A FLEXIBLE RATE CONTROL MECHANISM FOR INTERACTIVE MPEG-2 VIDEO SERVICES OVER ATM NETWORKS

Olivier Verscheure, Philippe Chabloz and Jean-Pierre Hubaux

TCOM Laboratory, Telecommunication Services Group
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
E-Mail: verscheure@tcom.epfl.ch
URL : http://tcomwww.epfl.ch/~verscheu

ABSTRACT

In this paper, we present a new rate control mechanism for delay sensitive MPEG-2 video services over ATM networks. The rate control constraints the video bitstream to respect the three negotiated VBR ATM traffic contract parameters modelled by a dual leaky bucket. We first develop several new mathematical relationships that permit to predict the bit rate on a frame-by-frame basis. Then a rate control algorithm, based on both a sustainable-rate control and an instantaneous-rate control, is derived. The contribution mainly lies on the rate control *flexibility*. Indeed, the resulting mechanism is independent of the GOP structure and length. Moreover, a macroblock-based adaptive quantization scheme may be incorporated without degrading its performance. Simulations were performed on several TV-resolution sequences. The proposed mechanism appeared to behave consistenly; for instance, the quantizer scale, derived from the rate control on a frame-by-frame basis, converges very quickly to a stable value.

1. INTRODUCTION

The Asynchronous Transfer Mode (ATM) technology is reaching a certain level of maturity that permits its deployment in local as well as in wide area networks. One of the key points of such a technology is that it may provide statistical guarantees on performance. In other words, each user may agree with the network on a set of performance parameters (i.e., Quality of Service -QoS- parameters) which will have to be met. It is therefore mandatory to implement traffic control procedures in broadband networks to ensure that the traffic characteristics are as declared. The traffic descriptors considered by the ATM Forum include the peak cell rate (PCR), the sustainable cell rate (SCR) and the intrinsic burst tolerance (IBT) defined with respect to the Generic Cell Rate Algorithm [1].

Audiovisual applications (e.g., video conferencing, video on demand, teleteaching, etc.) are foreseen as one of the major users of such broadband networks. At the heart of this revolution is the digital compression of audio and video signals. The biggest advantage of compression resides in data rate reduction which results in reduction of transmission costs. The choice of the compression algorithm mostly depends on the available bandwidth or storage capacity and the features required by the application. The MPEG-2 [1] standard [2], a truly integrated audio-visual standard developed by the International Organization for Standards (ISO), is capable of compressing NTSC or PAL video into an average bit rate of 3 to 6 Mbits/s with a quality comparable to analog CATV [3].

Video coding naturally produces a varying bit rate since not all video pictures have the same entropy [4]. It means that variable bit rate (VBR) coding of video performs usually better than the traditionally used constant bit rate (CBR) scheme. Furthermore, VBR transmission may provide more efficient use of network bandwidth through statistical multiplexing. However, the evolution of the entropy is not known a priori. Therefore, it seems particularly difficult to guarantee the QoS in a VBR video connection by characterizing the encoder output. However, for certain applications such as video conferencing, studies have shown that it was possible to characterize the output succinctly in terms of a small number of parameters (e.g., [5]). Another approach consists in modifying the encoder output so that

[1] MPEG stands for Moving Picture Experts Group

it becomes compliant with the negotiated traffic characteristics [6, 7, 8]. Rate control is considered as an important issue in video coding since it significantly affects video quality.

This work focuses on a flexible rate control mechanism for interactive MPEG-2 video streams to be transmitted over an ATM network. The paper is organized as follows: Section 2 first introduces the MPEG-2 video standard and then briefly presents traffic control for VBR connections in ATM networks. The proposed MPEG-2 VBR rate control mechanism is the subject of Sec. 3 which starts with some preliminary mathematical studies. Section 4 presents some experimental results and, finally, concluding remarks are given in Sec. 5.

2. VBR MPEG-2 VIDEO OVER ATM

2.1. The MPEG-2 Video Standard

An MPEG-2 video stream is hierarchically structured. The stream consists of a sequence which is composed of several pictures. Each picture is composed of slices which are, by definition, a series of macroblocks. Each macroblock (16×16 pixels) contains 4 blocks (8×8 pixels) of luminance and 2, 4 or 8 blocks of chrominance depending on the chroma format.

The MPEG-2 video syntax defines three different types of pictures :

- Intra coded pictures or I-pictures are coded without reference to preceding or upcoming pictures in the sequence. The picture is actually divided into 8×8 blocks of pixels. The 2-D Discrete Cosine Transform (DCT) is then applied on each block. The resulting DCT coefficients are finally quantized. The quantizer, applied to each DCT coefficient, comes from the multiplication of a *Quantizer Scale*, Q [2], and the corresponding element of a *Quantizer Matrix*. A different Q value may be used for each macroblock. This whole process is referred to the DCT-based intraframe encoding technique. I-pictures provide moderate compression ratios while producing potential random access points into the compressed video data.

- Predicted (or P-) pictures are coded with respect to the nearest previous I or P-picture. The predictions of the best-matching prediction macroblocks are indicated by motion vectors that describe the displacement between them and the target macroblocks (which in this case correspond to the macroblocks of the nearest previous I or P-picture). The prediction error is then encoded using the DCT-based intraframe encoding technique summarized previously. This technique is called the motion compensation technique.

- Bidirectional (or B-) pictures use both a past and a future picture as a reference which may be an I- or a P-picture. The motion compensation technique is also applied. However, in this case, two motion vectors may be associated to each macroblock since B-pictures are using two reference pictures. B-pictures provide the highest compression ratio.

The use of these three picture types allows MPEG-2 to be robust (I-pictures provide error propagation reset points) and efficient (B- and P-pictures allow a good overall compression ratio). Furthermore, the MPEG-2 standard does not specify how I-, P- and B-pictures have to be mixed together (i.e., the standard allows Group of Pictures, GOPs, to be of arbitrary structure and length).

2.2. Traffic control in ATM Networks

ATM networks offer the possibility to support source coding at a variable bit rate, which involves the potential provision of consistent picture quality, bandwidth savings and delay reduction. However, in order to guarantee the QoS inherent to the ATM philosophy, a traffic control must be implemented. The ATM Forum recommendations for the user-network interface specify a mechanism for describing the traffic flowing through a virtual circuit connection. This mechanism is the Generic Cell Rate Algorithm (GCRA), which is equivalent to a leaky bucket (LB). The LB scheme is a widely accepted policing function [9]. Figure 1 represents the leaky bucket, LB(R_s, B_{max}), control diagram. Packets accumulate at the input traffic rate in a bucket (virtual buffer) that can store up to B_{max} packets and empty at a constant rate R_s. When a packet arrives and the bucket is not full, the bucket size is incremented by the size of a packet and the packet is accepted. However, if the bucket is full when the packet comes in, the packet is either marked as being of low priority or rejected.

[2]For reasons of convenience, we decided to refer the Quantizer Scale by Q and not by $MQUANT$

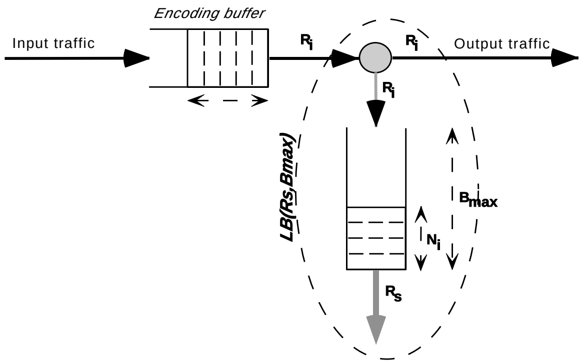

Figure 1: *Leaky Bucket Control Diagram.*

As stated in the introduction, it appears difficult to find a suitable stochastic model to characterize the MPEG-2 encoder output. We then considered the alternative approach consisting in enforcing its deterministic characteristics in order to respect the dual leaky bucket [3] parameters corresponding to the negotiated service contract.

For reasons of convenience, we will consider the three DLB parameters, (R_p, R_s and B_{max}), as being expressed in bits per frame period. Therefore, from the coding side, the following equations must be verified in order to avoid the buckets to overflow:

$$N_i < B_{max} \iff \begin{cases} R_i < B_{max} - N_{i-1} + R_s \\ R_i < R_p \end{cases}$$

where (R_p, R_s, B_{max}) are the DLB parameters, N_i represents the bucket fullness after the ith frame transmission and R_i is the number of bits associated to the encoded version of frame i.

3. THE PROPOSED RATE CONTROL MECHANISM

3.1. Preliminary Studies

The design of a rate control mechanism would become straightforward if it were possible to predict precisely, at any time, the number of bits a frame, of a certain type, would require to be coded at a given perceptual quality. However, the non-stationnary nature of video signals makes it almost impossible. Therefore, we have to approximate this number of bits by, for instance, exploiting the relationship that exists between both the bit rate associated to a given frame [4] and the average quantizer scale, Q, used to encode this frame.

Let $R(GOP_k, Q_k)$ be the bit rate of the kth GOP when the quantizer scale, applied to all the frames and macroblocks of this GOP, is set to Q_k. Hamdi and Roberts [6] observed that $R(GOP_k, Q_k).Q_k$ was approximately independent of Q_k for MPEG-1 video streams. However, the MPEG-2 standard has been developed for video sequences of higher motion and higher resolution than that for which MPEG-1 was designed. Furthermore, our purpose is to elaborate a rate control which operates on a frame-by-frame basis. Therefore, when considering MPEG-2 video rates at the frame level $(R(frame_k, Q_k))$, this relationship is no longer accurate as illustrated in Fig. 2. Indeed, this graph shows that the $R(frame_k, Q_k).Q_k$ product is no longer equal to a constant value χ_k. Figure 3 presents the same result in another way. The log-log scale clearly exhibits the relationship between $R(frame_k, Q_k)$ and Q_k. Indeed, if $R(frame_k, Q_k).Q_k = \chi_k$ then $LOG[R(frame_k, Q_k)] = LOG[\chi_k] - LOG[Q_k]$. However, we see that the slope is not equal to 1. It is therefore necessary to modify the existing equation by introducing a new parameter, β_k, in:

$$R(frame_k, Q_k).{Q_k}^{\beta_k} = \chi_k, \qquad (1)$$

where χ_k represents a complexity measure of frame k.

For a given frame i, this equation becomes:

$$R1_i.Q1_i^{\beta_i} \simeq R2_i.Q2_i^{\beta_i} \simeq R3_i.Q3_i^{\beta_i} \simeq ...$$

The optimal β value, β^{opt}, for each frame may easily be found by means of a least squares method. However, it would involve several loops in the encoding of each frame which is not feasible in an interactive video service. Therefore, we need to find a way to estimate this value by means of another measure. Since β is intimately related to the picture's complexity, we considered the SITI (Spatial Information-Temporal Information) metric introduced in [10]. Basically, the SI and TI measures for the kth frame are defined as :

$$\begin{aligned} SI[Frame_k] &= STD_{space}\{Sobel[Frame_k]\}, \\ TI[Frame_k] &= STD_{space}\{\Delta Frame_k\}, \end{aligned}$$

[3] The dual leaky bucket (DLB) is equivalent to the LB but uses a second bucket (i.e., LB(R_p, 0)) which imposes the maximum packet rate at which the source may transmit the packets

[4] The bit rate of a frame is defined as the number of bits associated to this frame divided by the time slot it occupies (e.g., $\frac{1}{25}^e$ s. for the PAL format)

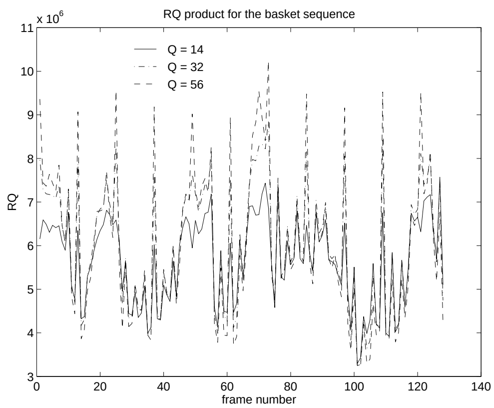

Figure 2: $R(frame_k, Q_k).Q_k$ *product versus the frame number for the BasketBall sequence. Three different* Q_k *have been used: 14, 32 and 56.*

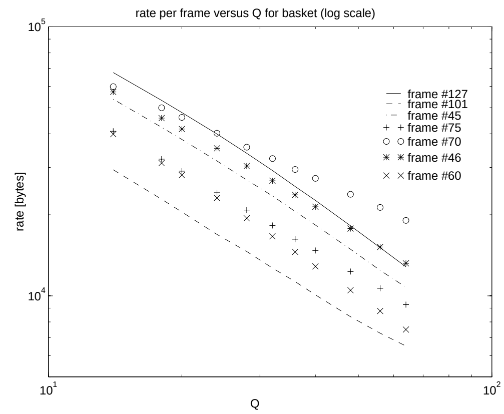

Figure 3: $R(frame_k, Q_k)$ *versus* Q_k *in a log-log scale for seven frames of the BasketBall sequence. Eleven* Q_k *have been used from 14 to 56.*

where, STD_{space} is the standard deviation operator over the horizontal and vertical spatial dimension in a frame, *Sobel* is the Sobel filtering operator and $\Delta Frame_k = Frame_k - Frame_{k-1}$.

The best estimation of the β value has been determined by applying a linear interpolation between $SI[Frame_k]$ and the optimal β_k^{opt} for each $Frame_k$. For this purpose, we used four different sequences (BasketBall, Ski, Mobile&Calendar and FlowerGarden) in order to cover a very large set of SI values. Basically, we then obtained β as a linear function of SI for each picture type (I, P and B).

Furthermore, the quantizer value may be modulated over the picture around a constant Q^* while introducing a negligible error of prediction. The equivalent quantizer scale, Q, is therefore computed, using Eq. 1, by:

$$Q = q.\left[\frac{R(frame, q)}{R(frame, Q^*)}\right]^{\frac{1}{\beta}},$$

where Q^* is the quantizer scale around which the modulation is performed and q corresponds to any quantizer scale (e.g., $q = 28$).

This flexibility allows, for example, the utilization of a macroblock-based adaptive quantization scheme like the one introduced in the MPEG-2 Test Model v5 specification.

Up to now, we proposed a method to predict the bit rate of a frame coded at a quantizer scale $Q1$ based on the bit rate of the *same* frame coded at $Q2$. In other words, we need to encode each frame twice which still does not meet the delay requirement. Therefore, we elaborated some other equations to predict the bit rate of a frame which will be coded with a quantizer $Q1$. This prediction is based on the bit rate of the *previous* frame of the same type (I, P or B) and the TI values of these both frames.

The performance achieved by the new relationships presented in this section are shown in Fig. 4. It presents the relative error of the frame-based rate prediction for $Q^* = 56$ given the bit rate at $Q^* = 14$ of the previous frame of the same type. The sequence is composed of 400 frames of the BasketBall, Ski, Mobile&Calendar and FlowerGarden sequences. We used the same macroblock-based modulation as the one introduced in the MPEG-2 TM5 which is based on a simple activity metric. These three graphs show the improvement obtained in the bit rate prediction for I-, P- and B-frames by introducing the β and TI parameters. Moreover, quantizer scales of 14 and 56 may be considered as two extremes since an average quantizer scale of 56 usually gives a bad perceptual quality while $Q = 14$ produces a decoded video sequence very acceptable for a user.

3.2. The Rate Control Mechanism

We now introduce the main features of the proposed MPEG-2 rate control mechanism based on the DLB parameters presented in Sec. 2.2. It makes use of the new equations developed in the previous section. One of the key aspects of this mechanism resides in its flexibility mainly due to

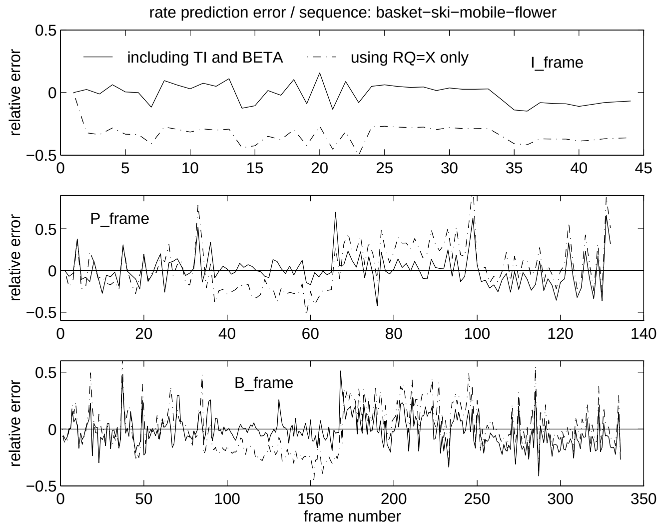

Figure 4: *Relative error of the frame-based bit rate prediction for a $Q = 56$ given the bit rate at $Q = 14$ of the previous frame of the same type (I, P or B). This sequence is composed of 400 frames of the BasketBall, Ski, Mobile&Calendar and FlowerGarden sequences.*

its frame-based design [5]. For example, the rate control is independent of the GOP structure and length. Moreover, this mechanism may be used in a real-time encoding scheme due to its low complexity.

The rate control algorithm is mainly divided into two parts: the sustainable-rate (SR) control which adapts the quantizer scale, Q, by observing the leaky bucket fullness and the instantaneous-rate (IR) control which modifies Q when the latter has been under-evaluated by the previous control. The IR control enables the bucket not to overflow in case of emergency and is then only activated if necessary. The SR control operates every N frames (N is related to the number of times the IR control has to be activated). For this purpose, we used a PID controller based on the bucket fullness history on the last N frames. Both the SR control and the IR control intensively use the equations introduced in Sec. 3.1 for bit rate predictions.

When a scene change occurs, the SR control is immediately activated and a new Q is computed. A scene change is detected by means of the TI metric.

The proposed rate control mechanism allows for the use of a macroblock-based adaptive quantization scheme without deteriorating its performance. The adaptive quantization may be, for instance, based on a perceptual activity metric such as the one presented in [11]. Moreover, no a priori has been made concerning the GOP structure which means that a GOP-selection mechanism may easily be incorporated without modifying the rate control behavior.

4. EXPERIMENTAL RESULTS

We implemented the rate control mechanism in the TM5 MPEG-2 software encoder . Simulations were performed on several sequences (e.g., BasketBall, Mobile&Calendar, Ski, FlowerGarden) conforming to the ITU-R BT.601 format (720*576, 25 fps). These sequences were coded as interlaced video with different DLB parameters, with a structure of 12 images per GOP and 2 B-pictures between every reference picture, and a single slice per line. The TM5 macroblock-based adaptive quantization scheme has also been used.

During these tests, we never experienced any buffer overflows or other chaotic behaviors. Moreover, we noted a good stability and a very quick convergence of Q. Since the Q value is stable within a given scene, we obtained an interesting consistent quality. However, the global perceptual quality is not necessarily constant. This is mainly due to the rate control design which does not incorporate a quality control mechanism. We believe that such a mechanism has to take place on top of the rate control mechanism. Indeed, the main goal of a rate control mechanism is to optimize the leaky bucket fullness since the DLB parameters (i.e., the traffic descriptors) will be directly related to the service price. The addition of a quality control mechanism depends mostly on the user requirements.

Figure 5 illustrates the performance of the presented mechanism on a 400 frames long composite sequence built with 100 frames of each of the following sequences respectively: BasketBall, Mobile&Calendar, Ski and FlowerGarden. This sequence was coded with a $R_p = 8$ MBits/s, $R_s = 3.5$ MBits/s and $B_{max} = 600$ kBits. These three graphs respectively show the controlled bit rate, the buffer occupancy and Q as a function of the frame number. In addition to what has already been pointed out, we may notice, for instance, that the bucket is very well utilized since it varies consistently between 0 and its maximal value. Moreover, we see that, after each scene change, Q is very quickly adapted to a stable

[5] The prediction method has been designed on a frame by frame basis

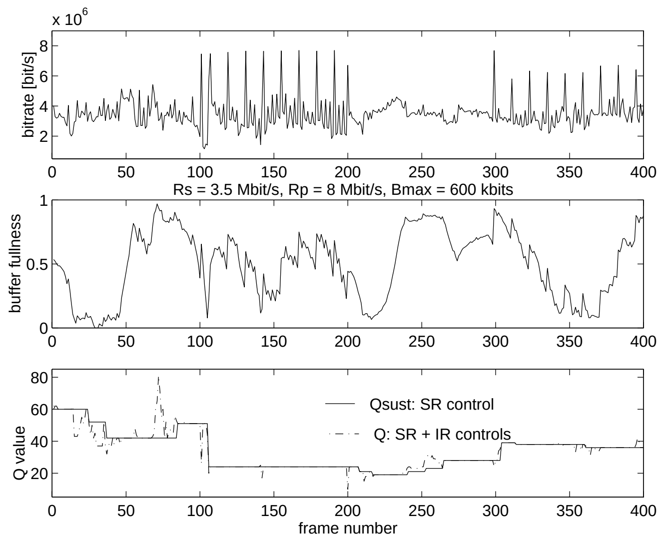

Figure 5: *Controlled bit rate, buffer occupancy and Q versus the frame number. The sequence is composed of 400 frames of the BasketBall, Ski, Mobile&Calendar and FlowerGarden sequences.*

value (after an average of 4 frames).

5. CONCLUSION

In this paper, we have presented a new frame-based rate control mechanism for delay sensitive MPEG-2 video services over ATM networks. The key design of this mechanism is that it imposes the video bitstream to optimally utilize the three negotiated VBR ATM traffic contract parameters.

We first developed several mathematical relationships to accurately predict the bit rate of a frame given the bit rate of the previous frame. We then introduced the proposed rate control mechanism mainly composed of both a sustainable-rate control and an instantaneous-rate control. The resulting mechanism appeared to behave consistenly for several TV-resolution sequences.

Further works will be carried out in order to elaborate a perceptual quality control mechanism to be placed on top of the proposed rate control mechanism.

6. REFERENCES

[1] T. A. Forum, "Traffic Management Specification." April 1996.

[2] I. J. 1, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1, 2 and 3.* ISO/IEC JTC 1, 1994.

[3] Barry G. Haskell, Atul Puri and Arun N. Netravali, *Digital Video: an Introduction to MPEG-2.* Digital Multimedia Standards Series, Chapman and Hall, 1997.

[4] D. Reininger, et al., "Variable bitrate mpeg video : Characteristics, modeling and multiplexing," in *Proceedings of the ITC-14*, pp. 295–306, Elsevier, 1994.

[5] D. P. Heyman, A. Tabatabai and T. V. Lakshman, "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, pp. 49–59, March 1992.

[6] M. Hamdi and J. W. Roberts, "QoS Guaranty for Shaped Bit Rate Video Connections in Broadband Networks," in *IEEE Computer Society Press*, pp. 153–162, september 1995.

[7] Wei Ding and Bede Liu, "Joint Encoder and Channel Rate Control of VBR Video over ATM Networks," in *SPIE Electronic Imaging - Digital Video Compression*, vol. 2668, January 1996.

[8] Chi-Yuan Hsu, Antonio Ortega and Amy R. Reibman, "Joint Selection of Source and Channel Rate for VBR Video Transmission under ATM Policing Constraints," *IEEE Journal on Selected Areas in Communications*, 1997. accepted for publication.

[9] M. Dirksen and K. van der Wal, "Safety Margins in ATM Policing Function Algorithms," *IEEE Transactions on Communications*, vol. , 1992.

[10] A. Webster, C. Jones, M. Pinson, S. Voran and S. Wolf, "An Objective Video Quality Assessment System Based on Human Perception," in *SPIE - Human Vision, Visual Processing and Digital Display*, vol. 1913, pp. 15–26, 1993.

[11] O. Verscheure and C. van den Branden Lambrecht, "Adaptive Quantization Using a Perceptual Visibility Predictor," in *Proceedings of the International Conference on Image Processing*, (Santa Barbara, CA), October 26-29 1997. available on `http://tcomwww.epfl.ch/verscheu/`.