

OPTIMAL RATE CONTROL FOR VIDEO TRANSMISSION OVER VBR CHANNELS BASED ON A HYBRID MMAX/MMSE CRITERION

Sang-Yong Lee and Antonio Ortega

Integrated Media Systems Center, Signal and Image Processing Institute,
Department of Electrical Engineering-Systems,
University of Southern California
3740 McClintock Ave., Los Angeles, CA 90089-2564

ABSTRACT

In this paper, we consider the problem of rate control for video transmission over variable bit rate (VBR) channels. We focus on finding off-line optimal rate control for VBR transmission with a token bucket policing function. To ensure a maximum minimum quality is obtained over all data units, we introduce a minimum maximum distortion (MMAX) criterion in this channel-constrained problem. We show that, due to the channel constraints, a MMAX solution leads to a relatively low average distortion, because the total rate budget is not completely utilized. Therefore, after finding a MMAX solution, an additional minimization of average distortion (MMAX+) criterion is proposed to increase overall quality of the data sequence by using remaining resources. The proposed algorithms lead to an increase in average quality with respect to the MMAX solution, while providing a much more constant quality than MMSE solutions. Moreover we show how the MMAX+ approach can be implemented with low complexity.

1. INTRODUCTION

Future high bandwidth video applications, such as video-on-demand (VOD) will require transmission over the network of video compressed at a variable rate. Thus, a rate control has to be used, based on objectives such as coded video quality or data rate. Also, video transmission requires delay constraints to guarantee real time playback, since video frames that arrive too late are useless.

To transmit VBR encoded data, VBR transmission is preferable to CBR (constant bit rate) transmission since VBR transmission needs lower end-to-end delay and a smaller buffer size [1, 2, 3]. However, due to the limited network resources, negotiation between each user and the network is indispensable in order to ensure QoS (quality of service) guarantees, where the parameters specified to define QoS can be delay jitter, bandwidth, end-to-end delay and so on. In addition, policing mechanisms are used to alert the network if there are users who violate the agreed upon transmission parameters.

VBR video transmission through ATM (Asynchronous Transfer Mode) networks with a leaky bucket policing function has been studied in the literature [1, 4, 2, 3]. In [1, 2, 3], VBR transmission under both encoding and decoding buffer constraints and channel constraints is studied. In [4], multiple leaky bucket policing is introduced to regulate peak rate. It may be desirable to supplement

a traffic policing function with a traffic shaping policy where traffic shaping is used to smooth out a traffic flow. One simple traffic shaping approach is “token bucket” (TB) policing. At the encoder side, leaky buckets and token buckets with the same parameters (i.e., the bucket size and the token rate) are equivalent. However, at the decoder side, the incoming data rate produced by each of these two approaches can be different.

Token buckets are also specified in next generation Internet Protocol (IP) networks. The Internet Engineering Task Force (IETF) has defined the Guaranteed Service (GS) in order to provide QoS to real time applications and token bucket policing is recommended as a traffic shaping method for GS [5].

After channel and source constraints such as channel bandwidth, peak transmission rate, limited delay, total bit-budget or the size of codec buffers have been determined, a target quality measure should be chosen. Most previous work for image and video coding has been based on minimization of average distortion (MMSE). As a consequence, optimal bit allocation under various constraints for the MMSE criterion has been widely studied in the literature. Examples include bit allocation for arbitrary inputs and a discrete set of available quantizers [6], bit allocation under buffer constraints [7] and bit allocation for video transmission over ATM networks [2] [3]. A main drawback of the MMSE criterion is that the quality difference between frames can be large and some frames may be coded at relatively low quality even though the average quality is high. A minimum maximum distortion (MMAX) criterion has been proposed to prevent this heavy fluctuation of source quality and, based on this criterion, optimal bit allocation under a bit-budget constraint has been studied [8]. Using this criterion, coding units having a significantly lower than average quality can be avoided. However, when multiple constraints are present, the MMAX criterion by itself may be inefficient. This is because the MMAX optimization is terminated as soon as it cannot decrease the maximum overall distortion.

A criterion for minimizing distortion in lexicographical sense (MLEX) has been proposed as a modified MMAX approach to increase overall quality [9]. This criterion is used to find optimal bit allocation under CBR constraints in [9] where quantizer levels are used as a distortion measure and it is shown that the optimal solution is determined by using constant quantizer segments. But, in general, if a distortion measure can take any arbitrary values, the proposed algorithm cannot be easily applied.

As an alternative approach to increase overall quality after finding a MMAX solution, we propose to use the MMSE criterion for the remaining bit-budget. We denote this criterion MMAX+,

This work has been supported in part by NASA under grant AIST-0122-0005.

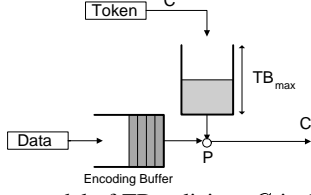


Fig. 1. System model of TB policing. C is the token rate and C_i is the transmission rate of the i^{th} frame interval. TB_{max} and P indicate the size of a token bucket and the peak rate for each. In this policing, one byte data can be transmitted per token.

because it adds additional targets to the MMAX criterion. Since MMAX+ explicitly targets average distortion it will lead to better average mean square error (MSE) than MLEX. We have also proposed the MMAX+ criterion for video transmission over a CBR channel in [10]. In this paper, we extend the idea to video transmission over a VBR channel with token bucket policing. While in the CBR case a MMAX+ technique improves quality by preventing encoder buffer underflow, in the VBR case the increase in quality is achieved by preventing token buffer overflow. First we develop an algorithm to find the optimal MMAX solution. We introduce MMAX and MMAX+ criteria in this channel constrained problem with a discrete set of quantizers available to code each frame, so that the best minimum quality of all frames is provided by the MMAX criterion and good overall quality is achieved by the MMAX+ criterion. We also propose an algorithm to reduce the complexity of finding the MMAX+ solution. Simulation results show that the solution of our proposed method gives much better minimum quality than the MMSE solution, with comparable average quality at a much lower complexity.

This paper is organized as follows: in section 2, an algorithm to find the optimal solution under a MMAX criterion is presented. In section 3, algorithms to find the optimal MMAX+ solution with reduced complexity are proposed. Experimental results are provided in section 4. Conclusions are provided in section 5.

2. OPTIMAL RATE CONTROL IN A MMAX CRITERION

Video transmission is constrained by the maximum delay allowable, the encoder and decoder buffers and channel constraints, such as channel rate and channel policing functions. In a VBR transmission case, preventing encoder buffer overflow does not guarantee that a decoder buffer is not in overflow or underflow. Therefore, the decoder buffer state has to be considered as a constraint in a rate control scheme for VBR transmission [2] [3]. In this paper, to maximize channel utilization, we assume that transmission is constrained by the maximum delay, rather than by the size of encoder and decoder buffers. In other words, the size of encoder and decoder buffers is assumed to be large enough to always store all the data that it will be possible to transmit under the given delay constraint (for the given channel rate constraint). Thus, our goal is finding the optimal rate control for MMAX and MMAX+ criteria under the given maximum delay and network policing constraints.

In this paper, TB policing is used as a policing constraint. TB policing is defined with 5 parameters named transmission specification (Tspec) in [5]. In the specification, the amount of data sent is constrained not only by the available tokens but also by the peak rate. This peak rate constraint is used to put a limit on the size of data bursts. Therefore, as shown in Fig. 1, the constraints of our problem are token bucket parameters (the token rate (C) and the

size of token bucket (TB_{max})), the peak rate (P) and the maximum delay (M), where the peak rate and the token rate are measured in bytes per frame interval. Then the problem we are trying to solve using a MMAX criterion can be formulated as follows:

$$\min_{q_i} (\max(D_i)) \quad (1)$$

$$\text{s.t. } B_i \leq \min(TB_i + M \cdot C, M \cdot P), \quad (2)$$

$$\text{where } TB_i = \min(TB_{max}, TB_{i-1} + C - C_{i-1}), \quad (3)$$

$$B_i = B_{i-1} + R_i - C_{i-1}, \quad (4)$$

$$\text{with } C_{i-1} = \min(B_{i-1}, TB_{i-1} + C, P). \quad (5)$$

In the above equations, B_i indicates the buffer occupancy after the encoded i^{th} frame ($1 \leq i \leq S$, S is the number of frames) is moved to the buffer, TB_i indicates the TB state just before starting the i^{th} frame interval (the interval between the i^{th} frame and the next frame) transmission, and M indicates maximum delay in frame units. R_i and D_i indicate the rate and distortion of the i^{th} frame, which are determined by the selection of a quantization level q_i ($1 \leq q_i \leq Q_i$). Initial transmission rate (C_0) is zero and initial encoder buffer and TB state can be any values between zero and their maximum value. The number of tokens in a TB (TB_i) and the number of data in an encoder buffer (B_i) cannot be negative; this is guaranteed by (5). Note that among many possible channel rate selection policies, in (5), we select the maximum available transmission rate at each frame interval. This selection guarantees a performance as good as the best, since it tends to minimize the token overflow probability (because it uses rate as fast as it can.)

As shown in (2), the encoder buffer state is restricted by the maximum amount of data that can be sent during the next M frame intervals. Therefore the minimum size of the encoder buffer (B) that will prevent additional buffer constraints from arising is

$$B = \min(TB_{max} + M \cdot C, M \cdot P). \quad (6)$$

In other words if the physical buffer size is greater than B in (6) then we need not consider the additional buffer constraint (but still consider the delay constraint as in (2)). The algorithm to find the optimal MMAX solution can then be defined as follows:

Algorithm 1: Optimal bit allocation in a VBR channel with channel constraints under a MMAX criterion

[Step 0]: Initialize buffer occupancy (B_i) by quantizing all frames with the coarsest quantization available to each frame.

[Step 1]: Find the frame that has maximum distortion and decrease the quantization step size of that frame.

[Step 2]: If the buffer state satisfies the condition in (2) for all i then go to Step 1, otherwise STOP. The frame that has maximum distortion is the frame whose quantization changed just before STOP.

Fig. 2 shows two simple examples of VBR transmission with TB policing in a MMAX criterion. As shown in (a), even though more data are stored in the buffer, the amount of data transmitted is always lower than the middle diagonal lines in Fig. 2 since the available maximum rate is determined by remaining tokens in the TB and new incoming tokens in a frame interval (i.e., $TB_i + C$). In (b), the constant slope between the 2^{nd} frame and the 6^{th} frame indicates the peak rate. As shown in the figure, the amount of transmitted data is limited even though more tokens are available.

The MMAX solution may result in TB overflow, especially when easily compressed frames are coded successively. Note that

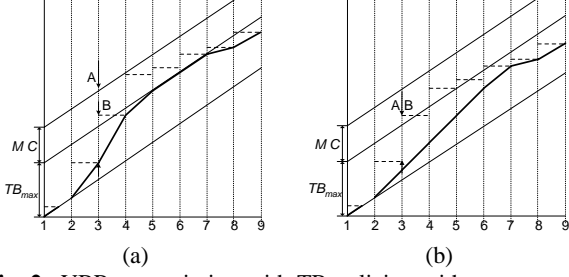


Fig. 2. VBR transmission with TB policing with parameters (C , TB_{max} , P) in a MMAX criterion. Horizontal axes indicate time in frame units and vertical axes indicate the size of transmitted data. Horizontal dashed lines indicate the solutions in a MMAX criterion and the slopes of thick lines indicate transmission rate of each frame interval. (a) is the case that the peak rate is high enough not to be a constraint (i.e., $M \cdot P \geq TB_{max} + M \cdot C$) and (b) is the case that the peak rate is used as a constraint.

tokens that are dropped due to TB overflow cannot be used for future data transmission. In this paper, we propose to use “spare” tokens due to TB overflow in order to decrease MSE. We term this the MMAX+ approach as the MMAX solution is improved upon with an additional MSE criterion.

3. OPTIMAL RATE CONTROL IN A MMAX+ CRITERION

After finding the MMAX solution, the problem we are trying to solve can be formulated as

$$\min_{q_i} \left(\sum_{i=1}^S D_i \right) \text{ s.t. } B_i^M \leq B_i \leq B_i^U \text{ for all } i, \quad (7)$$

where B_i^U represents the right side of (2) and B_i^M represents the buffer state at the i^{th} frame interval when the MMAX solution is used. Note that the additional step we propose can only increase the number of bytes used since the MMAX solution is used as the initial condition. For example, in Fig. 2 (a) and (b), “A” and “B” indicate B_3^U and B_3^M respectively (in (b), “A” and “B” are same.).

Note that although B_i^U in (7) is lower than B this does not mean we can allocate an additional $B_i^U - B_i^M$ bytes. This is because the trace B_i^M already incorporates the effect of transmitted data, and thus additional data allocated by the encoder do not result in additional channel rate (i.e., they need to exploit instances of token buffer overflow when available transmit capacity was wasted.) Thus any increase to B_i over B_i^M leads to decreasing tokens in the TB for all $i' \geq i$, so that token underflow could occur for i' , even if it does not occur for i . To reduce the upper bound of the buffer state of a frame, we introduce the concept of effective buffer size (EBS), where the EBS of a frame is the maximum additional rate that can be used to increase the quality of the frame and such that no token underflow and no violation of the condition in (2) occur. To find the EBS, we first consider the case that P is not a constraint. We denote the remaining tokens after transmitting the data determined by the MMAX solution for a frame interval as RT (i.e., $RT_i = TB_i + C - C_i^M$, where RT_i is the RT of the i^{th} frame and C_i^M is the transmission rate in the i^{th} frame interval determined by the MMAX solution) then obviously EBS_i (the EBS of the i^{th} frame) is smaller than or equal to RT_i , which varies from frame to frame. Examples of computation of the EBS are shown in Fig. 3. In the figure, the EBS of frame “b” is determined by the minimum RT of all frames from “b” onwards. This is because

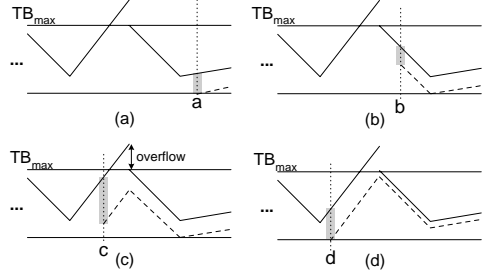


Fig. 3. Example of EBS computation of different frames. The solid line represents the TB states of a MMAX solution. The height of the gray box is the EBS of the given frame and dashed lines show that the determined EBS does not induce TB underflow. The EBS of frame “a” is determined by the remaining tokens of the frame and that of frame “b” is determined by the remaining tokens of a following frame. The EBS of frame “c” is determined by the sum of the amount of TB overflow and the EBS of a frame after overflow. The EBS of frame “d” is determined by the remaining tokens of the frame since it is smaller than the EBS of frame “c”.

additional bytes used at “b” will reduce the remaining tokens of all following frames (dashed lines in Fig. 3 (b)). However, if TB is in overflow at the i^{th} interval then the amount of token overflow (TO_i , where $TO_i = \max(TB_i + C - C_i^M - TB_{max}, 0)$) can also be added to the bit-budget of the i^{th} frame without affecting the TB state of future frames (See Fig. 3 (c)). As shown in these examples, given EBS_{i+1} , EBS_i is computed as $EBS_i = \min(RT_i, TO_i + EBS_{i+1})$ since RT_i and EBS_{i+1} guarantee no TB underflow in the current frame and all future frames, respectively. Therefore the EBS can be computed from the last frame by using the following equation.

$$EBS_i = \begin{cases} RT_i & : i \text{ is the last frame,} \\ \min(RT_i, TO_i + EBS_{i+1}) & : \text{otherwise,} \end{cases} \quad (8)$$

Under the peak rate constraint, the encoder buffer size is restricted by $M \cdot P$ and the residual buffer size after finding the MMAX solution of the i^{th} frame (RBS_i) is computed by $RBS_i = M \cdot P - B_i^M$. Therefore the minimum of RBS_i and RT_i can be added to the MMAX solution without any violation in the current frame. After changing RT_i in (8) to $\min(RT_i, RBS_i)$, EBS_i can still be computed by (8).

After computing the EBS for all frames, the problem in a MMAX+ criterion is redefined as

$$\min_{q_i} \left(\sum_{i=1}^S D_i \right) \text{ s.t. } B_i^M \leq B_i \leq EBS_i + B_i^M \text{ for all } i, \quad (9)$$

This new formulation now guarantees that increasing B_i does not lead to TB underflow and any violation of the condition in (2).

The optimal rate control problem under the MMAX+ criterion can be solved by using a dynamic programming (DP) method [3] or a Lagrangian optimization method [2]. Obviously other techniques are possible to find faster approximate solutions but in this paper we provide a result with an optimized DP method to provide a fair comparison between MMAX+ and MSE solutions. In the DP method which we use in this paper, given the buffer and channel constraints due to our goal to preserve the MMAX solution, the number of states can be reduced significantly by computing the EBS. The complexity of the MMSE algorithm is proportional to the number of states (B and TB_{max}) and quantization levels

Table 1. Performance (PSNR) comparison of the proposed MMAX and MMAX+, and MMSE optimal solutions. Used token rate is 1.25M/sec (i.e., 625K per a GOP interval), the maximum delay is 4 GOP intervals and the size of a TB is 2.5 Mbytes. Initial and final TB and buffer states are at mid-buffer. In CBR transmission, the token rate is used as channel rate.

Method	Avg.	Std. Dev.	Min.	Max.
MMAX	38.422	0.140	38.373	39.440
MMAX+	38.583	0.489	38.373	40.800
MMSE	38.733	1.505	35.260	42.753
CBR MMAX	38.209	0.137	38.153	39.440
CBR MMAX+	38.598	0.847	38.153	42.567

Table 2. Performance (PSNR) comparison when the maximum delay, TB size and peak rate are changed with respect to the settings of Table 1. In the “Method” column, M indicates that the maximum delay is half that in Table 1, TB indicates TB size half that in Table 1, and P indicates that the peak rate is $1.5 \cdot C$. In each case the remaining parameters are not modified.

Method	Avg.	Std. Dev.	Min.	Max.
MMAX (M)	38.312	0.158	38.260	39.440
MMAX+ (M)	38.593	0.804	38.260	42.567
MMSE (M)	38.732	1.502	35.260	42.567
MMAX (TB)	38.280	0.163	38.227	39.440
MMAX+ (TB)	38.590	0.831	38.227	42.567
MMSE (TB)	38.730	1.492	35.260	42.753
MMAX (P)	38.209	0.175	38.153	39.440
MMAX+ (P)	38.599	0.902	38.153	42.567
MMSE (P)	38.724	1.497	35.260	42.753

(Q) [7]. Since the EBS and the number of allowable quantization levels are much smaller than B (and $T B_{max}$) and Q respectively, the complexity to find the optimal solution under a MMAX+ criterion is much lower than that under a MMSE criterion.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In order to verify the performance of the proposed algorithm, we implement this algorithm and test it with 1800 frames from the “Fire Birds” movie sequence. We use the Group of Pictures (GOPs) in MPEG as the basic data unit and the “closed GOP” option in MPEG2 is used to code each GOP independently. To gather R-D data of GOPs, each GOP is coded by using 159 different rates roughly between 125 Kbytes and 1.585 Mbytes (The difference between steps is roughly 10K bytes.). Total sequence has 120 GOPs (each GOP has 15 frames) and each GOP is coded by using the MPEG2 TM5 rate control.

Table 1 shows the experimental results for each criterion. As expected, the minimum PSNR of the MMAX solution is higher than that of the MMSE solution. The MMAX+ criterion improves the average PSNR around 0.16 dB, where the difference between the average PSNR of the MMAX solution and that of the MMSE solution is 0.31 dB. The standard deviation of the MMAX solution shows that the PSNR of each GOP is very similar but the maximum PSNR is relatively high. The reason for this is that some GOPs have simple contents and so the PSNR of these GOPs at minimum rate (125 Kbytes) determines the maximum PSNR. The minimum PSNR of the MMAX solution in VBR transmission is higher than that in CBR transmission under the same maximum

delay constraint since tokens can be stored in the TB for future use (CBR transmission can be viewed as VBR transmission with TB policing, where TB size is zero.). Although the complexity to find the MMAX+ solution highly depends on the token bucket states of the MMAX solution, in this experiment, the complexity of the MMAX+ algorithm is roughly 20 times lower than that of the MMSE algorithm.

In Table 2, the performance of each method under lower maximum delay, smaller token bucket size, lower peak rate is compared. Because these parameter changes give higher constraints to the problem, the allowable encoder buffer size is reduced as in (2) and (3) and severe local fluctuation of the bit-rate of GOPs cannot be absorbed. Therefore the minimum PSNR of the MMAX solution is decreased and the additional bit-budget available due to the TB overflow is increased. Since the bit-budget for the MMAX+ solution is increased, the average PSNR can be improved and so the difference of average PSNR between MMAX+ and MMSE solutions is decreased.

5. CONCLUSIONS

In this paper, we developed the optimal bit allocation algorithm of VBR transmission in MMAX and MMAX+ criteria. The MMAX+ criterion is introduced to improve total quality by using the remaining channel bandwidth under the MMAX criterion. Also an algorithm for finding the effective buffer size is proposed. The effective buffer size is used to reduce the number of possible states of each frame and as a result, the complexity of the algorithm to find the optimal MMAX+ solution is reduced.

6. REFERENCES

- [1] A. R. Reibman and B. G. Haskell, “Constraints on variable bit-rate video for ATM networks,” *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 2, pp. 361–372, Dec. 1992.
- [2] J.-J. Chen and D. W. Lin, “Optimal bit allocation for coding of video signals over ATM networks,” *IEEE JSAC*, vol. 15, pp. 1002–1015, Aug. 1997.
- [3] C.-Y. Hsu, A. Ortega, and A. Reibman, “Joint selection of source and channel rate for VBR video transmission under ATM policing constraints,” *IEEE Journal on Select Areas in Communications*, vol. 15, pp. 1016–1028, Aug. 1997.
- [4] A. Ortega, M. W. Garrett, and M. Vetterli, “Rate constraints for video transmission over ATM networks based on joint source/network criteria,” in *Annales des Télécommunications*, Jul.-Aug. 1995, vol. 50, pp. 603–616.
- [5] S. Shenker, C. Partridge, and R. Guerin, “Specification of guaranteed quality of service,” *Internet Engineering Task Force RFC 2212*, Sep. 1997.
- [6] Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers,” *IEEE Trans. on Signal Proc.*, vol. 36, pp. 1445–1453, Sept. 1988.
- [7] A. Ortega, K. Ramchandran, and M. Vetterli, “Optimal trellis-based buffered compression and fast approximation,” *IEEE Trans. on Image Proc.*, vol. 3, pp. 26–40, Jan. 1994.
- [8] G. Schuster, G. Melnikov, and A. Katsaggelos, “A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers,” *IEEE Trans. on Multimedia*, vol. 1, pp. 3–17, Mar. 1999.
- [9] D. T. Hoang, E. L. Linzer, and J. S. Vitter, “Lexicographic bit allocation for MPEG video,” *J. of Visual Commun. Image Presentation*, vol. 9, no. 4, pp. 384–404, Dec. 1997.
- [10] S.-Y. Lee, “Color processing and rate control for storage and transmission of digital image and video,” *Ph.D dissertation proposal, University of Southern California*, 2002.