Minimum Delay Content Adaptive Video Streaming over Variable Bitrate Channels with a Novel Stream Switching Solution

Tanır Özçelebi, M. Reha Civanlar
College of Engineering
Koç University
İstanbul, Turkey
{ tozcelebi, rcivanlar }@ku.edu.tr

A. Murat Tekalp
Department of Electrical and Computer Engineering
University of Rochester
Rochester, New York
mtekalp@ku.edu.tr

Abstract — We present a new channel adaptive stream switching solution for variable bitrate video transmission, where the receiver buffer status is used for making switching decisions and instantaneous transmission rate is determined by means of explicit channel feedback or TCP friendly rate control. Receiver buffer status is used in selecting from a set of pre-encoded bitstreams avoiding buffer underflows and overflows. A pre-roll delay to buffer data at the receiving side is necessary in order to compensate for variations in the channel throughput and the encoding bitrate. For each stream, content dependent video coding parameters are chosen by means of dynamic programming such that the maximum overall video quality and minimum pre-roll delay are achieved for a finite number of transmission rates. Experimental results show that buffer violations are successfully avoided using the proposed stream switching framework as opposed to a regular non-adaptive streaming case.

Keywords-stream switching; multiple objective optimization; content adaptation; delay-distortion optimization

I. INTRODUCTION

Over the last decade, video streaming over low and variable bitrate (VBR) channels has become a very popular research area due to significant developments in wireless technologies. Since most of the practical transmission systems do not provide Quality-of-Service (QoS), efficient ways of video streaming that improve user satisfaction are required.

Guaranteeing continuous video playout at the receiving side considering pre-roll delay and buffer issues is still a challenging problem. There have been a limited number of attempts to solve this problem until now. The adaptive media playout (AMP) framework described in [2] relies on the adaptation of video playout speed to the instantaneous buffer occupancy level and channel conditions. In [2], the playout speed of the already available video information at the decoder buffer is adaptively changed, so that possible buffer underflows and overflows are either delayed or prevented. The use of AMP in improving the rate-distortion (R-D) optimized video streaming performance is addressed in [3]. Especially due to the bursty nature of packet losses on multi-user VBR networks, there is a certain limit to the extra playout time that can be gained by employing AMP. Therefore, the AMP framework can provide only limited protection against long time periods with insufficient throughput, and decoder buffer underflows are still possible. When this happens, the receiver should resynchronize at the next intra-coded (I) frame, and re-buffer data until a required level of buffer fullness is reached again. If the underflow occurs due to network congestion and the encoding bitrate is not lowered to an appropriate level, the congestion may get even worse harming other flows, in particular, congestion sensitive TCP flows against TCP friendliness [1].

In packet networks, congestion causes the packets to be dropped at intermediate routers and end-points. Also, if there is high traffic, queuing and scheduling delays in the network cause some packets arrive at the receiving side much later than their corresponding frames' decoding/display time making them effectively useless. This must be considered as another packet loss scenario. In addition to these, in wireless communication links, bit errors may occur as well. Error resilience techniques such as Forward Error Correction (FEC) and Automatic Repeat reQuest (ARQ), and proper error concealment methods can be employed to alleviate the effects of bit errors, but none of these techniques can guarantee error free delivery. An end result of packet losses due to any of the three reasons mentioned above may be considered as reduced throughput which has an effect on the receiver buffer occupancy.

In video streaming, user satisfaction can be increased significantly by making use of content adaptive optimization of video encoding parameters. Some of the existing streaming techniques, which consider content analysis and content adaptive encoding of video, are given in [4] and [5]. In [4], an input soccer video is divided into segments with two levels of importance. Segments with high importance level are encoded at full quality, while one still image per unimportant segment is transmitted so that the playout freezes at the client side during these unimportant shots. In our earlier work, we presented a content-aware multi-objective optimization framework that minimizes overall distortion and pre-roll delay necessary for continuous playback at the client side, under maximum distortion and maximum buffer size constraints for constant bitrate (CBR) channels [5], where the group-of-pictures (GoP) are defined to match shot boundaries determined by content

0-7803-9134-9/05/$20.00 ©2005 IEEE

Authorized licensed use limited to: ULAKBIM UASL - KOC UNIVERSITY. Downloaded on April 16,2026 at 11:14:45 UTC from IEEE Xplore. Restrictions apply.

analysis and an appropriate relevance level is assigned to each semantically defined GoP. In these techniques, instantaneous encoding rate fluctuates from GOP to GOP, hence complicating receiver buffer management for continuous playback even for CBR channels. Most common networks, however, are not CBR and network throughput variations must also be addressed.

To this effect, this paper extends the content adaptive approach presented for CBR channels in [5] to *variable bitrate (VBR)* channels based on a stream switching framework. The new method selects the most appropriate encoding bitrate among a set of predetermined rates based on receiver buffer feedback. The transmission rate is determined based on channel feedback or TCP friendly rate control (TFRC) [1]. Multi-objective optimization (MOO) method is used for joint optimization of pre-roll delay and visual quality of each pre-encoded bitsream under maximum distortion per GoP and maximum buffer size constraints; assuring continuous video playout at the decoder for a given bitrate.

The paper is organized as follows: Section II describes the multiple-objective optimization formulation. Section III presents adaptation of the MOO solution to VBR channels. Section IV demonstrates the improvements gained over the regular streaming case with experimental results. Finally, Section 5 presents the conclusions.

II. Optimization Formulation for CBR Channels

On a CBR channel, the optimal encoder parameters are determined as follows [5]. Let R_i^j, D_i^j and TD_i stand for the i^{th} temporal segment's (shot's) encoding bitrate, allowed visual distortion and time duration, respectively. The available maximum client buffer size is denoted by *Buffer_Size*. The delay-distortion optimization (DDO) formulation that minimizes the pre-roll delay (T_P) and overall visual distortion (D) of the streamed video for the k^{th} ($1 \le k \le L$) a constant throughput level (B_k) is given by:

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

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

jointly subject to

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

and

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

The best set of encoding parameters is chosen to be the one that belongs to the best compromise solution among pre-roll delay and visual quality. Detailed explanation on how to determine the best compromise (optimal) solution can be found in [5].

III. Adaptation to Channel Conditions

In the new framework, we assume that our transmission channel sustains one of L average rates between its lowest and highest throughputs for a short time period. We, then, generate different bitstreams which are optimized for each of these L average channel rates and for each video content type. That is, assuming that the input video gets segmented into N semantically defined temporal segments, $N \times L$ bitstreams must be generated by using the delay-distortion optimization (DDO) formulation outlined above and discussed in [5]. Semantic analysis and temporal segmentation of the input video can be accomplished manually or using proper analysis techniques such as those given in [6]-[9] as in [5].

In the proposed method, the actual instantaneous transmission rate is determined by channel or network conditions. In IP networks, we use TFRC to determine the encoder's transmission rate. In wireless networks, a network analyzer at the encoder side keeps checking the communication channel conditions for a specific user with period T_s, until the end of the communication session. Based on this, the encoder can determine how much data can be transferred to the receiver side at all times. As an example, such communication network awareness is possible in IS-856 CDMA/HDR [10] standard via use of channel signal-to-noise ratio (SNR) feedbacks from mobile stations, which effectively determine the instantaneous channel throughput values for a specific user subject to a maximum probability of error bound. In CDMA/HDR system, in which all the transmission power of the base station (BS) is allocated to only one user at a time in a time-multiplexed manner (opportunistic multiple access), the server side maps the SNR feedback values from each mobile station to quantized channel capacity levels. Therefore, given the applied scheduling regime, the server side knows the current channel condition of each user logged on in the system exactly at each time slot (i.e. in periods of T_s =1.667 msec).

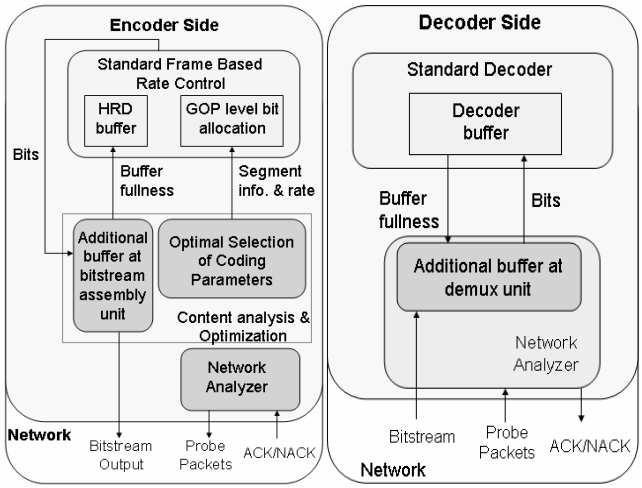
The selection of the particular bitstream to be transmitted is accomplished by monitoring the receiver buffer to guarantee continuous playback. The receiver buffer occupancy gets conveyed to the transmitter regularly. Assuming that the current stream that is being transmitted corresponds to the k^{th} channel rate,

i) When there is an important degradation in the channel throughput due to network congestion or increased wireless channel bit error rates and corresponding throughput falls, the decoder buffer occupancy level starts to fall. No action is taken until the buffer level goes below a certain thresholdvalue, Th_{k_AMP}. On the other hand, if this threshold is exceeded, Adaptive Media Playout (AMP) technique described in [2] and [3] *may be* employed in order to prevent buffer underflow. If the buffer level still continues depleting and crosses a second critical threshold, $Th_{k_lowRate}$, the encoder side understands that the channel conditions are getting even worse and switches to one of the pre-optimized bitstreams at the nearest lower rate point to k, for providing continuous playout at the client side.

Authorized licensed use limited to: ULAKBIM UASL - KOC UNIVERSITY. Downloaded on April 16,2026 at 11:14:45 UTC from IEEE Xplore. Restrictions apply.

ii) If the channel conditions get better for the specific user and buffer level goes up to a certain value ($Th_{k_highRate}$), the encoder side switches to a bitstream optimized for a higher channel throughput, hence avoiding or delaying a possible buffer overflow.

In our framework, due to the variable rate data transmission driven by TFRC or the use of SNR feedback from the channel [10] and unequal bitrate allocation across different GoP's (shots) [5], additional buffers may be required at both ends of the communication link. The required additional buffers at the encoder and the decoder are illustrated in Fig. 1. The additional encoder buffer is placed at the bitstream assembly unit, where the bitstreams coming from independently and unequally encoded GoP's (shots) are brought together. The additional decoder buffer is placed at the demultiplexer unit, where the input bitstream from the network is divided back into GoP's for proper display, since different GoP's may have different spatial and/or temporal resolutions.



(a) **(b)**

Fig. 1. Additional units needed at the encoder (a) and decoder (b) sides.

A possible switching scenario across bitstreams is demonstrated in Fig. 2, where there are three different assumed channel throughput levels ($L=3$).

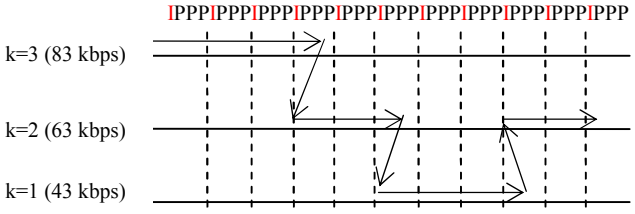


Fig. 2. Switching across bitstreams according to network conditions.

When the algorithm decides to switch to another bitstream at some point, the encoder side starts transmitting the new bitstream starting with the last SI or SP frame (for H.264 encoder [11]) before the current frame. While the encoder starts transmitting another bitstream, the client side still continues playout of the already available visual information at the decoder buffer. The choice of which bitstream to switch depends on the current buffer fullness.

IV. Experimental Results

In our experiments, we use a 60 seconds soccer video which is divided into 12 shots with different semantic properties, where content analysis (shot classification) has been made manually. These shots have been assigned semantic importance weights according to their semantic content (goal possessions, audience shots, etc.) as explained in [5]. Each shot is then encoded optimally for three constant channel throughput values at 43 kbps, 63 kbps and 83 kbps, for which the instantaneous bitrates are shown in Fig. 3.

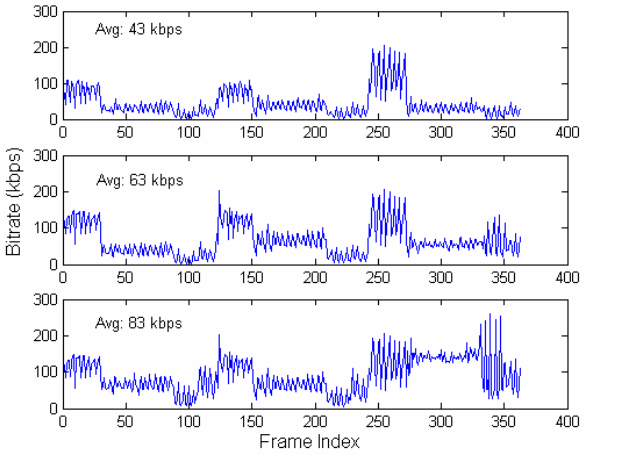


Fig. 3. Bitrate versus frame index (6.25 fps) plot for the three bitstreams.

The experiments are performed on a wireless channel simulator. We determine the instantaneous channel throughput for a variable bitrate CDMA/HDR (1xEV-DO) system at each time slot by making use of the Advanced Design System (ADS 2004A) software, the network simulation package of Agilent Technologies. The physical layer simulations have been carried out using ITU Pedestrian-A test environment in a 3-tier cellular layout and for 32 users, where the users are uniformly randomly and repetitively dropped into the cell with periods of 1 second. In this system, using the proportionally-fair (PF) scheduler [12], available channel throughput graph for a randomly selected user is found as in Fig. 4. Note that the channel throughput values observed along the duration of 54000 time slots (60 seconds) have a bursty nature due to the opportunistic access scheme used.

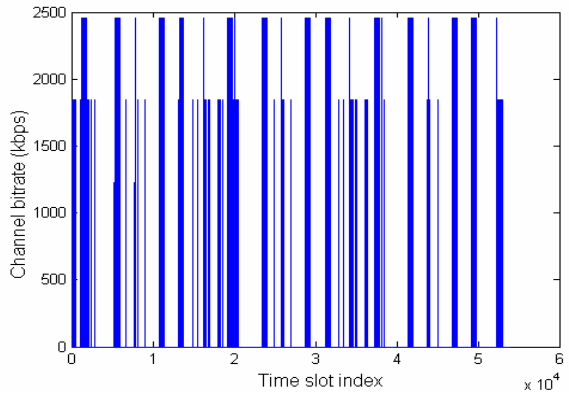


Fig. 4. Channel throughput versus time slot index (T_s=1.667 msec).

Authorized licensed use limited to: ULAKBIM UASL - KOC UNIVERSITY. Downloaded on April 16,2026 at 11:14:45 UTC from IEEE Xplore. Restrictions apply.

Fig. 5 shows the decoder buffer occupancy plots for the proposed stream switching solution and the regular streaming case where a single bitstream optimized for a CBR channel with 53 kbps throughput is used. It is observed that the buffer goes empty many times in the regular streaming case, while the playout is kept continuous in the proposed system. Here, a maximum buffer size of 1000 kbits and 2 seconds of maximum pre-roll delay are assumed. For our experiments, we set the upper threshold for the $(k-1)^{th}$ quantized buffer level to the lower threshold for the k^{th} quantized buffer level as follows:

$$Th_{k_lowRate} = Th_{(k-1)_highRate} \quad (3)$$

Since the number channel throughput quantization levels is three for our experiments, two thresholds for stream switching are used:

$$Th_1 = Th_{2_lowRate} = Th_{1_highRate} = 0.2*Buffer_Size$$

$$Th_2 = Th_{3_lowRate} = Th_{2_highRate} = 0.8*Buffer_Size$$

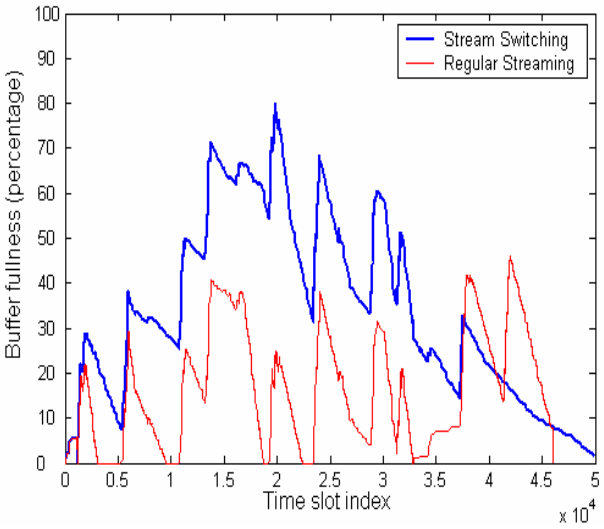


Fig. 5. Buffer occupancy graphs for regular streaming and the proposed stream-switching solution.

V. Discussion and Conclusions

In this paper, we propose a novel method for content and channel adaptive video streaming over variable bitrate (VBR) channels. The main target of the proposed algorithm is to provide continuous playout of the streamed video with very small pre-roll delay under maximum buffer size constraints, while offering the best possible viewing quality of the video content at the client side. The proposed algorithm makes use of intelligent switching between different bitstreams optimized for a set of pre-selected channel throughput values, in a channel adaptive manner. Our experimental results show that buffer overflows and underflows can be avoided by this method as opposed to the direct streaming case.

As a future extension of this work, it is possible to employ a scalable video coder, such as the SVC reference codec, and extract the bitstream at the desired instantaneous rate from a scalable video bitstream instead of using L non-scalable streams at pre-selected bitrates.

References

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

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

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

[4] S.F. Chang, D. Zhong, and R. Kumar, "Real-Time Content-Based Adaptive Streaming of Sports Video", IEEE Workshop on Content-Based Access to Video/Image Library, pp.139-146, Hawaii, December 2001.

[5] 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, October 2004.

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

[7] M.G. Brown, J. T. Foote, G. J. F. Jones, K. S. Jones and S. J. Young, "Automatic content-based retrieval of broadcast news," Proceedings of the 3rd International Conference on Multimedia (ACM Multimedia 95), pp. 35-43, ACM Press, New York, 1995.

[8] V. Tovinkere and R. J.Qian, "Detecting semantic events in soccer games: Toward a complete solution," in Proceedings of IEEE ICME, August 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, April 21-23, 2004.

[10] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth efficient high-speed wireless data service for nomadic users," IEEE Communications Magazine, vol. 38, no. 7, pp. 70-77, July 2000.

[11] 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.

[12] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of cdma-hdr: A high efficiency high data rate personal communications system," in IEEE 51st Vehicular Technology Conference, Tokyo, Japan, May 2000.

Authorized licensed use limited to: ULAKBIM UASL - KOC UNIVERSITY. Downloaded on April 16,2026 at 11:14:45 UTC from IEEE Xplore. Restrictions apply.