ISSUES RELATED TO SLICE BASED MODELING OF MPEG VBR ENCODED VIDEO

Michael R. Izquierdo
IBM Corporation
RTP, North Carolina 27709

Douglas R. Reeves
Dept. of Elect. and Comp. Engineering
North Carolina State University
Raleigh, North Carolina 27695

Abstract - The objective of this paper was to explore the possibility of developing a slice based model whose random variable is independent and would fit classical distributions. We analyzed four MPEG-1 VBR encoded high quality, color video sequences. None of the sequences contained audio. We present three types of slice based models and discuss the merits of each. We show the distributions given by each of the models and show their fit to the Gamma and Pareto distributions using the QQ plot.

INTRODUCTION

The Asynchronous Transfer Mode (ATM) Network has gained much attention as an effective means to transfer voice, video and data information over computer networks. The use of a fixed size, fifty-three byte cell to transfer data makes ATM well suited to support isochronous type services like voice and video [1]. The small cell size makes it possible to interleave cells from multiple sources over a common communications link; thereby, providing low end-to-end latency. Much work has been done in the area of transporting compressed video over ATM addressing such issues as bandwidth allocation, source modeling, multiplexing, encoding methods and quality of service (QoS).

Variable bit rate (VBR) encoding has several advantages over constant bit rate (CBR) encoding such as consistent and subjectively better video quality, simpler encoder design and increased multiplexing gain. One paper compared the luminance signal-to-noise ratio (SNR) of CBR and 1-layer VBR and showed significant reductions in SNR of up to 7 dB [2]. Another paper showed that the statistical multiplexing of multiple VBR sources provided a gain of a factor of two over CBR [3]. Recently a gain of slightly higher than four was found to be possible with cell loss probabilities of 10^{-6} [4].

One of the main drawbacks of VBR is that its burstiness increases the probability of cell loss by making it difficult to determine bandwidth requirements. Burstiness is caused by the fact that the encoder is not controlling the quantization scale dynamically in order to maintain a constant bit rate. In this sense, VBR is referred to as an *open-loop* encoding method.

The user must specify bandwidth requirements when establishing a connection in order for the ATM network to determine if enough resources, such as buffers and communications links, exist. This is a preventative congestion control method in that flow control is done at the source in an attempt to avoid congestion [5]. The user can specify bandwidth at the peak rate, but this would waste a significant amount of bandwidth. If bandwidth is improperly specified, high cell loss could occur. For this reason, it is important to develop effective models which developers and researchers can use to determine the effective bandwidth (bandwidth required for a given cell loss) and the multiplexing gain for compressed video sources such as MPEG.

We focused our attention on the video server output cell stream and assumed that the video was located on a local file system. The objective was to characterize the cell generation process at the output of the Segmenter and determine an appropriate model. Once this is done, then the effective bandwidth (C) for a certain buffer size (B) can be determined.

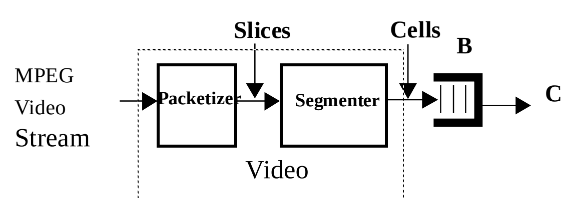

FIGURE 1. VBR video server model to analyze cell generation statistic

MPEG OVERVIEW

MPEG stands for the Motion Pictures Experts Group and is an ISO standard for the coding of video and audio [6]. It was defined to encode video and audio at rates of 1.5 Mbps using a SIF frame resolution of 352x240 pixels. Video is compressed spatially using the Discrete Cosine Transform (DCT) and temporally using

Motion Compensation (MC). The frame (or picture) type determines which method of compression is used.

A video sequence is broken up into a series of sequential Group-of-Pictures (GOPs) with each GOP consisting of at least one anchor frame and a number of reference frames. There are three frame types in MPEG, Intracoded (I frame), Predictive coded (P frame) and Bidirectionally predictive coded (B frame). I frames are encoded using the DCT and do not use MC. Since I frames do not reference any other frames, they are useful in preventing the propagation of any distortion errors Each GOP is required to begin with an I frame. P frames use MC and reference previous I or P frames. I and P frames are typically referred to as anchor frames. B frames use MC and reference a past and/or future I or P frame. B frames typically offer the highest compression.

Each GOP consists of a number of I, B and P frames determined by the parameters N and M, where N equals the number of frames in a GOP and M equals the number of B frames between anchor frames. This defines a frame sequence pattern like IBBPBBI where N=6 and M=2 which repeats itself for the whole video sequence.

PREVIOUS WORK IN VIDEO MODELING

Simulations require effective source models in order to understand and identify the impact of compressed video sources on computer networks. Most of the work has focused on video conference type sequences and consequently did not include frames encoded using B frames. The reasoning was that B frames required too much time to encode for real time video conference type applications. The majority of the models have also been frame based in that cells for a particular frame are presented to the network at the beginning of a frame interval and transmitted within the interval with either random or uniform cell spacing. This is less accurate than a slice source model, but less demanding on simulation resources. Frame based models do not capture the effects of spatial content within a frame on queueing performance. It would be interesting to see if a slice based model produced significant changes in queueing performance when compared to a frame based model. In any case, a model must capture both the distribution and autocorrelation functions of the source in order to be useful.

Early work used AR processes to model single VBR sources [4] [7] [8]. AR(2) models have been found to be sufficiently accurate for traffic studies; however, a DAR process based on a discrete multi-state Markov Chain was found to be more accurate for video conference data [4]. Markov Chains have been used to model multiple VBR sources in a multiplexed environment [9] [10] [11]. More sophisticated models have been developed for VBR sequences with scene changes which are not adequately modeled with single AR process. One model used two AR processes and two complementary processes (used to determine the occurrence of a scene change) which are modulated by a three state Markov Chain [12]. Another used multiple AR processes, one for the number of block per field and a second for the number of bits per block [13]. In all of these models only I and P frames are included, not B frames. One paper suggested that ignoring B and P frames could severely under estimate cell loss rates [14]. Recent work has focused on modeling each frame type individually, cycling through each model based on the video frame sequence pattern (e.g. IBBPBBP...) [15] [14].

SLICE BASED MODELING

We studied four video sequences obtained from the Portable Video Research Group called *Bike*, *Flowg*, *Tennis* and *UnderSiege*. Each video sequence was encoded using a quantization scale (q) I/B/P triplet of (4,8,4). All sequences contained 15 slices per frame and 150 frames except for *UnderSiege* which contained 731 frames. Each sequence had SIF resolution (352x240 pixels) with N=6 and M=2. None of the sequences contained audio and all contained color.

Model I: We first investigate a model whose random variable is the number of cells/slice. This random variable is used for all frame types. We show in Figure 2 the distribution of this random variable for the *Flowg* sequence. One can see from the QQ plots shown in Figure 3 that the Pareto function provides a better fit than Gamma (The QQ plot is commonly used to determine if a given sample data fits a known classical distribution. A linear plot indicates a fit). It might be possible to develop a slice based model using the Pareto distribution; however, the samples need to be distorted in a such a way so as to produce a similar autocorrelation function.

The advantage of this model is that it is simple in the sense that only one random variable is required to represent the number of cells/slice irrespective of frame type. However, accuracy is compromised since this model does not take the I/B/P frame sequence pattern into account. The frame type sequence pattern has a big impact on the autocorrelation function.

Figure 4 show that the autocorrelation function for *Flowg* is quasi-periodic with negative decay. This was also true for the other sequences as well. This differs

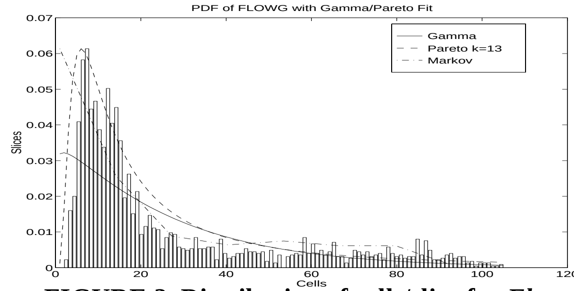

FIGURE 2. Distribution of cells/slice for *Flowg* sequence.

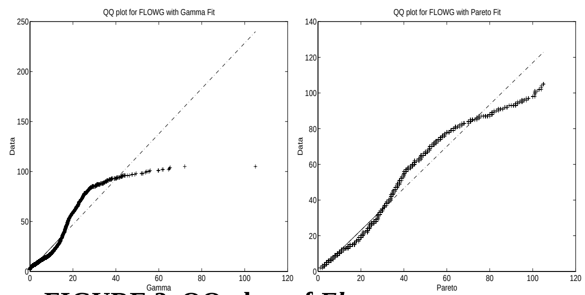

FIGURE 3. QQ plots of *Flowg* sequence showing Gamma fit (left) and Pareto fit (right).

from the autocorrelation functions seen at the frame layer (without B frames) which typically show either exponential or hyperbolic decay [15]. This indicates that special considerations should be taken with I/B/P sequences in the areas of multiplexing and dynamic bandwidth allocation [16].

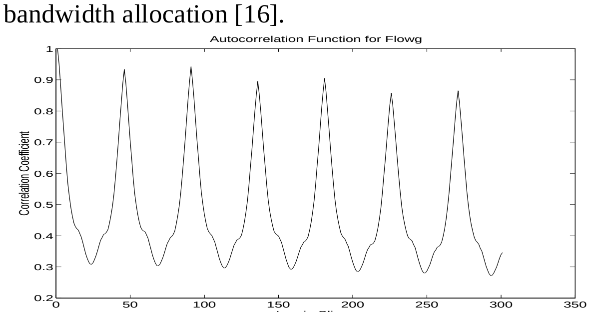

FIGURE 4. Autocorrelation function of *Flowg*.

Model II: It would be difficult for a model to produce such an autocorrelation function. As a consequence, we look at using a slice based model based on frame type similar to [17] [18]. The histograms for the number of cells/slice within I frames for all four sequences is shown in Figure 5. In this case, we will determine fit by inspection. We can see that *Bike* appears to fit Gamma well, whereas *Flowg* would probably fit a Uniform distribution better and *Tennis* a Gaussian distribution. The fourth sequence, *UnderSiege*, does not appear to fit any distribution, however it is distorted due to the large number of slices with low cell counts (3 cells). This was due to the black borders on the top and bottom of the frames so slices 1,2 and 14,15 had 3 cells for all frames.

This type of model is promising in that the frame sequence pattern is taken into account. However, one of the drawbacks is that it does not take into account the spatial behavior within a frame. For example, a single random variable could be used to represent the number of cells/slice within an I frame. Fifteen samples from this random variable, one for each slice position within a frame, might be more random than the original sample data.

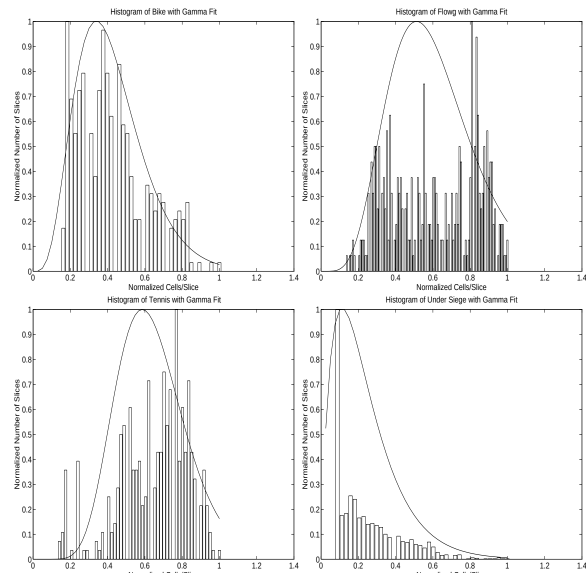

FIGURE 5. Normalized histogram of I frame cells/slice for *Bike*, *Flowg*, *Tennis* and *UnderSiege*.

Model III: Next we looked at a model which is based on the slice position within a frame. In this case, one could have different distributions for each slice position within a frame. We looked at I frames only for the video sequence *UnderSiege*.

The distribution for the number of cells per slice position was deterministic for rows 1,2 and 14,15. As was mentioned earlier, this was due to dark borders on the top and bottom of each frame. This was probably done to correct the aspect ratio. While this might not be a common thing to do, it does point out the drastic effects video content can have on the statistics.

The slice position distributions were not smooth indicating that more samples are needed. However, slice positions 4-10 appeared to be Gamma shaped while 3 appeared to more uniformly distributed and 11-13 appeared to fit Pareto due to its longer tail.

This model is the most complicated requiring a random variable for each slice position per frame type. For

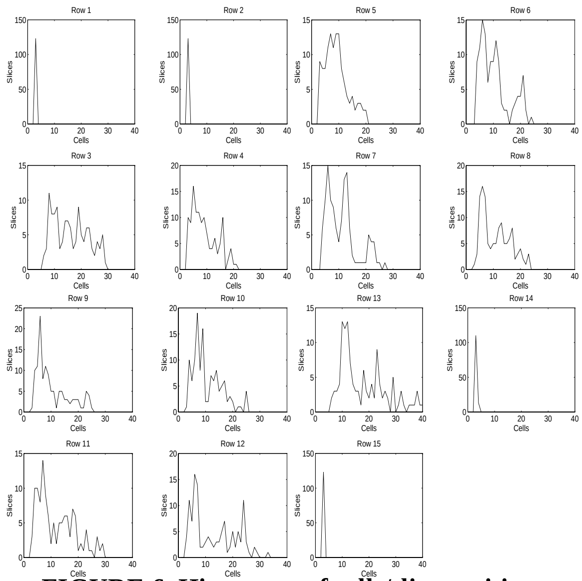

FIGURE 6. Histogram of cells/slice position within a frame for *UnderSiege*.

these video sequences, that would mean 45 random variables!

CONCLUSION

We discussed three types of slice based models and discussed their merits. Future work would involve simulating these different models to determine their merits over frame based models. We would also like to determine which of the three types of models work best for video traffic studies to determine multiplexing gain and effective bandwidth.

REFERENCE

[1] M. Kawarasaki and B. Jabbari, “B-ISDN Architecture and Protocol,” IEEE J. Select. Areas Commun., vol. 9, no. 9, pp. 1405-1415, December 1991.

[2] D.G. Morrison, “Variable Bit Rate Video Coding for Asynchronous Transfer Mode Networks,” Br. Telecom Technol. J. (UK), Vol. 8, No. 3, July 1990, pp 70-80.

[3] W. Verbiest, L. Pinno, and B. Voeten, “The Impact of the ATM Concept on Video Coding,” IEEE J. Select. Areas Commun., vol.6, no. 9, pp. 1623-1632, December, 1988.

[4] D. Heyman, A. Tabatabai, and T. V. Lakshman, “Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks,” IEEE J. Select. Areas Commun., vol. 2, no. 1, pp. 49-59, March 1992.

[5] M. Sidi, et al.,”Congestion Control Through Input Rate Regulation,” in Proc. IEEE GLOBECOM ‘89, pp 49.3.1-49.3.5.

[6] Motion Pictures Expert Group, “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/sec,” ISO 2-11172 rev, Part 2: Video, November 22, 1991.

[7] M. Nomura, T. Fujii, and N. Ohta, “Basic Characteristics of Variable Rate Video Coding in ATM Environment,” IEEE J. Select. Areas Commun., vol. 7, no. 5, pp. 752-760, June 1989.

[8] R. Grunenfelder, J. Cosmas, S. Manthorpe, and A. Odinma-Okafor, “Characterization of Video Codecs as AutoRegressive Moving Average Processes and Related Queueing System Performance,” IEEE J. Select. Areas Commun., vol. 9, no. 3, pp. 284-293, April 1991.

[9] P. Sen, B. Maglaris, N.E. Rikli, and D. Anastassiou, “Models for Packet Switcjing of Variable-Bit-Rate Video Spurces,” IEEE JSAC, Vol. 7, No. 5, pp.865-869, June 1989.

[10] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins, “Performance models of statistical multiplexing in packet video communications,” IEEE Trans. Commun., vol. 36, no. 7, pp. 834-844, July 1988.

[11] P. Skelly, M. Schwartz and Sudhir Dixit, “A Histogram-Based Model for Video Traffic Behaviour in an ATM Multiplexer,” IEEE/ACM Transactions on Networking, Vol. 1, No. 4, pp. 446-459, August 1993.

[12] F. Yegenoglu, B. Jabbari, Y. Q. Zhang, “Motion-Classified Autoregressive Modeling of Variable Bit Rate Video,” IEEE Trans. on Circuits and Systems for Video Tech., vol. 3, no. 1, pp 42-53, February 1993.

[13] B. Jabbari, F. Yegenoglu, Y. Kuo, S. Zafar, and Y. Q. Zhang, “Statistical Characterization and Block-Based Modeling of Motion-Adaptive Coded Video,” IEEE Trans. on Circuits and Systems for Video Tech., vol. 3, no. 3, pp 199-207,June 1993.

[14] D. P. Heyman, A. Tabatabai and T.V. Lakshman, “Statistical Analysis of MPEG2-Coded VBR Video Traffic,” 6th Intl. Workshop on Packet Video, pp B2.1 - B2.5.

[15] M.W. Garrett and W. Willinger, “Analysis, Modeling and Generation of Self-Similar VBR Video Trafic,” SIGCOMM ‘94, pp269-280.

[16] M.R. Izquierdo and D.R. Reeves, “Statistical characterization of MPEG VBR video at the Slice layer,” To appear in the IS&T/ Proceedings, Feb. 1995.

[17] J. Enssle, “Modelling and Statistical Multiplexing of VBR MPEG Compressed Video in ATM Networks,” To Be Published ‘94.

[18] D.P. Heyman and T.V. Lakshman, “Source Models for VBR Broadcast-video Traffic,” INFOCOM