Improved Multiuser Wireless Video Streaming by Considering Influence on a Subsequent Frame

Slavche Pejoski and Venceslav Kafedziski

Abstract—**This paper gives contribution to improvement of h.264 downlink multiuser video streaming over wireless channels. The improvement is acheaved by including the influence of packet loss on the subsequent frame in a real time distortion estimation system, during the packet ordering procedure. In order to take into account the probability of unsuccessful reception of the video packets of the subsequent video frame, required for the ordering procedure, constant value probability is used. Performance improvement comes at the price of increased complexity.**

Index Terms—**h.264 video, multiuser video streaming, subsequent frame influence.**

I. Introduction

THE great expansion of internet has made video traffic almost inseparable part of our lifes, thus video transmission is envisioned to be the next leading service in wireless communications. The wireless environment is a nonreliable medium, so robust video coding must be taken into consideration. On the other hand the fluctuating nature of the wireless channel gain combined with the diversity of the video frames content gives the opportunity of improving the quality of the transmitted video when multiuser video transmission is considered. Thus, optimal resource allocation is a very important issue, and combined with source coding leads to the concept of cross layer multiuser video transmission. The advantage of cross layer multiuser video transmission system where the resources are dedicated based on the needs of the particular video frame has been research topic in a vast amount of literature, such as [1], [2] and [3]. In all this work, the influence of the video encoding process only on the currently encoded video frame is taken into account, and the influence on the future video frames is not investigated or the procedure for calculation of the functions that represent quality does not follow the losses in the channel in real time. Exploring these possibilities will be the subject of this paper.

The multiuser video transmission includes two main services: real time video transmission, and video streaming. Even though they have differences in their delay restrictions and in the time instance when the video is coded, due to the techniques of scalable coding and transcoding described in [4] and [5] the two separate video services are converging and real time video streaming becomes a very popular research subject. Here we consider scheduling in a multiuser real time video streaming system.

Previous research has shown that consideration of the dependences among the video frames of the transmitted video can boost the performance of the received video in terms of PSNR. This is the case in [6], where the dependences among various types of coded frames are taken into consideration by means of directed acyclic graph and diminishing importance is given to the future frames. The approach considered in [6] is strictly model based and does not account for the specific content of the video frames. On the the other hand, when video streaming is in question, an assessment of the influence of video packets, based on their content, can be obtained. Approach based on directed acyclic graph has been also used in [7], [8] namely RaDiO (Rate-Distortion Optimization) framework. The penalized greedy algorithm discussed in [8] offers a scheduling algorithm based on the influence of a video packet on all future frames. This influence is computed offline, as the increase of the expected distortion in the sequence due to the loss of that particular packet, when all its ancestors are received without errors. This kind of distortion calculation does not require high computational complexity, but the loss of accuracy can be severe, especially if a loss of some important ancestor has occurred. Other disadvantage of this distortion calculation technique is that error concealment is not considered.

Different approach is used in [9] and [10]. In this work packet ordering has been proposed, that takes into consideration the influence of the video packet on the distortion reduction caused by specific packets. The authors in [10] consider joint scheduling and physical layer parameters optimization in CDMA environment and authors in [9] consider error free CDMA environment. The authors use recursive per pixel algorithms proposed in [11] to estimate the influence of the specific packet based on the history of the transmission process, so the estimation is real time. The scope of their research takes into consideration only the influence of the packet on the video frame where the packet belongs. The novelty of our work comes from modifying the calculation algorithm in [9] in order to take into account the influence on the subsequent video frame. In order to calculate the influence on the subsequent video frame, estimation of the probability of unsuccessful reception of the packets in the future video frame is necessary. Packet can be unsuccessfully received due to two reasons. The first one is when the system does not have sufficient resources to send the packet and the second one is due to transmission errors. If a packet has unsuccessful reception, the part of the video frame that

S. Pejoski, Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Republic of Macedonia (phone: +38923099191; fax: +38923064262; e-mail: slavchep@feit.ukim.edu.mk).

V. Kafedziski, Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Republic of Macedonia (phone: +38923099120; fax: +38923064262; e-mail: kafedzi@feit.ukim.edu.mk).

is contained in the specific packet is recovered by the error concealment procedure in the video decoder. Here we model the probability of unsuccessful reception of the packets in the subsequent video frame as a constant value. We also show the performance gain achieved by the multiuser diversity as a result of transmitting different video sequences.

The paper is organized as follows: in section 2 we present an overview of the method, in section 3 we show the simulation setup and simulation results and in section 4 we give some concluding remarks.

II. System

Our system is made of a server that contains the video sequences and users to whom the server sends video sequences. These sequences have been previously encoded into video packets and stored at the server. Each packet contributes to a reduction of the distortion of the received video at the receiver. The server knows the length in bytes for every video packet. All users are placed in the same wireless network, and there is a scheduler at the point where the wireline network is connected to the wireless network. In our system the packets are sent from the server to the scheduler without delay and transmission errors. Based on a certain logic, the scheduler sends some of the packets, and, due to resource restrictions discards some of them. The system is shown in Fig. 1.

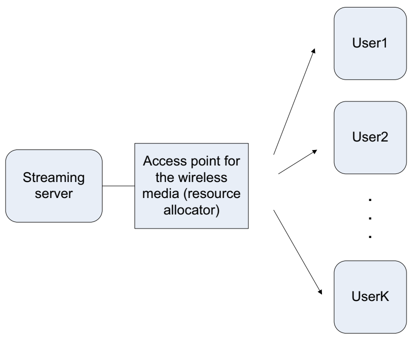

Fig. 1. Multiuser video streaming system

In order to obtain the best quality of the received video, the scheduler grants resources to packets based on their importance which can be estimated as in [9], i.e. by the average reduction of the expected distortion per bit obtained by the reception of the packet. The distortion per pixel can be estimated using one of the recursive methods for expected distortion estimation, such as those described in [11] or [12]. The expected distortion is calculated as:

$$D = E\{(f_n^i - \tilde{f}_n^i)^2\} \tag{1}$$

In (1), $\tilde{f}_n^i$ stands for the anticipated decoded value of the i^{th} pixel of the n^{th} video frame at the encoder, and f_n^i stands for the value of the i^{th} pixel of the n^{th} uncoded video frame. Considering that packet m will be successfully received with probability p_m, the expected distortion due to packet m will be:

$$D_m = p_m D_{m_s} + (1 - p_m) D_{m_{us}}. \tag{2}$$

In (2) D_{m_s} is the expected distortion when successful reception occurs, and $D_{m_{us}}$ is the expected distortion for unsuccessful reception of packet m.

The scheduler allocates the available resources based on:

$$\min_{r_i} \sum_{i=1}^{K} \sum_{m=1}^{M} D_{i,m}. \tag{3}$$

In (3) r_i stands for resources allocated to user i, K is the number of users in the system, M is the number of video packets used to encode a single video frame and $D_{i,m}$ is the expected distortion for packet m of user i if r_i resources are allocated to user i.

In our system we assume a block fading channel model, with partial channel state information at the server, that includes the pdf's of the channel gains and the durations of the coherence intervals of all the users. Moreover, we assume that strong channel codes are used and that transmission errors occur only due to outage events and that the number of bits in a packet does not influence the packet error probability as long as the packet is sent in a constant number of coherence intervals. Here we assume that the scheduler does not change the parameters of the physical layer and uses previously defined number of bits R in every coherence interval that corresponds to outage probability p in the same interval.

Our approach is based on the approach in [9] that solves (3), so we give a short explanation of that algorithm. Let m video packets $\psi_1, \psi_2, ..., \psi_m$, be sent from a particular user, where $\psi_1, \psi_2, ..., \psi_m$ is a permutation of the user's packets. Let Ψ be the set of packets of the particular user that are already sent and $\Phi = \{\phi_1, ..., \phi_M\}$ be the set of all the packets of the same user that are produced by encoding a single frame. Ψ is a subset of Φ and M is the total number of packets used to encode the n-th video frame. Only the packets that have not been already sent are included in the packet ordering procedure. At each step, this procedure finds the packet that results in the largest reduction in the expected distortion, i.e. performs the following optimization procedure:

$$\max_{i \in \{\Phi \backslash \Psi\}} \frac{E\{D|\Psi\} - E\{D|\Psi \cup i\}}{B_i} \tag{4}$$

In (4) B_i denotes the number of bits required for sending packet i, $E\{D|\Psi\}$ is the expected distortion of the video frame where the packet i belongs in situation when packets in Ψ are already sent, and $E\{D|\Psi \cup i\}$ is the expected distortion of the same video frame when packets in Ψ and the packet i are already sent. Then, the packet i is added to Ψ. The procedure is carried out until all unsent video packets are sorted. This procedure must be completed for every user. Then the scheduler creates a virtual buffer for every user where every virtual buffer contains the sorted packets and the packet with the largest reduction of the expected distortion per bit is set as head of line for the appropriate virtual buffer. Then the packet that has the largest reduction of the expected distortion per bit from the head of lines of all virtual buffers obtains the appropriate resources for transmission. In the ordering procedure the scheduler has an accurate information about the reception of the video packets at the users. This information is available after certain previously defined delay. If any packet is unsuccessfully received then it is reinserted in the virtual buffer and the ordering procedure for that virtual buffer is carried out again.

Here, we extend the concept from [9] by calculating the reduction of the expected distortion not only in the current frame, but also in the subsequent frame, so that (4) becomes:

$$\max_{i\in\{\Phi\setminus\Psi\}} \frac{E\{D_n|\Psi\}+a\cdot E\{D_{n+1}|\Psi\}}{B_i} - \frac{E\{D_n|\Psi\cup i\}+a\cdot E\{D_{n+1}|\Psi\cup i\}}{B_i}. \tag{5}$$

In (5) $n+1$ denotes that the expected distortion is computed for the subsequent video frame. The coefficient a allows us to change the contribution of the subsequent video frame to the expected distortion per bit. When calculating (5) the probability of unsuccessful reception of a particular packet from the subsequent video frame is needed. We model the probability of unsuccessful reception as constant value p_1, equal for all packets in a single video sequence and for all users.

We must point out that the improvement due to inclusion of the influence on the subsequent video frame comes at a price of complexity increase (the complexity is inceased approximately twice), but the procedure is intended for multiuser downlink system that has loose constraints on the processing power which makes the method attractive for such systems.

III. Simulations

We base our simulations on h.264/AVC JM v.16 video coder, publicly available at [13]. Five different QCIF video sequences (foreman, news, hall monitor, mother and daughter, carphone) are encoded at a frame rate of 30 fps. Each video sequence is divided into slices, independently encoded. Every slice consists of one row of macroblocks. Resynchronization markers are used in every slice. The first video frame is encoded INTRA and is available at the receiver with no transmission errors and all the other video frames are encoded INTER. In order to mitigate the effects of error propagation 15 random macroblocks from every video frame are encoded INTRA and the rest of them are encoded INTER. The video sequences are encoded using CABAC (Content Adaptive Binary Arithmetic Coding). For encoding each video sequence a constant number of quantization levels that produces a reconstructed video sequence of overall quality of 35 dB for 300 video frames of each video sequence, is used.

The video transmission is intended for symmetrically placed users with equal mean channel gain. In our system the physical and the application layer are separated and only one option at physical layer is offered to the application layer. The system supports a rate of 500 kbps over an error prone channel. The coherence interval of the channel is set to be 1/6 of the duration of a single video frame transmission. The outage probability is set to $p = 0.1$ for each coherence interval.

At the receiver we use an error concealment method that uses the median motion vector from the macroblocks at the up and left, up, and up and right position relative to the current macroblock. If the aforementioned motion vectors are not available, the zero motion vector is used for the error concealment procedure. In order to calculate the expected distortion we use the ROPE algorithm from [11] and it's distance adaptive correlation calculation and quantization theory - based rounding error compensation versions from in [14]. The time after which the scheduler receives an accurate information about the reception of video packets for all users is set to two coherence intervals.

First we show the gain from multiuser diversity due to video sequences. In order to do so we compare the multiuser system where the influence of the packet on the subsequent video frame is not considered, to the system where every user uses equal resources i.e. every user uses 100kbps over channel with coherence interval of 1/6 of the video frame transmission time and outage probability of $p = 0.1$ in every coherence interval. The system in which every user uses equal resources also does not consider the influence of the packet on the subsequent video frame. All simulations in this paper were carried out 23 times with respect to different channel realizations for every user.

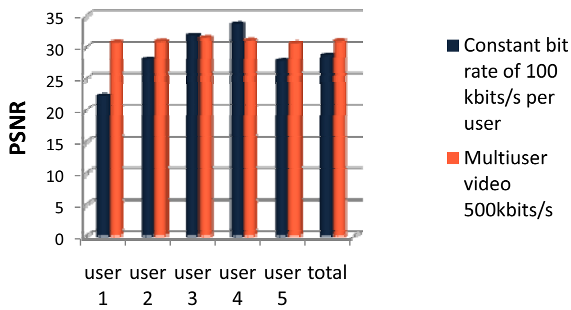

Fig. 2. Average PSNR for different users/sequences

In Fig. 2 the average quality of the received video measured in PSNR is shown for the multiuser system and the system where every user uses equal resources. The averages are taken over 150 video frames and over all the simulations. The order of the sequences is the same as in the description given at the beginning of the section. The figure shows that multiuser system outperforms the system where every user uses equal resources, which can be seen from the last column, which gives the average performance over all users. The performance gain is of the order of 2 dB. Notice that some users might have worse performance than in the system with equal resources.

The performance of the two systems averaged over all users is shown in In Fig. 3. In every video frame, the multiuser system outperforms the system where every user uses equal resources.

Next, we compare our proposed method with a multiuser video system that does not include the influence on the subsequent video frame. As previously described in this paper, the probability of unsuccessful reception is set to a constant value equal for all packets in a single video sequence and for

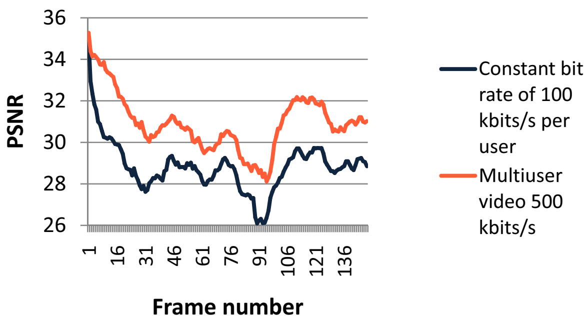

Fig. 3. Average PSNR for different users/sequences

all users. We used the value of $p_1 = 0.1$. We set the parameter to a value $a = 1$.

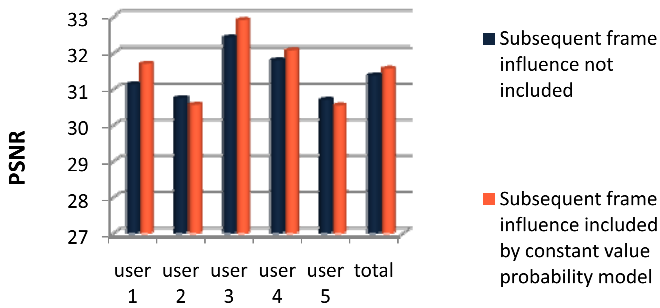

Fig. 4. Average PSNR for different users/sequences

In Fig. 4 the average quality of the received video measured in PSNR for every user is shown for the two models. The averages are taken over 69 video frames and over all the simulations. The order of the sequences is the same as in the description given at the beginning of the section. The figure shows that our proposed method outperforms the method that does not include the influence on the subsequent frame. The performance gain is of the order of 0.25 dB. As can be seen, our method shows superior performance for all sequences, except for the news video sequence where the deterioration is very small.

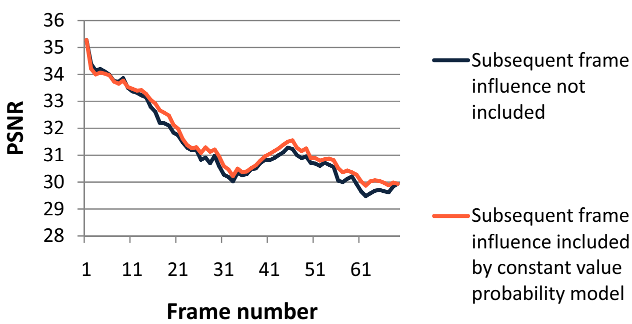

Fig. 5. Average PSNR in terms of the video frame order

In Fig. 5 the performance of the multiuser system in terms of the video frame order is shown. Again, the proposed model shows performance which is superior to that of a system not including the influence on the subsequent video frame. Although the average improvement of the system behavior is of the order of 0.25 dB it can be seen from Fig. 5 that for some frames the improvement of the PSNR averaged over all users reaches 0.5 dB. In Fig. 5 a small deterioration can be noticed for low frame order, but this behavior can be explained by the inaccuracy of the constant value model.

It should be noted that taking into account the influence on future video frames that are further away from the current video frame is expected to result in modest additional performance gain, at the expense of even higher complexity.

IV. Conclusion

We propose an improvement to the real time multiuser video streaming in wireless channels. The improvement is due to the packet ordering procedure that includes the influence on the packets in the subsequent frame. To estimate the probability of unsuccessful reception of video packets in the subsequent video frame, a simple constant value model is used. The proposed method outperforms the standard multiuser video streaming algorithm by 0.25 dB, at a price of increased complexity.

References

[1] L. U. Choi, W. Kellerer, and E. Steinbach, "On cross-layer design for streaming video delivery in multiuser wireless environmentsy," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, pp. 55–65, 2006.

[2] G. M. Su, Z. Han, M. Wu, and K. J. R. Liu, "A Scalable Multiuser Framework for Video Over OFDM Networks: Fairness and Efficiency," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 16, no. 10, pp. 1217–1231, 2006.

[3] Z. Guan, D. Yuan, and H. Zhang, "Optimal and Fair Resource Allocation for Multiuser Wireless Multimedia Transmissions," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, 2009.

[4] G. M. Su and M. Wu, "Efficient bandwidth resource allocation for low-delay multiuser video streaming," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 15, no. 9, pp. 1124–1137, 2005.

[5] M. Xia, A. Vetro, H. Sun, and B. Liu, "Rate-distortion optimized bit allocation for error resilient video transcoding," in *ISCAS (3)*, 2004, pp. 945–948.

[6] F. Fu and M. van der Schaar, "A systematic framework for dynamically optimizing multi-user wireless video transmission," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 308–320, 2010.

[7] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, 2006.

[8] C. Vleeschouwer, J. Chakareski, and P. Frossard, "The Virtue of Patience in Low-Complexity Scheduling of Packetized Media With Feedback," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 348–365, 2007.

[9] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-Aware Resource Allocation and Packet Scheduling for Video Transmission over Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 749–759, 2007.

[10] E. Maani, P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Resource Allocation for Downlink Multiuser Video Transmission Over Wireless Lossy Networks," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1663–1671, 2008.

[11] R. Zhang, S. Regunathan, and K. Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience," *IEEE Journal on Selected Area in Communications*, vol. 18, no. 6, pp. 966–976, 2000.

[12] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint Source-Channel Rate-Distortion Optimization for H.264 Video Coding Over Error-Prone Networks," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 445–454, 2007.

[13] *Available at: http://iphome.hhi.de/suehring/tml/.*

[14] H. Yang and K. Rose, "Advances in recursive per-pixel estimation of end-to-end distortion for application in h.264," in *ICIP (2)*, 2005, pp. 906–909.