Adaptive Resource Allocation and Frame Scheduling for Wireless Multi-User Video Streaming

Yang Peng, Shoaib Khan, Eckehard Steinbach
Media Technology Group
Technische Universität München
Munich, Germany

Marco Sgroi, Wolfgang Kellerer
DoCoMo Communications Laboratories GmbH
Munich, Germany

***Abstract*—We propose an application-driven multi-user resource allocation and frame scheduling concept for wireless video streaming. Our approach is based on joint optimization of the application layer, the data link layer and the physical layer. For this, key parameters from these three layers are abstracted. The abstracted parameters at the application layer describe the rate-distortion characteristics of the pre-encoded video streams. At the lower layers they describe the current transmission characteristics of all users. The outcome of the joint optimization leads to adaptive resource allocation at the lower layers and an adaptive decision on which frames to send on the application layer. We show that for our scenario the expected video quality at the client side can be described analytically which leads to low complexity joint optimization. The performance of our approach is demonstrated using a real-time testbed implementation.**

I. INTRODUCTION

Wireless multimedia communication is challenging due to the time varying transmission characteristics of the wireless channel and the dynamic application QoS requirements. In order to address these challenges joint or cross-layer adaptation can be performed at all OSI layers. The application layer can adapt to the varying network characteristics by adequate processing, such as for instance dynamic rate variation at the video encoder [5] or decoder [6], or joint adaptation of application source rate and FEC code rate [7]. The wireless network can adapt to the application QoS requirements by adequate processing at the physical, data link and network layers.

Most of the ongoing research on joint optimization focuses on the physical layer and data link layer in the protocol stack. [8] provides an overview of the cross-layer paradigm shift that is beginning to take place as wireless communication evolves from a circuit-switched to a packet-based infrastructure.

We propose a cross-layer and multi-user optimization concept for wireless video streaming. Information is exchanged across several layers in the protocol stack and the optimization is carried out jointly for multiple video streaming users. The work presented in this paper is based on the cross-layer architecture proposed in [4] where the joint optimization is based on abstracted parameters from the application and physical/link layer. We show that for our scenario the expected video quality at the client side can be described analytically which leads to low complexity joint optimization. A real-time video streaming testbed is designed and implemented to demonstrate the benefits of the proposed cross-layer optimization approach.

II. SYSTEM OVERVIEW

Our testbed for wireless multi-user video streaming consists of five components: a streaming server, a cross-layer optimizer, a radio link layer parameter generator, a wireless channel emulator and multiple streaming clients (Fig. 1).

The streaming sever receives requests from the streaming clients and delivers the video content using RTSP and RTP/RTCP. The source video data is pre-encoded using MPEG-4. The side information required for the optimization is extracted during the encoding process. Both the compressed video data and the side information are stored on the streaming server. Based on the decision of the optimizer, the compressed video files are further processed (e.g., packetized, protected by FEC, scheduled, etc.) and transmitted. Previous frame concealment is performed at the clients when frames are lost in the wireless channel. The wireless channel emulator is capable of emulating a mobile multi-user scenario, in which each user has an independent wireless channel with different transmission characteristics.

Based on the decision of the optimizer, the radio link layer parameter generator provides the corresponding channel parameters to the wireless channel emulator. The cross-layer optimizer interfaces the streaming server and the radio link layer parameter generator and carries out the optimization for each Group of Pictures (GOP) of each user.

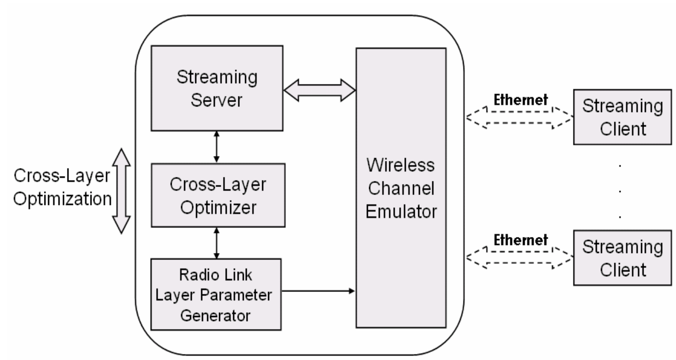

Fig. 1. Cross-layer optimization testbed.

III. Channel Model

For mobile radio channels, Rayleigh fading is a widely accepted model. The Gilbert-Elliot packet erasure channel model has been shown to model this kind of fading channel with sufficient accuracy [1]. In our work, we also use the Gilbert-Elliot model to simulate the packet error behavior of the wireless channels. The two states of the GE model are denoted as G (good) and B (bad). In state G, packets are assumed to be received correctly and timely, whereas in state B, packets are assumed to be lost. This model can be described by the transition probabilities p from state G to B and q from state B to G. The steady-state probability of being in the good state and bad state, respectively, is given by

$$P_G = \frac{q}{p+q}, \quad P_B = \frac{p}{p+q}. \tag{1}$$

IV. Video Streaming Application

The expected video quality at the client side can be determined analytically by the optimizer from application and radio link layer parameters. Necessary information include: 1) video source rate, GOP structure and rate-distortion profile [9] from the application layer, and 2) data packet size, transmission data rate and transition probabilities of the GE channel model from the radio link layer.

In the following, we assume that the video is encoded with GOPs of 15 frames, including one I-frame and 14 P-frames. Therefore, one frame in the GOP can only be successfully decoded when all the previous frames are received correctly. Each frame is divided into several packets, where the number of packets depends on the frame size and the packet size. It is further assumed that a repetition code is used at the application layer for forward error correction (FEC) and the number of repetitions is determined according to the available transmission data rate.

V. Optimization

For a single user scenario, four key parameters are abstracted at the radio link layer: 1) transmission data rate d, 2) transmission packet error rate e, 3) data packet size s, and 4) channel coherence time t. These four parameters form the abstracted parameter tuple $\tilde{\mathbf{r}}_i = (d_i, e_i, s_i, t_i)$ at the radio link layer. The transmission data rate d is influenced by the modulation scheme, the channel coding, and the multi-user scheduling. The transmission packet error rate e is influenced by the transmit power, channel estimation, signal detection, the modulation scheme, the channel coding, etc. The channel coherence time t of a user is related to the user velocity and its surrounding environment, while the data packet size s is usually defined by the wireless system standard. In a K user scenario, this parameter abstraction can be extended for each user, which results in the parameter tuple $\tilde{\mathbf{r}}_i$ that contains *4K* parameters: $\tilde{\mathbf{r}}_i = (d_i^{(1)}, e_i^{(1)}, s_i^{(1)}, t_i^{(1)}, \ldots, d_i^{(K)}, e_i^{(K)}, s_i^{(K)}, t_i^{(K)})$. In order to effectively use the GE model at the radio link layer, the two transition probabilities (p and q) have to be set up properly. Following [2], the two transition probabilities can be obtained as

$$p = \frac{es}{td}, \quad q = \frac{(1-e)s}{td}. \tag{2}$$

For a given packet size s, the two parameters (p and q) of the GE channel model, together with the transmission data rate d, form the parameter tuple

$$\tilde{\mathbf{r}}_i = (d_i^{(1)}, s_i^{(1)}, p_i^{(1)}, q_i^{(1)}, \ldots, d_i^{(K)}, s_i^{(K)}, p_i^{(K)}, q_i^{(K)}) \tag{3}$$

which is the input to the optimizer from the radio link layer. Given the abstracted parameters from both the streaming server and the radio link layer, the cross-layer optimizer carries out the optimization for each GOP of each user with respect to a particular objective function [4]. Since the channel model is a random process, the expected value of the user perceived video quality is used for optimization. In order to carry out the cross-layer optimization in real time, an analytical solution is proposed for the optimizer.

A. Transmission without repetition

When the current transmission data rate is equal to the video source rate, the GOP can only be transmitted once. In this case, all the packet losses can be categorized into 16 different patterns, as shown in Fig. 2, where n_i (i = 1,…,15) denotes the number of packets in the i-th frame, which can be determined from the rate vector and the packet size.

Pattern 1 represents the cases of packet loss where at least one packet in the I-frame is lost and therefore the I-frame is not decodable. Because of the frame dependencies, all the frames in the current GOP cannot be decoded and will be replaced by the last decoded frame in the previous GOP. Pattern 2 includes all the cases where all the packets in the I-frame are received correctly but at least one packet in frame P_1 is lost. The rest may be deduced by analogy. Pattern 16 represents the case without any packet loss. Given the transition probabilities (p and q) of the GE model, the probability of each pattern p_i can be computed by

$$\begin{aligned}
p_1 &= 1 - P_G(1-p)^{(n_1-1)}; \\
p_2 &= P_G(1-p)^{(n_1-1)} - P_G(1-p)^{(n_1+n_2-1)}; \ldots \\
p_i &= P_G(1-p)^{(n_1+\cdots+n_{i-1}-1)} - P_G(1-p)^{(n_1+\cdots+n_i-1)}; \ldots \\
p_{15} &= P_G(1-p)^{(n_1+\cdots+n_{14}-1)} - P_G(1-p)^{(n_1+\cdots+n_{15}-1)}; \\
p_{16} &= P_G(1-p)^{(n_1+\cdots+n_{15}-1)}
\end{aligned} \tag{4}$$

where P_G denotes the steady-state probability of being in the good state in (1). In addition, we can also compute the resulting reconstruction distortion D_i for each loss pattern from the distortion matrix proposed in [9]. Once the loss pattern probability p_i and resulting reconstruction distortion D_i are obtained, the expected reconstruction distortion D_{exp} can be computed by

$$D_{\exp} = \sum_{i=1}^{16} p_i D_i. \tag{5}$$

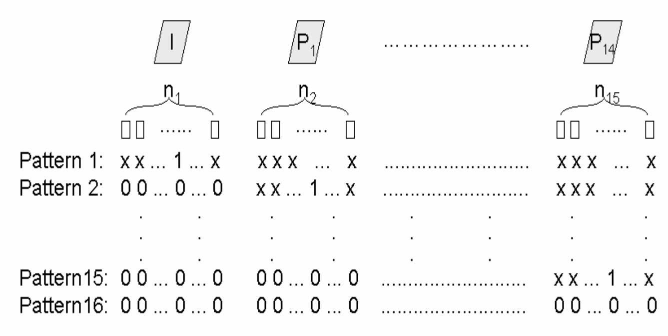

Fig. 2. Packet losses categorized into 16 loss patterns.

B. Transmission with full repetition

When the transmission data rate is an integer multiple of the video source rate, all the packets in the GOP are repeated. Let n denote the number of packets in the GOP and k denote the number of repetitions. If any one out of these k copies is received correctly, this packet will be successfully received. Since the packet size in the wireless system is very small, n is normally large enough that we can ignore the dependency between the GOPs. Therefore, the effect of k repetitions is equivalent to a parallel structure with k branches, as shown in Fig. 3.

Based on the equivalent parallel structure, we can map the original GE model to an equivalent GE model, which essentially reflects the effect of the repetition. Fig. 4 shows this mapping process. From the parallel structure, we can derive the new average packet loss probability $P_e^* = P_e^k$ and the probability of staying in state B for the new GE model

$$P(state\,B \mid state\,B) = 1 - q^* = (1-q)^k . \tag{6}$$

The transition probabilities of the equivalent GE model then become

$$p^* = q^* \frac{(1-P_G^*)}{P_G^*};\; q^* = 1-(1-q)^k \tag{7}$$

where P_G^* is given by

$$P_G^* = 1 - P_B^* = 1 - P_B^{\,k} = 1 - \left(\frac{p}{p+q}\right)^k . \tag{8}$$

Using the equivalent GE model approach, we can calculate the expected reconstruction distortion in the same way as for the case without repetition. For this, the GE model parameters in (4) are replaced by the parameters of the equivalent GE model.

C. Transmission with partial repetition

For partial repetition the more important packets are repeated until the transmission data rate is reached. The resulting equivalent parallel structure is shown in Fig. 5.

As an example, it is assumed that the packets in the first i frames are transmitted twice, while the others are transmitted only once. The value of i is determined by the transmission data rate and the video source rate. For the packets in the first i frames, we use the equivalent GE model to reflect the repetition. For the other packets in the GOP, we use the original GE model. Therefore, we compute the probability of each pattern p_i by

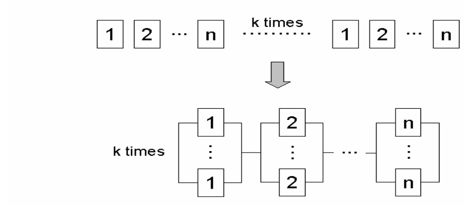

Fig. 3. Equivalent parallel structure for full repetition.

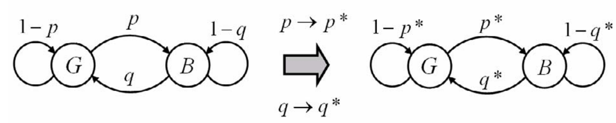

Fig. 4. Equivalent GE model approach.

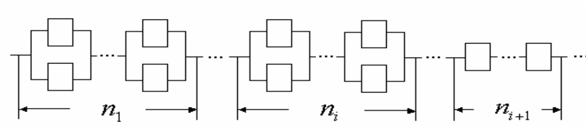

Fig. 5. Equivalent structure for partial repetition.

$$\begin{aligned}
p_1 &= 1 - P_G^*(1-p^*)^{(n_1-1)};\\
p_2 &= P_G^*(1-p^*)^{(n_1-1)} - P_G^*(1-p^*)^{(n_1+n_2-1)};\ldots\\
p_i &= P_G^*(1-p^*)^{(n_1+\cdots+n_{i-1}-1)} - P_G^*(1-p^*)^{(n_1+\cdots+n_i-1)};\\
p_{i+1} &= P_G^*(1-p^*)^{(n_1+\cdots+n_i-1)}(1-P_G(1-p)^{(n_{i+1}-1)});\ldots\\
p_{15} &= P_G^*(1-p^*)^{(n_1+\cdots+n_i-1)}(P_G(1-p)^{(n_{i+1}+\cdots+n_{14}-1)} - P_G(1-p)^{(n_{i+1}+\cdots+n_{15}-1)});\\
p_{16} &= P_G^*(1-p^*)^{(n_1+\cdots+n_i-1)}(P_G(1-p)^{(n_{i+1}+\ldots+n_{15}-1)})
\end{aligned} \tag{9}$$

where p^* and P_G^* is given by (7) and (8). Substituting (9) into (5), we obtain the expected reconstruction distortion.

VI. SIMULATION

A multi-user scenario with three QCIF test video sequences (foreman, carphone, mother-daughter) is considered in our simulation. Each sequence has 300 frames and the frame rate is 30 fps. All three videos are encoded at the target rate of 100kbps using the Xvid codec. Each GOP has one I-frame and 14 P-frames. On the radio link layer, it is assumed that the total transmission symbol rate in the system is 450k symbols/s. The data packet size is equal to 54 bytes, which is the specified packet size of the IEEE802.11a or HiperLAN2 standard. The channel coherence time is assumed to be 50ms for all the three users, which approximately corresponds to a pedestrian speed (for 5GHz carrier frequency). The residual packet error rate can be described as a function of the average SNR [3]. This residual packet error rate is used as the parameter e in (2). User position dependent path loss and shadowing commonly observed in wireless links are taken into account by choosing

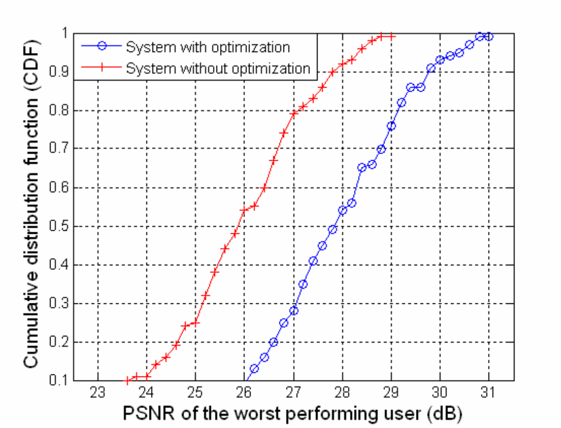

Fig. 6. Performance comparison with and without cross-layer optimization. CDF for 1000 simulation runs.

the corresponding average SNR for each user. In addition, two parameters at the radio link layer can be optimized: 1) modulation scheme, and 2) multi-user scheduling. Two different modulation schemes (BPSK and QPSK) and seven cases of air time arrangement in a time-division multiplexing based multi-user scheduling (as shown in Table I) are assumed. Given the modulation scheme and the case of time arrangement, the resulting transmission data rate can be determined for each user. For example, if a user uses BPSK and 2/9 of the total transmission time is assigned to it, its transmission data rate is equal to 100kbps. A user can have a transmission data rate as high as 400kbps when QPSK is used and 4/9 of the transmission time is assigned. For simplicity, we assume that every user uses the same modulation scheme. Therefore, the two parameters that can be optimized lead to 14 cases of resulting transmission data rate, as shown in Table II.

In our simulation, the three videos are synchronized at the GOP level. At the beginning of the transmission of each GOP the average SNR is generated independently for each user. For each of the 14 cases the SNR is translated into the packet error rate. We then transform the parameters of the radio link layer into the transition probabilities of the GE model. In the next step the expected reconstruction distortion of the current GOP for each user is computed from (5). The distortion is then expressed as PSNR and the multi-user objective function is defined based on the individual PSNR values. In our experiment the optimization maximizes the quality of the worst performing user in the system. We compare the proposed cross-layer optimization approach with a system without optimization. In the system without joint optimization, the same amount of transmission time is assigned to all users and every user uses BPSK (i.e., Case 1 in Table II). It is assumed that all the three users are moving. Two of them have good channel conditions, while the other one has a relatively bad channel. The SNR of the two users with good channel conditions is changing randomly between 10 and 20dB, and that of the other user between 0 and 10dB. It can be seen from Fig. 6 that the PSNR of the worst performing user improves significantly in the system with cross-layer optimization. For example, there is a 50% chance that the PSNR of the worst performing user is larger than 28dB in the system with optimization, which improves about 2dB compared to the system without optimization.

TABLE I: MULTI-USER SCHEDULING

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
User 1	3/9	4/9	4/9	3/9	2/9	3/9	2/9
User 2	3/9	3/9	2/9	4/9	4/9	2/9	3/9
User 3	3/9	2/9	3/9	2/9	3/9	4/9	4/9

TABLE II: TIMING ARRANGEMENTS AND RATE OF DIFFERENT CASES

	Modulation Scheme	Time Arrangement	Resulting Transmission Data Rate (kbps)		
			User 1	User 2	User 3
Case 1	BPSK	A_1	150	150	150
Case 2	BPSK	A_2	200	150	100
…	…	…	…	…	…
Case 14	QPSK	A_7	200	300	400

VII. CONCLUSION

In this paper we have presented a multi-user resource allocation and frame scheduling concept for wireless video streaming. After introducing the real-time multi-user testbed implementation for cross-layer optimization, we show that for our scenario the expected video quality at the client side can be described analytically, which leads to low complexity joint optimization. The performance of the proposed cross-layer optimization approach is evaluated in the real-time testbed for a three-user scenario. The simulation results show that already in simple scenarios, significant improvements can be achieved by the proposed joint optimization approach.

REFERENCES

[1] H. S. Wang, P. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," IEEE Transactions on Vehicular Technology, vol. 45, no. 2, pp. 353- 357, May 1996.

[2] M. T. Ivrlac, "Parameter selection for the Gilbert-Elliott model," Technical Report TUM-LNS-TR-03-05, Institute for Circuit Theory and Signal Processing, Munich University of Technology, May 2003.

[3] W. Kellerer, Lai-U Choi, and E. Steinbach, "Cross-layer adaptation for optimized B3G service provisioning", Proc. 6th Int. Symp. on Wireless Personal Multimedia Communications (WPMC), Japan, Oct. 2003.

[4] Lai-U Choi, W. Kellerer, and E. Steinbach, "Cross-layer optimization for wireless multi-user video streaming," ICIP-2004, Singapore, Oct. 2004.

[5] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, "Video coding for streaming media delivery on the Internet," IEEE Trans. CSVT, vol. 11, no. 3, pp. 269-281, Mar. 2001.

[6] E. Steinbach, N. Faerber, and B. Girod, "Adaptive playout for low-latency video streaming," ICIP-2001, pp. 962-965, Thessaloniki, Greece, Oct. 2001.

[7] K. Stuhlmüller, N. Färber, and B. Girod, "Analysis of video transmission over lossy channels," IEEE Journal on Selected Areas in Communications, vol. 18, no. 6, pp. 1012-1032, June 2000.

[8] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," IEEE Communications Magazine, vol. 41, no. 10, pp. 74-80, Oct. 2003.

[9] W. Tu, W. Kellerer, and E. Steinbach, "Rate-distortion optimized video frame dropping on active network nodes," Packet Video Workshop, Irvine, California, Dec. 2004.