Bottom-Up Approach to Cross-layer Design for Video Transmission over Wireless Channels

Lai-U Choi[1] Michel T. Ivrlač[2] Eckehard Steinbach[1] and Josef A. Nossek[2]

[1]Institute of Communication Networks
Media Technology Group
Munich University of Technology
80290 Munich, Germany
email: {laiuchoi, eckehard.steinbach}@tum.de

[2]Institute for Circuit Theory and Signal Processing
Wireless Communications
Munich University of Technology
80290 Munich, Germany
email: {ivrlac,nossek}@tum.de

***Abstract*— In this paper[1], a bottom-up approach to cross-layer design is proposed which is based on the partitioning of the optimized parameters into private and interfacing parameters. The optimization is divided into an intra-layer optimization of private parameters and an inter-layer optimization of interfacing parameters. This concept keeps as much of the optimization local to the layers as possible and thereby reduces the amount of additional information exchange and associated delay times caused by cross-layer design. Intra- and inter-layer optimization are illustrated for a simplified multi-user wireless communication system which offers a streaming video service application to its mobile clients.**

I. Introduction

Highly efficient use of available system resources, like bandwidth and power, is emerging as a paramount requirement in design of future wireless communication systems. This is partly due to the ever higher demands of new services, such as streaming or conversational applications. These increasing demands may be difficult to meet by a system design which separates the functionality into essentially independent layers. Cross-layer design is a promising way to increase the efficiency of the use of available resources [1]. This is due to inter-layer information exchange.

In general, there are two fundamental approaches: "top-down" and "bottom-up" cross-layer design. The "top-down" approach is similar to ordering custom made furniture at a carpenter shop. The application is communicating more or less precise specification of its requirements downwards in the protocol stack, while the lower layers try to fulfill the given requirements with minimum cost [2]. In the "bottom-up" approach, the carpenter shop is offering a selection of furniture for the customer to choose from. Likewise, the lower layers are communicating upwards in the protocol stack, sets of system parameters which can be implemented with a given cost. The application is free to choose the most suitable system parameters from the set. While the "bottom-up" approach is maximizing quality of service for a given cost, the "top-down" approach is minimizing cost for a given quality. In this paper, we focus on the "bottom-up" approach.

In order to minimize the amount of additional information flow caused by cross-layer optimization, we propose to divide the system parameters into two groups: private and interfacing parameters. The private parameters (operating modes) are not visible to other layers and optimized locally by an *intra-layer* optimization. Only interfacing parameters (operating points) are exchanged between the layers and subject to an *inter-layer* optimization. We concentrate on the problem of transmission of compressed streaming video over wireless lossy channels in a multi-user environment. In this way, it is necessary to include parts of the physical (PHY), media-access control (MAC) and application layer into the optimization process. The physical layer interacts with the wireless channel directly and has to cope with its impairments by proper signal processing and channel coding, while the media-access control layer takes care of sharing the resources between the multiple users of the system. Finally, end to end quality of service can only be defined in the application layer.

After an introduction to the proposed bottom-up approach in Section II, the intra-layer optimization is presented for the PHY/MAC layers in Section III, optimizing air-time and transmit power. Inter-layer optimization is discussed in Section IV. It is based on sophisticated models for the video distortion caused by compression with the H.264/AVC video encoding standard and by loss of frames due to transmission errors. The optimization is illustrated by sample results in Section V.

II. Cross-layer Design Concept

Cross-layer design aims at optimizing the overall system performance, which requires an inter-layer information exchange about the specific needs and capabilities of the processing implemented in the different layers of the communication system. In order to minimize the amount of additional information flow caused by cross-layer design, it is desirable to keep as much of the optimization local to the layers as possible. This necessitates to divide the optimization parameters into two groups [2]:

- private parameters (operating modes)
- interfacing parameters (operating points)

The private parameters are local to the processing of a given layer and not visible to other layers. We refer to these private parameters as *operating modes*. They take part in cross-layer optimization only in the sense of an *intra-layer* optimization as illustrated in Figure 1. That is, these operating modes are optimized within each layer. On the other hand, the interfacing parameters are visible to other layers and subject to *inter-layer* optimization. We refer to these interfacing parameters as *operating points*. In order to accomplish inter-layer optimization,

[1]This work is supported partially by an Alexander von Humboldt (AvH) Research Fellowship.

0-7803-8887-9/05/$20.00 (c)2005 IEEE

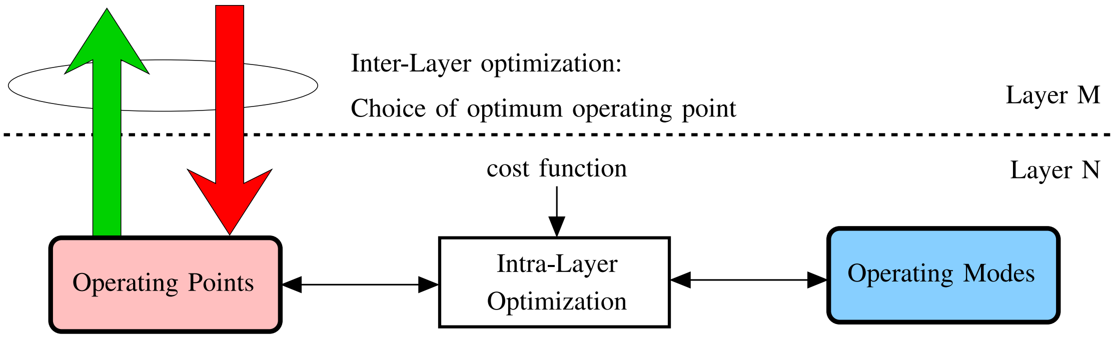

Fig. 1. Intra- and inter-layer optimization

information about these operating points has to be exchanged between the layers. Intra- and inter-layer optimization have to be matched to each other, such that either

- a given quality of service is provided with minimum cost, or
- the quality of service is maximized for a given cost.

The former is called *top-down*, while the latter is referred to as the *bottom-up* cross-layer optimization. Depending on which of these two approaches is used, the functions of intra- and inter-layer optimization change [3]. It is also possible to combine both approaches [4].

III. Intra-Layer Optimization

The purpose of the intra-layer optimization is to establish a mapping between operating points and operating modes, which is optimum in the sense of a defined cost function. In the PHY layer, costs can for instance be defined in terms of necessary transmit power. In the bottom-up approach, the intra-layer optimization establishes a so-called equivalence class $\mathcal{P}_c$ of operating points which all lead to the same cost c [2]. Moreover, it maps each operating point out of $\mathcal{P}_c$ to a *cost-efficient* operating mode. An operating mode that implements an operating point at cost c is called cost-efficient, if and only if there are no other operating modes that can implement this operating point at lower cost. Therefore, the result of the bottom-up intra-layer optimization is

- a cost-efficient mapping between operating points (interfacing parameters) and operating modes (private parameters)
- an equivalence class $\mathcal{P}_c$ of operating points which can be implemented cost-efficiently with a given cost c.

This equivalence class $\mathcal{P}_c$ is communicated upwards to the higher layer, which thereafter can initiate an inter-layer optimization, as will be discussed in the next Section.

Let us have a look at an illustrative example for a bottom-up intra-layer optimization for the simplified MAC/PHY layer processing in a wireless communication system as shown in Figure 2. Two mobile users of a wireless communication system are served by their base-station in time division multiple access (TDMA)[2]. The transmission data of the users are protected against influence of channel noise by forward error control (FEC) encoding. The encoded signals of the two users are transmitted in packets with powers $P_{\mathrm{T},1}$ and $P_{\mathrm{T},2}$, respectively. After transmission over the wireless channels, the received data are decoded with a residual error measured in terms of the packet error probabilities PEP_1 and PEP_2 for the two users, respectively. The information rates R_1 and R_2 and the corresponding packet error probabilities are visible to higher layers. Therefore, the tuple

$$(R_1, R_2, \mathrm{PEP}_1, \mathrm{PEP}_2)$$

forms an operating *point*. Due to the time varying nature of the wireless channel, the packet errors occur in bursts. To simplify analysis, we assume a stochastic block fading channel model. Thereby, the channel is described by its channel capacity which is modelled as constant for a time duration T_{dec} (decorrelation time [3]) and then abruptly changes to an independent random realization. The average burst length of packet errors is proportional to the decorrelation time and given by $T_{\mathrm{dec}}/\mathrm{PEP}$. Since the burstiness may influence the quality of service in higher layers, the decorrelation time has to be made known to higher layers in addition to the operating points.[3]

The TDMA scheme assigns the relative airtime $\alpha \in]0;1[$ to the first user, leaving a fraction of $(1-\alpha)$ of the air-time for the second user. Let us call

$$\beta = P_{\mathrm{T},1}/P_{\mathrm{T},2} \tag{1}$$

the ratio of the respective transmit powers of the two users. The set of operating *modes* can then be given by the collection of all possible ordered pairs (α, β), which is kept private to the PHY/MAC layers.

The bottom-up intra-layer optimization now has to find out a cost-efficient mapping between $(R_1, R_2, \mathrm{PEP}_1, \mathrm{PEP}_2)$ and (α, β). We define cost in terms of average transmit power

$$\mathrm{cost} = \overline{P_{\mathrm{T}}} = \alpha P_{\mathrm{T},1} + (1-\alpha) P_{\mathrm{T},2}. \tag{2}$$

[2]In this paper, we use TDMA for its conceptual simplicity. The design approach can however be directly applied to other multiple access techniques, such as OFDMA, CDMA or SDMA.

[3]In this context, T_{dec} is called a *side-effect* [2].

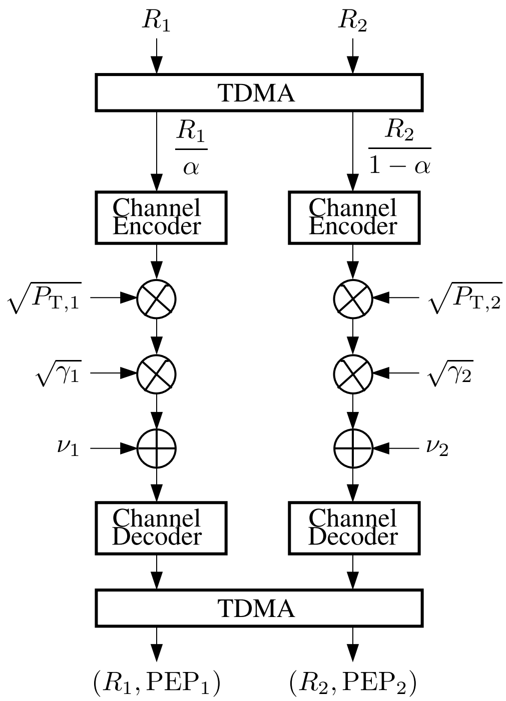

Fig. 2. MAC and PHY layer processing

For the sake of simplicity, let us assume that the wireless links can be modelled as Rayleigh fading AWGN channels. Their channel capacities C_i during a decorrelation time are then given by

$$C_i = B \cdot \log_2\left(1 + \frac{P_{\mathrm{T},i}}{\sigma_{\mathrm{n},i}^2}\gamma_i\right), \tag{3}$$

where γ_i with $i \in \{1,2\}$ are i.i.d. exponentially distributed random variables, which we normalize to $\mathrm{E}[\gamma_i] = 1$. Hence, the probability density function (pdf) of γ_i is given by

$$\mathrm{pdf}_{\gamma_i}(\gamma_i) = \begin{cases} \exp(-\gamma_i) & \text{for} \quad \gamma_i \geq 0 \\ 0 & \text{else} \end{cases}. \tag{4}$$

The radio bandwidth and receiver noise powers are denoted by B and $\sigma_{\mathrm{n},i}^2$, respectively. For capacity approaching channel coding, a decoding error occurs only if the instantaneous capacity C_i falls below the rate at which information is transfered over the i-th AWGN channel. This assumes that the duration of a codeword is less than the decorrelation time (no interleaving). The packet error probability for the first channel is therefore given by

$$\mathrm{PEP}_1 = \Pr[C_1 < R_1/\alpha]. \tag{5}$$

If the codewords are spread over m decorrelation times (interleaving), the corresponding expression would be given by $\mathrm{PEP}_1 = \Pr[\sum_{i=1}^{m} C_{1,i} < m \cdot R_1/\alpha]$ instead, where the $C_{1,i}$ are the channel capacities of the first channel at successive decorrelation times. In the following we will assume $m = 1$ for simplicity. Due to the exponential distribution of γ_1 in (4), it follows from (3) and (5) that the transmit power

$$P_{\mathrm{T},1} = \sigma_{\mathrm{n},1}^2 \cdot \frac{1 - 2^{R_1/(B\alpha)}}{\log_{\mathrm{e}}(1 - \mathrm{PEP}_1)} \tag{6}$$

is necessary to establish a net rate of R_1 with a packet error probability of PEP_1 for the first user, given a particular relative airtime α. With (2), the transmit power available for the second user then becomes

$$P_{\mathrm{T},2} = \frac{\overline{P_{\mathrm{T}}} - \alpha P_{\mathrm{T},1}}{1-\alpha}. \tag{7}$$

If there is at least one $\alpha \in]0;1[$, such that $0 < P_{\mathrm{T},2} < \infty$, the pair (R_1, PEP_1) can be implemented. All implementable packet error probabilities for the second user are elements of the set

$$\mathcal{S} = \{\ \Pr[C_2 < R_2/(1-\alpha)] \mid \alpha \in]0;1[\ \}, \tag{8}$$

where

$$\Pr[C_2 < R_2/(1-\alpha)] =$$

$$1 - \exp \frac{\sigma_{\mathrm{n},2}^2 \left(1 - 2^{R_2/(B(1-\alpha))}\right)(1-\alpha)}{\overline{P_{\mathrm{T}}} - \sigma_{\mathrm{n},1}^2 \dfrac{1 - 2^{R_1/(B\alpha)}}{\log_{\mathrm{e}}(1 - \mathrm{PEP}_1)} \cdot \alpha}\ . \tag{9}$$

The cost-efficient value of PEP_2 is then given by

$$\mathrm{PEP}_2 = \min_{p \in \mathcal{S}} p, \tag{10}$$

since any lower value of PEP_2 would require an increase in average transmit power, and hence increase in cost. The cost-efficient mapping between operating points and operating modes is then given by setting $\alpha = \alpha_{\mathrm{opt}}$, where α_{opt} is the relative airtime correspoding to the cost-efficient implementation of PEP_2. Its value can be determined numerically by minimizing the expression in (9). With (6), (7) and the optimum airtime α_{opt}, the transmit power ratio β from (1) has the corresponding optimum value given by:

$$\beta_{\mathrm{opt}} = \frac{1 - \alpha_{\mathrm{opt}}}{\dfrac{\overline{P_{\mathrm{T}}}}{\sigma_{\mathrm{n},1}^2} \dfrac{\log_{\mathrm{e}}(1 - \mathrm{PEP}_1)}{1 - 2^{R_1/(B\alpha_{\mathrm{opt}})}} - \alpha_{\mathrm{opt}}}. \tag{11}$$

This completes the intra-layer optimization. To this end, note that the cost-efficient equivalence class of operating points $\mathcal{P}_{\overline{P_{\mathrm{T}}}}$ is formed by all tuples $(R_1, R_2, \mathrm{PEP}_1, \mathrm{PEP}_2)$ which are implementable with average transmit power $\overline{P_{\mathrm{T}}}$, where PEP_2 is chosen according to (10). A tuple is implementable if $P_{\mathrm{T},2}$ from (7) is finite and positive for $\alpha = \alpha_{\mathrm{opt}}$. Finally, note that the resulting individual transmit powers can be computed from the average transmit power and the optimum operating mode:

$$P_{\mathrm{T},2} = \frac{\overline{P_{\mathrm{T}}}}{1 + \alpha_{\mathrm{opt}}(\beta_{\mathrm{opt}} - 1)}, \quad \text{and} \quad P_{\mathrm{T},1} = \beta_{\mathrm{opt}} P_{\mathrm{T},2}. \tag{12}$$

IV. Inter-layer Optimization

The purpose of inter-layer optimization is to select the optimal operating point $\mathrm{OP}_{\mathrm{opt}} \in \mathcal{P}_c$ out of the equivalence class $\mathcal{P}_c$ of cost-efficient operating points, which is communicated from the lower layer. The optimization criterion is based on maximizing quality of service (QoS). A QoS objective function ($Q : \mathcal{P}_c \rightarrow \mathcal{R}$) is defined on the set of cost-efficient operating points.[4] The optimum operating point is given by

$$\mathrm{OP}_{\mathrm{opt}} = \arg\max_{p \in \mathcal{P}_c} Q(p). \tag{13}$$

The QoS objective function depends on the service application. In the following, we use a streaming video service as an example application. Let us briefly discuss the main features of the streaming video that are relevant in our example. The video is encoded using a standard video compression scheme H.264/AVC, and the corresponding video stream is stored on the streaming server. When the stream is requested by the client, it is packetized and sent to the user. The number of bits to be sent for each video frame depends on what kind of encoding mode has been selected. We distinguish so-called I-frames which are encoded without reference to previous frames, and P-frames that are encoded by forming a prediction from previous frames. While I-frames can be decoded without receiving the previous frames, P-frames typically can not. In order to allow fast forward and interactive scene selection, I-frames are typically introduced every 0.5 to 1.0 seconds. An I-frame and all of the following P-frames up to and exluding the next I-frame are referred to as a group of pictures (GOP). In the following, we will focus on a IPP$\cdots$P - structure, where each GOP consists of one I-frame followed by $(F-1)$ P-frames. Calling T_{GOP} the duration of a GOP, we have a frame rate of F/T_{GOP} frames per second. This encoding structure is prone to error propagation due to inter-frame dependencies introduced by the predictive encoding. A typical measure to mitigate the effect of frame loss is error concealment. In the simplest, yet commonly used way, an incorrectly decoded frame and all subsequent frames in the GOP are replaced by the most recently correctly received frame. In this paper, we assume this type of error concealment, which is referred to as *previous frame error concealment*.

There are two major effects that influence the quality of the displayed streaming video:

- Source distortion
- Packet loss distortion .

The source distortion is due to lossy compression, which is a result of fairly high compression ratios needed for streaming video. Even when the source decoder is connected directly to the source encoder, the decoded video stream is different from the original. The difference between those two streams is called *source distortion* and is usually measured by the mean square error D_{S}, or the *peak signal to noise ratio*

$$\mathrm{PSNR}_{\mathrm{S}} = 10 \cdot \log_{10} \frac{255^2}{D_{\mathrm{S}}}. \tag{14}$$

The source distortion depends on the data rate R that is generated by the source encoder. The relationship between the rate and the distortion is defined by the *distortion-rate* (D-R) function. The D-R function also depends on the content of the video sequence. An accurate sequence-level D-R model is developed in [5] as:

$$\mathrm{PSNR}_{\mathrm{S}}(R) = a + b \cdot \sqrt{\frac{R}{c}}\left(1 - \frac{c}{R}\right), \tag{15}$$

where a, b, and c are sequence specific coefficients which are determined from three pairs of measured rate and distortion.

The loss distortion on the other hand is due to transmission errors, which lead to inter-frame error propagation. Due to previous frame error concealment and the IPP...P-structure, the average distortion D_i introduced by a lost frame depends only on the number $i \in \{0, 1, \ldots, F-1\}$ of the *first* frame which cannot be decoded correctly in a GOP. Here $i = 0$ refers to the I-frame of the GOP. Calling P_i the probability that this happens, the loss distortion D_{L} is defined as:

$$D_{\mathrm{L}} = \sum_{i=0}^{F-1} D_i \cdot P_i, \tag{16}$$

which is the *expected distortion* due to frame losses concealed by previous frame error concealment. In the following, we use an analytical model for D_i and P_i which is developed in [6]:

$$D_i = (F-i) \cdot \frac{F \cdot i \cdot D_{\min} + (F-i-1) \cdot D_{\max}}{(F-1)F}. \tag{17}$$

The values $D_{\max}$ and $D_{\min}$ depend on the video sequence and are determined by measurement. The event probabilities P_i are derived in [6] as:

$$P_i = \begin{cases} (1-\mathrm{PEP})\left[\mathrm{e}^{-\gamma_{i-1}\mathrm{PEP}} - \mathrm{e}^{-\gamma_i \mathrm{PEP}}\right], & i \geq 1 \\ 1 - (1-\mathrm{PEP})\ \mathrm{e}^{-\gamma_0 \mathrm{PEP}} & , \ i = 0 \end{cases}, \tag{18}$$

where

$$\gamma_i = \frac{T_{\mathrm{GOP}}}{T_{\mathrm{dec}}} \cdot \frac{A+i}{F+A-1}. \tag{19}$$

Herein A is the ratio of the average number of bits encoded in an I-frame to the average number of bits encoded in P-frames in a video sequence. Since I-frames are bigger than P-frames, we have $A > 1$. The overall distortion is given by [7]:

$$D_{\mathrm{tot}} = D_{\mathrm{S}} + D_{\mathrm{L}}, \tag{20}$$

where D_{S} is the source distortion which can be obtained from (15), while D_{L} is the loss distortion from (16), (17) and (18). Therefore, the QoS objective function for the streaming video service can be set to

$$Q = 10 \log_{10} \frac{255^2}{D_{\mathrm{tot}}}. \tag{21}$$

For multiple users (say K), we can define

$$Q = f(Q_1, Q_2, \ldots, Q_K), \tag{22}$$

where the Q_i with $i \in \{1, 2, \ldots, K\}$ are the QoS objective functions for the individual users given by (21) and f is a suitable combining function. A possible choice for f is for instance the minimum function, i.e.

$$f(Q_1, Q_2, \ldots, Q_K) = \min\left(Q_1, Q_2, \ldots, Q_K\right), \tag{23}$$

which leads with (13) to maximization of the worst-case user performance (minimax approach).

[4]The symbol $\mathcal{R}$ refers to the set of real numbers.

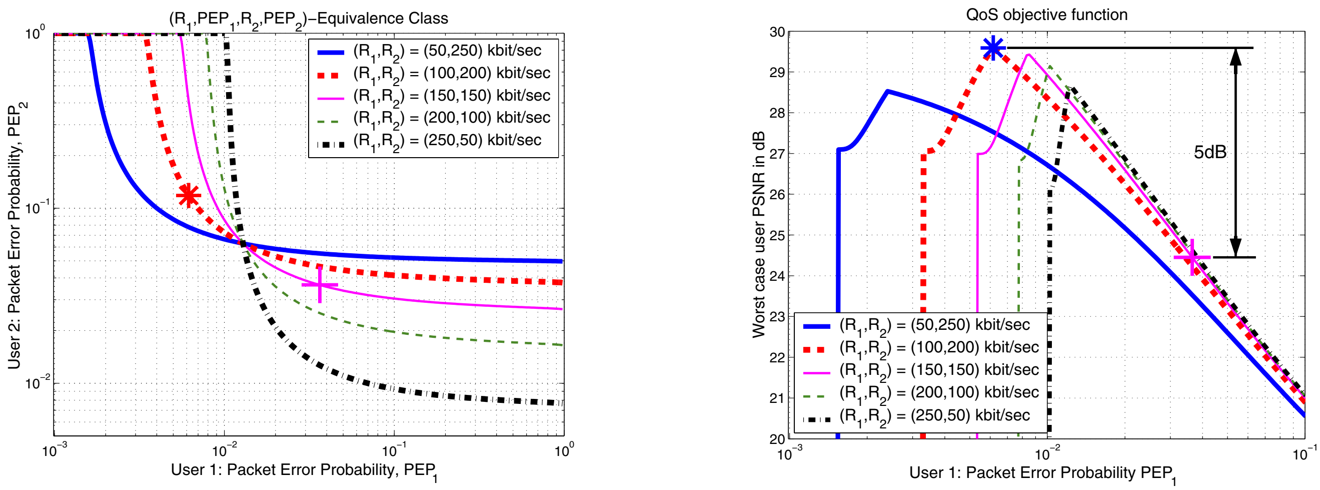

Fig. 3. *Left*: Cost-efficient equivalence class $\mathcal{P}_{\overline{P}_{\mathrm{T}}}$ of operating points. *Right*: QoS objective function (worst-case user PSNR).

V. Sample Results

In this Section, we illustrate the proposed bottom-up intra- and inter-layer optimizations by a concrete example. Two different video test sequences are transmitted to the users ("Foreman" (FM) for the first and "Mother&Daughter" (MD) for the second user). Both sequences use $F = 15$ frames per GOP which has a duration of $T_{\mathrm{GOP}} = 0.5$ seconds (30 frames per second). The following model parameters for the source and the loss distortion were obtained from measurement:

	a	b	c	$D_{\min}$	$D_{\max}$	A
FM	36.3	4.65	200kbps	15	1175	6.07
MD	45.0	4.65	449kbps	0.87	123	12.3

where a, b, and c are the coefficients used in the D-R function from (15), while $D_{\min}$, $D_{\max}$ and A are the corresponding coefficients for the loss distortion as used in (17) and (19). The decorrelation time is set to $T_{\mathrm{dec}} = 0.055$ seconds, which corresponds to a velocity of about v=4 km/h at a carrier frequency of 2GHz. The video sequences are transmitted by the system depicted in Figure 2. We set the bandwidth to $B = 250$kHz and let $\sigma^2_{\mathrm{n},1} = 1$, $\sigma^2_{\mathrm{n},2} = 5$ and $\overline{P_{\mathrm{T}}} = 100$. The results from the intra- and inter-layer optimization are shown in Figure 3. The left hand side displays the cost-efficient equivalence class $\mathcal{P}_{\overline{P}_{\mathrm{T}}}$ of operating points $(R_1, R_2, \mathrm{PEP}_1, \mathrm{PEP}_2)$. For given pairs (R_1, R_2) the lowest achievable PEP_2 (see (10)) is shown as a function of PEP_1. Decrease of PEP_1 always leads to an increase in PEP_2. Based on the equivalence class $\mathcal{P}_{\overline{P}_{\mathrm{T}}}$, the quality of service objective function Q from (22) and (23) is maximized by inter-layer optimization. The right hand side of Figure 3 shows the respective values of Q for different operating points from $\mathcal{P}_{\overline{P}_{\mathrm{T}}}$. The optimum operating point is indicated by the "*"-mark. It turns out that this optimum operating point corresponds to the optimum operating mode $(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}}) = (0.46, 1.88)$. To get an idea about the gain in QoS which is obtained by cross-layer optimization, we also look at the case where the operating mode is set to R_1=R_2 and $\mathrm{PEP}_1 = \mathrm{PEP}_2$. This setting provides both users with equal transmission properties. If nothing is known about the service application, this may be a reasonable choice. The results are shown by the "+"-mark in Figure 3. The gain obtained by the cross-layer optimization is about 5 dB in terms of worst-case user PSNR, which corresponds to a 70% reduction in the mean square error distortion.

VI. Conclustion

An approach to cross-layer design consisting of an intra-layer optimization of operating modes and an inter-layer optimization of operating points is presented. The intra-layer optimization results in a cost-efficient equivalence class of operating points which is communicated to the higher layer. From this equivalence class, an optimum operating point is chosen by inter-layer optimization which aims at maximizing quality of service.

VII. Acknowledgement

The authors would like to express their sincere thanks to the Alexander von Humboldt (AvH) foundation for kindly supporting this research.

References

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

[2] M. T. Ivrlač, and J. A. Nossek, "Cross Layer Design - An Equivalence Class Approach", *Proc. IEEE International Symposium on Signals, Systems, and Electronics*, ISSSE-04, Linz, Austria, 2004.

[3] M. T. Ivrlač, "Wireless MIMO Systems - Models, Performance, Optimization", *Doctoral Dissertation*, Munich University of Technology, Munich, Germany, 2005.

[4] L. U. Choi, W. Kellerer, and E. Steinbach, "Cross-Layer Optimization for Wireless Multi-User Video Streaming", *Proc. IEEE International Conference on Image Processing*, ICIP 2004, Singapore, Oct. 2004.

[5] L. U. Choi, M. T. Ivrlač, E. Steinbach, and J. A. Nossek, "Sequence-Level Models for Distortion-Rate Behaviour of Compressed Video", *submitted to IEEE International Conference on Image Processing*, ICIP 2005.

[6] L. U. Choi, M. T. Ivrlač, E. Steinbach, and J. A. Nossek, "Analysis of Distortion Due to Packet Loss in Streaming Video Transmission over Wireless Communication Links", *submitted to IEEE International Conference on Image Processing*, ICIP 2005.

[7] K. Stuhlmüller, N. Färber, M. Link, 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.