

A Practical Model for VBR Video Traffic with Applications

Tamás Borsos

Traffic Analysis and Network Performance Laboratory,
Ericsson Hungary, Laborc u. 1, 1037 Budapest
`Tamas.Borsos@eth.ericsson.se`

&

Budapest University of Technology and Economics,
Dept. of Computer Science and Information Theory,
Budapest, Hungary

Abstract. Video traffic compressed with variable bit rate coding scheme is known to possess high variations and multiple time scale characteristics. This property makes parsimonious video modeling a difficult task. A possible way of describing this traffic is via self-similar models, which also produce high variations on many time scales. However, these are general traffic models and do not represent many important characteristics of video. In this paper we show that video traffic has well-separable time scales. Based on this result, a new model is presented, which is capable of capturing the main properties of VBR video. The concept is scene-oriented, while a larger time scale - called epoch - is introduced. The main contribution of this paper is that the presence of multiple time scales seem to be the real reason for the slowly decaying autocorrelation function rather than heavy tailed level durations. Finally, the application of the model is shortly discussed for dimensioning, admission control and simulation purposes.

1 Introduction

The emerging multi-service architectures speed up the widespread adoption of video applications in communication networks. Streaming video – in which viewers can begin watching content almost as soon as it starts downloading – is now developing as a mainstream technology on the Internet. Wireless networks like GPRS and UMTS will offer mobile users a convenient way to access these services. However, the efficient support of high quality video services over packet switched wireline or wireless channels is still a challenging task.

Modern effective coding standards use hybrid coding (e.g., MPEG, H.263, H.263+) which comprises lossy intraframe coding and motion compensation to exploit both spatial and temporal redundancy. In this paper MPEG-4 traces are analyzed, since - due to the similarity of the coders - the results can be generalized straightforward to MPEG-2 or H.263 traffic and they also serve as a basis for understanding the bandwidth requirement of compressed video streams.

There are different coding schemes that can be used depending on the actual application (e.g., constant bit rate, variable bit rate, adaptive rate coding). This paper is concerned with the variable bit rate (VBR) scheme, which provides the best compression for given image quality. Although, from the point of view of performance analysis, it is not easily tractable due to its high variability on many time scales.

There are a number of different approaches that can be found in the literature. In the discrete autoregressive DAR(1) model [10] a finite-state Markov chain is used to generate the sequence of the number of bits associated with each video frame. This process stays at a constant level for geometrically distributed period, then steps onto another level independently of the current state. The transition matrix of this Markov chain is composed of identical rows equal to the marginal frame size distribution. The use Markov Modulated Process [17] is similar to the previous approach, but it is extended by state-dependent transitions and varying bit rate during the scenes. Clearly, this model requires quantized average scene bit rates. In addition, the number of quantization levels can not be large because all of the elements of the transition matrix must be estimated from the trace. Another type of models [6], [11] capture the marginals and the autocorrelation function (ACF) of a video trace. In this case, video traffic is modeled as a self-similar process. The frame (or group of pictures) size distribution can be arbitrary (in [6] it is a mixture of Gamma and Pareto distributions), while the correlation structure is caught through the Hurst parameter. Thus, beside the marginal distribution, only the Hurst parameter needs to be estimated, which determines the decay rate of the autocorrelation function. The shifting level process (SLP) (also called space-time renewal process) is a very simple model [5], [16], [7], [8]; it consists of constant bit rate levels, which may represent the scenes. Both the mean and the length of these levels are independent, identically distributed random variables. By choosing the length distribution appropriately, any ACF structure can be captured [12]. Thus, the process is completely described by two distributions. In [14] the compound ACF behaviour is captured by a special family of functions, the stretched exponential, instead of polynomial (long range dependence) or exponential (Markovian). The video model with such a correlation structure is based on $M/G/\infty$ input process, where the actual bit rate is proportional to the number of customers being in the $M/G/\infty$ system. By varying G many forms of time dependence can be displayed, while it affects the bit rate marginal distribution only through its mean. The marginal will be Poisson distribution, which is further transformed into the appropriate frame size distribution.

The temporal behaviour of a VBR coded video has always been the most debated question in the literature. Its bandwidth requirement is often regarded as a long range dependent (LRD) process because of its slowly decaying autocorrelation function. Thus, traditional short range dependent Markovian models may fail to predict QoS parameters in certain circumstances [6]. In [18] the use of LRD models is questioned in the context of realistic traffic engineering, i.e., using short network buffers.

A simple Markovian model is indeed able to capture only one time scale. The scene level characteristics of video is a well known property, which also has an obvious physical meaning (i.e., scene cuts in a movie). However, the Markov chain driven scene levels are still not descriptive enough for the whole video trace. Therefore, self-similar models are used, which are able to capture infinite number of time-scales. In that model the marginals and autocorrelation functions are properly matched to the video trace. This approach is not necessarily effective due to the following reasons:

- there is no physical meaning of the parameters and the model provides no support about how to change them for different video traces and coding standards;
- without the actual trace there is no way of estimating the parameters;
- very different video traces may have similar models;
- when estimating the ACF the traces were assumed to be stationary which may not be true [15];
- second order statistics may be irrelevant from the queueing point of view [12,2].

In the light of these criteria, we propose a model in which the time scales of VBR video are explicitly represented, having a physical meaning for each of them. The model is scene-oriented, but a larger time scale – called *epoch* – is introduced. This result suggests that the large tail of the ACF is not necessarily due to the heavy tailed level durations (e.g. scenes), but seems to be the sign of *large time scale level shifts*.

The paper is organized as follows. In Section 2 the structure of the examined MPEG video traffic is described. Section 3 contains the scene level analysis, while in Section 4 the epoch concept is introduced and validated. In Section 5 we look at three application areas of this model, i.e., dimensioning, admission control and simulations. Section 6 concludes the paper.

2 MPEG Video Structure

The modeling approach in this paper is intended to describe the common properties of all of the widely used *hybrid coding* standards, rather than a specific coding standard. Hybrid coding, like MPEG or H.263, comprises *lossy intraframe coding* and *motion compensation* to exploit both spatial and temporal redundancy. The subject of the actual investigation is a set of VBR MPEG-4 traces taken from [4]. Similar results were obtained for the traces of Rose [17].

A standard MPEG encoder generates three types of compressed frames. I-frames are compressed using intra frame information only. P-frames are coded similarly to I-frames but with the addition of motion compensation with respect to the previous I- or P-frame. B-frames are similar to P-frames except that the prediction is bidirectional. Typically I-frames are the largest followed by P-frames, while B-frames require the lowest bandwidth. After coding, the frames are arranged in a deterministic order, which is called group of pictures (GOP,

e.g., 'IBBPBBPBBPBB'). The GOP pattern is not specified by the standard and coders may use different patterns for subsequent GOPs. However, since many sequences are being coded with regular GOP patterns (often to simplify the codec design), developing a traffic model for such sequences has its merits.

There are two main levels of modeling: *frame level* and *GOP level*. The former attempts to catch the size of each frame, while the latter takes a GOP as a whole and is not interested in the individual frame sizes. In order to eliminate the deterministic alternation of frame types, this paper is concerned with the GOP level, thus making the model independent of the MPEG coding scheme. The model can then be further refined to frame level.

3 Scene Level Analysis

It has been observed by many authors (e.g., [10], [12]) that the fluctuation of frame sizes exhibit different behaviour at different time scales. Within small locality, I-frames have relatively small fluctuation about some average level, which itself varies at larger time scale. These *level shifts* are usually attributed to *scene changes* [5]. The scene can be defined as a portion of video without sudden changes in view. Such behaviour can be captured using a mean-nonstationary time series or some kind of modulated process. These models are called *scene-based models*. Incorporating the 'scenic' component in the traffic model gives its predictions a physical meaning.

A simple implementation of such a model is the Shifting Level Process (SLP), which consists of constant bit rate levels. The bit rate Y_i and the duration T_i of these levels are independent, identically distributed random variables. The GOP size is clearly distributed according to Y . It has also been shown [12] that the ACF ρ_k is exactly equal to the integrated tail of the renewal distribution function F_T . That is

$$\rho_k = 1 - \frac{1}{m} \sum_{u=0}^k (1 - F_T(u)),$$

where m is the mean scene length. One can see that, despite the simple structure of the SLP, it provides an arbitrarily exact match of the first- and second-order statistics. Thus, it is able to match subexponential ACF and to capture long range dependence. The model can be extended to involve bit rate variations during scenes. This modification has little impact on the tail characteristics of the ACF. According to this reasoning the slow decay of the ACF is due to the heavy tailed scene length distribution. However, our investigations show that the scene durations follow Weibullian distribution with a mean of 2-3 seconds.

In order to obtain scene statistics one has to find scene boundaries in a trace. To make the estimation of the distribution robust, two different methods were used for this purpose. The first one was performed with change point detection using the trace data. In this case, the well known intra-coded Star Wars movie was used, since this coding scheme provides more reliable basis for scene change detection. In the second case, the scene identification was performed on the basis

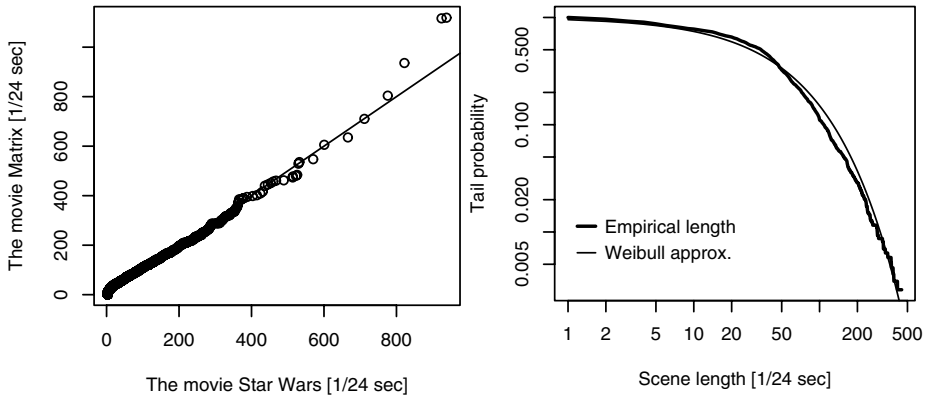


Fig. 1. Q-Q plot for the two different types of scene length estimates and the scene length distribution with a Weibullian approximation on a log-log scale

of a real MPEG source (the movie Matrix) instead of its trace. The frames were decoded and the difference of the luminance bitmaps of every two consecutive frames was stored. This is a visual identification and is supposed to be more reliable than the frame size based method. Figure 1 shows that the character of the two empirical distributions are surprisingly similar, despite that two different movies were processed by two different methods. This validates the obtained result. The tail of the scene length distribution can also be seen in Figure 1 with its Weibullian approximation.

The fact that scene bit rates are not independent can be easily checked: by shuffling the scenes in a video trace the time dependence structure is destroyed. Grasse *et al.* [7] suggest that the Y_i and T_i processes should be modeled by fractionally differenced white noise passed through an ARMA(1,1) filter. This solution, however, results in a general traffic model beyond the scene level, with little physical meaning of the filter parameters.

4 Epoch Time Scale

After extensive statistical and graphical tests [1] concludes that video traffic exhibits long range dependence, which is very important from the point of view of queueing behaviour. When calculating the empirical autocorrelation function of several traces, one may see that the tail behaviour is really far from exponential.

Figure 2 shows the empirical ACF of the Star Wars trace. It can be seen [6] that the initial part of the curve can be accurately matched to an exponentially decaying function, but only up to 5-10 sec lag. Beyond that it decreases slower than exponentially, up to c.a. 40 sec lag. It then adopts a quite erratic behaviour decaying toward zero but extremely slowly. This composite shape of the ACF (i.e., exponential start and slow decay later) is a property of many other video



Fig. 2. Empirical ACF of the Star Wars trace (intra frame coded)

traces, as well, suggesting the property of long range dependence. This long time dependency is often explained by the heavy tail of scene lengths (e.g., Pareto distribution). Although our statistical analysis reinforces the subexponentiality of scene length duration, it does not seem the main cause of the large tail of the autocorrelation function, since the tail of the ACF is not simply slowly decaying but remains high (~ 0.4) for large time lags, beyond the scene time scale.

In our concept the complex behaviour of the ACF is *due to the presence of multiple time scales instead of heavy tailed level durations*. Analyzing the VBR video traces one can observe that the level shifts, which are the characteristics of the scenic behaviour, appear on a larger time scale, as well (Figure 3). These long term level shifts constitute the modulating process of the scene level process. These larger time scale levels – we called them *epochs* – are introduced, which

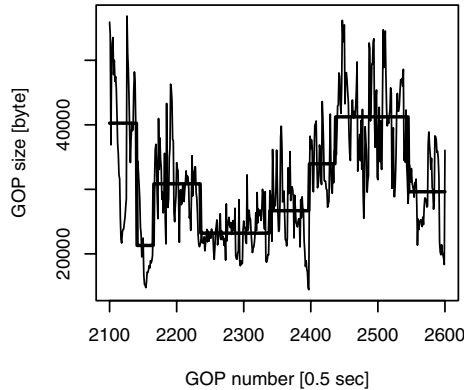


Fig. 3. Level shifts on large time scale in the GOP trace of the movie MrBean

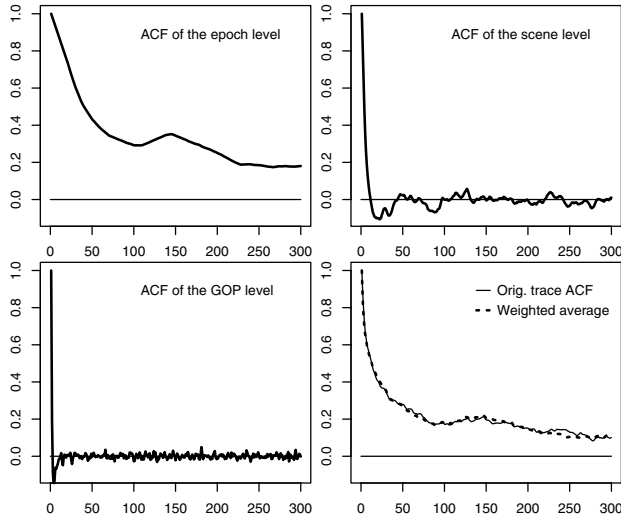


Fig. 4. The ACF of the epoch, scene, GOP levels and the ACF of the trace (GOP numbers, up to 150 sec)

serve as an explanation for the large tail of the ACF. Therefore, our model has 3 level characteristics (the index i refers to the GOP number):

- GOP size variation during a scene N_i ;
- level shifts due to scene changes S_i ;
- level shifts due to epochs (a group of scenes) E_i .

S_i and N_i are zero-mean processes. The GOP trace T_i is then simply the sum of the three level processes: $T_i = E_i + S_i + N_i$. Note that this concept contains no specific information about the size or duration distributions, or about the correlation structure of the individual level processes. Moreover, *long range dependency is not a requirement to produce the desired ACF function*.

The hypothesis that N_i , S_i and E_i are independent is not completely true, but it proved to be a reasonable modeling assumption. The weighted average of their ACFs gives the ACF of the whole trace, as can be seen in Figure 4. This structure explains the quickly decaying start (due to S_i and N_i) and the slowly decaying tail (due to E_i), which has also been observed by several authors [6]. The stochastic behaviour beyond some time lag may be the result of non-stationarity, i.e., the epoch process varies too slowly to be evaluated from a "short" 1 hour trace.

The concept of defining epochs in a movie has some practical rationale, since consecutive scenes may be filmed in a similar environment and background. Thus, this modeling approach is based on the real traffic behaviour, and it is probably easier to generalize to other coding schemes. On the other hand, level shifts on the large time scale mean sudden jumps in bandwidth requirement

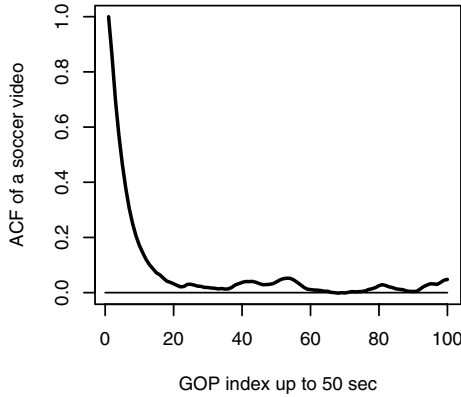


Fig. 5. The ACF of a "stationary" soccer video

instead of slow changes. This can be a significant property in adaptive environments, like, e.g., TCP background traffic, WCDMA or measurement based admission control. Finally, the three-level model is supported by the fact that movie traces always possess large ACF tail (up to even 400 sec lag), while those TV broadcasts that contain "stationary" environment, like tennis or football, have significant tail only up to ~ 20 sec lag, as can be seen in Figure 5. This phenomenon indicates that the epoch time scale in this video type is missing. This means that the distribution of the epoch mean is already meaningful and can be used as a video traffic descriptor. (One can also see that the negative correlation values of Figure 4 have also disappeared in Figure 5, since these were the result of making empirical averages for the level processes.)

The next step is characterizing the three level processes. As mentioned above, the marginal distribution of the epoch level E_i is the main descriptor of a VBR video. On the other hand, the temporal behaviour of E_i is of no importance at all in most of the applications. In particular, it is not interesting from the queueing point of view, since this time scale is too large to be smoothed in a buffer of reasonable size [9].

The scene level process S_i can already be smoothed in case of streaming video applications (e.g., video on demand or Internet streaming applications), where a delay of 10-20 sec is allowed. We have not developed a parsimonious model for this process yet, but a short example shows that despite the Weibullian scene length distribution, a Markovian model may be an appropriate description. The process S_i is quantized to 50 levels and a 50 state Markov chain is fitted to S_i (note that it is not for modeling purpose, only for demonstration). The queue tail distribution of the original trace S_i and a randomly generated trace S_i^* is illustrated in Figure 6 for different service rates on logarithmic scale.

The GOP level process N_i is similar to S_i but appears on a smaller time scale. It should be accounted for in case of low-delay video applications.

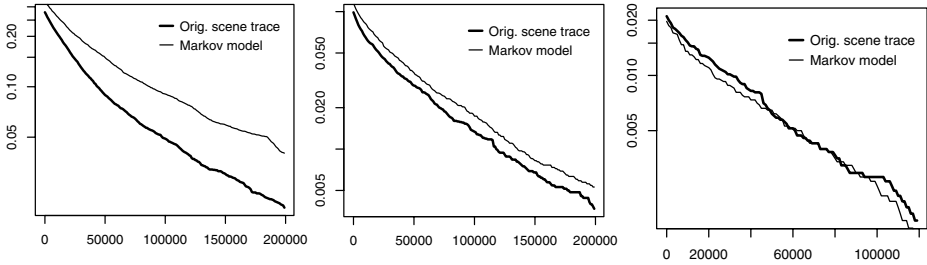


Fig. 6. The queue tail distribution of S_i and a Markovian equivalent S_i^* for different service rates (increasing from left to right) on logarithmic scale. S_i^* is usually an upper bound.

5 Applications

Dimensioning for streaming video. In this application the GOP time scale process N_i is not necessary to appear in the model, since its average is approximately zero on 4-5 sec interval, while the smoothing buffer is typically larger. The epoch process E_i affects the bandwidth requirement only through its marginal distribution. After some analysis of the 20 available video traces the Gamma distribution seems to be a good conservative approximation (see Figure 7). For the estimation of the traffic description parameters a large video database should be used. The required capacity can be calculated separately for the epoch level processes (with bufferless multiplexing) and the scene level processes (effective bandwidth with a given buffer [13]).

Admission control for streaming video. Based on the above reasoning a robust admission control method can be constructed [3]. While a simple measurement based admission control (MBAC) is not a reliable method for VBR video traffic due to its long term variations, a compound, measurement based and parametric admission control is already suitable. While the bursts coming from the scene time scale (and GOP) bit rate variations can be buffered and thus, its temporal behaviour must be accounted for by the measurement based scheme, the large time scale variation affects the bandwidth requirement only through its marginal distribution. The two time scales can be separated in real-time by applying moving average. Thus, the algorithm is measurement based on small time scales and parametric on the large time scale.

Video traffic simulations. With the proposed model it is possible to generate video traces for simulations that have similar characteristics to the real video traffic. The main impact of the model on packet level simulations is that the large time scale variation actually appears as level shifts. This causes a sudden jump in the bandwidth requirement occasionally. These level jumps may have serious effects on the performance of adaptive background applications, like TCP or adaptive video. In these applications, besides the bandwidth oscillation, it also causes bursty packet losses.

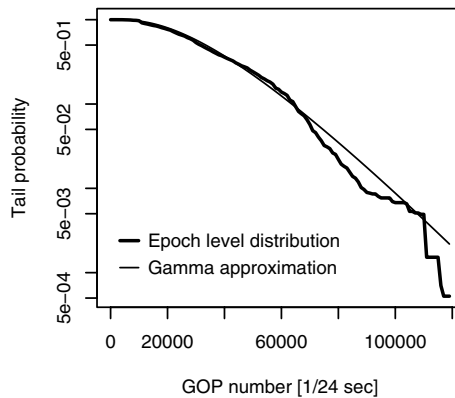


Fig. 7. The marginal distribution of epochs with its Gamma approximation

6 Conclusion

VBR video traffic is known for its high correlation on large time scales. In this paper we argued for multiple time scale video models instead of long range dependent models. We described a modeling approach which explicitly captures the time scales of video traffic. The model is scene-oriented, but large time scale level shifts are introduced, called epochs. This time scale process has a physical meaning and can be used for parsimonious characterization of video traffic. The model also explains the compound structure of a typical video traffic auto-correlation function without the use of long range dependent models. Although the model has a composite structure, certain time scales are irrelevant in the performance analysis, thus, it remains parsimonious for a given application. In particular, the time scales beyond the epoch levels and their temporal behaviour is of practically no importance. Examples were also shown that the model can be used in a wide range of applications.

Acknowledgment. The author wishes to thank Prof. László Györfi and András György for useful comments. He would also like to thank the Telecommunication Networks Group at the Technical University of Berlin for making their video traces publicly available.

References

1. J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Communications*, 43:1566–1579, 1995.
2. T. Borsos, L. Györfi, and A. György. On the second order characterization in traffic modeling. Proc. 9th IFIP Working Conference on Performance Modelling and Evaluation of ATM and IP Networks, pages 41–49, July 2001.
3. T. Borsos and A. György. Robust admission control for video traffic. Preprint.

4. F. H. P. Fitzek and M. Reisslein. MPEG-4 and H.263 video traces for network performance evaluation. Technical Report TKN-00-06, Technical University of Berlin, Telecommunication Networks Group, 2000.
5. 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. *IEEE Trans. Circuits and Systems for Video Technology*, 4:521–526, 1994.
6. M. W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar VBR video traffic. In *Proc. ACM SIGCOMM '94*, pages 269–280, Aug. 1994.
7. M. Grasse, M. R. Frater, and J. F. Arnold. Implications of non-stationarity of MPEG2 video traffic. *COST257TD(97)10*, 1997.
8. M. Grasse, M. R. Frater, and J. F. Arnold. Origins of long-range dependence in variable bit rate video traffic. In *Proc. ITC 15*, pages 1379–1388, June 1997.
9. M. Grossglauser and J. C. Bolot. On the relevance of long-range dependence in network traffic. *Computer Communications Review*, 26:15–24, 1996.
10. D. P. Heyman and T. V. Lakshman. Source models for VBR broadcast-video traffic. *IEEE/ACM Trans Networking*, 4:40–48, 1994.
11. C. Huang, M. Devetsikiotis, and I. Lambadaris A. R. Kaye. Modeling and simulation of self-similar variable bit rate compressed video: a unified approach. In *Proc. ACM SIGCOMM '95*, pages 114–125, Aug. 1995.
12. P. Jelenkovic, A. A. Lazar, and N. Semret. The effect of multiple time scales and subexponentiality in MPEG video streams on queueing behavior. *IEEE Journal on Selected Areas in Communications*, 15:1052–1071, 1997.
13. F. P. Kelly. *Stochastic Networks: Theory and Applications*, chapter Notes on effective bandwidths, pages 141–168. Oxford University Press, 1996.
14. M. Krunz and A. Makowski. Modeling video traffic using M/G/infinite input processes: a compromise between Markovian and LRD models. *IEEE Journal on Selected Areas in Communications*, 16:733–748, 1998.
15. J. T. Lewis, N. O'Connell, R. Russel, and F. Toomey. Predicting QoS for traffic with long range fluctuations. Technical Report DIAS-STP-94-31, Dublin Institute for Advanced Studies, Applied Probability Group, 1994.
16. B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski. TES-Based video source modeling for performance evaluation of integrated networks. *IEEE Trans. Communications*, 42:2773–2777, 1994.
17. O. Rose. Simple and efficient models for variable bit rate MPEG video traffic. Technical Report TR No. 120, Institute of Computer Science, University of Wurzburg, 1995.
18. B. K. Ryu and A. Elwalid. The importance of long-range dependence of VBR video traffic in ATM traffic engineering: myths and realities. *ACM Computer Communication Review*, 26:3–14, 1996.