A CROSS-LAYER APPROACH FOR MINIMUM DELAY CONTENT ADAPTIVE VIDEO STREAMING OVER VARIABLE BITRATE CHANNELS

Tanır Özçelebi[1], A. Murat Tekalp[1,2], M. Reha Civanlar[1], M. Oguz Sunay[1]

[1]Koç University, College of Engineering, 34450, Istanbul, Turkey
[2]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627
{tozcelebi, mtekalp, rcivanlar, osunay}@ku.edu.tr

ABSTRACT

A novel cross-layer scheduling and stream switching algorithm for content and channel adaptive video streaming over 1xEV-DO (CDMA-HDR), where stream switching is done according to the receiver buffer level is presented. The instantaneous transmission rate to each user is determined by a multi-objective optimized scheduler, maximizing network throughput and individual receiver buffer levels simultaneously, while the instantaneous coding rate for each user is determined according to buffer status feedback. The transmitter always switches to the most suitable pre-optimized bitstream with minimum delay and visual distortion calculated for various channel capacity values and under constant bitrate assumption. The main target of the proposed framework is to guarantee continuous playout of the transmitted content at all user devices while providing maximum system throughput, minimum overall distortion and minimum pre-roll delay. Experimental results show that decoder buffer overflows and underflows that cause pauses in the playout are prevented as opposed to the case without stream switching.

1. INTRODUCTION

Due to present developments in wireless communication and mobile device technologies, finding efficient ways of data transmission over low-bandwidth systems has now become a very hot research topic. As a result of the vast customer demand, one of the most popular applications of this research area is the streaming of video content over low capacity variable bitrate (VBR) channels, in which, abrupt channel variations leading to buffer underflows and/or overflows are quite probable, as opposed to its constant bitrate (CBR) counterpart. Since these buffer violations lead to undesired pauses during video playout at the receiving side, it is a far more difficult task to provide continuous playout for VBR channels, which constitute most of the real-life communication systems. Therefore, system specific and careful management of channel resources is required, especially for the VBR case.

Thus far, several streaming methods that try to provide continuous playout of received video have been introduced in the literature. In [1], we presented a content-adaptive delay-distortion optimization (DDO) framework that minimizes overall distortion and pre-roll delay (pre-buffer time) necessary for continuous playback at the client side, under maximum distortion and maximum buffer size constraints for CBR channels. The pre-roll delay, i.e. the initial waiting time to pre-fetch video data, can not be kept too long as it is limited by the receiving buffer size. In [1], the group-of-pictures (GoP) were defined to match shot boundaries determined by content analysis and an appropriate relevance level was assigned to each GoP according to user's choices.

The number of actions that can be taken at the receiving side for continuous playout is very limited. The adaptive media playout (AMP) framework described in [2] adaptively changes the video play-back speed at the decoder side according to the instantaneous buffer status and channel state. In that work, buffer underflows are postponed or not permitted by decreasing the playout speed of the present data stream at the decoder buffer, whenever there is a shortage in the buffer occupancy level of the decoder. Likewise, foreseen buffer overflows are prevented by increasing the video playout speed. The AMP algorithm can also be used in improving the rate-distortion (R-D) optimized video streaming performance as shown in [3]. On the other hand, the AMP algorithm alone is not enough to solve the playout interruption problem encountered in VBR channels. Especially when there are frequent packet losses in the network, there is a limit to the extra playout time that can be achieved by using AMP. Therefore, the AMP framework behaves weak against long term performance sufferings, and decoder buffer underflow or overflow occurs sooner or later. Buffer underflow causes the decoder side to re-buffer data so that a required level of buffer fullness is again achieved.

In cases of network congestion, if the encoding rate is not lowered to a proper level, the congestion gets worse. In order to maintain TCP friendliness [4], the channel resource sharing algorithm that the encoder side uses has to avoid flooding the communication network.

In this paper, a cross-layer, TCP friendly framework for streaming over *variable bitrate (VBR)* channels (i.e. CDMA-HDR, 1xEV-DO standard) considering fair slot scheduling and source coding issues simultaneously is presented. For each pre-assumed channel condition, the proposed method employs delay-distortion optimization (DDO) technique of [1] for joint optimization of encoding rate and coding parameters under maximum distortion per GoP, maximum buffer size and continuous playback constraints. After this, according to the behavior of the varying channel conditions and buffer status of each user logged on in the network, a scheduling policy and a stream switching strategy maximizing the overall quality of experience (i.e. continuous playout, maximum visual quality, minimum pre-roll delay) is given.

The paper is organized as follows: The multi-objective optimization formulation used in the CBR case is described in Section 2. The proposed scheduling algorithm is presented in Section 3 and the proposed stream switching policy is explained in Section 4. The improvements gained over the content adaptive solution using the standard H.264 [5] encoder are shown with experimental results in Section 5. Finally, discussions are made in Section 6.

2. OPTIMIZATION FORMULATION FOR THE CBR CHANNEL

In the proposed framework, the communication channel capacity is quantized into a number of throughput levels (L) to come up with different bitstreams that are optimized for the quantized average bitrates. Assume that the input video is divided into N semantically defined temporal segments manually or using one of the automatic video content analysis techniques in the literature [6]-[9]. Semantic analysis and temporal segmentation of the input video is beyond the scope of this paper. The encoding parameters of the N temporal segments are optimized for each of these L throughput levels, by using the delay-distortion optimization (DDO) framework of [1] for CBR channels.

The DDO formulation that minimizes the pre-roll delay (T_P) and overall distortion (D) for the k^{th} ($1 \le k \le L$) throughput level (B_k) is given by:

$$\min_j(T_p) = \min_j\left\{\sum_{i=1}^{N}\frac{R_i^{\,j} - B_k}{B_k}\cdot TD_i\right\} \tag{1}$$

$$\min_j(D) = \min_j\left\{\sum_{i=1}^{N} w_i \cdot D_i^{\,j} \cdot TD_i\right\} \tag{2}$$

jointly subject to

$$D_i^{\,j} \le D_i^{\max},\quad i = 1,\ldots,N$$

and

$$Buffer\_Size \ge T_p \cdot B_k - \sum_{i=1}^{n}(R_i^{\,j} - B_k)TD_i \ge 0,\quad n = 1,\ldots,N$$

Here, R_i^j, D_i^j and TD_i stand for the i^{th} temporal segment's (shot's) encoding bitrate, visual distortion and time duration, respectively. *Buffer_Size* denotes the available client buffer size.

3. SCHEDULING ALGORITHM

The recent 1xEV-DO (IS-856) system is a VBR packet switched network, in which, all channel resources are dedicated to only one user at a time (the user with the best channel conditions). SNR feedbacks received from individual users logged on in the system are used to determine the maximum transfer rate for that specific user in a given time slot. In this system, the quality of video segment received by each user can be drastically improved by employing cross-layer optimized time slot scheduling and stream switching between previously optimized bitstreams.

As shown in [10], in order to maximize the overall channel throughput in a multi-user wireless system, all the transmission power must be dedicated to one user at a time in a time-multiplexed manner. In 1xEV-DO, scheduling intervals (*Ts*) are defined to be 1.667 msec. Such a time slot scheduling method is called *opportunistic multiple access scheme*. Although this scheme is optimal in terms of channel throughput, its performance is quite poor in terms of service fairness, since a user with relatively bad channel conditions is likely to suffer from long durations of not being served.

In order to prevent unfairness, other scheduling algorithms that give priority to the users whose QoS requirements have not been met for a long time, such as the proportionally fair (PF) scheduler and exponential scheduler were introduced [11]. Another scheduling algorithm for the HDR system that offers better statistical QoS guarantees is presented in [12]. In [13], we introduced the multi-objective optimized (MOO) scheduler, where the buffer status of each user ($B_j(t)$) is continuously monitored via decoder buffer feedback and the user with the best compromise among the maximum throughput and minimum buffer occupancy is scheduled at all time slots. The overall performance of the MOO scheduler was found to be much higher than the above state-of-the-art schedulers in terms of channel throughput and fairness issues. In the proposed technique, MOO scheduler is used along with the stream switching strategy whose details are drawn in Section 4. The AMP technique described in [2] and [3] can easily be incorporated into this framework.

4. STREAM SWITCHING

Assume that the k^{th} bitstream is being transmitted to the receiver i at time t. If the channel throughput of user i experiences a long enough shortage, the decoder buffer occupancy level decreases. After the buffer level goes below a certain threshold, Th_{k_AMP}, AMP technique is used to avoid buffer underflow. In case the use of AMP does not increase the buffer level, after the buffer level goes below a second threshold, $Th_{k\_lowRate}$, the encoder side switches to one of the pre-optimized bitstreams encoded at a lower bitrate than bitstream k, in order to provide continuous playout at the decoder side. If the reverse happens and channel conditions get better for the i^{th} user and its buffer level exceeds a certain value ($Th_{k\_highRate}$), the transmitter switches to another optimized bitstream encoded at a higher average bitrate, avoiding buffer overflow.

An example switching scenario across bitstreams is demonstrated in Figure 1, where there are three different channel throughput levels (L=3). When bitstream switching happens, the transmitter starts sending the new bitstream beginning with the last I, SI or SP frame (for H.264 encoder) prior to the current frame, while the decoder side continues playout of the already available visual information from the previous bitstream. The choice of new bitstream to be transmitted is made according to the current decoder buffer status.

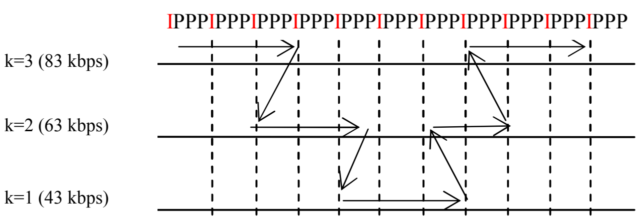

Figure 1. Stream switching across bitstreams according to buffer status.

5. EXPERIMENTAL RESULTS

We divided a 1500-frames long (60 seconds long and 25 frames per second) soccer video sequence into 12 semantically defined shots using the technique given in [6], and encoded optimally at 43 kbps, 63 kbps and 83 kbps (Figure 1) using the DDO framework of [1] for three distinct throughput values.

In order to determine the instantaneous channel SNR (channel throughput) for each user at all times, the physical layer simulations were carried out using the Advanced Design System (ADS) of Agilent Technologies. In [13], we had shown that the MOO scheduler performs better in terms of continuous playout even without the stream switching. Here we show that the results are further improved with decoder-buffer-aware stream switching. For the non-stream switching case, the second stream (k=2) whose average bitrate is 63 kbps was transmitted. The average bitrate over 32 users for the stream switched (43, 63 and 83 kbps as in Figure 2) cross-layer solution is found as 58.7 kbps. The decoder buffer occupancy graphs in Figure 3 show that, as opposed to the case when stream switching is not used, our algorithm does not allow buffer underflow for any of the 32 users in the network, hence providing continuous playout.

6. DISCUSSION

In this paper, we proposed a novel cross-layer approach for content and channel adaptive video streaming over variable bitrate (VBR) channels. The main target of the proposed algorithm is to provide continuous video playout at the receiving side under maximum buffer size and maximum distortion constraints, by considering application layer stream switching and physical layer time slot scheduling simultaneously. The experimental results show that buffer overflows and underflows can be prevented by this method as opposed to the regular streaming case, where stream switching is not employed. The proposed framework employs intelligent switching between different bitstreams optimized for different channel throughput values.

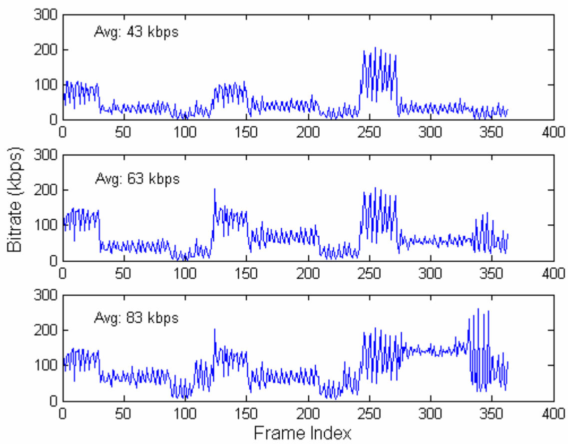

Figure 2. Bitrate versus frame index (6.25 fps) plot for the three bitstreams.

7. REFERENCES

[1] T. Özçelebi, A. M. Tekalp, M. R. Civanlar, "Optimal rate and input format control for content and context adaptive video streaming," in Proceedings of IEEE International Conference on Image Processing (ICIP), Singapore, Oct. 2004.

[2] E. Steinbach, N. Farber, and B. Girod, "Adaptive playout for low-latency video streaming," Proc. IEEE ICIP, Thessaloniki, Greece, Oct. 2001.

[3] M. Kalman, E. Steinbach, and B. Girod, "R-D optimized media streaming enhanced with adaptive media playout," Proc. IEEE ICME, Lausanne, Switzerland, Aug. 2002.

[4] M. Handley, S. Floyd, J. Padhye and J. Widmer, "TCP friendly rate control (TFRC): Protocol specification," RFC-3448, Jan. 2003.

[5] T. Wiegand, G. Sullivan and A. Luthra, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," May, 2003.

[6] A. Ekin, A. M. Tekalp and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Trans. on Image Processing, vol. 12, no. 7, pp. 796-807, June 2003.

[8] M.G. Brown et al., "Automatic content-based retrieval of broadcast news," Proc. 3rd Int'l Conf. Multimedia (ACM Multimedia 95), pp. 35-43, ACM Press, New York, 1995.

[9] V. Tovinkere and R. J.Qian, "Detecting semantic events in soccer games: Toward a complete solution," in Proc. IEEE Int. Conf. Mult. Expo (ICME), Aug. 2001.

[9] M. Bertini, A. Del Bimbo, A. Prati, R. Cucchiara, "Semantic annotation and transcoding for sport videos" in Proceedings of International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004), Lisboa, Portugal, Apr. 21-23, 2004.

[10] R. Knopp, P.A. Humblet, "Information capacity and power control in single-cell multiuser communications," in Proceedings of the IEEE ICC Conference, pp. 331-335, Seattle, USA, 1995.

[11] M. Oguz Sunay and Ali Eksim, "Fair scheduling for spectrally efficient multi-service data provisioning," Wiley's International Journal of Communication Systems, August 2004.

[12] Ferit O. Akgul and M. Oguz Sunay, "Efficient resource allocation for statistical QoS assurances in HDR based wireless packet data," to appear in the Proceedings of the IEEE PIMRC Conference, Barcelona, Spain, September 5-8, 2004.

[13] T. Özçelebi, O. Sunay, A. M. Tekalp, M. R. Civanlar, "A cross-layer optimized minimum-delay scheduling algorithm for real time video streaming to multiple users over 1xEV-DO," submitted to IEEE GlobeCom 2005.

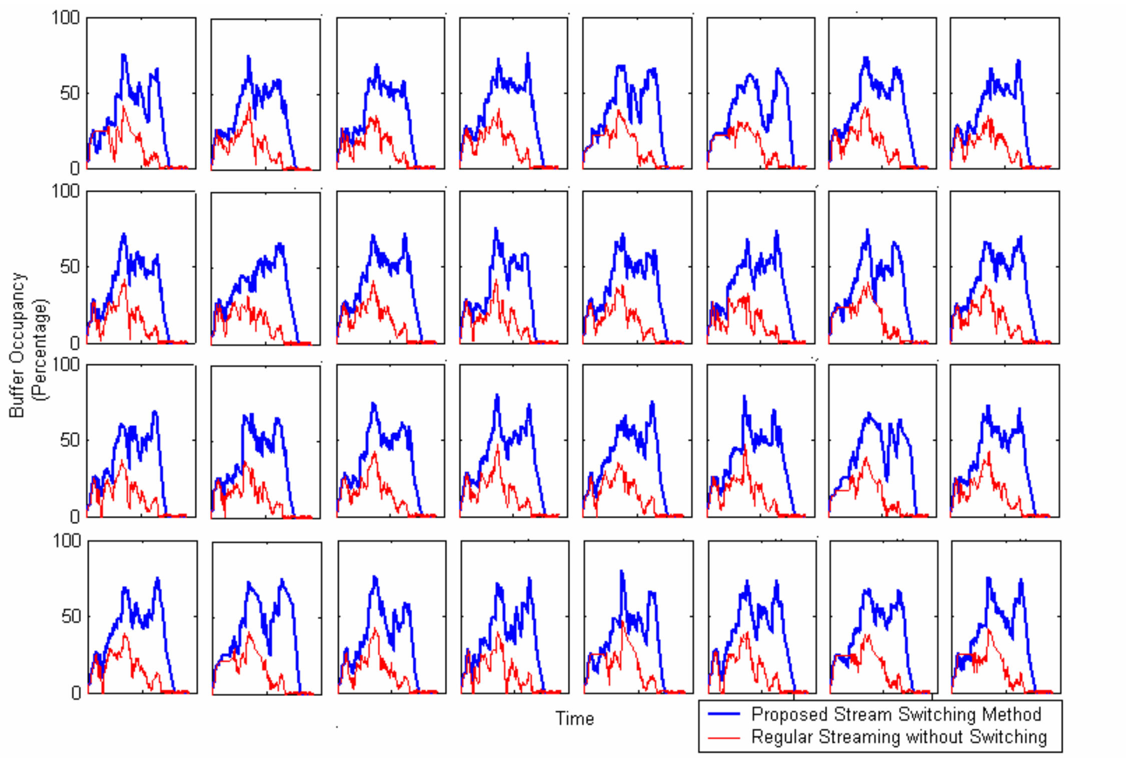

Figure 3. Buffer occupancy level change of the 32 users in time.