

Video Traffic Modeling for Resource-Constrained Clients

Roland Osborne, John Villasenor
Electrical Engineering Department
University of California, Los Angeles
Email: roland@icsl.ucla.edu, villa@icsl.ucla.edu

Abstract—We present a new approach to variable bit rate (VBR) video traffic modeling designed to capture the multi-scale characteristics that are critical in determining buffering requirements. Much of the previous work in video traffic modeling has focused on real-time multimedia systems that of necessity involve queuing models with small buffers. Today there is growing interest in understanding systems in which high quality video is downloaded in non-real-time from a source such as the Internet, and then distributed at a later time in a local environment to a very low-cost and thus resource-constrained device. Smooth playback and avoidance of data loss are of paramount importance. The work here analyzes the tradeoffs between buffering resources and the video quality in such environments, and provides a framework that is directly relevant to a large class of emerging applications and systems.

I. INTRODUCTION

The importance of variable bit rate (VBR) video traffic is growing rapidly over a wide variety of high speed networks. VBR video places significant demands on the elements of video distribution systems that include storage, transmission, decoding and display. In all such systems one of the main design goals is to maximize the quality of the displayed video while minimizing the burden on network and hardware resources. Buffering plays a central role in this tradeoff. If buffers are not properly sized, overflow or underflow can create video disruptions. Liang [1] developed encoding techniques for minimizing the effect of these errors. Larger buffers can also reduce or eliminate this problem but at the cost of memory. Traffic models that accurately reflect local and non-local time variations in instantaneous video data rate are crucial to understanding buffering requirements. These buffering requirements are critical in determining system cost, particularly in many recent and emerging low cost consumer electronic applications.

Video traffic modeling has been an issue for many years. It has been considered extensively in the literature and utilized in the design of a large range of commercially deployed systems. These systems including DVD players, direct broadcast satellite, and real time video streaming applications. Thus, traffic characteristics as they impact the design of systems for these environments are quite well understood. Very recently, however, there has been significant growth in very inexpensive video display devices in which there are particularly severe manufacturing cost constraints, resulting in severe hardware resource limitations. These devices are often

designed to receive video from a nearby wireless server, and must utilize a minimum of memory and processing power. While minimization of memory and processing power is also a limitation on devices such as DVD players, in a DVD player the tasks of reading and then decoding the compressed data are performed inside a single device. By contrast, in the environment under consideration here, the compressed data is delivered to the resource-constrained client over a bandwidth-limited and potentially lossy link. Furthermore, in contrast with DVD content, low-cost video display devices are called upon to receive and decode content covering a large range of video parameters, most of which was encoded without regard for the limitations imposed by these systems. Although the emerging remote clients are resource constrained, and buffers should be kept as small as possible, they can still be large enough to enable storage of several seconds or more of data. This contrasts with real-time two-way systems such as video-conferencing which targets delays, and therefore maximum buffer sizes. Delays are typically kept to tens or hundreds of milliseconds.

In the present work we present a framework for characterizing buffering requirements in light of the wide variety of content and distribution systems. We then describe a new VBR traffic model that specifically aims to capture the multi-scale characteristics that impact design of resource-limited display devices. The model and associated framework are validated using a set of several hundred video sequences each with a duration of about 45 minutes, and reflecting a wide variety of content. The rest of this paper is organized as follows: Section II discusses previous work on VBR video models and places the new model in context. The buffering metric and VBR traffic model are described in sections III and IV respectively. Experimental results are contained in Section V and conclusions in Section VI.

II. BACKGROUND

Much of the previous work on video traffic modeling has been performed in the context of networks delivering real time content and characterized by small buffers. Thus, many studies have emphasized the short range dependence (SRD) (i.e. the short range correlations in the instantaneous data rate) over the long range dependence (LRD) of video data. Examples of well known SRD models include GBAR [2] and DAR [3], which aim to simulate the autocorrelation and distribution of VBR

video teleconferencing traffic. LRD has also been examined by researchers including Beran et al., who noted in [4] that LRD is an inherent characteristic of general VBR video traffic. Ahn [5] described isolating the effects of SRD and LRD on queuing behavior when transmitting a full movie and determined that LRD is especially critical at large buffer sizes. LRD models allow for the characterization of more general video traffic behavior. Huang [6] developed a multi-scale model capturing a range of dependencies. Since the average bit rate of a movie is generally higher during scenes with high motion, accurately modeling the behavior at a scene level (where a “scene” is a segment with statistically consistent content) requires an LRD model. Krunz [7] observes that as the buffer size increases, the effects of SRD on the performance diminish, and the average bit-rate over a wider range in time will play a greater role in the required buffer size.

Typical metrics used in validating the models include the autocorrelation function, frame size distribution, and ATM queuing performance [2], [8], [6], [9]. While these metrics are appropriate for the environments they address, they say little about the behavior of other systems such as home or other local area networks involving different quality of service requirements and larger buffers.

We utilize an alternative approach to traffic modeling based on spatial renewal random processes (SRRP). These are random processes that can be divided into segments in which both the segment duration and the average value during that segment are drawn from samples of i.i.d. random variables. In the context of video content modeling, this involves partitioning a video sequence into statistically appropriate segments, and then utilizing an appropriate model within segments. Liew et al. [10] use an SRP based model for characterization of two full feature films. Lucantoni [11] also uses a renewal process for VBR video traffic modeling.

III. BUFFERING METRIC

Many emerging systems involve a powerful content server streaming a unicast audio, video and overlay stream to a resource-limited display client. The link speeds between the client and server can vary greatly. A wired Ethernet connection can provide hundreds of Mbits per second of actual throughput while two hops on an 802.11G network (i.e. from a server to an access point and then to a client) limits the actual TCP throughput to 7 Mbits/sec. In addition, while the average bit rate of the content distributed in such systems must be less than the link capacity between the client and server, there are frequent bursts in data rate that can lie well above the link capacity. The client display device must provide a buffer to guarantee smooth reliable playback of the content. The amount of buffering required, in turn, is dependent on the network environment, nature of the content, and the tolerance for potential display disruptions due to buffering problems.

In the present discussion, we assume a fixed link capacity between the client and server and consider only video, as it consumes most of the overall bandwidth in a multimedia stream. The client operates by filling its buffers and then

beginning playback at the variable content data rate. Each frame period the client draws a variable amount of data from the buffers and is able to store an amount of data equal to the frame period multiplied by the link capacity, provided that buffer overflow is avoided. The required buffer size is the minimum needed in order to prevent buffer underflow during playback. We express buffer size in terms of number of seconds of average video data rate and link capacity as a multiple of the average video data rate. This normalizes both to an underlying attribute of the content and system under consideration, and readily allows application across a wide range of network environments and video standards.

Fig. 1 provides a plot of frame size, measured in bytes after compression, for a typical video sequence. This sequence has a total duration of approximately 45 minutes, and contains a total of 61,210 frames at a frame rate of 24 frames/sec. The first 30,000 frames are displayed in the plot. The compression used was MPEG-4 using the DivX 5 codec (www.divx.com), with an average data rate of 1 Mbit/sec. The video trace contains many bursts in data rate. For our purposes we will define a burst as beginning when the client’s buffer begins to get depleted to the point where it has reached its minimum. A single dominating burst is responsible for determining the required buffer size at any given link capacity. Fig. 2 illustrates two bursts in a video trace. Each burst imposes a minimum value for the client buffer size equal to the area above link capacity. The buffering requirement is largest with Burst-A at low link capacities and largest with Burst-B at high link capacities. At a given link capacity the required buffer size for playback of the full trace is therefore determined by the maximum of all bursts.

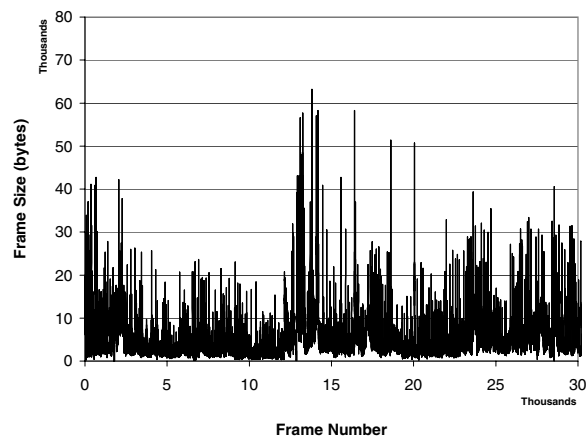


Fig. 1. Example video trace from sample set.

Fig. 3 plots the relationship between link capacity and buffer sizes for the sequence in Fig. 1. For each value of link capacity, the buffer size is the minimum amount of memory needed to ensure that there will be no underflow events in which the decoder in the client device attempts to read coded video data that has not yet arrived in the buffer. The plot shows, for example, that for this video sequence delivered over a 3 Mbits/sec link, which is typical of what would be

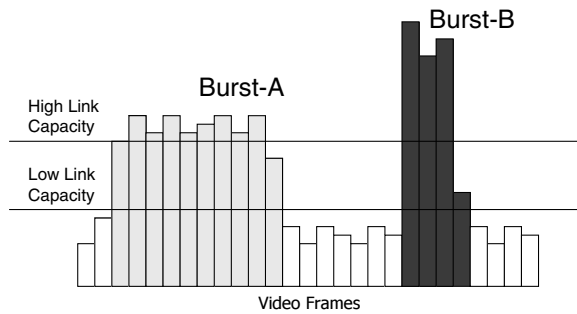


Fig. 2. Video trace illustration with two bursts in data rate.

encountered on an uncongested 802.11B network with two hops, a buffer size of approximately 5.3 Mbits, or equivalently, 0.7 Megabytes would be required.

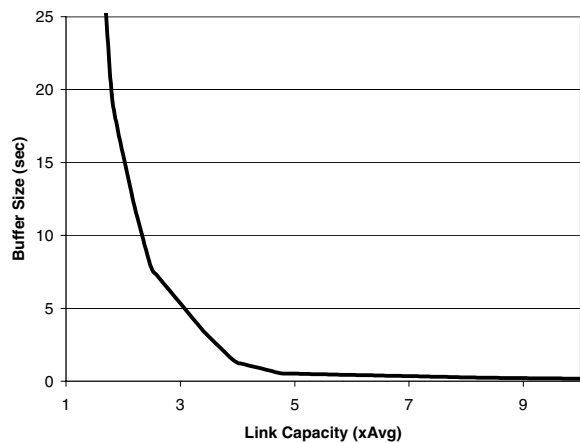


Fig. 3. Minimum buffer size vs. link capacity. Buffer size is expressed in seconds at the average video rate; link capacity is expressed as a multiple of the average video bit rate.

The piecewise linear structure of the curve in Fig. 3 is a consequence of the nature of the video data, and more specifically of the fact that buffer limits are determined by specific high-rate burst events in the video data. The abrupt transition in slope corresponds to a transition in dominating bursts within the video trace. This is illustrated more specifically in Fig. 4, which shows the negative of the derivative of Fig. 3 and more clearly illustrates the linear segmentation. The slope values in Fig. 3, which become horizontal lines in Fig. 4, give the duration of the burst that furnishes the buffer limitations. For example, for a link capacity at 3 times the average content data rate, the buffer plot has a slope of -4.7 . The limits at this link capacity are therefore furnished by a burst lasting 4.7 seconds. Viewed another way, the slope indicates the rate at which buffer size requirement is reduced for a given increase in link capacity. It also indicates that the burst leading to the slope value of -4.7 would deplete the buffer over a period of 4.7 seconds.

As noted above with reference to Fig. 3 and given the content was encoded at 1 Mbit/sec, a link rate of 3 Mbits/sec requires a 5.3 Mbit buffer. Thus, over 4.7 seconds the system

needs the capacity to handle $3 \times 4.7 + 5.3 = 19.4$ Mbits, corresponding to an average rate of 4.1 Mbits/sec. This is much higher than the average content data rate of 1 Mbit/sec. For link capacities above 8.6 Mbits/sec, the required buffer size is determined by a single frame. This can be seen in Fig. 4 where the burst governing buffer size has a duration of 0.042 sec, which is the duration of a single frame in a 24 frames/second video sequence.

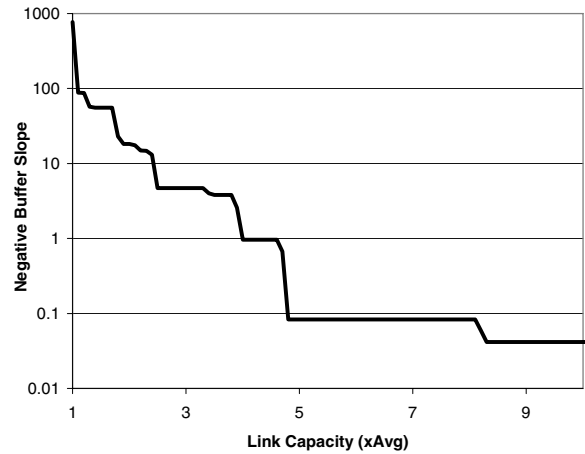


Fig. 4. Negative Slope of Fig. 3.

As seen in Fig. 4 the dominating bursts range from close to 100 seconds at low link capacities to fractions of a second at high link capacities. A video traffic model therefore needs to account for the multi-scale behavior of video data in order to accurately characterize the behavior at a wide range of link capacities.

IV. TRAFFIC MODEL

The above results identify multiple levels of dependency governing the buffering requirements for VBR video transmission. Long range dependencies are realized in bursts lasting up to several thousand frames, while shorter range bursts last only several frames. To capture this wide range of dependency, we utilize a model composed of several levels of a spatial renewal random process (SRRP). The highest level generates segments with the longest duration. The average bit rate and duration of these segments are drawn from an independent and identically distributed (i.i.d.) random variable. At the next level, another SRRP partitions each long duration segment into segments of a shorter duration. Again, the average bit rate and duration are drawn from an i.i.d. random variable. This recursive division continues until all dependency levels of interest are generated.

The parameters, or random variable distributions, of each SRRP level can be determined from a set of realizations (plots in the form of Fig. 1). Each dependency level must be first identified by applying a windowed average to the original data trace. Consider Fig. 5 where a windowed average is applied to a video trace at two different window sizes. The larger window size retains less of the variability, and the remaining peaks and valleys are due only to the longer dependency

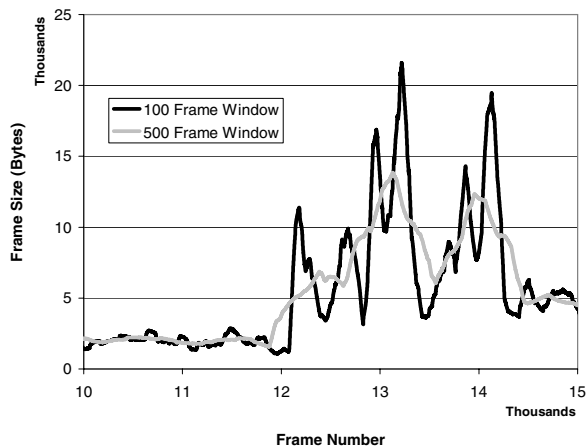


Fig. 5. Windowed averages of a video trace.

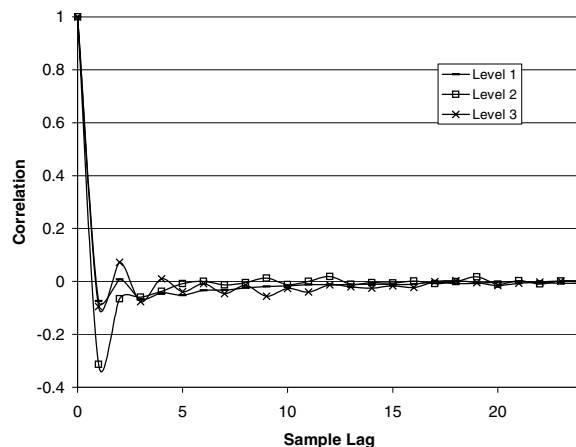


Fig. 7. Autocorrelation of the data.

behavior. A windowed average with a large window size is therefore generated for parameter extraction of the highest level. All frames are grouped with the nearest identifiable peak, plateau or valley, where a distribution of the average value and duration can be estimated. Fig. 6 is an example of an SRRP level estimated from a window size of 240 frames. The estimated SRRP level is then subtracted from the original trace, and the next level's parameters are estimated from the difference. This recursive extraction continues for all dependency levels of interest.

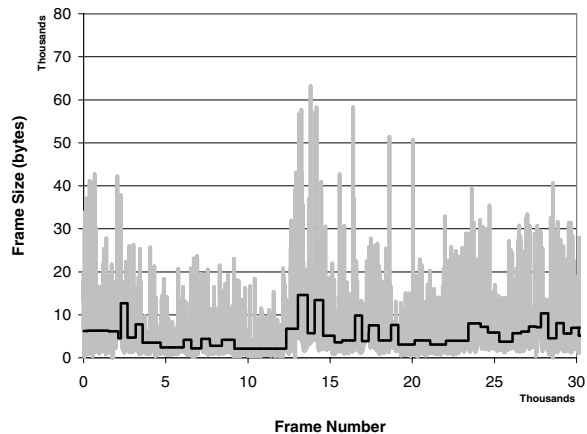


Fig. 6. SRRP level estimated from Fig. 1 with a window size of 240 frames.

Each SRRP level assumes appropriately randomized segment duration and average values. Each window size is therefore selected to be the largest value from which i.i.d. segments are extracted, as inferred by the autocorrelation. Fig. 7 plots the autocorrelation of the segment bit rates for the window sizes used. The fast correlation dropoff supports the use of independent realizations for the model.

Based on extensive experiments, three spatial renewal levels are found to be sufficient in modeling most video data. No shorter dependency levels are needed since their effect is visible only at very high link rates and very small buffer sizes. For efficient use of both memory and network resources, the

link capacities of interest range from 2 to 5 times the content bit rate. In the results shown here, the levels are based on window sizes of 24, 360, and 1080 frames, through some variation in window sizes is possible.

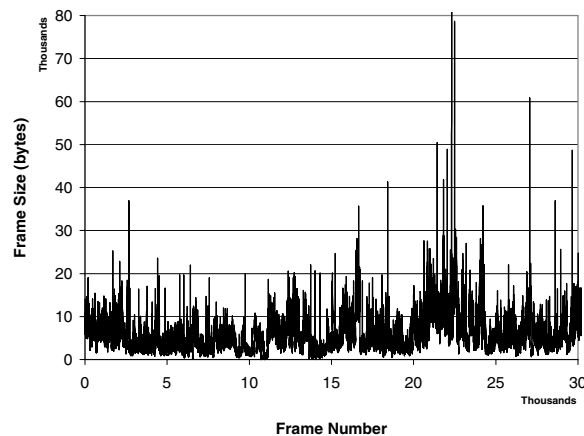


Fig. 8. Generated video trace.

With only three levels of SRRP the individual frame characteristics were ignored. While these have no effect at the link capacities of interest, it does effect the appearance of the simulated trace. Gaussian perturbations were then added to simulate the various individual frame sizes. An example of synthetically generated data is provided in Fig. 8 with its associated buffer plot in Fig. 9. Both figures are consistent in appearance with the actual data.

V. EXPERIMENTAL VALIDATION

The traffic model was applied to a set of several hundred randomly selected 1 Mbit/sec video sequences each with duration of about 45 minutes. The system described earlier, in which data arrives in the buffer at a regular rate and is drawn from the buffer at a varying rate, is utilized. The required buffer sizes for different reliability levels is given in Fig. 10. A reliability level refers to the percentage of full-length (45 minutes in the test set used in these experiments)

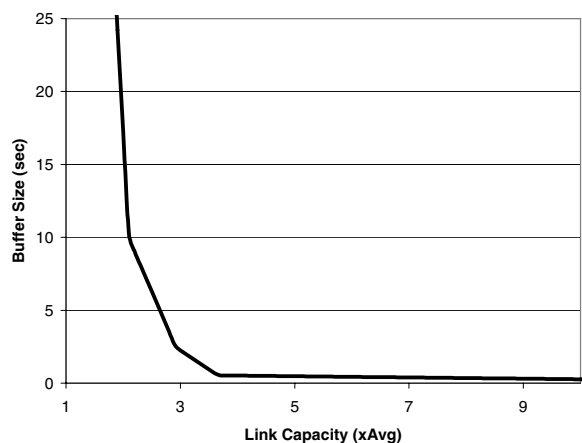


Fig. 9. Plot of minimum buffer size for no underflow vs. link capacity for Fig. 8.

video sequences that can be delivered with no buffer underflow events. For example, Fig. 10 shows that achieving error-free delivery of 99% of the video sequences over a 7 Mbits/sec link requires a buffer with about 30 seconds of capacity, corresponding to 3.5 MBytes given the average data rate of 1 Mbit/sec. Reducing the buffer size to 1 Mbyte would reduce the reliability to 95%.

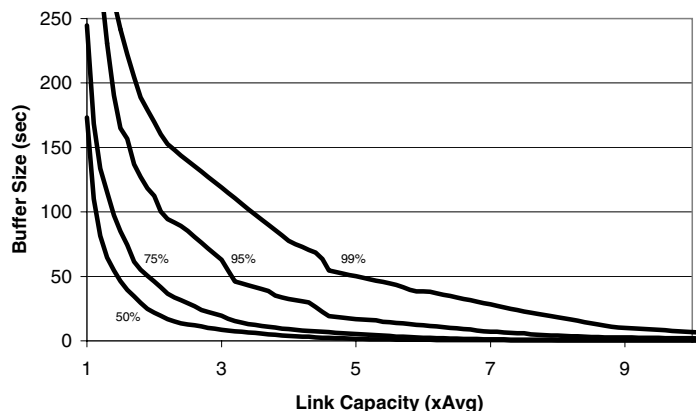


Fig. 10. Required buffer sizes for reliability levels of 50%, 75%, 95% and 99%.

Fig. 11 presents a cdf of the reliability as a function of buffer size for link capacities of 2, 3 and 6 times the average video data rate. For each link capacity, the gray curve gives the performance for the real data and the black curve shows the performance as predicted by the model. This shows that the model accurately enables prediction of buffer sizes aimed at achieving specific target reliabilities.

VI. CONCLUSION

We have presented a spatial renewal video traffic model involving multiple levels of dependency that accurately characterizes both short and long range instantaneous video rate characteristics. The accuracy of the model was validated using a large set of long duration (45 minutes) video sequences, and

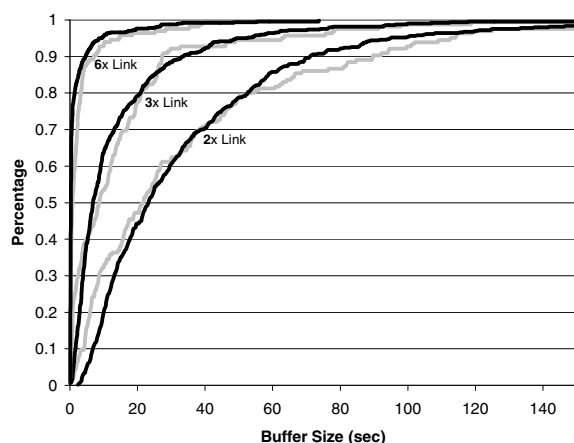


Fig. 11. Reliability for video delivery without buffer underflow at normalized link capacities 2, 3 and 6. The gray curves show the actual data and the black curves show the model.

shown to accurately predict the relationship between buffer size and reliability of video delivery. The methods presented here utilized normalized relationships among average video data rate, network link capacity, and buffer size, allowing them to be generally applied in systems with wide ranges in any of these parameters.

REFERENCES

- [1] Y. Liang, E. Setton, and B. Girod, "Network-Adaptive Video Communication Using Packet Path Diversity and Rate-Distortion Optimized Reference Picture Selection," *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 41, no. 3, pp. 345-354, Nov. 2005.
- [2] D. Heyman, "The GBAR Source Model for VBR Videoconferences," *IEEE/ACM Transactions on Networking*, vol. 5, no. 4, pp. 554-560, August 1997.
- [3] L. Yan-ling, W. Peng, W. Wei-ling, "A Steady Source Model for VBR Video Conferences," in *Proc. ITCC'03*, pp 268-273, April 2003.
- [4] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566-1579, February-March-April 1995.
- [5] H. Ahn, J. Kim, S. Chong, B. Kim, and B.D. Choi, "A Video Traffic Model based on the Shifting-Level Process: the Effects of SRD and LRD on Queuing Behavior," in *Proc. IEEE INFOCOM 2000*, pp. 1036-1045, March 2000.
- [6] X. Huang, Y. Zhou, and R. Zhang, "A Multiscale Model for MPEG-4 Varied Bit Rate Video Traffic," *IEEE Trans. on Broadcasting*, vol. 50, no. 3, pp. 323- 334, September 2004.
- [7] M. Krunz, and A. Ramasamy, "The Correlation Structure for a Class of Scene-Based Video Models and Its Impact on the Dimensioning of Video Buffers," *IEEE Trans on Multimedia*, vol. 2, no. 1, pp. 27-36, March 2000.
- [8] R. Narashimha, and R. Rao, "Discrete-Time Self-Similar Systems and Stable Distributions: Applications to VBR Video Modeling," *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 65-68, March 2003.
- [9] B. Melamed, "Modeling Compressed Full-Motion Video," in *Proc. 1997 Simulation Conference*, pp 1368-1374, December 1997.
- [10] C.H. Liew, C.K. Kodikara and A.M. Kondoz, "MPEG-Encoded Variable Bit-Rate Video Traffic Modeling," *IEE Proc.-Commun.*, vol. 152, no. 5, pp. 749-756, October 2005.
- [11] D. Lucantoni, M. Neuts, and A. Reibman, "Methods for Performance Evaluation of VBR Video Traffic Models," *IEEE/ACM Trans on Networking*, vol 2, no 2, pp 176-180, April 1994.