INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS
Int. J. Commun. Syst. 2001; **14**:125–146

The transmission of MPEG-2 VBR video under usage parameter control

Lou Wenjing, Chia Liang Tien*,†, Lee Bu Sung and Wang Bin

School of Applied Science, Nanyang Technological University, Singapore 639798, Singapore

SUMMARY

This paper studies the transmission of MPEG-2 VBR video over ATM network under usage parameter control. The idea is to seek a compromise between the network utilization and the quality of video service by applying UPC-based rate control strategies to the video source. A modified leaky bucket algorithm is proposed to calculate the constraints on the bit-rate guaranteeing conformance to peak cell rate, sustainable cell rate and burst tolerance usage parameters. Two rate control strategies, one for real-time generated video coding and the other for pre-compressed video, are proposed for MPEG-2 VBR video. The rate control strategies control the video source to generate traffic conforming to the constraints on the bit rate. The experimental results show that both the UPC-based rate control strategies can provide lossless transmission from the source perspective as well as to reduce the burstiness of the traffic. To keep within the bit-rate allowed, the control method uses coarser quantization to maintain better picture quality than that by removing the number of AC transformed coefficients. The slight degradation of picture quality caused by the source rate control is preferable than the severe drop of picture quality caused by the cell loss at UPC.

KEY WORDS: ATM networks; usage parameter control; leaky bucket algorithm; variable bit rate; MPEG-2; video; source rate control

1. INTRODUCTION

A lot of current research has been focused on VBR video traffic transmission strategy in ATM networks. A key problem is how to achieve the two main conflicting goals: good quality of service and high network utilization. Due to its burstiness and delay-sensitive nature, it is fundamentally difficult to achieve both goals with VBR video traffic. On the one hand, if the user requires a bandwidth much more than the traffic actually offered, the quality of service will be satisfied while the network utilization suffers. On the other hand, since a VBR traffic source generates various amount of data during different time periods, it may occasionally generate a very big

* Correspondence to: Chia Liang Tien, Division of Computing Systems, School of Applied Science, Nanyang Technological University, Nanyang Avenue, Singapore 6397998, Singapore.
† E-mail: asltchia@ntu.edu.sg

Received 4 October 1998
Accepted 8 September 1999

amount of data. If the aggregate amount of traffic exceeds the bandwidth allocated to the connection, cells not conforming to the traffic contract will be tagged or dropped at UPC, which will in turn cause the application's quality of service to suffer.

Concerning real-time generated VBR video transmission, there are two basic approaches to such traffic control problem: either find a suitable characterization of coder output enabling QoS prediction or oblige the coder to make its output conform to predefined traffic contract. For pre-compressed VBR video transmission, there are similar approaches: either to define a suitable traffic contract according to the characteristics of the existing video to guarantee the QoS, or conduct source rate control during some abnormally bursty periods to lower down the bandwidth requirements of the whole connection duration.

There has been some work done based on the first approach. The output of certain applications such as video conferencing may be characterized by a Markov model which needs a small number of parameters to evaluate the performance of a network multiplexer [1]. The traffic characteristics of pre-compressed video such as in a video-on-demand (VoD) environment can be evaluated in advance [2]. However, to characterize less stereotyped video is particularly difficult. Indeed, long video sequences seem to systematically exhibit long-range dependence whose significant detrimental impact on performance is beginning to be well understood [3]. The alternative approach, which is to control the output rate of the encoder so that it conforms to certain traffic constraints, has been considered by fewer researchers [4–6]. Our contribution belongs to this category.

It is generally recognized that some forms of source traffic control should be conducted to enhance the performance of the whole transmission. In this paper, the source rate control of MPEG-2 VBR traffic is studied. For real-time generated VBR video, we tend to seek a compromise between pure open-loop coding and rate controlled coding which maintains available picture quality while satisfying certain traffic constraints. For pre-compressed video, we try to selectively discard some less important components from the video stream during the transmission so that a set of more efficient traffic constraints can be observed. The traffic constraints used in this study are imposed by the UPC, which is taken by the ATM networks to monitor and control traffic in terms of traffic offered and validity of the ATM connection at the user access. A modified leaky bucket algorithm is proposed to monitor the bit-rate generated and dynamically calculate the constraints on the consequent bit-rate based on certain pre-defined traffic parameters, including peak cell rate (PCR), sustainable cell rate (SCR), and burst tolerance (BT). A new rate control strategy, named re-quantization rate control strategy, is proposed for real-time video applications to generate the conforming bit rate. Another rate control strategy, consisting of an AC-removal PCR rate control method and a selectively frame dropping rate control method, is proposed to conduct the rate control purpose for pre-compressed video.

The remainder of this paper is organized as follows. In Section 2, the UPC function in ATM network and the leaky bucket algorithm explanation of the generic cell rate algorithm are briefly described. A modified leaky bucket algorithm is proposed to calculate the rate constraints imposed by the UPC function. Next, in Section 3, the two new rate control strategies, respectively, for real-time generated video coding and pre-compressed video, are described and applied to MPEG-2 VBR video streams. The purpose of both the schemes is to limit the bit-rate transmitted to the network within the bandwidth constraints. Some experimental results will be shown in Section 4. The performance of the controlled VBR transmission is compared with pure VBR in terms of network utilization and quality of picture. Finally, in Section 5, the conclusion is drawn.

2. USAGE PARAMETER CONTROL AND CONSTRAINTS ON SOURCE BIT RATE

2.1. Usage parameter control and leaky bucket algorithm

Traffic contracts are an integral element of high-speed ATM networks that can guarantee the QoS for user. A contract will constrain the associated traffic in one, and perhaps multiple ways, e.g. specifying traffic parameters such as peak cell rate (PCR), SCR and BT. The traffic management of ATM network uses UPC functions to monitor cell streams, checks the conformity between the actual cell stream and the nominal cell stream (specified by the traffic parameters) and take necessary action when non-conformity is detected. The main purpose is to protect network resources from malicious use as well as unintentional misbehaviour which can affect the QoS of other already established connections. The non-conforming cells will be either discarded immediately or tagged as low-priority cells and be discarded at a congestion point. The discarded cells will increase the cell loss ratio and severely degrade the service quality. It is important to ensure traffic to be in conformance before submitting to the network.

ATM Forum has defined the generic cell rate algorithm (GCRA) to determine whether the cell is conforming with the traffic contract of the connection [7]. The GCRA depends only on two parameters: the increment I and the limit L. The notation 'GCRA(I, L)' means the generic cell rate algorithm with the value of the increment parameter set equal to I and the value of the limit parameter set equal to L. The GCRA can be interpreted as a virtual scheduling algorithm or a continuous state leaky bucket algorithm. The two algorithms are equivalent in the sense that for any sequence of cell arrival times, the two algorithms determine the same cells to be conforming and thus the same cells to be non-conforming. In this study, the leaky bucket algorithm which has gained wide spread acceptance is used.

The continuous-state leaky bucket algorithm can be viewed as a finite-capacity bucket whose real-valued content drains out at a continuous rate of 1 unit of content per time unit and whose content is increased by the increment I for each conforming cell. When a cell arrives and the content of the bucket is less than or equal to the limit value, L, then the cell is conforming. The capacity of the bucket is $L + I$.

A UPC process implements one or more leaky bucket algorithms to police each connection. Each leaky bucket mechanism has two parameters:

- The increment parameter corresponds to the inverse of the compliant rate (fill rate of the bucket).
- The limit parameter corresponds to the number of cells that can burst at a higher rate (size of the bucket).

When more than one traffic parameter is used for a connection, multiple leaky buckets are cascaded, with the highest rate being policed first.

2.2. Modified leaky bucket algorithm and constraints on source bit rate

Peak cell rate (PCR), sustainable cell rate (SCR), and burst tolerance (BT) are declared as the source traffic parameters for a connection. To ensure that the cells will conform to the usage parameters at the physical layer service access point (PHY-SAP) within the equivalent terminal, the source needs to offer traffic conforming to one GCRA or a set of GCRAs [7]. For the PCR traffic parameter, the source needs to offer traffic that is conforming to GCRA(T, 0), where

$T = 1/\mathrm{PCR}$; for the SCR and BT traffic parameters, the source needs to offer traffic that is conforming to $\mathrm{GCRA}(T_S, \tau_S)$, where $T_S = 1/\mathrm{SCR}$ and $\tau_S = \mathrm{BT}$.

The interval to monitor the cell rate in the ATM standardization is expected to be much shorter than one video frame interval (e.g. 33 or 40 ms). The definition of the cell rate is an instantaneous rate calculated from an interval between two successive cells. However, this definition is obviously too strict for the bit rate of video source. We introduce traffic shaping as a solution for this problem. We assume that the video source segments data into cells and transmits the cells for a frame at evenly spaced intervals over the duration of the frame. This is the so-called uniform transmission mode in the study of ATM traffic management. Under this assumption, the constraints on the cell rate can be expressed in terms of frame size. However, it should be pointed out that this kind of traffic shaping would introduce additional delay due to buffering, in this case one frame interval delay is introduced. More sophisticated way to suppress burstiness of a bit rate remains for possible further study.

We propose a modified leaky bucket algorithm (MLBA) to define the constraints on the cell rate of a source. The MLBA is actually a more intuitive interpretation of the leaky bucket algorithm. The MLBA depends on two parameters: the leaky rate R_C and the limit L_C. The $\mathrm{MLBA}(R_C, L_C)$ can be interpreted as a finite-capacity bucket whose real-valued content, counted in cells, drains out at a fixed and continuous rate of R_C cells per time unit. The generated bits of a video encoder are segmented into cells and the cells for a frame are fed into the bucket with even space during the frame interval before they are transmitted to the network. For each incoming cell, the content of the bucket is increased by 1. If at any cell arrival the content of the bucket is less than or equal to the limit value, L_C, then we consider the MLBA (R_C, L_C) as being satisfied. The capacity of the bucket is $L_C + 1$ cells. Figure 1 illustrates the MLBA model.

According to the explanation of the GCRA, the $\mathrm{GCRA}(I, L)$ will be satisfied if the $\mathrm{MLBA}(R_C, L_C)$ is satisfied when $R_C = 1/I$ and $L_C = L/I$.

In order to compute the constraints on the output bit rate of the encoder imposed by the UPC function, we define r_n as the cell rate in frame n. To simplify the notation, we define r_n as the number of cells generated and transmitted in frame n. Similarly, we also define PCR and SCR in the number of cells per frame interval, and BT in the number of frame intervals.

With respect to the peak cell rate of the traffic, $\mathrm{GCRA}(T, 0)$ need to be satisfied, where $T = 1/\mathrm{PCR}$. Interpreted as $\mathrm{MLBA}(\mathrm{PCR}, 0)$, we consider a one-cell capacity leaky bucket whose

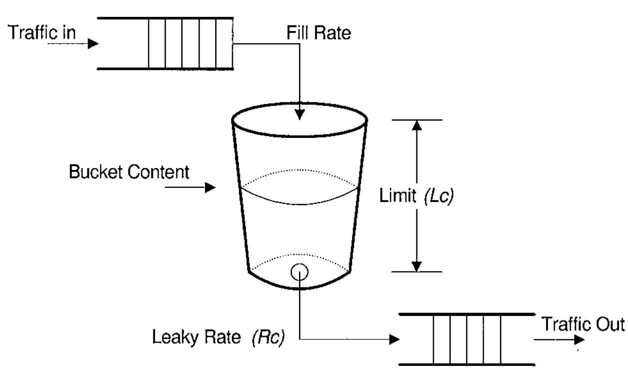


Figure 1. Modified leaky bucket algorithm model.

content drains out at a continous rate of PCR cells per frame interval. In order to avoid the overflow, the following constraints have to be put on the cell rate of the input traffic, namely the output cell rate of the video encoder.

$$r_n \leqslant \mathrm{PCR} \tag{1}$$

Similarly, with respect to the sustainable cell rate and the burst tolerance of the traffic, GCRA(T_S, τ_S) need to be satisfied, where $T_S = 1/\mathrm{SCR}$ and $\tau_S = \mathrm{BT}$. Interpreted as MLBA(SCR, SCR $*$ BT), we consider a finite-capacity leaky bucket whose content drains out at a continous rate of SCR cells per frame interval and its capacity is (SCR $*$ BT $+ 1$) cells. In order to avoid the overflow, the following constraints have to be put on the input cell rate of the bucket, namely the output cell rate of the video encoder.

$$L_n + r_n - \mathrm{SCR} \leqslant \mathrm{SCR} * \mathrm{BT}$$

which equals

$$r_n \leqslant \mathrm{SCR} * \mathrm{BT} - L_n + \mathrm{SCR} \tag{2}$$

where L_n denote the bucket content at the beginning of the nth frame interval. L_n is never negative and can be calculated as follows:

$$L_1 = 0$$

$$\vdots$$

$$L_n = \max\{0, L_{n-1} + r_{n-1} - \mathrm{SCR}\}$$

Under the assumption of traffic shaping, if the video encoder generates frames satisfying expressions (1) and (2), then the video traffic will conform to the declared PCR, SCR and BT traffic parameters at the PHY-SAP of the equivalent-terminal. Although the traffic shaping will introduce additional delay to the traffic, the modified leaky bucket algorithm itself does not introduce any delay. The bucket is a virtual one and the only purpose of the algorithm is to calculate the constraints on the output rate of the encoder.

3. RATE CONTROL STRATEGY

3.1. Rate control strategy for real-time generated video

In this section a rate control strategy is proposed to force the output rate of the MPEG-2 encoder to conform to the UPC-based cell rate constraints. The idea behind the rate control algorithm is to ensure that the VBR traffic generated conform to the pre-defined traffic parameters PCR, SCR and BT, thus no cells will be discarded at UPC. The expected scenario is that the encoder basically works in open loop mode using the fixed quantization scale while at the same time monitoring the output rate using the proposed modified leaky bucket algorithm. In case the cells generated in a frame, r_n, are detected as exceeding the constraints described by expressions (1) and (2), the encoder switches to controlled coding mode temporarily. In the controlled coding mode, the encoder will slow down the bit rate to meet the constraints by adjusting the quantization parameters. The variation of quantization parameters is said to be the reason resulting in the inconsistent picture quality. Thus, in order to maintain the picture quality as consistently as possible, the rate control strategy aims to change the quantization parameters at frame level, namely, the quantization parameters will keep fixed in each frame even under the controlled coding mode.

3.1.1. Relationship between quantization parameter and output bit rate. It is well known that the adjustment of the quantization scale is a major and efficient way to implement rate control in video coding. In MPEG-2, the quantization factors range from 1 to 112. Small value of quantization factor means fine quantization and big value of quantization factor means coarse quantization. Figure 2 shows the effect of the different quantization factor on picture quality. The generated bit rate decreases with an increase to the quantization parameter. It would be good if the bit rate could be precisely predicted and the proper quantization parameter could be derived from the desired bit rate. However, the relationship between the quantization parameter and corresponding bit rate varies in time and depends on instantaneous complexity and activity of the scenes. Although a lot of research has been done on this topic, so far there is no algorithm to accurately predict the bit rate and describe the quantization parameter as a function of desired bit rate at frame level.

We study the relationship between the quantization parameter (Q) and the frame size (S). (In MPEG, the interframe interval is a fixed value, thus the frame size can be considered as the bit rate of the frame.) Figure 3 plots the reciprocal of frame size ($1/S$) as a function of Q ranging from 1 to 64 for 12 sample I pictures, 12 sample P pictures and 12 sample B pictures, respectively. The sample pictures are randomly chosen from six different sequences with different scenarios, different picture complexities and activities, and different motion types and degrees. Each curve in the figure corresponds to one picture.

From both Figures 2 and 3, we can conclude that a quantization factor of between 0 and 40 would change the frame size and yet maintain a reasonable picture quality. Any quantization factor greater than 40 has a small effect on the change in frame size and the picture quality at such high quantization factor would be blocky and quite unacceptable.

From Figure 3, it can be shown that the most effective portion of the graph (quantisation factor between 0 and 40) show reasonable linear correlation. We approximate them by linear functions and experimentally express them as

$$Q_i = k_i \frac{1}{S_i} \tag{3}$$

where k_i is a constant that depends only on the scene complexity and activity. k_i can be derived by

$$k_i = Q_{i0} S_{i0}$$

where $(Q_{i0},\ 1/S_{i0})$ is a point in line i.

Obviously, the simple linear function (3) may introduce big error for those curves which are less linear. In this case, when two points in a curve are available, the curves can be further approximated by piecewise linear function

$$Q_i = \frac{1}{S_i}\frac{Q_{i0} - Q_{i1}}{1/S_{i0} - 1/S_{i1}} + \frac{S_{i0}Q_{i0} - S_{i1}Q_{i1}}{S_{i0} - S_{i1}} \tag{4}$$

(a) Original, Size=829440 bytes (b) QF= 4, Size=140463 bytes

(c) QF = 24, Size=28501 bytes (d) QF = 48, Size=20443 bytes

Figure 2. Effect of quantization factor on picture quality.

(e) QF = 64, Size=18665 bytes (f) QF = 112, Size=16609 bytes

Figure 2. *Continued.*

where $(Q_{i0}, 1/S_{i0})$ and $(Q_{i1}, 1/S_{i1})$ are two different points in curve i. The applicability of expressions (3) and (4) depends on the availability of the (Q_i, S_i) pairs.

3.1.2. Time consumption in MPEG coding. Another observation is that in an MPEG system, the whole encoding process can be divided into several stages and executed dependently following certain order. Among the stages, the quantization (Q) stage is the only lossy stage and it determines the final output bit rate. After the Q stage, there is only one stage, variable length coding (VLC) stage, which must be executed before the final accurate output rate can be obtained.

We study the time consumption of the different stages. Table I shows the approximate CPU time consumed by each stage when running a software MPEG-2 VBR encoder on a Sun Sparc 20.

The data indicate that the total time for encoding an I frame is much less than that for encoding a P or B frame because the encoding of I frames does not involve the most time-consuming stage, the motion-estimation (ME) stage. And the time consumed by Q and VLC stage is only a small portion of the total time spent in the whole process of encoding a frame. For the least time consuming I frames, the Q and VLC stage consumes about $\frac{1}{4}$ of the total time. And for P and B frames, it consumes only about $\frac{1}{17}$ and $\frac{1}{14}$ of the total time, respectively. Thus, it is feasible to repeat the Q and VLC stage even in real-time applications. Although one repeat of the Q and VLC stage may introduce $\frac{1}{4}$ of the encoding time for a I frame, it may not affect the encoding of an I frame in real time since the encoding of I frames is the less time-consuming process. Thus, repeating the Q and and VLC stages in the I frames is sensible and possible for real-time application. For the most time consuming P and B frames, our experiments show that both the average and the maximum time consumption for encoding a frame under re-quantization rate control with at most five iterations are similar or even less than those for encoding a frame under the normal adaptive CBR rate control strategy, because the fixed quantization used in re-quantization strategy speeds up the motion estimation stage. Thus, this strategy is applicable for real-time video applications.

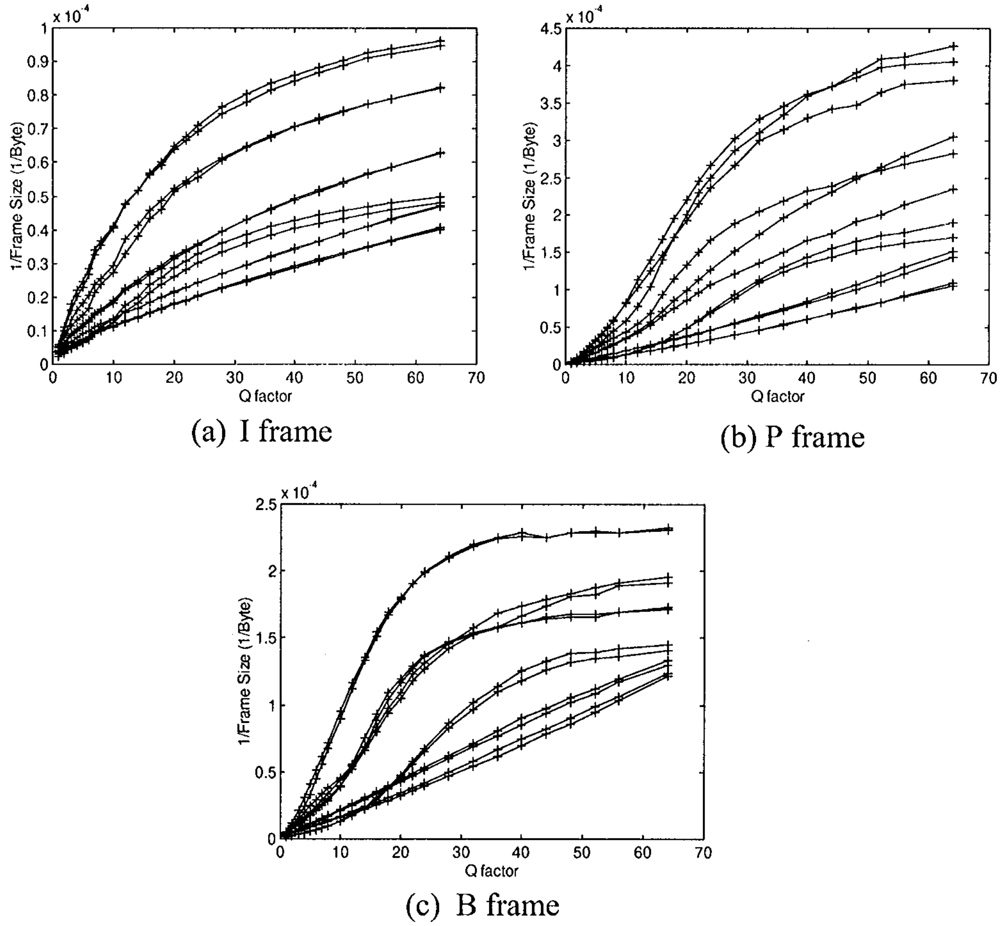


Figure 3. The relationship between quantization factor and frame size.

Table I. Time consumption in MPEG coding (Unit: minisecond).

Picture type	ME	DCT	Q and VLC	IQ and IDCT	Total
I	80	2130	863	529	3808
P	11 436	2146	866	358	15 000
B	8161	2197	839	324	11 725

3.1.3. Re-quantization rate control strategy. Based on the above two studies, namely the relationship between quantization parameter and frame size and the time consumption of different stages in MPEG-2 video coding, we propose the following re-quantization rate control strategy to ensure the lossless transmission of the video traffic.

Figure 4 depicts the structure of the developed MPEG-2 VBR encoder using the proposed UPC-based re-quantization rate control strategy.

Suppose that the encoder is generating VBR traffic which need to conform to the declared usage parameters PCR, SCR and BT, and the specified quantization parameters are Q_I, Q_P, and Q_B for I, P, B frames, respectively. Under the proposed rate control scheme, the modified leaky bucket algorithm will be implemented after the VLC stage inside the encoder to monitor the bit-rate generated. The encoder will initially work in open-loop mode. Scenes with reasonable activity and duration will pass the leaky bucket safely. When criteria (1) or (2) is violated, it indicates that the frame currently being encoded has exceeded the constraints on the bit rate. In this case, the bits generated for this frame will be discarded and the Q and VLC stage will be repeated with a new quantization parameter. The new quantization parameter is derived from expression (3) for the first time or from expression (4) when the Q and VLC stage has been repeated more than once. The quantization parameter used and bits generated in the previous process will be used as available (Q_i, S_i) pairs. Once the quantization parameter for a frame has been changed, then the quantization parameters for subsequent frames within the same group of picture (GoP) will also be changed proportionally. This principle takes into account the continuity of the scene and the high correlation of the consecutive pictures. High complexity and activity scene seldom last for only one picture and usually result in a series of oversized pictures. In addition, the quantization scale for P and B frames, especially the quantization scale for B frames which are never used for references for other frames, is not necessary to be finer than that for I frames.

With this strategy, the encoder uses the open-loop mode during scenes with reasonable complexity and motion, this is the expected mode of operation for the main duration of a sequence. While during the overload periods (where r_n exceeds the constraints of the traffic contract), the encoder switches to the controlled mode using the proposed rate control scheme as described above. Only pictures which will cause cell tagging by the UPC function are shaped. The picture quality will degrade slightly during these controlled periods due to the coarser quantization scale.

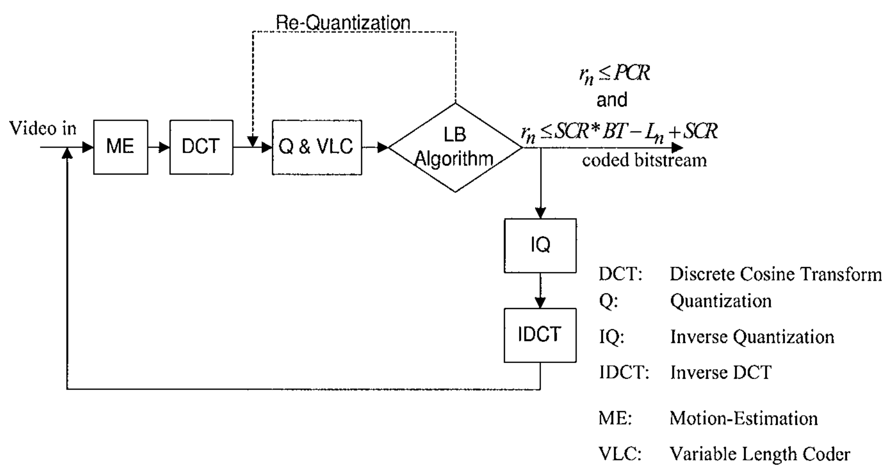


Figure 4. UPC-based MPEG encoder structure.

3.2. Rate control strategy for pre-compressed video

The pre-compressed video has less flexibility than real-time generated video in source rate control. A simple method is to selectively discard frames to temporarily lower down the bit rate. This kind of rate control is effective in maintaining SCR, but it does not contribute to the PCR control. Peak rate usually occurs at I frames, and the drop of I frames will cause significant and continuous degradation of picture quality since the consequent frames in the same GoP lose its reference. Thus, in this section, a new method is studied to perform PCR control for MPEG-2 video streams. The new method reduces the bit rate of I frame by removing a certain number of AC coefficients, thus denoted as AC-removal rate control algorithm.

DCT is used by MPEG-2 to reduce the spatial redundancy within a frame or motion compensation errors. A two-dimensional DCT is performed on every 8×8 block of the frame or motion compensation errors to produce blocks of DCT coefficients. The magnitude of each DCT coefficient indicates the contribution of a particular combination of horizontal and vertical spatial frequencies to the original frame block. The coefficient corresponding to zero horizontal and vertical frequency is called the DC coefficient, others are called AC coefficients. The transformation tends to concentrate the energy into the low-frequency coefficients, and many of the other coefficients are near zero. The higher the frequency of the AC coefficients are, the less they contribute to the original frame because human eye is less sensitive to the AC coefficients with high frequency.

Figure 5 shows the effect of the number of AC coefficients on picture quality. The pictures (zoom in 260×334 resolution of a 720×576 picture) are all I frames but with different number of AC coefficients.

It is observed that even when there are only 8 or 16 AC coefficients, the picture quality is quite acceptable, while the frame size could be reduced by one-half or two-thirds of the original frame size with the full 64 AC coefficients. Thus, it is assumed that removing a certain number of less important AC coefficients is an applicable method to maintain the peak cell rate (coming from I frames) of the traffic within certain constrains.

In MPEG coding, after the DCT, quantization, zig-zag scan, and run length coding, the information of all the DCT coefficients within a block are placed in a one-dimensional array, ordered according to perceptual significance of each coefficients. The least important coefficients should be removed first. In this study, the removal of n number of AC coefficients means the removal of the last n number of the AC coefficients. In addition, to simplify the computation as well as to maintain consistent picture quality, to remove a certain number of AC coefficients of a frame means to remove the same number of AC coefficients of all the blocks within the frame.

A program was written to parse the MPEG-2 video streams according to the MPEG-2 syntax. It implements the AC-removal rate control strategy for PCR control in such a way that it checks the size of every frame in the whole sequence and applies the AC-removal algorithm to those I frames whose size violate expression (1). The first scan of the algorithm calculates all the corresponding frame sizes under different situations: removing last 1 AC coefficient, last 2 AC coefficients, …, till all the 64 AC coefficients. Then the biggest value of frame sizes satisfying expression (1) and its corresponding number of AC coefficients removed are selected. The second scan removes the selected number of AC coefficients block by block from the frame.

The AC-removal algorithm is less efficient for P and B frames since DCT is only used in the motion compensation errors in the P and B frames, so the quantity of AC coefficients which could be removed from P or B frames is quite limited. In addition, the motion compensation error information are usually at a higher frequency and removing AC coefficients here is not a suitable method. Therefore, for the rate control of P and B frames, the selectively frame dropping method is used. The size of every frame will be monitored by the MLBA. If the size of a frame is found to violate the expression (2), one or more frames will be selectively dropped according to the following method. If the frame is a B frame, then it is discarded. If the frame is a P frame, then the

Figure 5. Effect of number of AC coefficients on picture quality.

(e) AC=4, Size=38831 (f) AC=2, Size=29242

(g) AC=1, Size=22788 (h) AC=0, Size=15102

Figure 5. *Continued.*

modified leaky bucket algorithm goes back one step and discards the most recent B frame. Only in cases when the most recent frame is also a P or I frame will the current P frame be discarded. Similarly, if the frame is an I frame, the MLBA goes back to discard the most recent B frames first. Only when all the B frames in the previous GoP have been discarded and there is still not enough room for the coming I frame, the most recent P frames will be discarded. I frames should be the last one to be dropped and in normal case, it is not possible for I frames to be discarded unless the

PCR, SCR and BT parameters are set unreasonable. This method takes into account the significance of the different types of frames. The backward method is reasonable since it is for the pre-compressed video and the MLBA can be applied before the real transmission.

4. EXPERIMENTAL RESULTS

In this section, we show some experimental results of the UPC-based rate control for real time generated video transmission and UPC-based rate control for pre-compressed video transmission. The results are compared with pure VBR transmission in terms of network performance and video quality.

4.1. Source characteristics

Experiments were carried out using 1580 images captured from the movie 'Rain Man' in 720×576 format. The scenes consist of zooming in, zooming out, and panning. There are eight scene changes in the sequence, respectively, in frame 154, 220, 641, 708, 1110, 1147, 1282 and 1498. Figure 6 presents the frame size trace of the test sequence coded using open-loop mode with GoP structure as IBBPBBPBBPBB and quantization parameters set to 10 for I pictures, 12 for P pictures and 14 for B pictures.

Owing to the combination of intra- and inter-frame coding technologies, the traffic is rather bursty and periodic peaks caused by the intra-frames can be observed in the bit rate. In addition, due to the open-loop coding scheme, the bit-rate generated varies in time and depends on the instantaneous picture complexity and motion. Higher bit rate can be observed for scenes from frames 641 to 707 and 1282 to 1497. As expected, significant variations in bit rate often occur at scene changes.

4.2. Lossless transmission with UPC-based rate control

The objective of UPC-based control at the video source is to avoid the effects of ATM network UPC policing on VBR quality. Acting at the source, the rate control provides a mechanism for

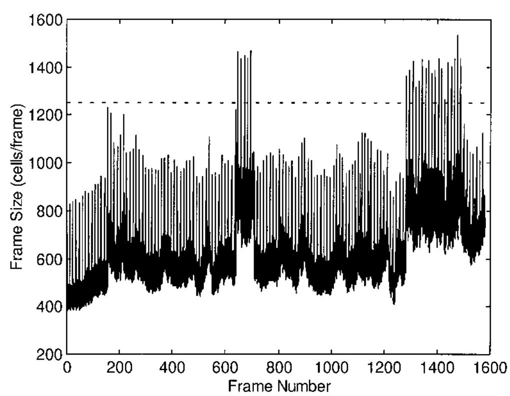


Figure 6. Cells per frame time series (pure VBR).

staying within agreed-upon ATM network VBR bit-rate parameters without severe degradation in image quality due to cell loss. In practice, the source and the network usually exchange a fixed set of easily identifiable parameters for the purpose of the UPC at connection setup. In this section, the results of the source rate control under certain predefined usage parameters are shown.

4.2.1. Network performance. In this study, PCR control is used as the control to satisfy only expression (1) while the SCR control is required to satisfy both expressions (1) and (2). We denote the sequence generated using open-loop coding as sequence pure-VBR. The sequence generated under real-time re-quantization rate control strategy and with only PCR control is denoted as sequence rt-PCR, and the sequence under same rate control strategy but with SCR control as sequence rt-SCR1, rt-SCR2, etc. Similarly, the sequence generated under the AC-removal PCR control strategy is denoted as sequence pre-PCR and the sequences generated under both AC-removal PCR control and selectively frame dropping method are denoted as pre-SCR1, pre-SCR2, etc. Table II summarizes the predefined traffic parameters of the sequences.

Figure 7 shows the frame size series for the test sequences generated under the UPC-based rate control strategies with various predefined traffic parameters.

It can be clearly observed that the proposed PCR rate control method can maintain the resulting traffic peak rate within the agreed-upon traffic contract. The SCR rate control method is also proved to be effective in keeping traffic satisfying certain SCR and BT parameters. Lossless transmission can be guaranteed from the source perspective.

According to the experiment, for the re-quantization rate control strategy, the PCR control is activated more often than the SCR control. This is because PCR control not only works on the enforcement of peak cell rate but also contributes a lot to the maintenance of SCR due to the principle that the change of quantization factor in one frame will propagate to the rest of the frames in the same GoP. In the SCR control of the pre-compressed video, all the frames dropped in the experiments are B frames. During the two bursty periods, one period from frame 641 to 707 and another one from frame 1282 to 1487, totally including 283 frames, the total number of frames dropped is 85 for pre-SCR2 sequence and 64 for pre-SCR3 sequence. The worst scenario is to drop one B frame from every three frames (IBB or PBB).

Table II. Pre-defined traffic parameters of the test sequences.

		Pre-defined traffic parameters		
Rate control strategy	Sequence	PCR (cells/frame interval)	SCR (cells/frame interval)	BT (frame intervals)
	pure-VBR	—	—	—
Re-quantization	rt-PCR	1250	—	—
	rt-SCR1	1250	697	3
	rt-SCR2	1250	625	3
	rt-SCR3	1250	625	15
Ac-removal plus frame dropping	pre-PCR	1250	—	—
	pre-SCR1	1250	800	3
	pre-SCR2	1250	625	3
	pre-SCR3	1250	625	15

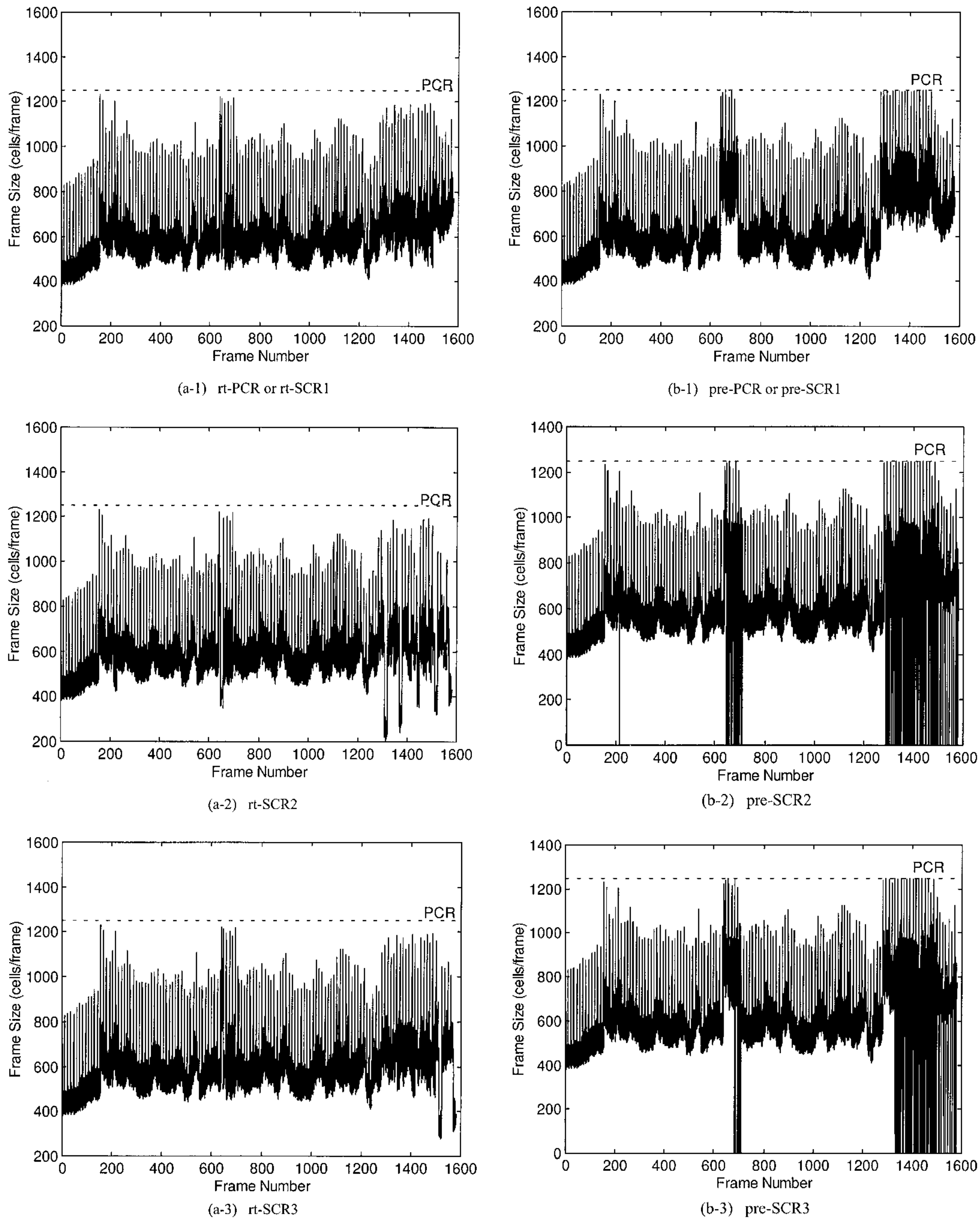


Figure 7. Cells per frame time series (controlled VBR).

Table III. Traffic characteristics of the test sequences (ACR: average cell rate).

	Characterization of generated traffic			Cell loss ratio (%)		
Sequence	PCR	ACR	PCR/ACR	PCR = 1250	PCR = 1250 SCR = 625 BT = 3	PCR = 1250 SCR = 625 BT = 15
pure-VBR	1536	632	2.43	0.37	6.29	4.76
rt-PCR	1230	602	2.04	0	1.56	0.69
rt-SCR1	1230	602	2.04	0	1.56	0.69
rt-SCR2	1230	586	2.10	0	0	0
rt-SCR3	1230	596	2.06	0	1.05	0
pre-PCR	1250	630	1.98	0	5.9	4.4
pre-SCR1	1250	630	1.98	0	5.9	4.4
pre-SCR2	1250	592	2.11	0	0	0
pre-SCR3	1250	601	2.08	0	1.6	0

Table III summarizes the traffic parameters of the generated sequences.

The result shows that both the rate control methods slightly reduce the average cell rate while significantly reducing the peak cell rate of the traffic. Therefore the ratio of PCR to average cell rate is also reduced. From the network perspective, smaller value of the ratio of PCR to average cell rate is preferable since this will lead to less bursty traffic in the network.

The most significant advantage of the UPC-based rate control coding lies on the elimination of the lost cells at UPC. It can be observed that with the UPC-based rate control, the cell loss ratio can be reduced to 0 from the source perspective. The lost cells in the video streams will cause the failure in the reconstruction of the sequence. The lost cells in the I frames will not only affect the reconstruction of the I frame itself but also destroy the reconstruction of all the frames in the same GoP because the I frame is used as a reference frame for all the other frames in the GoP. Considering the propagation of the error in I frames, the cell loss ratio of the pure VBR is even higher.

4.2.2. Quality of picture. Obviously, all the benefits from the UPC-based rate control schemes come at the cost of the instantaneous degradation of the picture quality due to the coarser quantization scale or the dropped AC coefficients. Figure 8 compares the SNR of the various sequences. Figure 8(a) plots the SNR of all the I frames in the pure-VBR, rt-PCR, and rt-SCR2 sequences to show the degradation of picture quality caused by coarser quantization scale. Figure 8(b) plots the SNR of all the I frames in the pure-VBR and pre-PCR sequences to show the degradation of picture quality caused by the removal of AC coefficients. It can be observed that the values of SNR drop during the controlled periods. By comparing the two figures, it can also be noticed that in order to achieve the same bit rate, using coarser quantization can maintain better picture quality than removing certain number of AC coefficients. In real-time generated sequences, the degradation of picture quality caused by SCR control is more severe than that caused by PCR control. While for the pre-compressed video, since the SCR control is conducted by dropping an entire frame, the degradation of picture quality cannot be evaluated by SNR, thus it is not shown in the figure. According to the experiments, since the worst scenario for the frame

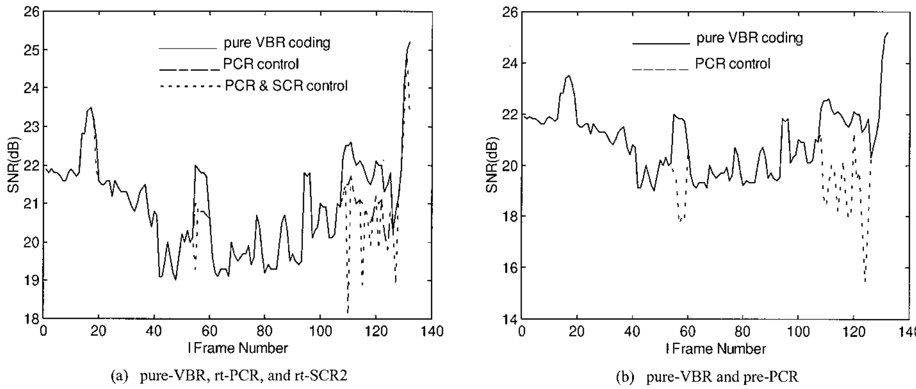


Figure 8. Comparison of signal to noise ratio.

dropping scheme is to drop one B frame out of every three frames (IBB or PBB), it is simply assumed here that the quality drop is not so serious because human eyes are not so sensitive to 30 frames per second or 20 frames per second.

Figure 9 illustrates the degree of degradation of the picture quality by comparing the worst case rt-SCR2 picture with its original pure-VBR picture. Figure 9(a) shows the reduced picture which causes the most severe quality drop in the experiment. Figure 9(b) shows an enlarged portion of the picture captured from sequence pure-VBR while Figure 9(c) shows the same portion of the picture captured from sequence rt-SCR2. It is observed that the worst case picture quality is still quite good. For the pre-compressed video, the worst-case situation in this experiment is keeping only 19 AC coefficients in the I frame. As illustrated in Figure 5, the picture quality of I frames with 19 AC coefficients is also quite acceptable. Thus, compared with the severe quality degradation caused by cell loss under UPC, it is preferable to conduct the source rate control with the proposed rate control strategies.

Here it should be pointed out that the cell tag rate is regarded as the worst-case cell loss rate for the UPC function. Although cells tagged by the UPC will be transmitted, they may be discarded at a later stage at any congested node in the network. Since there is no guarantee of tagged cells being delivered to the user, these cells, as a worst-case scenario, can be treated as dropped cells. So in this study, all the non-conforming cells are considered as discarded cells by the UPC function.

4.3. Leaky bucket parameters

Ideally, if the source submits traffic that conforms to the usage parameters (e.g. PCR, SCR and BT) which have been negotiated with the UPC function, no losses will occur at UNI by UPC for traffic generated by that source. For the LB mechanism, lossless transmission entails choosing the combination of values of the bucket capacity and leaky rate. The required value of PCR for lossless transmission is simply decided by the peak rate of the source. The required values of SCR and BT which determine the values of leaky rate and bucket capacity are correlated and depicted in Figure 10.

The SCR is an upper bound on the possible conforming average cell rate of an ATM connection, where the average cell rate is the number of cells transmitted divided by the duration of the connection. The average cell rate is the least value of the SCR. Figure 10 plots the required

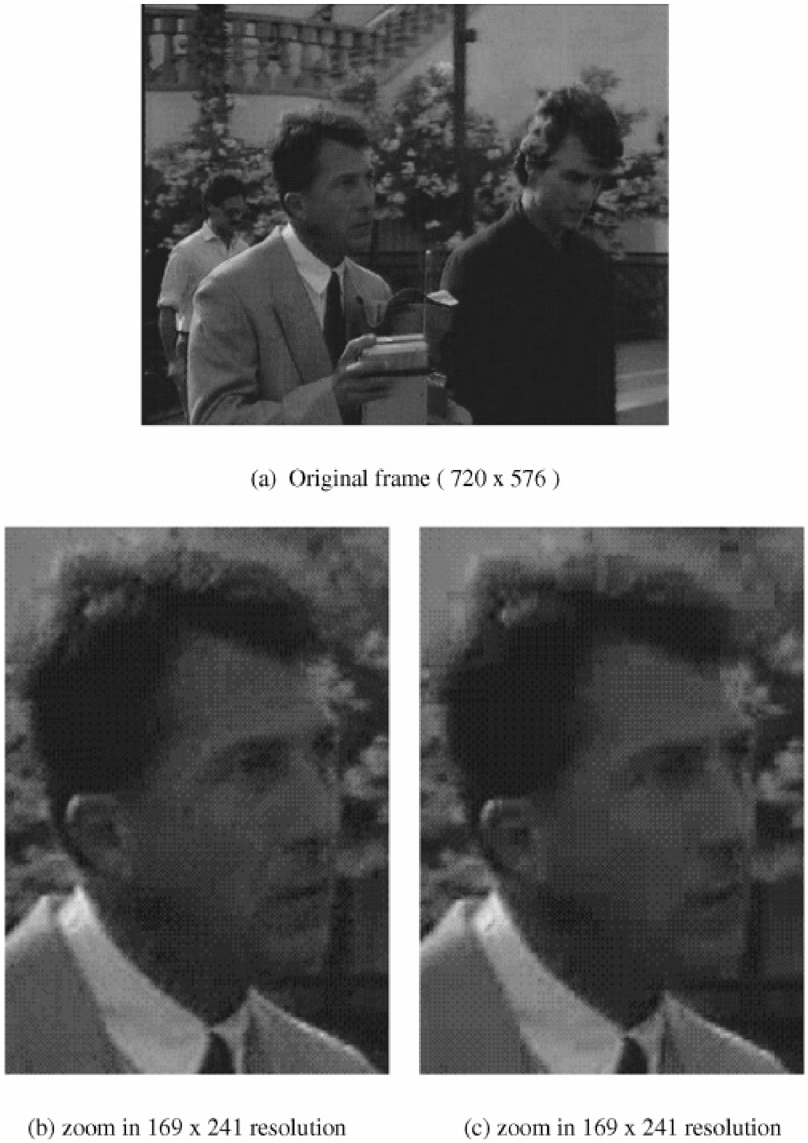


Figure 9. Illustration of worst case picture quality.

BT for lossless transmission as a function of SCR ranging from average cell rate to PCR for the four test sequences with different rate control options. To make the values of cell rate independent of the magnitude of the source traffic, the values of SCR here are presented by the ratio of SCR to average cell rate. Generally speaking, the curves indicate that lower SCR which is equal or close to the average cell rate results in very high BT. Higher BT usually means more bursty traffic and longer delay of the traffic. For delay-sensitive video traffic, information with large delay is useless. On the contrary, lower BT needs higher ratio of SCR to average cell rate, while higher ratio of

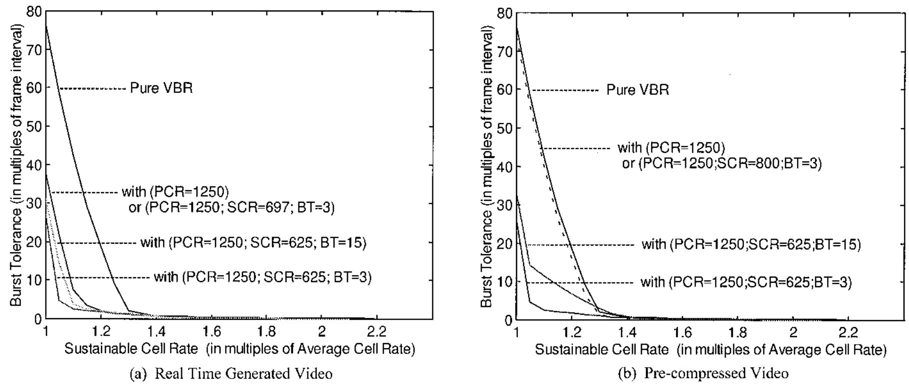


Figure 10. Usage parameters for lossless transmission.

SCR to average cell rate implies less network bandwidth utilization. From the network point of view, curve with both less ratio of SCR to average cell rate and less BT is preferable due to higher network utilization and being less bursty. Compared with the four curves in each figure, all the three curves representing sequences with the UPC-based rate control have better traffic shape than the pure VBR stream.

5. CONCLUSION

In this paper the transmission of MPEG-2 VBR video over ATM network under usage parameter control is studied. The idea is to seek a compromise between the network utilization and the quality of video service by applying UPC-based rate control strategies to the video source. A modified leaky bucket algorithm is proposed to calculate the constraints on the bit-rate guaranteeing conformance to PCR, SCR and BT usage parameters. A new rate control strategy, named re-quantization rate control strategy, is proposed to ensure the real-time encoder output compliance to constraints on the bit rate. Another new rate control strategy, consisting of the AC-removal PCR rate control strategy and selectively frame dropping method, is proposed for pre-compressed video transmission to ensure the traffic submitted to the network compliance to the constraints on the bit rate. Both the UPC-based rate control strategies are implemented in software for MPEG-2 VBR video. The experimental results show that both the schemes can provide lossless transmission from the source perspective. The control on the peak cell rate reduces the burstiness of the traffic. The VBR picture quality keeps unchanged for scenes with reasonable complexity and activity. Only when the bit-rate generated at open-loop mode violates the traffic contract will the rate control apply and the picture quality degrade. The results also show that both the rate control strategies are effective in lowering down the bit rate while keeping the necessary picture quality. To reduce the bit rate to the same level, the rate control method by using coarser quantization can maintain better picture quality than that by removing a number of AC coefficients. On the whole, taking into consideration the network resources saved, the degradation of picture quality caused by the source rate control is preferable than the severe drop of picture quality caused by the cell loss at UPC.

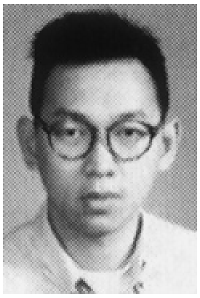
In this paper, the experiments are performed with a set of reasonably pre-defined UPC parameters. For the re-quantization rate control strategy as well as the AC-removal rate control strategy, the quantity of bit rate that can be reduced is limited. If the pre-defined UPC parameters are too low compared with the real traffic, it may go beyond the capability of the rate control strategies. For pre-compressed video, the selection of a set of reasonable pre-defined UPC parameters is quite easy [2]. For real-time generated video, an experimental prediction may be needed. Even though, the UPC parameters defined at the connection setup may be found unreasonable, either too high or too low, sometimes during the entire connection time. In this case, we propose that the video source re-negotiates its traffic parameters with the network. ATM networks provide re-negotiations of the traffic specifications and QoS via ATM signaling protocol during the connection time. Thus, if the current UPC parameters are found to be unreasonable, a new set of reasonable UPC parameters can be predicted and exchanged with the network. This method will guarantee the proposed rate control strategies work in the normal situation. The specific way for the re-negotiation is left for further study.

REFERENCES

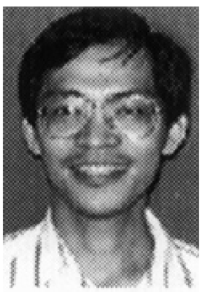
1. Heyman DP, Tabatabai A, Lakshman TV. Statistical analysis and simulation study of video teleconference traffic in ATM networks. *IEEE Transactions on Circuits and Systems for Video Technology* 1992; **2**(1).
2. Wang B, Gupta AK. Traffic characterization of VBR video streams using on–off transmission in ATM networks. *Proceedings of Performance and Control of Network Systems, part of SPIE's Voice, Video, and Data Communications*, Dallas, November 1997.
3. Garrett MW, Willinger W. Analysis, modeling and generation of self-similar VBR video traffic, *Proceedings of SigComm*, ACM Press: New York, September 1994.
4. Hamdi M, Roberts JW, Rolin P. GOP-scale rate shaping for MPEG transmission in the B-ISDN. *International Symposium on Multimedia and Video Coding*, NY, October 1995.
5. Harasaki H, Yano M. A study on VBR control under usage parameter control. *Proceedings of Fifth International Workshop on Packet Video*, *F2.1*, Berlin, Germany, June 1993.
6. Kawashima M, Tominaga H. A study on VBR video transmission under the usage parameter control. *Proceedings of Fifth International Workshop on Packet Video*, *F3.1*, Berlin, Germany, June 1993.
7. ATM Forum: ATM User-Network Interface Specification (Version 3.1). September 1994.

AUTHORS' BIOGRAPHIES

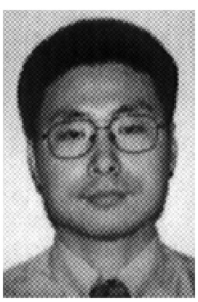
Wenjing Lou received the BE degree and ME degree in Computer Science and Engineering from Xi'an Jiaotong University, China, in 1993 and 1996, respectively. She received the MASc degree in Computer Communications from Nanyang Technological University, Singapore, in 1998. From December 1997 to July 1999, she was working as a Research Engineer in Network Technology Research Center, Nanyang Technological University. She was involved in the projects of Singapore Advanced Research and Education Network (SingAREN) and Asia Pacific Advanced Network (APAN). She is currently pursuing the PhD degree in Computer Engineering at New Jersey Institute of Technology. Her research interests are in the areas of TCP/IP network and ATM network, including the network performance evaluation, multimedia transmission and network security.



Dr Chia Liang Tien (Clement) is a Lecturer in the Division of Computing Systems, School of Applied Science at Nanyang Technological University. He gained a first degree in Electrical Engineering and a PhD from Loughborough University of Technology in 1990 and 1994, respectively. His current research interests are in video communication, bandwidth management and error concealment techniques.



Dr Lee Bu Sung (Francis) graduated with BSc (Honours) and PhD from Loughborough University, in 1982 and 1986. He is presently working as a Senior Lecturer in the School of Applied Science, Nanyang Technological University. He also holds the position of Deputy Director of the Network Technology Research Center in NTU. Current research areas are ATM traffic management, multimedia networks, reliable multicast over ATM, distributed interactive simulation.



Bin Wang received the BE degree and ME degree in Computer Science and Engineering from Xi'an Jiaotong University, China, in 1993 and 1996, respectively. He received the MASc degree in Computer Communications from Nanyang Technological University, Singapore, in 1998. From January 1998 to April 1999, he was an R & D Engineer in R & D Center, JVC Asia Pte Ltd. He was working on data broadcasting of digital TV. Currently he is a Software Engineer in Emerging Technology Center, IBM Singapore Pte Ltd. His research interests include network applications, bandwidth management and multimedia trasmission.

