Submitted to ICIP'97

MULTIPLEXING VBR VIDEO SEQUENCES ONTO A CBR CHANNEL WITH LEXICOGRAPHIC OPTIMIZATION

*Dzung T. Hoang**

Digital Video Systems, Inc.
2710 Walsh Ave., Ste. 200
Santa Clara, CA 95051
dth@dvsystems.com

Jeffrey Scott Vitter†

Department of Computer Science
Duke University, Box 90129
Durham, NC 27708-0129
jsv@cs.duke.edu

ABSTRACT

We apply a novel lexicographic framework for bit allocation to the multiplexing of multiple VBR video streams onto a single CBR channel. In the lexicographic framework, the maximum distortion is minimized, then the second highest distortion, and so on, resulting in nearly constant quality. With a suitably constructed multiplexing model, we show that the multiplexing problem reduces to a single-stream CBR bit allocation problem, to which we apply the lexicographic framework. This method has applications for video servers, especially for near-video-on-demand.

1. INTRODUCTION

There are several scenarios where multiple compressed video streams are to be transmitted through a common channel. Two obvious examples are networked video and digital video broadcasting. In these types of applications, the transmission channel is typically bandwidth-limited. With the available bandwidth, we would like to provide as much video programming as possible without having to sacrifice quality.

Video coding systems in common use, such as the MPEG standards [1, 2], employ variable-length coding techniques that produce a number of bits that varies from frame to frame. To transmit a single coded video stream over a constant-bit-rate (CBR) channel, an encoder buffer is typically used to smooth out the variations in bit rate. The fullness of the buffer is monitored by a rate controller to prevent the buffer from overflowing or underflowing. In this type of coding system, the encoder generates a bitstream whose bit rate is constant in the long term, with short-term variations limited by the size of the encoder buffer. Such a bitstream is commonly referred to as a constant-bit-rate (CBR) stream. In contrast, a bitstream whose average bit rate is not constant over time is called a variable-bit-rate (VBR) stream.

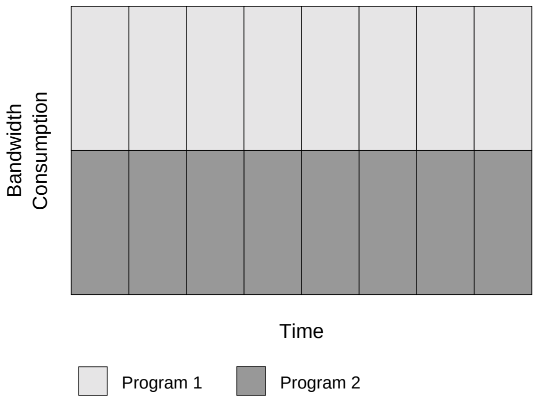
1.1. Statistical Multiplexing

When multiple video streams are to be transmitted over the same channel, it is no longer required for any single bitstream to be coded at a constant rate. An often-cited motivation for VBR video coding is that it could potentially allow for the simultaneous transmission of more video streams over a common channel than CBR coding at the same quality. The main reasoning is provided through a concept called *statistical multiplexing*. Statistical multiplexing is based on the observation that the bit rate of constant-quality video is highly variable from frame to frame. In order to achieve image quality that is *not less* than that of a constant-quality VBR encoding, a CBR encoding would require an average bit rate that would correspond to the peak rate of the VBR encoding. Since a VBR encoding typically requires the peak rate for only a small percentage of time, it uses less bandwidth on average than a comparable CBR encoding. Furthermore, assuming that the VBR streams have independent bit-rate characteristics, we can transmit more VBR streams than CBR streams over a common channel with a low probability that the combined instantaneous bit rate would exceed the channel rate.

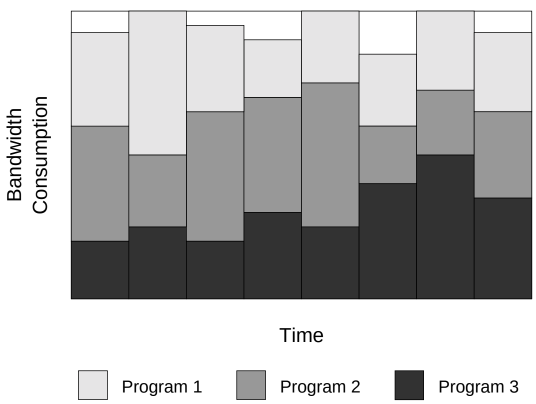
As an example, consider a channel with a bandwidth of 100 Mbit/s. Suppose that for a desired level of quality, a peak rate of 10 Mbit/s is required for coding a suite of video programming. Using CBR encoding,

*While the author was a visiting scholar at Duke University, support was provided in part by Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants F49620–92–J–0515 and F49620–94–1–0217, by Army Research Office grant DAAH04–93–G–0076.

†Support was provided in part by Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants F49620–92–J–0515 and F49620–94–1–0217, by Army Research Office grant DAAH04–93–G–0076, and by an associate membership in CESDIS.



(a) Multiplexing of 2 CBR Bitstreams



(b) Multiplexing of 3 VBR Bitstreams

Figure 1: These plots show an example of how three VBR bitstreams can be multiplexed into the channel as two CBR bitstreams, for a statistical multiplexing gain of 1.5.

up to 10 video streams can be transmitted through the channel simultaneously. Since the peak rate is required only a small percentage of the time, suppose that the actual average rate is only 5 Mbit/s. Then using VBR encoding with a peak rate of 10 Mbit/s and average rate of 5 Mbit/s, we can *potentially* transmit 20 simultaneous video treams. This would correspond to a *statistical multiplexing gain* of 2.

The advantage of VBR encoding over CBR is shown in a simple example in Figure 1. In this example, three VBR streams are shown multiplexed using at most the same bandwidth required by a CBR encoding of only two of the sources.

In order to transmit the 20 VBR sequences simultaneously, however, the instantaneous bit rate for the 20 sequences must not exceed the channel capacity for an extended period of time, which is determined by the amount of buffering present. Assuming that the bit rates of the different video streams are uncorrelated in time, there is a low probability that the channel capacity would be exceeded in any time interval. Quantifying and minimizing this probability are central themes of research.

1.2. Lexicographic Rate Control

Another motivation for VBR video coding is that it allows for constant-quality encodings. In [3, 4], a new framework that uses a novel lexicographic optimality criterion is proposed for bit allocation in an MPEG video coder. The framework consists of 1) a bit production model for the input frames, 2) a set of bit-rate constraints imposed by the MPEG Video Buffering Verifier (VBV) for both CBR and VBR encoding, and 3) a novel lexicographic criterion for optimality.

In contrast to existing optimization frameworks [5, 6] that minimize a sum-distortion measure, such as mean squared error, the lexicographic framework first minimizes the maximum distortion, then the second highest distortion, and so on. Under this framework, the authors derive necessary and sufficient conditions for optimality that lead to efficient algorithms based on dynamic programming.

In this paper, we show how the lexicographic bit allocation framework can be applied to multiplex several VBR streams over a common CBR channel. In contrast to typical statistical multiplexing techniques, as exemplified in Figure 1, our method allocates bits to the VBR streams in a *deterministic* manner, making full use of all the available channel bandwidth. However, our method is based on a global optimization and requires off-line processing. This requirement is not prohibitive for stored-video applications such as near-video-on-demand (NVOD).

2. MULTIPLEXING MODEL

We first elaborate a model for multiplexing multiple VBR streams onto a CBR channel. Since the lexicographic framework is deterministic and uses look-ahead, we assume that complete statistics of the multiple video sources are available to the bit allocation algorithm. This requirement can be met by providing a centralized encoder for the multiple sources, as depicted in Figure 2. In the figure, M video sources enter a encoder/multiplexer that produces a single multiplexed stream for transport over a CBR channel. On the receiving end, a demultiplexer/decoder performs demultiplexing and decoding to reproduce the M video

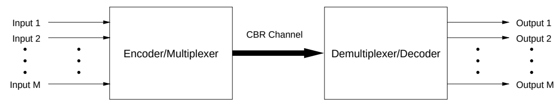


Figure 2: System for transmitting multiple sequences over a single channel.

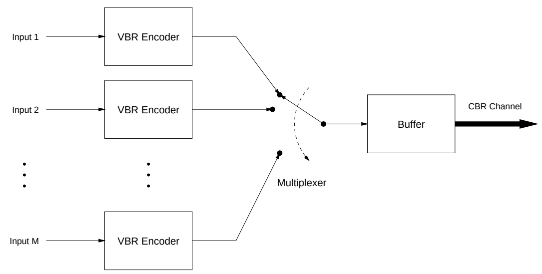


Figure 3: Block diagram of encoder/multiplexer.

sequences. This multiplexing model is similar to that proposed in [7].

This model is applicable to applications such as a video server where the video sequences to be multiplexed are known in advance. An especially noteworthy case is that of near-video-on-demand (NVOD), where a single sequence is to be transmitted simultaneously with different starting times. For example, 20 copies of a 2-hour long movie can be multiplexed so that one can view the beginning of the movie every six minutes.

The encoder/multiplexer block is expanded in Figure 3. As shown, the input video sources are encoded individually and time-division multiplexed and stored in a buffer before being output to the channel at a constant bit rate. The encoders are synchronized so that they output the encoding of a picture at the same time every T seconds. The multiplexer then concatenates the multiple encodings in order as shown in Figure 4.

The demultiplexer/decoder block is diagrammed in Figure 5. The demultiplexer/decoder mirrors the operation of the encoder/multiplexer. Incoming bits from the channel are stored in a decoding buffer. In intervals of T seconds, the demultiplexer instantaneously removes from the buffer all bits needed to decode the next picture of all sequences and routes the bitstreams to the appropriate decoders, which then output the reconstructed video.

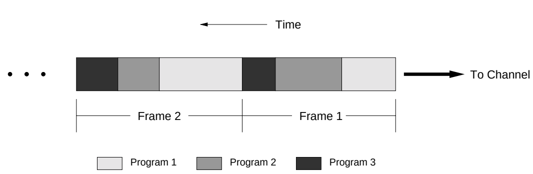


Figure 4: Operation of multiplexer.

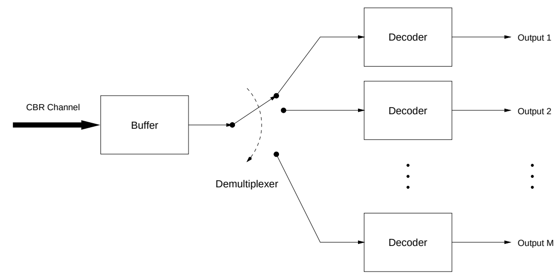


Figure 5: Block diagram of demultiplexer/decoder.

The multiplexing model described above resembles the operation of the single-stream encoder and decoder system implied by the MPEG Video Buffering Verifier when operating in CBR mode. If we view the different input sources as providing "slices" of the same picture, the resemblance would be very close indeed. This construction is intentional and allows us to apply the lexicographic framework to allocate bits optimally to the multiple VBR streams.

3. LEXICOGRAPHIC CRITERION

Before we can apply the lexicographic bit allocation framework, we need to define an optimality criterion. Since there are multiple sources, a lexicographical criterion based on the encoding of a single video sequence is certainly not appropriate. We need to consider the quality of all the sequences. A simple way to do this is to consider the concatenation of all the sequences and define a lexicographic criterion on the concatenated video sequence. By doing this, we are putting equal weight to each picture of each video sequence. We can also consider extending the lexicographic criterion to compare vectors of length M instead of scalar quantities. However, as we will see, this is equivalent to just considering the concatenation of the M sequences.

4. EQUIVALENCE TO CBR BIT ALLOCATION

We now show that in the multiplexing model put forth above, the problem of optimal bit allocation for multiple VBR streams reduces to a CBR bit allocation problem for a single stream.

The multiplexing model is intentionally set up to follow the operation of the MPEG VBV in CBR mode. We can view the buffer, the demultiplexer, and the bank of M decoders in Figure 5 as comprising a single VBV. The lexicographic framework of [4] can then be applied.

In [4], it is assumed that perceptual quantization is used so that the perceived distortion can be equated to a "nominal" quantization scale. We make the same assumption and seek to equalize distortion by equalizing quantization scale.

For display interval i, we need to consider the quantization scales used to code picture i of each of the M video sequences. Given a fixed bit budget for coding picture i of each video sequence, it is easy to show that the same quantization scale must be used to code picture i of all sequences to achieve lexicographic optimality: If the quantization scales differ, we can always shift bits around to reduce the highest quantization scale by increasing a lower quantization scale. This result also holds if we formulate the lexicographic criterion using vectors of quantization scales.

By using a combined bit production model that is the sum of the bit production models for each sequence, we can then allocate bits to the sequences using the CBR Algorithm in [4].

While the above technique guarantees that the demultiplexer buffer in Figure 5 does not overflow or underflow, it should be noted that doing so *does not* guarantee MPEG VBV compliance for the individual video streams, except when the individual VBV buffers are at least the size of the demultiplexing buffer. A multiplexing model that explicitly includes individual decoder buffers is certainly possible. However, analysis of this situation is not as straightforward as with the above model and remains an open problem.

5. RELATED WORK

A buffered rate control scheme for multiplexing VBR sources onto a CBR channel is described in [8]. This work is based on the rate-distortion framework of [5] and [9] and uses a multiplexing model very similar to Figure 3. As described, the basic allocation unit is taken to be a GOP.

6. REFERENCES

[1] ISO. Cd11172-2: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbits/s, November 1991.

[2] ISO-IEC/JTC1/SC29/WG11/N0802. Generic coding of moving pictures and associated audio information: Video, November 1994. MPEG Draft Recommendation ITU-T H.262, ISO/IEC 13818-2.

[3] D. T. Hoang, E. Linzer, and J. S. Vitter. Lexicographically optimal rate control for video coding with MPEG buffer constraints. Technical Report CS-1996-02, Duke University, Dept. of Computer Science, 1996.

[4] D. T. Hoang, E. L. Linzer, and J. S. Vitter. A lexicographic framework for mpeg rate control. In *Proceedings 1997 Data Compression Conference*, 1997.

[5] A. Ortega, K. Ramchandran, and M. Vetterli. Optimal trellis-based buffered compression and fast approximations. *IEEE Transactions on Image Processing*, 3(1):26–40, January 1994.

[6] C.-Y. Hsu, A. Ortega, and A. R. Reibman. Joint selection of source and channel rate for VBR video transmission under ATM policing constraints, 1997. To appear.

[7] B. G. Haskell and A. R. Reibman. Multiplexing of variable rate encoded streams. *IEEE Transactions on Circuits and Systems for Video Technology*, 4(4):417–424, August 1994.

[8] D. Park and K. Kim. Buffered rate-distortion control of MPEG compressed video channel for DBS applications. In *Proceedings IEEE Internation Conference on Communications*, volume 3, pages 1751–1755, 1995.

[9] J. Choi and D. Park. A stable feedback control of the buffer state using the controlled lagrange multiplier method. *IEEE Transactions on Image Processing*, 3(5):546–557, September 1994.