TRAFFIC MODELING FOR LAYERED VIDEO

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 and Engineering
University of Texas at Arlington
Arlington, TX 76019-0015, USA
Email: iahmad@cse.uta.edu

ABSTRACT

In this paper we propose a traffic model for delivering scalable video encoded with multiple layers on heterogeneous networks. The model is based on Markovian arrival process with marked transitions. The state of the underlying Markov chain of the video arrival process is derived from the correlation feature found in the video data. The base layer and enhancement layer video frame size pairs are decided by a cluster detection algorithm; each cluster corresponds to one state of the Markov chain. 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 on the traffic model data and the video trace data is carried out and compared with the model. The results show that the proposed traffic model can predict the network performance with good accuracy.

1. INTRODUCTION

Layered video encoding is widely adopted for multimedia video applications in heterogeneous network environment with diverse bandwidth and loss behavior. A key purpose of video source modeling is to employ the traffic model to predict the network performance. To maintain an acceptable visual quality for real-time video, sufficient network bandwidth need be allocated to reduce the delay. In such situation, the short range dependence property of video data is more important to predict the network performance [4]. Thus, traditional Markov based models are widely used to model video data directly, or used as a building block to model more complex video traffic. A source model based on Markov modulated Auto-Regressive process for two layer video was proposed in [2]. The model, however, can not capture the cross layer correlation since the base layer is assumed to be CBR. For layered video, the base layer and the enhancement layer(s) video data are often correlated because of cross layer predictive coding; further, the base layer can be VBR. Thus traffic model taking the above factors into consideration is desired, which is the motivation of this

---

*Correspondent author.

paper. Specifically, we propose a traffic model for video source with no (or very few) scene change and scalably encoded in multiple layers. The model is based on the discrete time batch Markovian arrival process (DBMAP) [1] with marked transitions. Since the model can grasp the cross layer correlation, it can be more precise in predicting the network performance; further, since the model is Markov based, the involved queueing system is solvable.

2. THE LAYERED VIDEO TRAFFIC MODEL

Markovian arrival process (MAP) [5] is widely used in modeling heterogeneous data traffic due to the versatility. Scalable video data are organized in more than one layers, with different significance in affecting the decoded video quality. This naturally leads to the representation of such video traffic by using an MAP with marked transitions [3], in a way that different type of transitions correspond to different type of traffic arrivals.

2.1. DBMAP with Two Types of Arrivals

We first briefly introduce the DBMAP process with marked transitions. Consider an n-state DBMAP with two types of arrivals, which can be from class-1 or class-2 traffic. Let the maximum batch size for class-1 and class-2 traffic arrival to be b_1 and b_2, respectively. The marked DBMAP is defined by $\{D_{00}, D_{01}, ... D_{b_1 b_2}\}$, each $D_{i_1 i_2}$ is an $n \times n$ transition matrix. Suppose that at time $t, t \geq 0$, the underlying Markov chain of the DBMAP process is in state $j, 1 \leq j \leq n$. Then at time epoch $t + 1$, with conditional probability $D_{i_1 i_2}(j, j')$, $0 \leq i_1 \leq b_1$, $0 \leq i_2 \leq b_2$, 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. The transition probability matrix for the underlying Markov chain of the arrival process is given by $D = \sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} D_{i_1 i_2}$.

Assume the arrival process is in stationary state and the initial probability vector is $\pi = [\pi_1, \pi_2, ..., \pi_n]$, let e be a column vector with all elements being 1's, then we have
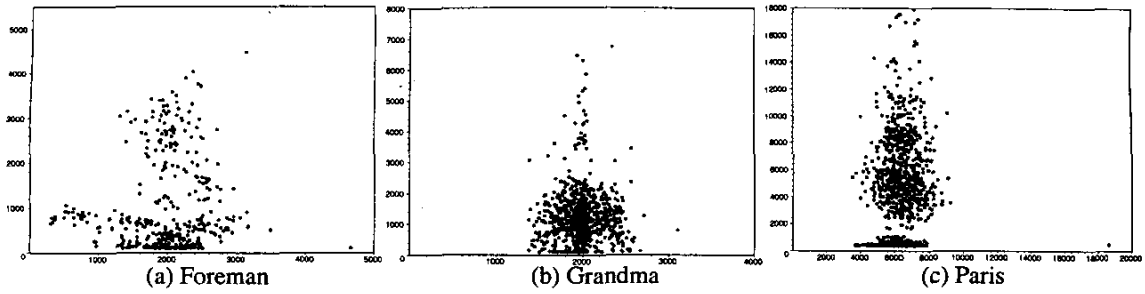

ICME 2003

(a) Foreman   (b) Grandma   (c) Paris

Fig. 1. Rate pairs as point set on 2-dimensional plane

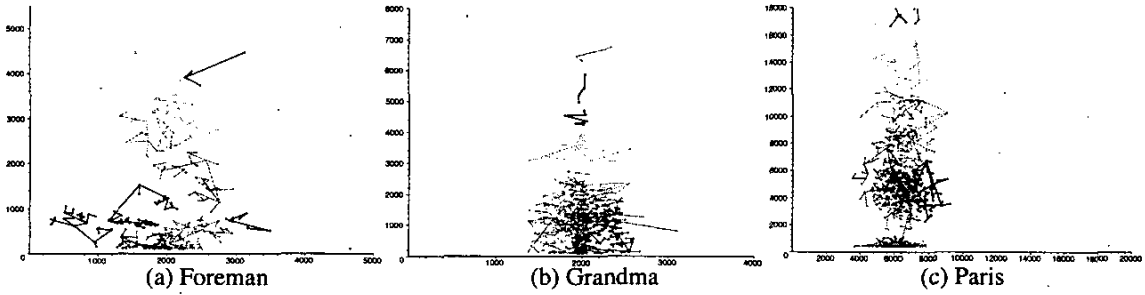(a) Foreman   (b) Grandma   (c) Paris

Fig. 2. Clustering results for rate point sets

$\pi e = 1$ and $\pi D = \pi$. Class-1 traffic arrival rate is given by $\lambda_1 = \pi(\sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} i_1 D_{i_1 i_2})e$, class-2 traffic arrival rate is given by $\lambda_2 = \pi(\sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} i_2 D_{i_1 i_2})e$, and the total traffic arrival rate is given by $\lambda = \lambda_1 + \lambda_2$.

2.2. Markov Based Video Source Model

We study 3 short video sequences, *Foreman, Grandma* and *Paris*. *Foreman* contains 400 frames and the picture size is 176 × 144 in pixels (QCIF). *Grandma* contains 870 frames and the picture size is also 176 × 144. *Paris* contains 1000 frames and the picture size is 352×288 (CIF). *Foreman* contains one time of scene change and the video pictures have a comparatively larger degree of movement, while *Grandma* and *Paris* contain no scene change, and the video mainly consist of pictures with slow motion.

We assume the I-frames are triggered by scene change resulting in an arbitrary number of P-frames following, as in [2]. Thus all the three video sequences are encoded in IPPPPP..: pattern without any fixed GOP, i.e., only the first video frame is encoded in I-frame, and all the subsequent frames are encoded in P-frames. We encode the video sequences in two layers. The base layer data is encoded with TM5 rate control, while the enhancement layer is encoded in MPEG-4 FGS [6]. In the analysis we only consider the first sublayer of the enhancement layer video data. Our aim is to develop a tractable and accurate traffic model for video

sequence without or with very few scene change. We expect that by incorporating scene detection and modeling methods, longer video sequence can be modeled. We model the layered video data in the following four steps.

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 $R_b(t)$ are viewed as the x and y coordinates of the corresponding point. 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 quite intuitive, since if there is no scene change, contents of the successive video pictures are very similar, and the encoded data rates should have small variation. This is clearly depicted in Fig. 1. We observe that all the three video sequence *Foreman, Grandma* and *Paris* show a strong clustering characteristics.

In the second step, for each video sequence, we do cluster analysis on the corresponding rate pair point set. We adopt the hierarchical cluster detection algorithm. Basically

$$\begin{bmatrix}
0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.67 & 0.23 & 0.01 & 0.01 & 0.0 & 0.03 & 0.0 & 0.0 & 0.03 & 0.02 \\
0.0 & 0.0 & 0.23 & 0.52 & 0.16 & 0.02 & 0.0 & 0.01 & 0.04 & 0.02 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.12 & 0.18 & 0.44 & 0.22 & 0.04 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.17 & 0.0 & 0.50 & 0.11 & 0.22 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.02 & 0.03 & 0.02 & 0.05 & 0.82 & 0.05 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.02 & 0.0 & 0.0 & 0.0 & 0.05 & 0.84 & 0.09 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.19 & 0.0 & 0.0 & 0.0 & 0.03 & 0.75 & 0.03 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.40 & 0.40 & 0.0 & 0.0 & 0.0 & 0.0 & 0.20 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.44 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.56 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0
\end{bmatrix}
\qquad
\begin{bmatrix}
0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.78 & 0.005 & 0.03 & 0.125 & 0.016 & 0.046 \\
0.0 & 0.0 & 0.17 & 0.83 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.59 & 0.0 & 0.41 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.20 & 0.0 & 0.0 & 0.66 & 0.12 & 0.02 \\
0.0 & 0.0 & 0.035 & 0.0 & 0.0 & 0.148 & 0.127 & 0.690 \\
0.0 & 0.0 & 0.085 & 0.0 & 0.0 & 0.020 & 0.375 & 0.520
\end{bmatrix}$$

Fig. 3. Transition probability matrix for Foreman (left) and Paris (right)

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. We apply the complete linkage clustering criterion, and take a heuristic approach to stop the aggregation process. The clustering results are shown in Fig. 2. We obtain 12 clusters for *Foreman*, 9 clusters for *Grandma*, and 8 clusters for *Paris*[1].

In the third step, we construct a Markov chain based on the above clustering results. We view each cluster as one state for the underlying Markov chain of the video arrival process. We estimate the transition probability matrix for each Markov chain as follows:

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

In this way we obtained the transition probability matrix for *Foreman* and *Paris* as shown in Fig. 3[2].

In the fourth step, for each state of the Markov chain, we estimate the joint $\langle base, enhance\rangle$ frame size distribution. Since we applied the complete linkage clustering algorithm, the resulting clusters have compact and nearly circular shape. This naturally leads to approximate the joint frame size distribution in each state of the Markov chain by bivariate normal distribution. More general or accurate estimation of the rate distributions can be applied by using other multivariate density estimation methods.

Thus 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 rates follow the corresponding two dimensional normal distribution. We next show that the above traffic model belongs to the 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 ar-

---

[1] In what follows we only present results for *Foreman* and *Paris* because of space limitation. For detailed results please contact with the corresponding author and refer to our technical report.

[2] In our calculation, the p_{ij} has a precision near 10^{-3}.

Video sequence	Video length	Base layer frame size	Enhance 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. Statistic comparison of trace data and model data

bitrary state i, $1 \le i \le n$, of the Markov chain into discrete levels, with a level size of b_i. Let $f_i(u, v)$ be the two dimensional normal density function for state i, and let $F_i(x, y) = \int_0^{b_i x} \int_0^{b_i y} f_i(u, v)dudv$, then the 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: $D_{i_1 i_2} = (F_i(i_1, i_2) - F_i(i_1 - 1, i_2 - 1))D$, for $i_1 > 0$, $i_2 > 0$, $D_{00} = D - \sum_{i_1>0}\sum_{i_2>0} D_{i_1 i_2}$, and $D = T$.

3. SIMULATION RESULTS

We validate the traffic model by simulation study. First, we generate sample data from the traffic model, and compare the statistics between the generated video traffic and the video trace. The results are shown in Table 1. From the table it can be seen that the statistics of the two data sets, including the mean and the standard deviation to average ratio (SDA) are comparatively close. In particular, the frame size probability density functions (pdf's) for the model data and the trace data are well matched for all the three video sequences. This is shown in Fig. 4 and Fig. 5.
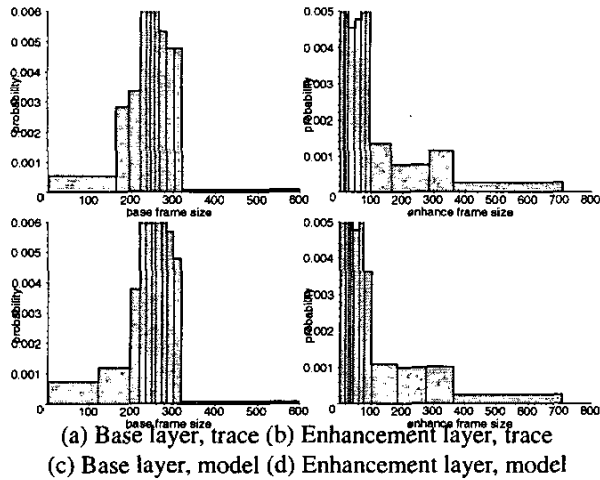
(a) Base layer, trace (b) Enhancement layer, trace
(c) Base layer, model (d) Enhancement layer, model

Fig. 4. Foreman frame size pdf: trace vs model

(a) Base layer, trace (b) Enhancement layer, trace
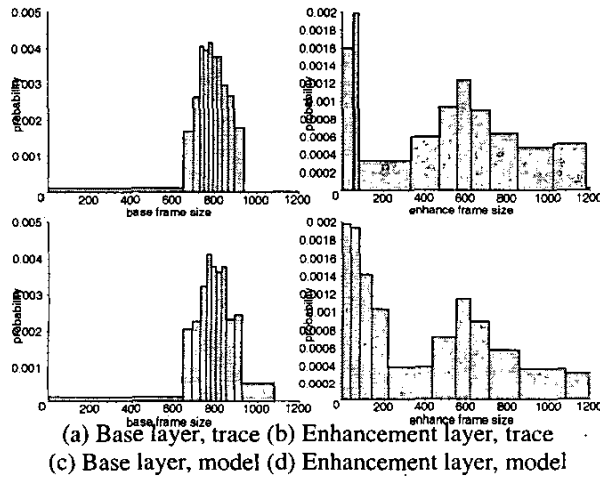(c) Base layer, model (d) Enhancement layer, model

Fig. 5. Paris frame size pdf: trace vs model

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, and compare the queueing behavior of the network transmission buffer. We divide the network bandwidth into two parts, in proportion to the mean arrival rate for the base and enhancement layer video data. We assume the network transmission buffer has infinite size. Since the base layer encoding is under rate control and the base layer traffic has a relatively small variation, we only show the enhancement layer queue length cumulative distribution function (cdf) in Fig. 6. We observe that the enhancement layer queue length cdf curves for the model data and the trace data are closely matched, especially in the range where the queue size is large. The
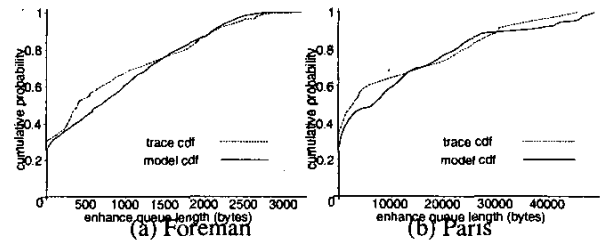

(a) Foreman (b) Paris

Fig. 6. Enhancement layer queue length cdf: trace vs model

results demonstrate that, in terms of network queueing impact, the traffic model can emulate the video trace data with a good accuracy.

4. CONCLUSIONS

In this paper we proposed a traffic model for layered video with no or few scene changes. Simulation study shows that the proposed traffic model can predict the network performance with good accuracy. In our modeling example we only analyzed video traces coded in 2 layers. The reason is that we lack scalable video trace coded in more than 2 layers. The modeling approach, however, is general and applicable for n-layer case. For example, if the video is encoded in 3 layers, a data tuple $\langle base, enhance_1, enhance_2 \rangle$ 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 2.2. We are interested in performance evaluation of different network scheduling algorithms for layered video data transmission, and in extending the proposed traffic model for long video sequence with scene changes, as well as for layered video encoded in fixed GOP pattern.

5. REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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," Adv. Appl. Prob., vol. 22, pp. 676-705, 1990.

[6] 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.