DELIVERY OF PRE-STORED VIDEOS FOR MOBILE CLIENTS WITH INSUFFICIENT PLAYBACK BUFFER AND BOUNDED DELAY GUARANTEE

Chu-Chuan Lee * and Pao-Chi Chang

ABSTRACT

A deterministic service model assisted by a sufficiently large playback buffer space can provide bounded delay guarantees for video packets and simplify network resource management. However, many popular mobile terminals do not have sufficient memory capacity for deterministic video services since the embedded memory is limited and needs to be shared by numerous software programs and masses of personal data. This paper improves the traditional deterministic modeling approach for delivering pre-stored videos to mobile clients with QoS guarantees. The limitation of playback buffer space, the network delay jitter, the processing load of resource management, and the QoS guarantee are considered in the proposed mechanism. Some traffic smoothing operations are integrated into the proposed mechanism for reducing the playback buffer demand and data rate variation. This paper further proposes a smart video frame skip algorithm, originating at the sender for preventing possible overflow problems due to insufficient playback buffer space. The algorithm can determine the most suitable temporal range for skipping frames and prevent arbitrary discarding from inappropriate video frames such as I-frames on the client side. Simulation results reveal that the proposed mechanism can effectively remedy situations of insufficient playback buffer space while still maintaining the advantages of deterministic services.

Key Words: mobile terminal, resource management, ATM, deterministic service model.

I. INTRODUCTION

With improved capabilities of network infrastructure and video compression techniques, the availability of rich multimedia applications such as live video news and streaming contents to users are dramatically boosted (Burleson *et al*., 2002; Schmitz, 2002; Wei *et al*., 2002). Compressed Variable Bit Rate (VBR) video, one of the major components of most multimedia contents, is extremely vulnerable to packet delay and loss (Koenen, 1999). However, the burstiness of VBR video traffic makes it difficult to determine the amount of resource required. If a video sender requests more bandwidth than it actually uses, the overall network utilization drops. Conversely, packet loss and delay will be significant if the video sender directly reserves bandwidth according to the average encoded rate.

Two transmission models, namely the stochastic service model (Frost and Melamed, 1994) and the deterministic service model (Recker *et al*., 2003), are commonly used for video applications. A deterministic service model is one in which a QoS aware network, assisted by deterministic traffic modeling, can provide bounded delay and loss free guarantees to packets. The major difference between the two service models is that the former utilizes statistical properties to achieve high bandwidth utilization, while the latter utilizes deterministic bandwidth reservation to obtain the QoS guarantee and the advantage of being parameterized easily.

The literature is full of studies in this area. Esaki

*Corresponding author. (Tel: 886-3-4245661; Fax: 886-3-4244888; Email: chuchu@cht.com.tw)

C. C. Lee is with the Chunghwa Telecommunication Laboratories, Taoyuan, Taiwan 326, R.O.C.

P. C. Chang is with the Department of Communication Engineering, National Central University, Taoyuan, Taiwan 320, R.O.C.

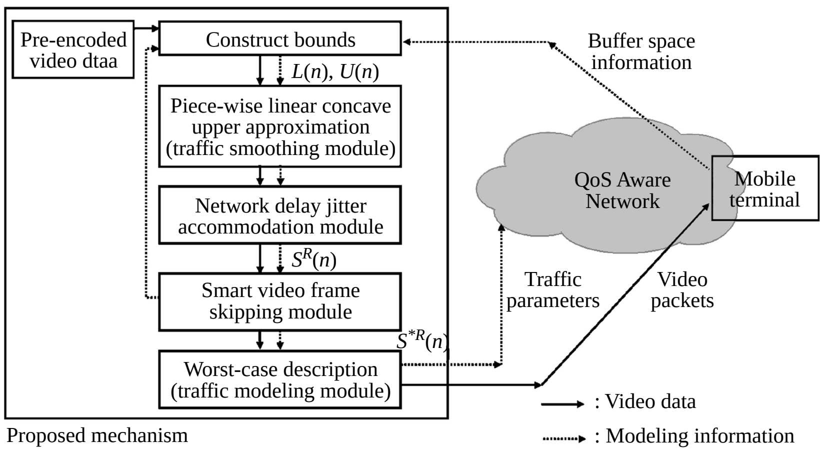

Fig. 1 Architecture of the proposed framework

(1992) considered a call admission control solution in ATM networks for real-time multimedia applications. Chen *et al.* (2003) analyzed the performance of dynamic resource management with finite buffers in cellular networks. Cheng *et al.* (1999) utilized a neural fuzzy technique to provide the QoS guarantees for multimedia high-speed networks. Mokhtar *et al*. (2003) used the theory of effective bandwidth to obtain a new deterministic traffic model. The model is derived from a worst-case analysis of the traffic on a single-node basis. Grossglauser *et al.* (1997) proposed a Renegotiated Constant Bit Rate (RCBR) mechanism. Using RCBR, the description of VBR traffic, which is named *transmission schedule*, comprises a series of bandwidth requests. Chang and Lin (2001) developed a filtering theory for deterministic traffic regulation and service guarantee. A traffic regulator was implemented by a linear time-invariant filter. Wrege *et al*. (1996) proposed an Empirical Envelope Modeling (EEM) scheme, which belongs to the deterministic service model, to characterize video traffic. A deterministic service model must have the worst-case characterization of the video source to provide an absolute upper bound, which is described as follows. If the actual traffic of a connection is given by a function A such that $A[\tau, \tau + t]$ denotes the traffic arrivals in the time interval $[\tau, \tau + t]$, an upper bound on A can be given by a traffic constraint function A^* if for all times $\tau \geq 0$ and all interval lengths $t \geq 0$, the following holds,

$$A[\tau, \tau + t] \leq A^*(t). \quad (1)$$

The EEM scheme first uses the worst-case description to get an empirical-envelope function E^* from the cumulative transmitted data, and then generates a transmission schedule from E^* using the piece-wise linear upper approximation. Consequently, a sufficiently large playback buffer space at the client and a large network bandwidth in the initial transmission stage are required due to the worst-case description. Unfortunately, the memory capacity of most popular mobile terminals may be insufficient because the embedded memory is limited and needs to be shared by numerous software programs and masses of personal data (Tomimori and Nakamoto, 2002; Hartwig *et al*., 2000). In addition, Rexford *et al.* (1997) proposed a Pre-Recorded Video Smoothing (PRVS) strategy to utilize effectively the client buffer space and reduce the rate variability for pre-stored videos. However, PRVS alternatively increases and decreases the transmission rate by a wide margin to solve the problem of insufficient decoder buffer capacity. Events that increase transmission rate in half the delivery duration may complicate the resource management of a cellular network and cause unpredicted packet loss or latency, particularly when the available bandwidth is nearly exhausted.

This paper improves conventional deterministic traffic modeling schemes for delivering pre-stored video contents to mobile clients with QoS guarantees. Herein the limitation of playback buffer space, the network delay jitter, the processing load of resource management, and the QoS guarantee are considered in the proposed mechanism. The architecture of the proposed framework is shown in Fig. 1. This study assumes that an asynchronous transfer mode (ATM) or a resource reservation protocol (RSVP) is used in the QoS aware network. In the context of QoS-guaranteed network communication, an efficient traffic modeling mechanism to quantify the video traffic characteristics is very important. The proposed framework is implemented on the video sender's side and the major responsibility of the mobile client is to provide feedbacks of available buffer space to video sender. When a mobile client requests a video service, the proposed mechanism first constructs the bounds for generating a transmission schedule. Based on the available playback buffer space of clients and the characteristics of video frames, this study utilizes an adequate traffic smoothing operation, i.e., the piece-wise linear concave approximation, to reduce playback buffer demand and rate variation before executing the process of deterministic traffic modeling. During the above approximation work, the accommodation of network delay jitter is considered. However, the results of piece-wise linear concave approximation may occasionally cause undesired playback buffer overflow if the playback buffer space is insufficient. To solve the problem, we further propose a smart video frame skip algorithm on the sender side. Skipping an inappropriate frame not only reduces the received picture quality, but also complicates the deterministic traffic modeling operation. The proposed algorithm can determine the most suitable temporal range for skipping frames and prevent arbitrary discarding from inappropriate video frames such as I-frames on the client side. Given the regulated video bitstream, this work finally uses the worst-case description scenario to generate a series of monotonically decreasing bandwidth demands, which can simplify network resource management and obtain QoS guarantees.

The rest of this paper is structured as follows. In Section II, the improved deterministic traffic modeling mechanism and the proposed smart video frame skip algorithm are described in detail. In Section III, performances of the proposed mechanism are evaluated through simulations. Finally, section IV concludes this paper.

II. TRAFFIC MODELING WITH INSUFFICIENT PLAYBACK BUFFER

In this section, a video sequence with N frames is considered and a discrete-time model with the unit of frame time is used, where the encoded-bits of frame i are denoted as f_i for $i = 1, 2, \cdots, N$. This work assumes that the available playback buffer space is given from the client in the service request stage. The detailed operations are described below.

1. Construct Bounds

At the beginning of the proposed traffic modeling work, an estimated bound $L(n)$ for the cumulative consumed data of decoder is constructed and expressed at the video sender side,

$$L(n) = \begin{cases} 0, & 0 \le n \le D \\ \sum_{i=1}^{n-D} f_i, & D+1 \le n \le D+N \end{cases} . \tag{2}$$

PROCEDURE Transmission_schedule ( ) {

$k = 1, J_1 = 0;$

WHILE $(N - J_k > 0)$ {

$\rho_k = \max_{J_k < \alpha \le N} \left\{ \frac{L(\alpha) - L(J_k)}{\alpha - J_k} \right\};$

$\sigma_k = L(J_k) - \rho_k \times J_k;$

$P_k = \max\{\alpha \mid \rho_k \cdot \alpha + \sigma_k = L(\alpha)\};$

OUTPUT $(\rho_k, \sigma_k, J_k, P_k);$

$J_{k+1} = P_k;$

$k = k + 1;$ }

}

Fig. 2 The algorithm of piece-wise linear concave upper approximation to bound $L(n)$

where D is the pre-loading time limited by the playback buffer space and $n = 0$ is the time instant that the first video packet arrives at the decoder. From the client viewpoint, a feasible transmission schedule must always be larger than or equal to $L(n)$ for avoiding buffer underflow. In other words, $L(n)$ is the lower bound for any possible transmission schedule. Moreover, for preventing possible buffer overflow due to insufficient playback buffer space, this work further adds an upper bound into the process of deterministic traffic modeling. The upper bound $U(n)$ is then defined as

$$U(n) = \min\{L(n) + B, L(D + N)\}, \tag{3}$$

where B represents the available playback buffer space of mobile clients.

2. Execute Linear Concave Upper Approximation

Rather than traditional deterministic traffic modeling approaches, this study utilizes the piece-wise linear concave upper approximation to bound $L(n)$ before the worst-case description work is executed. The main reason is that the linear concave upper approximation implies a traffic smoothing effect, which can effectively decrease the rate variation of video data and reduce the playback buffer demand of worst-case representation. The detailed algorithm is presented in Fig. 2. The k-th linear segment, which bounds $L(n)$ in the interval $[J_k, P_k]$, can be found by

$$\rho_k = \max_{J_k < \alpha \le N} \{ \frac{L(\alpha) - L(J_k)}{\alpha - J_k} \} \tag{4}$$

$$\sigma_k = L(J_k) - \rho_k \times J_k, \tag{5}$$

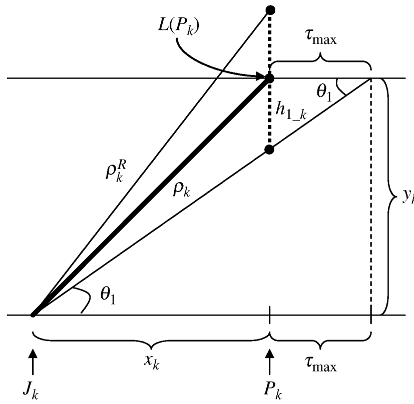

Fig. 3 Jitter accommodation at the k-th linear segment with ρ_k

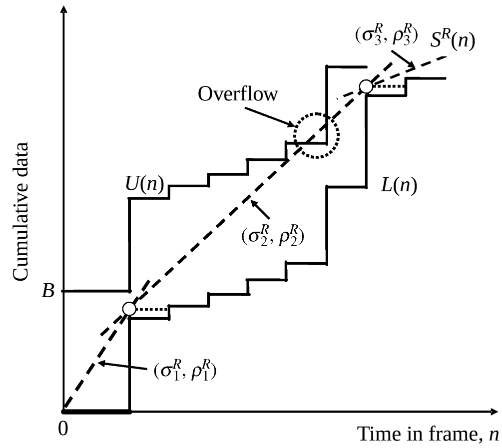

Fig. 4 Buffer overflow due to insufficient buffer space B

where ρ_k is the minimum requested token rate (i.e., the slope of k-th linear segment) that does not cause buffer underflow, and σ_k is the k-th corresponding token depth. Besides, J_k and P_k denote the beginning and end time of the k-th linear concave upper approximation, respectively.

3. Accommodate Delay Jitter

Although this study targets the deterministic services model, in which a QoS aware network assisted by deterministic traffic modeling, can provide bounded delay guarantees to packets, the network delay jitter still may affect the received picture quality. Herein we assume that the maximum network delay jitter τ_{max} is specified at the call setup stage. Given the τ_{max} and the k-th linear segment with ρ_k, we recalculate the requested token rate in $[J_k, P_k]$ by

$$\rho_k^R = \frac{L(P_k) + h_{1\_k} - L(J_k)}{P_k - J_k}, \tag{6}$$

where

$$P_k = \max\{\beta | \rho_k \cdot \beta + \sigma_k = L(\beta)\}, \quad \text{for } J_k < \beta \le N \tag{7}$$

$$h_{1\_k} = \tau_{max} \cdot \tan \angle \theta_1 \tag{8}$$

$$\theta_1 = \tan^{-1}\left(\frac{y_k}{(P_k - J_k) + \tau_{max}}\right) \tag{9}$$

$$y_k = \rho_k \cdot (P_k - J_k). \tag{10}$$

After the above computations, we updated $L(n)$ by

$$L(n) = \max\{L(P_k) + h_{1\_k}, L(n)\}, \text{ for } n \ge P_k. \tag{11}$$

The relationship among ρ_k, ρ_k^R, τ_{max} and h_{1_k} is plotted in Fig. 3. The proposed mechanism deals with the delay jitter problem at the end time of each linear segment. The main reason is that transmitting video data using ρ_k implies a pre-loading operation for those video frames in $[J_k, P_k]$, which can effectively reduce the impact of delay jitter. However, the pre-loading effect always stops at the end time of each linear segment. In other words, the probability that buffer underflow occurs in the neighborhood of P_k due to jitter is higher than that of other temporal ranges in $[J_k, P_k]$, and a corresponding strategy is required at the end time of each linear segment.

The above procedures are repeated and a raw transmission schedule $S^R(n)$ that consists of multiple piecewise linear segments is then generated. These linear segments can be expressed as a set of traffic parameter pairs $\{(\sigma_k^R, \rho_k^R) | k = 1, 2, \cdots, M^R\}$, where M^R represents the maximum number of parameter pairs. The relationship between $S^R(n)$ and its corresponding traffic parameter pairs can be further described by

$$S^R(n) = \min_k(\rho_k^R \cdot n \cdot T_f + \sigma_k^R), \tag{12}$$

where n is the frame number and the frame time T_f is the reciprocal of frame rate.

When the playback buffer space is insufficient, we observe that $S^R(n)$ may cause playback buffer overflow, as shown in Fig. 4. In such a situation, many traditional strategies may alternatively increase and decrease the transmission rate by a wide margin to cope with the problem. However, any midproblem request to increase the used bandwidth may complicate resource management and cause unpredicted packet loss or latency, particularly for practical cellular networks. Since the dynamic video source rate control is not suitable for pre-stored videos, this work develops a Smart video Frame Skipping (SFS) algorithm to solve the above problem while still maintaining the advantages of deterministic services.

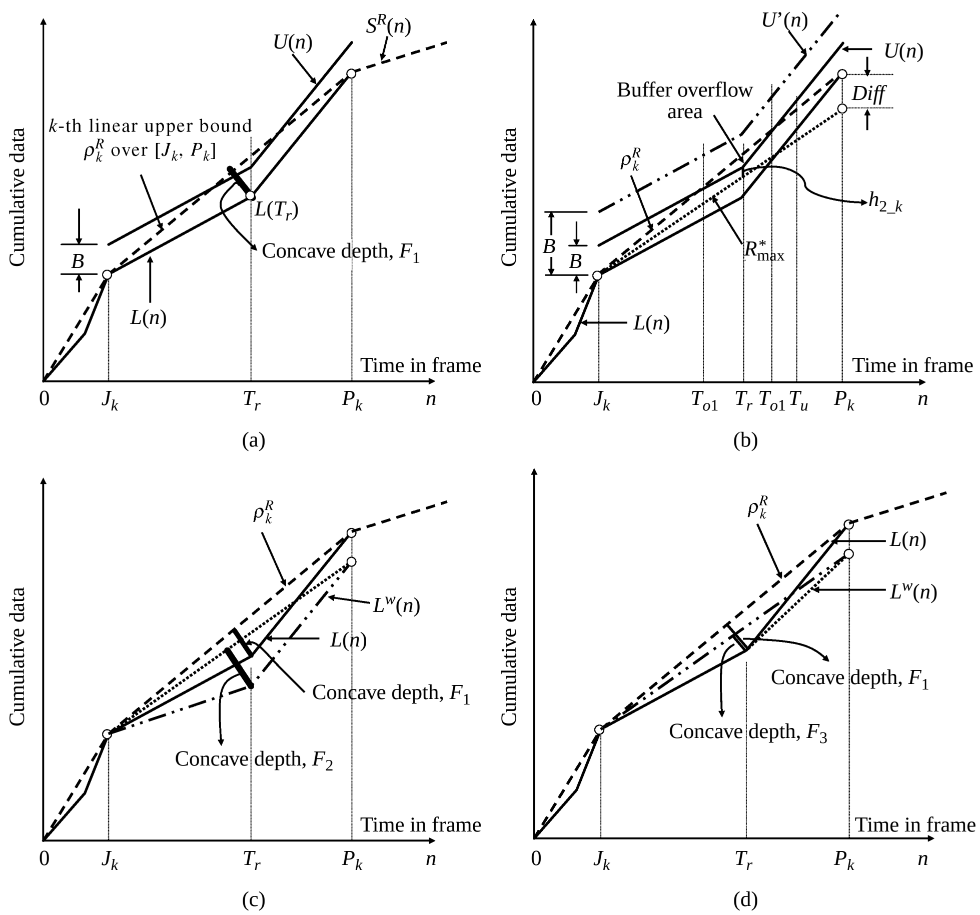

Fig. 5 Illustrations of smart video frame skip: (a) simplified $L(n)$ & $U(n)$ with related parameters; (b) the region of buffer overflow and underflow; (c) the case of skipping frame in $[J_k, T_r]$; (d) the case of skipping frame in $[T_r, P_k]$

4. Smart Video Frame Skipping

For explanation, both $L(n)$ and $U(n)$ are redrawn by simplified piece-wise linear segments, as shown in Fig. 5(a). Here the token rate ρ_k^R is used in the interval $[J_k, P_k]$ of $L(n)$. In Fig. 5(a), we define the *concave depth* F_1 as the minimum distance between the point $L(T_r)$ and the k-th linear upper bound. A large concave curvature is associated with a large concave depth, which may cause a high possibility of buffer overflow in the case of insufficient playback buffer space. Therefore, whenever a possible playback buffer overflow is detected during the traffic modeling process, the target of SFS is to reduce the concave depth efficiently.

From Fig. 5(b), if a client has a sufficiently large playback buffer space B' that exceeds B, no buffer overflow occurs in $[J_k, P_k]$ and the finally required token rate ρ_k^{*R} in $[J_k, P_k]$ is determined to be ρ_k^R directly. However, when the playback buffer space is limited to B bits, which is insufficient in this case, a buffer overflow in $[T_{o1}, T_{o2}]$ is detected by means of Eq. (3). Meanwhile, SFS is activated and a new required token rate, without causing playback buffer overflow in $[J_k, P_k]$, must be recalculated by

$$R_{\max} = \min_{J_k < \alpha \leq P_k} \{\frac{U(\alpha) - L(J_k)}{\alpha - J_k}\}, \tag{13}$$

where the time instant T_r that determines the value of $R_{\max}$ can be further expressed by

$$T_r = \{\tau | U(\tau) = R_{\max} \cdot (\tau - J_k) + L(J_k)\}, \quad \text{for } J_k < \tau \leq P_k. \tag{14}$$

Considering the network delay jitter $-\tau_{\max}$ that may cause buffer overflow and using concepts similar to those of Fig. 3, we refine Eq. (13) as

$$R^*_{\max} = \max_{J_k < \alpha \leq P_k} \{\frac{U(\alpha) - L(J_k)}{\alpha - J_k}\}, \tag{15}$$

where

$$h_{2\_k} = \tau_{\max} \cdot \tan \angle\, \theta_2 \tag{16}$$

$$\theta_2 = \tan^{-1}\left(\frac{y_k}{T_r - J_k}\right) \tag{17}$$

$$y_k = R_{\max} \cdot (T_r - J_k). \tag{18}$$

Although using $R^*_{\max}$ can prevent buffer overflow, there still exists a buffer underflow in $[T_r, P_k]$. The main reason is that $R^*_{\max}$ is less than ρ^R_k which is the minimum rate without causing buffer underflow in $[J_k, P_k]$. Now, the first time instant T_u that causes the buffer underflow due to $R^*_{\max}$ can be formulated by

$$T_u = \min_{J_k < \tau \le P_k} \{ \tau \left| \frac{L(\tau) - L(J_k)}{\tau - J_k} \ge R^*_{\max} \right. \} . \tag{19}$$

To solve the buffer overflow problem while avoiding unnecessary underflow, some video frames must be skipped in $[J_k, P_k]$ to eliminate the difference *Diff* between the cumulative arrived data using $R^*_{\max}$ and the lower bound $L(P_k)$. Notably, the selection of skipped frames is very important. An inappropriate selection not only degrades the received picture quality but also worsens the above problem. The relationship among $R^*_{\max}$, ρ^R_k, T_r, T_u and *Diff* is presented in Fig. 5(b).

To identify a suitable temporal range for skipping video frames in $[J_k, P_k]$, SFS first considers one of the possible temporal ranges, i.e., $[J_k, T_r]$. After skipping video frames in $[J_k, T_r]$, a new lower bound $L^w(n)$ that has the concave depth F_2 in $[J_k, P_k]$ is regenerated, as shown in Fig. 5(c). In $[J_k, T_r]$, the slope of $L^w(n)$ is smaller than that of $L(n)$ because of frame skipping, but the slope of $L^w(n)$ in $[T_r, P_k]$ is the same as that of $L(n)$. From Fig. 5(c), the concave depth F_2 is larger than F_1. That is, the operation of skipping frames in $[J_k, T_r]$ cannot achieve the target of reducing the concave depth but even worsens the problem. The above statements can be easily proved by simple triangle geometry. SFS then considers another possible temporal range $[T_r, P_k]$ for effectively skipping frames. Using similar derivations, it can be easily proved that the concave depth F_3 in $[J_k, P_k]$ is effectively reduced if skipped video frames are located at interval $[T_r, P_k]$, as shown in Fig. 5(d). Summarizing the above results, this study shows that the most suitable temporal range for skipping frames in $[J_k, P_k]$ is limited to the interval $[T_r, P_k]$.

From the viewpoint of temporal scalability in scalable video coding, skipping B-frames can minimize the influence on received picture quality. Therefore, some video frames with low significance are skipped in $[T_r, P_k]$ for removing the difference *Diff*. The frame skip operation is repeated until the difference *Diff* at P_k disappears. After accomplishing the frame skip operation in $[T_r, P_k]$, a finally required traffic parameter pair $(\sigma^{*R}_k, \rho^{*R}_k)$ for $[J_k, P_k]$ is

PROCEDURE smart_frame_skip ( ) {

$R_{\max} = \min_{J_k < \alpha \le P_k} \left\{ \frac{U(\alpha) - L(J_k)}{\alpha - J_k} \right\}$;

$T_r = \{ \tau \,|\, U(\tau) = R_{\max} \cdot (\tau - J_k) + L(J_k) \}$;

$R^*_{\max} = \frac{(U(T_r) - h_{2\_k}) - L(J_k)}{T_r - J_k}$;

$\sigma^*_{\min} = L(J_k) - R^*_{\max} \times J_k$;

$Diff = (\rho^R_k - R^*_{\max}) \cdot P_k \cdot T_f + (\sigma^R_k - \sigma^*_{\min})$;

WHILE $(Diff > 0)$ {

SKIP a B-frame with size D_B in $(T_r\ ,\ P_k]$;

$Diff = Diff - D_B$;

UPDATE $L(n)$ and $U(n)$; }

OUTPUT $L(n)$ and $U(n)$; }

Fig. 6 Smart frame skip algorithm

then determined and added to the mature transmission schedule $S^{*R}(n)$, where $\rho^{*R}_k = R^*_{\max}$ and $\sigma^{*R}_k = L(J_k) - R^*_{\max} \cdot J_k$. Fig. 6 details the proposed frame skip algorithm.

5. Model Regulated Traffic with Worst-Case Description

From the network perspective, the arrival data from video sender is the regulated $S^{*R}(n)$ instead of the original VBR video data. Therefore, we execute the worst-case description work for $S^{*R}(n)$ and observe an important property, i.e.

$$S^{*R}[n, n + m] \le S^{*R}(m),\ \forall n \ge 0,\ m \ge 0. \tag{20}$$

That is, the worst-case description result of $S^{*R}(n)$ is identical to $S^{*R}(n)$ itself. From the definition of (1), $S^{*R}(n)$ satisfies the requirements of deterministic service model in nature. The main reason is that those transmission rates of $S^{*R}(n)$ are monotonically decreasing.

As mentioned in Section I, a deterministic service model can provide bounded delays and loss free guarantees to video packets. However, many traditional deterministic modeling methods such as (Wrege *et al.*, 1996) do not consider the factor of playback buffer limitation and many conventional traffic smoothing schemes such as (Rexford *et al.*, 1997) alternatively increase and decrease the transmission rate by a wide margin in cases of insufficient playback buffer space. In contrast, the proposed mechanism not only considers the playback buffer limitation, but also provides a deterministic QoS guarantee to the regulated video traffic. More importantly, no in the middle request for increasing bandwidth occurs in $S^{*R}(n)$, which can effectively simplify the resource management of cellular networks. Besides, the computational complexity of the proposed mechanism is O(MN), which is less than that of EEM with O(N^2). M represents the number of piece-wise linear segments of a transmission schedule with N video frames.

6. Call Admission Test and Rate Renegotiation

In this paper, the proposed mechanism is associated with an RSVP or ATM network in which the embedded call admission control (CAC) scheme requires a worst-case bound on individual sources and the call admission test must be able to determine how a collection of sources will interact. Currently, there are rich research results about the CAC method such as (Cruz, 1991; Liebeherr *et al.*, 1996; Knightly and Zhang, 1997). By using the fact that $S^{*R}[n, n + m] \leq S^{*R}(m)$ together with other manipulations in (Knightly and Zhang, 1997), delay can be upper bounded by

$$d_{upper} = \frac{1}{L}\max_{n \geq 0}\{\sum_{j=1}^{N} S_j^{*R}(n) - n \cdot T_f \cdot L + PS\}, \quad (21)$$

where

d_{upper}	upper bound of delay (sec);
N	total number of admissible connections;
L	link bandwidth (bps);
$S_j^{*R}(n)$	transmission schedule for connection j (bits);
PS	maximum packet size.

Eq. (21) is used as a CAC test in this paper, where it can determine if a new connection can be admitted so that each packet of all admissible connections, including the new one, can be delivered to clients within the guaranteed delay bound. After passing the CAC test, the source then delivers the regulated video data and renegotiates its service rate with the network according to the given transmission schedule $S^{*R}(n)$. The operation of renegotiation consists of sending a signaling message requesting an increase or decrease in the current service rate. When the ATM network is considered, the source can reuse the resource management (RM) cell mechanism for lightweight signaling. The explicit rate (ER) field of the RM cell contains the difference between the old and new service rates. When receiving an RM cell, the ATM switch checks if the current port utilization plus the rate difference is less than the port capacity. If the request is feasible, the network allows the renegotiation. When the integrated services of the Internet using the RSVP signaling protocol are considered, a route reservation for a video flow is created and periodically updated in two stages: the source multicasts PATH messages containing traffic characteristics, then RESV messages containing resource reservation requests are forwarded from the receiver along a reverse path (Grossglauser *et al.*, 1997).

Table 1 Statistical properties of two encoded video traces

	Frame size		Bit rate	
	Mean (Kbits)	Peak (Kbits)	Mean (Kbps)	Peak (Kbps)
Jurassic Park I	10.08	54.50	302.5	1634.8
The firm	6.70	42.14	201.1	1264.3

III. SIMULATION RESULTS AND DISCUSSION

In the following simulation scenarios, this study implements the proposed mechanism on the video sender side with a multiplexer whose line rate is DS-3 (44.736 Mbps). This work assumes that all admissible video connections are homogeneous with identical priority and the end-to-end delay is nonzero but fixed by the assistance of QoS aware networks. We use the MPEG-4 codec to generate three compressed test video sequences, a standard raw video sequence "*Foreman*", a part of the movie "*Jurassic Park I*", and a part of the movie "*The Firm*". All test video sequences have the following properties: 1) A GOP consists of 15 frames with frame rate of 30 fps and its pattern is set to I-B-B-P format. 2) Considering the characteristics of 3G video applications, all sequences are 3 minutes in length. 3) Both QCIF and CIF formats are applied to the simulations. Since "*Foreman*" is only 300 frames in length, it is repeated cyclically to form the video sequence. Besides, the properties of "*Jurassic Park I*" with 300 kbps and "*The Firm*" with 200 kbps are shown in Table 1 where the *frame size* represents the bit count of an encoded video frame and the *bit rate* represents the encoded bit rate of video sequence.

This study first evaluates the minimum required playback buffer demand of the proposed method with the traditional EEM method (Wrege *et al.*, 1996) for achieving bounded delay and loss free guarantees, as shown in Fig. 7. Both video sequences, "*Jurassic Park I*" and "*The Firm*", are used with CIF format and various target encoding bit rates. Although current 3G mobile terminals can only support the QCIF format video, video services with CIF format for mobile users are expected in the near future. We observe that the proposed method requires less playback buffer space than that of EEM, regardless of the

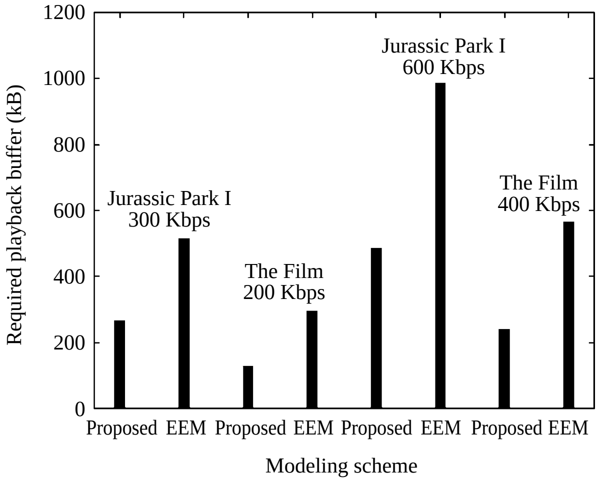

Fig. 7 Minimum required playback buffer space for achieving the bounded delay guarantee

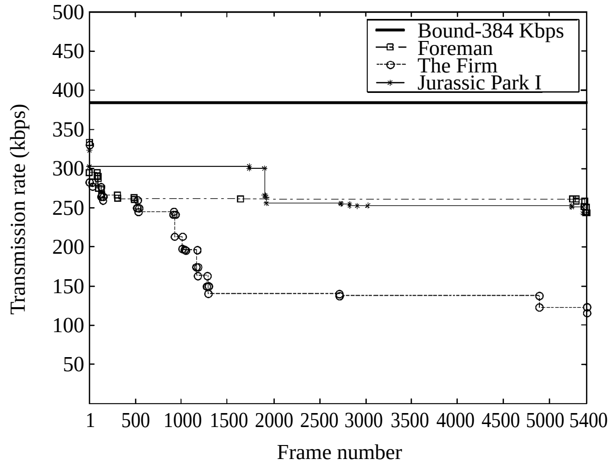

Fig. 9 Transmission schedules generated by the proposed method for various video sequences

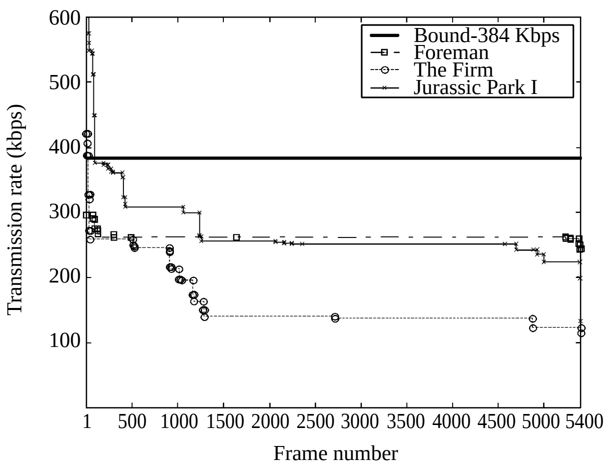

Fig. 8 Transmission schedules generated by EEM for various video sequences

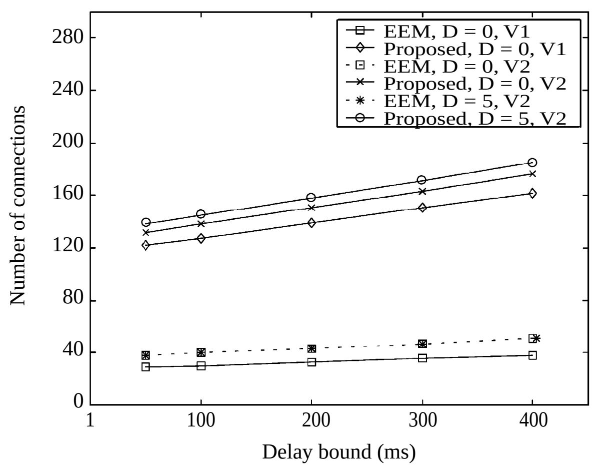

Fig. 10 Maximum number of admissible connections as a function of delay bound d

use of various video sequences and encoded bitrates in this simulation. When a mobile client can only provide limited playback buffer space for video contents, the proposed approach can effectively reduce the demand of playback buffer space while maintaining the bounded delay guarantee.

On the assumption that the playback buffer space is infinite, this work compares the initial bandwidth requirement of EEM with that of the proposed method. Three video sequences are encoded with QCIF format for expanding the discussion scope. From the results in Figs. 8 and 9, we observe that the initial bandwidth requirement generated by the proposed method is significantly smaller than that generated by EEM. The curves of Fig. 8 show that the initial bandwidth requirement generated by EEM may exceed the outdoor upper bound of 384 Kbps in current 3G networks, particularly for complex video sequences such as "*Jurassic Park I*". In contrast, the proposed approach can effectively reduce the initial bandwidth requirement while maintaining the advantages of the deterministic service model. Note that both approaches generate a series of bandwidth requests that decrease monotonically.

Moreover, this study utilizes Eq. (21) to evaluate the link utilization of the proposed method with EEM. Fig. 10 plots the maximum number of admissible connections as a function of delay bound d. "*Jurassic Park I*" and "*The Firm*" with CIF format are denoted as "V1" and "V2", respectively. To examine the impact of pre-loading on the link utilization, simulations are executed with two pre-loading times, i.e. D = 5 and D = 0. From Fig. 10, the proposed method can significantly increase the link utilization compared with the traditional EEM method, regardless of the settings of various delay

Table 2 PSNR performances of the proposed mechanism with various video sources and playback buffer spaces

	Video source	Buffer space	Received PSNR	Hurt I-frame	Hurt P-frame	Hurt B-frame
PRVS	The Firm	300KB	30.16 dB	15	163	354
		500KB	31.02 dB	5	70	147
	Video-A	200KB	31.28 dB	11	133	273
		400KB	32.02 dB	2	33	72
Proposed without SFS	The Firm	300KB	29.90 dB	0	0	573
		500KB	30.81 dB	0	0	290
	Video-A	200kB	31.25 dB	0	0	331
		400KB	31.81 dB	0	0	148
Proposde with SFS	The Firm	300KB	31.24 dB	0	0	236
		500KB	31.65 dB	0	0	89
	Video-A	200KB	32.10 dB	0	0	150
		400KB	32.41 dB	0	0	45

bounds and pre-loading times in this simulation. Moreover, considering the impact of pre-loading, we observe that two curves of EEM with D = 5 and D = 0 are coincident with each other. This means that the link utilization of EEM is independent of the pre-loading operation in our simulation cases. The main reason is that EEM only uses the worst-case description method to generate the transmission schedule. The pre-loading operation for reducing the huge initial bandwidth demand of EEM is invalid if a complex video frame section appears in the middle of a video sequence. In contrast, an appropriate pre-loading operation in the proposed method can effectively improve the link utilization.

Finally, Table 2 evaluates the received picture quality of the proposed mechanism using two methods. The first method is the traditional PRVS scheme (Rexford *et al.*, 1997) and the second method is the proposed mechanism without the SFS function. In addition to the test video "*The Firm*", another video sequence *Video-A*, which was constructed by concatenating sixteen commonly used segments is applied here. The sixteen segments are *Table*, *Container*, *Mobile*, *Salesman*, *Coastguard*, *Paris*, *Singer*, *Claire*, *Akiyo*, *Stefan*, *Children*, *Dancer*, *Silent*, *Foreman*, *Tempete* and *Hall monitor*. The target encoding bitrate of both video sequences is set to 400 Kbps. From the viewpoint of video coding characteristics, all frames belonging to the same GOP are hurt due to error propagation in the decoding process if an I-frame is lost. Moreover, all succeeding frames belonging to the same GOP are also hurt if a P-frame is lost. This phenomenon causes significant degradation of received picture quality. This work thus defines the *hurt frame* as a video frame that is lost or hurt in quality because of error propagation. The scenario assumes that all admissible connections are simultaneously served and a midjob request for increasing bandwidth during the playback is rejected if the available bandwidth of the network is nearly exhausted. From Table 2, the amount of *hurt frames* using the proposed solution is observed to be less than that using PRVS in each case of the simulation scenario. Note that no I-frame or P-frame is dropped when the proposed mechanism is used. Moreover, the received PSNR of the proposed mechanism exceeds that of PRVS up to 1.1dB in the case of using "*The Firm*" with 300KB buffer limitation. The PSNR improvement of the proposed framework increases when the available playback buffer space decreases. The main reason is that the proposed mechanism can actively select suitable frames to be skipped if necessary, instead of passively and arbitrarily skipping frames congested or delayed by network or decoder.

Table 2 also evaluates the performance of the proposed SFS scheme. When the SFS function of the proposed mechanism is disabled, the selection of skipped B-frames for the calculated token rate $R^*_{\max}$ in $[J_k, P_k]$ is arbitrary. Under the same proposed mechanism, the number of skipped B-frames without SFS is significantly larger than that with SFS. The main reason is that the operation of arbitrarily skipping B-frames in $[J_k, T_r]$ cannot achieve the target of reducing the concave depth in $[J_k, P_k]$, as described in Section II-D.

IV. CONCLUSIONS

Regarding the delivery of pre-stored video data, the proposed mechanism has the same advantages as traditional smoothing and deterministic modeling schemes when the decoder buffer space of mobile clients is sufficiently large. More importantly, the proposed mechanism maintains sustained playback and outperforms traditional schemes in cases of insufficient decoder buffer space. Herein, the limitation of playback buffer space, the network delay jitter, the processing load of resource management, and the QoS guarantee are considered in the proposed mechanism.

REFERENCES

Burleson, W., Thampuran, S., and Ramaswamy, N., 2002, "Multimedia Systems: Enabling Computer Engineering Education," *Proceedings of the 32nd Annual Frontiers in Education*, Vol. 1, pp. T2F1-T2F6.

Chang, C. S., and Lin, Y. H., 2001, "A General Framework for Deterministic Service Guarantees in Telecommunication Networks with Variable Length Packets," *IEEE Transactions on Automatic Control*, Vol. 46, Issue 2, pp. 210-221.

Chen, W. Y., Wu Chen, J. L., and Shin, H. Y., 2003, "Performance Analysis of Dynamic Resource Allocation with Finite Buffers in Cellular Networks," *Proceedings of the 11th IEEE International Conference on Networking*, pp. 641-646.

Cheng, R. G., Chang, C. J., and Lin, L. F., 1999, "A QoS-Provisioning Neural Fuzzy Connection Admission Controller for Multimedia High-Speed Networks," *IEEE/ACM Transactions on Networking*, Vol. 7, Issue 1, pp. 111-121.

Cruz, R., 1991, "A Calculus for Network Delay, Part I: Network Elements in Isolation," *IEEE Transactions on Information. Theory*, Vol. 37, Issue 1, pp. 114-131.

Esaki, H., 1992, "Call Admission Control Method in ATM Networks," *Proceedings of the IEEE International Conference on Communications*, Vol. 3, pp. 1628-1633.

Frost, V., and Melamed, B., 1994, "Traffic Modeling for Telecommunications Networks," *IEEE Communications Magazine*, Vol. 32, Issue 3, pp. 70-81.

Grossglauser, M., Keshav, S., and Tse, D. N. C., 1997, "RCBR: A Simple and Efficient Service for Multiple Time-Scale Traffic," *IEEE Transactions on Networking*, Vol. 5, Issue 6, pp. 741-755.

Hartwig, S., Luck, M., Aaltonen, J., Serafat, R., and Theimer, W., 2000, "Mobile Multimedia-Challenges and Opportunities," *IEEE Transactions on Consumer Electronics*, Vol. 46, Issue 4, pp. 1167-1178.

Knightly, E. W., and Zhang, H., 1997, "D-BIND: An Accurate Traffic Model for Providing QoS Guarantees to VBR Traffic," *IEEE/ACM Transactions on Networking*, Vol. 5, Issue 2, pp. 219 -231.

Koenen, R., 1999, "MPEG-4 - Multimedia for Our Time," IEEE Spectrum, Vol. 36, pp. 26-33.

Liebeherr, J., Wrege, D., and Ferrari, D., 1996, "Exact Admission Control for Networks with Bounded Delay Services," *IEEE/ACM Transactions on Networking*, Vol. 4, Issue 6, pp. 885-901.

Mokhtar, H. M., Pereira, R., and Merabti, M., 2003, "An Effective Bandwidth Model for Deterministic QoS Guarantees of VBR Traffic," *Proceedings of the IEEE International Symposium on Computers and Communication*, Vol. 2, pp. 1318-1323.

Recker, S., Geisselhardt, W., and Wolff, I., 2003, "Dimensioning of Traffic Engineered Trunks for Deterministic Service Guarantees in Mobile Data Networks," *Proceedings of the IEEE International Conference on Communications*, Vol. 3, pp. 1669-1674.

Rexford, J., Sen, S., Dey, J., Feng, W., Kurose, J., Stankovic, J., and Towsley, D., 1997, "Online Smoothing of Live, Variable-Bit-Rate Video," *Proceedings of the IEEE International Workshop on Network and Operating System Support for Digital Audio and Video*, pp. 235-243.

Schmitz, P., 2002, "Multimedia Goes Corporate," *IEEE Multimedia*, Vol. 9, Issue 3, pp. 18-21.

Tomimori, H., and Nakamoto, Y., 2002, "An Efficient and Flexible Access Control Framework for Java Program in Mobile Terminals," *Proceedings of Distributed Computing Systems Workshops*, pp. 777-782.

Wei, G., Petrushin, V. A., and Gershman, A. V., 2002, "The Community of Multimedia Agents Project," *Proceedings of the IEEE International Conference on Multimedia and Expo.*, Vol. 2, pp. 289-292.

Wrege, D. E., Knightly, E. W., Zhang, H., and Liebeherr, J., 1996, "Deterministic Delay Bounds for VBR Video in Packet-Switching Networks: Fundamental Limits and Practical Trade-Offs," *IEEE/ACM Transactions on Networking*, Vol. 4, Issue. 3, pp. 352-362.

Zhang, H., and Ferrari, D., 1994, "Improving Utilization for Deterministic Service in Multimedia Communication," *Proceedings of the International Conference on Multimedia Computing and Systems*, Boston, MA, USA, pp. 295-304.

Manuscript Received: Sep. 07, 2005
Revision Received: Mar. 07, 2006
and Accepted: Mar. 30, 2006