CROSS LAYER OPTIMIZATION FOR WIRELESS MULTI-USER VIDEO STREAMING

Lai-U Choi[1]*, Wolfgang Kellerer*[2]*, and Eckehard Steinbach*[1]

[1]Media Technology Group
Institute of Communication Networks
Technische Universität München

[2]DoCoMo Communication Laboratories
Europe GmbH Munich

ABSTRACT

A cross-layer optimization concept for wireless multi-user video streaming is proposed. We describe a cross layer optimizer that interfaces the video streaming application and the radio link layer by means of parameter abstraction. The optimizer maximizes the end-to-end quality of the video streaming service jointly for all users while efficiently using the wireless resources. Our simulation results for video streaming in a multi-user environment show the performance improvements achievable with this concept. We demonstrate that even for a small number of users and a small number of degrees of freedom in the optimization significant quality improvements can be obtained.

1. INTRODUCTION

Cross layer design in mobile communication has recently gained much attention in the context of multimedia service provisioning. The concept of cross-layer design is based on inter-layer information exchange across the protocol stack with the aim of joint optimization of the communication on two or more layers. Although this concept can be employed in all communication networks, it is especially relevant in wireless networks because of the unique challenge of the wireless environment. The time-varying and fading nature of the wireless channels together with user mobility lead to random variation in network performance and connectivity. In addition, the demanding quality of service (QoS) requirements for multimedia support makes mobile multimedia communication even more challenging in system design. This challenge is hard to meet with a conventional layered design approach, which separates system design into essentially independent layers.

In this work, we propose a cross-layer optimization approach for wireless multi-user video streaming that jointly considers the application layer and the radio link layer. We refer to the radio link layer as the physical layer and the data link layer in the protocol stack. We include the video streaming application in the joint optimization because it has direct information about the impact of each successfully decoded piece of video data on the perceived quality. We also include the physical layer and the data link layer in our consideration because the unique challenge of mobile wireless communication results from the nature of the wireless channel, which these two layers have to cope with.

Previous work mainly concentrates on optimizing the performance at a single layer, such as the adaptation of the application to the transport, network, data-link and physical layer characteristics (bottom-up approach) and the adaptation of the physical, data link or network layers to the application requirements (top-down approach). Most of the on-going research in cross layer design focuses on joint optimization of the physical layer and data link (or MAC) layer (e.g. [1],[2]). Only recently approaches that explicitly include the application in the cross layer optimization appeared (e.g. [3],[4]).

2. SYSTEM ARCHITECTURE

We consider a video-streaming server located at the base station and multiple mobile streaming clients. K streaming clients or users are assumed sharing the same air interface and network resources but requesting different video content. At the base station, the architecture shown in Figure 1 is proposed to provide end-to-end quality of service optimization.

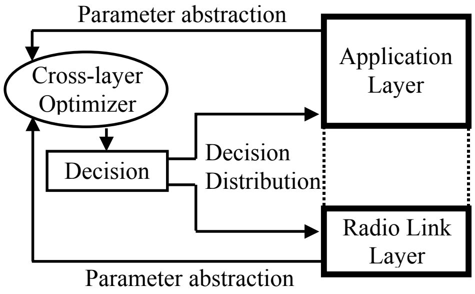

Figure 1: Cross-layer optimization architecture.

This figure illustrates the tasks and information flows related to the proposed joint optimization concept. Necessary state information is first collected from the application layer and the radio link layer through the process of *parameter abstraction*. The process of parameter abstraction results in the transformation of *layer specific parameters* into parameters that are comprehensible for the *cross-layer optimizer*. The optimization is carried out with respect to a particular objective function. From a given set of possible cross-layer parameter tuples, the tuple optimizing the objective function is selected. After the decision on a particular cross-layer parameter tuple is made, the optimizer distributes the decision information back to the corresponding layers.

3. PARAMETER ABSTRACTION

In order to carry out the joint optimization, state information has to be abstracted from the selected layers and provided to the

0-7803-8554-3/04/$20.00 ©2004 IEEE.

cross-layer optimizer. This is necessary because layer specific or technology specific parameters may be incomprehensible or of limited use to other layers and the optimizer.

3.1. Radio Link Layer

The physical layer deals with issues like transmit power control, channel estimation, synchronization, signal shaping, modulation and signal detection, while the data link layer is responsible for radio resource allocation and error control. Since both of these two layers are closely related to the unique characteristics of the wireless channel, it is useful to consider them together. In the following, we refer to their combination as the *radio link layer*. Since there are many technology specific parameters in the radio link layer parameter abstraction is necessary. To be more specific, we follow the approach proposed in [6] and define the set $\mathcal{R}=\{\mathbf{r}_1,\mathbf{r}_2,...\}$ of tuples $\mathbf{r}_i=(r_i^1,r_i^2,...)$ of radio link layer specific parameters r_i^j (e.g., modulation alphabets, code rate, air time, transmit power, coherence time). Since these radio link specific parameters may be variable, the set $\mathcal{R}$ contains all possible combinations of their values and each tuple $\mathbf{r}_i$ represents one possible combination. In order to formalize the process of parameter abstraction, we define the set $\tilde{\mathcal{R}}=\{\tilde{\mathbf{r}}_1,\tilde{\mathbf{r}}_2,...\}$ of tuples $\tilde{\mathbf{r}}_i=(\tilde{r}_i^1,\tilde{r}_i^2,...)$ of abstracted parameters $\tilde{r}_i^j$. We call the mapping between $\mathcal{R}$ and $\tilde{\mathcal{R}}$ *radio link layer parameter abstraction*. For a single user scenario, for example, four key parameters can be abstracted. They are transmission data rate d, transmission packet error rate e, data packet size s, and the channel coherence time t. This leads to the abstracted parameter tuple $\tilde{\mathbf{r}}_i=(d_i,e_i,s_i,t_i)$. In a K user scenario, one can extend the parameter abstraction for each user. The parameter tuple $\tilde{\mathbf{r}}_i$ then contains $4K$ parameters $\tilde{\mathbf{r}}_i=(d_i^{(1)},e_i^{(1)},s_i^{(1)},t_i^{(1)},...,d_i^{(K)},e_i^{(K)},s_i^{(K)},t_i^{(K)})$, in which a group of four parameters belongs to one user. 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, the current user position, 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 normally defined by the wireless system standard.

Alternatively, it is possible to transform the transmission packet error rate e and the channel coherence time t into the two parameters of the two-state Gilbert-Elliott model, which are the transition probabilities (p and q) from one state to another. The transformation is given by [5]

$$p=\frac{es}{td} \text{ and } q=\frac{(1-e)s}{td} \tag{1}$$

where p is the transition probability from the good state to the bad state and q is the transition probability from the bad state to the good state. The abstracted parameter tuple now becomes $\tilde{\mathbf{r}}_i=(d_i^{(1)},p_i^{(1)},s_i^{(1)},q_i^{(1)},...,d_i^{(K)},p_i^{(K)},s_i^{(K)},q_i^{(K)})$. The main advantage of this parameter abstraction step is that the resulting parameter tuple $\tilde{\mathbf{r}}_i$ is no longer technology specific and only captures the key characteristics of the radio link layer.

3.2. Application Layer

The video streaming application compresses, packetizes, and schedules the data for transmission. The key parameters to be abstracted for the cross-layer optimization are related to the characteristics of the compressed source data. For a formal description, let us define the set $\mathcal{A}=\{\mathbf{a}_1,\mathbf{a}_2,...\}$ of tuples $\mathbf{a}_i=(a_i^1,a_i^2,...)$ of application layer specific parameters a_i^j. Since these application layer specific parameters may be variable, the set $\mathcal{A}$ contains all possible combinations of their values and each tuple $\mathbf{a}_i$ represents one possible combination. We further define the set $\tilde{\mathcal{A}}=\{\tilde{\mathbf{a}}_1,\tilde{\mathbf{a}}_2,...\}$ of tuples $\tilde{\mathbf{a}}_i=(\tilde{a}_i^1,\tilde{a}_i^2,...)$ of abstracted parameters $\tilde{a}_i^j$. We call the mapping between $\mathcal{A}$ and $\tilde{\mathcal{A}}$ *application layer parameter abstraction*

The abstracted information fed to the cross-layer optimizer in this work is the encoding distortion and the distortion profile for lost frames. Figure 2 shows an example of the distortion profile of lost frames and the encoding distortion for 3 different videos, each of which is composed of a group of pictures (GOP) with 15 frames, which corresponds to 0.5 seconds at a frame rate of 30 frames per second. The video sequences are encoded at a mean data rate of 100 kbps. Each GOP starts with an independently decodable Intra-frame. The following 14 frames are Inter-frames, which can only be successfully decoded if all previous frames of the same GOP are decoded error-free. The distortion is quantified by the mean squared reconstruction error (MSE), which is measured between the displayed and the original video sequence. The index in Figure 2 indicates the loss of a particular frame. It is assumed that as part of the error concealment strategy all following frames of the GOP are not decodable and the most recent correctly decoded frame is displayed instead of the non-decoded frames. The index 16 gives the MSE when all frames are received correctly, which we refer to as the encoding distortion. As expected, losing the first frame (I-frame) has the most dramatic influence on the reconstruction quality. Losing the last P-frame (index 15) of a GOP leads to very little increase in distortion in comparison to the error-free case.

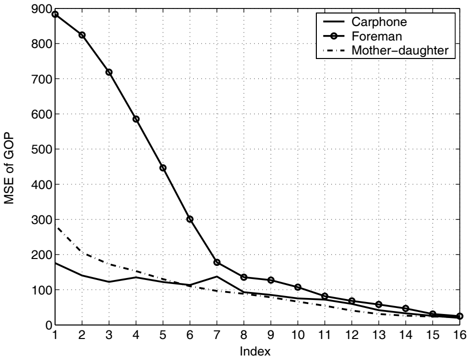

Figure 2: Distortion profile for lost frames for a GOP for three different videos.

4. CROSS-LAYER OPTIMIZATION

The abstracted parameter sets ($\tilde{\mathcal{R}}$ and $\tilde{\mathcal{A}}$) from both the application layer and the radio link layer form the input to the cross-layer optimizer. Since any combination of the abstracted parameter tuples from the two input sets is valid, it is convenient to define the cross-layer parameter set

$$\tilde{\mathcal{X}} = \tilde{\mathcal{R}} \times \tilde{\mathcal{A}} \tag{2}$$

which combines the two input sets into one input set for the optimizer. The set $\tilde{\mathcal{X}} = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots\}$ consists of tuples $\tilde{\mathbf{x}}_n = (\tilde{\mathbf{r}}_i, \tilde{\mathbf{a}}_j)$ and $|\tilde{\mathcal{X}}| = |\tilde{\mathcal{R}}| \cdot |\tilde{\mathcal{A}}|$. The optimizer selects from the input set $\tilde{\mathcal{X}}$ a true non-empty subset $\hat{\mathcal{X}}$ that is the output of the optimizer. In the following, we assume $|\hat{\mathcal{X}}| = 1$, that is the output of the optimizer is a single tuple and $\hat{\mathcal{X}} = \tilde{\mathbf{x}}_{opt} \in \tilde{\mathcal{X}}$. The decision or output of the cross-layer optimizer $\tilde{\mathbf{x}}_{opt}$ is made with respect to a particular objective function

$$\Gamma : \tilde{\mathcal{X}} \rightarrow \Re \tag{3}$$

where $\Re$ is the set of real numbers. Therefore, the output of the optimizer can be expressed as

$$\tilde{\mathbf{x}}_{opt} = \arg \min_{\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}} \Gamma(\tilde{\mathbf{x}}) \tag{4}$$

The choice of a particular objective function Γ depends on the goal of the system design and the output (or decision) of the optimizer might be different for different objective functions. In the example application of streaming video, one possible objective function in a single user scenario is the MSE between the displayed and the original video sequence. For a multi-user situation, different extensions of the MSE are possible. For example, the objective function can be the sum of MSE of all the users. That is,

$$\Gamma(\tilde{\mathbf{x}}) = \sum_{k=1}^{K} \mathrm{MSE}_k(\tilde{\mathbf{x}}) \tag{5}$$

where $\mathrm{MSE}_k(\tilde{\mathbf{x}})$ is the MSE of user k for the cross-layer parameter tuple $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$. This objective function will optimize the average performance among all users. Another useful definition of the objective function

$$\Gamma(\tilde{\mathbf{x}}) = \max_{k=1,\ldots,K} \mathrm{MSE}_k(\tilde{\mathbf{x}}) \tag{6}$$

optimizes the performance of the worst performing user.

5. NUMERICAL RESULTS

In this section, we provide simulation results to evaluate the performance of the proposed joint optimization concept. Throughout this section, we assume 3 users (user 1, 2, and 3), each of which requests a different video. User 1, 2, and 3 request the *Carphone*, *Foreman*, and *Mother-daughter* video, respectively. We choose the peak-signal-to-noise ratio (PSNR) as our performance measure. We use the objective function given in (6), which maximizes the worst-case user's performance. The cross-layer optimizer chooses the parameter tuple that maximizes the minimum of the PSNR among the users.

In the simulation, it is assumed that the data packet size at the radio link layer is equal to 54 bytes, which is the same as the specified packet size of the IEEE802.11a or HiperLAN2 standard. The channel coherence time is assumed to be 50 ms for all the three users, which approximately corresponds to a pedestrian speed (for 5 GHz carrier frequency). Since the transmission data rate is influenced by the modulation scheme, the channel coding, and the multi-user scheduling, two different modulations (BPSK and QPSK) are assumed and it is further assumed that there are 7 cases of air time arrangement in a time-division multiplexing based multi-user scheduling as shown in Table 1.

Table 1: Seven cases of air time arrangement.

	case 1	case 2	case 3	case 4	case 5	case 6	case 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

A user's transmission data rate is assumed to be equal to 100kbps when BPSK is used and 2/9 of the total transmission time is assigned to it. Therefore, if QPSK is used and 4/9 of the total transmission time is assigned, the user can have a transmission data rate as high as 400kbps. The transmission error rate on the other hand depends on the transmission data rate, the average SNR and the error correcting capability of the channel code. Usually, the performance of a channel code is evaluated in terms of the residual error rate (after channel decoding) for a given receive SNR. In our simulation, we assume a convolutional code of code rate ½ and a data packet size of 54 bytes. The residual packet error rate is a function of SNR. However, in the wireless link, the receive SNR is not constant, but fluctuating around the mean value (long term SNR), which is due to fast fading caused by user mobility. In this way, the receive SNR can be modeled as a random variable with a certain probability distribution, which is determined by the propagation property of the physical channel (e.g., Rayleigh distribution, Rice distribution). The residual packet error rate in a fading wireless link is computed by averaging the packet error ratio with the fading statistics. The resulting average packet error rate is used as the parameter e in (1) in our simulation. User position dependent path loss and shadowing commonly observed in wireless links are taken into account by choosing the long-term average signal-to-noise ratio randomly and independently for each user uniformly within the range from 1 to 100 (0 dB to 20 dB).

On the application layer, it is assumed that the video is encoded using the emerging H.264 video compression standard with 15 frames per GOP. The video sequences have been pre-encoded at two different target rates (100 kbps and 200 kbps) and both versions are stored on the streaming server. We can switch from one source stream to the other at the beginning of a GOP. In each GOP, the first frame is an I-frame and the following 14 frames are P-frames. We use the measured distortion profile of a particular lost frame and the encoding distortion for the 3 requested videos as shown in Figure 2. It is assumed that each video frame (or picture) is packetized with maximum size of 54 bytes and each packet only contains data from one frame. Figure 3 provides simulation results of three scenarios. In scenario 1, we restrict that only BPSK modulation is used at the radio link layer and only the source rate with 100kbps is available at the application layer. Therefore, only one constant abstracted parameter tuple (with 100 kbps for all 3 users) is provided by the application layer (i.e., $|\tilde{\mathcal{A}}|=1$) in this scenario, while the radio link layer provides 7 abstracted parameter tuples (i.e., $|\tilde{\mathcal{R}}|=7$), which results from the 7 cases of air time arrangement shown in Table 1. The cross-layer optimizer selects one out of the 7 combinations of the input parameter tuples ($|\tilde{\mathcal{X}}|=|\tilde{\mathcal{R}}|\cdot|\tilde{\mathcal{A}}|=7$) such that our objective function given in (6) is optimized. Please note that the seven different cases of air time arrangement in Table 1 offer the possibility to send data packets more than once for some of the users. This improves the chances to get the important data over the wireless channel. As the K different users see different channel qualities, the 7 different cases of air time arrangement allow us to optimize the resource allocation such that our objective function is optimized.

The MSE of the reconstructed video is a random variable controlled by the two factors discussed above, namely fast fading and user position dependent path loss and shadowing. In general, fast fading takes place in a much smaller time scale than the path loss and shadowing. In this paper, we evaluate the MSE averaged over fast fading by taking the expected value of the MSE with respect to the fast fading for a particular position of the users or equivalently for a particular long term SNR. Based on this value the cross-layer optimizer makes its decision. We also look at its statistical properties for an ensemble of user positions. Therefore, the cumulative density probability function (CDF) of this average MSE is chosen to show the performance. The simulations are performed for two modes, the first one working without a retransmission of lost packets (*Forward Mode)* and a second one with retransmission of lost packets (*ARQ Mode*). The performance of the worst performing user in the system with the proposed joint optimization (w/ JO) is compared with that in a system without joint optimization (w/o JO). The performance gain is shown in terms of $\Delta PSNR$. A system without joint optimization is assumed to assign the same amount of transmission time to all the users (i.e., Case 1 in Table 1) and use BPSK modulation, while the source data rate is fixed to 100 kbps. It can be seen from Figure 3 that the PSNR of the worst performing user improves significantly in the system w/ JO. For instance, there is about 50% of the chance that the PSNR of the worst performing user is improved by at least 1dB in the system w/ JO in Forward Mode.

A similar trend of improvement can be observed for scenario 2 and 3. In scenario 2, the same abstracted parameter tuple as in scenario 1 is assumed at the application layer but the radio link layer provides 14 abstracted parameter tuples, which result from the 7 cases of time arrangement with BPSK and another 7 cases of time arrangement with QPSK. For scenario 2, an improvement of 4 dB or more is observed in 50% of the cases. In scenario 3, it is assumed that the two different source rates of 100 kbps and 200 kbps for each of the 3 users are provided by the application layer (resulting in $2^3 = 8$ parameter tuples). The same abstracted parameter tuples as in scenario 2 are provided by the radio link layer. The performance improvement for scenario 3 is almost identical with scenario 2. This means that the additional freedom of sending either 100kbps or 200kbps video is not selected by the cross-layer optimizer. For the ARQ mode it can be observed that the largest performance gain is achieved for scenario 3 which offers the largest number of degrees of freedom to the cross layer optimizer.

From these experiments it can be observed that choosing a suitable set of abstracted parameters tuples is important in order to obtain large performance improvements while optimizing at low complexity. Also, the experiments show that it is important to identify all degrees of freedom that are available on the individual layers and to consider the important ones in the cross-layer design. Our experiments show that even for a small number of users and a small number of degrees of freedom significant quality improvements can be obtained by our cross-layer optimization concept for the considered wireless multi-user video streaming scenario.

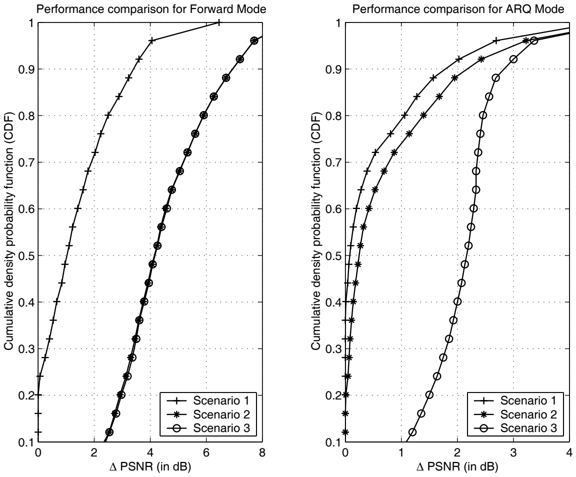

Figure 3: Performance improvements obtained with cross-layer optimization for three different scenarios.

6. REFERENCES

[1] S. Shakkottai, and T. S. Rappaport, "Research challenges in wireless networks: a technical overview," *Proc. of the Fifth International Symposium on Wireless Personal Multimedia Communication*, Honolulu, HI, Oct. 2002.

[2] T. Holliday, and A. Goldsmith, "Optimal power control and source-channel coding for delay constrained traffic wireless channels," *The IEEE International Conference on Communications 2003*, Anchorage, Alaska, USA, May 2003.

[3] Y. Shan, and A. Zakhor, "Cross layer techniques for adaptive video streaming over wireless networks", in *Proc. IEEE ICME*, August 2002.

[4] S. Krishnamachari, M. V. D. Schaar, S. Choi, and X. Xu, "Video streaming over wireless LANs: a cross-layer approach," *The 13th International Packetvideo Workshop 2003*, Nantes, France, April 2003.

[5] 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.

[6] M. T. Ivrlac, and F. Antreich, "Cross OSI layer optimization – an equivalence class approach," *Technical Report TUM-LNS-TR-03-09*, Institute for Circuit Theory and Signal Processing, Munich University of Technology, May 2003.