

Quality Control for VBR Video over ATM Networks

Wenjun Luo and Magda El Zarki, *Member, IEEE*

Abstract—Uncontrolled variable-bit-rate (VBR) coded video yields consistent picture quality, but the traffic stream is very bursty. When sent over ATM networks, cell losses may be incurred due to limited buffer capacity at the switches; this could cause severe picture quality degradation. Source rate control can be implemented to generate a controlled VBR bit stream which conforms to specified bit rate bounds and buffer constraints. However, source rate control could result in picture quality degradation too. Hence, for real-time video services, an important issue to address is whether the picture quality degradation incurred by source rate control is within acceptable levels or how to choose the appropriate coding parameters to make it so. In this paper, we establish quantitatively the relationship between picture quality and source rate control for the case of guaranteed service with different combinations of allocated bandwidth, buffer size, and other key video-coding parameters of MPEG-2. In addition, quality control in the context of two-layered scalable video service (basic and enhanced quality) is also considered. Our study reveals that, in order to maximize both the basic and the enhanced quality, source rate control should be implemented on both layers. The relationships between the two types of quality and different combinations of allocated bandwidths, buffer sizes, and some key coding parameters are also established quantitatively for MPEG-2 SNR scalability.

Index Terms—MPEG, packet video, rate control, scalability.

I. INTRODUCTION

NETWORKS based on asynchronous transfer mode (ATM) [1] are able to support variable-bit-rate (VBR) sources. However, because of limited buffer capacities at the switches, very bursty traffic sources may cause congestion, and hence buffer overflow, when multiple connections transmit at peak rate simultaneously. In order to avoid such congestion, traffic control is implemented in ATM networks [1], [2].

Traffic control in ATM networks mainly incorporates two functions: connection admission control (CAC) and usage parameter control (UPC). CAC is implemented during the call setup procedure to ensure that the admission of a call will not jeopardize the quality of service (QoS) of the existing connections. If the connection is admitted, a service contract is established specifying the amount of resources reserved according to the source traffic descriptor and the required QoS. The service contract also specifies the traffic behavior to which the input bit stream should conform in order to achieve the desired QoS. UPC is performed during a connection's lifetime to monitor the input traffic. Its main purpose is to protect network resources from malicious as well as unintentional

misbehavior which can affect the QoS of other connections by detecting violations of negotiated parameter values.

In this study, we assume that all the cells conforming to the UPC parameters are guaranteed delivery, i.e., guaranteed service is used. Nonconforming cells are tagged, and can be sent as best effort traffic. But the network provides no guarantees for their delivery as no network resources are reserved for these cells. They are discarded if the network is congested. Therefore, our analysis in this paper assumes the worst case scenario—all tagged cells are not available at the receiver (i.e., they are all lost).

Uncontrolled variable bit-rate (VBR) coded video bit streams exhibit dramatic long- and short-term bit-rate variations [3], [4]. Moreover, the bit-rate requirement can change dramatically during the transmission due to other factors, e.g., the end user changes the quality requirement (i.e., resolution, frame rate, etc.). A static CAC/UPC cannot efficiently accommodate such variations in bit rate. Studies [5], [6] have shown that a well-designed dynamic resource renegotiation scheme can match the bit-rate variation, and hence maintain a high level of video quality and good bandwidth utilization.

However, renegotiation is not a deterministic process; renegotiation delay may be incurred or the request may be rejected. Between two successful renegotiation events, the UPC parameters are fixed, and the source has to conform to what is available. In this paper, we study the impact of traffic control on VBR video transmission, and identify the strategy that achieves the best quality service given the allocated resource constraints.

We first show that uncontrolled VBR coded video is very bursty. For high-end real-time video services, in order to achieve low cell tagging rates ($<10^{-5}$), either nearly peak-rate bandwidth or impracticably large buffers need to be reserved. If high bandwidth utilization and small buffer size are required and network resources are reserved accordingly, the resultant cell tagging rate will exceed the limit ($<10^{-5}$) under which error concealment techniques can effectively conceal the errors. In other words, the picture quality will still be very poor even if error concealment techniques have been applied [7].

Therefore, certain actions need to be taken at the source to control the burstiness of the VBR video traffic before it is input to the network. There are typically two ways in which one can do this: 1) simple traffic smoothing by buffering a certain number of frames at a time, and then sending them at the average rate of those frames and 2) controlling the coding procedure to generate a controlled VBR bit stream so that the traffic burstiness is limited and fits into specified bandwidth and buffer constraints. This approach is known as source rate control.

Manuscript received May 1, 1996; revised September 30, 1996. This work was supported by the National Science Foundation under Grant NCR94-05601.

The authors are with the Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA.

Publisher Item Identifier S 0733-8716(97)04191-7.

The first approach introduces buffering delay. For real-time interactive video services, the smoothing period needs to be small as the maximum tolerable end-to-end delay can be at most 10–15 frames [8]. A previous study [9] has shown that for uncontrolled VBR video traffic, the first approach does not yield any substantial gain when the smoothing period is limited to be less than 15 frames.

In this paper, we investigate the second approach in more detail. Generally, source rate control will cause picture quality degradation. Hence, the important issue is whether the picture quality degradation incurred by source rate control is within acceptable levels or how to dimension the coding parameters to make it so. Based on the MPEG-2 [10] Test Model 5 (TM5) [11] source rate control scheme, we establish quantitatively the relationship between the video sequence's picture quality and rate control with different combinations of allocated bandwidth, buffer size, and other key coding parameters for MPEG-2 video.

One of MPEG-2 video's major extensions over MPEG-1 is the addition of scalability, its purpose being to support applications beyond that supported by single-layer video. One such application involves the provision of several different qualities of service to accommodate various classes of users. A two-layered source coding scheme generates a base layer which contains the basic video information. This layer can be solely decoded to provide a basic quality of service. The residual information is coded as an enhancement layer so that, when received and added to the base layer, the enhanced video quality can be obtained.

We extend our study in this paper to the quality control issue for the two-layered scalable video service. The base layer is critical for the reconstruction of both qualities, and hence needs to be source rate controlled to avoid substantial cell tagging. We show in this paper that the enhancement layer has a profound impact on both the basic and the enhanced quality, and that when source rate controlled too, produces the best results. The relationships between the two types of quality and different combinations of allocated bandwidths, buffer sizes, and other key coding parameters are also established quantitatively for MPEG-2 SNR scalable video.

This paper is organized as follows. In Section II, we show how VBR video is affected by traffic control in ATM networks. We then analyze the source rate control scheme in detail, and present results of its performance for single-layer MPEG-2 video in Section III. In Section IV, we extend the quality control issue for MPEG-2 SNR scalable video. A summary is given in Section V.

II. IMPACT OF TRAFFIC CONTROL ON UNCONTROLLED VBR MPEG-2 VIDEO

In this section, we show the impact that traffic control has on uncontrolled VBR coded MPEG-2 video. An overview of the ATM traffic control model based on the leaky bucket scheme is briefly described in Section II-A. In Section II-B, we first obtain values for network resources that need to be reserved such that there are no cells tagged. We then demonstrate how the video quality is affected if bandwidth allocation equal to average source rate and small buffer sizes is enforced.

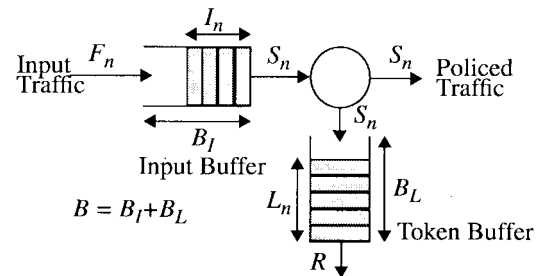


Fig. 1. ATM traffic control diagram based on the leaky-bucket algorithm.

A. ATM Traffic Control Model

1) *UPC Function and Traffic Monitoring*: The UPC function is implemented to control the input traffic according to the service contract. The essential function of the UPC mechanism is to separate conforming cells from nonconforming cells. We chose to implement this function using the leaky-bucket scheme, which has been shown to be superior to other policing schemes such as the moving window and the jumping window [12]. We also assume a certain adaptation layer between the video source and the ATM network which packetizes bits into cells with 47 bytes of payload.

The generic leaky-bucket model we consider (see Fig. 1) consists of an input buffer of size B_I and a token buffer of size B_L . The input buffer is used to smooth the traffic before it is sent into the network. The token buffer is used to monitor and police the input traffic. The basic working mechanism is described as follows.

At the beginning of the n th frame period, the real-time video source (video server sending stored video sequences in real time or real-time encoder) places F_n cells (the number of cells for the n th frame) into the input buffer. During this frame period, a certain number of cells (designated as S_n) are taken out of the input buffer and sent to the network with a rate equal to S_n/T , where T is the frame period. The input buffer occupancy at the end of the n th frame period is denoted as I_n , and can be expressed as

$$I_n = \max\{0, I_{n-1} + F_n - S_n\}. \quad (1)$$

The token buffer is drained at a constant leak rate. We denote the number of tokens leaked out in the T period to be R . Each cell sent from the input buffer attempts to add a token into the token buffer, and the cell is allowed to pass if the token buffer is not full. If the token buffer is full, the arriving cell is designated as nonconforming. The token buffer occupancy after S_n cells are sent out is denoted as L_n , and can be expressed as

$$L_n = \max\{0, L_{n-1} - R + S_n\}. \quad (2)$$

The conforming cells can be deterministically guaranteed delivery by allocating bandwidth equal to R/T and reserving buffer space at each ATM switch of size equal to the token buffer. When network congestion occurs, the network buffer is able to accommodate the conforming cells since the token-leaking mechanism ensures that the buffer will not overflow. Certain tradeoffs can be made between bandwidth and buffer, i.e., by allocating more bandwidth, less buffer needs to be reserved at the switches [13].

2) *Traffic Descriptors*: An important part of a CAC function is to find out how much network resources (in terms of B_L and R) are required based upon the source traffic descriptors and QoS requirement of the service. In this section, we illustrate the relationship between B_L, R and the three standard source traffic descriptors defined in [2]. The three descriptors defined for VBR services are: 1) the sustained cell rate Λ_{SCR} ; 2) the peak cell rate Λ_{PCR} ; and 3) the maximum burst size (MBS).

The relationship between Λ_{SCR} and R is straightforward: $\Lambda_{SCR} = R/T$. MBS is the maximum number of cells the source can send at peak rate, and is equal to $B_L + R$. The peak cell rate is equal to the maximum rate at which the source can send out cells, and can be assigned to be $(MBS)/T$.

B. Network Resource Requirements of MPEG Video

In this section, we use trace information from prerecorded video sequences with the objective to quantify the order of magnitude of the resource requirements for an uncontrolled VBR coded video stream, i.e., given a particular bandwidth allocation, how much buffer size is needed, or given a buffer size, how much bandwidth is needed, such that no cells are tagged by the UPC mechanism.

1) *Computing the Resource Requirements*: First, let us look at the constraints imposed by the requirement that the token buffer never overflows. From (2), during the n th frame period, the maximum number of cells that can be sent to the network should conform to

$$S_n \leq B_L + R - L_{n-1}. \tag{3}$$

For the input buffer, the condition for the prevention of buffer overflow can be expressed as

$$F_n \leq B_I + S_n - I_{n-1}. \tag{4}$$

Combining (3) and (4), we obtain the upper bound for F_n as

$$F_n \leq (B_I + B_L) + R - (L_{n-1} + I_{n-1}). \tag{5}$$

The similarity among (3)–(5) indicates that we can view the two-buffer model as a single virtual buffer with capacity $B = B_I + B_L$, output of R cells per frame period, and occupancy of $L_n + I_n$ in the n th frame period. By preventing this virtual buffer from overflowing, we can ensure that neither the token buffer nor the input buffer will overflow. A detailed analysis can be found in [14].

Therefore, given the trace information about F_n and an allocated bandwidth R/T , we can always find the minimum required B by ensuring that (5) is met for every frame in the sequence so that neither buffer overflows at any moment.

2) *Simulation Results*: In this section, we present the B – R relationship to guarantee that no cells are tagged for two video sequences coded with different combinations of coding parameters N and M . N is the intra-coded (I) frame to the I frame distance; hence, there are $N - 1$ predictive-coded (P) or bidirectional-coded (B) frames between two I frames. M is the P -to- P distance (hence, there are $M - 1$ B frames between two P frames). The following (N, M) combinations are used in the simulation.

- $N = 1$ and $M = 1$: The sequences use only I frames (IIII ...).

TABLE I
AVERAGE NUMBER OF CELLS PER FRAME

(N,M)	(1,1)	(12,1)	(12,2)	(12,3)
<i>Indiana Jones</i>	390	287	227	219
<i>The Larry King Show</i>	343	176	168	166

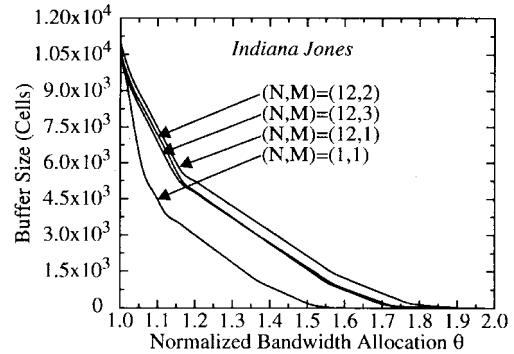


Fig. 2. B – R relationship for *Indiana Jones*.

- $N = 12$ and $M = 1$: Both I and P frames are used with 11 P frames between two I frames (IPPPPPPPPPPI...).
- $N = 12$ and $M = 2,3$: B frames are also used with, respectively, one, two B frames between two consecutive P frames or between I and P frames (IBPBPBPBPBPI... or IBBPBBPBBPBBI...).

The quantization scale is fixed at four for all the coded sequences.

In our simulations, we let R change from the average number of cells per frame for the sequence to the peak number of cells per frame, then for each R value, we find the minimum B value necessary to prevent cell tagging.

We define a normalized bandwidth allocation measurement variable θ as R/γ , where γ is the average number of cells per frame obtained from the trace. If θ is equal to one, it is ideal since the bandwidth utilization is the highest. The larger the θ , the lower the bandwidth utilization.

In Figs. 2 and 3, we show the relationship between θ and B , respectively, for two clips: one taken from the movie *Indiana Jones*, and the other one from the TV talk show *The Larry King Show*. The first clip consists of a series of consecutive high-motion scenes; this is typical of a high-action movie. Each scene lasts for about 2–5 s. The second one is similar to a videoconference-type sequence with a single person talking and a fixed background. Both clips are 30-s long, which is sufficient for the period between two successful bandwidth renegotiation events [6]. The resolution of the sequences used is 512×480 pixels. The γ values for the two sequences coded with different combinations of (N, M) are listed in Table I.

The first observation we can make is that there are some differences between the various (N, M) combinations. The I/P sequence ($N = 12, M = 1$) appears to be the most bursty in terms of buffer requirement given the same normalized

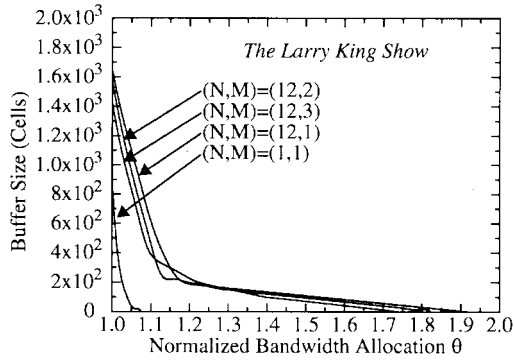


Fig. 3. B - R relationship for *The Larry King Show*.

bandwidth allocation. The pure I ($N = M = 1$) sequence is the least bursty, and the $I/P/B$ sequences lie in between. This can be explained in terms of the different coding modes used. When the sequence contains long periods of high activity (e.g., continuous high motion and/or many scene changes), the P coding mode does not work very efficiently. The motion is either too high or there are too many new elements in the current frame. This results in high bit rates even though the P mode is used. On the other hand, the B coding mode compensates for this inefficiency by referring to both previous and future frames. Hence, the $I/P/B$ mixed mode is more effective during periods of high activity. For the I only sequence, no prediction mode is used in the coding; hence, the overall average bit rate is much higher than the other two cases. The higher average rate results in a lower buffer requirement.

However, no matter what the values of N and M are, if we restrict the buffer size to be small, the required R is much higher than the average source rate. This results in inefficient bandwidth utilization. On the other hand, if we want a high bandwidth utilization, the buffer requirement is impracticably high. Generally, we want the buffers to be small so as to control the variance in the cell transfer delay. In particular, for real-time services, if the delay exceeds a certain threshold, a decoder regards the delayed cells as lost. So it is desirable to restrict the buffer size to be small.

We also note some differences between the two video sequences. *Indiana Jones* appears to be much more bursty as the buffer requirement is much larger. This is because the sequence contains a lot of motion and many scene changes. Unless explicitly specified, the results in the remainder of the paper are based on the *Indiana Jones* sequence.

In Fig. 4, we present the cell tagging rate versus B when R is fixed to be the average number of cells per frame. In Fig. 5, we present the cell tagging rate versus R when B is fixed to be equal to the peak number of cells per frame in the sequence. The results reveal that, under the above small B and R conditions, the cell tagging rate is on the order of (10^{-3} - 10^{-2}). Similar results were also presented in [12]. One may implement error concealment techniques at the decoder to reconstruct the corresponding lost components in the picture by interpolation in various ways. However, our earlier study on error concealment [7] showed that it can improve performance only when the cell loss rate (CLR) is lower than 10^{-5} . This would require large B and/or R to be assigned.

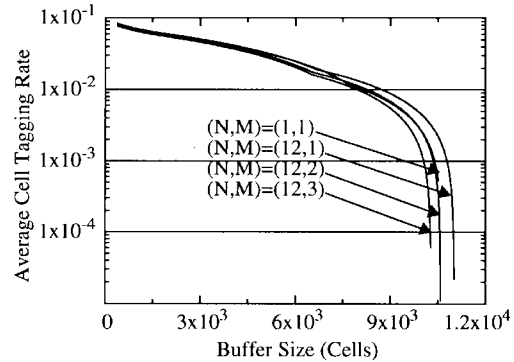


Fig. 4. Average cell tagging rate versus buffer size.

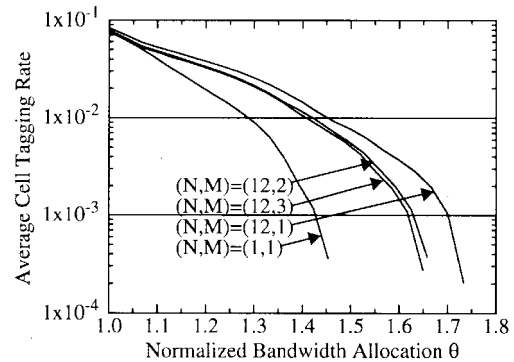


Fig. 5. Average cell tagging rate versus bandwidth allocation.

The above results indicate that controlling the VBR video traffic's burstiness is necessary if: 1) we want to restrict the B and R to achieve high bandwidth utilization or 2) only certain B and R are available.

III. SOURCE RATE CONTROL AND THE IMPACT ON VBR VIDEO QUALITY

Source rate control is used to control the coding procedure of real-time encoding, or the readaptation procedure of prerecorded sequences, such that the generated video traffic fits within the B and R constraint specified in (5) so as to guarantee that the buffers never overflow. This section focuses on the rate control for real-time encoding. There are also several studies on the readaptation of prerecorded sequences [15]–[17].

Source rate control could result in picture quality degradation. Hence, the important issue is how to dimension B and R to ensure that the picture quality degradation is within the desired quality level. We need to establish quantitatively the relationship between picture quality and source rate control in order to assess its performance and justify its use.

In Section III-A, we briefly look into the problem of how to do source rate control to prevent buffer misbehavior. Then in Section III-B, we show the impact of source rate control on video quality quantitatively by means of simulation results for different B and R values.

A. Source Rate Control Algorithm

1) *Basic Concept*: A real-time source rate control scheme should achieve efficient use of the buffer capacity to maximize

the picture quality (hence reduce picture quality degradation) and to prevent the buffer from overflowing.

In the MPEG-2 Test Model 5 (TM5) [11], a real-time source rate control algorithm is described. It consists of three steps which can essentially be decomposed into two phases: 1) bit allocation for each frame at the sequence level and 2) bit-rate control for each frame to achieve the target bit allocation obtained in phase 1. The scheme offers a dramatic improvement to the simulation model (SM) method used for MPEG-1 due to more sophisticated preanalysis and postanalysis routines while maintaining simplicity. There are still several known limitations. An important one is that in the first phase, there is no mechanism to avoid the buffer from overflowing, which is one of the essential functions of a source rate control scheme. In Sections III-A2) and III-A3), we briefly analyze the TM5 source rate control scheme and present ways to amend it.

2) *Sequence Level Bit Allocation*: The primary function of this phase is to efficiently allocate bits among the frames such that the buffer will never overflow and picture quality is as uniform as possible.

In the TM5 source rate control scheme, the bit allocation is based on the frame complexity, coding mode (I, P or B) for the current frame, and the remaining number of frames in the group of pictures (GoP). I pictures are generally assigned the most bits. B pictures are generally assigned the least bits since the information is not used by other frames. P frames are generally assigned more bits than B frames and fewer bits than I frames. Frame target setting then allocates target bits for a frame based on the frame complexity and the remaining number of frames in the GoP.

As we mentioned above, it is possible that the TM5 algorithm will result in an F_n that is larger than the upper bound calculated in (5). Unless we decrease F_n , the input buffer will overflow even when we send the largest possible number of cells that meets the upper bound in (3) to prevent the token buffer from overflowing.

To maintain uniform picture quality as much as possible, we should allow for as much variation as possible in the bit allocation scheme. When this situation occurs, the fewer bits we decrease by, the higher the picture quality that can be maintained.

Setting F_n to be the upper bound in (5) when overflow would occur means the smallest degradation in the current frame, but the subsequent frames will have fewer bits available. However, the frames that incur potential buffer overflow are generally I frames according to the TM5 algorithm; the subsequent frames are either P or B frames, which get a much lower bit rate allocation by the TM5 algorithm anyway. Therefore, in our simulation, we always set F_n to be the upper bound in (5) whenever potential overflow might happen.

It is to be noted that the upper bound in (5) is determined only by B and R . Hence, the encoder only needs to know the total buffer size B and bit rate R to do source rate control, regardless of how B is divided into B_L and B_T .

The scheme used in TM5 is certainly not the only way to do bit allocation. For example, a bit allocation scheme based on the Viterbi algorithm is proposed in [18] and [19] to achieve

minimum mean square error. But one of the appealing features of the TM5 scheme is its relative simplicity.

3) *Frame Level Bit Rate Control*: Once we have obtained the target number of bits for the current frame, we need to control the coding procedure to achieve this target. The TM5 algorithm attempts to adjust bit allocation if there is a significant difference between the number of target bits (anticipated bits) and the number of actual coded bits for a macroblock, and then recomputes the macroblock quantization factor according to the activity of the block against the normalized activity of the frame. The effect of this approach is to roughly assign a constant number of bits per macroblock.

Note that there is still room for improvement. For example, some macroblocks are visually more important than others, and hence we could allocate bits among the macroblocks according to their visual importance based on certain human visual system (HVS) models [20]. The resulting picture quality can be perceptually more uniform. However, the tradeoff between performance and complexity needs to be taken into consideration when devising such an algorithm. For simplicity and conformance with the public domain encoder model [21], we decided to conform with the TM5 algorithm for this phase of the process in our simulations.

B. Impact of Source Rate Control on Picture Quality

Source rate control introduces picture quality degradation. Therefore, an important issue is whether the resultant picture quality degradation can stay within acceptable levels given certain B and R or how to choose appropriate B and R to make it so. To study this, we looked into the following three scenarios.

In the first scenario, we fixed the allocated bandwidth R/T , and let B vary within a reasonable range. This allows us to determine the relationship between picture quality and B . The impact of B is important because larger B generally means higher allowed burstiness, and hence is supposed to increase the picture quality uniformity.

In the second scenario, we fix B and let R vary. This enables us to determine the relationship between R and picture quality. The impact of R on picture quality is complex because the higher the R , the higher the average picture quality in the video sequence, but the smaller the room for bit-rate variation. That may, in turn, affect the picture quality uniformity of the video sequence.

Finally, in the third scenario, we let B and R vary simultaneously and proportionally; this allows us to observe the overall impact of varying both B and R on the picture quality given a fixed maximum buffering delay.

In our simulations, we used three (N, M) combinations: 1) $N = 1$ and $M = 1$, 2) $N = 12$ and $M = 1$, and 3) $N = 12$ and $M = 2$. We adopted three metrics to measure the quality of a video sequence: average peak signal-to-noise ratio (PSNR), PSNR standard deviation (σ), and a grade point (GP) generated by an objective video quality assessment system [22] based on the HVS (by comparing the decompressed video sequence with the original video sequence in both the spatial and temporal domains). The average PSNR is used to measure the general coding quality of a video sequence, the σ indicates

TABLE II
QUALITY OF UNCONTROLLED VBR CODED SEQUENCES

(N,M)	σ (dB)	PSNR (dB)	Grade Point
(1,1)	0.99	39.26	4.65
(12,1)	0.91	39.30	4.69
(12,2)	0.92	39.32	4.69

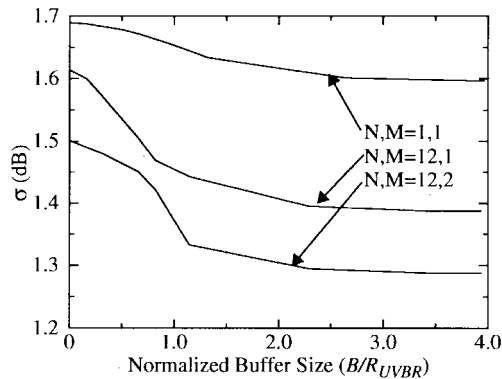


Fig. 6. σ versus buffer size.

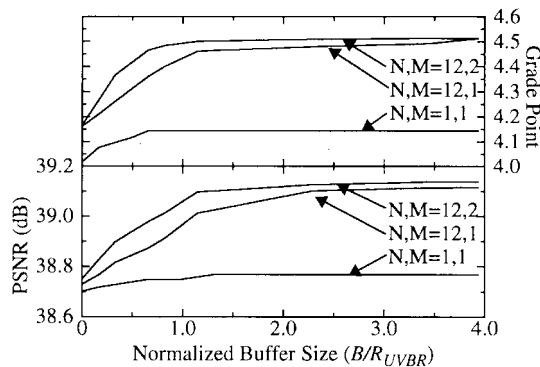


Fig. 7. PSNR and grade point versus buffer size.

the degree of variation in picture quality in the video sequence, and the grade point reflects the overall viewing quality as perceived by a human being. The grade point ranges between 1.0 and 5.0, with 5.0 meaning excellent quality, 4.0 meaning good quality, 3.0 meaning acceptable quality, 2.0 meaning bad quality, and 1.0 meaning absolutely unacceptable quality.

In order to compare the performance of source rate control with that of uncontrolled VBR coding (quantized with fixed Q of 4), we list in Table II the quality of the uncontrolled VBR coded sequences using different (N, M) combinations.

1) *Scenario 1*: In this scenario, we study the impact of B on the picture quality given a fixed R . In our simulations, we let R be R_{UVBR} , the average number of cells per frame of the uncontrolled VBR coded sequences (listed in Table I). The B value changes from zero to four times R_{UVBR} .

In Fig. 6, we present the σ results. The lower half of Fig. 7 shows the PSNR results, and the upper part shows the grade point results. The buffer size is normalized by R_{UVBR} . We can make the following observations.

a) For a pure I sequence ($N = M = 1$; there is only one frame in each GoP), B does not have too much effect on

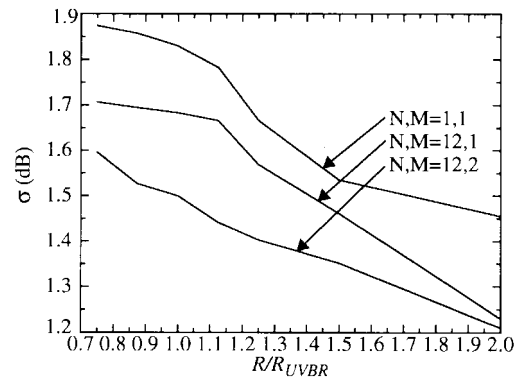


Fig. 8. σ versus bit rate.

the picture quality; all three metrics do not vary too much as B increases. This is due to the fact that the average rate over a GoP (hence, each I frame) is controlled to be close to R in the sequence level bit allocation.

b) Contrary to the pure I sequence, the impact of B is more obvious for the I/P sequence, especially in the region of $0 \leq B \leq R_{UVBR}$. For the $I/P/B$ sequence, the impact of B is even more substantial in the same region. This can be explained again by the different coding modes. Picture quality degradation mainly happens in the high-activity periods, during which the bit rate has to be cut in order to prevent buffer misbehavior. I/P sequences still have the tendency to generate high bit rates during these periods due to the inefficiency of the P mode in the high-activity scenes. Hence, a larger buffer can only allow more bits for the first several frames at the start of the high-activity periods. The buffer saturates very soon, and then the following frames are enforced to generate fewer bits than necessary to keep constant quality. For the $I/P/B$ sequence, as the bidirectional coding mode can generally bring down the bit rate effectively, the larger B can be utilized by the I and P coded frames to generate more bits to keep a high picture quality.

c) The increase in quality is limited for both I/P and $I/P/B$ sequences after B becomes larger than R_{UVBR} . Actually, we observe that when the value of B exceeds $3R_{UVBR}$, the encoding procedure that decreases the bit allocation F_n occurs very rarely. This is again due to the fact that the TM5 source rate control enforces the average rate conformance over a GoP period. The sequence level bit allocation automatically lowers the frame bit rate when the high-activity period spans more than one GoP. This indicates that the TM5 scheme can be further amended by incorporating the impact of B . The GoP level conformance is not necessary as long as we can ensure that the buffer does not overflow.

2) *Scenario 2*: In this scenario, we fix B , and let the target bit rate R vary over the range from 0.75 to 2 times R_{UVBR} to study the impact of R on picture quality. We choose the value of B to be $3R_{UVBR}$, which limits the buffer delay for interactive video services.

In Fig. 8, we present the σ results. The lower half of Fig. 9 shows the PSNR results, and the upper part shows the grade point results. We can make the following observations.

a) For the pure I sequence, the improvement in quality is substantial in contrast to the first scenario.

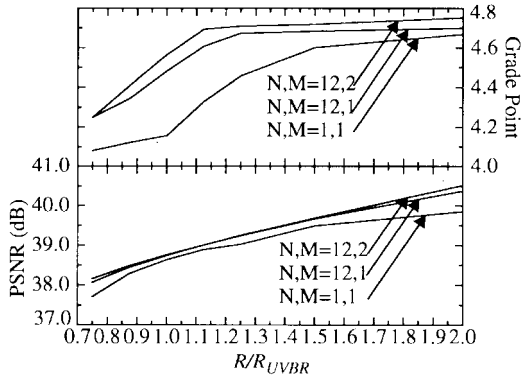


Fig. 9. PSNR and grade point versus bit rate.

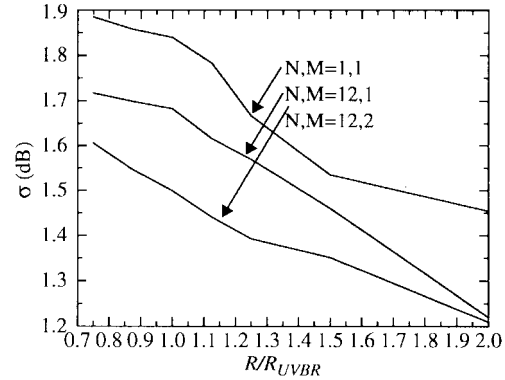


Fig. 10. σ versus bit rate and buffer.

b) For the I/P sequence, we notice an interesting phenomenon for the behavior of σ : there is a cutoff point before which it does not decrease substantially. Two things are happening at the encoder as we increase R : first, a larger target bit rate may introduce a larger bit-rate variation as everything is scaled up, but with a fixed B , the buffer's ability to accommodate bursts is relatively small. This can have an adverse impact on the σ even though the average PSNR increases. Second, since the allocated bandwidth is equal to the target bit rate, increasing R means that we can accommodate more long bursts caused by high-activity scenes. This tends to increase the picture quality for the long burst periods, and hence reduce the picture quality deviation.

From the figure, we conclude that the first offsets much of the effect of the second until the bit-rate variation no longer increases substantially with R , and then the impact of the larger R on the reduction of the long bursts becomes dominant. After that cutoff point, the σ goes down steadily as R is increased.

For the $I/P/B$ sequence, the cutoff effect for σ is less obvious. This is primarily due to the coding efficiency of the B mode during times of high-activity periods. The higher allocated bandwidth can then be fully utilized for the I and P frames in the long bursts to improve their picture quality.

c) For both the I/P and $I/P/B$ sequences, PSNR and grade point increase steadily with R , showing that the target bit rate plays a dominant role in the average picture quality of a video sequence.

d) For I/P and $I/P/B$ sequences, although the σ is always larger than that of the uncontrolled VBR coding, the PSNR values are close to those of the uncontrolled VBR coded sequences (with average rate of R_{UVBR}) in the region between $R = 1.1R_{UVBR}$ and $R = 1.2R_{UVBR}$. This basically means that with a small increase in the bit rate, the average quality of a controlled VBR sequence is close to that of the uncontrolled VBR sequence. It is to be noted that if that same increase in bandwidth were allocated to the transmission of the uncontrolled VBR sequence, the resultant cell tagging rates would still be very high (see Figs. 4 and 5). For I sequences, however, a much higher increase in bit rate is required to achieve the same average quality of the uncontrolled VBR sequence.

3) *Scenario 3*: Finally, in this scenario, we let B and R vary simultaneously and proportionally to observe the overall

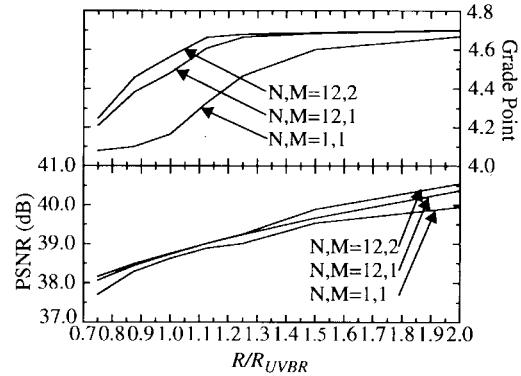


Fig. 11. PSNR and grade point versus bit rate and buffer.

impact of varying both B and R on the picture quality. Again, we choose the value of B to be $3R$, which limits the buffering delay. R ranges from 0.75 to 2 times the value of R_{UVBR} .

In Fig. 10, we present the σ results. The lower half of Fig. 11 shows the PSNR results, and the upper part shows the grade points.

For I sequences, the results are almost the same as those of the second scenario as R is the dominant force for this kind of sequence. For I/P and $I/P/B$ sequences, the results of PSNR and grade point are also very similar to those of Scenario 2. The σ also goes steadily down as expected since the buffer constraint is eliminated as compared with Scenario 2.

As can be seen from the three scenarios outlined here, although allocated bandwidth plays an important role, the impact of source rate control on the picture quality of a video sequence is a complex issue, and is also related to different coding options. Hence, the user should take into consideration all of the factors to make the proper choice of parameters that matches the individual situation the best.

IV. QUALITY CONTROL FOR SCALABLE VIDEO TRANSMISSION

One of MPEG-2 video's major extensions over MPEG-1 is the addition of scalability. It is designed to support applications beyond that supported by single-layer video such as: accommodating different classes of users, interworking of different video standards, video service hierarchies with multiple spatial, temporal and quality resolutions, etc.

In this paper, we focus on a scalable scheme based on two layers only. As for our earlier study on single-layer video

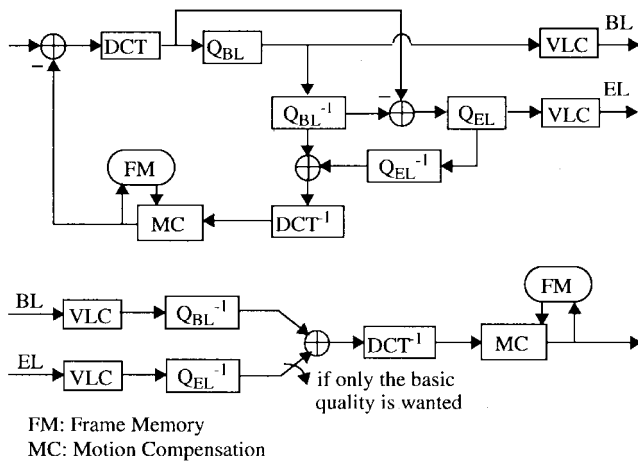


Fig. 12. Encoder and decoder diagram for MPEG-2 SNR scalability.

transmission, we assume the use of a guaranteed service for the delivery of UPC conforming cells of both layers and that all tagged cells are lost. Below, we discuss some interesting issues related to the provision of scalable video services.

A. Basic Concepts Related to Scalable Video Coding

Although a simple solution to scalable video is to produce multiple independently coded streams of video, a more efficient alternative is scalable coding. In scalable coding, a given reproduction of video is partially reutilized in the coding of the next reproduction of video. The four basic scalabilities defined in MPEG-2 are: data partitioning, SNR scalability, spatial scalability, and temporal scalability. For example, two-layered SNR scalability generates two video layers of the same spatial resolution, but with different video qualities. The base layer is coded typically using a larger quantization scale to generate the roughly quantized picture, and thus provides the basic video quality service. The residual error, which is typically coded with a smaller quantization scale, results in the enhancement layer. When added back to the base layer, it regenerates a higher quality reproduction of the input video sequence, and hence provides an enhanced quality service.

Fig. 12 shows the basic coding/decoding scheme for SNR scalability. The base layer's encoding procedure is identical to that of a single-layer video. The quantized discrete cosine transform (DCT) coefficients from the base layer are inversely quantized and then subtracted from the unquantized DCT coefficients. The residual quantization error is then quantized with Q_{EL} to produce the enhancement layer. In this paper, we will study SNR scalability.

A very important feature of SNR scalability is that the motion-compensated encoding (for P and B frames) is generally based on the reference frame reconstructed from both layers to ensure the correct decoding for the enhanced quality service. However, for the basic quality service, the motion-compensated decoding of the base layer is based solely on the reference frame reconstructed from the base layer (see Fig. 12). This results in so-called *drift* errors, and has a profound impact on both the basic and the enhanced quality services.

Note that since MPEG-2 only specifies the standard decoding procedure (as illustrated in Fig. 12), it is possible to confine the coding loop only to the base layer, i.e., the motion-compensated coding is based on only the base layer to prevent drift errors for the basic quality service. However, when the allocated bit rate of the base layer is small, this will cause a serious drift problem for the enhanced quality service as the MPEG-2 standard mandates that its decoding loop be based upon both layers. We chose to base our study on the approach that uses both layers at the encoder so that we can maintain a high quality for the enhanced quality service no matter what bit rate is used for the base layer.

B. Difference in the Impact of the Base and the Enhancement Layers on Video Quality

For SNR scalability, the base layer provides the basic quality service. When the decoder only receives the base layer bit stream, it treats it as a single-layer bit stream to reconstitute the video sequence with the basic quality. The enhancement layer is used to enhance the video quality. When both layers are present at the decoder, an enhanced quality video service is obtained.

The different coding methodologies and information content of the two layers make the impact that the base layer and the enhancement layer each have on the basic and the enhanced video quality very different. In this section, we look at their difference from the perspective of video quality control for both the basic and enhanced quality services.

The base layer contains the basic video information necessary for the reconstruction of both the basic and enhanced video quality. Furthermore, in order to reduce the coding overhead, certain critical information such as the motion vector information is only put into the base layer. Both the basic and the enhanced video quality services are vulnerable to the cell losses in the base layer. If base layer cell loss occurs, the enhancement layer cells may be useless even when there is no loss in this layer. This means that implementing source rate control on the base layer is necessary so as to control the cell losses.

The enhancement layer contains the residual error signal between the base layer and the original video information. The question is whether or not source rate control should be implemented for this layer.

If the enhancement layer is coded as controlled VBR to conform to the allocated network resources for this layer, picture quality degradation can result due to a less efficient encoding procedure. This is because motion-compensated encoding is based on frames reconstructed from both layers. Source rate control results in the loss of some information. Therefore, when comparing the current frame to the reconstructed previous frame at the encoder for motion-compensated encoding, a larger error signal is produced that requires a larger number of bits for encoding. Source rate control reduces this number of bits by reducing the quality (using a higher quantization step). This results in picture quality degradation for both the basic and the enhanced quality services in the encoding procedure.

Hence, for the basic quality service, if the enhancement layer is coded as uncontrolled VBR, there may be less picture

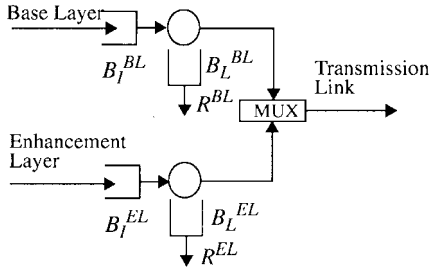


Fig. 13. Scalable video transmission model.

quality degradation incurred by the encoding process. However, compared with the controlled VBR-coded enhancement layer, the drift error we mentioned in Section IV-A is larger since there is more information in the uncontrolled VBR-coded enhancement layer, and hence the drift caused by excluding this part of the information from the decoding loop is worse.

For the enhanced quality service, if the enhancement layer is coded as uncontrolled VBR, a higher quality image may be obtained due to the more efficient motion-compensated encoding. But some cells of the enhancement layer could be tagged since it is not rate controlled to fit within the guaranteed service parameters. Should cell losses occur, the video quality will be degraded. If this layer is coded as controlled VBR, there will be no quality degradation due to potential cell losses, but picture quality is reduced due to less efficient encoding.

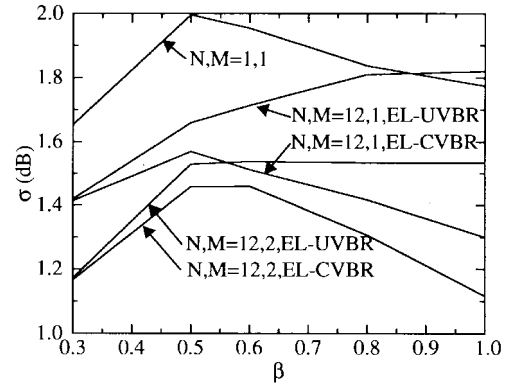
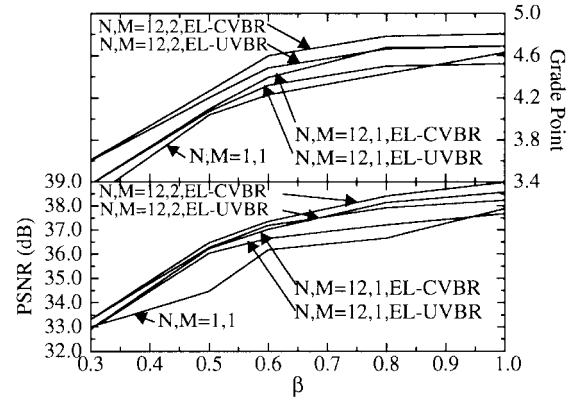
Therefore, it is not a clear choice whether or not to implement source rate control for the enhancement layer. In the next section, we will compare these two different approaches, and show how they affect the basic and the enhanced quality services. Note that, as mentioned earlier, we always assume worst case scenario, i.e., all tagged cells are lost (as we have no way of knowing what the CLR will be).

C. Quality Control for Scalable Video Transmission

Fig. 13 shows the basic model for scalable video transmission using guaranteed service. The only difference when compared to the single-layer video transmission is the addition of another leaky-bucket traffic control module for the enhancement layer. We denote the bandwidth allocated for the base layer as R^{BL} , the reserved buffer for the base layer as B^{BL} (B_I^{BL} and B_L^{BL}); similarly, the parameters for the enhancement layer are R^{EL} and B^{EL} (B_I^{EL} and B_L^{EL}), respectively. We also define a partition parameter β which measures the portion of resources allocated to the base layer given fixed total resources for both layers:

$$b = \frac{R^{BL}}{R} = \frac{B^{BL}}{B}.$$

In the last section, we presented two possible choices with regard to the coding of the enhancement layer: 1) the enhancement layer is coded as controlled VBR, denoted as EL-CVBR and 2) the enhancement layer is coded as uncontrolled VBR, denoted as EL-UVBR. For both schemes, the base layer is always coded as controlled VBR. Given the two choices, we would like to know which scheme results in better video quality given different combinations of R^{BL} , B_I^{BL} , B_L^{BL} , R^{EL} , B_I^{EL} , and B_L^{EL} .


 Fig. 14. σ versus β .

 Fig. 15. PSNR and grade point versus β .

1) *Basic Quality*: Since we implement source rate control on the base layer and use guaranteed service for its transmission, we assume that the quality of the reconstructed sequence at the receiver is identical to that at the output of the encoding procedure.

From our discussions in Section IV-B1), we know that for the EL-UVBR approach, higher coding efficiency can be achieved, and that for the EL-CVBR approach, the drift error is smaller. Hence, whether the EL-CVBR scheme or the EL-UVBR scheme yields higher basic video quality really depends on which force has a larger impact.

In Fig. 14, we present the σ for the three N/M combinations, respectively. The PSNR and grade point results are shown in Fig. 15.

The first observation we can make is that there is no difference between the EL-UVBR and EL-CVBR for the pure I sequence. The reason is obvious: no motion-compensated coding is used for I frames. For I/P and $I/P/B$ sequences, judging from all of the metrics, it appears that the EL-CVBR outperforms the EL-UVBR approach. Especially, when β increases, the difference in performance widens. This is because when β is small, the impact of the drift error is dominant for both approaches. As more bandwidth is allocated to the base layer, the impact of the drift error becomes relatively smaller for the EL-CVBR approach, while the drift error still causes quite some deterioration for the uncontrolled VBR approach. This is particularly true when $\beta = 1.0$; then there is no enhancement layer for the EL-CVBR, and hence

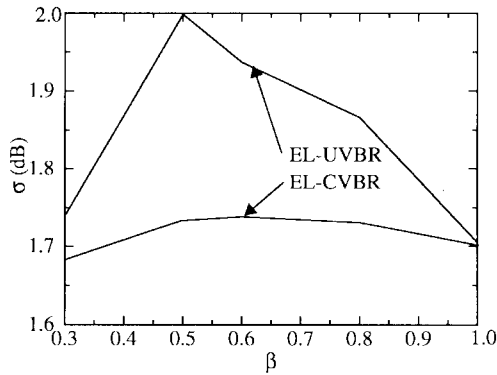


Fig. 16. σ versus β ($N, M = 1, 1$).

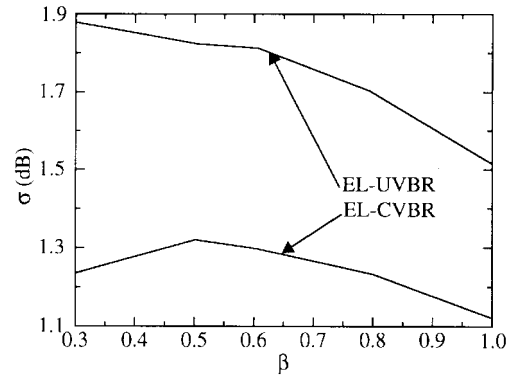


Fig. 18. σ versus β ($N, M = 12, 2$).

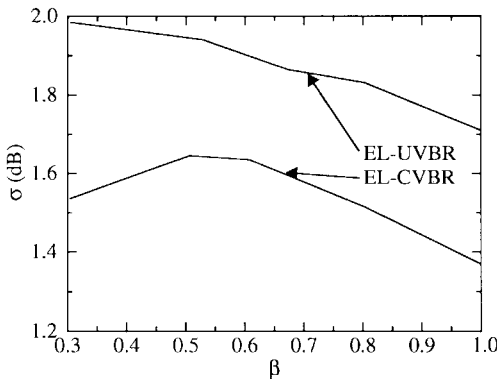


Fig. 17. σ versus β ($N, M = 12, 1$).

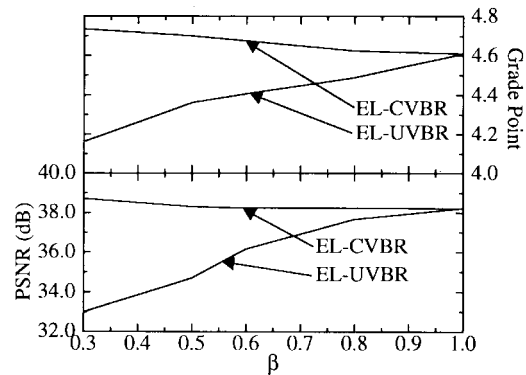


Fig. 19. PSNR and grade point versus β ($N, M = 1, 1$).

no drift since the allocated bandwidth (target bit rate) for the enhancement layer is zero. However, for EL-UVBR, there is still an enhancement layer (there will always be an error signal resulting from the difference between the original image and the rate-controlled base layer; this error signal constitutes the enhancement layer), and hence drift errors result.

If we further compare the I/P and $I/P/B$ sequences, once again the $I/P/B$ sequence displays better performance due to its coding efficiency during periods of high activity.

Noteworthy are the σ results. For the EL-UVBR approach, the σ increasing monotonically with β as drift is the dominant force. However, for the EL-CVBR approach, the drift problem becomes relatively smaller as β is increased.

2) *Enhanced Quality*: For the EL-CVBR approach, no cell tagging occurs in either the base or the enhancement layers. Hence, the received quality is assumed to be identical to that of the output of the encoder.

For the EL-UVBR approach, enhancement layer cells are tagged whenever the traffic does not fit within the guaranteed service parameters. This, in turn, causes quality degradation at the decoder end (as we assume all tagged cells are lost). So the question is whether the advantage in coding (more efficient than EL-CVBR) can offset the degradation incurred by cell losses.

In Figs. 16–18, we present the σ results for three N/M combinations. The PSNR and grade point results are shown in Figs. 19–21.

The answer to the above question is obvious judging from all three metrics: the EL-CVBR outperforms the EL-UVBR

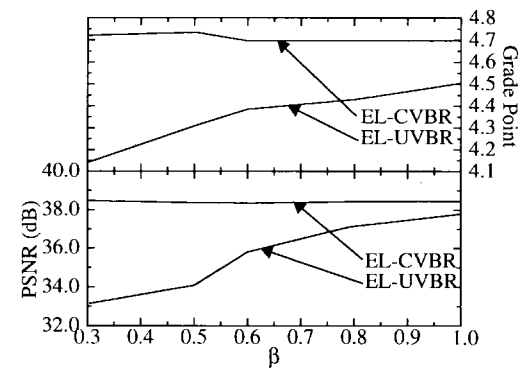


Fig. 20. PSNR and grade point versus β ($N, M = 12, 1$).

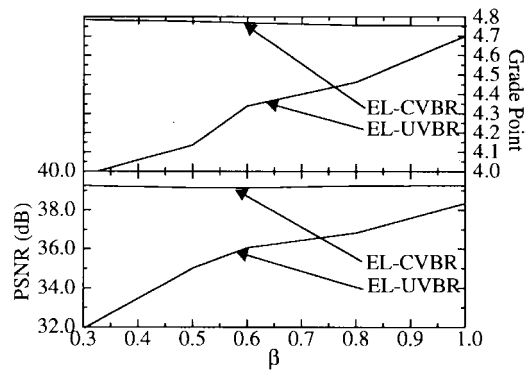


Fig. 21. PSNR and grade point versus β ($N, M = 12, 2$).

approach. It appears that cell losses play a dominant role in the reconstruction of the enhanced quality. For the EL-UVBR approach, as β is increased, most enhancement layer cells are tagged; hence, the received quality is lowered to be close to the basic quality decoded only from the base layer.

In this section, we showed that, in order to achieve better quality for both the basic and enhanced quality services, both the base and the enhancement layers need to be controlled VBR coded.

V. SUMMARY

ATM-based networks support VBR video services. However, because of the highly bursty nature of VBR video, during periods of congestion, the resultant cell losses due to buffer overflow could incur severe quality degradation. Source rate control can be implemented to generate a controlled VBR bit stream whose burstiness is constrained to fit within certain bounds. An important issue is whether the picture quality degradation incurred by source rate control is within the desired level or how to do control to make it so. In this paper, we establish quantitatively the relationship between the picture quality degradation and different combinations of allocated bandwidth, buffer size, and other key coding parameters for MPEG-2 video. We also look into the quality control problem for scalable video coding and transmission. Our study reveals that, in order to maximize the quality of both the basic and the enhanced services, source rate control should be implemented for both the base layer and the enhancement layer. The relationships between the two types of quality and different combinations of bandwidth allocations, buffer sizes, and other coding parameters are also established for scalable MPEG-2 video.

An important issue that has not been addressed in this paper is statistical multiplexing. Theoretically, if statistical multiplexing is used, fewer resources need to be reserved in the network. But the network should always guarantee the delivery of conforming cells with a very high probability. Whether this is possible for highly bursty VBR video sources or how to achieve this is still a challenging issue.

REFERENCES

- [1] M. De Prycker, *Asynchronous Transfer Mode: Solution for Broadband ISDN*, 2nd ed. New York: Ellis Horwood, 1993.
- [2] *Traffic Control and Congestion Control in B-ISDN*, ITU Study Group 13, Perth, Australia, Nov. 1995.
- [3] G. Karlsson, "Asynchronous transfer of video," *IEEE Commun. Mag.*, vol. 34, pp. 118–126, Aug. 1996.
- [4] M. W. Garrett, "Contributions toward real-time services on packet networks, Ph.D. dissertation, Columbia Univ., New York, NY, May 1993.
- [5] P. Pancha and M. El Zarki, "Bandwidth allocation schemes for variable bit rate MPEG sources in ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 190–198, June 1993.
- [6] D. Reininger and D. Raychaudhuri, "Bandwidth renegotiation for VBR video over ATM networks," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 1076–1086, Aug. 1996.
- [7] W. Luo and M. El Zarki, "Analysis of error concealment techniques for MPEG-2 video transmission over ATM based networks," in *Proc. SPIE Visual Commun. Image Processing '95*, Taipei, Taiwan, May 1995, pp. 102–108.
- [8] J. P. Leduc, *Digital Moving Pictures—Coding and Transmission on ATM Networks*. New York: Elsevier, 1994, pp. 300–301.

- [9] P. Pancha and M. El Zarki, "Leaky bucket access control for VBR MPEG video," in *Proc. IEEE INFOCOM '95*, Boston, MA, Apr. 1995, pp. 796–803.
- [10] *Generic Coding of Moving Picture and Associated Audio*, ISO/IEC 13818-2, MPEG-2 Draft International Standard, May 1994.
- [11] *Test Model 5*, ISO/IEC/JTC1/SC29/WG11, Apr. 1993.
- [12] E. Rathgeb, "Policing of realistic VBR video traffic in ATM networks," *Int. J. Digital Analog Commun. Syst.*, vol. 6, pp. 213–226, 1993.
- [13] A. Elwalid, D. Mitra, and R. Wentworth, "A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 1115–1127, Aug. 1995.
- [14] A. Reibman and B. Haskell, "Constraints on VBR video for ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 361–372, Dec. 1992.
- [15] D. G. Morrison, M. E. Nilsson, and M. Ghanbari, "Reduction of the bit-rate of compressed video while in its coded form," in *Proc. 6th Int. Workshop Packet Video*, Portland, OR, session D-17, Sept. 1994.
- [16] A. Eleftheriadis and D. Anastassiou, "Constrained and general dynamic rate shaping of compressed digital video," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Washington, DC, Oct. 1995, pp. 396–400.
- [17] R. J. Safranek, C. R. Kalmanek, Jr., and R. Garg, "Methods for matching compressed video to ATM networks," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Washington, DC, Oct. 1995, pp. 13–16.
- [18] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximation," *IEEE Trans. Image Processing*, vol. 3, pp. 26–40, Jan. 1994.
- [19] C. Hsu and A. Ortega, "Joint encoder and VBR channel optimization with buffer and leaky bucket constraints," in *Proc. Int. Symp. Multimedia Commun. Video Coding*, New York, NY, Oct. 1995, pp. 104–105.
- [20] W. Luo and M. El Zarki, "Adaptive layered source coding for MPEG-2 video transmission," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Washington, DC, Oct. 1995, pp. 17–20.
- [21] MPEG Software Simulation Group, MPEH-L@netcom.com.
- [22] A. Webster, C. Jones, M. Pinson, S. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," in *Proc. SPIE Human Vision, Visual Processing, Digital Display TV*, vol. 1913, San Jose, CA, Feb. 1993, pp. 15–26.
- [23] P. Salama, N. Shroff, E. Coyle and E. Delp, "Error concealment techniques for encoded video streams," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Washington, DC, Oct. 1995, pp. 9–12.



Wenjun Luo received the B.S. degree in electronic engineering from the Jiao Tong University, Shanghai, China, in 1992, and the M.S. degree from the Department of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1994.

He is now a Ph.D. candidate in the same department, where he also serves as a Teaching Assistant and Research Assistant. His current research interests include video compression, error concealment, QoS control for packet video transmission in broadband networks, and multimedia systems.



Magda El Zarki (S'81–M'88) received the B.E.E. degree from Cairo University, Cairo, Egypt, in 1979, and the M.S. and Ph.D. degrees in electrical engineering, both from Columbia University, New York, NY, in 1981, and 1987, respectively.

From 1981 to 1983, she worked as a Communications Network Planner in the Department of International Telecommunications, Citibank, New York, NY. She joined Columbia University in 1983 as a Research Assistant in the Computer Communications Research Laboratory. Currently, she holds the position of Associate Professor in the Department of Electrical Engineering, University of Pennsylvania, Philadelphia, where she is involved in doing research in telecommunication networks. She also holds a secondary appointment in the Department of Computer and Information Sciences.

Dr. El Zarki is a member of the Association for Computing Machinery and Sigma Xi. She is an Editor for several journals in the telecommunications area and is actively involved in many international conferences. She is currently on the Board of Governors of the IEEE Communications Society, and also is the Vice Chair of the IEEE Technical Committee for Computer Communication.