論文 98-23-3-9

Source Traffic Smoothing Algorithm for a Real-time VBR MPEG Video Service and Its Performance Analyses

Jin-Soo Kim*, Jae-Kyoon Kim* *Regular Members*

실시간 VBR MPEG 영상 정보원을 위한 트래픽 평활화 기법과 성능 분석

正會員 김 진 수*, 김 재 균*

※본 논문은 한국통신 연구개발 본부의 지원을 받았음

ABSTRACT

In this paper, we propose a new source traffic smoothing algorithm, that can be effectively used in various live video applications such as video conferencing, where the sender and receiver buffer sizes are usually fixed prior to transmission. For these applications, we describe the necessary constraints imposed by the delay bound, and by the sender/receiver buffer sizes. Then, based on these constraints, a new source traffic smoothing scheme is designed in such a way as to smooth maximally the transmission rate, while controlling so that the buffer overflow and underflow rates may be avoided.

Experimental results show that the proposed method has led to low loss rates at the moderately small buffer sizes. Then, the smoothing performances of the proposed scheme are investigated in terms of the buffer sizes and the delay bound. Finally, based on simple queueing results, the performances of the proposed scheme are evaluated. Through these experiments, it is confirmed that the proposed method is effective in reducing both short-term and mid-term burstiness of the transmission rate for visual conferencing applications.

요 약

본 논문에서는 영상 회의와 같이 송신측과 수신측의 버퍼 크기가 고정된 실시간 영상 통신용에 효과적으로 적용될 수 있는 적응적 정보원 트래픽 평활화 기법을 제안한다. 이를 위해 먼저, 지연과 버퍼 크기에 의해 부여된 제

*한국과학기술원 전자공학과

論文番號:97333-0918
接受日字:1997年 9月 18日

한 조건들을 정의하고, 이들의 제한 조건에 기초하여 버퍼의 넘침 현상과 고갈 현상을 회피하는 방식으로 트래픽 평활화 기법을 설계한다.

제안한 트래픽 평활화 기법은 버퍼의 넘침으로 인하여 발생되는 손실 특성, 전송되는 트래픽의 첨두치율, 시간 변이 특성, 단위 시간당 일정 비트율의 전송 기법을 도입하는 방법에 의한 재협상의 실패율, 그리고 등가 대역을 구함으로써 분석된다. 이와 같은 분석을 통하여, 제안한 기법은 실시간 VBR MPEG 영상 통신을 위한 정보원의 평활화에 효과적으로 적용될 수 있음을 알 수 있다.

Ⅰ. Introduction

VBR(variable bit rate) video coding and transmission are known to be advantageous in various points of views such as an constant image quality due to the evenly distributed distortion and an efficient bandwidth utilization due to the statistical multiplexing. Recently, as one way to provide VBR-coded services, some literatures have dealt with source traffic smoothing methods[1-5], which are based on the introduction of delay and rate buffering between the video encoding and decoding processes. The main purposes of source traffic smoothing are to transmit coded-data by conserving VBR characteristics of the original video sequence itself and to alleviate the network load without wasting excessive bandwidth.

Several transmission techniques of a pre-coded video source such as VoD applications have been addressed in [6-9]. In these applications, by capitalizing on prior knowledge of the coded-frame sizes for the entire video, these techniques can smooth traffic on a large time scale by prefetching frames into a buffer in advance of bursts of large frames. In contrast to pre-coded video applications, interactive video applications, such as video conferencing, typically have the limited knowledge of coded-data sizes and finite buffer sizes in the sender and receiver sides.

As a result, live video service applications require dynamic techniques that can react quickly to changes in coded-frame sizes and the available buffer sizes. For live VBR video services, in [1], Reibman and Haskell studied constraints on buffer sizes and transmission rates. But, they focused on the encoder rate control to prevent overflow and underflow at the encoder and decoder. In [2], a lossless smoothing scheme is suggested not to yield a bursty traffic by using the continuous service property and the delay bound. His scheme is not restricted by the buffer sizes. And in [3-5], several VBR traffic smoothing schemes have been dealt with VBR video coding and transmission on an ATM network.

In this paper, we consider an adaptive source traffic smoothing algorithm that can be effectively applicable for video conferencing applications with the finite and small buffer sizes. The proposed smoothing algorithm is based on the constraints that are imposed by the delay bound, by the sender/receiver buffer sizes. By using these constraints, we propose an adaptive source traffic smoothing algorithm which minimizes the overflow and underflow rates due to finite buffer sizes, while ensuring that the transmitted data is being maximally smoothed.

Ⅱ. System Model and Notations

The approach of this paper is specified at an interface node where the user will be in a position to control the instantaneous transmission rate and buffer occupancies under a given delay bound. In Fig. 1, the situation in an adaptive transmission rate control is shown as a simplified block diagram. Let $s(j)$ be the number of generated bits in the interval $[(j-1)\Delta t,\ j\Delta t)$, where Δt is the duration corresponding to one uncoded frame interval. It is assumed that $s(j)$ is uniformly fed to the sender buffer during this time interval. Additionally, the size of decoding data package is assumed to be same as the generated-data size by encoding per frame.

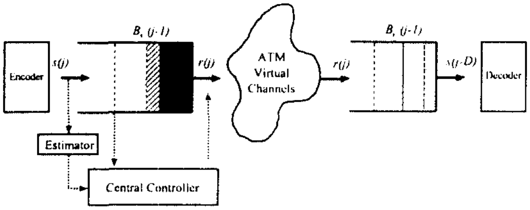


Fig. 1 An overview figure for an adaptive source traffic smoothing.

Similarly, let $r(j)$ be the number of bits that are uniformly transmitted into the network in the interval $[(j-1)\Delta t,\ j\Delta t)$. The value of $r(j)$ depends on the nature of the sender and receiver buffer fullness as well as the delay bound. Let us define $D\Delta t$ as the allowable delay bound with which a specific coded-frame data can be stayed in the sender buffer[1]. It means that the sender side must terminate transmitting a specific coded-frame data within $D\Delta t$ into network, after the first bit of that data is generated. We assume that the network itself is transparent to transmit video service, i.e., the network is considered to be a source of no delay and no loss. Under this assumption, the receiver must wait, at least, $D\Delta t$ before starting to decode. As expressed in [1], based on the above definitions and assumptions, the sender and receiver buffer fullness at time $j\Delta t$, $B_s(j)$ and $0 \le B_s(j) \le B_s^{size}$ can be expressed as

$$B_s(j)=\sum_{i=1}^{j} s(i)-\sum_{i=1}^{j} r(i) \qquad \text{for } j>0 \qquad (1)$$

$$B_r(j)=\sum_{i=1}^{j} r(i), \qquad \text{for j} < \text{D}$$

$$B_r(j)=\sum_{i=1}^{j} r(i)-\sum_{i=1}^{j-D} s(i), \qquad \text{for j} > \text{D} \qquad (2)$$

From (1) and (2), two different forms of $B_r(j)$ can be written for $j>D$ as follows

$$B_r(j)=\sum_{i=j-D+1}^{j} s(i)-B_s(j) \text{ for } j>D \qquad (3)$$

$$B_r(j)=\sum_{i=j-D+1}^{j} r(i)-B_s(j-D) \text{ for } j>D \qquad (4)$$

We note that in order to guarantee lossless smoothing, the conditions of $0 \le B_s(j) \le B_s^{size}$ and $0 \le B_r(j) \le B_r^{size}$ must be satisfied for all j, where $B_s(j)$ and $B_r(j)$ represent the sizes of sender and receiver buffers, respectively. Hence from (3) and (4), we obtain simple boundaries of partial sums over $D\Delta t$ intervals.

$$\sum_{i=j-D+1}^{j} s(i) \le B_s^{size}+B_r^{size} \qquad (5)$$

$$\sum_{i=j-D+1}^{j} r(i) \le B_s^{size}+B_r^{size} \qquad (6)$$

It means that overall buffer size($B_s^{size}+B_r^{size}$) must be large enough to absorb the variability of data generated, or transmitted during the delay bound.

Ⅲ. Proposed Source Traffic Smoothing Algorithm

3.1 Constraint on the Receiver Buffer Size

If the receiver buffer size is sufficiently large, the selection range of $r(j)$ can be wide. But, in the real case that the receiver buffer size is small and fixed, the maximum value of $r(j)$ is restricted by the overflow prevention condition of the receiver buffer. From (4), since $B_r(j) \le B_r^{size}$ must be satisfied for all j, that is,

$$\sum_{i=j-D+1}^{j} r(i)-B_s(j-D) \le B_r^{size}\text{, for all } j \qquad (7)$$

must be maintained. Hence, the maximum value of $r(j)$ is bounded as follows,

$$r^{\max}(j)=\max\left\{B_r^{size}+B_s(j-D)-\sum_{i=j-D+1}^{j} r(i),\ 0\right\} \qquad (8)$$

or by using (1) in (8),

$$r^{\max}(j) = \max\left\{B_r^{size} + B_s(j-1) - \sum_{i=j-D+1}^{j} s(i),\ 0\right\} \quad (9)$$

3.2 Constraint on the Delay Bound

Fig. 2 shows the queueing structure in the sender buffer at time $(j-1)\Delta t$. The coded-frame data that waits for transmission in the sender buffer, can not wait longer than the delay bound($D\Delta t$). In other words, the time interval for each batch of coded-frame data, between the time of the first bit entering into the sender buffer and the time of last bit departing the buffer must be less than $D\Delta t$, as specified in section 2, i.e.

$$(j-1)\Delta t + \frac{\bar{s}(j-J)}{r(j)}\Delta t - b_{j-J}\Delta t \le D\Delta t \quad (10)$$

where the new symbols are defined as follows.

J: this value is an integer less than D and represent the queueing length at time $(j-1)\Delta t$ in the sender buffer(see Fig. 2).

$\bar{s}(j-J)$: the first data size which will be transmitted at $(j-1)\Delta t$ and the remainder of $s(j-J)$ which was not transmitted before $(j-1)\Delta t$.

$b_{j-J}\Delta t$: this value represent the instantaneous time when the first bit of has begun entering into the sender buffer.

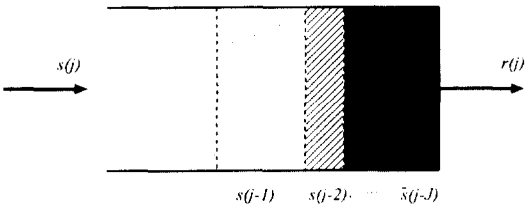


Fig. 2 Queueing structure in the sender buffer.

The departing time of $\bar{s}(j-J)$ may be the transmission starting time of $s(j-J+1)$. As soon as the transmission of $\bar{s}(j-J)$ is terminated, in order to transmit $s(j-J+1)$ continuously at the same rate, $r(j)$, the following inequality must be simultaneously satisfied as follows.

$$(j-1)\Delta t + \frac{\bar{s}(j-J)}{r(j)}\Delta t + \frac{s(j-J+1)}{r(j)} - b_{j-J+1}\Delta t \le D\Delta t \quad (11)$$

By extending this procedure to p frame data, a lower inequality for $r(j)$ can be formally defined as

$$r^{L1}(j,0) \equiv \frac{\bar{s}(j-J)}{D + b_{j-J} - j + 1} \quad (12)$$

$$r^{L1}(j,p) \equiv \frac{\bar{s}(j-J) + \sum_{i=1}^{p} s(j-J+i)}{D + b_{j-J+p} - j + 1},\ 1 \le p \le J-1 \quad (13)$$

$$r^{L1}(j,p) \equiv \frac{B_s(j-1) + \sum_{i=J}^{p} \hat{s}(j-J+i)}{D + b_{j-J+p} - j + 1},\ J \le p \quad (14)$$

where if p is equal to, or greater than J, $\hat{s}(j-J+i)$ is a value which is estimated by using the characteristics of pseudo-periodic patterns in the MPEG-coded video sequence. The coded-data size estimation method is presented at subsection 3.4, in detail.

3.3 Constraint on the Sender Buffer Size

In deciding the transmission rate, it is required that the sender buffer overflow and underflow are avoided. For these purposes, we describe the constraints imposed by sender buffer size. (1) can be recursively written as

$$B_s(j) = B_s(j-1) + \hat{s}(j) - r(j),\ for\ j > 0 \quad (15)$$

Since $0 \le B_s(j) \le B_s^{size}$ must be satisfied for all j, the following inequalities can be written by

$$B_s(j-1) + \hat{s}(j) - B_s^{size} \le r(j) \le B_s(j-1) + \hat{s}(j) \quad (16)$$

To obtain an identical transmission rate for the incoming h(t intervals, the following inequalities should be simultaneously satisfied.

$$B_s(j-1) + \sum_{k=0}^{h-1} \hat{s}(j+k) - B_s^{size} \le h \cdot r(j) \le B_s(j-1)$$

$$+\sum_{k=0}^{h-1}\hat{s}(j+k),\ h\geq 1 \tag{17}$$

For convenience of explanation, from (17), let us define the upper bounds and the second lower bounds of $r(j)$, which are a function of lookahead interval h, respectively.

$$r^{L2}(j,h)\equiv\max\left\{\frac{B_s(j-1)+\sum_{k=0}^{h-1}\hat{s}(j+k)-B_s^{size}}{h},0\right\} \tag{18}$$

$$r^{U1}(j,h)\equiv\frac{B_s(j-1)+\sum_{k=0}^{h-1}\hat{s}(j+k)}{h} \tag{19}$$

where $\{\hat{s}(j+k),\ \mathrm{k}=0,1,\cdots\}$ are estimated values by using the proposed method described at next subsection, in detail. It is noted that as h increases, the effect of the sender buffer size disappears gradually. In the case that B_s^{size} is not sufficiently large, $r^{L2}(j,h)$ tends to be more dominant than $r^{L1}(j,h)$ as h increases. But, in some applications with a sufficiently large sender buffer size, $r^{L1}(j,h)$ is always greater than $r^{L2}(j,h)$ for all p and h.

3.4 Simple Coded-frame Size Estimation

In this paper, it is assumed that $\{s(j+k),\ \mathrm{k}\geq 0\}$ are not known at time $(j-1)\Delta t$. Accordingly, to efficiently predict the coded-frame sizes for future frames, k-step linear predictor is used[13]. That is, $s(j+k)$ is predicted by using a linear combination of the previous values of $s(j)$ and denoted by. $\hat{s}(j+k)$. Thus, L-th order linear predictor has the form:

$$\hat{s}(j+k)=\sum_{l=0}^{L-1}w(l)\cdot\hat{s}(j-l) \tag{20}$$

where $w(l)$, $l=0,1,\cdots,L-1$, are the linear prediction filtering coefficients. The optimal linear predictor in the mean square sense is one that minimizes the mean square error. The $\{w(l)\}_{l=0,1,\cdots,L-1}$ is found by adaptively solving the *Wiener-Hopf* equations. Since I, P, and B frame types have different statistical characteristics, we separate them and predict $s(j+k)$ based on the each frame type.

3.5 Adaptive $r(j)$ decision algorithm

Fig. 3 shows one typical example for rate bounds described in the above subsections. First, the data which waits for transmission within the sender buffer have to satisfy the delay constraints as well as the receiver buffer constraints (Step 1 and Step 2). Second, the data which will arrive at the sender buffer does not have to give rise to overflow or underflow (Step 3 and Step 4). In this order, a general purpose $r(j)$ decision rule is designed by using (8), (12), (13), (17) and (18).

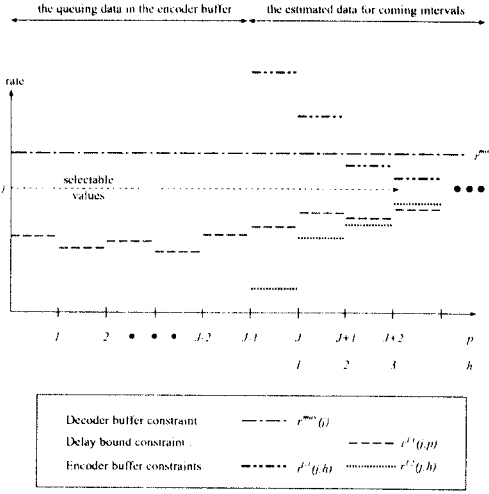


Fig. 3 typical examples for rate bounds.

Step 1: If there exists p satisfying $r^{\max}(j)<r^{L1}(j,p)$ as p increases from 0 up to $J-1$, let us denote

$$P=p,\ r^{\min}(j)=\max\left\{r^{L1}(j,p),\ p=0,\cdots,p-1\right\}$$

$$\text{else } r^{\min}(j)=\max\left\{r^{L1}(j,p),\ p=0,\cdots,J-1\right\}$$

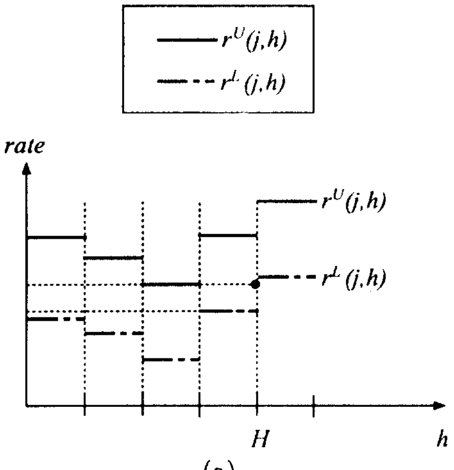
and go to Step 2.

Step 2: Let us define $r^{U}(j,h)\equiv\min\left\{r^{U1}(j,h),r^{\max}(j)\right\}$,

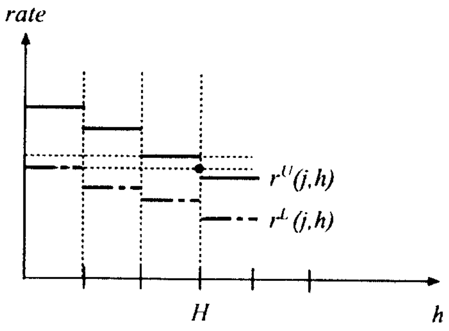
$$r^L(j, h) \equiv \max\left\{ r^{L1}(j, J+h-1),\ r^{L2}(j, h),\ r^{\min}(j) \right\},$$

If $r^L(j, 1) > r^U(j, 1)$, $r(j) = r^{\min}(j)$ and stop.
else go to Step 3.
Step 3: Find the largest integer $H(\le GOPsize)$ such that $\max_{1\le h\le H} r^L(j, h) \le \min_{1\le h\le H} r^U(j, h)$.
If $H = GOPsize$, go to Step 4.
else
if $\max_{1\le h\le H} r^L(j, h) > r^U(j+H+1)$, $r(j) = \max_{1\le h\le H}(j, h)$
else $r(j) = \min_{1\le h\le H} r^U(j, h)$
and stop.
Step 4: $r(j-1) \le \max_{1\le h\le GOPsize} r^L(j, h)$, $r(j) = \max_{1\le h\le GOPsize} r^L(j, h)$
else if $\max_{1\le h\le GOPsize} r^L(j, h) \le r(j-1) \le \min_{1\le h\le GOPsize} r^U(j, h)$, $r(j) = r(j-1)$
else $r(j) = \min_{1\le h\le GOPsize} r^U(j, h)$
and stop.

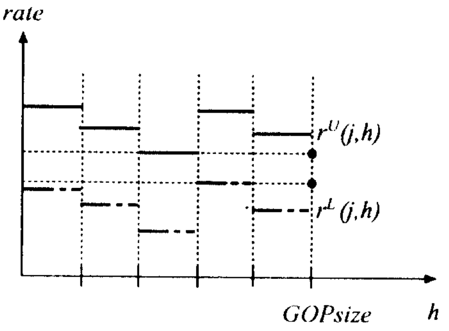
In Step 1 and Step 2, if $r^{\max}(j) < r^{L1}(j, 0)$, the receiver buffer will overflow and then some data will be lost. On the other hand, the sender buffer overflow and underflow are incurred, mainly, due to the inaccurate estimated data size or the obrupt scene change. Sender buffer overflow lead to data loss, while sender buffer underflow gives rise to a bursty or some dummy data transmission (Fig. 4).



(a)



(b)



(c)

Fig. 4 Rate decision bounds for Step3 and 4.

Ⅳ. Experimental Results and Performance Evaluations

The trace used in our experiments is a "Star-wars" and the GOP pattern is given by "IBBPBBPBBPBB". For more details about the encoder parameters, refer to [14].

4.1 Smoothing Loss Rates and Buffer Size

This paper determines the sender buffer size, and the receiver buffer size, B_r^{size} and the receiver buffer size, B_r^{size}, as flows,

$$B_r^{size} = \alpha_s \cdot \max_{\forall j}\left\{ \sum_{i=j-D+1}^{j} s(j) \right\}$$

$$B_r^{size} = \alpha_r \cdot \max_{\forall j}\left\{ \sum_{i=j-D+1}^{j} s(j) \right\}$$

where α_s and α_r represent the multiplicative factors, which are introduced to investigate the smoothing effects that arise when the sender buffer size and the receiver buffer size change.

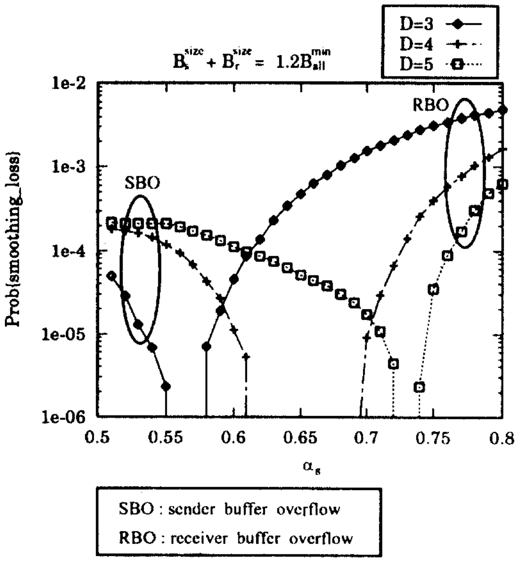
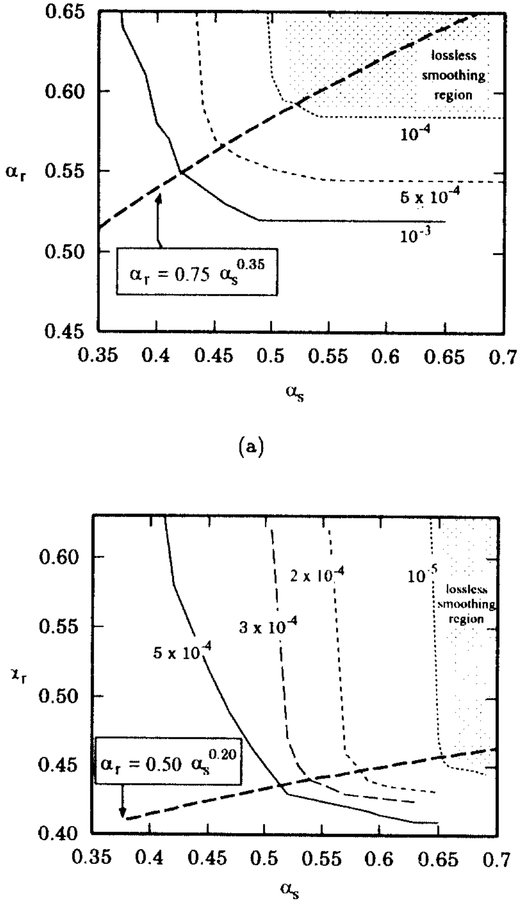


Fig. 5 Smoothing loss rates for different delay bounds : $\alpha_s + \alpha_r = 1.2$.

Simulation experiments have been performed using the proposed method for various buffer sizes. Fig. 5 shows the loss probabilities incurred by the sender buffer overflow or the receiver buffer overflow for different pairs of (α_s, α_r), with fixed at $B_r^{size} + B_r^{size} = 1.2 B_{all}^{min}$. The small values of α_s gives mainly rise to overflow at the sender buffer, while the large values of α_r to overflow at the receiver buffer. From these results, it is observed that loss probabilities can be minimized, if overall buffer sizes are slightly larger than the left term of (5) and the appropriate pairs (α_s, α_r) are selected in $\alpha_s \leq 1$ and $\alpha_r \leq 1$. Contour plots of the loss characteristics are obtained from computer experiments as shown in Fig. 6. Points along the different curves are those pairs of (α_s, α_r) which results in a fixed percentage of data overflowing the sender buffer or the receiver buffer. Since these figures are representing the probability of buffer overflow as the z-axis coming out of the page, it is shown that contour lines decreases steeply in height as α_s and α_r increase, simultaneously. The steepest decreasing lines determine the optimal buffer size relationships that minimize loss probabilities, for a fixed overall buffer size. And it is shown that the larger the delay bound is, the larger the sender buffer size is required, relative to the receiver buffer size.



(a)

(b)

Fig. 6 Contour plots of the loss charactersitics and the steepest decreasing lines : (a) $D = 3$. (b) $D = 5$.

From the above observations, it is confirmed that the proposed algorithm can be used in smoothing and transmitting for VBR-coded MPEG services, in cases that overall end systems have extremely small buffers. On the other hand, in this paper, smoothing loss is mainly caused by the inaccurate coded-data size estimation due to abrupt data size changes, such as fast moving frames, scene changes etc. Hence it is expected that if these informations are provided from the encoding side a priori, both loss probabilities and buffer sizes can be significantly reduced.

4.2 The Characteristics of Smoothed Traffic

In this subsection, the performance characteristics of smoothed traffic are investigated as functions of buffer sizes and the delay bound. The comparative smoothing schemes are listed as

(i) *SLD*: This is a sliding window smoothing scheme without adopting any coded-frame size estimation method for the incoming frames. For more details, see [3, 4]

(ii) *PROP_S*: The proposed $r(j)$ selection rule is used and this scheme adopts a simple repetition of one GOP interval for the coded-frame size estimation. That is, an estimating value for the coded-frame size of the incoming frame is simply replaced as the coded-frame size generated by the most recently encoded frame with the same coding type.

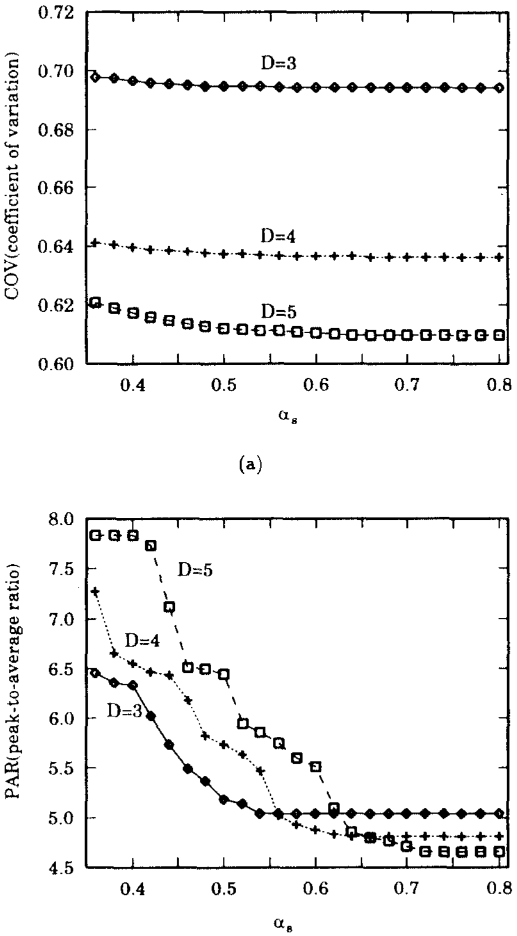
(iii) *PROP_M*: The proposed $r(j)$ selection rule and the proposed coded-data size estimation method described in section 3.4 are used.

(iv) *PROP_I*: The proposed $r(j)$ selection rule is used and, in this scheme, we assume that all coded-frame sizes are known a priori.

4.2.1 The effects of buffer sizes

- The coefficient of variation(COV)

The coefficient of variation is introduced as a measure to investigate and evaluate the temporal variations of the transmission rate[2, 12, 4]. Fig. 7 (a) shows the characteristics of COV as a function of α_s, where is determined by the steepest decreasing lines that are approximated in Fig. 6. The COV of the smoothed traffic decreases slightly as the buffer sizes increase, particularly, in $D=5$. But, beyond some critical buffer size, the performance is independent of the buffer size, because the delay bound becomes the main limitation on the ability to determine $r(j)$.



(a)

(b)

Fig. 6 Characteristics of *COV* and *PAR* for buffer sizes: for a given α_s, α_r is obtained by the steepest decreasing lines which are approximated in Fig.6.

- The peak-to-average ratio(PAR)

As shown in Fig. 7(b), the PAR is steeply decaying for the changes of buffer sizes. It is shown that as the sender and receiver buffer sizes increase, the PAR tends to decrease and get closer to a constant value which is independent of the buffer sizes. It is noted that the PAR is significantly dependent on the buffer sizes in regions that overall buffer sizes are extremely small ($\alpha_s < 0.6$). Because the burstiness of the transmission rate in these regions is significantly affected by the abrupt coded-frame size changes such as scene changes.

4.2.2 The effects of delay bounds

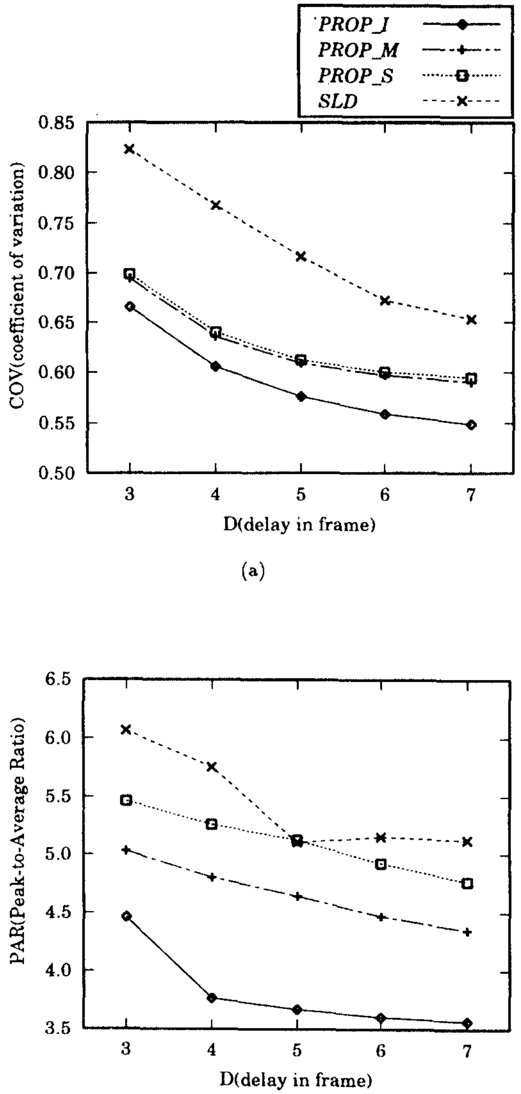
For experiments of the proposed scheme i.e. (ii)-(iv), we use $\alpha_s = 0.8$, $\alpha_r = 0.65$ and for the sliding-window scheme $\alpha_s = 1.0$, $\alpha_r = 1.0$, respectively.

- The coefficient of variation

Fig. 8(a) shows the COV performances of each smoothing schemes as a function of delay bound. As expected, the temporal variations of the smoothed traffic monotonically decay as the delay bound increases. For each delay bound, the sizes of COV are in order that '*PROP_I*' < '*PROP_M*' < '*PROP_S*' < '*SLD*'. Particularly, it is noted that, for low delay bound regions ($D \leq 4$), performance difference between the proposed estimation and the ideal one is small, but as the delay bound increases these differences become large since the estimated coded-frame sizes of incoming frames may not be relatively exact, compared to low delay bound regions.

- The peak-to-average ratio

The superiority of '*PROP_M*' over '*PROP_S*' is more obvious for the PAR-metric in Fig. 8(b), since '*PROP_M*' estimates the incoming data sizes more exactly than '*PROP_S*'. Even with $D = 3$, '*PROP_M*' reduces the peak rate by 51.7% over the original video traces. For comparison, '*PROP_I*' reduces the peak rate by 57.2%, while '*PROP_S*' the peak rate by 47.7%. Accordingly, for practical visual conferencing communications (in this paper, $D \leq 7$), it can be concluded that the proposed algorithm is remarkably effective in removing the short-term and mid-term burstiness in the underlying video stream.



(a)

(b)

Fig. 8 Charactersitics of COV and PAR for delay bounds: $\alpha_s = 0.80$, $\alpha_r = 0.65$ are used for the proposed scheme and $\alpha_s = 1.0$, $\alpha_r = 1.0$ for the sliding-window scheme.

4.3 The Performances Comparison for Different Service Strategies

4.3.1 The piece-wise CBR renegotiation performances

Fig. 9 shows a conceptual multiplexing overview for piece-wise CBR services. It is assumed that if $r(j)$ is calculated at time $(j-1)\Delta t$ and a rate increase or decrease would happen, renegotiations are initiated. Simply, we assume that a renegotiation process takes no delays. There is always a possibility that a renegotiation for a new bandwidth request can fail due to the likelihood that the aggregate cell arrival rate exceeds the service rate of the transmission link. In this paper, we estimate the renegotiation failure probability by using Chernoff's approximation and Bahadur-Rao's refinement [15-18].

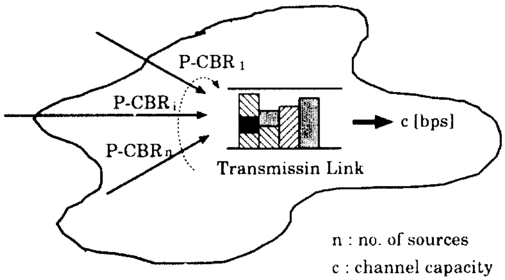
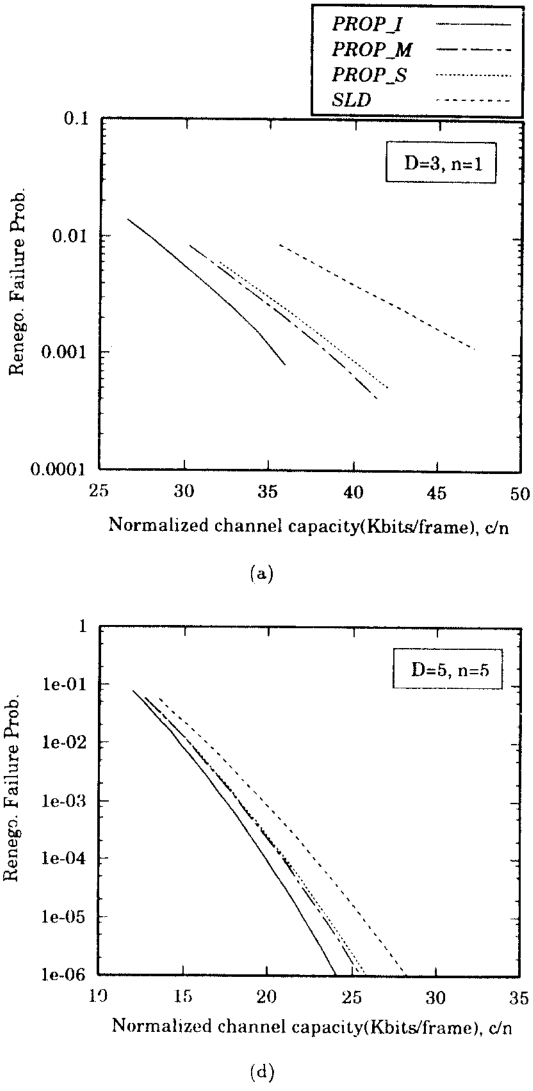


Fig. 9 Conceptual multiplexing overview for piece-wise CBR services.

Fig. 10 shows the results of renegotiation failure probabilities in terms of c(channel capacity) and n (number of multiplexed sources). These graphs show that increasing of the delay bound reduce the renegotiation failure probabilities. It is also known that as the number of sources and channel capacity increase, the temporal smoothing effects disappear due to statistical multiplexing (spatial averaging) in networks.



(a)

(d)

Fig. 10 Renegotiation failure probabilities for piece-wise CBR services: (a) $D=3$, $n=1$. (b) $D=5$, $n=5$, where $\alpha_s=0.80$, $\alpha_r=0.65$ are used for the proposed scheme and $\alpha_s=1.0$, $\alpha_r=1.0$ for the sliding-window scheme.

4.3.2 The effective bandwidth

Fig. 11 shows a discrete-time queue model with constant service rate in ATM networks. The large deviations effective bandwidths (simply called "effective bandwidth", or also "equivalent bandwidth") estimate statistically the network throughput required to transmit the video service under a cell loss probabilities(CLP) with switch buffer(B). For more details refer to [15-18]. By simplifying the results of [17], for multiplexed transmission rates with $r(j)$ $\{j=1,\cdots,N\}$, the effective bandwidth(C_{eff}) is computed as

$$C_{eff}=\frac{\log\sum_{i=1}^{N}\exp\{r(j)\cdot\theta/N\}}{\theta}$$

where $\theta=-\frac{\log(CLP)}{B}$. B is a switch buffer size and CLP is a tolerable cell loss rate as depicted in Fig. 11.

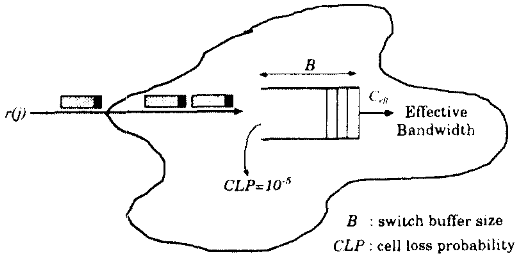


Fig. 11 Conceptual overview for effiective bandwidth allocations.

If switch buffer size is small($B=1$Kbits, $CLP=10^{-5}$), large bandwidths are required for each scheme. In this case, the graphs of effective bandwidths are similar to Fig. 8(b). The intervals with a lot of cells have a great influence in the loss probability although they are not frequent. Because, in the case that buffer sizes are small, short-term burstiness of transmission rate is dominant in deciding the effective bandwidth. Even with $D=3$, the effective bandwidth drops by 16.4%, 25.5%, 29.8% and 38.4% for 'SLD', '$PROP\_S$', '$PROP\_M$' and '$PROP\_I$', respectively, compared to unsmoothed original traffic.

On the other hand, for a large buffer size($B=100$ Kbits, CLP$=10^{-5}$), long term burstiness of smoothed traffic appears. Increasing of switch buffer size dramatically reduces the network bandwidth requirements for transmitting the video stream, since the buffer sizes allocated in the networks links can absorb easily short term bursts. Accordingly, the graph pattern is very similar to that of COV. For comparison, unsmoothed original traffic requires the transmission of 20.0 Kbits per frame interval.

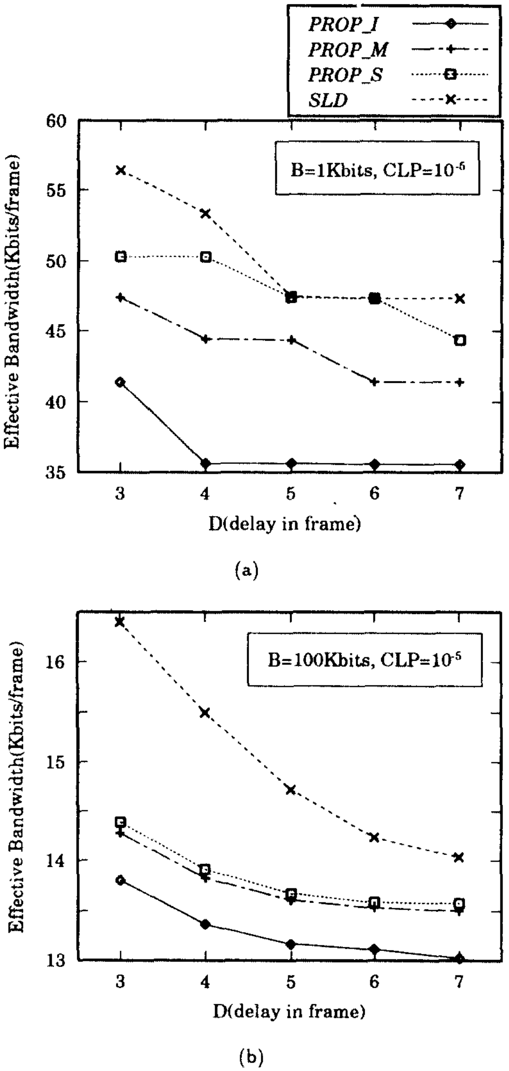


Fig. 12 Effective bandwidths for delay bounds: (a) $B=1Kbits$. (b) $B=100Kbits$, where $\alpha_s=0.80$, $\alpha_r=0.65$ are used for the proposed scheme and $\alpha_s=0.80$, $\alpha_r=0.65$ are used for the proposed scheme and $\alpha_s=1.0$, $\alpha_r=1.0$ for the sliding-window scheme.

Ⅴ. Conclusion

For the purpose of transmitting efficiently a VBR MPEG-coded video data and alleviating network management load, an adaptive source traffic smoothing method is proposed. Different from the conventional methods, the proposed scheme is based on the physical constraints imposed not only by the delay bound, but also by the sender buffer size and the receiver buffer size. By using these constraints, the proposed algorithm is designed in such a way as to smooth maximally the transmission rate, while controlling so that the buffer overflow and underflow may be avoided. Through computer experiments, it is shown that the proposed method is effective in reducing short-term and mid-term burstiness of the transmission rate. Particularly, by developing more exact coded-data size estimation method, it is shown that the proposed method has reduced significantly the burstiness as well as the temporal variation of the transmission rate.

From the practical implementation point of view, it is expected that the proposed method can be effectively used in live video applications that the overall buffer sizes are small, or asymmetrically placed. Furthermore, the proposed scheme can be applied for an important set of emerging multimedia applications, such as live videocasts of course lectures or television news that lie between two extremes of interactive and pre-coded video services.

References

1. A.R.Reibman and B. G. Haskell, "Constraints on Variable Bit Rate Video for ATM Networks," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.2, no.4, pp.361-371, Dec. 1992.
2. S.S.Lam, S.Chow, and D.K.Y.Yau, "An Algorithm for Lossless Smoothing of MPEG Video," *in Proc. of ACM/SIGCOMM*, pp.281-293, Aug. 1994.
3. W.Ding, "Joint Encoder and Channel Rate Control of VBR video over ATM Networks," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.7, No.2, April 1997.
4. K.Joseph and D.Reininger, "Source Traffic Smoothing and ATM Network Interfaces for VBR MPEG Video Encoders," *IEEE International Conference on Communication* (Seatle, U.S.A., June 1995), pp. 1761-1767.
5. S.Dixt and P.Skelly, "Video Traffic Smoothing and ATM Multiplexer Performance," *IEEE GLOBECOM'91* (Arizona, U.S.A., Dec. 1991), pp.239-243.
6. J.Lauderdale and D.H.K.Tsang, "A New Technique for Transmission of Pre-Encoded MPEG VBR Video Using CBR Service," *IEEE International Conference on Communication* (Texas, U.S.A., June 1996), pp.1416-1420.
7. J.Ni, T.Yang and D.H.K.Tsang, "CBR Transportation of VBR MPEG-2 Video Traffic for Video-On-Demand in ATM Networks," *IEEE International Conference on Communication* (Texas, U.S.A., June 1996), pp.1391-1395.
8. J. M.McManus and K.W.Ross, "Video on Demand over ATM:Constant-rate Transmission and Transport," in *Proc. IEEE INFOCOM*, pp.1357-1362, March, 1996.
9. M. Grossglauser, S.Keshav, and D.Tse, "RCBR:A Simple and Efficient Service for Multiple Timescale Traffic," in *Proc. ACM SIGCOMM*, pp.219-230, Aug./Sept. 1995.
10. MPEG-2 Test Model5;*Document ISO/IEC JTC1/SC29/WG11/93-400*;Test Model Editing Committee, Apr.1993.
11. ITU-T SG15 AVC, *Draft Recommendation H.222.0*, June 1994.
12. N.Ohta, *Packet Video:Modeling and Signal Processing*, Aretech House, 1994.
13. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1991.
14. Rose. O, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems," *University of Wuerzburg, Institute*

of Computer Science Research Report Series. Report No. 101. February 1995.

15. A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, Jones and Bartlett Publishers, 1992.
16. F. P. Kelly, "Effective Bandwidths of Multi-class Queues," *Queueing Systems*, Vol. 9, No. 1, pp. 5-16, 1991.
17. R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High-speed Networks," *IEEE Jour. on Selected Areas in Communications*, Vol. 9, No. 7, pp. 968-981, 1991.
18. G. Kesidis, J.Walrand, and C.S.Chang, "Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources," *IEEE/ACM Trans. On Networking*, 1(4), pp. 424-428, Aug. 1993.

김 진 수(Jin-soo Kim) 정회원

1967년 9월 9일생

1991년 2월:경북대학교 전자공학과 졸업(공학사)

1993년 2월:한국과학기술원(KAIST) 전기 및 전자공학과 졸업(공학석사)

1993년 3월~현재:한국과학기술원 전기 및 전자공학과 박사과정 재학중

※주관심분야:ATM Network Adaptation, Traffic Smoothing and Multiplexing

김 재 균(Jae-kyoon Kim) 종신회원

1938년 9월 17일생

1962년 2월:한국항공대학 응용전자과 졸업(공학사)

1967년 2월:서울대학교 전자공학과 졸업(공학석사)

1971년 8월:미국 남가주대학교 전자공학과 졸업(공학박사)

1972년 4월~1973년 3월:미국 우주과학연구소(NASA) GSFC연구원

1993년 9월~현재:한국과학기술원(KAIST) 멀티미디어 통신 공동 연구센터장

1973년 4월~현재:한국과학기술원 교수

※주관심분야:Information Theory, Video Coding, Visual Communication