TRAFFIC MODEL FOR LAYERED VIDEO: AN APPROACH ON MARKOVIAN ARRIVAL PROCESS

Ji-An Zhao[1], *Bo Li*[1*], *Ishfaq Ahmad*[2]

[1]Department of Computer Science
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon, Hong Kong, China
Email: {zhaojian,bli}@cs.ust.hk

[2]Department of Computer Science & Engineering
University of Texas at Arlington
Arlington, TX 76019-0015, USA
Email: iahmad@cse.uta.edu

ABSTRACT

Video transmission usually has stringent requirements on the network bandwidth, packet loss rate and the experienced delay. To better study the impact of layered video traffic on the network performance, accurate and tractable traffic model for layered video source is important. In this paper we propose a traffic model for scalable video encoded in multiple layers. The model is based on Markovian arrival process with marked transitions. The states of the Markovian arrival process are derived from the correlation feature found in the video data. The base layer and enhancement layer video frame size pairs are classified by cluster detection algorithm. Each cluster corresponds to one state of the underlying Markov chain of the video traffic arrival process. The joint base and enhancement layer video frame size distribution for each state of the Markov chain is approximated by multivariate normal distribution. Simulation study shows that the proposed traffic model can predict the network performance with high accuracy.

1. INTRODUCTION

Due to the importance of video application, video traffic modeling had raised great research attention. Video data is usually encoded in variable bit rate (VBR), or nearly in constant bit rate (CBR) if a proper rate control algorithm is applied by the video encoder. One development in video technology is the emergence of layered video encoding (also called scalable encoding), which has found way in multimedia applications in heterogeneous network environment with diverse bandwidth and loss behavior. Generally there are three types of scalability, i.e., temporal scalability, spatial scalability and SNR scalability. In all the cases, the base layer pictures are encoded based on sub-sampling either with less frame rate (for temporal scalability), or with smaller picture size (for spatial scalability), or with coarser picture quality (for SNR scalability). Combining both layers can generate full quality video. Recently, a new video coding framework called fine grained scalability (FGS) is proposed for MPEG-4 [23]. The FGS base layer can be compressed using any motion compensation video encoding methods, often with a proper rate control. The enhancement layer data is encoded in a fine-granular way by adopting the bit-plane DCT-based FGS coding method [16]. This offers the flexibility to arbitrarily truncate the enhancement layer bit stream when the network bandwidth is variable, and makes the FGS framework well suited for Internet video streaming.

Video transmission usually has stringent requirements on the network bandwidth, the packet loss rate and the experienced delay. A key purpose of video source modeling is to employ the traffic model to predict the network performance by queueing analysis or computer simulation. To better study the impact of layered video traffic on the network performance, accurate and tractable traffic model for layered video source is desirable. In the literature there are plenty of statistical source models for VBR compressed video. A survey of the video source models can be found in [14]. Most video traffic models can be categorized into three classes, i.e., Markov process based, transform expand sample (TES) process based, and self-similar process based. TES based models [20] have the advantage to closely fit both the marginal distribution and the autocorrelation function of the empirical data. Thus they had been used to model the video traffic with a high accuracy in [21, 22]. It had been shown in [2] that long range dependence (LRD) is an inherent feature of the VBR video data, and that the LRD feature can lead to challenge in network traffic engineering. Source models with LRD property for VBR video were developed based on self-similar process in [7, 13]. TES and self-similar based models, however, are not tractable for queueing analysis, and the involved computation cost is

---

*Please direct all correspondence to Bo Li, Computer Science Department, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, CHINA. Fax: +852 2358 1477. E-mail: BLI@ust.hk. His research was supported in part by a grant from Research Grant Council (RGC) under contracts AoE/E-01/99.

rather high. Studies in [12] showed that although the LRD property of data traffic may have a strong negative impact on the network performance, the effects are significant only if the LRD causes the busy period of the network channel to be long enough such that the long lag traffic can accumulate and come into play. Real-time VBR video often has stringent QoS requirements on delay and loss. So sufficient network bandwidth must be allocated for video traffic and the buffer size need to be limited. Thus the traffic intensity and the busy period can not be very high. In such situation, the short range dependence property of video data is more important for predicting the network performance. This means that for real-time VBR video, traditional Markov based models are still valuable. For example, Markov chains were applied to model video conference traffic in [10, 19]. In addition, Markov chain had been widely used as a basic building block for more complex video traffic models, either in modeling scene change as in [22, 26], or in modeling group-of-picture (GOP) pattern as in [17].

Although in the literature there are large number of traffic models for VBR video, few is for layered video. A model for one and two layer video was proposed in [5]. In the model, the base layer was assumed to be coded in CBR and the enhancement layer is VBR. The model was targeted for video sequence with no (or few) scene change and encoded in IPPPPP... pattern without any fixed GOP, i.e., the first video frame is encoded in I-frame, and all the subsequent frames are encoded in P-frames. The correlation between successive P-frames was studied by cluster analysis on the P-frame bit rate pairs $\langle R(t), R(t+1)\rangle$, where $R(t)$ represents the bit rate for the t-th P-frame. Based on the clustering result, a finite state Markov chain was constructed. For each state of the Markov chain, an auto-regressive process was estimated. The model was validated by matching statistical features and examining the loss behavior in a leaky bucket policing function. The model, however, can only handle the case when the base layer is encoded in CBR, or in VBR with perfect rate control. For scalable video when both the base layer and enhancement layer are encoded in VBR, a general traffic model is still needed, which is the motivation of this paper.

In this paper we propose a traffic model for video source with no (or very few) scene change and scalably encoded in multiple layers[1]. The model is based on Markovian arrival process with marked transitions. The state of the Markovian arrival process is derived from the correlation feature found from the video data. The video frame size distribution in each state of the underlying Markov chain is modeled by a multivariate normal distribution[2] in order to capture the dependence between the successive video frames within the same video layer, as well as for the dependence across the base layer and the enhancement layer. Simulation study on the proposed model and the video trace is carried out. The results show that the proposed traffic model can predict the network performance with high accuracy.

2. TRAFFIC MODEL BASED ON MARKOVIAN ARRIVAL PROCESS

Due to the versatility of the Markovian arrival processes (MAPs) [18], they are widely used in modeling computer data traffic [4]. MAP was used to model the performance of superposition of VBR video sources in [3]. In recent time, MAP also found application in modeling aggregated Internet traffic [15]. It has been recognized that a compressed video stream has high peak-to-mean ratio, and video traffic is highly correlated in nature [6]. Modern video system, e.g., MPEG-4 video [1], is defined in the form of multiple video objects (VOs), which can be independently encoded, multiplexed and transmitted. This provides the potential for content-based interactivity, and also results in high efficient encoding, which makes it particularly appealing in running over low bit rate and wireless networks. Further, the video contents are usually encoded in more than one layers, with different significance in affecting the final decoded video quality. This naturally leads to the representation of such video traffic by using a MAP with marked transitions [9], in a way that different type of transitions correspond to different type of traffic arrivals. In this section, we develop general traffic model for VBR video with layered encoding. The model is based on the discrete time batch Markovian arrival process (DBMAP) with marked transition. For simplicity, also due to the lack of scalable video trace coded in more than 2 layers, we consider a video source with *two* layers in this paper. The model, however, is generally applicable to any hierarchically encoded video streams.

2.1. DBMAP with Two Types of Arrivals

We first briefly introduce the DBMAP with marked transitions. The process is first defined in [27], and is called the marked DBMAP process hereafter. We consider an n-state DBMAP with two types of arrivals, class-1 and class-2 traffic, respectively. Let the maximum batch size[3] for a class-1 traffic arrival to be b_1, and the maximum batch size for a class-2 traffic arrival to be b_2. The correspondent parameter matrices for the arrival process are given by $\{D_{00}, D_{01}, ... D_{b_1 b_2}\}$, each $D_{i_1 i_2}$ is an $n \times n$ matrix. Suppose that at time $t, t \geq 0$, the underlying Markov chain of the DBMAP pro-

[1] In our modeling example we analyzed video traces coded in 2 layers. One reason is that we lack scalable video trace coded in more than 2 layers. The modeling approach in this paper, however, is general and applicable for n-layer ($n > 2$) case.

[2] Since we only analyze the 2-layer case, we actually estimate the joint video frame size distribution by bivariate normal distribution.

[3] In real applications, the maximum arrival batch size at an instant for a stochastic arrival process is usually limited.

cess is in state $j, 1 \leq j \leq n$. Then at time epoch $t+1$, with conditional probability $D_{i_1 i_2}(j, j')$[4], where $0 \leq i_1 \leq b_1$ and $0 \leq i_2 \leq b_2$, the process transits to state $j', 1 \leq j' \leq n$, which is triggered by an arrival from class-1 traffic with batch size of i_1, and an arrival from class-2 traffic with batch size of i_2, simultaneously. Note that i_1 and i_2 might be 0.

Let the transition probability matrix of the underlying Markov chain for the arrival process to be D (also an $n \times n$ matrix), and we have $D = \sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} D_{i_1 i_2}$, i.e., an element $D(j, j')$ in the matrix D is given by

$$D(j, j') = \sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} D_{i_1 i_2}(j, j').$$

Now let us focus on the arrival process in a state (j), it may have arbitrary number of arrivals from the two traffic classes, and results a transition to another state (j'). This is given by the conditional probability $D_{i_1 i_2}(j, j')$, therefore we have

$$\sum_{i_1=1}^{b_1} \sum_{i_2=1}^{b_2} \sum_{j'=1}^{n} D_{i_1 i_2}(j, j') = 1.$$

In another word, when the arrival process is in the state (j), the next arrival is determined by this probability $D_{i_1 i_2}(j, j')$[5]. Notice that the above essentially implies $De = e$, in which e is a column vector with all elements being 1s.

We assume the arrival process is in stationary state and the initial probability vector is $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]$, which satisfies

$$\begin{aligned} \alpha e &= 1, \\ \alpha D &= \alpha. \end{aligned}$$

Therefore, we can derive the average arrival rate for each traffic class, λ_1 and λ_2. The arrival rate of class-1 traffic λ_1 is given by

$$\lambda_1 = \alpha(\sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} i_1 D_{i_1 i_2})e,$$

the class-2 arrival rate λ_2 is given by

$$\lambda_2 = \alpha(\sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} i_2 D_{i_1 i_2})e,$$

and the total traffic arrival rate is given by

$$\lambda = \lambda_1 + \lambda_2.$$

[4]We use the notation $D_{i_1 i_2}(j, j')$ to represent an element in the matrix $D_{i_1 i_2}$.

[5]More precisely, this should be the arrivals from each of the two traffic classes, with batch size of i_1 and i_2, and state transition from (j) to (j'). There are total $n(1+b_1)(1+b_2)$ such probabilities.

2.2. The Layered Video Source

Our aim is to analyze and model video source with no (or very few) scene change. We study 3 short video sequences, *Foreman*, *Grandma* and *Paris*, as well as 2 long video sequences, *Silence of the Lambs* and *Terminator One*. *Foreman* contains 400 frames with picture size of 176×144 in pixels (QCIF), *Grandma* contains 870 frames with picture size of 176×144 in pixels (also QCIF), and *Paris* contains 1000 frames with picture size of 352×288 in pixels (CIF). *Foreman* contains one time of scene change and the video pictures have a comparatively large degree of movement, while *Grandma* and *Paris* contain no scene change. *Paris* mainly consists of pictures with slow motion, while pictures in *Grandma* only have very few motion. We encode the video sequences with a layered video encoder supporting fine granular scalability (FGS) [23]. Since scene change can usually introduce significant variation of video content, encoding the immediate frame after a scene change in P-frame can lead to poor performance. We assume the I-frames are triggered by scene change resulting in an arbitrary number of P-frames following, as in [5]. Thus all the three video sequences are encoded in IPPPPP... pattern without any fixed GOP. The base layer is encoded with TM5 rate control, while the enhancement layer is encoded in MPEG-4 FGS. In the analysis we only consider the first sublayer of the enhancement layer video data. Our aim is to propose a tractable and accurate traffic model for video sequence without or with very few scene change. We expect the traffic model developed in this paper can be used as a basic building block to model long video sequence with scene changes. For example, by incorporating the scene detection and scene modeling methods in [11, 22], longer video sequence can be modeled.

2.3. The Layered Video Traffic Model

In this section we introduce the Markov based traffic model for layered video. We model the layered video data in the following four steps.

1. In the first step, we analysis the rate clustering feature of the video data. We view the encoded 2 layer video frame sequence as a vector time series along the frame index, i.e., $\langle R_b(t), R_e(t) \rangle, t = 1, 2, 3, \ldots$. Here $R_b(t)$ denotes the frame size of the t-th base layer video frame, and $R_e(t)$ denotes the frame size of the t-th enhancement layer video frame. We draw all the $\langle R_b(t), R_e(t) \rangle$ pairs as *points* on the 2-D plane, where $R_b(t)$ and $R_b(t)$ is viewed as the x-coordinate and the y-coordinate of the corresponding point, respectively. For example, for the first video frame in *Foreman*, the encoded base layer frame size is 11215 bits, and the encoded enhancement layer frame size is 3832 bits, we draw a point $[x = 11215, y = 3832]$

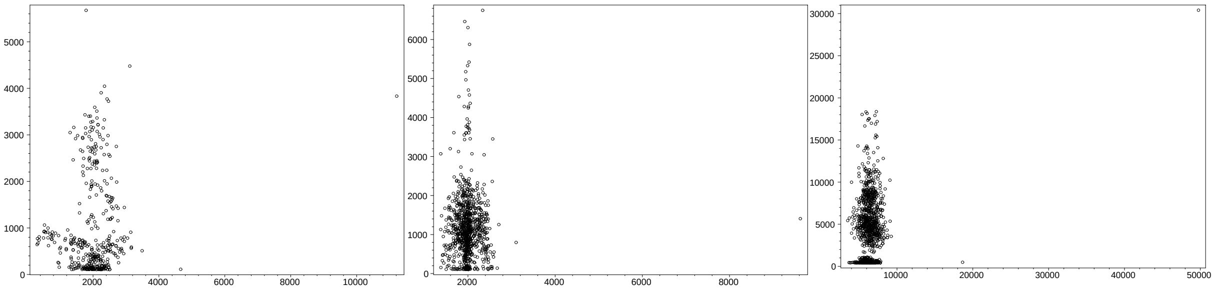


Fig. 1. Rate pairs as point set on 2-D plane
(a) Foreman IPPP (b) Grandma IPPP (c) Paris IPPP

on the two-dimensional plane. Since the raw video sequence has very few scene change, and the encoding pattern is IPPPPP..., we expect that the rate pairs $\langle R_b(t), R_e(t)\rangle$ and $\langle R_b(t+1), R_e(t+1)\rangle$ should be located very close from each other on the 2-D plane. Thus when all the pairs are drawn, the graph should have a clustering feature. This is clearly depicted in Figure 1. We see from the figure that all the three video sequences *Foreman*, *Grandma* and *Paris* show a strong clustering characteristics.

2. In the second step, for each video sequence, we apply cluster analysis on the corresponding rate pair point set. We take the hierarchical clustering approach [8]. Basically this works as follows: for a set of n points, we start with n clusters, each containing a single point; we then recursively aggregate the nearest two clusters into one; while combining cluster i and cluster j, we apply the complete linkage clustering algorithm, which means the distance between the two clusters is defined to be the greatest distance between an arbitrary member of cluster i and an arbitrary member of cluster j. The above method produces tight clusters with similar characteristics, and the shape of the cluster tends to be circular form. It is well known that there is no general criteria to determine the optimal number of resulting clusters. Thus we take a heuristic approach to stop the above aggregation process. The results of the above cluster analysis are shown in Figure 2. We obtained 12, 9 and 8 clusters for *Foreman*, *Grandma* and *Paris*, respectively.

3. In the third step, we construct a Markov chain based on the above clustering analysis results. For each clustering result, we view one cluster as one state for the Markov chain. Thus the Markov chain for *Foreman*, *Grandma* and *Paris* contains 12, 9 and 8 states, respectively. We estimate the transition probability matrix $P = [p_{ij}]$ for each Markov chain in the following way:

$$p_{ij} = \frac{\text{number of jumps from state } i \text{ to state } j}{\text{number of jumps out of } i}.$$

In this way we obtained the transition probability matrix T_F, T_G and T_P for *Foreman*, *Grandma* and *Paris*, respectively, as shown in Figure 3. In our calculation, the p_{ij} has a precision near 10^{-3}.

4. In the fourth step, for each state of the Markov chain, we estimate the joint $\langle base, enhance\rangle$ 2-layer frame size distribution. Since we applied the complete linkage clustering algorithm in the above step two, the resulting clusters have compact and nearly circular shape. This naturally lead to approximate the joint frame size distribution in each state of the Markov chain by bivariate normal distribution. Of course, this does not preclude more general or accurate estimation of the rate distributions using other advanced methods [25].

By the above four steps, we obtain a video traffic model based on a *Markov modulated process with correlated batch arrivals*. In the model the arrival process evolves according to the underlying Markov chain. In each state of the Markov chain, the base and enhancement layer data arrival rate follows the corresponding two-dimensional normal distribution. The two-dimensional normal distribution has 6 parameters, i.e., the mean of the base layer rate, the mean of the enhancement layer rate, the covariance matrix, which is a 2×2 matrix containing 4 parameters. All the 6 parameters for each state of the Markov chain can be easily estimated from the video trace and the clustering results. For the case of an n state Markov chain, the whole traffic model contains $6n$ parameters for the rate distributions, and one $n\times n$ transition probability matrix for the underlying Markov chain.

Clearly, the above traffic model belongs to a marked DBMAP process introduced in Section 2.1. Denote the transition probability matrix of the underlying Markov chain as T. Suppose that we divide the range of data rate for an arbitrary state of the Markov chain into discrete levels, with

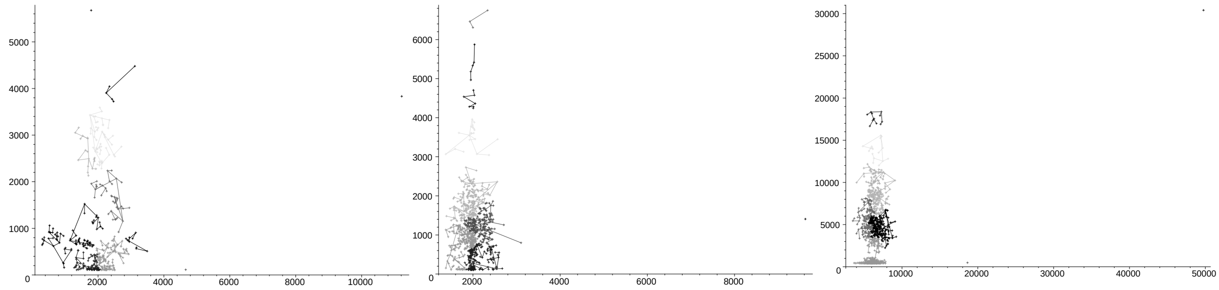


Fig. 2. Clustering results for rate point sets
(a) Foreman IPPP (b) Grandma IPPP (c) Paris IPPP

$$T_G = \begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.38 & 0.33 & 0.02 & 0.0 & 0.27 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.30 & 0.43 & 0.24 & 0.005 & 0.02 & 0.005 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.32 & 0.58 & 0.01 & 0.06 & 0.0 & 0.01 & 0.02 \\ 0.0 & 0.0 & 0.05 & 0.05 & 0.85 & 0.05 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.31 & 0.02 & 0.05 & 0.0 & 0.62 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.67 & 0.33 & 0.0 \\ 0.0 & 0.0 & 0.20 & 0.40 & 0.0 & 0.0 & 0.0 & 0.40 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.29 & 0.14 & 0.0 & 0.0 & 0.14 & 0.43 \end{bmatrix}, T_P = \begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.78 & 0.005 & 0.03 & 0.125 & 0.016 & 0.046 \\ 0.0 & 0.0 & 0.17 & 0.83 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.59 & 0.0 & 0.41 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.20 & 0.0 & 0.0 & 0.66 & 0.12 & 0.02 \\ 0.0 & 0.0 & 0.035 & 0.0 & 0.0 & 0.148 & 0.127 & 0.690 \\ 0.0 & 0.0 & 0.085 & 0.0 & 0.0 & 0.020 & 0.375 & 0.520 \end{bmatrix}.$$

$$T_F = \begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.67 & 0.23 & 0.01 & 0.01 & 0.0 & 0.03 & 0.0 & 0.0 & 0.03 & 0.02 \\ 0.0 & 0.0 & 0.23 & 0.52 & 0.16 & 0.02 & 0.0 & 0.01 & 0.04 & 0.02 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.12 & 0.18 & 0.44 & 0.22 & 0.04 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.17 & 0.0 & 0.50 & 0.11 & 0.22 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.02 & 0.03 & 0.02 & 0.05 & 0.82 & 0.05 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.02 & 0.0 & 0.0 & 0.0 & 0.05 & 0.84 & 0.09 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.19 & 0.0 & 0.0 & 0.0 & 0.03 & 0.75 & 0.03 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.40 & 0.40 & 0.0 & 0.0 & 0.0 & 0.0 & 0.20 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.44 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.56 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \end{bmatrix},$$

Fig. 3. Transition probability matrix for *Grandma* (T_G), *Paris* (T_P) and *Foreman* (T_F)

a level size of b^6. Let $f_i(u,v)$ be the two-dimensional normal density function in state $i, 1 \le i \le n$ of the underlying Markov chain, and let

$$F_i(x,y) = \int_0^{bx} \int_0^{by} f_i(u,v)dudv,\ 1 \le i \le n$$

be the two-dimensional normal distribution function[7] for state i, the corresponding parameter matrices for the marked DBMAP

$$\{D, D_{00}, \ldots D_{i_1 i_2}, \ldots, D_{b_1 b_2}\}$$

can be derived as follows:

$$\begin{cases} D &= T, \\ D_{i_1 i_2} &= [(F_i(i_1,i_2) - F_i(i_1-1, i_2-1))D[i,j]], \\ & \quad \text{for } i_1 > 0,\ i_2 > 0,\ 1 \le i \le n,\ 1 \le j \le n, \\ D_{00} &= D - \sum_{i_1>0} \sum_{i_2>0} D_{i_1 i_2}, \end{cases}$$

[6]Different states may have different b values.

[7]In fact this is a *scaled* distribution function since the actual distribution function is $F_i(x,y) = \int_0^x \int_0^y f_i(u,v)dudv$.

where $D[i,j]$ is the (i,j)-th entry of matrix D. Subsequently, $[(F_i(i_1,i_2) - F_i(i_1-1,i_2-1))D[i,j]]$ is an $n \times n$ sub-stochastic matrix.

3. SIMULATION RESULTS AND DISCUSSIONS

In this section we make simulation study on the traffic model obtained in Section 2. We validate the traffic model in two aspects: 1) we generate sample data from the traffic model, and compare the statistics between the generated video traffic and the video trace; 2) we inject the video trace data and the generated video data into the network and compare the queueing behavior of the network transmission buffer. The comparison of average video statistics, including the mean and the standard deviation to average ratio (SDA), for the trace data and generated video data samples are shown in Table 1. It can be seen from the table that the two set of data share very similar statistics. In particular, the frame size density functions for the model data and the trace data are well matched for all the three video sequences. This is

Video sequence	Video length	Base layer frame size	Enhancement layer frame size
Foreman **trace**	400 frames	X=248.78 (bytes) SDA=0.3635	X=126.98 (bytes) SDA=1.0398
Foreman **model**	400 frames	X=245.70 (bytes) SDA=0.3739	X=128.42 (bytes) SDA=1.0382
Grandma **trace**	870 frames	X=1997.93 (bits) SDA=0.1653	X=1210.64 (bits) SDA=1.1474
Grandma **model**	870 frames	X=1991.11 (bits) SDA=0.1209	X=1190.96 (bits) SDA=1.0168
Paris **trace**	1000 frames	X=798.56 (bytes) SDA=0.1562	X=592.20 (bytes) SDA=0.8664
Paris **model**	1000 frames	X=792.72 (bytes) SDA=0.1348	X=615.44 (bytes) SDA=0.9024

Note: X=mean, SDA=standard deviation to average ratio.

Table 1. Comparison of average video statistics for trace and model

shown in Figure 4 for base layer traffic, and in Figure 5 for enhancement layer traffic. We also compute the autocorrelation functions (acf) for the three video sequences, and find that the base layer and enhancement layer acf's for the trace data and the model data are closely fitted for all the three cases. We also find that the correlation after lag 100 is rather small and thus the long range dependence feature for the three video sequences is not very strong. Due to the space limitation, we can not show the acf curves in this paper.

We next take a simple approach to compare the network performance for the trace data and the model data. We inject data generated from the traffic model and data recorded in the video trace file into the network, respectively. We divide the network bandwidth into two part, in proportion to the mean traffic arrival rates for the base and enhancement layer video data. We assume the network transmission buffer has infinite size. We then simulate the queueing behavior of the network transmission buffer and compare the cumulative queue length distribution functions (cdf's) for inputs from the model data and the trace data. Since the encoder has rate control in the base layer, the base layer video traffic has a relatively small variation. Thus the base layer data buffer only experience occasional queueing and the buffer remains idle with a high probability. Therefore, we only show the enhancement layer queue length cdf for the three video sequences, as in Figure 6. From the figure, it can be seen that for the enhancement layer traffic, the whole queue length distributions for the model data and the trace data are closely matched, especially in the range when the queue size is large. The results demonstrate that, in terms of network queueing impact, the traffic model can emulate the video trace data with a relatively high accuracy.

The traffic model developed in this paper is general. First, although the model is targeted for video sequence with no or very few scene change, for long video sequence with lots of scene changes, a hierarchical model, based on scene detection and modeling technique [11, 22], can be built based on this model. Second, data clustering is an inherent feature for video traffic. This is quite intuitive, since if there is no scene change, content of the successive video pictures are very similar, and thus the encoded data rates should have small variation. We study two long video traces, *Silence of the Lambs* and *Terminator One*, which are publically available from [24]. The above two video sequences are encoded in 2 layers with spatial scalability, without rate control in either the base or the enhancement layer. We display the first 1600 $\langle base, enhance\rangle$ frame size pairs for *Silence of the Lambs* on the 2-D plane, as shown in Figure 7(a), and the first 2200 frame size pairs for *Terminator One*, as shown in Figure 7(b). From the figure it is obvious the rate pairs have strong clustering feature for both cases. Third, note that for the definition of the marked DBMAP process in Section 2.1, it can be extended to support $k, k \geq 2$ classes of traffic by extending the $D_{i_1 i_2}$ parameter matrix to $D_{i_1 i_2 \dots i_k}$ with a k-dimensional index. This means mathematically the model developed in this paper can be applied to video data encoded in $k, k \geq 2$ lay-

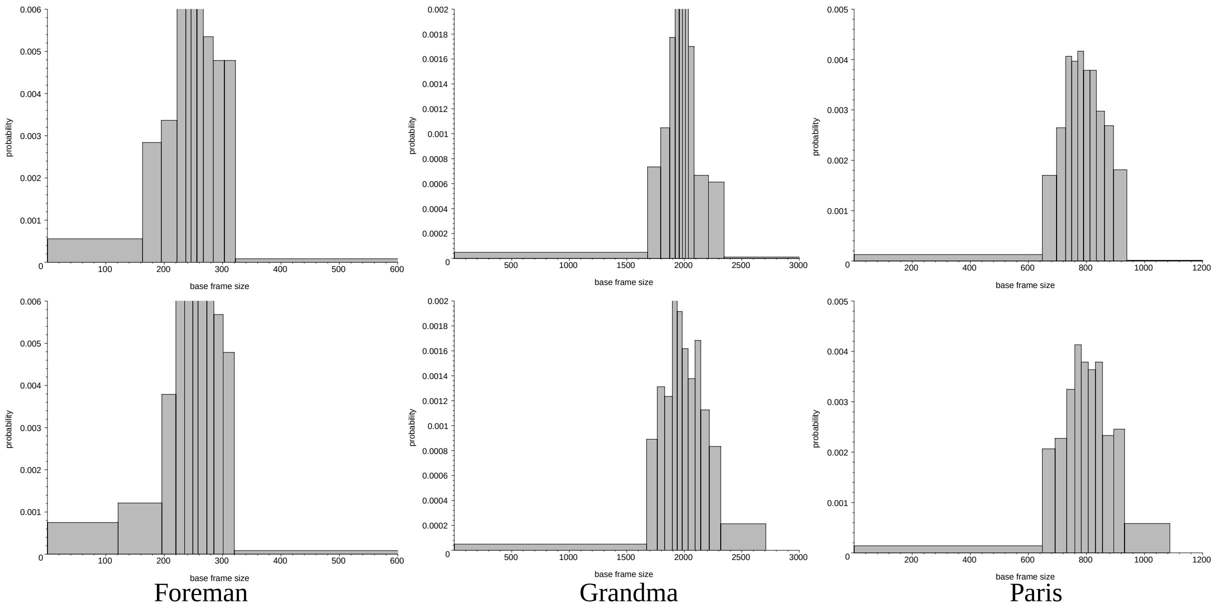


Fig. 4. Base layer frame size pdf comparison: trace (above) vs model (below)

ers. For example, if the video is encoded in 3 layers, then a data tuple $\langle base\_rate, enhance\_rate_1, enhance\_rate_2\rangle$ would appear to be a three-dimensional point in the 3-D space. Subsequently, 3-D cluster analysis algorithm can be applied and the underlying Markov chain can be estimated, and a traffic model can be developed in a similar approach as in Section 2.3.

In scalably encoded video data there exists a cross layer dependence between the base and the enhancement layer(s). This is because the enhancement layer data are usually coded by prediction from the base layer data. A large base layer frame size often indicates that the video content is complex, and the enhancement layer will need more bits to code the residual video signals, thus the enhancement layer frame size will also tend to be large. To demonstrate this, we study the long video sequence *Terminator One*. We display the first 1400 base layer and enhancement layer video frame size in Figure 8(a) and Figure 8(b). From the figure we can discover that whenever the base layer data rate tend to be high, the enhancement layer data rate also follows a similar trend. This indicates that there exists a strong correlation between the base layer and the enhancement layer video traffic. We also compute the cross correlation coefficients (ccf) for the video trace the result is shown in Figure 8(c). The ccf curve reveals that the cross layer correlation lasts for a long period in terms of frame lags. For example, the ccf value after 150 lags remains higher than 0.5. The existence of cross correlation suggests that in order to precisely model layered video data, we must study the data tuple $\langle base\_rate, enhance\_rate\rangle$ as a whole for each $\langle base, enhance\rangle$ frame pair, rather than to study the data statistics of individual layer separately. In this sense, the model proposed in [5] lacks the ability to capture the cross layer correlation, since the base layer is assumed to be CBR and independent of the enhancement layer. Thus the model is actually built on the enhancement layer data statistics only. The model proposed in this paper, however, can grasp the inter-layer dependence by properly estimating the joint $\langle base, enhance\rangle$ data rate pdf.

4. CONCLUSIONS

In this paper we proposed a traffic model for layered video data with no or few scene change. The model is based on Markovian arrival process with marked transitions. The state of the Markovian arrival process is derived from the correlation feature found from the video data. The base layer and enhancement layer video frame size pairs are analyzed and grouped into clusters. Each cluster corresponds to one state of the underlying Markov chain of the arrival process. The video frame size distribution for each state of the Markov chain is modeled by two-dimensional normal distribution. Simulation study on both the proposed model and the video trace is carried out. The results show that the proposed traffic model can predict the network performance with good accuracy. One important property of the traffic model is that it is tractable for queueing analysis in studying the performance of video transmission over the network. Our previous work in [27] showed that video transmission over wireless network can be modeled by a $DBMAP/PH/1$ priority queue, which have been solved in [28] by adopting matrix analytic methods. We are inter-

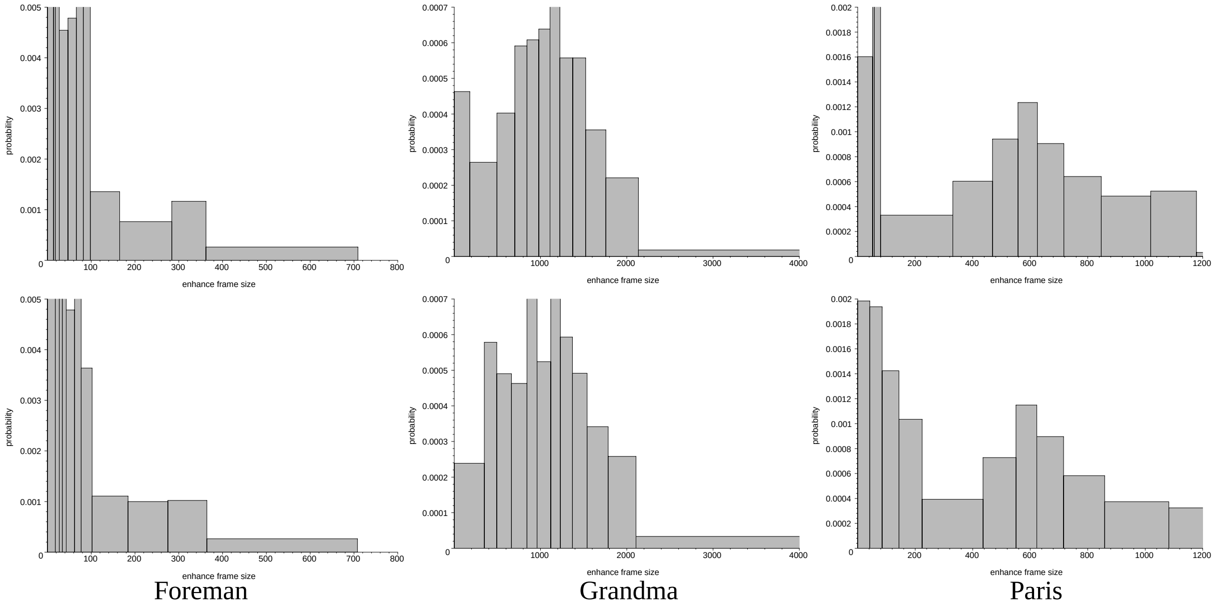


Fig. 5. Enhancement layer frame size pdf comparison: trace (above) vs model (below)

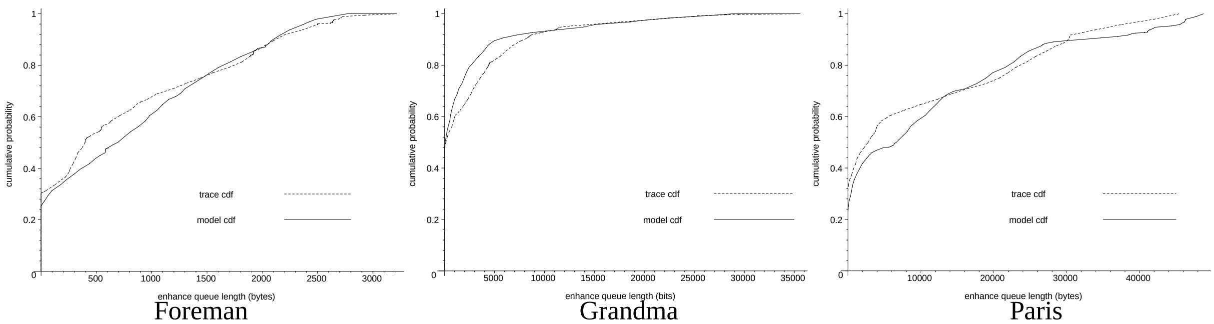


Fig. 6. Queue length cdf for enhance layer data: trace vs model

ested in performance evaluation of various scheduling algorithms for layered video data transmission, and in extending the proposed traffic model for long video sequence with numerous scene changes, as well as for layered video encoded with fixed GOP pattern.

5. REFERENCES

[1] R. Koenen, *MPEG-4 Overview* - (V.18 - Singapore Version), ISO/IEC JTC1/SC29/WG11 N4030, March 2001.

[2] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Transactions on Communications*, vol. 43, pp. 1566-1579, 1995.

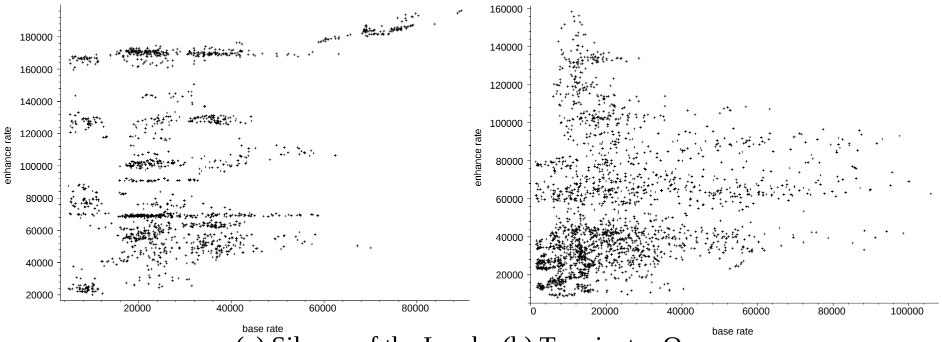
[3] C. Blondia and O. Casals, "Statistical multiplexing of VBR sources: a matrix-analytic approach," *Performance Evaluation*, vol. 16, pp. 5-20, 1992.

[4] C. Blondia, "A discrete time batch Markovian arrival process as B-ISDN traffic model," *Belgian Journal of Operations Research, Statistics and Computer Science*, vol. 32, no. 3-4, 1993.

[5] K. Chandra and A. R. Reibman, "Modeling one- and two-layer variable bit rate video," *IEEE/ACM Transactions on Networking*, vol. 7, no. 3, pp. 398-413, June 1999.

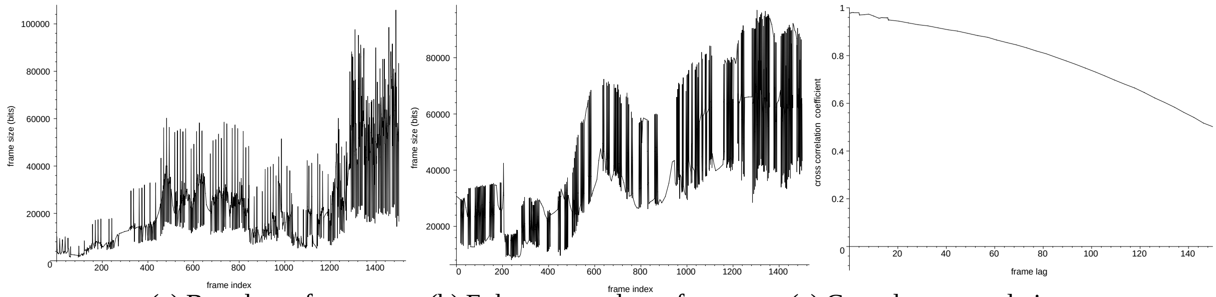
[6] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40-54, Nov/Dec 2001.

[7] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," *Pro-*



(a) Silence of the Lambs (b) Terminator One

Fig. 7. Clustering feature for layered video data



(a) Base layer frame rates (b) Enhancement layer frame rate (c) Cross layer correlation

Fig. 8. Terminator One coded in 2 layers

ceedings ACM SIGCOMM 94, pp. 269-280, London, UK, August 1994.

[8] J. A. Hartigan, *Clustering algorithms*, Wiley, New York, 1975.

[9] Q.-M. He and M. F. Neuts, “Markov chains with marked transitions,” *Stochastic Processes and Their Applications*, vol. 74, no. 1, pp. 37-52, 1998.

[10] D. P. Heyman, A. Tabatabai and T. V. Lakshman, “Statistical analysis and simulation study of video teleconference traffic in ATM networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, no. 1, pp. 49-59, 1992.

[11] D. P. Heyman and T. V. Lakshman, “Source models for VBR broadcast-video traffic,” *IEEE/ACM Transactions on Networking*, vol. 4, no. 1, pp. 40-48, Feb. 1996.

[12] D. P. Heyman and T. V. Lakshman, “What are the implications of long-range dependence for VBR-video traffic engineering?”, *IEEE/ACM Transactions on Networking*, vol. 4, pp. 301-317, June 1996.

[13] C. Huang, M. Devetsikiotis, L. Lambadaris and A. Kaye, “Modeling and simulation of self-similar variable bit rate compressed video: A unified approach,” *Proc. of ACM SIGCOMM*, Cambridge USA, August 1995.

[14] M. Izquierdo and D. Reeves, “A survey of statistical source models for variable-bit-rate compressed video,” *Multimedia Systems*, vol. 7, no. 3, pp. 199-213, 1999.

[15] A. Klemm, C. Lindemann and M. Lohmann, “Traffic modeling of IP networks using the batch Markovian arrival process,” *12th International Conference on Modeling Tools and Techniques for Computer and Communication System Performance Evaluation* (Tools 2002), London, April 2002.

[16] F. Ling, W. P. Li and H. Q. Sun, “Bit-plane coding of DCT coefficients for image and video compression,” *Proceedings of SPIE VCIP’99*, San Jose, Jan. 25-27, 1999.

[17] A. Lombardo, G. Morabito and G. Schembra, “An accurate and treatable Markov model of MPEG-video traffic,” *Proc. IEEE INFOCOM’98*, vol. 1, pp. 217-224, 1998.

[18] D. M. Lucantoni, K. S. Meier-Hellstern and M. F. Neuts, “A single-server queue with server vacations

and a class of non-renewal arrival processes," *Advances in Applied Probability*, vol. 22, pp. 676-705, 1990.

[19] D. M. Lucantoni, M. F. Neuts and A. R. Reibman, "Methods for performance evaluation of VBR video traffic models," *IEEE/ACM Transactions on Networking*, vol. 2, no. 2, April 1994.

[20] B. Melamed, "An overview of TES processes and modeling methodology," *Performance evaluation of computer and communication systems*, Springer-Verlag, pp. 359-393, 1993.

[21] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski, "TES-based video source modeling for performance evaluation of integrated networks," *IEEE Transactions on Communications*, vol. 42, no. 10, pp. 2773-2777, 1994.

[22] B. Melamed and D. Pendarakis, "Modeling full-length VBR video using Markov-renewal-modulated TES models," *IEEE Journal on Selected Areas in Communications*, vol. 16 no. 5, pp. 600-611, June 1998.

[23] H. M. Radha, M. V. D. Schaar and Y. W. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Transactions on Multimedia*, vol. 3, no. 1, pp. 53-68, March 2001.

[24] Martin Reisslein, "Video Traces for Network Performance Evaluation," publically available at http://peach.eas.asu.edu .

[25] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*, Wiley, New York, 1992.

[26] F. Yegenoglu, B. Jabbari and Y.-Q. Zhang, "Motion-classified autoregressive modeling of variable bit rate video," *IEEE Transactions on Circuit Systems for Video Technology*, vol. 3, no. 1, pp. 42-53, 1993.

[27] J.-A. Zhao, B. Li, C.-W. Kok and I. Ahmad, "Priority scheduling with ARQ control: performance model for packet video in wireless networks," *2002 International Symposium on Performance Evaluation of Computer and Telecommunication Systems*, SPECTS'2002, San Diego, USA, July 2002.

[28] J.-A. Zhao, B. Li, X.-R. Cao and I. Ahmad, "Matrix analytic solution for $DBMAP/PH/1$ priority queues," submitted to *Queueing Systems, Theory and Applications*.