CONTENT-AWARE RESOURCE ALLOCATION FOR SCALABLE VIDEO TRANSMISSION TO MULTIPLE USERS OVER A WIRELESS NETWORK

Peshala V. Pahalawatta, Thrasyvoulos N. Pappas, Randall Berry and Aggelos K. Katsaggelos

Northwestern University
EECS Department
2145 Sheridan Rd, Evanston, IL, 60208

ABSTRACT

Wireless video transmission is prone to unpredictable degradations due to time-varying channel conditions. Such degradations are difficult to overcome using conventional video coding techniques. Scalable video coding offers a flexible bitstream that can be dynamically adapted to fit the prevailing channel conditions. Within a scalable video coding framework, we develop simple packet prioritization strategies, which, when combined with a reasonable error concealment scheme and a content-aware resource allocation technique, provide for robust video transmission over time-varying channels. The packet prioritization as well as the calculation of the content-aware scheduling metric can be performed offline and signaled to the wireless scheduler.

***Index Terms*—** Scalable video coding, wireless video streaming, cross-layer design

1. INTRODUCTION

Wireless video streaming is a topic of increasing interest in mobile telecommunications. The key requirement for video streaming is that high per-user data rates be achieved in the network. Recent advances in channel dependent resource allocation, which exploit multiuser diversity, have significantly improved the achievable throughputs in wireless networks. These advances have led to the development of new generations of wireless technologies such as HSDPA, and IEEE 802.16, which allow for dynamic allocation of resources among users based on fast channel feedback.

In past work, we used application-layer side information, in a multiuser video streaming framework, to improve the performance of resource allocation strategies [1]. Specifically, we considered the case when a combination of TDM and CDMA/OFDM is used to share resources among users at each transmission opportunity. Here, we extend the methods developed in [1] to a scalable video coding scheme. The methods in [1] depend on the improvement in quality of a received video sequence given that the transmitted video packets have been prioritized according to their importance. The prioritization needed to occur in real-time in order to be optimal. In this paper, we explore means by which a scalable coded bitstream can be prioritized offline without knowledge of the specific channel realization. The proper method for prioritization is not immediately clear due to the effects of temporal scalability and error concealment.

Previous work on multi-user video streaming can be found in [2] and [3]. They, however, do not specifically address the issue of scalable video encoding. In [4], temporal scalability, in the form of hierarchical Bi-prediction, and SNR scalability, in the form of progressive refinement through FGS (Fine Granularity Scalability), is considered. Comparing the emerging scalable coding extension of H.264/ MPEG4-AVC, termed SVC (Scalable Video Coding), to the AVC standard without scalable video coding, the authors show that significant improvement can be made in wireless multiuser video streaming through the use of SVC. A simple packet dropping strategy is used for buffer management and maximum throughput scheduling is used at the air interface. In this paper, we use a different packet dropping strategy than that in [4] combined with a media-aware scheduling scheme that requires minimal side information. We show that such an approach can further enhance the quality of scalable video bitstreams under a variety of channel conditions.

The rest of the paper is organized as follows. In the next section we provide a brief overview of the system and describe the packet ordering strategies that can be used in conjunction with SVC. In Sec. 3, we describe the problem formulation and the derivation of the content-aware scheduling metric. In Sec. 4, we show some simulation results that validate the proposed scheduling scheme. The final conclusions are presented in Sec. 5.

2. SYSTEM OVERVIEW

2.1. Scalable Video Coding

We consider a system similar to [1] in which a media server contains multiple pre-encoded video sequences. We assume that the sequences are encoded using the scalable extension of H.264/MPEG4-AVC. As in [4], we assume that the base

This work was supported by the Motorola Center for Seamless Communications at Northwestern University

layer is AVC compatible, that temporal scalability is achieved through hierarchical Bi-prediction [5], and that SNR scalability is achieved through fine granularity scalability layers [6]. We do not allow spatial scalability. The basic structure of such a video stream is shown in Fig. 1. Each frame in the sequence consists of a base layer encoded at low quality and one or more progressive refinement (PR) enhancement layers. The sequence is divided into multiple groups of pictures (GOPs), in which the base layer of a key picture (type I/P in Fig. 1), is either intra coded, or predictively coded using only the base layer of the key picture in the previous GOP. Temporal scalability can be achieved by dropping frames in their reverse decoding order. The key benefit of using FGS is that the bitstream may be truncated at any point within a progressive refinement (PR) layer to achieve an image quality in accordance with the number of bytes transmitted.

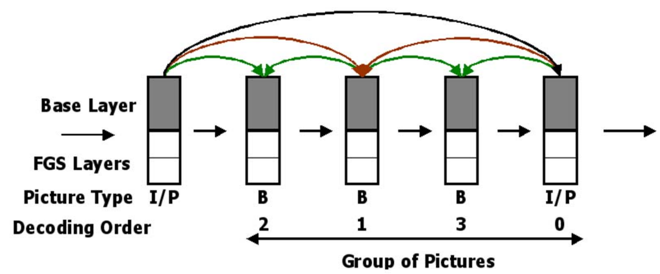

Fig. 1. Structure of scalable video stream.

2.2. Video Packet Prioritization

We assume the video sequence is packetized such that each PR layer of each frame is contained in one or more packets but no two layers are contained in one packet. Once a particular sequence is requested by a client, the packetized video is transmitted through a backbone network (assumed to be lossless and of high bandwidth) to a wireless base station servicing multiple clients. Each video packet also contains a decoding deadline by which time it must be received by the client. A scheduler located at the base station allocates resources across users based on channel feedback and a content-aware scheduling metric. The available resources may not be sufficient to transmit every video packet to every user within the decoding deadline and those packets that remain past their deadline are dropped. Therefore, a packet prioritization strategy must be used in order to ensure that the packets that are dropped are the ones with the least impact on video quality.

Scalable video coding offers a natural packet prioritization strategy that can be exploited to greatly simplify the buffer management policy at the scheduler. We have considered a few different schemes for prioritization.

- Method I - This is the same as that proposed in [4] in which the PR packets and base layer of the highest temporal level of the GOP are dropped first in that order.
- Method II - The PR packets of the highest temporal level are dropped first, and after all PR packets in the GOP have been dropped, the base layer of the highest temporal level is dropped. The main difference between Method I and Method II is that in Method I the entire picture, including the base layer, of the highest temporal level is dropped prior to dropping packets from the next highest temporal level.
- Method III - The base layer of the key picture is given the highest priority. Subsequent packets are ordered such that the next highest priority is given to the decodable packet (decodable given only the higher priority packets are received) that provides the largest reduction in distortion per bit.

Method I sacrifices temporal resolution to maintain the quality of the transmitted video frames. Method II maintains temporal resolution at the cost of image quality, by keeping the base layer packets of the entire GOP for as long as possible. Method III is the most flexible strategy but is difficult to implement and our results show little gain compared to Method II.

3. PROBLEM FORMULATION

3.1. Scheduling Metric

Given any prioritization strategy in Sec. 2.2, we define a content-aware scheduling metric as described below. Let user i's transmission queue be, $\Pi_i = \{\pi_{i,1}, \pi_{i,2}, ..., \pi_{i,M_i}\}$, where $\pi_{i,1}$ is the packet with the highest priority and M_i is the number of packets in the current GOP. Also, let $D_i[k_i]$ be the distortion given k_i packets in Π_i are transmitted and the remaining $M_i - k_i$ packets are dropped. We use the MSE metric to calculate distortion, and, in order to avoid complexity at the decoder, we assume that if an entire frame is lost, it is concealed by copying the previously decoded frame (A more complex concealment strategy may also be used with minor modifications to the calculation of the metric [1]). Now, the scheduling metric, denoted $u_i[k_i]$ can be calculated as,

$$u_i[k_i] = \frac{D_i[k_i] - D_i[k_i + 1]}{b_{i,k_i+1}}, \tag{1}$$

where b_{i,k_i+1} is the size in bits of π_{i,k_i+1}. Note that $u_i[k_i]$ is the gradient of the distortion-bits function of user i after k_i packet transmissions. We emphasize that if the base layer packet of each key frame is received, then the values of $u_i[k_i]$ can be calculated offline with a reasonable degree of accuracy at the media server. If not, the values can be recalculated at the scheduler.

3.2. Packet Fragmentation

The transport packets received from the media server are too large to be transmitted over the wireless channel. Therefore, the scheduler must account for packet fragmentation when determining the scheduling metric. For a base layer packet to be successfully decoded, the entire packet must be received, unless the decoder uses a sophisticated error concealment scheme. Therefore, the scheduling metric for all fragments of a base layer packet must be kept unchanged. For PR packet fragments, however, the performance of the system can be improved by re-calculating the scheduling metric based on the number of transmitted bits.

Figure 2 shows a typical distortion-bits curve for the PR packets of a GOP taken from the *carphone* sequence using ordering method II. The markers indicate packet fragment boundaries. The dotted line shows an approximation of the distortion-bits curve of the form, $\beta_{k_i} e^{\alpha_{k_i} R}$, where R is the number of bits, for each P frame PR packet, and a linear approximation of the curve for the B frame PR packets. We have noted similar behavior for the PR packets of other sequences as well. Clearly, this approximation is sufficient for the calculation of the distortion gradients of all PR packet fragments. The constants, β_{k_i}, and α_{k_i}, are the only required additional data per PR layer. The approximation also helps smooth the distortion-bits curve and prevents the scheduling algorithm from moving towards a spurious solution.

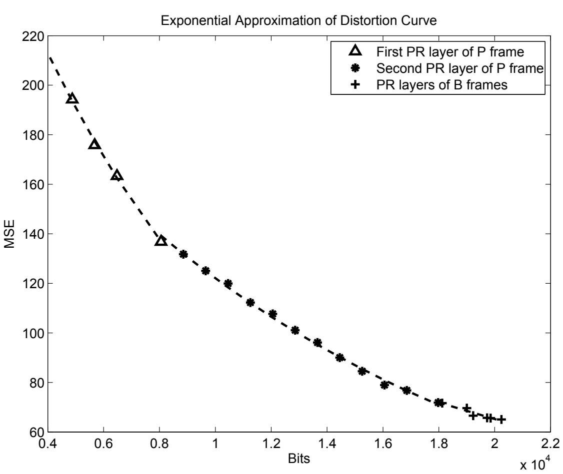

Fig. 2. Approximation of the distortion-bits curve for one GOP in carphone sequence.

3.3. Gradient-Based Scheduling

The gradient-based resource allocation problem is described in the context of multiuser video streaming in [1]. Therefore, in this paper, we only provide a summary of the technique. We consider a system in which a combination of TDM and CDMA, as in HSDPA, is used. At each transmission opportunity, the scheduler must assign the available resources to be shared among multiple users. The resources consist of a total number of spreading codes, N, and total power P. N is specified by the wireless standard, and the maximum number of spreading codes allowable per user, N_i, may be less than N due to device constraints. P needs to be constrained in order to minimize interference among users within a cell. Assuming K total users, the above constraints can be written as,

$$\sum_{i=1}^{K} p_i \leq P, \ \ \sum_{i=1}^{K} n_i \leq N, \ \ n_i \leq N_i. \qquad (2)$$

Given a channel state vector, $\mathbf{e}$, obtained from channel feedback, where e_i represents the normalized *Signal to Interference Plus Noise Ratio* of user i's channel, the gradient-based scheduling method consists of maximizing the sum of each user's achievable data rates weighted by the gradients of a utility function in order to determine the optimal resource allocations at each transmission opportunity. Using the scheduling metric calculated in (1) as the utility gradient, this can be written as,

$$\max_{(\mathbf{n},\mathbf{p})\in\chi} \sum_{i=1}^{K} u_i[k_i] n_i B \log(1 + \frac{p_i e_i}{n_i}), \qquad (3)$$

where $\mathbf{n}$ and $\mathbf{p}$ are the spreading code and power allocation vectors respectively, B denotes the symbol rate per code, χ represents the constraints on $\mathbf{n}$ and $\mathbf{p}$, and k_i is the number of packets transmitted to user i up to the current time slot. Further details on the derivation and solution of (3) are provided in [7], and are beyond the scope of this paper.

4. SIMULATIONS

We used 6 bitstreams containing 150 frames each of the varied QCIF sequences (Foreman, Mother and Daughter, Carphone, News, Hall Monitor and Silent) encoded using the JSVM 3 software. The GOP size was set at 4, with a base layer and two PR layers. The sequences were coded such that at the highest rate, and with no losses, they had a decoded Y-PSNR of about 35dB. The wireless network was modeled as an HSDPA system with multipath fading. N_i was set at 5, and N at 15. In addition to the constraints in (2), we assumed a max SINR per code constraint for each user of 1.76dB. The results were averaged over 5 channel realizations.

Figure 3 compares the proposed metric, utilizing the packet prioritizing strategy in Method II, and the approximation for PR fragments in Sec. 3.2, to a queue-length dependent metric (similar to the M-LWDF scheme in [8]), and a maximum throughput scheduling scheme, both of which also first use Method II for packet ordering. The queue-length metric achieves comparable performance when the power is increased but degrades dramatically as the power, and therefore, the available channel rates decrease. The distortion gradient metric shows a more graceful degradation with the loss of power. The performance of the maximum throughput scheme is significantly worse, as expected in a multiuser scheme where users request varied content. Figure 4 shows the cumulative reduction in distortion for a particular GOP where the reduction in MSE at each transmission opportunity is summed over all the users. Note that, in general, multiple transmission slots are available for the transmission of packets in each GOP. The figure

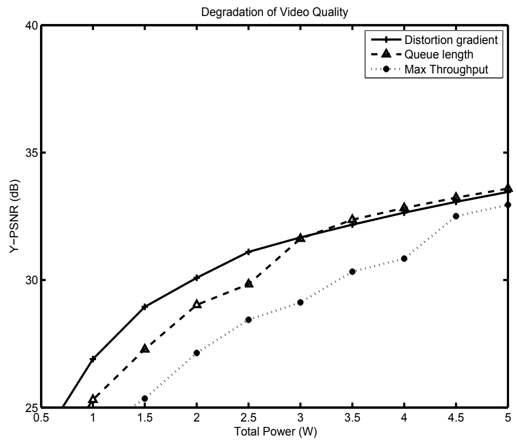

Fig. 3. Degradation of average quality with decrease in total power.

shows that given the available power, the content-dependent metric reaches a reasonable level of quality earlier than the other metrics, and that the degradation in quality as the number of available time slots decreases is slower.

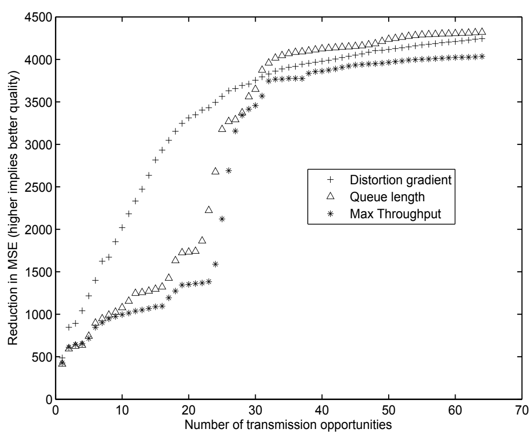

Fig. 4. Cumulative improvement in quality over all users (Higher means greater reduction of MSE) (P=5.0W).

The different ordering methods are compared in Table 1. Method IV in the table represents Method II using the approximate distortion-bits curve as specified in Sec. 3.2. We can see that Method I shows the worst performance, while Method III, which is complex to implement, does not provide any advantage over Methods II and IV. The performance loss in Method I is due to the loss of temporal resolution.

5. CONCLUSIONS

We have shown that a content-aware scheduling metric which uses the gradient of the distortion-bits curve of each video sequence in conjunction with channel state information can be easily implemented within a scalable video coding architecture. Such a scheme would significantly reduce the degradation of video quality in a congested network. We also show that SVC allows for a simple packet prioritization strategy which can be implemented offline and signaled to the real-time scheduler.

Table 1. Comparison of Ordering Methods

	Avg PSNR (dB)	Var of PSNR
Method I	29.6	44.0
Method II	33.4	2.2
Method III	33.4	2.2
Method IV	33.4	2.1

6. REFERENCES

[1] P.V. Pahalawatta, R. Berry, T.N. Pappas, and A.K. Katsaggelos, “A content-aware scheduling scheme for video streaming to multiple users over wireless networks,” in *Proc. of European Signal and Image Processing Conference*. EUSIPCO 2006, 2006.

[2] G. Liebl, H. Jenkac, T. Stockhammer, and C. Buchner, “Radio link buffer management and scheduling for wireless video streaming,” *Telecommunications Systems*, vol. 30/1-3, pp. 255–277, November 2005, Springer Science & Business Media B.V.

[3] G. Liebl, M. Kalman, and B. Girod, “Deadline-aware scheduling for wireless video streaming,” in *Proc. of IEEE Int. Conf. on Multimedia and Expo*, 2005.

[4] G. Liebl, T. Schierl, T. Wiegand, and T. Stockhammer, “Advanced wireless multiuser video streaming using the scalable video coding extensions of H.264/MPEG4-AVC,” in *Proc. of IEEE Int. Conference on Multimedia and Expo*, 2006.

[5] H. Schwarz, D. Marpe, and T. Wiegand, “Analysis of hierarchical B pictures and MCTF,” in *Proc. of IEEE Int. Conf. on Multimedia and Expo*, 2006.

[6] H.M. Radha, M. van der Schaar, and Y. Chen, “The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP,” *IEEE Transactions on Multimedia*, vol. 3, pp. 53–68, March 2001.

[7] R. Agrawal, V. Subramanian, and R. Berry, “Joint scheduling and resource allocation in CDMA systems,” *IEEE Transactions on Information Theory*, to appear.

[8] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, “Providing quality of service over a shared wireless link,” *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, Feb. 2001.