IEEE/ACM Trans. Networking v7, (3), p398-413, June 1999

Modeling One- and Two-Layer Variable Bit Rate Video

Kavitha Chandra [1] and *Amy R. Reibman*[2]

Center for *Advanced Computation* and *Telecommunications*[1]
Department of Electrical and Computer Engineering
University of Massachusetts Lowell
Lowell, MA 01854
Kavitha_Chandra@uml.edu

AT&T Labs − Research[2]
100 Schultz Drive
Red Bank, NJ 07701
amy@research.att.com

ABSTRACT

This paper presents a source model for variable bit rate (VBR) video traffic. A finite state Markov chain is shown to accurately model one- and two-layer video of all activity levels on a per source basis. Our model captures the source dynamics, including the short-term correlations essential for studying network performance. The modeling technique is shown to be applicable for both H.261 and MPEG2 encoded video of a variety of activity levels. The traffic model is shown in a simulation study to be able to accurately characterize both the single-source buffer occupancy over a wide range of buffer sizes and the multiplexing behavior. The VBR video model is also used to model the enhancement layer of two-layer SNR scalable video. We show that two-layer encoding has significantly better statistical multiplexing gains than one-layer video, particularly when the network admits calls based on a leaky-bucket characterization.

1. INTRODUCTION

Video communication services such as entertainment based video on demand, home shopping, videoconferencing, telemedicine and distance learning are emerging applications for broadband integrated services digital networks (B-ISDN). Analog video signals are digitized, compressed and encoded prior to transmission over communication networks. Current compression techniques for digital video can achieve video bit rates of acceptable quality in the range of 1-5 Mbps. In a network carrying voice, video and data, the video traffic component can consume the dominant portion of the available bandwidth. In order to evaluate the performance of B-ISDN networks, traffic models of proposed services are necessary. This work addresses the problem of developing a model for variable bit rate video traffic, which we then apply to both one- and two-layer video.

Video can be transported either with a constant bit-rate (CBR) or with a variable bit-rate (VBR). VBR video has several potential advantages over traditional CBR video: improved image quality and shorter delay. In addition, through statistical multiplexing, improved channel allocation may be obtained compared to CBR transport. However, statistical multiplexing invariably results in buffering delays and losses, which can significantly degrade video quality. To minimize the amount of delay and loss, the net-working community has focused on the development of effective and implementable congestion control schemes, including connection admission control (CAC) and usage parameter control (UPC). To minimizethe impact of delay and loss, the video community has focused on developing good error concealment algorithms and designing efficient two-layer coding algorithms [1,2] for use in combination with the dual-priority transport provided by ATM networks. For example, while one-layer MPEG-2 [3] produces generally unacceptable video quality with a cell loss ratio of 10 ?3 , losses at this rate with SNR scalability (one of the four standardized layered coding algorithms of MPEG-2) are generally invisible, even to experienced viewers [4].

To determine a set of techniques appropriate for CAC and UPC, traffic models that accurately represent the statistical nature of very high-speed bursty services are necessary. A variety of application and coding-specific models of one-layer VBR video have been proposed in the literature. Existing models can be categorized into models for

low to medium activity, such as in videophones and videoconferencing, and models for high activity as in entertainment video. Distinctions can also be found between models for single sources and those used the characterize the aggregation of traffic sources.

For low activity videophone signals, Maglaris et. al. [5] analyzed 10 second sequences and proposed a first order autoregressive process (AR) as a model for successive frame bits. The traffic stream resulting from multiplexed sources was also modeled by Maglaris [5] as a discrete finite state, birth-death Markov model. Longer sequences of moderate activity videoconference data were modeled by Heyman et. al. [6]. In this model a first order discrete state autoregressive model (DAR) was augmented with a Markov chain structure to model the transitions between states. Lucantoni et. al. [7] compared the DAR model to a Markov renewal process model (MRP). The MRP was shown to better capture the burstiness in the low-activity video data, although it was still inadequate in matching the tail probabilities at large buffer sizes. Further, the DAR model has been found to be more favorable for modeling multiplexed VBR teleconferencing traffic [8].

Relatively fewer analyses exist for full motion video. Garrett and Willinger [9] propose a self-similar process to model long-range dependence features found in the analysis of the Star Wars movie sequence. This model is very parsimonious in the number of parameters, but does not lend itself well to analytical studies. A source model for full motion video was presented by Yegenoglu et. al. [10], where an autoregressive process with time-varying coefficients adapted by a finite state Markov chain was constructed by constraining to match the histogram of the video frame bits. Gruenenfelder et. al. have also used an autoregressive moving average process for modeling VBR video [11]. In [12], Heyman and Lakshman propose a model for the scene changes that characterize broadcast video traffic, in which the DAR was shown to overestimate the cell loss rate for certain high-activity sequences.

Most of these previous studies on modeling one-layer VBR traffic have identified models by matching statistical features such as the autocorrelation function (acf) and/or the probability density function (pdf) of the video frames. Frater et. al. [13] additionally verify the performance of a non-Markovian model for full motion video based on scene characterization by matching the cell loss probabilities at different buffer sizes. Krunz and Hughes [14] modeled MPEG, with distinct models for each different frame type. In the present work, the goal is to understand how a traffic model can incorporate features and changes in the encoding process and correspondingly influence the network performance while simultaneously matching the statistical descriptors of the measured data. In particular, we show that these requirements can be met by a Markovian model.

As noted, earlier studies have used different modeling techniques based on the level of activity in the video, while the Markov chain model developed in this paper is general enough to accurately model video of varying activity levels, both for a single source and for large buffer sizes. We place particular emphasis on selecting the number of states required to model the dominant modes in the data trace accurately. Thus, instead of selecting the number of model states to match static descriptors such as the steady-state density function and its moments, we choose the number of states so as to capture the dominant eigenvalues characterizing the source dynamics. Adequate spectral content in single sources will be necessary to understand the scaling aspects of Markov models for superposed sources.

In addition to providing a more compact parametric representation than measurement traces, the queueing behavior of Markov chain based sources can be evaluated using fluid-buffer analysis [38], where the delay and loss probability distributions are explicit functions of the eigenvalues and eigenvectors of the Markov transition matrix. We place emphasis on deriving a traffic model that can offer a more fundamental understanding of encoder and source dynamics than can be obtained from trace-driven studies. The structure of the resulting Markov chain captures some aspects of the source video. For example, the near tri-diagonal structure of the Markov transition matrix indicates the clustering of bit-rateswithin a scene, while the number of model states indicates the range of scene types. Thus, the model provides a means to compare among video sources. The model is also useful for generating small tail probabilities for queue loss and delays which cannot be obtained from a limited set of measurement traces. The stochastic model we present here can therefore provide a number of insights that link source characteristics and encoder features with model structure and parameters which can be useful both for practical encoder implementations and for QOS in call admission and policing.

Compared to one-layer video, much less is known about the statistical characteristics of two-layer video. In two-layer coding algorithms, the base layer can be decoded independently to produce a lower quality picture, and should be transported at high priority with negligible loss. The enhancement layer, which contains the remaining information, can be transported at low priority. Pancha and El Zarki [15] explore the statistical characteristics of a non-standard approach for splitting MPEG bitstreams similar to data partitioning, while Ismail et. al. [16] examine modeling MPEG-2 data partitioning using TES. However, both of these assume that the base layer is VBR. This poses a problem for CAC and UPC on the important base layer. While CAC and UPC are important for a VBR enhancement layer as well, they are more vital, and hence more difficult to do well, for a VBR base layer because

the required cell loss rate is more stringent. Earlier work [17],[18],[19] suggests that the overhead incurred by two-layer coding is large enough to offset any gains in improved cell loss resilience. However, these studies are based on less efficient compression algorithms than are currently standardized in MPEG-2. Furthermore, despite the prevailing assumption in the community that two-layer video provides invaluable cell loss resilience, and despite the appearance of two-layer video in several international compression standards, it is still unknown whether or not the additional bandwidth required to transport two-layer video is offset by its improved resilience to losses.

In this work, we focus on modeling SNR scalability, in which the base layer consists of a coarsely quantized version of the video, and the enhancement layer contains the refinement information. We choose SNR scalability because it provides the best trade-off between error resilience and complexity among the standardized layered coding algorithms [4]. Our goal is to characterize the statistical process of the VBR enhancement-layer traffic, given that the base layer is CBR, and examine the resulting multiplexing performance.

Understanding the impact of choosing different parameter values in a two-layer compression algorithm can be facilitated by using a model instead of using simple trace-driven simulations, in part because of the large number of free parameters to choose. The parametric model we present here is particularly useful for studying the impact of rate-control algorithms and encoding parameter changes on the video traffic statistics, both for one- and two-layer encodings. On the contrary, trace driven studies can become quite prohibitive as the number of free parameters and data gathering requirements increase substantially. Further, the requirement that multiplexed sources be independent typically limits trace-driven simulations to about 10-20 sources, while the multiplexing behavior of hundreds of sources can be studied using a model.

Because the output of the video source will be policed at the network interface, a video system will restrict its traffic to fit within the declared traffic descriptors [32]. In general, the coder that produces a bitstream conforming to constraints will not have the same statistical characteristics of an unconstrained coder [20]. However, statistical models of unconstrained coders are still useful in a variety of roles [7]. In this paper, we use a model of an unconstrained VBR coder to evaluate the burstiness of the one- and two-layer coders, as well as to provide insight into both useful traffic descriptor parameter values and the resulting multiplexing gains if these traffic descriptor parameter values are used for call admission.

Section 2 describes the one- and two-layer video coding algorithms used in this paper. Section 3 presents the statistical features and characterization of measured VBR video data. Based on the statistical features identified, a source traffic model is proposed. Section 4 gives results of the source model simulation and validation for one-layer video. Section 5 addresses the application of the traffic model to two layer video, and presents a comparison of the multiplexing efficiency of one- and two-layer video. Section 6 concludes the paper.

2. VIDEO ALGORITHMS AND DATA COLLECTION

2.1 One-layer encoder

In this paper we model both video-conferencing data generated using the H.261 recommendation [21,22] and full-motion video coded using MPEG-2 [23,3]. Both standards allow the encoder to compress both spatially and temporally. A video frame is segmented into a fixed number of blocks. For each block in the current frame, a block in the previous frame is found that best predicts the current block. If the resulting motion-compensated prediction block is close enough to the current block, the block will be coded as a predictive (P) block; otherwise, the block will be coded as an intra (I) block. For P blocks, only the differential information between the original block and the predicted block is transmitted. P frames contain at least one P block, while I frames contain only I blocks. Since P blocks use previous information, P frames typically are characterized by moderate bitrates, while I frames are characterized by the highest range of rates. MPEG-2 also allows B frames to be generated. Typically, B frames are characterized by the lowest rates, since they can use prediction from both a previous and a future frame. In this work, we do not consider an encoder that uses B frames.

The video data collected for this work were encoded with H.261 for the teleconferencing video sessions and using MPEG-2 for the entertainment applications. In all cases, the data is collected by using a fixed quantizer step-size throughout the sequence. Detailed specifications for these encoding schemes are provided in the references [3,22,23]. In our H.261 encoder, the first frame is an I frame and subsequent frames are encoded as P frames. When we model the H.261 data, we ignore the single I frame at the beginning and only consider the sequence of P frames. The entertainment video corresponds to The Blues Brothers movie. In our MPEG-2 encoder, an I frame is generated only at a scene change. Therefore between I frames an arbitrary number of P frames can occur depending on the video source characteristics, in particular the scene duration. To limit error propagation, each P frame has two slices of blocks that are sent using intra-coding methods. Thus, each block is updated at least once every 15 frames. The frame rate is 1/30 th and 1/24 th of a second for the H.261 and the MPEG-2 data respectively.

Two sample paths of the videoconferencing data referred to as VC_1 and VC_2 are shown in Figures 1(a) and (b) respectively. The vertical axis represents the number of bits per frame r(n) and the horizontal axis represents the frame index n. The two traces correspond to videos of a head and shoulders scene of two different participants in the conference. Figure 1(a) is representative of a participant who shows a continuously changing range of activity in time. In comparison, the participant represented in Figure 1(b) shows less frequent activity. The trace for the entertainment video sequence comprising of I and P frames is shown in Figure 1(c), and is denoted VBR_ENT. Although, MPEG-2 is often coded with a periodic pattern of I and P frames, this periodicity is not required nor does it necessarily obtain the best compression efficiency. For the data considered here, the I frames were triggered by scene changes resulting in an arbitrary number of P frames following an I frame. Approximately 5 minutes of each videoconference data and 15 minutes of the entertainment sequence are analyzed.

2.2 Two-layer SNR scalable encoder

MPEG-2 also standardizes a variety of algorithms for two-layer video, including SNR scalability. SNR scalability provides a way of transmitting two layers at the same spatio-temporal resolution but with different qualities. The base layer encoding process is identical to that for a non-layered encoder. The quantized DCT coefficients from the base layer (after being dequantized) are subtracted from the input DCT block. The resulting quantization error from each block is next re-quantized more finely and encoded to form the enhancement-layer bitstream.

In the standardized decoder for SNR scalability, the dequantized base- and enhancement-layer DCT-coefficients are first obtained by independently processing the respective bitstreams. At this point the two sets of coefficients are summed blockwise, and the IDCT is applied to this sum. To this result is added the temporal prediction signal to produce the output pels, which are also fed back into the motion compensation loop. More details on the coding algorithm for SNR scalability can be found in [3,4].

In this work, we generate SNR-scalable data only for entertainment video, for the movie The Blues Brothers. Again, I frames are used only at scene changes, resulting in a variable number of P frames between adjacent I frames. We do not encode using B frames. Data is collected by keeping both the base and enhancement quantizer step-sizes constant throughout the fifteen-minute sequence. We gather data for a range of base-layer quantizer step-sizes, so that we can subsequently model in section 5.1 how changing the base-layer quantizer affects the enhancement-layer bit-rate. The enhancement-layer quantizer step-size is fixed throughout to produce essentially constant video quality.

3. STATISTICAL MODELING OF VBR VIDEO

In this section, we present the statistical model for VBR video and describe the testing performed to verify its statistical accuracy. In section 3.1, we separate the P and I frames in the data. Section 3.2 discusses the clustering of P frames into statistically significant regions. In section 3.3 the data within a region and the transitions between regions are modeled. Section 3.4 discusses the choice of the optimal number of regions to use.

The model we develop in this section applies to any form of VBR video, whether generated from a one-layer encoder, or from the enhancement layer of a two-layer encoder. To derive a traffic model that generates both P and I frames, we consider two features of the encoding process. First, the generation of I frames driven by scene changes is dependent on the video source and as such can be considered to be independent of the encoder dynamics. However, in the time period characterized by a sequence of P frames one may expect intuitively that there are no significant changes in the information in successive frames. Therefore, the bit rates characterizing successive P frames can be expected to be correlated or more generally clustered around an average value. These observations allow us to devise a technique that uses the correlation between successive frame bit rates to identify the I and P frames in the data.

3.1 Identification of I and P frames

The VBR video is characterized by the number of bits per video frame r(n), where n represents the frame number. In the encoding process, successive P frames carry differential information with respect to a previous scene change. Because of this feature, the bit rate for a given P frame depends on the bit rate of the previous P frame. The dependence feature can be observed by considering the two-dimensional state space of frame bit rates r(n)vs r(n + 1). The scatter diagrams in this state space for the data shown in Figures 1(a-c) are shown in Figures 2(a-c). The correlation between successive P frames is evident by the clustering of frames in the diagonal region of the state space. In the entertainment video data depicted in Figure 2(c), frames that occupy regions outside of the diagonal cluster represent those occurring at transitions to or from I frames. The span of the P and I frames are delineated by the two linear boundaries depicted in Figure 2(c). In Figure 2(c), the region labeled IP corresponds on the y axis to the P frames that follow an I frame. Such a P frame is followed by either another P frame or an I frame. A transition to

another P frame occurs in the region labeled PP. This may be followed by one or more P frames in the region PP or a transition to an I frame in the region PI.

The state space representation allows the identification of any I frames in the data. Assuming that the I frame bit rates are drawn from an independent identically distributed process (iid), the extracted frames are then fit to a distribution function that matches that of the data. This distribution is a Gaussian function. In the next stage we consider the characterization of the P frames in the state space diagrams of Figure 2. For the analysis of the P frame structure we choose to partition the region PP into a finite set of smaller domains, based on a spatial clustering analysis.

3.2 Cluster Analysis for P frames

We require our source traffic model to accurately model the dependence structure in the PP region depicted in Figure 2, which represents the temporal evolution of successive P and I frames observed over a finite time period. At least two techniques can be identified for modeling these observed dependence features. One could assume a single function that captures the correlation structure across the diagonal of the PP regime. For example, a first order autoregressive process could model the one lag dependence in the region PP. This is the basis for the DAR model [6]; however, it cannot capture short-term correlations in the video. A second approach would be to assume that the diagonal region is a composition of a finite set of clusters. Each individual cluster would be represented by a mean value and a variance that captures the extent of the cluster. Since the traffic from most video sources does not typically jump discontinuously from levels of high to low activity and vice-versa within time durations of one or two frames, we believe the second approach will characterize an individual video source better than the first, which may produce discontinuous jumps in the model output traffic.

The segmentation of the P frames in the PP region of state space is carried out using a K-means cluster detection algorithm [24,25]. More sophisticated clustering algorithms could have been used, but we show below that this method was sufficient to meet our modeling objectives. Let N_v be the number of samples of video frames in the data sequence considered. The pairwise sequence of P frames in this region is represented by

$$\Omega_p = \{\bar{x}_i\} \qquad i = 1, .. N_v - 1 \tag{1}$$

where $\bar{x}_i$ is a vector representing the frame rates $(r(i), r(i+1))$. The cluster detection is carried out by dividing the set of two-dimensional vectors into K clusters based on a local minimization of the within-cluster sum of squares criterion. The initial conditions for this algorithm are the number of clusters and a list of the estimates of the centers of these clusters. Let the number of clusters be represented by K. For a given value of K, the initial set of centers are taken to be the K coordinate locations selected at equal increments along the diagonal of the two dimensional histogram of the PP region.

The K-means clustering algorithm progresses by moving points from one cluster to another with subsequent refinement of the cluster centers until a local minima of the within-cluster sum of squares is determined. Each vector belonging to Ω_p is assigned a cluster label $\{1, \ldots K\}$. An important requirement is the choice of the optimal number of clusters chosen to represent the PP region of the state space. Before describing this process, we first describe the structure of the model.

3.3 A Finite State Markov Model for VBR Video

The K clusters that result from the K-means cluster detection process for the P frames combined with the iid distribution for the I frames result in a $K+1$ state model for the VBR video. For simplicity, we will just use K to differentiate between models with a different number of states.

Describing the model requires two pieces. The first describes how the modeled video makes transitions from one state to another. The second piece describes the output bit rate within a particular state.

3.3.1 Transition probabilities: A Markov chain with K states is used to model the transitions between the states. The transition probabilities are assumed to be time homogeneous and are evaluated empirically from the data as

$$p_{ij} = \frac{\textit{Number of transitions from } i \textit{ to } j}{\textit{Number of transitions out of } i} \tag{2}$$

where, i, j represent the K states of the model. The resulting transition probability matrix is represented as P_K. The Markov transition probability matrices that result for the data analyzed are characterized by diagonally

dominant transition probabilities within the states representing P-frames. Although, the probability spread increases with increase in K, approximately 90% of the total probability is concentrated within the set $\{p_{ii-1}, p_{ii}, p_{ii+1}\}$, for each i. While this feature models the temporal correlation effects imposed by the encoding process it also provides an advantage in considerably reducing the number of dominant parameters that require estimation.

3.3.2 Autoregressive model for P frames: Within a P-frame state, the autocorrelation of frame rates appears to have exponential decay. Thus, we consider a first order autoregressive process (ARP) to model the observed correlation structure. A first order ARP is

$$\tilde{r}_i(n) = \alpha_i \, \tilde{r}_i(n-1) + g_i(n) \tag{3}$$

where $i = 1, \ldots K$ represents a state index, $\tilde{r}_i(n)$ represents the frame rate in state i whose mean value is zero and α_i is the first order autoregressive coefficient. The residual noise term $g_i : N(0, \sigma_i^2)$ is a zero mean Gaussian distribution with a standard deviation σ_i. The modified covariance method [26] is used to evaluate the autoregressive parameters. We explored the general p-th order ARP, and determined using the Akaike information criterion [26] that the optimal order was $p = 1$ for this data. The ARP parameters are given by

$$\alpha = \frac{c_{rr}[1,0]}{c_{rr}[1,1]} \tag{4a}$$

and

$$\sigma^2 = c_{rr}(0,0) - \alpha \, c_{rr}(0,1) \tag{4b}$$

where

$$c_{rr}(j,k) = \frac{1}{N-1}\left(\sum_{n=1}^{N-1} \tilde{r}(n+j)\, \tilde{r}(n+k)\right) \tag{5}$$

is the modified covariance function and N represents the number of samples over which the covariance function is evaluated.

Next, we test the hypothesis that the ARP actually characterizes the data within a state. For each state, only sequences of runs greater than or equal to the mean run length are considered in the testing procedure. First, the residuals were tested for white noise criteria at the 95% confidence level. For each state, a number of run length sequences were tested for the correlation structure. It was found that on average, 90% of the sequences passed the autoregressive test. The value of α_i and σ_i^2 for each state was taken to be the average value of these parameters obtained over all sequences for a given state i. The frame rates are then constructed by adding the mean values for each state,

$$r_i(n) = m_i \, (1 - \alpha_i) + \alpha_i \, r_i(n-1) + g_i(n) \tag{6}$$

where m_i is the estimated value of the mean rate in state i.

In summary, the traffic model for VBR video given the number of states K can be characterized by the following set of equations and the transition probabilities estimated from the data. Assuming the process is in state j at time $n-1$ and in state i at time n,

$$r_i(n) = m_i \, (1 - \alpha_i) + \alpha_i \, r_j(n-1) + g_i(n) \qquad i = j \tag{7a}$$

$$r_i(n) = m_i + g_i(n) \qquad i \neq j \tag{7b}$$

where $i, j = 1, \ldots . K$ and $\alpha_i = 0$ for the state representing I frames.

3.4 Selecting the optimal number of clusters

The choice of the optimal number of clusters is based on examining the eigenvalues of the matrix P_K. It can be shown that the n step transition matrix P_K^n , can be algebraically represented in the following form [27],

$$P_K^n = \sum_{i=1}^{K} c_i \ \lambda_i^n \ \bar{X}^{(i)} \ \bar{Y}^{(i)} \tag{8}$$

where λ_i represents the $i-th$ eigenvalue, $\bar{X}^{(i)}$ and $\bar{Y}^{(i)}$ represent the left and right eigenvectors of P_K. The coefficients c_i are given by $c_i = 1/(\bar{X}^{(i)} \ \bar{Y}^{(i)})$. The elements p_{jk}^n of P_K^n are given by,

$$p_{jk}^n = \sum_{i=1}^{a} c_i \ \lambda_i^n \ x_j^{(i)} \ y_k^{(i)} \tag{9}$$

where $x_j^{(i)}$ and $y_k^{(i)}$ are the $j-th$ and $k-th$ elements of the eigenvectors $\bar{X}^{(i)}$ and $\bar{Y}^{(i)}$ respectively.

The transition probability matrix is characterized by having at least one eigenvalue of magnitude one. The rest are either real or complex conjugates such that $|\lambda_i| \leq 1$. Assume that the $\lambda_i, i = 1,..K$ are ordered to have decreasing magnitude, with $|\lambda_1| = 1.0$. Then, in the limit as n tends to infinity, the first term λ_1 is dominant and governs the values of the steady state probabilities. However, in the transient regime, when n is small, the n-step transition probabilities are governed by the dominant eigenvalues in the set $\{ \lambda_i \}$. Thus, the model intrinsically captures the dynamics of the video source for small values of n. It is this feature that captures the short timescale correlations that characterize the video source.

These temporal correlations are reflected in the state sojourn times and affect the loss probabilities in a queue. The optimum number of states should be that value of K that captures the dominant eigenvalues of the VBR video sequence. To identify this feature, starting with a value $K = 2$, we continue to increase K by one, and for each value of K we examine the magnitudes of the eigenvalues $\lambda_i, i = 1,..K$. For each value of K considered, the model behavior is also examined in terms of the losses the source experiences in a queue.

To determine the effect K has on queueing behavior, a leaky bucket policing function, which is a FIFO queue, is considered. The input rate is the varying source rate and the output or drain rate is a constant. This is a finite buffer system and the buffer or bucket size is expressed in terms of the time required to drain a full bucket. If the leaky bucket overflows, the input traffic is in violation. The policing is carried out for a discrete range of drain rates and bucket sizes. For each pair of leaky bucket parameters, the bit loss ratio is represented by

$$BLR = \frac{Number \ of \ bits \ lost}{Total \ number \ of \ bits \ in \ video \ sequence} \tag{10}$$

and is evaluated for the duration of the video sequence.

In Figures 3(a) and (b), the bit losses observed for various values of K are depicted for sources VC_1 and VC_2 respectively. The results are shown for a bucket size corresponding to one millisecond. The drain rate on the horizontal axis is varied from the average rate to the peak rate of the VBR video sequence. In the simulation setup, the arrival process was the variable bit rate video frames with a constant inter-arrival time corresponding to $1/30^{th}$ and $1/24^{th}$ of a second for the videoconference and entertainment sources respectively. Packetization effects were therefore not considered in this study. The model results correspond to the average of several ensembles of the loss curves for five minute sequences (9000 frames) of the video model. At the bit level, with an average of approximately 0.3 Mbps, five minutes of data result in an arrival of over 10^7 bits, leading to statistically significant results for bit loss probabilities considered in Figs. 3.

The results show that the number of states has significant impact in matching the losses experienced by the data. When the rate histogram is modeled by a small number of states, for ex. $K = 4$ in Figures 3, the rate process characterizing each state has a high variance. This reduces the temporal clustering of the frames in the tail probabilities of the rate histogram, leading to smaller loss probabilities for a fixed drain rate. The clustering effect can also be seen in the structure of the transition probability matrix, which is diagonally dominant. As the number of states are increased, the state sojourn times in the tail probabilities are more accurately captured. From Figures 3 (a) and (b), it can be seen that for $K = 16$ and $K = 18$ respectively, the model provides a good match to the losses experienced by

the data. As K is increased beyond these values, the additional eigenvalues that resulted from further partitioning were typically small in magnitude, of order 10^{-1} or less. These values will have negligible effect on the state-sojourn times and contribute minimally to the terms in Equation (9). The procedure described above was carried out for the three video sequences considered. The sources VC_1, VC_2 and VBR_ENT were characterized by $K = 16$ $K = 18$ and $K = 17$ respectively.

3.5 Design procedure

The traffic model for VBR video is summarized by the block diagram shown in Figure 4. To determine the appropriate parameters for the traffic model, we use the procedure depicted in Figure 4. While the figure contains more details, the basic approach is to apply the procedures in section 3.1, 3.2, 3.3.1, 3.4, and finally section 3.3.2. In the next section, we present the results of simulations driven by the parametric model described in this section.

4. ONE-LAYER VIDEO: Numerical Results

We first present results to verify that the model can accurately characterize a single one-layer video source, and then examine its ability to characterize the multiplexing performance.

4.1 Model Validation

For the selected number of states, K, characterizing the video source, the sequence of video frames are simulated using the model proposed in the previous section. The results of the simulation will be validated against the data by considering three different measures. The first measure is the loss occurring in a leaky bucket policing function, and has been described in the previous section as a constraint for selecting the number of states. The second and third measures, the quantile-quantile (QQ) plots and the expected value and variance of the bitrate over a given time interval, are described in more detail later.

The results for VC_1 and VC_2 are depicted in Figures 3(a) and (b). Corresponding results for the entertainment video are shown in Figure 5. Here a range of bucket sizes with values of $\{0.001, 0.3, 0.6\}$ seconds are considered, to show the ability of the model to characterize source behavior on different time scales. Figure 3 shows results specifically for a bucket size of 0.001 seconds, although the videoconferencing data is relatively insensitive to changes in the bucket size in the range of $0.001 - 0.6$ second. On the other hand, the entertainment video, which includes I frames, is sensitive to the variation in the bucket size. In all cases, the simulated video exhibits reasonably good agreement with the data, even for a bucket size more than twice that considered in [13].

We also examined the ability of the model to predict actual bit rates not inside the training set. We extended the measurements of the entertainment video to 20,000 frames. The model parameters were generated from the first 10,000 frames, and we evaluated whether the resulting model could characterize two successive 5000-frame segments of the trace. Space restrictions prevent a detailed description of our results. However, the model does very well in characterizing the bit-loss rates for the remainder of the trace as well.

The losses that characterize single source video occur primarily due to two reasons. In the entertainment video, the dominant effect is due to the presence of I frames. Although they correspond to a single frame occurring at random intervals in time, they are characterized by a sufficiently large number of bits and lead to a full buffer for several frame durations. In the videoconference application, the losses are primarily due to the temporal clustering of frames in the states generating bits rates greater than the drain rate. It is also the reason for the observed invariance of loss ratio to the increase in buffer size. A similar insensitivity to buffer size for video teleconferencing is pointed out in [30].

The model also provides insight for determining useful traffic descriptor parameters. The leaky-bucket losses can be determined with the data, but the Markov chain model in conjunction with the fluid buffer analysis [38] provides a computational means for determining the losses. The finite buffer delay distributions can be derived with the knowledge of the eigenvalues and eigenvectors of the source Markov chain.

A distinguishing feature of our Markov model, in comparison to previous models, is that the structure of the transition matrix adequately characterizes how the data clusters in a group of neighboring states. This feature is not accounted for in the Markov chain structure of the DAR [6] model. In the DAR model, while there is a high probability of the process remaining in the current state, its transition out of that state is typically towards the mean of the marginal rate distribution. This feature results in excessively high holding times in tail states, having an adverse impact on the accuracy in predicting the statistical gains and losses. For example, a twenty-state DAR model fit to the trace VC_2 was found to yield an average holding time of 140 frames in states generating greater than 10^4 bits per frame. As a result, the multiplexing gain predicted using the DAR model is smaller by more than a factor of 1.5. The eighteen-state Markov model developed here results in an average holding time of 39 frames in these states and

is comparable with the data trace which yields an average value of 32 consecutive frames in this region of the marginal rate distribution. The accurate modeling of short term correlation structure is therefore a key ingredient in video traffic models.

The other two measures we considered to validate our model are a visual agreement of the distributions of the frame bit rates as demonstrated by the quantile-quantile (QQ) plots [28], and the expected value and variance of the number of bits of the video source over a given time interval. Representing the time interval in terms of the number of frames, let $\overline{rs}(m)$ be the sequence of values representing the aggregate number of bits in m successive frames. The sequence is obtained by segmenting the data into non-overlapping time intervals of duration m and summing the number of bits in this window. The number of values in the sequence depends on the length of the video data analyzed. The expected value and variance of this parameter are denoted by $E[\overline{rs}(m)]$ and $Var[\overline{rs}(m)]$ respectively, and are evaluated for increasing values of m. The ratio of the variance and the expectation of $\overline{rs}(m)$ is referred to as the index of dispersion for counts. This is a second order descriptor that has been used to capture the burstiness properties of arrival processes [29]. It allows one to quantify the deviation of the source traffic from being a Poisson or a renewal process.

A comparison of the results from the simulation and the measured data is depicted in Figures 6 and 7. Figures 6 show the QQ plots and Figures 7 depict the variations of $E[\overline{rs}(m)]$ and $Var[\overline{rs}(m)]$ for the three video sequences considered. These statistical descriptors are seen to be in good agreement with the data. The window size m is varied from 5 to 100 in the results shown, which corresponds to $3-4$ seconds of consecutive arrivals. The Markov model, with its temporally varying state space characteristics, captures the degree of variability exhibited by the data over the time scales considered in the plots.

4.2 Statistical Multiplexing of One-Layer VBR Video

VBR video has the potential for statistical multiplexing, so that potentially more calls can be carried than with CBR transport. However, an accurate comparison is only possible if the subjective video quality is equivalent in both the VBR and CBR video [32,37]. Since to our knowledge, no one has yet found a systematic method to determine the CBR bandwidth necessary to produce subjectively equivalent video quality to a given VBR video stream, we use a rough rule of thumb here to estimate a rate between the peak VBR rate and the average VBR rate. We estimate the CBR rate r cbr as the maximum of the average of M successive video frames in the VBR traffic stream. That is,

$$r_{cbr} = \max_i[\sum_{i}^{i+M-1} r(i)], \qquad i = 1, \ldots N_v - 2 \tag{11}$$

The value chosen for M relates to the amount of physical buffering delay in a CBR system. As such, r_{cbr} approximates the drain rate of a physical encoder buffer in the CBR system, where the buffer is large enough to store M frames compressed at the average rate r_{cbr}. It can be expected that in a real CBR system with such a buffer, the video quality would not be lower than the VBR quality. Since it is generally agreed that viewers rate the overall quality of a sequence to be equal to its worst quality, this is a reasonable approach to approximate the CBR rate which would produce equivalent visual quality to the VBR trace.

For teleconferencing, we use $M = 3$ corresponding to a delay of 100 ms, motivated by the fact that in H.261 a buffer size of at least 3 frames is required. For entertainment video, physical delays may be longer, between $M = 6$ frames (0.25 seconds) and $M = 60$ frames (2.5 seconds). For the Blues Brother trace, the estimate of r_{cbr} for $M = 6$ is 14% higher than that for $M = 60$. We use $M = 6$ throughout this paper for the entertainment video.

In the simulation study, we consider multiplexing statistically identical video sources. The video sources are fed into a common buffer and served at a channel capacity C Mbps in first-in-first-out order. Multiplexing is carried out by admitting successive sources into the buffer every δt seconds where $\delta t = t_f/N$ and t_f is the frame interval and N sources are admitted. The buffer size is fixed to limit the maximum frame delay to twenty milliseconds. For each value of C and fixed delay, we determine the maximum number of VBR sources that can be carried for a bit loss ratio of less than 10^{-6}. The statistical gain is evaluated relative to CBR transport at rate r_{cbr} calculated above. Let $CBR_{\max}$ be the maximum number of CBR sources that can be carried by the link. This number is given approximately by $C/\hat{r}_{cbr}$, where $\hat{r}_{cbr}$ is the average of the equivalent CBR rates for the N streams. The statistical multiplexing gain G is defined as

$$G = \frac{VBR_{\max}}{CBR_{\max}} \tag{12}$$

The statistical multiplexing results are tabulated in Tables 1, 2 and 3 for the three sets of data considered. As the capacity increases, the multiplexing gain increases, until it converges to the maximum achievable gain. Also shown is the channel occupancy, which is given by $\rho = C/N * r_{avg}$ where r_{avg} is the average rate of the source in bits per second.

The videoconferencing sources are seen to exhibit higher statistical multiplexing gains of 2.5 and 3.7 compared to the entertainment video, where the gain is of the order of 1.5. The higher gains for the videoconferencing sources can be attributed to the relatively long holding times spent in a group of low-activity (transmitting below channel rate) states. This is captured in the model by the near tri-diagonal structure in the Markov transition probability matrix. In addition, since the diagonal transition probability is largest, the model output tends to have long holding times in neighboring states. In the entertainment video, this clustering effect is less obvious because each state of the Markov chain can transition to an I frame.

The results in Tables 1, 2, and 3 are optimistic because were the call admission procedure based solely on the negotiated traffic descriptors, fewer calls would be admitted than predicted by simply multiplexing sources until the cell loss rate becomes too great. In section 5.2.2, we show that using leaky bucket traffic descriptors for call admission under-utilizes the channel for one-layer video. However, we also show there that leaky bucket descriptors can more accurately characterize two-layer video.

5. MODELING TWO LAYER VIDEO

Here, we adapt the single-source model described in the previous section for VBR video to model the enhancement layer of a two-layer video encoder. Our primary goal is to explore the relative efficiency of transporting one- and two-layer video. Early results [17,18,19] indicated that while two-layer video could withstand higher cell loss rates than one-layer video, its overhead bitrate was large enough that it still had worse multiplexing performance. However, the two-layer compression algorithm used in this paper is more efficient, and hence may have higher multiplexing gains than one-layer video.

5.1 Two-layer model

Studying the characteristics of two-layer video, as opposed to one-layer video, vastly increases the number of free parameters. For example, each different base-layer bitrate produces different characteristics in the enhancement-layer bitrate. To study this effect without a model, it would be necessary to encode the video for each different base-layer bitrate of interest. Furthermore, the enhancement-layer bitrate characteristics also depend on the particular rate control algorithm used to enforce a constant rate in the base layer. If the impact of different rate control algorithms are to be studied, again, traffic mea-surements must be performed for each different algorithm at each different base-layer bitrate of interest. Clearly, the amount of data gathering quickly becomes prohibitive if actual traces are to be used.

In this paper, we generate a model of the two-layer video traffic based on traffic measurements gathered during a single set of encodings using different base-layer quantizer step-sizes. Both quantizers are fixed throughout an encoding, but the enhancement-layer quantizer is the same for each encoding so that each encoding has essentially equivalent video quality. We then use the model to characterize the enhancement layer traffic generated for a specific base-layer bitrate and rate control algorithm. The flexibility of the solution allows different base-layer bitrates and rate control algorithms to be easily analyzed without gathering additional data. Rodriguez-Dagnino et. al. used a similar approach in [33] where they proposed a set of fundamental parameters that characterize a bitrate process from which the effect ofdifferent encoding schemes can be determined.

In [7] we present a model for one-layer video that incorporates feedback for buffer control, as shown in Figure 8. This model enables characterization of a video source compressed for a constant-rate channel. The source model generates the output bit-rate of a constant quality VBR encoder with quantization parameter fixed at Q=4, while a multiplicative scaling factor accounts for variations in bit-rate caused by using a different quantization parameter to ensure no buffer overflow or underflow. The model in Figure 8 was shown to accurately capture the video quality when CBR transport is used [7]. This type of model can also be used to shape the one-layer bitrate to meet a given UPC constraint, for example to meet a given leaky bucket parameter description.

We extend this concept to obtain a model for two-layer video when the base layer is transported with CBR (or any other channel constraint including a leaky bucket constraint), as shown in Figure 9. Base and enhancement layer data measured using a constant quantizer step-size of $Q_b = 8$ in the base layer and $Q_e = 3$ in the enhancement layer are used as reference input for generating the two-layer video. Since the base layer is to be transported at a constant bit-rate, the quantization parameter actually used when encoding the base layer would have to be chosen such that the base-layer buffer neither overflows or underflows. By changing the actual Q_b to be different than 8, the base-

layer bit-rate varies, as does the enhancement-layer bit-rate. These variations are well characterized by a multiplicative scaling factor plus an offset, $R_Q = A_Q R_8 + B_Q$. The parameters A_Q and B_Q for these scaling functions are derived from measurements of base and enhancement layer signals encoded at a range of base layer quantization step-sizes varying from 4 to 30. A pair-wise regression analysis on these measurements provides the scaling parameters that map the bit rate corresponding to a reference quantization step size to the approximate bit rate for any other quantization parameter. Separate scaling functions are used for the I frames and the P frames. As depicted in Figure 9, when Q_b increases, less data is coded in the base layer and more is left to be coded in the enhancement layer.

The shapes of the scaling functions were examined for five successive sequences of 5000 frames and were found to be reasonably, although not completely, independent of the sequence. For the base-layer, the curves for A_q cluster tightly around the curve found from using the entire 25,000 frames, while there is slight variation across the base-layer B_q curves. For the enhancement layer, three of the five sets are well characterized by the combined A_q curves, while one curve is somewhat higher and one is somewhat lower for large quantization step-sizes. The B_q curves are correspondingly more spaced at these large quantization step-sizes.

The rate control algorithm used for buffer control is as follows. The quantization step (q-step) is adjusted based solely on the buffer fullness. When the buffer is empty, the minimum q-step of 4 is used, while when the buffer is full, the maximum q-step of 30 is used. Between these end-points, the chosen q-step is an exponential function of the buffer fullness. (We also considered a linear function of the buffer fullness and found no significant statistical difference in the results.) For each value of the base layer CBR rate chosen, the rate control algorithm generates a different sequence for the VBR enhancement layer. This data is then fit using the source model described in the previous section. The model-generated data is then used to compare the tradeoffs in one and two layer encoding.

5.2 Source Statistics and Statistical Multiplexing

We now evaluate the relative statistical characteristics of one- and two-layer encoded video and the corresponding impact these characteristics have on network performance. We begin by describing the single source statistics, and then explore the impact of multiplexing sources. In this study, the CBR base layer is sent with high-priority and the enhancement layer is sent with low priority. This is accomplished by subtracting the cumulative CBR rates from the overall capacity.

5.2.1 Single Source Statistics: We begin with a qualitative discussion of the two-layer source behavior, based only on an understanding of the video encoding algorithm and a coarse examination of the trace data. We then examine the statistical characteristics of the data produced by our model and relate these to our expectations.

First, we consider the average rates. It is clear that setting the CBR base-layer rate too large will increase the overall average rate because there will be times when the base layer rate will be greater than the instantaneous rate produced by a one-layer coder. Further, a cursory examination of the scaling functions computed in section 5.1 indicates that the overall average rate decreases as the quantizer step-size increases. Hence, a smaller base-layer rate decreases the overall average rate, since the rate-control results in a higher average quantizer step-size.

Next, we consider qualitatively the effect of varying the base-layer CBR rate on the peak rates and the associated burstiness of the video, as these are also important in characterizing the multiplexing performance. In this case, there are two competing influences. First, examining the scaling functions indicates that a larger base-layer quantizer step-size forces a larger rate increase in the enhancement layer I frames than in enhancement layer P frames. With a lower base-layer rate, the rate control uses a larger base-layer step-size. Thus a lower base-layer rate produces a greater difference between the rates of I and P frames in the enhancement layer. Therefore the burstiness increases as the base-layer rate *decreases*. On the other hand, we know that when the base-layer rate gets larger, the rate control will be able to fit more and more of the desired P-frame information into the base-layer, lowering the mean of the enhancement layer. The I frames will still be coded with a large enough base-layer quantizer to produce large peaks in the enhancement layer. Thus, the burstiness of the enhancement layer as measured by its peak-to-mean ratio will increase, but as the base-layer rate *increases*.

Using only data gathered from the encoding algorithm, it would require a large number of encoding passes to determine which of these two opposing effects for the enhancement layer burstiness dominates at any given CBR base-layer rate. To this end, we modeled the two-layer video for base-layer CBR values ranging from 1.5 to 3.5 Mbps. This range translates to a factor of 16-37% of the r_{cbr} for one-layer video (9.6 Mbps). The encoder uses a buffer for the base layer large enough to hold six frames at the average base-layer bitrate. The single source statistics for both one and two-layer encoding are tabulated in Table 4, in terms of the overall mean rate of the video signal (CBR base rate + mean rate of VBR enhancement layer), the variance of the frame rates and the peak to mean ratio of the enhancement layer.

As was predicted, the overall average rate increases as the base-layer rate increases. For the Blues Brothers example considered here, for values of base rate up to around 2 Mbps, the overall average rate of the two-layer signal remains comparable to that of the one-layer video. At this base rate, the rate-control is able to use the reference quantizer step-size. For larger base-layer rates, the average quantizer step-size decreases from the reference, and the two effects described above become more pronounced, causing the overall rate increases more quickly.

The burstiness of the enhancement layer can be characterized either by its variance, or by its peak-to-mean ratio. We see from Table 4 that the enhancement-layer variance decreases monotonically as the base-layer rate increases. On the other hand, the peak-to-mean ratio of the enhancement layer decreases to a minimum when the base-layer rate is 2 Mbps, but then increases as the base-layer rate continues to increase. As discussed above, the 2 Mbps value represents the transition point where the peak rate of the enhancement layer drops significantly, causing a reduction in the peak to mean ratio. Below and above this value of the base layer rate, the two competing influences discussed above affect the mean and peak rate statistics differently but the net effect is to increase the burstiness of the video. A further characterization of the traffic processes will be necessary to understand the effect that the aforementioned features have on multiplexing gains.

Thus, we consider next the optimum number of states required to adequately capture the temporal characteristics. Figure 10 shows the magnitude of the eigenvalues of the Markov transition matrix for the one-layer encoding and for two two-layer encodings with different base-layer bitrates. The two-layer encoding with a base rate of 1.5 Mbps has eigenvalues that decay more slowly than the one-layer eigen-values. Therefore, there are more dominant eigenvalues and subsequently lower temporal correlation in the traffic sequence. Hence, the multiplexing gains are greater. As the base layer rate is increased, a value is reached where there is little differentiation in the bit rates between frames in different scenes. Therefore fewer states are required to capture the traffic dynamics, resulting in fewer dominant eigenvalues and accounting for the decreasing gains beyond such a threshold value.

The impact of reduced variance resulting from two-layer encoding is also important in that traffic parameters for the UPC function can be more robust relative to the one layer video. This feature is illustrated in Figure 11 by comparing the bit loss ratios for the two-layer and the one-layer VBR video signal. The two-layer result is for a base CBR rate of 1.5 Mbps. This figure also demonstrates good agreement between the two-layer measurements and the model-generated data. For a fixed bucket size of 1 ms, the rate of decay of the bit losses with increasing drain rate is significantly faster than that exhibited by the one-layer data.

5.2.2 Statistical Multiplexing The expected advantage of VBR video coding is the bandwidth savings a network provider could achieve by statistically multiplexing the video sources. There are two ways to evaluate statistical multiplexing. The first, presented later in Table 5, simply examines how many sources can be multiplexed together before cell losses exceed a certain value. This implicitly assumes that the network has a complete knowledge of the statistics of the source. However, in practice, this is rarely the case, as it is expected that the network knows nothing more than that the admitted traffic satisfies the negotiated traffic descriptor. Therefore, the second method to evaluate the statistical multiplexing, presented later in Table 6, is to compute the number of worst-case sources satisfying the traffic descriptor that can be admitted. Here, we examine the statistical multiplexing gain (SMG) of one- and two-layer video using both methods.

The multiplexing gain is determined in terms of the maximum number of sources that can be multiplexed together in a finite buffer before bit losses exceed 10^{-6} or 10^{-3} for one- or two-layer video, respectively. Although the loss thresholds are different for the two encoding schemes, the resulting video has comparable quality [4], since two-layer video can tolerate more losses in the enhancement layer than one-layer video.

We first verify that simulations using our model produce results similar to those obtained by multiplexing segments of the actual measurements. For the latter, we consider a segment size of three hundred frames, which is at least an order of magnitude larger than the temporal correlation length of the traffic sequence. By randomly selecting the starting points of these segments, longer source streams (15000 frames) are constructed for the multiplexing simulation. For a fixed buffer size of 20 milliseconds and channel capacity varying from 45 to 600 Mbps, the number of sources that can be multiplexed for the specified bit loss probabilities are determined. Figure 12 depicts the bandwidth required per source obtained from averaging thirty ensembles of the simulation. The variation about the average was at most plus or minus one additional source stream. It is seen that the results based on the Markov model are in excellent agreement with the results obtained from the measurements.

With the buffer size of 20 millisecs above, results are dominated by the short-term correlations. To examine if the model captures the long-range dependence frequently reported in video [9] we also examined the impact of large buffer sizes, for a fixed waiting time. For the 600 millisecond case, the buffer size is approximately 12 Mbits, which is sufficient to accommodate up to 12 consecutive frames at the peak rate of the one-layer source. Our results indicate that the loss probabilities are valid for such a large buffer size and therefore that the model captures the

relevant aspects of the source's correlation.

Next, we use the model to examine the statistical multiplexing behavior for a range of base-layer bitrates, for two different channel capacities. Table 5 shows the results when sources are added to the multiplexer until losses exceed 10^{-6} or 10^{-3} for one- and two-layer video, respectively. (Recall these two-layer video can withstand higher loss rates for the same quality.) The first observation is that statistical multiplexing efficiency increases with increasing channel capacity for one-layer video. The required per-source bandwidth decreases by 0.4 Mbps as the capacity increases from 155 to 600 Mbps. The corresponding difference for two layers is about half that. This happens because the overall variability of the two-layer VBR signals is reduced, thus requiring fewer sources to be multiplexed for convergence of the statistical gain. The second observation is that for base rates below 2.5 Mbps, two-layer coding has a lower per-source bandwidth requirement than one-layer coding. Again, because the statistical gain converges faster with increasing capacity for two layers, the bandwidth savings is higher for the lower channel capacity. As the base rates become larger than 2.5 Mbps, the multiplexing performance becomes comparable to the one-layer case. This is a result of the rate-control process which, as the base layer rate increases, tends to smooth the enhancement layer signal by absorbing the variability caused by I frames into the base signal. This is most pronounced as the base layer rates increase beyond 2.5 Mbps. The results suggest an optimal ratio between the base and enhancement layer rates.

Finally, we assume the network knows nothing about the source except the negotiated traffic descriptor parameters, namely the leaky bucket size and drain rate. Then, call admission is based on determining how many worst-case sources can be accepted while obtaining acceptable cell loss ratios. Previous work [32] has shown that the maximum SMG when assuming the worst case ON-OFF source is obtained by choosing the drain rate of the leaky bucket to be the actual average rate of the source. We assume that cell losses occur as soon as the network capacity is exceeded; therefore, the size of the bucket does not affect the SMG. However, in general, the bucket size should be chosen to ensure that the frequency of overflowing the bucket does not exceed some fixed value, like 10^{-3}. (In a real system, the video would shape its traffic to ensure no bucket overflow [34].) Using the peak and mean values of each source, shown in Table 4, Table 6 shows the statistical multiplexing gain. The main observation from Table 6 is that two-layer coding has significantly higher SMG than one-layer coding when we assume the network admits calls based on the traffic descriptor parameters. Furthermore, the number of two-layer video sources admitted by this call admission algorithm is significantly closer to the multiplexing achievable if the source is completely characterized (Table 5), relative to the results for one layer. Therefore, the network utilization will be higher with two-layer video than with one-layer video. This is caused by the reduced peak-to-mean ratio for the enhancement layer signals which reduces the margin of error in the assumption of worst-case sources for leaky-bucket policing.

This clearly demonstrates the advantages of two-layer video for ATM networks. Because the enhancement layer can tolerate a higher percentage of losses than one-layer video, and because of the low overhead associated with two-layer coding *for a low CBR base layer rate*, more two-layer video calls can be multiplexed than one-layer video calls of the same quality; hence, network utilization is higher with two-layer video than one layer.

Finally, we consider the impact of smoothing on the multiplexing performance. The two-layer CBR+VBR studied so far has more buffering than the one-layer video, which can reduce the variability. To determine the impact of buffering on the enhancement layer signals we used the optimal smoothing algorithm of [32] to smooth the data from the model and the trace and explore the resulting multiplexing performance. For one-layer video, smoothing only increases the number of multiplexed calls by one, using the method of Table 5 to measure multiplexing capability. While this makes it marginally more competitive to two layers than it was without smoothing, two-layer encoding still outperforms one-layer encoder. If we were instead to use traffic descriptor parameters for call admission, we obtain the results shown in parentheses in Table 6. Now, smoothing greatly increases the number of one-layer calls that can be carried, but it also increases the number of two-layer calls such that two-layer encoding still outperforms one-layer encoding. In general, the multiplexing efficiency is governed both by the peak-to-average ratio and by the length of time spent in states above the average channel rate. Smoothing lowers the peak and hence lowers the peak-to-average ratio, but has little impact on the length of time spent in high-rate states. For high channel capacities and for many multiplexed sources, the latter plays a more significant role in multiplexing performance. Therefore, taking advantage of the available buffering by smoothing does not change the relative multiplexing gains for one- and two-layer video.

6. CONCLUSIONS

A traffic model for variable bit rate video has been presented. The model is characterized by a finite state Markov chain. The temporal correlations in the video are modeled by a diagonally dominant transition probability matrix. By selecting the optimal number of states in the Markov chain to capture the dominant eigenvalues, we can model the short time-scale behavior relevant for network performance studies. The model is validated by matching

statistical features such as the moments and the histogram of the video frame rates. Additional validation is provided by matching the conformance to a leaky bucket policing function. Statistical multiplexing gains obtained by simulation for one-layer video range from 1. 5 to 3. 7.

The traffic model has been used to statistically characterize and evaluate the relative network performance of one- and two-layer video. The Markov model, in terms of the number of states and the corresponding magnitudes of its eigenvalues is shown to capture the two-layer encoding features. Having a model using a Markov chain is useful for the network performance studies, because it is analytically and computationally tractable for addressing admission control issues. Fluid flow models for determining loss and delay statistics based on dominant eigenvalues have been discussed in several previous papers [35,36].

Two-layer encoding has statistically smoother variations in its enhancement layer relative to one layer VBR video. This feature allows a tighter characterization of UPC parameters. Two-layer encoding also results in a significant per-source bandwidth savings compared to one-layer video, particularly when the network admits calls based on a leaky-bucket characterization.

In this paper, we considered the case where variable-bit-rate video would be the only traffic on the network. Clearly, this will not be the case in practice. Indeed, in the final scenario, the unused bandwidth from real-time VBR streams will be exploited by available bit-rate (ABR) and unspecified bit-rate (UBR) services.

ACKNOWLEDGEMENTS

The authors would like to thank D. Heyman (AT&T) for his valuable discussions in the revision of this work and the reviewers for their feedback which strengthened the paper.

REFERENCES

[1] M. Ghanbari, "Two-layer coding of video signals for VBR networks," IEEE J. Select. Areas Commun., vol. 7, p771-781, June 1989.

[2] S. Tubaro, "A two layers video coding scheme for ATM networks," Signal Processing: Image Communication, vol. 3, p129-141, June 1991.

[3] ISO/IEC 13818-2 | Rec. ITU-T H.262, "Generic coding of moving pictures and associated audio," November 1994.

[4] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable coding algorithms", IEEE Trans. Circuits and Syst. for Video Tech., vol. 6, p426-435, October 1996.

[5] B. Maglaris, D. Anastassiou, P. Sen and G. Karlsson, "Performance models of statistical multiplexing in packet video communications," IEEE Trans. Comm., vol. 36, p834-843, 1988.

[6] D.P. Heyman, A. Tabatabai and T.V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks," IEEE Trans. Ckts and Systems, vol. 2, p49-59, 1992.

[7] D.M. Lucantoni, M.F. Neuts and A.R. Reibman, "Methods for performance evaluation of VBR Video Traffic Models," IEEE/ACM Transac. on Networking, vol. 2, p176-180, 1994.

[8] A. Elwalid, D. Heyman, T.V. Lakshman, D. Mitra and A. Weiss, "Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing," IEEE J. Select. Areas Commun., vol. 13, p1004-1016, August 1995.

[9] M.W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR video traffic," Proc. ACM SigComm., London, Sept. 1994

[10] F. Yegenoglu, B. Jabbari and Y. Zhang, "Motion-Classified Autoregressive Modeling of Variable Bit Rate Video," IEEE Trans. Ckts. and Systems for Video Tech., vol. 3, p42-53, 1993.

[11] R. Gruenenfelder, J. Cosmas, S. Manthorpe, A. Odinma-Okafur, "Characterization of video codecs as autoregressive moving average processes and related queueing system performance", IEEE Journal on Selected Areas in Communications, vol. 9, no. 3, April 1991, pp. 284-293.

[12] D.P. Heyman and T.V. Lakshman, "Source models for VBR broadcast-video traffic," IEEE/ACM Transac. on Networking, vol. 4, p40-48, 1996.

[13] M.R. Frater, J.F.Arnold, and P. Tan, "A new statistical model for traffic generated by VBR coders for Television on the Broadband ISDN," vol. 4, p521-526, 1996.

[14] M. Krunz and H. Hughes, "A traffic model for MPEG-coded VBR streams", Performance Evaluation Review (Proc. of ACM SIGMETRICS ’95), vol. 23, no. 1, pp. 47-55, May 1995.

[15] P. Pancha and M. El Zarki, "Prioritized transmission of VBR MPEG video", GLOBECOM ’92, p1135-1139, 1992.

[16] M. R. Ismail, I. E. Lambadaris, M. Devetsikiotis, and A. R. Raye, "Modelling Prioritized MPEG video using TES and a frame spreading strategy for transmission in ATM networks", INFOCOM ’95, p762-769, April 1995.

[17] G. Morrison and D. Beaumont, "Two-Level Video Coding for ATM Networks,’’ Signal Processing: Image Communication, vol. 3, p179-195, June 1991.

[18] J. R. Louvion, "2-Layer Versus 1-Layer Video Codecs: A Network Performance Approach,’’ 4th International Workshop on Packet Video, Kyoto, Japan, August 1991.

[19] J. W. Roberts, editor, "Performance Evaluation and Design of Multiservice Networks,’’ COST 224 Final Report, Commission of the European Communities, Brussels, 1992.

[20] H. Heeke, "A traffic-control algorithm for ATM networks", IEEE Trans. Circuits and Systems for Video Technology, vol. 3, no. 3, pp. 182-189, June 1993.

[21] M.L. Liou, "Overview of the px64 kbps video coding standard," Commun. ACM, vol. 34, (4), April 1991.

[22] "Video Codec for Audiovisual Services at pX64 kbits/s," CCITT Recommendation H.261, 1990.

[23] D. Le Gall, "MPEG: A video compression standard for multimedia applications," Commun. ACM, vol. 34, p47-58, April 1991.

[24] J.A. Hartigan, **Clustering Algorithms,** Wiley, NY, 1975.

[25] J.A. Hartigan and M.A. Wong, "A K-Means Clustering Algorithm," J. Royal Statistical Society, Ser. C, Applied Statistics, vol. 28, p100-108, 1979.

[26] S.M. Kay, **Modern Spectral Estimation,** Chapt. 7, Prentice-Hall Signal Processing Series, 1988.

[27] W. Feller, **An Introduction to Probability Theory and its Applications**, vol. 1, Chapt. XVI, John Wiley & Sons, 1957.

[28] J.M. Chambers, W.S. Cleveland, B. Kleiner and P.A. Tukey,"Graphical Methods for Data Analysis," The Wadsworth Statistics/Probability Series, Duxury Press, Boston, 1983.

[29] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," IEEE J. Select. Areas Commun., vol.6, p833-846, 1986.

[30] D. Heyman, "The GBAR source model for VBR video conferencing", IEEE/ACM Transactions on Networking, vol. 5, no. 4, p554-560, Aug. 1997.

[31] E. Cinlar, **Introduction of Stochastic Processes,** Prentice-Hall, NJ, 1975.

[32] A. R. Reibman and A. W. Berger, "Traffic descriptors for VBR video teleconferencing,'' IEEE/ACM Transactions on Networking, vol. 3, p329-339, April 1995.

[33] R.M. Rodriguez-Dagnino, M.R.K. Khansari, and A. Leon-Garcia, "Prediction of bit rate sequences of encoded video signals," IEEE J. Select. Areas Commun., vol. SAC-9, pp.305-314, April 1991.

[34] A. R. Reibman and B. G. Haskell, "Constraints on variable bit-rate video for ATM networks,'' IEEE Transactions on Circuits and Systems for Video Technology, vol. 2, p361-372, December 1992.

[35] A.I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks." Proc. IEEE INFOCOM, vol. 1, p3.a.2.1 - 3.a.2.10, 1993.

[36] G. Choudhury, D.M. Lucantoni, and W. Whitt, "Squeezing the most out of ATM," IEEE Trans. Comm., vol. 44, p203-217, 1996.

[37] T. V. Lakshman, A. Ortega, and A. R. Reibman, "Variable bit-rate video: trade-offs and potentials", Proceedings of the IEEE, vol. 86, pp. 952-973, May 1998.

[38] M. Schwartz, **Broadband Integrated Networks,** Chapt.3, Prentice Hall, New Jersey, 1996.

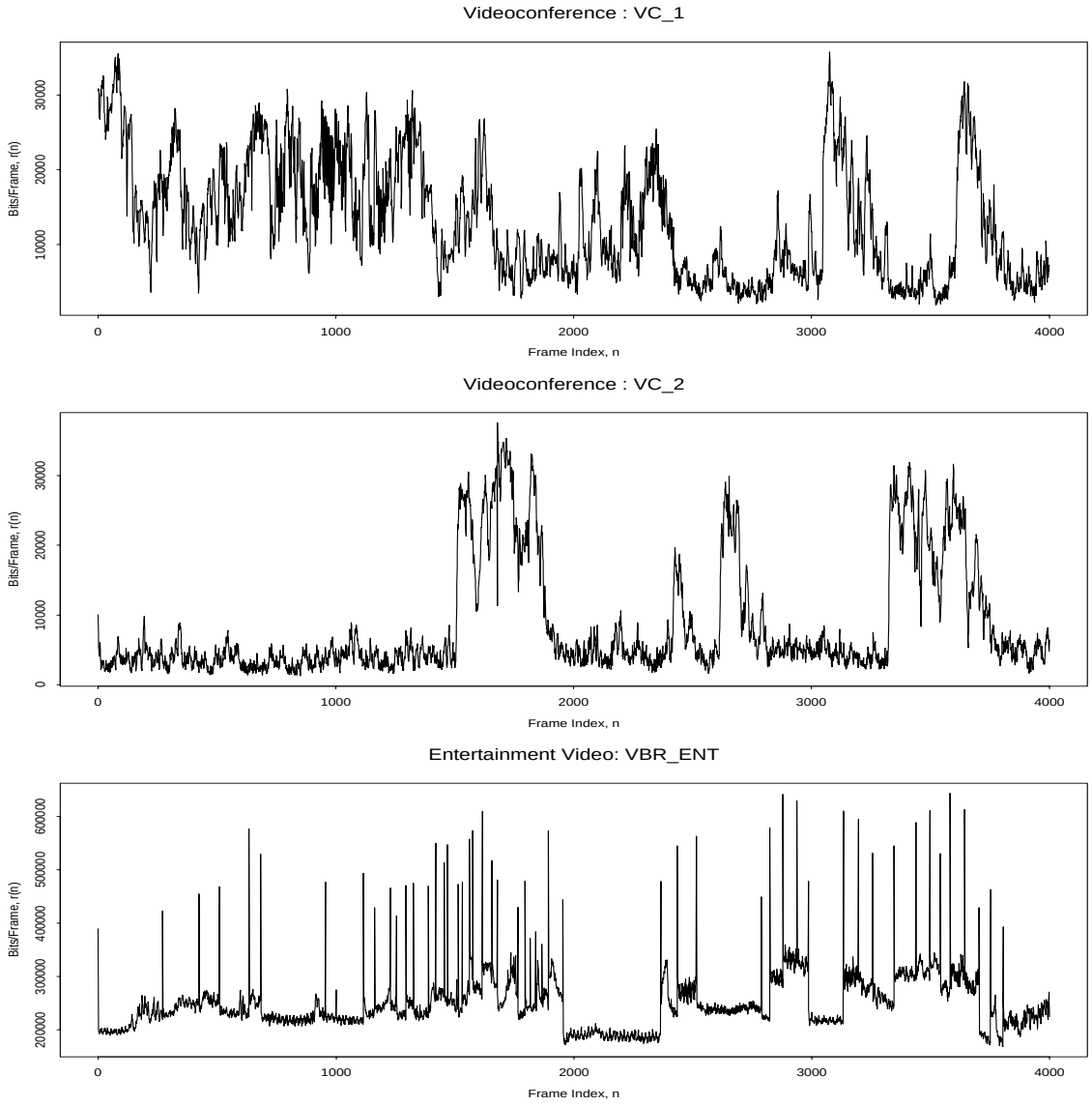
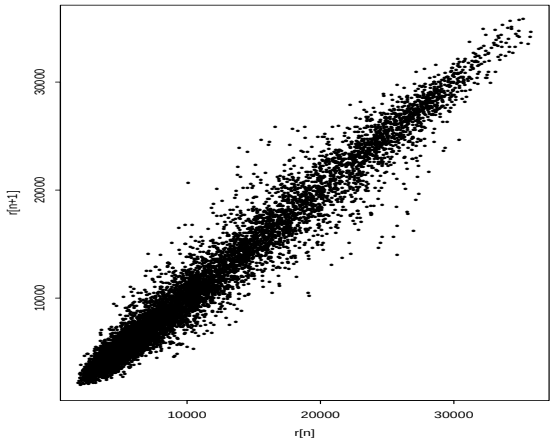
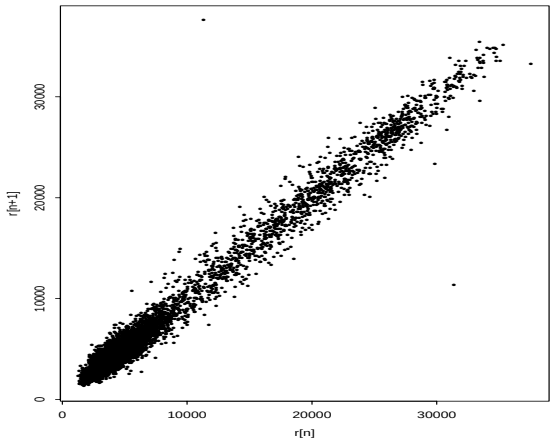


Fig. 1: Sample Paths of VBR video in videoconferencing and entertainment video applications.



(a) *VC*_1



(b) *VC*_2

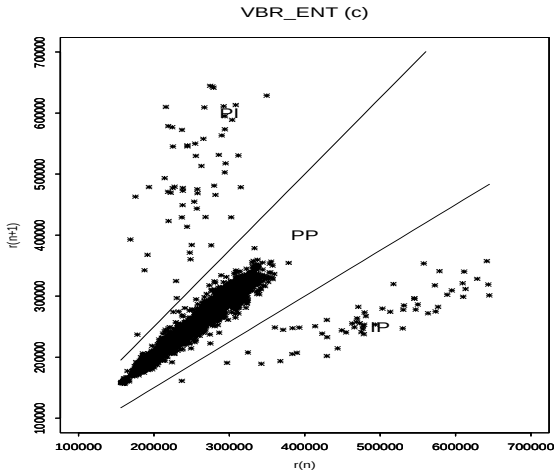


Fig. 2: State space representation of video frame rates.

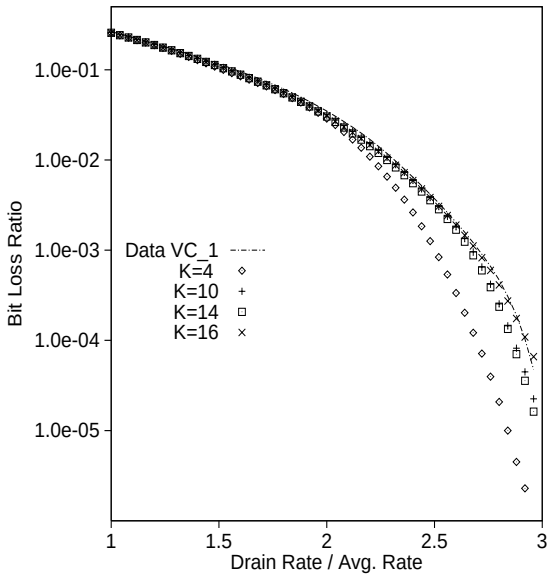


Fig. (3a) Bit loss probabilities as a function of number of states K for VC_1. Bucket size = 1 millisecond.

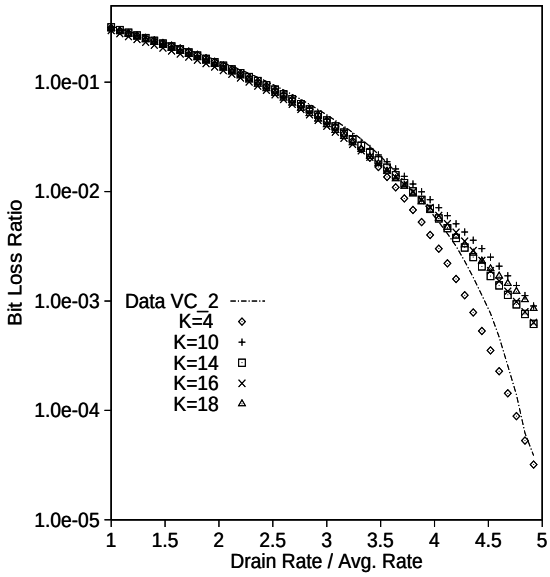


Fig. (3b) Bit loss probabilities as a function of number of states K for VC_2. Bucket size = 1 millisecond.

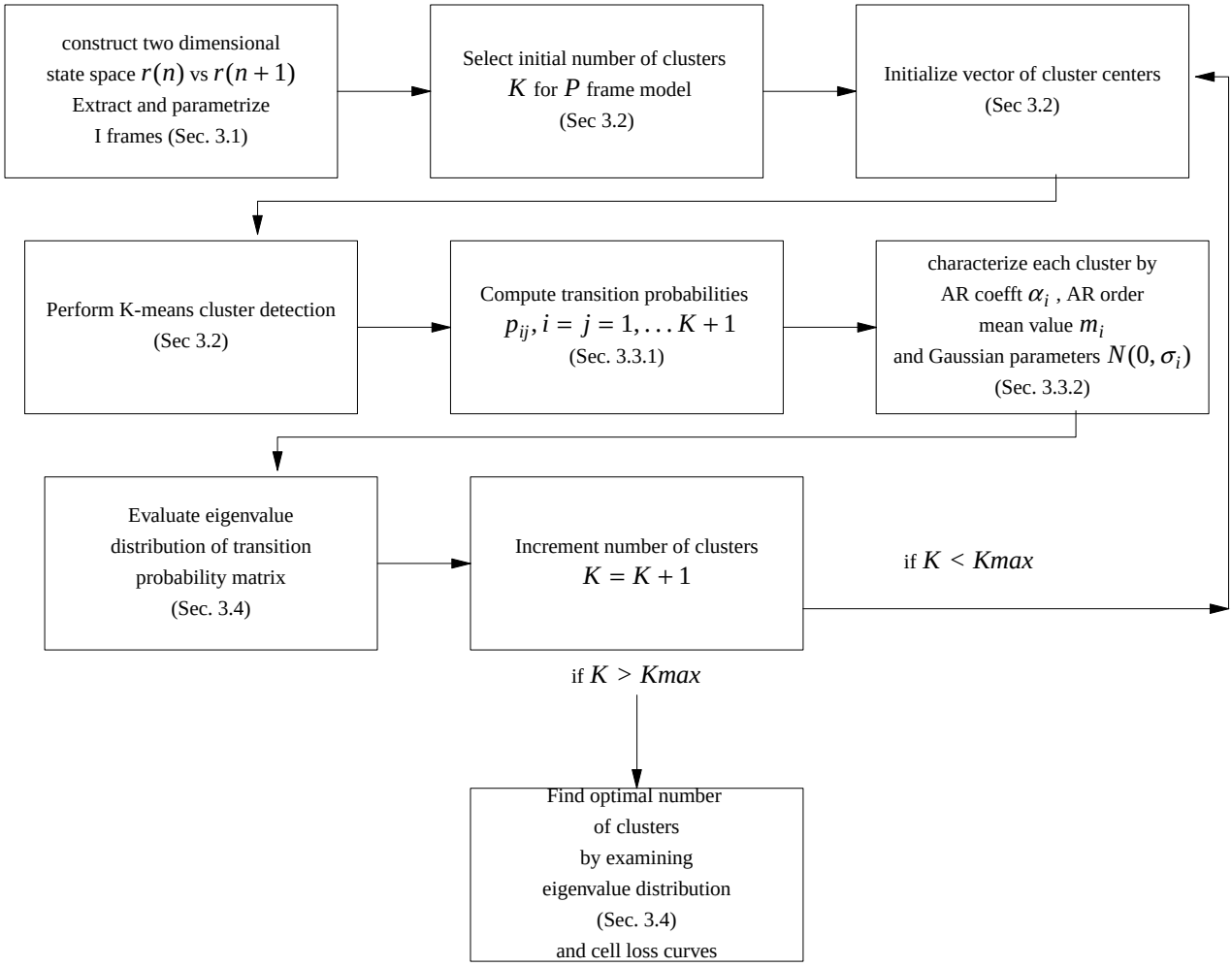


Fig. 4: Block diagram for traffic modeling from measurements

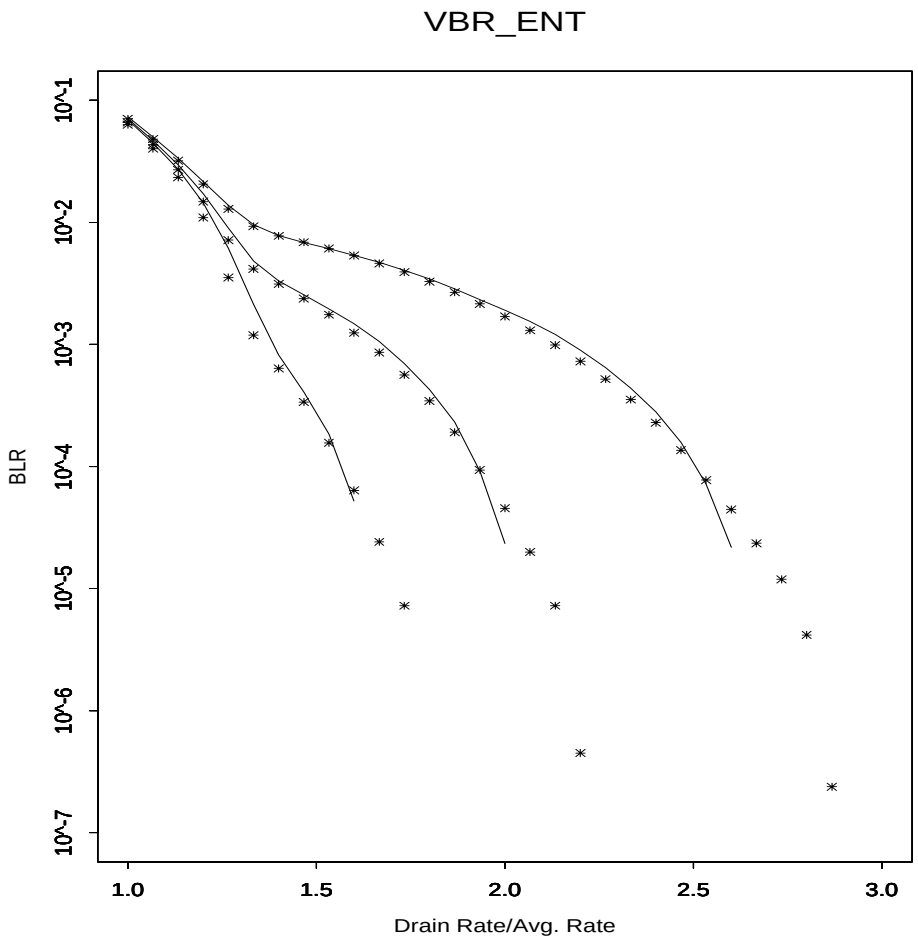
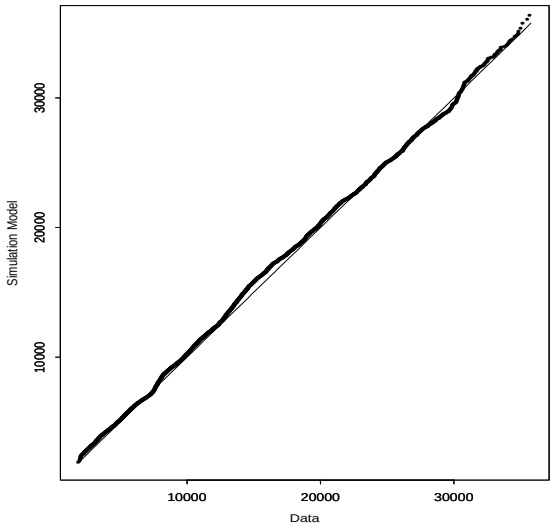
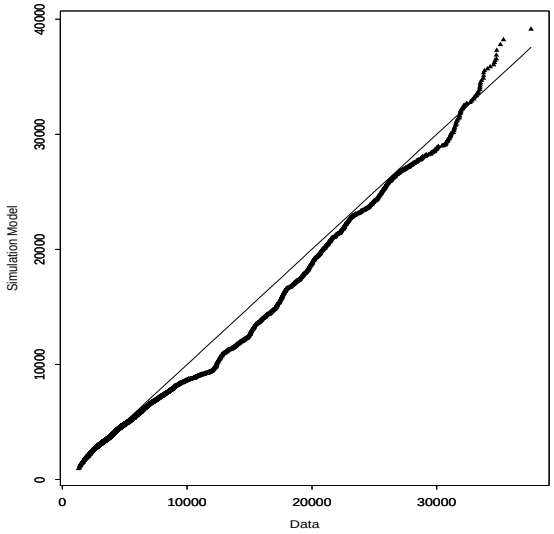


Fig. (5) Comparison of bit losses for entertainment video. Buffer size=0.001,0.3 and 0.6 seconds. (Simulation results in symbols)

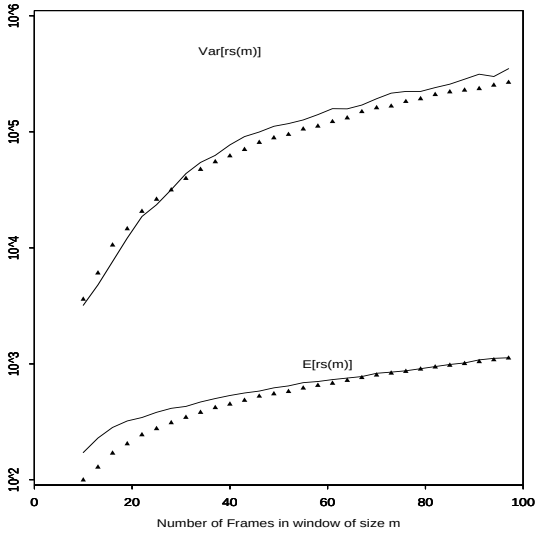


*VC*_1

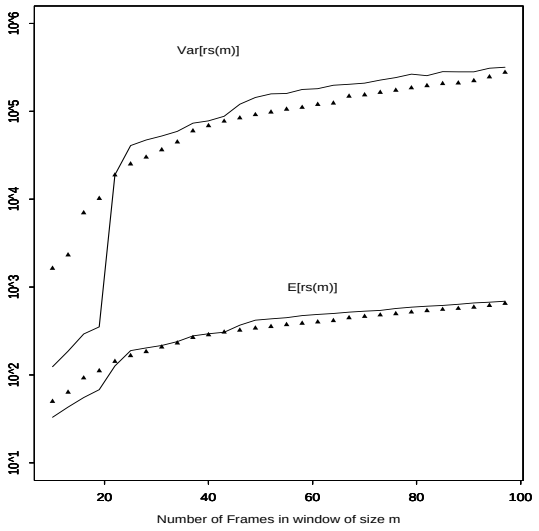


*VC*_2

Fig. (6a) QQ plots for measured data and simulated video.

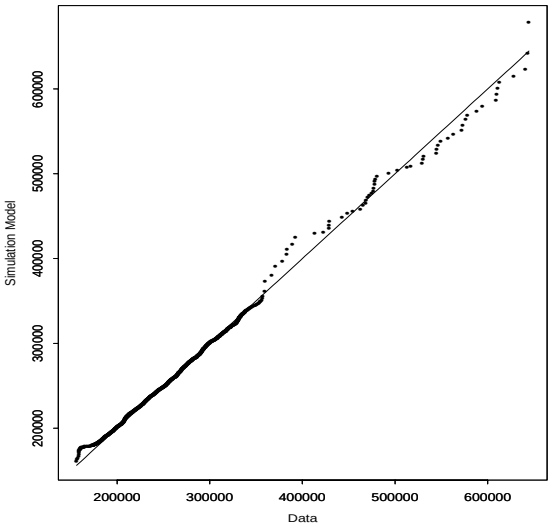


*VC*_1

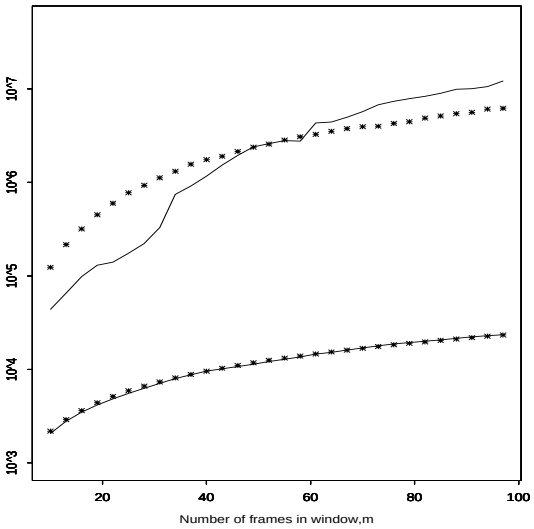


*VC*_2

Fig. (6b) Comparison of first and second order moments for data and simulation. Simulation results in symbols.



VBR_ENT: QQ plot



VBR_ENT: First and second order moments

Fig. (7) Comparison of statistical descriptors from simulation and data for entertainment video. Simulation results in symbols.

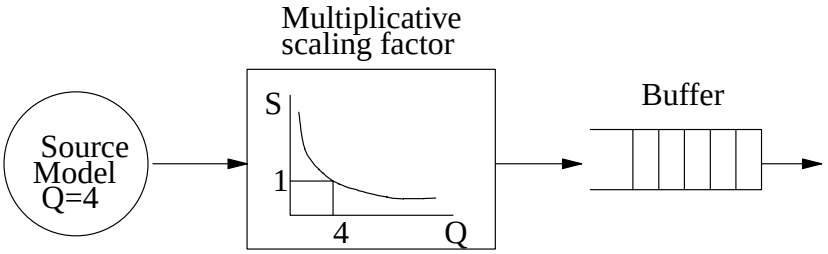


Fig. 8. Rate control for one-layer video coder

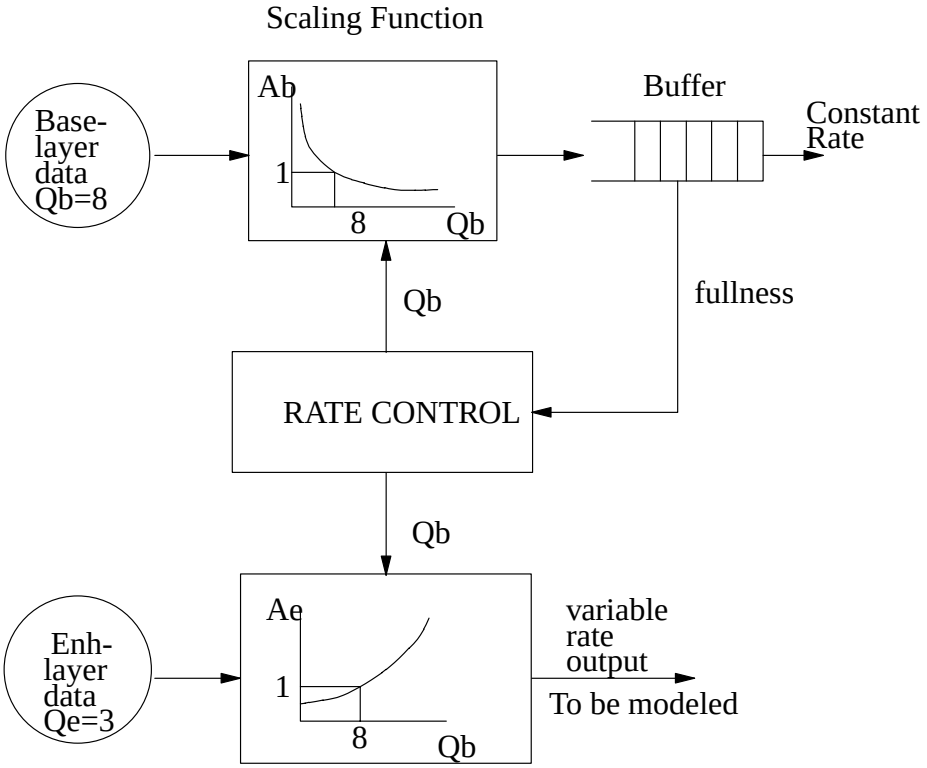


Fig. 9. Effect of base-layer rate control on enhancement-layer bit-rate

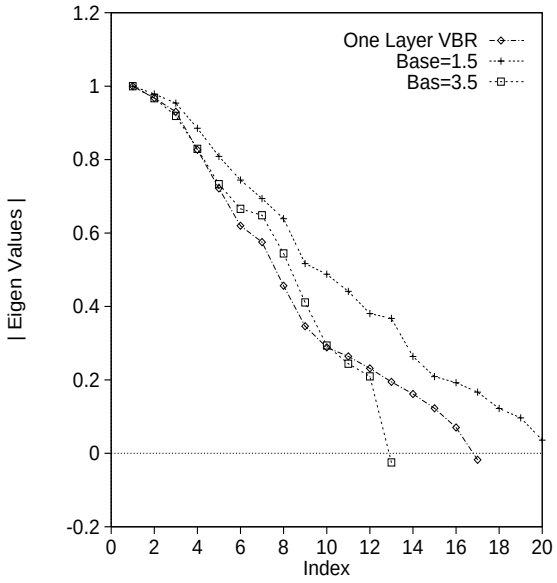


Fig. 10: The magnitudes of the eigenvalues of the Markov transition matrix for one and two layer video models

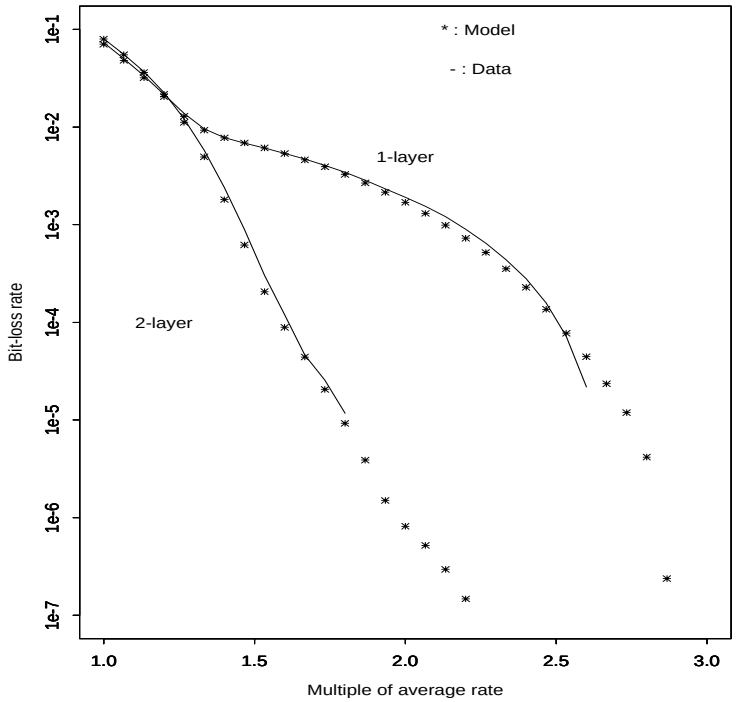


Fig. 11: Bit loss rate for one- and two- layer video. Buffer size = 1 millisecond.

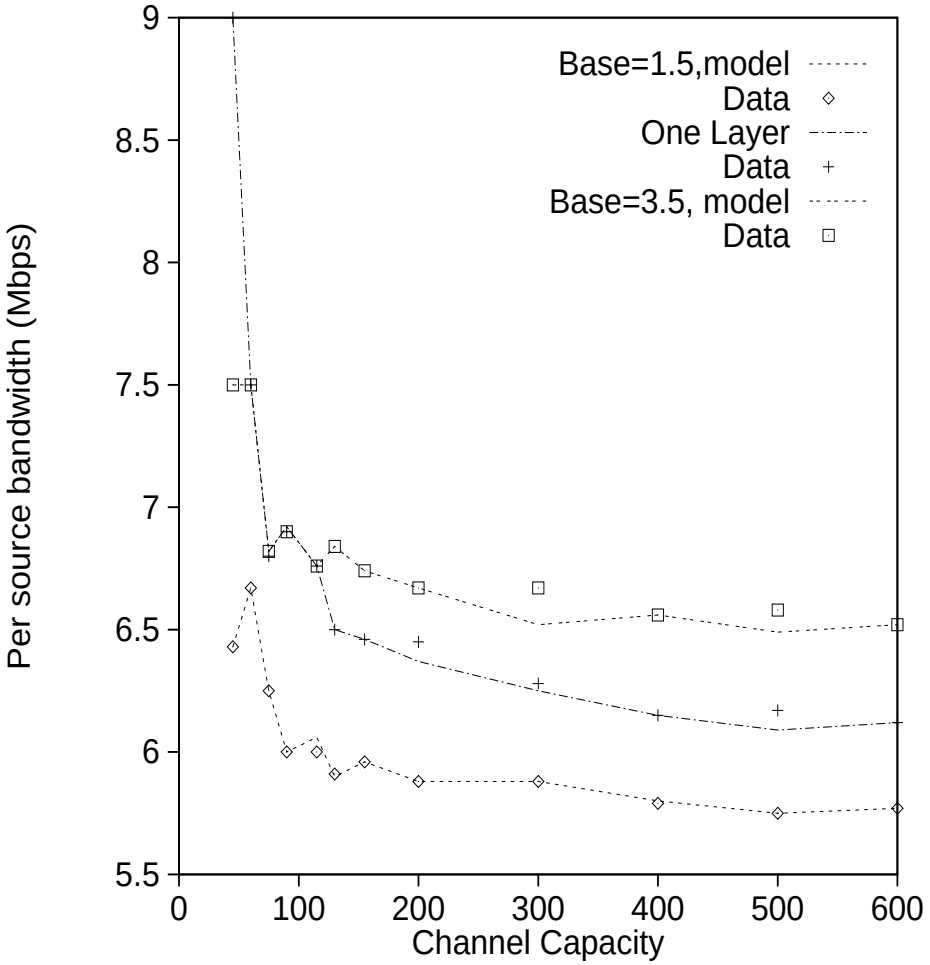


Fig. 12: Comparison of bandwidth per source results from measurements and model.

Table 1: Statistical Gains for VC_1

ource:VC_1	Buffer=20ms	$\hat{r}_{cbr} = 1.06$Mbps	$BLR = 10^{-6}$	
C(Mbps)	ρ	$CBR_{\max}$	$VBR_{\max}$	Gain
15	0.7	14	30	2.1
30	0.8	28	68	2.4
45	0.8	42	109	2.6

Table 2: Statistical Gains for VC_2

ource:VC_2	Buffer=20ms	$\hat{r}_{cbr} = 1.05$Mbps	$BLR = 10^{-6}$	
C(Mbps)	ρ	$CBR_{\max}$	$VBR_{\max}$	Gain
15	0.65	14	46	3.3
30	0.76	28	105	3.7
45	0.75	42	154	3.7

Table 3: Statistical Gains for VBR_ENT

ource:VBR_ENT	Buffer=20ms	$\hat{r}_{cbr} = 9.6$Mbps	$BLR = 10^{-6}$	
C(Mbps)	ρ	$CBR_{\max}$	$VBR_{\max}$	Gain
45	0.8	5	6	1.25
155	0.9	16	24	1.5
600	0.9	64	99	1.5

Table 4: Single Source Characteristics

One-layer	Two-layer base bit-rate (Mbps)					
		1.5	2.0	2.5	3	3.5
Mean bit-rate (Mbps)	5.6352	5.6520	5.8104	5.9856	6.1776	6.4152
Variance (enh. layer)	$2.48 * 10^9$	$1.17 * 10^9$	$1.05 * 10^9$	$9.4 * 10^8$	$8.69 * 10^8$	$7.24 * 10^8$
Peak/Mean (enh. layer)	2.64	1.85	1.77	1.86	1.95	1.97

Table 5: Statistical multiplexing gains assuming complete knowledge of source statistics

		One-layer (BLR=10^{-6})	Two-layer (BLR=10^{-3})				
			1.5	2.0	2.5	3	3.5
	# sources	24	26	26	25	24	23
	per source BW	6.46	5.96	5.96	6.2	6.46	6.73
	# sources	99	105	102	99	96	92
	per source BW	6.06	5.71	5.88	6.06	6.25	6.52

Table 6: Statistical multiplexing gains using leaky bucket characterization
Results from smoothing in parentheses.

		One-layer (BLR=10^{-6})	Two-layer(BLR=10^{-3})				
			1.5	2.0	2.5	3	3.5
	# sources	10 (22)	20 (23)	21 (22)	20 (22)	20 (22)	19 (21)
	per source BW	15.5 (7.05)	7.75 (6.74)	7.38 (7.05)	7.75 (7.05)	7.75 (7.05)	8.15 (7.38)
	# sources	64 (91)	93 (98)	92 (96)	90 (93)	87 (91)	85 (88)
	per source BW	9.375 (6.59)	6.45 (6.12)	6.52 (6.25)	6.67 (6.45)	6.90 (6.59)	7.06 (6.82)