one than the other discussed distributions.
- The third step is to model the approximation coefficients $\{A_k\}$ which are non-negligibly correlated and are not i.i.d. To preserve the correlation of approximation coefficients and to achieve the expected distribution, the coarsest approximation coefficients are modeled as dependent random variables with marginal Gamma distributions. The sub-steps involved are [1]
  a. Generate N dependent Gaussian variables x_i using a $k \times k$ correlation matrix, where N is the length of $\{A_k\}$ and the correlation lags l is chosen to be a value like the average scene length. The correlation matrix is obtained from the actual coefficients $\{A_k\}$.
  b. Apply the Gaussian CDF $F_G(x)$ directly to x_i to convert them into a uniformly distributed set of variables $F_G(x_i)$.
  c. Pass the result from the last step through the inverse Gamma CDF to generate (still dependent) Gamma random variables.
- Using the estimated approximation and detailed coefficients from above, the inverse wavelet transform is performed to generate the I-frame sizes.



(a) *SONY demo CIF*

(b) *Transporter2 1080p HD*

Fig. 2. The difference ACF plots for all estimated frames from five models vs the original traffic. The *unif_sametime* [1] plots indicate highest deviation from original ACF traffic, whereas our models *Hier_2Par_avg* and *Hier_2Par_weigh* have lowest deviation.

IV. Our Models for Estimating B-Frames

In the following, we outline our model, which uses a similar linear model as described in Equation (1) to model the

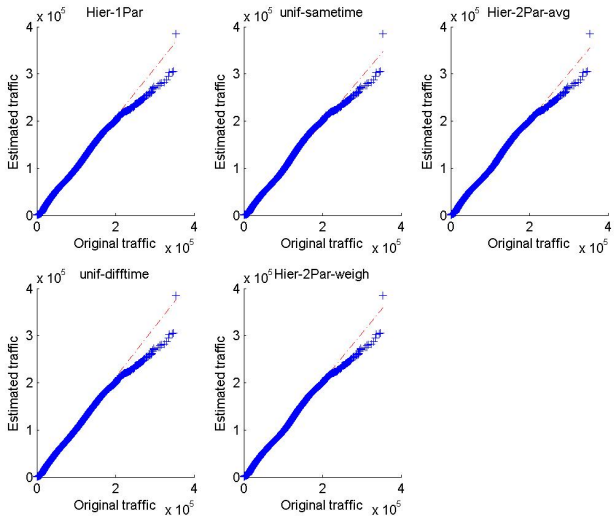


Fig. 3. QQ Plots for traffic generated by five models for *Transporter2* 1080p HD. Data generated by our model *Hier_2Par_avg* lies closest to the reference line, similar to *unif_sametime* [1].

correlation between I-frames and B-frames, except that there is a separate relation between I and each of the different B-frames belonging to separate hierarchical stages. We introduce j as the hierarchical B frame level and in turn derive

$$\phi_i^{B_j}(n) = a\tilde{\phi}^I(n) + \tilde{\upsilon}_B(n), \tag{5}$$

where i still signifies the different time instances of B-frames at a particular hierarchical level. In our model, $\phi_2^{B_3}(n)$ would hence represent the size of the second B-frame belonging to the third hierarchical B level.

The authors of [1] suggested that to estimate the residuals, a synthetic process $\tilde{\upsilon}_B(n)$ could be used. $\tilde{\upsilon}_B(n)$ is independent of the estimated I-frame sizes $\tilde{\phi}^I(n)$ and is estimated from the actual residuals $\upsilon_B(n)$ as

$$\upsilon_B(n) = \phi_i^{B_j}(n) - a\tilde{\phi}^I(n). \tag{6}$$

As in [1], we employ a generalized Gamma Distribution to estimate $\upsilon_B(n)$. We used the parameter estimation algorithm for the generalized Gamma distribution, described in [22], to estimate the parameters of the synthetic process. The pdf of a generalized gamma distribution with parameters γ, α, and β is

$$f(x) = \frac{|\beta|\,\gamma^{\alpha\beta}x^{\alpha\beta-1}e^{-(\gamma x)^{\beta}}}{\Gamma(\alpha)} \qquad \forall \quad x > 0, \tag{7}$$

where $\beta \in \mathbb{R}$ and $\alpha > 0$ are shape parameters, $\gamma \in \mathbb{R}$ is a scale parameter, and $\Gamma(\cdot)$ denotes the gamma function. To get comparable results, we have used the parameter estimation method used by Dai et al. [1] in this study.

To improve over the modeling accuracy of [1], we introduce additional estimations, in addition to the modeling of the B-frames from I-frames as shown above. We do this by considering the correlation between B-frames belonging to different hierarchical levels. In particular, we model the B-frame of a certain level (a child frame), by considering different combinations of the correlations between that child B-frame and its parent I/P and parent B-frames from a lower hierarchical level (considering the coarsest level to be that of the I-frames).

To model the B-frames of a higher level (B_j), from the lower hierarchical level B-frames (B_k), we follow a linear model similar to Equation (5). Specifically,

$$\phi_i^{B_j}(n) = b\tilde{\phi}_i^{B_k}(n) + \tilde{\omega}_{B_jB_k}(n), \tag{8}$$

where the value of j in $\phi_i^{B_j}(n)$ is always higher than the value of k in & $\tilde{\phi}_i^{B_k}(n)$, corresponding to the child and the parent B-frames of different hierarchical B-frames. B_j from Equation (5) should not be confused with B_j from Equation (8), as j takes different values 1,2,... based on the number of hierarchical levels. However, the i values in $\phi_i^{B_j}(n)$ & $\tilde{\phi}_i^{B_k}(n)$ may or may not be the same, depending on the hierarchical B-frame and GoP structures. b is defined as

$$b = r(0)\sigma_{B_j}/\sigma_{B_k}. \tag{9}$$

$\omega_{B_jB_k}(n)$ is a synthetic process and we model it as a generalized Gamma distribution as described above.

In this paper, we propose and evaluate three models, which differ in the estimation process of the correlation between children and parent B-frames as follows:

- *Hier_1Par* estimates the B-frames from only one of the two parents shown in Figure 1. The one parent is chosen to be the one with the higher correlation of the two parents.
- *Hier_2Par_avg* estimates the B-frames from both the parents by taking an average of the two estimations.
- *Hier_2Par_weigh* estimates the B-frames as the weighted sum of the frame index estimations of the two parents weighed by the absolute difference in frame index (distance). Frame index difference is considered since the correlation is observed to be heavily dependent on the distances of the frames.

When the estimations from two parents of the same child frame are averaged as in Equation (8), we get the values for our *Hier_2Par_avg* model. If these two estimations are weighted based on the frame numbers of the parents within the GoP, we get our model *Hier_2Par_weigh*, whereas if we just consider the estimation of one parent (with the higher correlation), we get the model *Hier_1Par*.

In addition we have also considered two cases for the traffic model developed by Dai et al. [1]. They have mentioned that the time-instance of B/P-frames does not matter, hence they calculate the correlation between I and any B/P-frame and use that for all the possible parent-child combinations. We call this original model as *unif_sametime*. We made a modification to this case to consider the correlation between I and B/P-frames of different time instances to be different, and not same as the previous case. We call this model as *unif_difftime*. Now we compare these five models in the next section.

V. RESULTS

We do not show any results for MPEG-4 video traces as shown in [1] for single layer video traces since MPEG-4 has decreased in significance. But we did run our models on MPEG-4 traces and results are similar to the ones shown here. Due to space constraints we show results for two video sequences, *SONY demo* and *Transporter2* [21]. *SONY demo* is of CIF (352 X 288) pixel resolution and has GoP structure of G16B7 at 30 fps and QP 10, whereas *Transporter2* is of 1080p HD (1920 X 1080) pixel resolution and has GoP structure of G16B15 at 24 fps and QP 22. *SONY demo* has different scenes of varied motion clubbed together for 10 minutes. *Transporter2* contains first 30 minutes of high motion and texture scenes from the motion picture. Hence we have lot of variations in our test conditions. Results for other combinations will be updated at [21]. We have used H.264-SVC Single layer encoding i.e JSVM 9.15.

We perform different studies to demonstrate accuracy of our three models and compare them with the model proposed by Dai et al. in [1] i.e *unif_sametime*. We propose a variation to the model proposed by Dai et al. i.e *unif_difftime*, which considers B frames at different time-instances. The order of models in figures and tables is related to number of parent frames and type of prediction.

A. ACF Plots

The first test is to plot the ACF function for all models. The ACF shows similarities between observations as a function of time separation between them [23]. ACF between frame sizes is generally used for traffic correlation. The difference between the original traffic ACF and a specific model's ACF is shown in Figure 2, to highlight differences between both ACFs. Difference correlation is plotted as a function of lag time. All models are shifted at short lag times in ACF plots, whereas they get close to the original ACF at longer lags. This is especially true for the *Transporter2* HD sequence. *unif_sametime* proposed by Dai et al. [1] fails to retrace smaller peaks, whereas our three models retrace the exact shape of original traffic ACF. As the difference ACF plots in Figure 2 indicates, *unif_sametime* has most deviations compared to the original ACF.

B. QQ Plots

The second test is Q-Q or QQ plots ("Q" stands for quantile). It is a statistical technique in which quantiles of two probability distributions are plotted against each other. The points lie close to a 45 degree reference line if they are similar to each other. So ideally we would like frame size points of estimated traffic models to be as close to the reference line as possible. These plots are shown in Figure 3 for *Transporter2*. The plots suggest that our model *Hier_2Par_avg* and the *unif_sametime* model proposed by Dai et al. [1] are very similar and represent the original data almost accurately. The QQ plots for *SONY demo* follow a similar pattern. The plots suggest that the estimated traffic from our models conforms with the distribution of the original traffic.

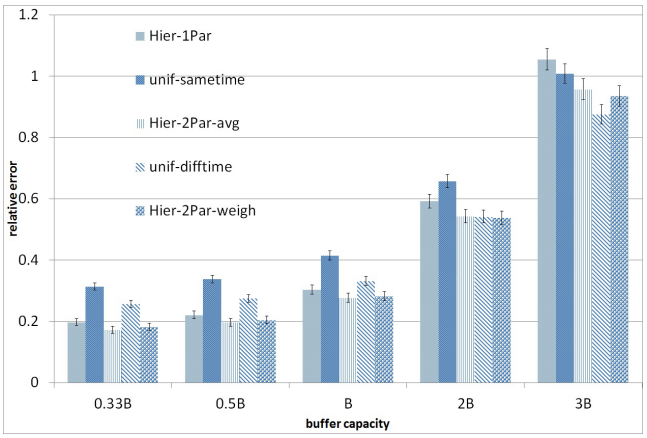
TABLE I
OVERFLOW DATA LOSS VALUES FOR FIVE MODELS VS ORIGINAL TRAFFIC. FOR CIF, OUR MODELS *Hier_2Par_avg* AND *Hier_2Par_weigh* HAVE DATA LOSS VALUES CLOSEST TO ORIGINAL TRAFFIC, WHEREAS FOR HD SEQUENCE OUR OTHER MODEL *Hier_1Par* PERFORMS BEST

Traffic type	Drain rate d				
	$\bar{r}$	$1.5\bar{r}$	$2\bar{r}$	$2.5\bar{r}$	$3\bar{r}$
SONY demo CIF					
Original	0.261	0.122	0.071	0.041	0.023
Hier-1Par	0.284	0.150	0.091	0.061	0.042
unif-sametime [1]	0.248	0.148	0.099	0.069	0.048
Hier-2Par-avg	0.271	0.144	0.089	0.061	0.042
unif-difftime	0.268	0.148	0.093	0.063	0.043
Hier-2Par-weigh	0.270	0.144	0.090	0.061	0.042
Transporter2 1080p HD					
Original	0.1891	0.0684	0.0270	0.0115	0.0052
Hier-1Par	0.1917	0.0638	0.0261	0.0117	0.0050
unif-sametime [1]	0.1642	0.0593	0.0269	0.0124	0.0054
Hier-2Par-avg	0.1819	0.0586	0.0246	0.0113	0.0050
unif-difftime	0.1855	0.0647	0.0270	0.0119	0.0051
Hier-2Par-weigh	0.1824	0.0592	0.0248	0.0114	0.0050

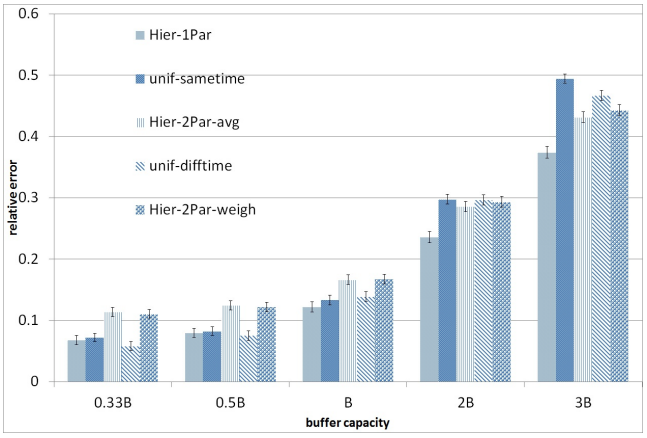
C. Buffer overflows

The third measure we have used is the buffer overflow test which verifies whether the models preserve the temporal information of the original traffic. We employed leaky-bucket simulation as a test [5] and used the simulation steps described in [24]. The test consists of sending traffic over a buffer with a capacity B at a drain rate d, which is usually varied as multiples of mean traffic rate $\bar{r}$ [1]. Buffer overflow data loss is calculated and plotted as a function of drain rates. The data loss numbers give a clearer picture than the plots, hence we have shown values in Table I for different drain rates. It was recommended in [24] that $B = 192$ kB for H.264 CIF videos. Please note that 1080p HD video requires larger buffer sizes than CIF videos. The values in Table I indicate that *Hier_2Par_weigh* and *Hier_2Par_avg* have data loss values closest to the original traffic and the *unif_sametime* model proposed by Dai et al. [1] has the most deviation. This indicates that our models preserve the temporal information of original traffic better than the model proposed by Dai et al. [1]. For each simulation we run many independent replications (N), until the 95% confidence interval (CI) value is less than 10% of the corresponding sample mean. For cases shown here N = 96.

We also calculate the relative error between the data loss of the original traffic and the estimated traffic generated by the models. Error e is the ratio of overflow data loss difference between original traffic and estimated traffic to that of the original data loss. We calculate e for all models and present values in Figure 4 where they are plotted as a function of the different buffer capacities. We have covered a wide range of buffer capacities, such as 50 ms to 1.5 s. For drain rate d = $2\bar{r}$, we plot the error bars for the 95% CI values obtained for N = 2000. The results from Figure 4 indicate that our models *Hier_2Par_weigh* and *Hier_2Par_avg* have the lowest relative error value for CIF sequences, whereas the *unif_sametime* model proposed by Dai et al. [1] has highest relative error



(a) *SONY demo CIF*



(b) *Transporter2 1080p HD*

Fig. 4. Relative error plots for five models for drain rate $d = 2\bar{r}$. (a) For CIF our models *Hier_2Par_avg* and *Hier_2Par_weigh* have lowest relative error. (b) for HD our model *Hier_1Par* has lowest relative error.

values. For the HD sequence, our model *Hier_1Par* and the variation model *unif_difftime* have the lowest relative errors.

VI. Conclusion

In this paper we proposed different traffic models which exploited the hierarchical prediction structures inherent in H.264. Our investigations indicate that our models *Hier_2Par_weigh*, *Hier_2Par_avg*, and *Hier_1Par* capture traffic autocorrelation characteristics more accurately than the model proposed by Dai et al [1], whereas the QQ plots are similar. Also, our models give favorable accuracy of buffer overflow probability for many scenarios; whereby our *Hier_1Par* model performs particularly well for HD sequences and our *Hier_2Par_weigh* model for CIF sequences. We are investigating these emerging different accuracies for different video resolutions in ongoing research. Currently we are working on extending this model to scalable and multiview traffic.

Acknowledgment

We are grateful to M. Dai and D. Loguinov for sharing their unified traffic model. This work was supported by the National Science Foundation Grant No. CRI-0750927.

References

[1] M. Dai, Y. Zhang, and D. Loguinov, "A unified traffic model for MPEG-4 and H.264 video traces," *IEEE Transactions on Multimedia*, vol. 11, no. 5, pp. 1010–1023, aug. 2009.

[2] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, pp. 560–576, Jul 2003.

[3] M. Krunz and S. K. Tripathi, "On the characterization of VBR MPEG streams," in *in Proc. ACM SIGMETRICS*, 1997, pp. 192–202.

[4] M. Nomura, T. Fujii, and N. Ohta, "Basic characteristics of variable rate video coding in ATM environment," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, pp. 752–760, Jun. 1989.

[5] U. Sarkar, S. Ramakrishnan, and D. Sarkar, "Modeling full-length video using markov-modulated gamma-based framework," *IEEE/ACM Transactions on Networking*, vol. 11, no. 4, pp. 638–649, 2003.

[6] D. Liu, E. Sara, and W. Sun, "Nested auto-regressive processes for MPEG-encoded video traffic modeling," *IEEE Trans. Circuits and Systems for Video Techn.*, vol. 11, no. 2, pp. 169–183, Feb. 2001.

[7] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems," in *Proceedings. 20th Conference on Local Computer Networks*, Oct. 1995, pp. 397–406.

[8] M. Mandjes, I. Saniee, and A. Stolyar, "Load characterization and anomaly detection for voice over IP traffic," *IEEE Transactions on Neural Networks*, vol. 16, no. 5, pp. 1019–1026, 2005.

[9] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[10] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking,*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[11] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *in Proc. ACM SIGCOMM*, 1994.

[12] A. Andersen and B. Nielsen, "An application of superpositions of two state markovian source to the modelling of self-similar behaviour," in *Proceedings of IEEE INFOCOM*, vol. 1, Apr. 1997, pp. 196–204.

[13] D. Reininger, D. Raychaudhuri, B. Melamed, B. Sengupta, and J. Hill, "Statistical multiplexing of VBR MPEG compressed video on ATM networks," in *Proceedings of IEEE INFOCOM*, 1993.

[14] F. Yegenoglu, B. Jabbari, and Y.-Q. Zhang, "Motion-classified autoregressive modeling of variable bit rate video," *IEEE Trans. Circuits and Systems for Video Techn.*, vol. 3, no. 1, pp. 42–53, Feb. 1993.

[15] S. Ma and C. Ji, "Modeling heterogeneous network traffic in wavelet domain," *IEEE/ACM Trans. Netw.*, vol. 9, no. 5, pp. 634–649, Oct. 2001.

[16] M. Ismail, I. Lambadaris, M. Devetsikiotis, and A. Kaye, "Modelling prioritized MPEG video using TES and a frame spreading strategy for transmission in ATM networks," in *Proceedings of IEEE INFOCOM*, vol. 2, Apr. 1995, pp. 762–770.

[17] K. Chandra and A. Reibman, "Modeling one- and two-layer variable bit rate video," *IEEE/ACM Trans. Netw.*, vol. 7, no. 3, pp. 398–413, Jun. 1999.

[18] J. Zhao and I. Ahmad, "Traffic model for layered video: An approach on," in *Markovian Arrival Process, Packet Video*. Company, 2003.

[19] D. Fiems, V. Inghelbrecht, B. Steyaert, and H. Bruneel, "Markovian characterisation of H.264/SVC scalable video," in *Proceedings of Int. Conf. on Analytical and Stochastic Modeling Techniques and Applications*. Heidelberg: Springer-Verlag, 2008, pp. 1–15.

[20] L. Rossi, J. Chakareski, P. Frossard, and S. Colonnese, "A non-stationary Hidden Markov Model of multiview video traffic," in *Proceedings of IEEE ICIP*, 2010, pp. 2921–2924.

[21] P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Communications Surveys and Tutorials, in print*, 2012, video traces available at http://trace.eas.asu.edu/.

[22] O. Gomes, C. Combes, and A. Dussauchoy, "Parameter estimation of the generalized gamma distribution," *Mathematics and Computers in Simulation*, vol. 79, no. 4, pp. 955–963, 2008.

[23] M. Reisslein, J. Lassetter, S. Ratnam, O. Lotfallah, F. Fitzek, and S. Panchanathan, "Traffic and quality characterization of scalable encoded video: a large-scale trace-based study, part 1: overview and definitions," in *Tech. Rep. Arizona State Univ.*, 2002.

[24] S. Srinivasan, J. Vahabzadeh-Hagh, and M. Reisslein, "The effects of priority levels and buffering on the statistical multiplexing of single-layer H.264/AVC and SVC encoded video streams," *IEEE Transactions on Broadcasting*, vol. 56, no. 3, pp. 281–287, sept. 2010.