# Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming[☆]

J. Gross[*], J. Klaue, H. Karl, A. Wolisz

*Telecommunication Networks Group, Technical University of Berlin, Einsteinufer 25, 10587 Berlin, Germany*

## Abstract

In this paper, we study the performance of a combined link- and physical-layer approach for the downlink of a dynamic OFDM-frequency division multiple access transmission system delivering MPEG-4 video streams. The approach consists of three different mechanisms—queue management, resource allocation, and subcarrier assignment—which can be used separately as well as in combination. This cross-layer approach allows to utilize channel-related knowledge and semantic information of the streams' packets. Consecutively adding these mechanisms to the studied system, we judge their respective impact on system performance. To do so, we present a new metric for quality assessment of long video transmissions (considerably longer than 10 s) and apply it to the simulation results. In particular, we introduce a local distortion threshold-based metric instead of an overall average judgment. We find that our combined system approach increases the number of supportable terminals per cell by up to 300%.
© 2004 Elsevier B.V. All rights reserved.

*Keywords:* Orthogonal frequency division multiplexing; Signal-to-noise ratio; Cross-layer optimization

## 1. Introduction

Users of modern telecommunication systems will expect the support of sophisticated services over wireless transmission. One particularly demanding example is video. Video transmission is characterized not only by a large required data rate, but also, when using up-to-date video encoding techniques, by a large variability of this required data rate over time, depending on the dynamics of a given scene. The data rate requirements are aggravated if multiple terminals demand video transmissions from a single access point, serving a given wireless cell.

The high data rate requirements of such applications require wireless transmission technologies that can support them. One especially promising approach here is orthogonal frequency division multiplexing (OFDM). Not only does it provide high data rates over a wireless channel, it also supports a notion of flexible resource sharing between multiple wireless terminals by dividing the wireless bandwidth into so-called 'subcarriers', which can be individually assigned to different terminals.

This multiplexing of subcarriers is primarily intended to combat the variability of the underlying wireless channel, but its flexible resource assignment can also be leveraged to support applications with varying resource requirements.

The question is hence whether these two sources of variability—the application and the wireless channel—can be handled in combination so that the system's performance characteristics are improved. For multimedia applications like video, these characteristics should not be measured in simple bits per second, but should take the user-experienced quality of a video transmission into account. Plausible figures of merit are therefore the user-experienced video quality for a given number of users, or the number of users that can be supported by a system at a given, minimally acceptable level of quality.

In this paper, we present three protocol mechanisms that are based on the usage of application- and channel-related information to influence the transmission of video streams from an access point to wireless terminals in a cell, with the goal of optimizing the above-mentioned figures of merit. The mechanisms work at link and medium access control (MAC) layer and can be used to some extent independently from each other. We study how they can be combined in a cross-layer optimization approach (Fig. 1). Cross-layer optimization in communications has been addressed by multiple studies in different scenarios [1,2]. The basic idea is
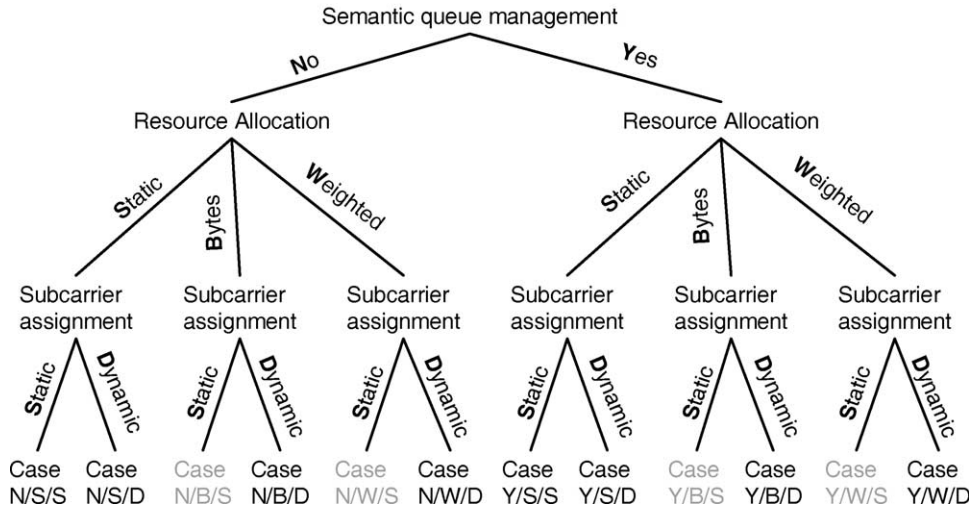
[*] Corresponding author.

Fig. 1. Overview of possible mechanism combinations (infeasible combinations are shown in light grey). Combinations are designated as 'semantic queue management type/resource allocation type/subcarrier assignment type'.

to design communication layers such that they can share and react to layer-specific information at different layers in the communication stack in order to improve transmission metrics.

Out of our three mechanisms, the one that works closest to the video application is *semantic queue management*. Suppose that video data packets are buffered in the access point in separate queues for each wireless terminal. Assuming that the access point has knowledge about the encoding scheme, it can decide which packets are more important than others and, e.g. give preference to important packets (like I-frames in MPEG-4 video) or drop less important packets (like B-frames) when transmission resources are scarce. In this sense, semantic information about the data in a given queue is taken into account.

The second mechanism, *resource allocation*, considers all queues in an access point and uses the amount of data in a queue relative to the other queues to decide how many resources, i.e. subcarriers, should be allocated to this terminal. The simplest version just uses the number of bytes in each queue; a more sophisticated version also considers some semantic knowledge and gives a larger weight to more important packets.

The third mechanism is *subcarrier assignment*. In a simple case, each terminal is assigned a given set of subcarriers. But as subcarriers can be differently attenuated with respect to multiple terminals, it makes sense to dynamically assign subcarriers such that each terminal has a set of subcarriers that are in a good or reasonable state; the sizes of these sets are computed by resource allocation.

We have hence 12 combinations of mechanisms at hand (including the trivial case for each of the three mechanisms),

only eight of which are reasonable (static subcarrier assignment cannot be combined with any dynamic resource allocation scheme).

We have considered some of these combinations already in other publications (using the terminology from Fig. 1, the cases N/S/S, N/S/D, N/B/D in Ref. [3], the cases N/S/S, N/S/D, Y/S/S, and Y/S/D in Ref. [4]). In this paper, we will complete these investigations and put the comparison of these combinations into perspective. Specifically, we will increase the complexity of the combinations from the lower layers towards the higher ones, representing an increasing amount of information becoming available about the semantic behavior of the video application—the particular combinations that we include here are shown in Fig. 2 in ascending order of complexity.

For this investigation, we shall proceed as follows: Section 2 describes the system model that we are using, Section 3 explains the three mechanisms in detail. Section 4 presents the performance evaluation, and Section 5 draws the conclusions and outlines future work.

## 2. System model

As system model we assume the following. $J$ wireless terminals move within a cell of a communication system of radius $R$. All data transmissions within this cell are managed by an access point. We focus on the downlink transmission direction, which is the one from the access point to the terminals. The access point receives data destined for terminals in its cell via a backbone. For transmission of this data to the terminals a total bandwidth $B$ is given with center
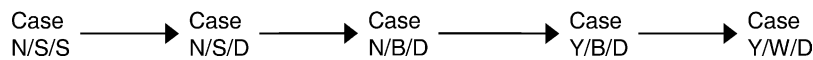


Fig. 2. Ascension of mechanisms described in this paper.

frequency $f_c$. As transmission scheme in the cell OFDM is employed, splitting the bandwidth into $S$ subcarriers.

Each terminal within the cell moves with a certain speed lower than $v_{max}$. Any transmission within the cell is subject to multi-path propagation where $\Delta\sigma$ characterizes the variance of the delay spread. Due to the multi-path propagation and the movement of the terminals, frequency- and time-selective fading constantly changes the subcarrier attenuations. In addition to fading, the sub-carrier attenuations are affected by shadowing and path loss, on larger time scales though (on the order of seconds, compared to milliseconds for fading). Subcarrier states of different wireless terminals are assumed to be statistically independent. The fading is assumed to be correlated both in time and in frequency where a Jakes-like power spectral density and an exponential power delay profile characterize the correlational structure of the fading.

For the downlink, we assume a dynamic OFDM-frequency division multiple access (FDMA) system with adaptive modulation, assigning each terminal a disjunctive set of subcarriers, which can be modulated differently. The sets are generated by the access point and are based on knowledge of the actual subcarrier attenuations. First the access point *allocates* each terminal a *number* of sub-carriers, afterwards the process of subcarrier *assignment* takes place, which decides on the disjunctive sets of subcarriers, given each terminal's allocated number. Then, per subcarrier different signal constellations may be used for data transmission, where the access point can assign one of $M$ possible constellation types. The access point picks for each subcarrier the constellation type which transmits the highest amount of bits while still providing a symbol error rate lower than $P_s$.

In order to perform these calculations, the access point has to have knowledge of all actual subcarrier signal-to-noise ratio (SNR) values regarding each wireless terminal. We assume the actual subcarrier states of each terminal to be known to the access point prior to the process of assigning sets of subcarriers to terminals and signal constellations to subcarriers. No form of power control is applied to the system, thus for each subcarrier the same transmission power $P_{tx}$ is used. Denote the attenuation (resulting from fading, path loss, and shadowing) at time $t$ for subcarrier $s$ regarding terminal $j$ by $a_{s,j}(t)$ and the noise power for each subcarrier by $n^2(t)$. Then, the SNR for subcarrier $s$ regarding terminal $j$, $x_{s,j}(t)$, results from Eq. (1).

$$x_{s,j}(t) = \frac{a_{s,j}^2(t)}{n^2(t)} P_{tx} \tag{1}$$

After the assignments have been generated by the access point, they have to be signaled to the terminals prior to the downlink transmission. For this purpose, an additional control channel is provided. Neither the cost of this
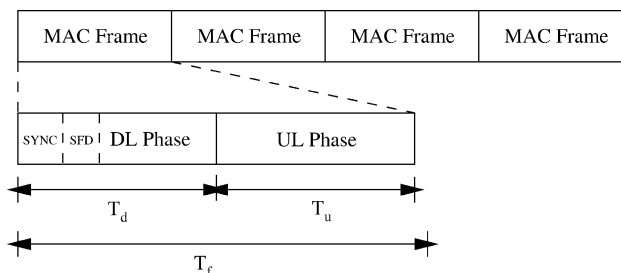


Fig. 3. Frame structure with downlink and uplink phase.

channel nor the influence of transmission errors on this channel are included in the system model.

Time is split into frames of equal length $T_f$. During such a frame first a phase for synchronization is provided, followed by a start frame delimiter. Then follows the payload transmission in downlink direction, which is followed by an uplink phase. The uplink phase is not considered further. The time span of uplink- and downlink phase are chosen to be equal. Fig. 3 shows the structure of the frame.

Signaled assignments and constellations are valid throughout the following downlink phase only. To adapt efficiently to the subcarrier attenuations, $T_f$ chosen to be smaller than the coherence time of the wireless channel at the given maximum speed $v_{max}$. Therefore, each subcarrier's SNR is modeled to be constant during one downlink phase. Thus, for the frame at time $t$ the system can be described by a matrix of all SNR values $x_{s,j}(t)$. Since for each SNR value exactly one signal constellation type is chosen, the system might also be described by a bit matrix with elements $b_{s,j}(t)$ denoting the number of bits transmitted per symbol by the signal constellation applied on subcarrier $s$ if it will be assigned to terminal $j$ for the next downlink phase.

As traffic model we assume data streams consisting of packets to arrive at the access point. The data sources are located outside of the cell. Each terminal receives one such stream that belongs to a video application on the terminal. The amount of data arriving at the access point per stream varies for a given time span. Each packet is queued at the access point upon arrival separately per terminal. Due to the encoding scheme of moving picture experts group-4 (MPEG-4), different packets have a different semantic relevance to the application at the terminal. We assume the access point to have full knowledge of the length of each packet, the semantic importance and the latest possible delivery time of each packet to the terminal. The remaining delivery time results from a given end-to-end delay minus the already consumed time for forwarding the packet from the source through the backbone to the access point. It is not assumed that the access point has knowledge regarding the streams, for example regarding their average bit-rate. Terminals receive in general different streams with different content and coding rate (heterogeneous stream mix).

As performance measure we are interested in the achievable application layer quality. Raw throughput or other network-layer-based performance metrics are not necessarily meaningful for multimedia applications, since these applications usually are loss tolerant and delay sensitive. Furthermore losses have different impacts on the quality the user perceives and delay causes additional losses.

Actually we used two metrics to measure the systems' performance in terms of application quality. The first one may be described as the *capacity* of the cell (where capacity has a different meaning than it has in information theory). By capacity we mean here how many terminals can be served in the cell, given a certain minimum video quality provided at any time span of length *l* during the video transmission. None of the transmitted video data streams is allowed to be of less quality over this time span. The second metric considered is the achievable video quality given that a certain number of terminals is present in the cell. Both quality metrics are explained in detail in Section 4.1.

## 3. Optimization approach

A lot of research regarding video quality improvement has been done in the last decade. A challenging problem with streaming video is its high demand on the QoS features of the transport medium. But bandwidth, delay, and loss rate vary over time (especially in a wireless system) and the current Internet lacks QoS support. To overcome these problems, various mechanisms have been developed. They can roughly be divided in feedback control, source-rate adaptation, packetization, and error control [5]. Feedback control is done by estimating available network bandwidth based on packet loss information at the receiver [6]. Source-rate adaptation is achieved with, e.g. frame skipping or the more sophisticated multi-layer coding (also known as fine-grained scalability (FGS)) or rate-distortion theory-based approaches [5,7,8]. Packetization schemes try to minimize overhead while maximizing robustness against losses. Effectively, inter-packet dependencies are minimized [5]. Error control approaches include forward error control (FEC), retransmission schemes and error-resilient encoding [9,10]. All these mechanisms address the problem of loss and delay during video transmissions and are, therefore, in principle applicable to wired and wireless connections. Many of these approaches require end systems that are aware of the actual link or end-to-end feedback or both.

Unlike these approaches, an access point connecting the Internet to a wireless environment can use its knowledge of both the wireless channel and the application requirements (packet importance). The classification of packets can be done with the schemes used in the differentiated services [11] approaches. The main advantage of prioritized transmission and priority drop [12], performed directly in the link-layer of the access point, is the independence of end-to-end feedback systems as well as from the actual end systems.

For these reasons we chose this approach. We seek to improve the situation by considering means the access point itself might employ, therefore involving the link-layer. In detail, we suggest and design a combined link- and physical-layer approach, where traditional layer separation is put aside and a new entity at the access point takes control of packet transmission management as well as subcarrier distribution, handling both link-layer and physical-layer issues. It involves exploiting channel as well as packet knowledge, including the semantical relevance of a packet. Our approach is based on two mechanisms: the one matching packets to subcarriers (where this approach is based on two components), while the other one relates to packet transmission management. Both parts can be used separately, but might also be connected via the semantic knowledge of packets in the queues of the access point. We first introduce the two separate parts of the combined layer approach, and show afterwards how they may be combined in order to boost the system even further. This yields the fully combined link- and physical-layer approach.

Sections 3.1–3.3 describe, in a top-down fashion, how our combined link- and physical-layer approach works by semantically managing the queues in the access point, allocating resources, and assigning subcarriers.

### 3.1. Video packet management

Recent video coding methods exploit both the spatial and the temporal redundancy in the source data. Spatial redundancy is reduced using block-wise discrete cosine transformation (DCT) and quantization followed by entropy coding of the remaining coefficients. This is known as intraframe coding. Temporal redundancy is lowered by coding only the differences between any two successive images. Several methods exist to perform temporal coding, such as frame differencing, motion estimation and motion-compensated prediction. These methods are known as inter-frame coding [13].

MPEG-4, a popular example of state-of-the-art video coding techniques, takes advantage of intra- and inter-frame coding methods. It distinguishes between three frame types, namely I-, P-, and B-frames. I-frames are solely intra-coded frames, P-frames are predicted frames depending on the previous I-frame, and B-frames are bidirectional 'predicted' frames (depending on the previous and following I- or P-frame). Frames are arranged in group of pictures (GOPs). A GOP consists of exactly one I-frame and some related P-frames and optionally some B-frames between these I- and P-frames [14]. In MPEG-4, I-frames contain by far the most information. Furthermore, loosing an I-frame would cause distortion of all following frames in a GOP. A P-frame loss would only influence the flanking B-frames and the loss of a B-frame would not influence any other frame.

The basic idea of the semantic queue management/scheduling algorithm is to exploit the significant differences between the MPEG-4 frame types regarding their information content and influence on the error propagation. The hypothesis is that prioritized treatment of semantically more important frames results in a much better video quality. The scheduling algorithm manages the order of transmission depending on the type of the frame and the time at which a packet will be dropped from the queue. If there are any packets in the queue containing parts of an MPEG-4 I-frame, they are transmitted first, followed by packets related to P-frames and B-frames. If there are no frames in the queue or there is still bandwidth left within the current downlink phase, other data possibly in the queue is transmitted.

The other scheduling parameter is the drop time. We define an overall maximum acceptable delay. Video packets, which cannot be transmitted within the time limited by the maximum delay, are dropped. A sensible approach is to derive this maximum delay from the play-out buffer size at the terminal and to drop a packet once the allowable delay is exceeded. While this would be straightforward for constant bit-rate traffic, the unpredictability of the amount of data per time in a variable bit-rate scheme requires an additional consideration: it is reasonable not to use the maximum allowable delay for all packets, since this could lead to unwanted loss of important data in the future. Therefore, we drop semantically less important packets earlier—P-frames are dropped 25% earlier than I-frames and B-frames are dropped 50% earlier. Fig. 1 refers to this mechanism as 'Semantic queue management (/Y/es or /N/o)'. If semantic scheduling is off, standard FIFO queue handling is used.

These parameters were not systematically optimized regarding their influence on the resulting video quality; this is an issue of further studies. But even these heuristically chosen values lead to a significant improvement compared to semantically 'blind' FIFO packet management.

### 3.2. Resource allocation

Resource allocation in OFDM-frequency division multiple access (FDMA) systems has to build subsets of subcarriers, which serve as channels for each active terminal in the cell. Building subsets dynamically based on actual subcarrier attenuation information has been investigated recently and turns out to be quite advantageous in terms of various transmission metrics such as power [15] or throughput [16].

In this paper, we follow the method suggested in Ref. [17] to first determine the size of each subset, called *allocation*, and select the specific subcarriers in a second step, called *assignment*. This two-step approach has a much lower complexity than one-step approaches; however, in some cases it turns out to be suboptimal. Due to complexity and available information at the access point, we do not follow the suggested methods in Ref. [17] but propose other ones.

Resource or, equivalently, subcarrier allocation, in the simplest case (static), gives each terminal a fixed number of subcarriers ('/S/tatic resource allocation' in Fig. 1). In a more sophisticated version, subcarrier allocation is influenced by the queue sizes at the access point. Such a scheme has been presented in Ref. [3]. The more data is queued at the access point for a certain terminal, the more subcarriers it will receive from the allocation process. The precise scheme works like this: at first, the access point allocates each terminal one subcarrier if its queue is not empty. Then, the remaining subcarriers are distributed according to the sizes of the queue. When $J_a$ terminals have a non-zero queue size, $S - J_a$ subcarriers remain not allocated after the first step of the scheme. Consider an overall amount of $d_j(t)$ bits to be actually stored in the queue for terminal $j$. Then, terminal $j$ receives an overall amount of subcarriers $a_j(t)$ given by Eq. (2).

$$a_j(t) = 1 + \left\lfloor (S - J_a)\frac{d_j(t)}{\sum_{\forall j} d_j(t)} + 0.5 \right\rfloor \qquad (2)$$

This allocation scheme has been called '/B/yte' in Fig. 1 as it takes into account only the queue length. It has multiple advantages. First, the access point does not have to be informed about the overall bit-rate of any stream. It also does not have to assign more subcarriers to terminals with an overall worse subcarrier state behavior. Everything is regulated automatically by the queue size. If a terminal's path loss is actually quite high due to its distance to the access point, its queue will simply become larger within the next couple of frames. Eventually, it will be allocated more subcarriers, depending on the queue dynamics of the other terminals, decreasing its queue size again. Furthermore, this scheme can take advantage of variable bit-rate streams by continuously varying the number of allocated subcarriers for each subcarrier. Due to statistical multiplexing this increases then the number of terminals, which might be served in the cell.

Still, an additional improvement is possible by using cross-layer information about the actual semantics of the data in the queues. Considering the fact that packets containing I-frame segments are more important for the resulting video quality than, e.g. packets with B-frame data, it is advisable to allocate more bandwidth for queues containing I-frames than for those containing B-frames. This idea requires some sort of layer interaction. We solve this by assigning certain weights for the three MPEG-4 frame types. I-frames get a weight of 4, P-frames a weight of 2 and B-frames a weight of 1. These weights are then multiplied with the amount of bytes of each frame type currently in the queue. These weighted queue sizes are calculated in the access point for each queue and used for the subcarrier assignment scheme. From this optimization approach (referred to as '/W/eighted queue size allocation'

in Fig. 1) we expect superior results compared to the simpler allocation scheme, where only the queue sizes matter.

Both approaches require cross-layer communication. In contrast, the '/B/' allocation scheme does not take semantic information into account and should, therefore, be easier to implement. Nevertheless, both queue size sensitive allocation approaches should lead to better performance than using no size information at all for the resource allocation.

### 3.3. Subcarrier assignment

Once the subcarrier allocation is found, the access point now has to assign each terminal this number of subcarriers. If the resource allocation is static (each terminal always gets the same number of subcarriers), the simplest choice is to also give to each terminal always the same subcarriers.

A dynamic subcarrier assignment (as required by a dynamic resource allocation scheme), on the other hand, tries to find specific subcarriers that are going to be used for a downlink data transmission of a certain terminal. For this task different algorithms have been suggested, namely one optimal approach and various heuristics [17,18]. Given the actual matrix of SNR values $x_{s,j}(t)$ or bits to be conveyed to each terminal on each subcarrier per symbol $b_{s,j}(t)$, the task is to find subcarrier assignments where each terminal receives the best possible subcarriers (either in terms of SNR or overall transmitted bits). Obviously, each subcarrier can only be assigned once.

Due to complexity reasons we choose here the heuristic approach given in Ref. [18]. In summary, this scheme assigns for each downlink phase each terminal a certain priority. It begins to assign subcarriers for the terminal actually holding the highest priority and assigns it the allocated number of best subcarriers out of all $S$ subcarriers. Next, the terminal with the second highest priority is assigned the allocated number of best subcarriers out of the remaining subcarriers from the first step, and so on, until all subcarriers have been assigned.

This assignment algorithm has input arrays by which the outcome for each terminal might be influenced. First, an array holding the subcarrier allocations $a_j(t)$ for all terminals $j$ at time unit $t$ and second, an array holding the priorities assigned to all terminals $j$ at time unit $t$, $p_j(t)$. In this work, the priorities are simply 'shifted' after each downlink phase, such that the terminal with the highest priority receives the second highest priority for the next phase and so on. However, the priorities might also be distributed differently by the access point.

## 4. Performance study

The above proposed system design has been studied by simulations. We first discuss the used metric in depth and how it relates to other metrics used for video transmission analysis (Section 4.1). Then we present in Section 4.2 the simulation scenario with all parameter settings. Our results are given in Section 4.3.

### 4.1. Performance evaluation metric

Video quality measurements must be based on the perceived quality of the actual video being received by the users of the digital video system. Such a perception-based evaluation is appropriate as the subjective impression of the user is what only counts ultimately. This intuitive impression of a human user watching a video is grasped by subjective quality metrics. These subjective metrics provide most information, but their determination is costly as real humans have to watch videos, making them highly time consuming and requiring high manpower and special equipment. Such subjective methods are described in detail by ITU [19], ANSI [20] and MPEG [21]. Describing the human quality impression by a subjective quality metric is usually done with a mean opinion score (MOS), on a scale from 5 (best) to 1 (worst) as in Table 1.

The expensive and complex subjective tests are often not affordable. Also many tasks in industry and research require automated methods to evaluate video quality. That is why objective metrics have been developed to emulate the quality impression of the human visual system. In Ref. [22], there is an exhaustive discussion of various objective metrics and their performance compared to subjective tests. The most widespread method is the calculation of the peak signal-to-noise ratio (PSNR) image by image. It is a derivative of the well-known SNR. The PSNR compares the maximum possible signal energy to the noise energy, which results in a higher correlation with the subjective quality perception than the conventional SNR [23]. Eq. (3) gives the definition of the PSNR of source image $s$ and destination image $d$ [24].

$$\text{PSNR}(s,d) = 20 \log \frac{V_{\text{peak}}}{\text{MSE}(s,d)} \text{ [dB]} \tag{3}$$

where

$V_{\text{peak}} = 2^k - 1, k$ bit color depth

$\text{MSE}(s,d) = $ mean square error of $s$ and $d$.

While the PSNR does not directly correspond to the MOS, there exist heuristic mappings of PSNR to MOS (subjective quality) as shown in Table 2 [25]. To evaluate the impact of

Table 1
ITU-R quality and impairment scale

| Scale | Quality | Impairment |
|---|---|---|
| 5 | Excellent | Imperceptible |
| 4 | Good | Perceptible, not annoying |
| 3 | Fair | Slightly annoying |
| 2 | Poor | Annoying |
| 1 | Bad | Very annoying |

Table 2
Possible PSNR to MOS conversion

| PSNR [dB] | MOS |
| --- | --- |
| >37 | 5 (Excellent) |
| 31–37 | 4 (Good) |
| 25–31 | 3 (Fair) |
| 20–25 | 2 (Poor) |
| <20 | 1 (Bad) |

the network (delay, loss) on the video quality, we need to compare the received (possibly distorted) video with the actually sent video. In fact, even the sent video is already distorted by the encoding process, e.g. MPEG-4, but this distortion cannot be avoided when striving for acceptable bit-rates in video streams.

Hence, we have a computational approximation of the subjective human impression of every single frame at our disposal. Based on this frame-by-frame MOS calculation, we define a metric, which reflects the user impression of the *entire* received video. This is necessary because even the replacement of PSNR with metrics correlating better with the subjective impression does not address the problem of how to assess entire video sequences.

A simple approach to extend a frame-wise metric like the PSNR to a metric for an entire video would be to calculate the average of all PSNR values of all frame. However, such an average PSNR (or MOS) value for the entire video does not map very well to the subjective impression, especially in the case of longer video clips. If, for instance, only the first 10 s of the video stream are highly distorted the user is not satisfied with the video quality while the average MOS would not reflect this. Fig. 4 shows the quality of five different video transmissions and the reference video quality. The average MOS is printed on top of each bar. Although this average MOS seems acceptable in all cases, it is likely that the frames with bad MOS grades happen to
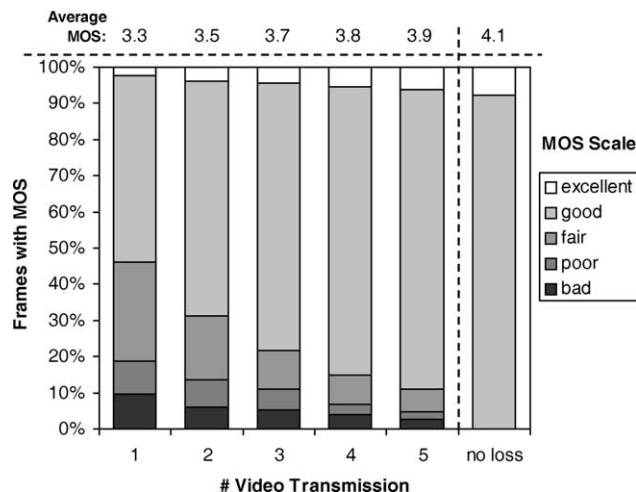
appear sequentially, if one considers fading channels for example. In this case, the subjective video quality will not match the average MOS.

To resolve this potential problem we follow a different approach here. We assume that the sent input video's quality is always acceptable for a human watching it, i.e. that the coding distortions are acceptable. When is the *received* video also acceptable, i.e. when are the transmission-related distortions compared to the sent video still acceptable?

Intuitively, some small amount of distorted frames is likely acceptable, as long as this number does not become too big. This 'too big' is formalized by requiring that, for any arbitrary part of the video, the number of received frames with a MOS smaller than that of the sent frame must not exceed, for example, 20% of the frames contained in that period—such a video transmission is called 'successful'. Formally, this is captured by the distortion in interval (DIV) metric we devised for this purpose. This metric maps an arbitrary interval of a video stream to the number of frames with reduced MOS value; a successful video is then a video where for all intervals of a fixed size, the DIV metric is smaller than a given threshold. Fig. 5 shows this metric. It shows if a video transmission was successful or not according to the metric defined above.

Applying this condition to any arbitrary period is necessary as videos which, e.g. completely loose their first 20% of all frames but perfectly transmit the others would not be acceptable. The actual threshold of 20% actually is a parameter of the evaluation process that has been adopted according to subjective quality comparisons. Also, the length of the interval is a parameter; with shorter intervals, the threshold is exceeded more often.

Note that this is a very stringent metric to satisfy: even a violation in a single interval suffices to render a video unacceptable—unlike in average-based metrics. But this stringency is particularly necessary for long videos,



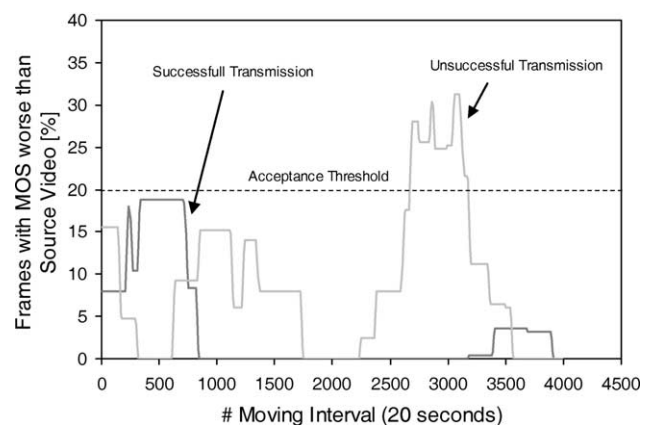Fig. 4. Exemplification of average MOS.



Fig. 5. Exemplification of video quality metric DIV; successful transmission means the percentage of frames worse than the original must be less than 20% within all intervals.

as time-averaging metrics would tend to smooth out lots of visually disturbing impediments.

### 4.2. Scenario parameters

For the simulation we parameterized the system model as follows. We chose a system with a bandwidth equivalent to IEEE 802.11a, thus the available bandwidth was $B = 16.25$ MHz, which was split into 52 subcarriers each with a bandwidth of 312.5 kHz, from which $S = 48$ subcarriers were available for data transmission. Corresponding to this, each OFDM symbol had a length of $T_s = 4$ μs from which $T_g = 0.8$ μs belonged to the guard interval. As center frequency we chose a channel from the U-NII lower band, which is located around 5.2 GHz.

The subcarrier attenuations changed constantly due to the movement of the terminals and the multi-path propagation environment. Wireless terminals moved within the cell with a random velocity, the maximum speed was set to $v_{max} = 1$ m/s. The considered cell's radius was $R = 100$ m. The effects influencing the subcarrier attenuations were path loss, shadowing and fading. Path loss was determined by the formula

$$\frac{P_0}{P_{tx}} = K \frac{1}{d^\alpha}$$

where $P_0/P_{tx}$ denotes the ratio between received and transmitted power, $d$ denotes the distance between transmitter and receiver, $K$ denotes the reference loss for the distance unit $d$ is measured in and $\alpha$ is the path loss exponent. We parameterized the reference loss with $10 \times \log(K) = 46.7$ dB and the path loss exponent with $\alpha = 2.4$. The shadowing was assumed to be log-normal distributed with a standard deviation of $\sigma = 5.8$ dB and a mean of 0 dB. For the fading, the power spectral density was chosen to have a Jakes-like shape [26] with a Doppler frequency depending on $v_{max}$. The multi-path propagation environment

was characterized by a delay spread of $\Delta\sigma = 0.15$ μs with an exponential power delay profile according to the large open space model of ETSI C [27]. An example environment corresponding to such a setting would be a large airport or exposition hall.

We set the frame length to $T_f = 2$ ms which corresponded to the frame length of HIPERLAN/2 systems. Uplink and downlink phases were considered to have an equal length, which left a time span of 1 ms for the downlink transmissions (where administrative tasks like synchronization still have to be performed). The noise power was assumed to equal $-117$ dBm per subcarrier. The transmit power was set to $-7$ dBm per subcarrier, according to the maximum allowed transmit power in the U-NII lower band, which is 40 mW per system. Together with the attenuation states of each subcarrier, an instantaneous SNR was generated and kept fixed for the whole frame length. The adaptive modulation system consisted of five different modulation types: BPSK, QPSK, 16-QAM, 64-QAM and 256-QAM. As maximum acceptable symbol error probability we chose the value $P_s = 10^{-2}$. As stated, depending on the SNR of the subcarrier the modulation type with the highest rate was chosen that still yielded a symbol error probability lower than $10^{-2}$.

We used two different video sources in our simulations. One was a source with low-motion video, and the other a sequence of a movie with very high-motion. These sources can be coded by MPEG-4 with different efficiency and they behave differently in case of packet losses (Section 3.1).

Both videos had 25 frames per second, were in common intermediate format (CIF) ($352 \times 288$ pixel) and were 4500 frames (3 min) long. They were encoded according to MPEG-4 with variable bit-rate and a fixed 12-GOP with two consecutive B-frames (IBBPBBPBBPBB) (Section 3.1). Fig. 6 shows the resulting bit-rate of the MPEG-4 encoded videos. Their average bit-rates were 723 and 951 kBit/s, respectively.
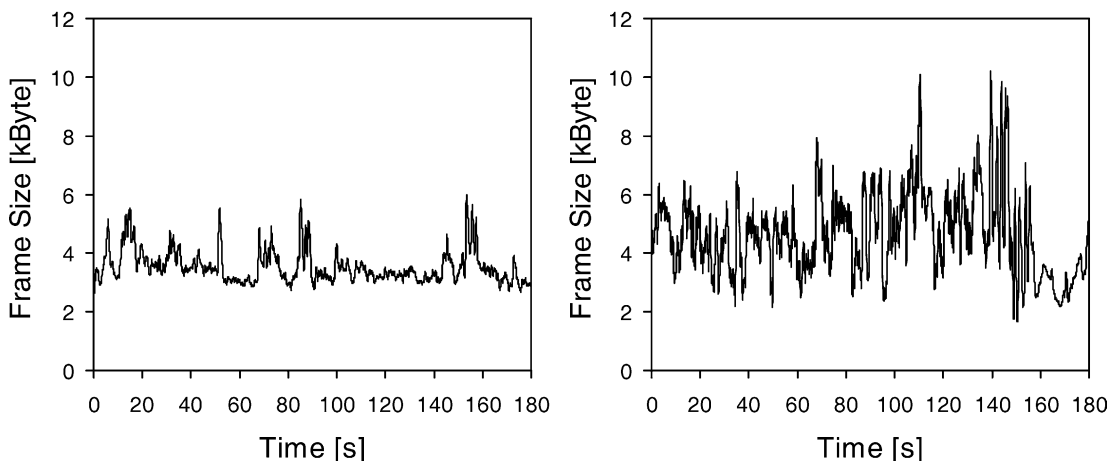


Fig. 6. Bit-rate of low-motion (left) and high-motion (right) video source used for the simulations.

### 4.3. Results

Using the first performance metric (how many terminals can be served with a certain video quality), we evaluated the five combinations of the optimization approaches shown in Fig. 2. In Figs. 7 and 8, the number of supportable terminals is shown for a homogeneous traffic case with either the high-motion or low-motion video source, respectively, for different optimization approaches and different deadlines. For instance, the white bar marked N/B/D at a deadline of 100 ms denotes that the system is able to serve 10 terminals with the high-motion stream (12 with low-motion) such that in any given interval of 20 s, the percentage of distorted frames is at most 20%. Increasing the complexity of methods (more sophisticated optimization approaches) at the access point leads to a better overall usage of the system resources. This applies to all simulated deadlines and to both classes of video sources. Comparing the simplest scheme (N/S/S) to the most sophisticated one (Y/W/D), the number of supportable terminals is about *four times* higher.

In a next step, we fixed the number of terminals in the cell to 12 and calculated the achievable video quality with every combination of methods. The maximum DIV value of each received video (the percentage of frames with reduced MOS, maximized over all intervals of a fixed size, small values are good) was averaged over all 12 terminals in the cell. Besides considering again the homogeneous traffic case with the high-motion as well as the low-motion video source, we also considered two cases with heterogenous traffic cases. In the first case, four terminals received the low-motion video source while the remaining eight terminals received the high-motion source. In the second case, this ratio between terminals receiving the high-motion streams and the ones receiving the low-motion streams was turned around. Fig. 9 shows the results for the homogeneous
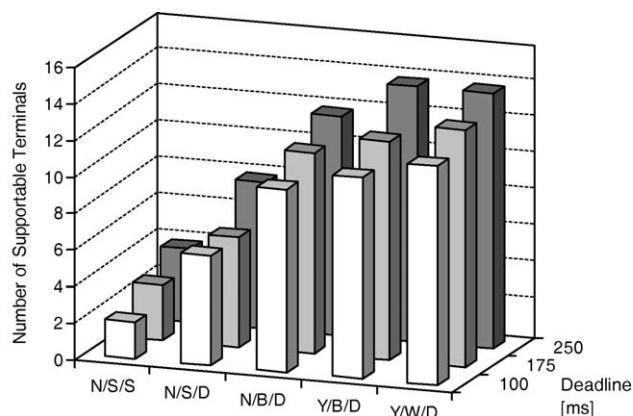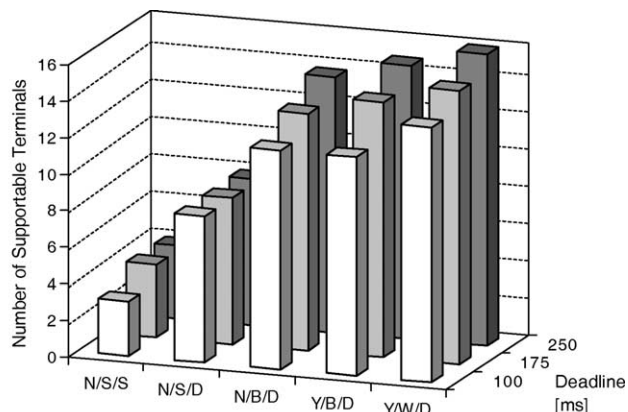
case where all terminals were served the low-motion video, Fig. 10 shows to the heterogenous case with $\frac{1}{3}$ of the terminals served the high-motion video, Fig. 11 belongs to the other heterogenous case with $\frac{2}{3}$ high-motion video and finally Fig. 12 belongs again to the homogeneous traffic case where all terminals received the high-motion video. These figures show a tremendous improvement: in Fig. 9, e.g. with the best combination of mechanisms there is a single 20 s interval, in which only 1.2 % of the video frames are worse than the original frames; all other intervals have an even better quality!

As expected, the achievable quality decreases with a higher fraction of terminals receiving the more variable video. For each setting, initially the ascending methods lead



|  | N/S/S | N/S/D | N/B/D | Y/B/D | Y/W/D |
|---|---|---|---|---|---|
| □ 100 | 3 | 8 | 12 | 12 | 14 |
| ▨ 175 | 4 | 8 | 13 | 14 | 15 |
| ■ 250 | 4 | 8 | 14 | 15 | 16 |

Fig. 8. Number of supported terminals using low-motion video source.



|  | N/S/S | N/S/D | N/B/D | Y/B/D | Y/W/D |
|---|---|---|---|---|---|
| □ 100 | 2 | 6 | 10 | 11 | 12 |
| ▨ 175 | 3 | 6 | 11 | 12 | 13 |
| ■ 250 | 4 | 8 | 12 | 14 | 14 |

Fig. 7. Number of supported terminals using high-motion video source.
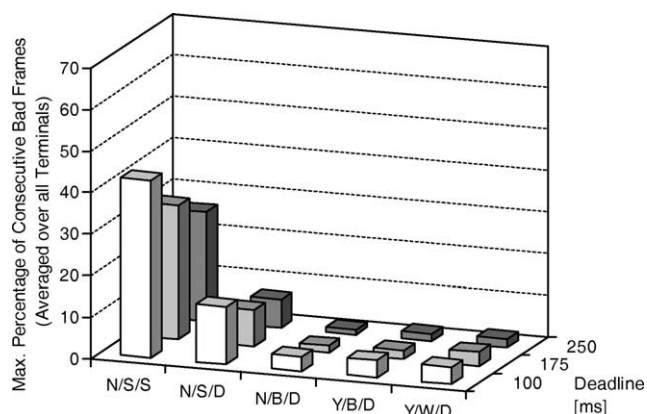


|  | N/S/S | N/S/D | N/B/D | Y/B/D | Y/W/D |
|---|---|---|---|---|---|
| □ 100 | 42.7 | 13.9 | 3.4 | 4.2 | 4.2 |
| ▨ 175 | 32.3 | 8.8 | 1.7 | 2.2 | 3.4 |
| ■ 250 | 26.2 | 6.9 | 1.2 | 1.8 | 2.1 |

Fig. 9. Average quality of 12 terminals receiving low-motion source.

| | N/S/S | N/S/D | N/B/D | Y/B/D | Y/W/D |
|---|---|---|---|---|---|
| ☐ 100 | 68.0 | 27.3 | 12.6 | 10.4 | 10.9 |
| ☐ 175 | 58.1 | 19.6 | 9.0 | 7.2 | 7.3 |
| ■ 250 | 54.3 | 15.0 | 7.8 | 6.4 | 6.7 |
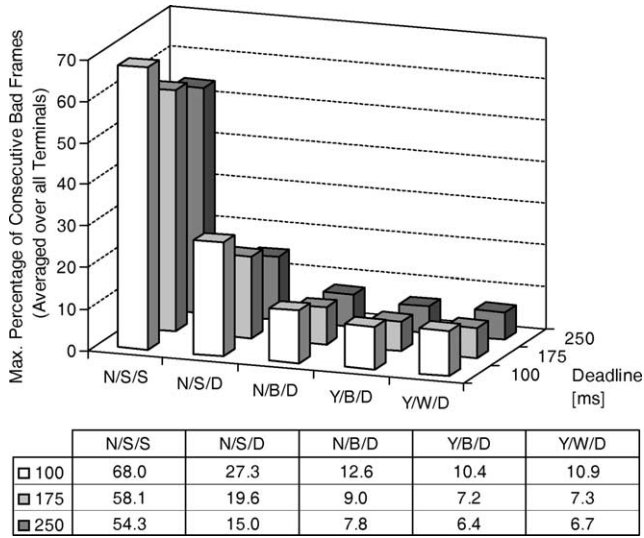
Fig. 10. Average quality of 12 terminals (1/3 receiving high-motion source).



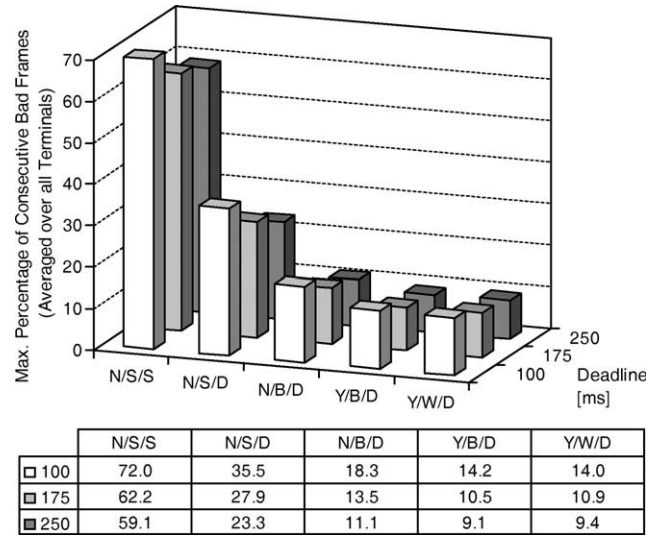| | N/S/S | N/S/D | N/B/D | Y/B/D | Y/W/D |
|---|---|---|---|---|---|
| ☐ 100 | 72.0 | 35.5 | 18.3 | 14.2 | 14.0 |
| ☐ 175 | 62.2 | 27.9 | 13.5 | 10.5 | 10.9 |
| ■ 250 | 59.1 | 23.3 | 11.1 | 9.1 | 9.4 |

Fig. 12. Average quality of 12 terminals receiving high-motion source.

to a significant average quality increase (always at least by a factor around 7). However, after some point a higher complexity of the optimization does *not always* lead to a higher quality as shown in Fig. 9. Note that this always occurs for average video quality values lower than 13%. Previously, applying the 20% quality limit to the maximum number of supportable terminals resulted in a performance increase for each step in the ascension of mechanisms.

This somewhat surprising result shows that the ad hoc chosen parameters of the semantic scheduling cannot be applied unconditionally. In these cases (and in contrast to the 'overloaded' settings in the cases with the maximum number of supportable terminals in the cell) the system provides enough bandwidth in principle, there is no reason

to drop frames with lower priorities earlier or to assign less subcarriers to queues with lower prioritized frames. This might lead to a slight quality degradation when activating the semantic queue management, for example. Although the chosen parameters for the semantic scheduling lead to improvements during phases with bad channel conditions and only little bandwidth available (compared to the queued data), either fine-tuning of these parameters in good situations or deactivating the semantic scheduling could resolve this behavior.

## 5. Conclusions and future work

We studied the performance of a cross-layer communicating system set up, trying to exploit channel-related information as well as packet-related information. This approach was studied using an OFDM-frequency division multiple access (FDMA) transmission system and considering the transmission of MPEG-4 encoded video streams. The approach consists of three mechanisms: dynamic packet management based on the packet related information, dynamic resource allocation depending on each terminal's queue, and dynamic resource assignment depending on the previously performed allocation and the states of the subcarriers.

We find that the suggested system setup is able to increase various performance metrics significantly. In terms of the cell's capacity, the fully optimized set up achieves a capacity increase by up to 300% for a given minimum acceptable video quality. Accordingly, when fixing the number of terminals in the cell this set up is able to eliminate almost all errors occurring within the video transmission.

Beside these performance related aspects, a clear advantage of this system concept is its simple set up



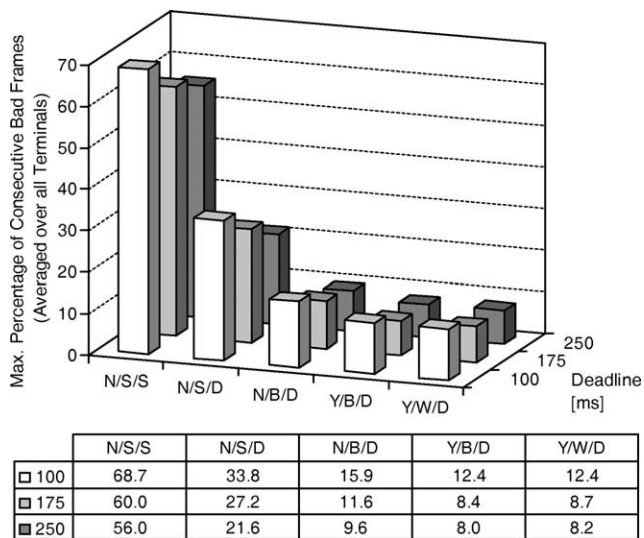| | N/S/S | N/S/D | N/B/D | Y/B/D | Y/W/D |
|---|---|---|---|---|---|
| ☐ 100 | 68.7 | 33.8 | 15.9 | 12.4 | 12.4 |
| ☐ 175 | 60.0 | 27.2 | 11.6 | 8.4 | 8.7 |
| ■ 250 | 56.0 | 21.6 | 9.6 | 8.0 | 8.2 |

Fig. 11. Average quality of 12 terminals (2/3 receiving high-motion source).

and modular structure. The set up is able to benefit from statistical variations of the video streams as well as from statistical variations of the subcarrier states. The 'history' of these variations influences the actual queue states on which the resource allocation is performed. The most recent changes of the stream variations are then handled by the packet management whereas the most recent changes of the subcarrier states are taken care of by the subcarrier assignment scheme. The success of these two mechanisms influences the queue states prior to the next downlink phase. If no packet information is available, the dynamic packet management can be simply excluded from the system; the same can be done with the dynamic resource management if no subcarrier state information is provided.

However, the considered system model does bear some optimistic assumptions. First of all, it is not guaranteed that the required information is available regarding either the actual subcarrier states or the packet-related information. While the packet-related information might be available in principle, it is not possible to provide the access point with the recent subcarrier states. Therefore, the access point would have to consider estimates instead which will lead to some performance decrease. Also, the dynamic subcarrier management requires a signaling system, which informs the terminals of their newest resource assignment prior to each data transmission. This task will decrease the system performance also, since signaling requires system resources either in time or in terms of bandwidth. While the impact of the signaling overhead is rather low, the provision of accurate channel estimates is quite important with a significant influence on the system performance. As further assumption the considered set up did not include the impact resulting from the increasing complexity at the access point: generating new subcarrier assignments requires some computational power since this is a real-time problem. For every frame, the assignments have to be available when the signaling phase starts. Although frame times are short in this study, the impact of complexity has already been studied and turned out to be quite low. Hence, the impact of complexity exists and will decrease performance, but not significantly.

As further work we seek to highlight the impediments of the system when making these assumptions more realistic. We actually consider this already via simulation; we are also interested in validating our results by applying the system set up to a real-life test bed. In addition to this, we also would like to extend this work to other stream types where transmission control protocol (TCP) streams are the most important ones. In this context, we will investigate heterogenous traffic scenarios where video streams and transmission control protocol (TCP) streams are transmitted in parallel such that the resource management unit now has to serve different stream types as good as possible.

## References

[1] G. Song, Y. Li, Utility-based joint physical-medium access control (MAC) layer optimization in OFDM, in: Proceedings of Global Telecommunications Conference (GLOBECOM), vol. 1, IEEE, 2002, pp. 17–21.

[2] Y. Zhang, L. Cheng, Cross-layer optimization for sensor networks, in: Proceedings of New York Metro Area Networking Workshop 2003, 2003.

[3] J. Gross, J. Klaue, H. Karl, A. Wolisz, Subcarrier allocation for variable bit rate video streams in wireless OFDM systems, in: Proceedings of Vehicular Technology Conference (VTC Fall), Florida, USA, 2003.

[4] J. Klaue, J. Gross, H. Karl, A. Wolisz, Semantic-aware link layer scheduling of MPEG-4 video streams in wireless systems, in: Proceedings of Applications and Services in Wireless Networks (ASWN), Bern, Switzerland, 2003.

[5] D. Wu, Y.T. Hou, W. Zhu, H.-J. Lee, T. Chiang, Y.-Q. Zhang, H.J. Chao, On end-to-end architecture for transporting MPEG-4 video over the internet, IEEE Transactions on Circuits and Systems for Video Technology 10 (6) (2000) 923–941.

[6] X. Lu, M.E.Z.R.O. Morando, Understanding video quality and its use in feedback control, in: Proceedings of Packet Video 2002, Pittsburgh, PA, USA, 2002.

[7] W. Li, Overview of fine granularity scalability in MPEG-4 Video Standard, IEEE Transaction on Circuits and Systems for Video Technology 12 (3) (2001) 301–317.

[8] ISO-IEC/JTC1/SC29/WG11, ISO/IEC 14496: information technology—coding of audio-visual objects, 2001.

[9] T. Wiegand, M. Lightstone, D. Mukherjee, T.G. Campbell, S.K. Mitra, Rate-distortion optimized mode selection for very low bit-rate video coding and the emerging H.263 Standard, IEEE Transactions on Circuits and Systems for Video Technology 6 (1996) 182–190.

[10] S. Wenger, G. Knorr, J. Ott, F. Kossentini, Error resilience support in H.263 +, IEEE Transactions on Circuits and Systems for Video Technology 8 (1998) 867–877.

[11] J. Shin, J. Kim, Performance evaluation of differentiated services to MPEG-4 FGS video streaming, Journal of the Korean Institute of Communication Sciences 27 (2002) 711–723.

[12] J. Huang, C. Krasic, J. Walpole, Adaptive live video streaming by priority drop, in: Packet Video, 2003.

[13] K.N. Ngan, C.W. Yap, K.T. Tan, Video coding for wireless communication systems, Signal Processing and Communications, Marcel Dekker, New York, 2001.

[14] ISO/IEC/JTC1/SC29/WG11, Overview of the MPEG-4 Standard, July 2000.

[15] C. Wong, R. Cheng, K. Letaief, R. Murch, Multiuser OFDM with adaptive subcarrier, bit and power allocation, IEEE Journal on Selected Areas of Communications 17 (10) (1999) 1747–1758.

[16] W. Rhee, J. Cioffi, Increase in capacity of multiuser OFDM system using dynamic subchannel allocation, in: Proceedings of Vehicular Technology Conference (VTC), 2000, pp. 1085–1089.

[17] H. Yin, H. Liu, An efficient multiuser loading algorithm for OFDM-based broadband wireless systems, in: Proceedings of IEEE Globecom, 2000.

[18] J. Gross, H. Karl, F. Fitzek, A. Wolisz, Comparison of heuristic and optimal subcarrier assignment algorithms, in: Proceedings of the 2003 International Conference on Wireless Networks (ICWN'03), 2003.

[19] ITU, ITU-T Recom. P.910/920/930—Subjective Video Quality Assessment Methods for Multimedia Applications/Interactive Test Methods for Audiovisual communications/Principles of a Reference Impairment System for Video, 1996.

[20] ANSI, T1.801.01/02/03-1996—Digital Transport of Video Teleconferencing and One-Way Video Signals, 1996.

[21] ISO-IEC/JTC1/SC29/WG11, Evaluation methods and procedures for july MPEG-4 tests, 1996.

[22] S. Wolf, M. Pinson, Video quality measurement techniques, Tech. Rep. 02-392, US Department of Commerce, NTIA, June 2002.

[23] L. Hanzo, P.J. Cherriman, J. Streit, Wireless Video Communications, Digital and Mobile Communications, IEEE Press, Piscataway, 2001.

[24] M.J. Riley, I.E.G. Richardson, Digital Video Communications, Artech House, Norwood, 1997.

[25] J.-R. Ohm, Bildsignalverarbeitung fuer Multimedia-Systeme, http://bs.hhi.de/users/ohm/download/bvm-kap1&2.pdf, 1999.

[26] J. Cavers, Mobile Channel Characteristics, Kluwer Academic, New York, 2000.

[27] J. Medbo, P. Schramm, Channel Models for HIPERLAN/2, ETSI EP BRAN document 3ERI085B, March 1998.