Bandwidth Estimation for Multiplexed Videos Using MMG-based Single Video Traffic Model

Wei Zhou, S. Ramakrishnan*, Dilip Sarkar, and Uttam K Sarkar
Department of Computer Science, *Department of Mathematics
University of Miami, Coral Gables, FL 33124
Emails: wzhou@cs.miami.edu,ramakris@math.miami.edu,sarkar@cs.miami.edu,uttam@cs.miami.edu

Abstract— **A *Video on Demand* (VoD) system is expected to transmit many movies over a single channel as demanded by the end users. When several videos are transmitted simultaneously over a link the effective bandwidth required per video is usually much lower than that needed by a single video because of *multiplexing gain*. Fast and accurate estimation of multiplexing gain is necessary for developing call admission control (CAC) algorithms. Known models which estimate queue size and effective bandwidth of multiplexed video system cannot capture frame size variations in different segments of a video, and are not very useful particularly when the number of videos is not too large.**

The multinomial model proposed in this paper is built on a Markov-modulated gamma (MMG)-based traffic model of a single video. It takes into consideration average frame size variations in different segments of a video and can predict multiplexing gain for any number of multiplexed videos. The model has been validated using MPEG traces of commercial movies.

I. Introduction

Video traffic is expected to dominate the bandwidth of broadband networks [MCS99], [CR99]. Due to its high bandwidth requirement any video traffic is usually encoded with MPEG-like data compression techniques resulting in *variable bit rate* (VBR) video. Significant past research efforts have developed many traffic models for single VBR video [HTL92], [Hey97], [WCJ95], [CR99], [RK98], [SRS01], [SRS03].

A broadband channel can and would simultaneously carry *several* videos because of its high available bandwidth. This high bandwidth will enable a *Video on Demand* (VoD) system to transmit many movies over a single channel as demanded by the end users. In fact, dozens of copies of the same popular movie could occupy a link at the same time. The relative positions of these movies under simultaneous transmission would be independent because of independent user requests.

Fortunately, when several videos are transmitted simultaneously over a link the effective bandwidth required per video for a given data loss ratio (DLR) is usually much lower than that needed by a single video for the same DLR. This is known as *multiplexing gain*. Multiplexing gain is observed because average frame size of a VBR video is usually different in different segments of a video. When many streams of same or different videos are transmitted simultaneously, some would require lower bandwidths and others higher bandwidths than the average bandwidth. Many empirical studies have confirmed that multiplexing gain actually occurs [Kru99a], [Kru99b], [KSH95], [KT97], [CR99]. Since estimation of bandwidth requirement can be used for call admission control (CAC), attempts have also been made to model multiplexing gain (see [Kru99b] and references therein). These models attempt to estimate two parameters, namely, queue size and effective bandwidth, assuming independent Markovian sources. They predict effective bandwidth quite accurately especially when the number of sources is very large since the models are developed for capturing asymptotic behavior. However, these approaches model a video source with too few parameters and hence cannot capture frame size variations in different segments of a video, and are not very useful particularly when the number of videos is not *too* large. The model proposed in this paper takes into consideration average frame size variations in different segments of a video and can predict multiplexing gain for any number of multiplexed videos.

We develop a novel analytical technique for estimation of bandwidth requirement for any number of multiplexed videos. It is based on an MMG-based traffic model [SRS01] of a single video. The MMG model is based on the fact that a video trace can be divided into segments based on frame sizes. Thus, segments associated with smaller frames would require lower transmission bandwidths and those with larger frame sizes would need higher bandwidths.

In Section II, we propose a multinomial model which estimates the bandwidth requirement of multiplexed videos under permissible data losses. For the validation of this model, discussed in Section III, multiplexed video traffic is generated using MPEG traces of commercial movies and the observed bandwidth requirement is compared with the values analytically computed from the multinomial model. Section IV describes the experimental setup and presents the findings. Our observations, possible uses of the proposed model, and future extensions of the work are summarized in Section V.

II. Multinomial Model of Multiplexed Video

The multiplexed video model proposed in this section makes use of the Markov-modulated gamma(MMG)-based model [SRS01] of a single video. The MMG-based model (see Section II-A) captures segment-level frame size variations in a single long-duration VBR video with a finite state Markov chain. The multinomial model, presented in Section II-B, shows how this finite state model can be effectively utilized in a multinomial framework in order to estimate bandwidth requirement when a number of videos are multiplexed on a broadband link.

A. MMG-based Model of Single VBR Video

The MMG-based model is a frame-size model that captures both short-term and long-term frame size variations in a single VBR video. Only the features essential for developing the multiplexed video model are outlined in this section — the detailed algorithms and analysis can be found in [SRS01].

The MMG-based model uses a finite state Markov chain. A state is associated with a sequence of video shots and is indicative of the bandwidth requirement whenever the video is in that state. The duration the video stays in a state depends on the distribution of shot lengths. The model isolates individual I, B, and P frames state-wise resulting in $3n$ sets of frame size data (one each for the I, B, and P frames of the n states). Each of these $3n$ data sets is modeled by a variant of Gamma distribution. An $n \times n$ matrix gives state transition probabilities. The stationary state probabilities indicate the chance of finding the video in any state.

According to this model the video starts in any of the n states. The time interval the video stays in the state follows the duration distribution. In a state, the frames continue to occur in the underlying MPEG GOP sequence (say, IBBPBB) at the designated rate (say, 30 frames per second) for the entire state-duration. The sizes of each of these I, B, and P frames follow the distribution for that frame type of the state. At the expiry of the duration in the state, a state-transition takes place following the Markov chain transition probability matrix. The process repeats for the entire duration of the video. Consequently, the model can be used to estimate bandwidth requirement and detailed data and frame loss patterns of a single video. The next section abstracts the essential parameters of the MMG model necessary to develop the proposed multinomial model.

A.1 Abstraction of MMG Model for Multinomial Model

Let $S = \{s_1, s_2, \cdots, s_n\}$ be a set of n states of the Markov model of a single video. Let p_i be the probability of a video being in the state s_i. Note that $\sum_{i=1}^{n} p_i = 1$, that is, the video system is always in one of the n states. There are data rates (and necessary bandwidth parameters) associated with each state s_i of the video. Let the minimum bandwidth necessary corresponding to the minimum transmission rate while the system is in state s_i be denoted by $Bmin_i$. Similarly, let average and maximum bandwidths necessary while in state s_i be denoted by $Bavg_i$ and $Bmax_i$, respectively. In general, let $Brep_i$ indicate the *representative bandwidth* for state s_i. This notation gives us flexibility as an appropriate choice of $Brep_i$ would allow studying a specific facet of bandwidth behavior. For an example, if $Brep_i$ is set to the 99 percentile of bandwidth interval associated with state s_i the model could make a conservative estimate (more of an upper bound) of the necessary bandwidth. Of course, $Bmin_i$, $Bavg_i$, and $Bmax_i$ could be special choices for $Brep_i$. We assume that states are indexed with increasing bandwidth requirements, that is, $Bmin_{i+1} > Bmin_i$, $Bavg_{i+1} > Bavg_i$, $Bmax_{i+1} > Bmax_i$, and $Brep_{i+1} > Brep_i$ for all i, $1 \leq i < n$. Thus, the minimum and maximum (transmission rates and) necessary bandwidths for the whole video is $Bmin_1$, and $Bmax_n$. The average transmission rate or bandwidth for the whole video is an weighted average of $Bavg_i$, $1 \leq i \leq n$ based on the probability of the states and is given by $B_{video} = \sum_{i=1}^{n} p_i Bavg_i$. **Since a Markov chain has the memoryless property the starting state doesn't have any impact in the long run and can be chosen at random.** This memoryless property also suggests the model should be useful in capturing the total bandwidth requirement of a number of independently multiplexed videos. This observation has led to the development of the proposed multinomial model for multiplexed video traffic.

B. A Multinomial Model for Multiplexed VBR Video

In this section we present the model when m videos with identical state and bandwidth parameters (as obtained from the MMG model) are simultaneously sent over a communication channel. These m videos may be m copies of the same video at independent states, or these may be similar videos with identical parameters. The general model of heterogeneous videos having different parameters can be obtained as an extension of this model and has not been included in this paper. We denote the m videos in the multiplexed system by $V_1, V_2, \ldots, V_m$.

B.1 States in the Multinomial Model

Let s_{t_j} denote the state of V_j at time t. Since the videos are transmitted to meet user demands originating from independent users at the receiving ends, we assume that at any point of time the state of V_j is independent of V_k, for all $j \neq k$. At time t, the state of the video system is characterized by an m-tuple identifying the combined states of these m videos, namely, the tuple $(s_{t_1}, s_{t_2}, \cdots, s_{t_m})$, where $s_{t_j} \in S$ for $1 \leq j \leq m$. The maximum number of states the video system can be in is clearly n^m.

B.2 Bandwidth and Probability of a State

At any instant t the minimum, average, maximum, and representative bandwidths of the multiplexed video system can be computed as, $Bmin(m) = \sum_{j=1}^{m} Bmin_{s_{t_j}}$, $Bavg(m) = \sum_{j=1}^{m} Bavg_{s_{t_j}}$, $Bmax(m) = \sum_{j=1}^{m} Bmax_{s_{t_j}}$, and $Brep(m) = \sum_{j=1}^{m} Brep_{s_{t_j}}$.

One can easily see that minimum bandwidth is necessary when all the m videos are in the state s_1 *and* also transmitting at the lowest possible rate $Bmin_1$. Similarly, maximum bandwidth is necessary, when all the m videos are in the state s_n *and* transmitting at the peak rate. The representative bandwidth follows from a similar reasoning. The average rates for m videos is simply m times that of a single video. Since the states of the individual videos at any point of time are independent the joint probability of

m-videos being in combined states $(s_{t_1}, s_{t_2}, \cdots, s_{t_m})$ would be $P(m)$, where

$$P(m) = \Pi_{j=1}^{m} p_{s_{t_j}} \tag{1}$$

There are n^m possible combined states of the video system. If $m > n$, by the pigeon-hole principle, two or more videos will have the same individual states. Even when $m \leq n$ some states may appear more than once because states of the videos are independent. Moreover, the number of distinct bandwidths for all states would be far fewer than n^m. In this model all states corresponding to the same representative bandwidth are merged to a single state and two distinct states would require two different bandwidths. This drastically reduces the number of states of the video system especially in situations where m is large.

B.3 Bandwidth Computation without Approximation

The representative bandwidth requirement of the video system consisting of m videos where each could be in any of n states can be given by the multinomial

$$M(p_1, p_2, \cdots, p_n, m, x) = (\sum_{i=1}^{n} p_i * x^{Brep_i})^m \tag{2}$$

The parameters in the multinomial are $p_1, p_2, \cdots, p_n$, and m. The number of terms in $M()$ gives the number of states in the multiplexed video system. The coefficient of x^b, say p_b, in $M()$ gives the probability with which the video system needs bandwidth b. The distribution of bandwidth b is governed by the number of ways in which the m videos can be in the n states. This can be obtained from the number of possibilities in which m_1 of the m videos stay in state s_1, m_2 in s_2, $\cdots$, and the remaining m_n videos stay in state s_n, where $0 \leq m_j \leq m$ and $\sum_{j=1}^{n} m_j = m$. The bandwidth b for this possibility would be

$$b(m_1, m_2, ..., m_n) = \sum_{j=1}^{n} m_j * Brep_j \tag{3}$$

and the associated probability p_b with which this event occurs is given by the multinomial distribution [Ric87]

$$p_b(m_1, m_2, ..., m_n) = \frac{m!}{m_1! m_2! \cdots m_n!} p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n} \tag{4}$$

where $\sum_{j=1}^{n} m_j = m$ and $0 \leq m_j \leq m$.

An exhaustive computation of $M()$ at all possible values of b for a given $p_1, p_2, \cdots, p_n$ and m for finding the bandwidth distribution would have an worst case time complexity $O(n^m)$. This complexity would be unmanageable especially when m is large. We now present a fast approximation technique which can compute the bandwidth without introducing any noticeable error.

B.4 Bandwidth Computation with Approximation

For $m \leq 20$ the computation time for multinomial evaluation is very small and can be integrated with CAC algorithms. For larger values of m the computational overhead would be significant and the multinomial distribution is approximated by first computing exact multinomial for k videos to get the frequency distribution of the bandwidth of k videos, where $k \leq 20$. Two such frequency distributions are now merged to obtain the bandwidth distribution of $2k$ videos. The process is repeated to obtain the distributions for $4k$, $8k$, $\cdots$ $2^{\lfloor \log_2 \frac{m}{k} \rfloor} k$ videos. Note that any two of these distributions, say for $2^i k$ and $2^j k$ videos, can be combined to obtain a single frequency distribution for $(2^i + 2^j)k$ videos. Let q and r be non-negative integers such that $m = qk + r$, where $r < k$. The approximated multinomial distribution for m videos for a given k is finally obtained as follows: q is expressed as a sum of distinct integral powers of 2 (the representation would be unique), say, $q = 2^i + 2^j + \cdots + 2^l$. The distributions for $2^i k$, $2^j k$, $\cdots$, $2^l k$, and r videos are now merged taking two at a time to obtain the distribution for $qk + r (= m)$ videos. Note that since $r < k$ an exact computation can be done for r videos. For example, if $m = 53$ and $k = 5$, we get $q = 10$ and $r = 3$. We now compute the exact multinomial for 5 and 3 videos and approximate multinomials for 10, 20, and 40 videos. We express q as $8 + 2$. Approximated multinomials for $40 (= 5 * 8)$ and $10 (= 5 * 2)$ and the exact multinomial for 3 videos are combined to obtain the distribution for 53 videos. In our study, values of k in the range 2 to 10 have been used in order to reach up to 200 and higher values for m.

B.5 How to Use the Model

The definition of $Brep_i$, the representative bandwidth for state s_i of an individual video, depends on the kind of bandwidth behavior that needs to be captured by the multinomial model. The present work aims at predicting the required bandwidth of the video system where the permissible data loss is low enough to keep visual distortion of received video within tolerable limits. This requires a state bandwidth be represented by its effective upper bound. Similarly, use of $Bmin_i$, $Bmax_i$, $Bavg_i$, or in general any specific percentile bandwidth value of state s_i for $Brep_i$ would reveal other bandwidth characteristics vis-a-vis data loss and video quality. For an example, this research shows a value of $Brep_i$ which the difference value of 99 percentile and mean bandwidths of s_i estimates the average bandwidth requirement well.

III. VALIDATION OF MULTINOMIAL MODEL

The model was validated addressing the need to answer the following questions - (i) How well the bandwidth requirement predicted by the multinomial model compares with observed multiplexed video data rate, and (ii) How the computationally extensive multinomial can be approximated for large values of m and how good the approximation is. To this end, the analytical model was validated by generating multiplexed video system traffic, observing the necessary bandwidth, and then computing the same from the multinomial model. The observed and predicted bandwidths are compared for a wide range of the number of videos multiplexed. A scheme for computing approximate bandwidth distribution of the multinomial model for large m is also given. The study methods used are described below. Specifications of experimental data and numerical results are given in Section IV.

A. Generation of Multiplexed Video Traffic

Suppose, the video system contains m copies of a video V of f_n frames. Two different multiplexing methods are used. In a *random multiplexing*, a random frame number between 1 to f_n in each copy of video is chosen. The data rates corresponding to all m frames are added to get the total bandwidth requirement at the instant. The frame number in each video is then incremented by one to get the next m frames and the data rate. The video is wrapped circularly when its end point is reached. The process is repeated for the desired number of frames. To avoid any bias of the initial frame selection, several such instances with random initial positions for each video are taken and the parameters of interest are averaged over all such observations. A *frame-aligned multiplexing* as opposed to *random multiplexing* is also performed to eliminate any possible effect of the number of I, B, and P frames which might get aligned over time in the multiplexed system. This is similar to *random multiplexing* except that the initial random frame positions are incremented by the fewest possible number of frames so that the i-th of m multiplexed videos starts with the i-th frame type in the underlying IBBPBBIBB$\cdots$ MPEG GOP sequence. Using MPEG-1 traces of commercial movies, multiplexed video traffic are generated for m different videos.

B. Bandwidth Requirement of Multiplexed Video Traffic

The bandwidth distribution of the multiplexed videos is obtained by computing the relative frequency distribution of the multiplexed video bandwidths generated as above over the entire bandwidth interval. The predicted bandwidth distribution of the multinomial model is computed according to the equations in Section II-B.3. The Markov chain transition probability matrix was obtained from the underlying MMG model using $7(=n)$ states for the video.

We used $Brep_i$, as the 99 percentile bandwidth of state s_i to analytically estimate the effective bandwidth required per video while ensuring a low DLR. For an identical DLR we also note the empirically observed bandwidth required per video. We can then show a comparison (see Fig. 1) between the two. The comparison would reveal how effectively the bandwidth requirement at a specific DLR can be analytically estimated and also how the bandwidth requirement decreases with in increase in the number of videos. We have also observed that if $Brep_i$ is chosen as the difference of 99 percentile value and mean bandwidth of state s_i then the observed bandwidth can also be predicted quite accurately.

IV. Experimental Results

A. Setup

Experiments were carried out with MPEG-1 traces of commercial movies including Crocodile Dundee[1], ET[1], Jurassic Park[2], Starwars[2], Terminator II[2], and The Silence of the Lambs[2] [Ros95]. For space constraints only some representative results for Crocodile Dundee are shown. The number of videos multiplexed was varied from 5 to 200. For the range 5-20 the exact multinomial coefficients were computed and beyond 20 those were approximated.

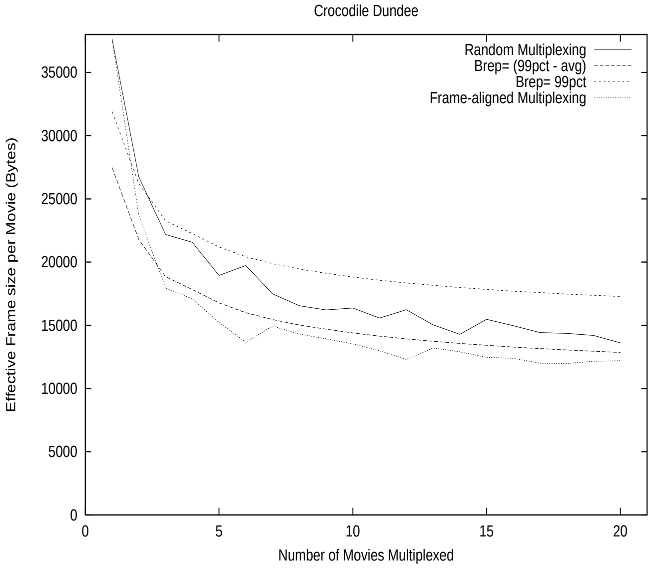

Fig. 1. Analytical Estimates and Observed Bandwidths per Movie

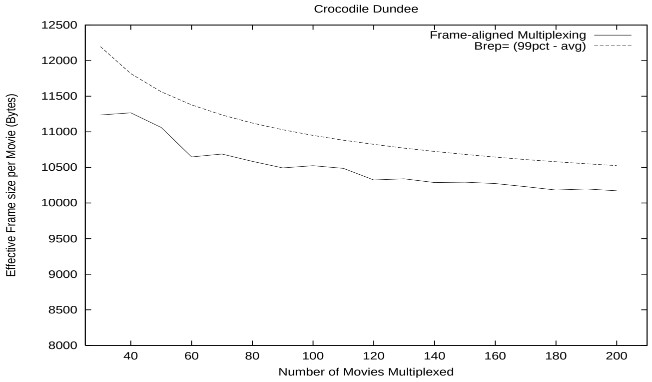

Fig. 2. Analytical Estimates and Observed Bandwidths per Movie

B. Numerical Results

For ease of visualization and to conserve space, we show only a few typical graphs. Three graphs showing accuracy of bandwidth estimation by multinomial model, and accuracy of approximation algorithm for the model are presented. Three plots in Fig. 1 show effective bandwidth per movie from random, frame-aligned, and multinomial model. As expected (and has been observed before

---

[1] The authors thank Wu-Chi Feng[1] and Oliver Rose[2] for the MPEG traces

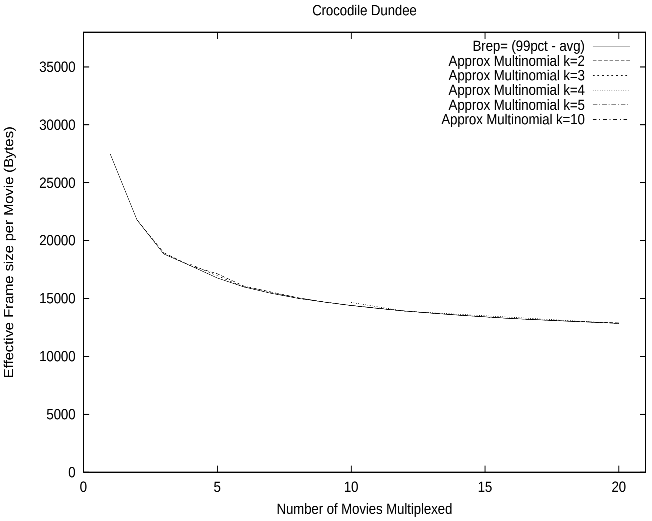

Fig. 3. Comparison of Bandwidth Estimates by Exact and Approximation of Multinomial Computation

[KT97]), frame-aligned multiplexing requires lower effective bandwidth than that of random multiplexing. The multinomial model prediction, being conservative estimates at 99% bandwidth values of the states, predicts a slightly higher bandwidth than observed from either multiplexing method.

Asymptotic effective bandwidth per movie when a large number of movies are multiplexed would be the mean bandwidth of a single movie. Thus, for a large number of movies, the conservative estimates ($Brep$ =99 pct in Figures) from the multinomial model would be higher and need adjustment. We have done this by subtracting from it the difference between this representative bandwidth and the mean bandwidth of a single movie ($Brep = 99pct - avg$ in Figures). Plots in figure 2 show the bandwidth predicted by the multinomial model after adjustment and the observed bandwidth from frame-aligned multiplexing of 20 to 200 movies. It is clear from these plots that the model predicts the effective bandwidth quite accurately while remaining conservative. The error in bandwidth prediction is low and decreases as the number of videos increases.

Figure 3 shows the estimated multinomial bandwidths up to 20 videos computed by exact and approximate means (see Section III-B) when k is varied from 2 to 10 — the plots tend to be indistinguishable indicating high accuracy of the method for useful values of m.

V. Discussion and Conclusions

In this paper, we have proposed a multinomial distribution based model for computing effective bandwidth per video in a multiplexed video system. It uses a finite state Markov chain based model of an individual video. Using our model one can estimate the bandwidth requirement when m copies of a video are transmitted over a channel and a low data loss needs to be guaranteed. The actual channel utilization can also be estimated by the model. The model can be integrated with bandwidth allocation and CAC algorithms for efficient utilization of broadband links.

We have given a theoretical foundation of the proposed model and have presented empirical results using MPEG traces of commercial movies for validating the model. Since exact computation of the model would be computationally expensive for large m, an approximation scheme for using the model is also proposed. The approximation scheme hardly introduces any error no matter how many videos are multiplexed.

In summary, we provide a model for understanding the bandwidth requirement behavior of a broadband video transmission link which would simultaneously carry several videos. For space constraints the model presented in this paper considered multiplexing identical videos — it's generalization for a heterogeneous video system is straightforward [ZRSS02]. The study involving a wide range of multiplexing scenarios is currently under progress.

References

[CR99] K. Chandra and A. R. Reibman. Modeling one- and two-layer variable bit rate video. *IEEE/ACM Trans. Networking*, 7(3):398 – 413, 1999.

[Hey97] D. P. Heyman. The GBAR source model for VBR videoconferences. *IEEE/ACM Trans. Networking*, 5(4):554 – 560, 1997.

[HTL92] D. P. Heyman, A. Tabatabai, and T. V. Lakshman. Statistical analysis and simulation study of video teleconferencing traffic in ATM. *IEEE Transactions on Circuits Syst. Video Technol.*, 2:49–59, 1992.

[Kru99a] M. Krunz. Bandwidth allocation strategies for transporting variable-bit-rate video traffic. *IEEE Communications*, 37(1):40–46, 1999.

[Kru99b] M. M. Krunz. Statistical multiplexing. In *Encyclopaedia of electrical and electronic engineering (ed. John Webster)*, volume 20, pages 479 – 492, 1999.

[KSH95] M. Krunz, R. Sass, and H. Hughes. Statistical characteristics and multiplexing of MPEG streams. In *Proceedings of the IEEE INFOCOM 1995*, pages 455–462, 1995.

[KT97] M. Krunz and S. K. Tripathi. Impact of video scheduling on bandwidth allocation for multiplexed MPEG streams. *Multimedia systems journal*, 5:347–357, 1997.

[MCS99] P. Manzoni, P. Cremonesi, and G. Serazzi. Workload models of VBR video traffic and their use in resource allocation policies. *IEEE/ACM Trans. Networking*, 7(3):387 – 397, 1999.

[Ric87] J. A. Rice. *Mathematical Statistics and Data Analysis.* Wadsworth and Brooks, 1987.

[RK98] Q. Ren and H. Kobayashi. Diffusion approximation modeling for markov modulated bursty traffic and its applications to bandwidth allocation in ATM networks. *Journal on Selected Areas in Comm*, 16(5):679–691, 1998.

[Ros95] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. In *Proceedings of the 20th annual conference on local area networks*, pages 397 – 406, 1995.

[SRS01] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar. Modeling full-length video using markov-modulated gamma-based framework. In *Proceedings of IEEE Globecom 2001, Multimedia QoS/Video Applications*, pages 1 – 5, 2001.

[SRS03] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar. Modeling full-length video using markov-modulated gamma-based framework. *IEEE/ACM Trans. Networking*, 11(4):638 – 649, 2003.

[WCJ95] J.-L. Wu, Y.-W Chen, and K.-C. Jiang. Two models for variable bit rate MPEG sources. *IEICE Trans. Commun.*, E78-B:737–745, 1995.

[ZRSS02] W. Zhou, S. Ramakrishnan, D. Sarkar, and U.K. Sarkar. Bandwidth estimation for multiplexed videos using an MMG-based traffic model of a single video. Technical Report CSC-Feb25-02, Department of Computer Science, University of Miami, Coral Gables, FL 33124, 2002.