Technical University Berlin

Telecommunication Networks Group

Scheduling of heterogeneous data streams in the downlink of a dynamic OFDM-FDMA wireless cell

Stefan Valentin, James Gross

{valentin, gross}@tkn.tu-berlin.de

Berlin, November 2004

TKN Technical Report TKN-04-017

TKN Technical Reports Series

Editor: Prof. Dr.-Ing. Adam Wolisz

Abstract

In this technical report the performance of a combined link- and physical layer approach [11] is studied for the downlink of a cellular OFDM-FDMA transmission system. This is done for two traffic load scenarios: In the heterogeneous scenario the simultaneous transmission of Web-pages, via HTTP and TCP, and MPEG-4 coded VBR video streams, via UDP, is considered, while in the homogeneous scenario only HTTP is transmitted. For the heterogeneous case a new scheduling scheme is proposed, which provides the adaption of the bit-rates of the transmitted video streams to the estimated load of the transmission system. For the applied filter semantic information, provided by the application layer, is considered. The presented performance analysis shows that this approach accelerates the downlink of a Web-page for up to 32% on the average in addition to the gain of up to 100%, found for the link- and physical layer approach.

Contents

Chapter 1

Introduction

The mobility provided by the transmission of speech and data without the need of a cable connection has lead to the wide integration of mobile telephones and computers in everydays life. With the increase in transmission speed and quality, provided by these techniques new interesting services, e.g. video telephony or the ubiquitous usage of the Internet can be provided to the mobile user.

However, the benefit of mobility adds a certain cost to the system. Unlike cable connections, wireless systems suffer from a high variance of the delay and the attenuation of the transmitted signals. This effect varies over time and frequency and rises with the data-rate. In order to assure the fault-tolerant transmission, while providing high data rates Orthogonal Frequency Division Multiplexing (OFDM) has been developed. With OFDM the spectrum is separated into small sub-bands, the so-called *subcarriers*, which are placed in order to assure a high utilization of the given bandwidth.

Although OFDM minimizes the distortion of the signal caused by the delay variance the attenuation of the channel still varies. Thus, the possibility of a bad channel state has to be considered in the design of a transmission system. To bound the resulting error-rate in conventional wireless systems typically a capacity-headroom, reserved by the use of a slow but fault-tolerant transmission scheme, is combined with an error-correction scheme. However, the overhead, i.e. error-correction complexity and headroom size, rises with the data-rate. A solution for this is the *channel-state* dependent scheduling approach. Here the scheduler distributes the channel resource according to the measured capacity in order to provide the optimal utilization of the wireless channel. For OFDM systems in combination with Frequency Division Multiple Access (FDMA) this has lead to the development of *dynamic OFDM-FDMA* systems, were subcarrier are dynamically distributed to the Wireless Terminals (WTs) according to their channel states. These systems, combined with a dynamic multiplexing scheme, were investigated in [11] for the transmission of MPEG video streams and for Web-page related traffic in [21].

However, performance studies in homogeneous scenarios do not represent the nature of the traffic in the Internet. Recent measurements have shown that up to 90% of the traffic on Internet backbones is related to the transmission of Web-pages while typically one third of the remaining proportion is related to video or audio streaming data [8]. In this technical report a *heterogeneous* scenario is considered, including both, the streaming of MPEG-4 coded video and the simultaneous transmission of Web-pages. For this scenario the performance

TKN-04-017 Page 1

of several combinations of the introduced dynamic OFDM-FDMA scheduling schemes is studied in Chapter 4. Furthermore a performance study for a homogeneous scenario is shown, considering the transmission of Web-pages only.

The second subject of this technical report is to propose an extension to the dynamic OFDM-FDMA system: the *adaptive Video Queue Management (VQM)*. The design of this new scheme takes heterogeneous traffic into account and provides the adaption of the scheduling policy to the varying system resources. This is known under the topics of *dynamic Quality Of Service (QoS)* and *adaptive filtering* [35] and is studied for distributed systems in wired networks in [14] and [15]. The basic idea of this adaption strategy is the manipulation of the video data according to the estimated system load. The objectives of adaptive VQM are to free capacity if high traffic load occurs and to balance the capacity requirement of the different flow types. The newly developed scheme considers cross-layer information for the filtering. This means, that semantic information from the video streams is provided by the application layer to the scheme on link layer. After the approach and the algorithm of this scheme are discussed in Chapter 5, its adaption capabilities and the resulting performance are analyzed for the heterogeneous case (Chapter 6).

In the following chapter the system model is described. The algorithms and scheduling policies of the considered dynamic OFDM-FDMA system are discussed in Chapter 3.

Chapter 2

System model and parameterization

2.1 Considered scenario and environment

We assume a single wireless cell of radius R containing a distinct number (J) of moving Wireless Terminals (WTs) and an access point with a fixed position. The access point, which is connected to the Internet via a cable link, forwards data between application servers in the Internet and the WTs using a wireless link. The characteristics of the radio channel and the employed transmission scheme OFDM are discussed in Section 2.2. In the model the communication in both directions, i.e. requests by the WTs to the application servers and the transmission of the requested data vice versa, is considered. However, in this report only the downlink transmission from the access point to the WTs is examined. For the downlink, we assume several combinations of static and dynamic OFDM-FDMA systems which are described in Section 2.3 and Chapter 3. The time is slotted into so-called Medium Access Control (MAC) frames of length T_f. In this report two scenarios are considered for the application layer: At first, in the homogeneous case, on each of the J WTs runs a HTTP client software, i.e. a Web-browser, and is requesting Web-pages from HTTP servers located in the Internet. Secondly we consider scenarios with heterogeneous traffic streams, where J_h WTs are running an HTTP client, while J_v WTs are receiving MPEG-4 video data streams from servers in the Internet. In Figure 2.1 an example setup of the considered system is illustrated, different colors for the servers and the WTs denote different types of application layer data. From setup to setup the amount of HTTP and video receiving wireless terminals is varied. However, during the simulation of a single setup the application type of a specific WT is fixed.
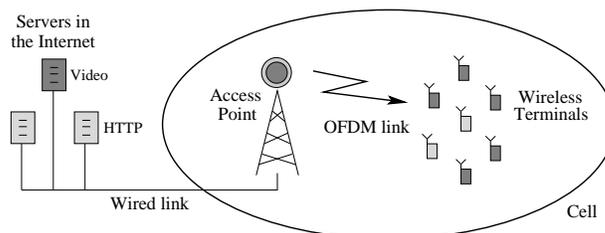
Figure 2.1: Example setup of the considered cellular system

2.2 Physical layer

In this section we will describe the modeling of the physical layer. This includes the behavior and the parameters of the radio channel as well as the employed modulation and transmission schemes.

During the terrestrial transmission of data, in the shape of electromagnetic waves, reflection on fixed or moving obstacles, like buildings, trees or cars, occurs. This leads to multi-path propagation, which means, that the transmitted signal reaches the receiver antenna on different paths each with a specific delay. The intensity of multi-path propagation is characterized by the delay spread, the difference between the transmission time of the direct path and the maximum delay of the reflected delay paths. With multi-path propagation the frequency response of the channel varies and therefore the attenuation (the absolute value of the frequency response, i.e. a power value) to the transmitted signal is frequency selective $(a(f))$. Since we consider moving WTs in our scenario, whose speed is assumed to be lower than a maximum value v_{max}, the attenuation also becomes time variant [4], which results in $a(f,t)$ Beneath the attenuation the transmitted signal suffers from channel noise (n), which also is time and frequency selective. The attenuation and the noise can be expressed in the channel gain G as in Equation 2.1 which leads to Channel Gain-to-Noise Ratio (CNR) values given in dB.

$$G(f,t) = \frac{a^2(f,t)}{n^2(f,t)} \tag{2.1}$$

The variation of the CNR in time and frequency due to multi-path propagation and movement is exemplarily shown in Figure 2.2. Considering moving WTs the CNR varies independently for each WT. The way how the CNR of a radio channel varies depends on the influence of several effects. For an analytical discussion these effects can be classified in path loss, shadowing and fading [1]. Path loss is a deterministic effect, which stands for the loss in power of a signal during the propagation from the sender to the receiver antenna. It is characterized by a reference loss K, which occurs over a reference distance and the exponent α, which characterizes how severe a distance increase attenuates the received signal. Both values α and K depend highly on the chosen environment. In this report we consider an urban
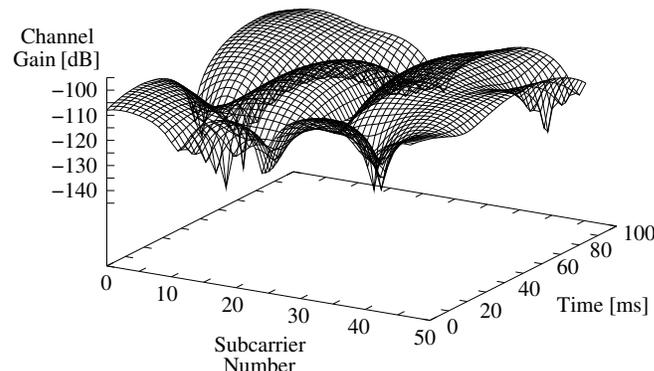
Figure 2.2: Example channel gain behavior of a frequency selective and time variant radio channel

environment, the related values are summarized in Table 2.1. Shadowing is an abstraction of several effects during the transmission of electromagnetic waves, e.g. reflection, scattering and diffraction. It influences the signal strength randomly, whereby the probability density function of the attenuation factor is known to be *log normal* distributed with a mean of 0 dB and a standard deviation depending again on the considered environment (Table 2.1). Fading stands for the interference of many scattered signals arriving at the receivers antenna. With fading the distortion of the transmitted signal rises significantly. Like shadowing its influence has a stochastic nature, but unlike path loss and shadowing the attenuation changes are more drastic and within smaller time intervals. While path loss and shadowing typically cause a time variance in the scale of a second, fading influences the attenuation in the time scale of milliseconds or even microseconds [4]. In our model the time selective behavior of fading is characterized by a *Jakes-like* power spectral density function, while the frequency selective fading is modeled using an exponential power delay profile. The used fading model is discussed in greater detail in [1].

The basic approach of Multi Carrier Modulation (MCM) schemes, like OFDM, is that the given bandwidth is separated into S subcarriers. During the transmission the digital data is modulated in the shape of digital symbols separately on each subcarrier. The symbols are transmitted in parallel on these subcarriers, during a given time slot, which is called the symbol time T_s. Compared to Single Carrier Modulation (SCM) schemes this has one advantage: In a SCM scheme the data is transmitted in serial, i.e. one symbol is transmitted per T_s cycle using the full available bandwidth. With higher data rates this leads to shorter T_s. In a frequency selective channel this raises the distorting influence of an effect called Intersymbol Interference (ISI), a smearing of the signals which arrive the receiver antenna on various paths [27]. For the longer symbol times of a MCM scheme this effect can be neglected. Orthogonal Frequency Division Multiplexing (OFDM) is a specific form of a MCM scheme. The primarily benefit of this implementation is the high spectral efficiency, which is achieved by the close, overlapping placement of the subcarriers. With OFDM the maximum of the power spectrum for each subcarrier is placed on the zero-crossings of the adjacent spectra (Figure 2.3). This prevents Intercarrier Interference (ICI) and thus there is no need for applying guard bands between the subcarriers, which saves bandwidth. The spacing claims orthogonal subcarrier frequencies and a constant interval between the carrier frequencies (Δf). Due to Equation 2.2 this leads to a constant T_s and therefore a fixed baud rate for all the subcarriers.

$$\Delta f = \frac{1}{T_s} \tag{2.2}$$

Channel Modeling Parameters	Value
Maximum Speed of the WTs (v_{max})	1 m/s
Delay spread standard deviation ($\Delta\sigma$)	0.15 μs
Path loss reference loss ($10 \log K$)	46.7 dB
Path loss exponent (α)	2.4
Shadowing standard deviation (σ)	5.8

Table 2.1: Values of radio channel modeling parameters for an urban environment
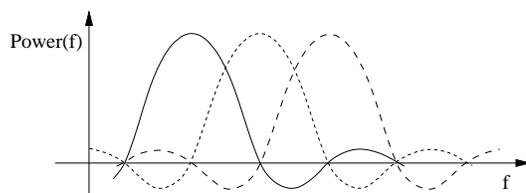
Figure 2.3: Illustration of the power spectrum of three OFDM subcarriers

A block of bits is represented by a single symbol in the modulation alphabet. The number of bits represented depends on the chosen modulation technique (Table 2.2). One criteria for the choice of the appropriate modulation technique is the expected CNR. If the CNR drops, then the probability of a symbol error P_s for a certain modulation technique rises. Since in this case the P_s for higher modulation techniques, i.e. those which represent higher amounts of bits per symbol, rises more drastical than for lower techniques (Figure 2.4) one would choose a lower class modulation technique to stay within appreciable boundaries for the symbol errors. Otherwise, with a high CNR one can employ a more powerful modulation technique, which allows the sender to transmit more bits per subcarrier during T_s. The problem with frequency selective and time variant radio channels is that the above mentioned effects, particularly the fading, permanently cause an intense variation of the CNR. Here it is profitable to adapt the chosen modulation technique to the CNR for a distinct time interval. In principle the assignment of power and modulation techniques is done best by bit loading algorithms [20]. However, we do not consider these algorithms in this report. Instead we assume the transmission power P_{tx} to be constant on each subcarrier for any modulation technique. Then we choose the highest modulation technique, which is able to provide a Symbol Error Probability (SEP) below a certain threshold P_s (Table 2.3). Basically this is equal to assigning each modulation technique a range of CNR values for which this modulation technique will be used (Figure 2.4). This procedure, which is called adaptive modulation, switches the modulation technique according to the measured CNR value x due to the CNR ranges given in Table 2.2. The switching and the measurement of the CNR is done once per MAC frame time T_f for each subcarrier independently. It is not assumed that the switching requires any gap time or any other form of cost. Supposing a variation of the CNR this adaption finally results in a variable bit-rate per subcarrier while a fixed baud-rate is used for all subcarriers.

Modulation Technique	Bits per Symbol	CNR Range
No modulation applied	0	$0 \leq x < 4$ dB
BPSK	1	$4 \leq x < 9$ dB
QPSK	2	$9 \leq x < 16$ dB
16-QAM	4	$16 \leq x < 22$ dB
64-QAM	6	$22 \leq x < 28$ dB
256-QAM	8	$28 \leq x < \infty$ dB

Table 2.2: Considered modulation types, which are applied for the P_s dependent CNR ranges, and the represented number of bits per symbol
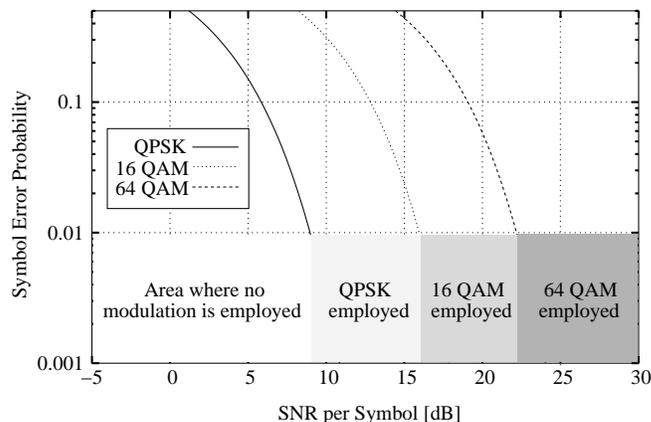
Figure 2.4: Symbol error probability and example CNR modulation ranges as applied for a maximum acceptable P_s of 0.01

The added flexibility due to adaptive modulation shows another benefit of separating the whole bandwidth into subcarriers using an MCM scheme. Further improvements are discussed under the topic of dynamic FDMA in the following chapters 2.3 and 3. The considered values for the physical layer parameters are summarized in Table 2.3. These values correspond to wireless local area networks following the IEEE 802.11a standard [16].

Physical Layer Parameters	Value
Total bandwidth (B)	16.25 MHz in the 5.2 GHz band
Number of subcarriers (S)	48
Maximum transmit power per subcarrier (P_{tx})	0.2 mW (-7 dBm)
Subcarrier spacing	312.5 kHz
Thermal noise power level (n_0)	-117 dBm
Symbol time (T_s)	4 μs
SEP threshold for adaptive modulation (P_s)	0.01

Table 2.3: Values of the physical layer parameters according to the IEEE 802.11a WLAN standard [16]

2.3 Multiple access scheme and medium access control

To provide the capacity of the wireless link to several WTs simultaneously a multiple access scheme is applied on top of the discussed modulation techniques. Due to the fact that in the considered system the access point decides about the amount of link capacity which is given to a specific WT it is denoted as a hierarchical system. There are several ways how these decisions can be made. For example with a Time Division Multiple Access (TDMA) scheme all OFDM subcarriers are assigned to one WT for a limited time interval, i.e. a multiple of a MAC frame (T_f). In this report we will basically consider FDMA whereby each WT obtains

(a) Time Division Multiple Access (TDMA)

(b) Frequency Division Multiple Access (FDMA) with $N = 2$
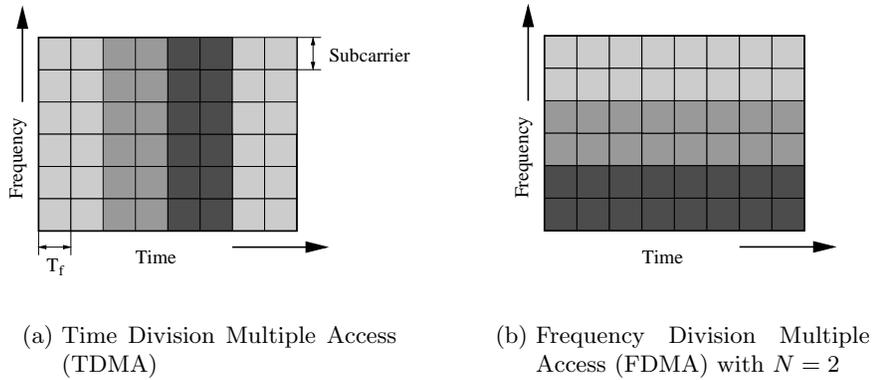
Figure 2.5: Static multiple access schemes with an OFDM system for 3 WTs

(a) Dynamic subcarrier assignment with $N = 2$

(b) Dynamic subcarrier allocation
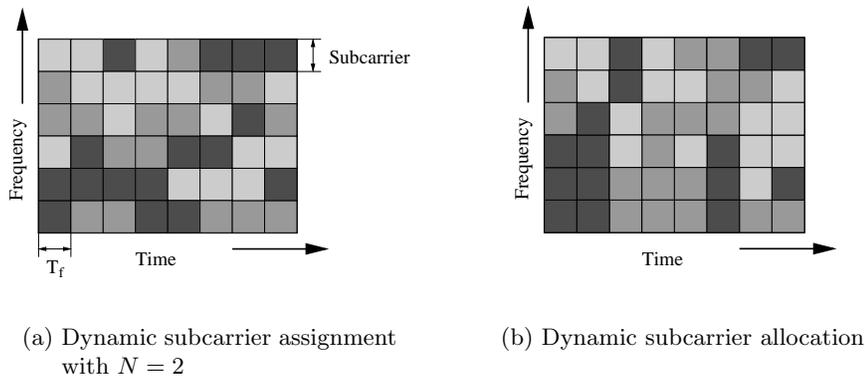
Figure 2.6: Dynamic OFDM-FDMA schemes for 3 WTs

a number of subcarriers (N) during the whole transmission time. From this starting point we will describe improved FDMA schemes, called dynamic OFDM-FDMA, while the described conventional multiple access schemes are called static schemes.

The subcarrier allocation techniques of the static schemes are illustrated in Figure Figure 2.5 for three WTs. The color of a single block denotes one subcarrier obtained by a specific WT for the duration of T_f. One benefit with FDMA is that it allows the dynamic assignment of these blocks to the WTs. While with static FDMA the amount and the spectral position of the OFDM subcarriers, each WT obtains, is fixed, it may vary with a dynamic OFDM-FMDA scheme. In Figure 2.6(a) a dynamic OFDM-FDMA scheme is shown where each of the three WTs receives the same number of subcarriers ($N = 2$) per T_f but their position in the frequency domain varies. In comparison to static OFDM-FDMA this flexible choice of the subcarrier assignments enables adaption to the varying CNR. The channel-state dependent scheduler can choose for each WT those subcarriers with the highest CNR and therefore, due to adaptive modulation, with the highest amount of Bits per symbol. An additional modifi-
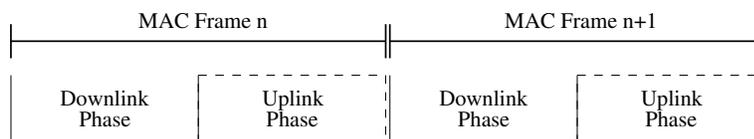
Figure 2.7: MAC frame phases

cation of N, as shown in Figure 2.6(b), enables another adaption technique, which is known as dynamic multiplexing. With the dynamic subcarrier allocation N is adapted to the traffic load. For example, if the size of all outstanding packets in the access point queue for a specific WT is high then more subcarriers will be allocated to this WT. An implementation of both approaches for dynamic OFDM-FDMA systems is explained in greater detail in Chapter 3.

As shown in Figure 2.7 the MAC frame is separated into two equally long phases. During the downlink phase data is transmitted from the access point to the WTs. The uplink phase is reserved for the transmission vice versa. The values for the basic timing structure of a single MAC frame are shown in Table 2.4. For the transmission of data during the downlink phase frame we assume the existence of a complete Data Link Control (DLC) layer including error control and synchronization schemes, which is described in [13]. To fulfill its tasks the DLC adds 8 symbols of overhead for its header and trailer to the downlink phase.

MAC Frame Element	Duration
MAC frame (T_f) [5]	2ms (500 Symbols)
Downlink- and Uplink phase	1ms (250 Symbols)
Downlink phase payload field	0.968ms (242 Symbols)

Table 2.4: Timing structure of a single MAC frame

Finally we will give a short overview of the characteristics of the wired link, i.e. the cable connection between the application servers and the access point. Since we focus on the effects the wireless system causes to the transmission of heterogeneous data streams the wired link is not assumed to be the bottleneck. Therefore, we assume an Ethernet connection, as in IEEE 802.3, between the application servers and the access point. The Maximum Transmission Unit (MTU), i.e. the amount of data transmittable within a single Ethernet frame, is 1500 Bytes. Since we assume the application servers located in the Internet, possible changes in the network structure or in the performance of the nodes may cause unpredictable transmission delays. This, in addition to the variance caused by the 802.3 MAC scheme (CSMA/CD), is modeled adding a gamma-distributed random value to the transmission delay for each Ethernet frame.

2.4 Network and transport layers

Before the access point forwards the data to the WTs via the wireless link, the packets are received from servers in the Internet. This supposes the TCP/IP suite to be employed to provide network and transport layer services. The network layer services are provided by

the Internet Protocol (IP) version 4 – the common standard in the Internet. Transport layer services are provided end-to-end, i.e. for the full transmission path between the sending application server and the finally receiving WT, by the Transmission Control Protocol (TCP) for HTTP transmission and by the User Datagram Protocol (UDP) for the transmission of MPEG-4 video streams. For the simulation of the TCP/IP suite we use the model developed at the university of Karlsruhe [34] based on the simulation software OmNet++ [32]. This model includes TCPs flow control mechanisms such as *slow-start* and the congestion avoidance scheme from [18], which are detailed described by Stevens in [30].

The used TCP version in our system is *New-Reno* [7], which is implemented in the most up-to-date operation systems [23]. Since in this report we focus on the effects of the wireless link, fragmentation on the cable link is avoided by limiting the size of the TCP packets and the UDP datagram. With the given Ethernet MTU of 1500 Bytes and the size of the headers for IP and TCP or UDP this results in 1460 Bytes for TCPs Maximum Segment Size (MSS) and 1472 Bytes data per output operation to UDP.

2.5 Heterogeneous application level traffic load

In the considered system the assumed *heterogeneous* application level traffic stands for the simultaneous transmission of Web-pages and video streams. For the transmission of Web-pages we assume the Hypertext Transfer Protocol (HTTP) version 1.0 as proposed in [2] at the application layer. It is the foundation protocol of the World-Wide-Web (WWW) and can be used to transmit data in the form of hypertext (e.g. HTML), plain text, images and files of any type, which is summarized using the term *Web-page* in this report. HTTP is a transaction oriented protocol, which relies on TCP. For each transaction, independently performed between the HTTP client and the Web-server, a new TCP connection is established. This stateless design avoids that the Web-server, which may serve many thousands of clients for a huge Web-site, has to store and manage the state information for each HTTP connection.

While the structure of a transaction has to be considered in the HTTP application layer model, the key aspect is the user behavior, since finally the user decides which Web-page is requested at which point in time. For the representation of the user behavior in this report basically the *Malaga* model as proposed in [28] is used. *Malaga* classifies the WWW traffic structure in session- and page-levels. A session is defined as a working session of an user from the starting time of the Web-browser until its termination. With *Malaga* sessions are considered to be non-overlapped. On the page level single Web-pages, containing a set of files (hypertext, images, etc.) are requested and received. On the session level the WWW traffic model defines the parameters session-inter-time, i.e. the time between two consecutive sessions, and the number of requested Web-pages per session. On the page level the page-inter-time, i.e. the time between two page requests, and the page size, i.e. the total amount of requested information per Web-page, are parameterized. The distribution functions and the related values for these parameters are summarized in Table 2.5. In this technical report we use the values obtained for the corporate environment which are more load intensive than those which *Malaga* lists for the educational environment. With this traffic model the desired traffic load on the whole system can be justified by changing the mean of the session-inter-time and the number of WTs receiving Web-pages (J_h). While J_h is varied in the performance

TKN-04-017   Page 10

Session Level Parameter	Distribution Function	Distribution F. Parameters
Session-inter-time	Exponential	$\mu = 3$ s
Pages per session	Log normal	$\mu = 25.807$ pages/session
		$\sigma = 78.752$ pages/session
Page Level Parameters	Distribution Function	Distribution F. Parameters
Page-inter-time	Gamma	$\mu = 35.286$ s
		$\sigma = 147.39$ s
Page Size	Pareto	$\alpha = 1.7584$
		$\beta = 30458$ Bytes

Table 2.5: Definition and values of the WWW traffic model as defined in [28]

studies we fix the mean for the session-inter-time to 3 seconds in this technical report.

For the transmission of the MPEG-4 coded video streams UDP is employed. Modern video coding methods, like MPEG-4, are taking advantage of two classes of redundancy within the video streams. At first the spatial redundancy within a single original picture, called a video frame, is exploited using the mechanisms as summarized in Figure 2.8. This encoding process, known as intra-frame coding, leads to encoded video frames which are called I-frames. Secondly the temporal redundancy, i.e. the difference between consecutive frames of the original video, is exploited by methods like motion estimation, motion prediction or frame differencing, summarized by the term inter-frame coding [24].

In addition to the I-frames the inter-frame coding with MPEG-4 leads to two frame types: P- and B-frames. Unlike the intra-frame coded I-frames, which are related to a single picture of the decoded video, these frames are predicted from other frames in the MPEG-4 video stream. Like illustrated by the arrows in Figure 2.9 P-frames are predicted by the use of the previous I- or P-frame, while the *bidirectional* predicted B-frames are depending on the previous and the following I- or P-frame. A consequence of these prediction methods is that I-frames contain the highest amount of information and are the basis for the reconstruction of the video stream. The entropy then decreases via the P-frames to the B-frames, which contain the least information. Thus, a single I-frame has to be transmitted using a much greater number of UDP-datagrams than the other two frame types. The structure in which the frames are arranged is called Group of Pictures (GOP). For the GOP, which is periodic repeated in the video stream, several types exist [6]. In this technical report the 12-element GOP as shown in Table 2.6, which is common with MPEG, is considered. The shape of this
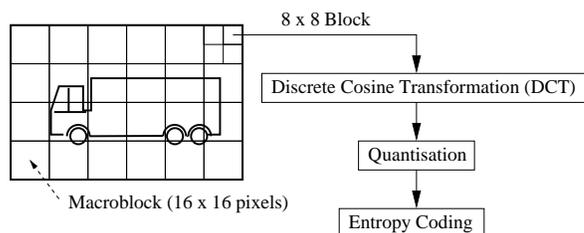
Figure 2.8: Intra-frame coding process

(a) Forward prediction of P-frames

(b) Bidirectional prediction of B-frames
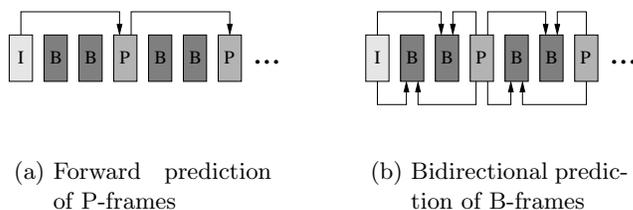
Figure 2.9: Prediction methods as applied in MPEG-4 and the resulting frame relationships

structure is interesting if errors occur in the video stream. We will come back to this aspect in Section 3.3 when semantic video queue management schemes are discussed.

For the video traffic model we consider a single high-motion Variable Bit-Rate (VBR) video stream to be sent to each of the J_v WTs. A high-motion video source, i.e. a stream containing strong fluctuations between consecutive frames, leads to a higher mean and variance of the bit-rate of the resulting MPEG-4 coded VBR data stream. The MPEG-4 video, with the length of 3 minutes, is in the Common Intermediate Format (CIF) format, which defines the frame size. The chosen values for the mentioned video parameters are summarized in Table 2.6

MPEG-4 Video Parameters	Value
GOP	12-GOP: IBBPBBPBBPBB
Mean bit-rate	951 KBps
Frame size	352×288 Pixel (CIF)
Frame-rate	25 frames/s
Video length	4500 frames, i.e. 180 s

Table 2.6: Applied MPEG-4 video parameters

2.6 Timing structure and encapsulation

The delays considered in the simulated system considering all the discussed layers of the simulated system are summarized in this section.

If a WT demands a video stream, a request is sent to the video server. Since no uplink is considered in the system model the request is sent directly from the access point to the video server for which a duration of T_i is needed. After a service delay T_h at the video server, e.g. for video encoding or hard-disc seeking purposes, the first part of the MPEG-4 coded video is send within an UDP datagram. Depending on the size of the transmitted video frame the UDP datagram may contain a full video frame or a part of it. After the encapsulation the resulting IP datagram is sent via the wired link to the access point were it is queued for the duration of T_q.

Depending on the results of the adaptive modulation a specific part of the IP datagram

can be sent to the WT during T_f. This part is modulated and the resulting symbols are transmitted simultaneously on the OFDM subcarriers via the wireless link. The transmission of the whole encapsulated IP datagram lasts T_w. As a simplification it is not assumed that the decapsulation and data processing, e.g. the decoding of the received video frame causes any delays on the WT. Thus, the delays which occur during the transmission can be summarized as follows:

$$RTT = 2 \times T_i + T_h + T_q + T_w. \tag{2.3}$$

In general the focus of this technical report is the investigation and optimization of T_q within the introduced system.

For the request and the transmission of a Web-page the timing structure is the same. However, since TCP is employed on the whole transmission path the whole Web-page or parts of it are encapsulated within a TCP segment.

Chapter 3

Cross layer optimization techniques

For the described transmission system several dynamic OFDM-FDMA schemes were developed at the TKN group [10], [9] to provide an increase of the downlink-capacity for all WTs in the cell. This is basically done at the access point by the use of adaption techniques. Although these schemes are located on the MAC layer, they depend on cross-layer-information provided by other layers. The *dynamic subcarrier assignment* enables the adaption of the spectral position of the subcarriers chosen per WT to the varying CNR. This channel-state-dependent scheduling approach, firstly investigated by Bhagwat et al. [3], makes use of the subcarrier states, as provided by the physical layer.

As introduced in Section 2.3 channel-state-dependent schedulers are often combined with dynamic- or statistic multiplexing schemes [21], [11], [26]. In this technical report the dynamic subcarrier assignment scheme is done in front of the *dynamic subcarrier allocation*. Here the number of subcarriers assigned to a specific WT (N) is adapted to the varying traffic load, which is obtained by measuring the queue length on the DLC layer. These schemes assume a signaling system to be present which provides the WTs with the resulting assignment- and allocation-decisions that are made at the access point. One possible system, assumed in this technical report is specified in [13] and investigated in [12].

In comparison to the dynamic OFDM-FDMA schemes, which provide the adaptive management of the channel resources, a *VQM* scheme enables the adaption of the scheduling policy to the semantics of the traffic stream, as resulting from modern video compression techniques as MPEG-4. Thus knowledge of application layer information on the link layer at the access point, where the VQM is located, is assumed. In Figure 3.1 the mentioned
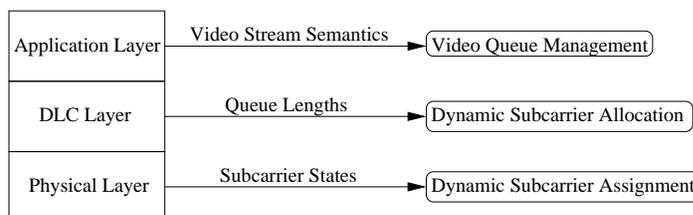
Figure 3.1: List of the considered optimization techniques and the used cross-layer-information

TKN-04-017
Page 14

optimization techniques and the related cross-layer-information are listed. In the following sections we discuss the basic algorithms of the OFDM-FDMA optimization techniques investigated in this technical report. Furthermore a functional description of a VQM scheme is given which is the building blocks for the new VQM method introduced in Chapter 5.

3.1 Dynamic subcarrier assignment

In Section 2.2 we introduced the channel gain (G), expressing the attenuation and noise levels of the radio channel by the use of CNR values. Due to the effects caused by multi-path propagation within the radio channel and due to movement, the channel gain varies over time and frequency $(G(f,t))$. Therefore the resulting CNR per subcarrier (s), regarding to a specific WT (j), can be given by $g_{j,s}(t)$ as in Equation 3.1.

$$g_{j,s}(t) = \frac{a_{j,s}^2}{n_{j,s}^2} \tag{3.1}$$

For all subcarriers regarding to all WTs, considering J WTs and S subcarriers, this leads to one CNR-matrix $\hat{G}(t)$ per MAC frame time of dimension $J \times S$:

$$\hat{G}(t) = \begin{pmatrix} g_{1,1}(t) & g_{1,2}(t) & \cdots & g_{1,S}(t) \\ g_{2,1}(t) & g_{2,2}(t) & \cdots & g_{2,S}(t) \\ \vdots & \vdots & & \vdots \\ g_{J,1}(t) & g_{J,2}(t) & \cdots & g_{J,S}(t) \end{pmatrix} \tag{3.2}$$

By multiplying each value of $\hat{G}(t)$ with the applied transmission power P_{tx}, that is constant for all the subcarriers during the whole transmission, the matrix $S(t)$ is obtained containing the Signal-to-Noise Ratio (SNR) values for the transmission towards all J WTs on all S subcarriers. From the SNR in $S(t)$ the adaptive modulation scheme derives the applied modulation technique, which results to a number of bits (c) represented by a symbol of the chosen modulation alphabet. Since the resulting c, called the *subcarrier state*, is specific for the considered WT and subcarrier, all $c_{j,s}$ can be expressed by the $J \times S$ matrix $C(t)$.

The matrix $C(t)$ is then used by the dynamic subcarrier allocation algorithm to calculate the subcarrier assignment for each WT for the next downlink phase $(t+1)$. Therefore the assignment algorithm has to derive the assignment from possibly out-dated subcarrier state information obtained by the measurements from the preceding downlink phase (Figure 2.7). For the assignment algorithm we consider the Advanced Dynamic Algorithm (ADA). The following functional description of ADA is an excerpt from [9], where the algorithm is described in greater detail: ADA makes use of priorities for each WT. For each WT a weight value (w) is calculated, where a single weight value for the specific WT j is obtained by the sum of the channel gain values of the i WTs with a lower priority than j:

$$w_{j,s}(t) = \sum_{\forall i} g_{i,s}(t). \tag{3.3}$$

After the calculation of the weights for all J WTs the assignment algorithm selects the subcarriers with the highest possible weight ratio, defined by

$$\frac{g_{i,s}(t)}{w_{i,s}(t)}. \tag{3.4}$$

This is done N times for each WT, starting with the WT with the highest priority. Therefore, each WT obtains N subcarriers, where to those with higher priorities the best subcarriers from the whole set of the size S are assigned. Finally the WT priorities are rotated to provide fairness.

Although ADA provides a heuristic and not the optimal solution for the subcarrier assignment itself it was shown that ADA are most only 5% off from the throughput gain achieved on the link layer by use of the optimal solution [9]. The throughput gain due to the adaption to the varying CNR with ADA in comparison to the static OFDM-FDMA scheme is from 30 to 40% [9]. We will see in Chapter 4 what gain is achievable on the application layer in combination with heterogeneous traffic load.

3.2 Dynamic subcarrier allocation

For the description of the dynamic subcarrier assignment scheme we supposed a constant number of subcarriers (N) given to each WT during the assignment procedure. However, with VBR traffic load this static allocation of subcarriers is far from being optimal. For the scenario where all WTs per cell are receiving MPEG-4 coded VBR video streams it was shown, that with dynamic subcarrier assignment combined with dynamic subcarrier allocation a capacity improvement of up to 300% can be achieved [10]. Basically this is due to the exploitation of the statistical multiplex, that is possible with packet switched networks, for the resource allocation. With a static allocation scheme each WT receives N subcarriers per downlink phase regardless of data has to be send to a WT or not. If, for example, no data is send to i WTs for a specific amount of downlink phases $i \times N$ subcarriers are lost for the transmission system. In contrast, a dynamic subcarrier allocation scheme then allocates these $i \times N$ subcarriers to those WTs which have to receive data. This will clearly result in an improvement of the throughput rates to these WTs.

In this technical report the dynamic subcarrier allocation scheme, as proposed in [10], is considered. It provides the adaption to the varying traffic load by measuring the queue length (d_j), i.e. the sum of all the packet sizes in the access point queue, for a single WT. The dynamic allocation scheme then compares the obtained d_j values to the sum of all d_j and calculates the amount of subcarriers allocated per WT (s_j) as given in Equation 3.5.

$$s_j(t) = 1 + \left\lfloor (S - J_a) \frac{d_j(t)}{\sum_{\forall J} d_j(t)} + 0.5 \right\rfloor \tag{3.5}$$

Since prior to this allocation, that is derived from the queue lengths, each WT which has data to receive during the next downlink phase, whose sum is denoted by J_a, obtains one subcarrier, only $S - J_a$ subcarriers can be weighted in Equation 3.5. The allocation is done once per MAC frame before the downlink phase starts, for which the achieved subcarrier allocations are applied.

In Section 4.4 we will see what benefits can be achieved for the transmission of HTTP using the combination of dynamic subcarrier assignment and dynamic subcarrier allocation. In Section 4.5 this combination is analyzed for heterogeneous traffic load, which consists of HTTP and VBR video streams.

3.3 Video queue management

In general we use the term Video Queue Management (VQM) to describe a number of functions employed on those queues, located at the access point, which contain packets of video stream data. It is supposed that even in scenarios with heterogeneous traffic the access point knows which queues contain video packets. Furthermore the access point is able to extract parameters from these packets, which have a semantic relevance for the related video stream. For example, with the MPEG-4 coded video stream in the considered system, this may be the type of the video frame which is (partially) included in an UDP datagram. Extracting this information the VQM on the access point knows whether this UDP datagram and the related IP datagram belongs to an I-,P- or B-frame and may adapt its scheduling policy in order to this information. In context of a VQM the scheduling policy decides which priority is given to the transmission of a particular video packet.

A simple form is a First-In-First-Out (FIFO) scheduling policy, where the access point transmits the packets in order of the reception. This is the basic approach of the first VQM we want to introduce: the *simple VQM*. The only function which distinguishes the simple VQM from a plain FIFO queue, as applied for non-video data transmission, is the removal of packets related to outdated video frames. This function is the consequence of the strict boundaries for the transmission delay on the downlink if videos in real-time application scenarios, e.g. video conferences, are considered. If the transmission delay exceeds 180 ms the distorting effect of acoustic echoes becomes annoying to the user [33]. Furthermore time lags in the video stream may become visible, since the video decoder has to wait until all packets per video frame have arrived. For the prevention of problems due to boundless transmission delays we assume that the video frames or parts of them, which are received later than a given global deadline ($\hat{D}$) are not decoded at the receiver anymore. Therefore the simple VQM may remove video packets whose transmission delay exceeds $\hat{D}$. Because the VQM is located at the access point it has no knowledge about the absolute end-to-end transmission delay of a packet. Therefore the value for $\hat{D}$ has to be much smaller than the overall acceptable transmission delay of 180 ms. In this technical report we chose $\hat{D} = 100$ ms, which is quite pessimistic. For the calculation of $\hat{D}$ using a timestamp the synchronization of the WT- and the access point-clock is assumed. The simple approach of this VQM can be formalized as in Figure 3.2. Considering a single packet with the sending time T_{tx} the algorithm calculates
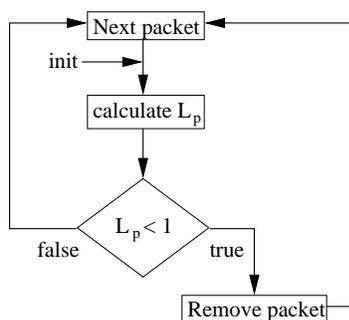
Figure 3.2: Algorithm of the simple VQM as applied for each access point queue

the *lifetime-index* L_p as in Equation 3.6.

$$L_p(t) = \frac{\hat{D}}{t - T_{tx}} \qquad (3.6)$$

If then $L_p < 1$, which means that the transmission delay is larger than the global deadline, the packet is removed from the access point queue. The algorithm is performed for each queue independently. It starts with the first packet at the tail of the queue and ends if all packets are inspected.

In addition to the errors which may occur during the encoding and the transmission of the video frame, the removal of IP datagrams due to exceeded deadlines, causes errors in the MPEG-4 video stream. If an error occurs the decoder has to resynchronize to the bit-stream and the missing information has to be replaced. Therefore, an *error concealment* scheme is performed. In the simplest case the replacement is done with monochrome blocks, while more complex error concealment schemes considering the data from the preceding and following video frames [24]. Usually this is done on the Macro-block level, which means that in case of an error, the whole affected Macro-block is rejected. Beneath the quality of the decoder the effect of errors to a MPEG-4 video stream highly depends on the type of the affected frame. If an error occurs in an I-frame it propagates through the video streams until the next I-frame appears. The reason for the so called *error propagation* is, that all P- and B-frames within the GOP are predicted from the preceding I-frame (Figure 2.9). For the typical 12-GOP this lasts into 11 also affected frames due to one erroneous I-frame. The effect of errors within a P-frame is similar: If the first P-frame in the 12-GOP is affected the error propagates through all B-frames and the following P-frame. In case of an error in the second P-frame of the 12-GOP only the two last B-frames are also affected (Figure 2.9(b)).

One approach for a more sophisticated VQM scheme is to avoid the removal of I-frames and then P-frames due to the exceeded deadline. This means that semantic information of the video stream, i.e. the video frame type, is considered within the scheduling policy. This leads to the second VQM, discussed in this technical report, which is called *semantic VQM*. The basic idea of the semantic VQM is, as proposed in [19], that semantically more important frames are prioritized. Therefore the VQM manages the order of the transmission of the IP datagrams in the access point queue: If the queue contains datagrams related to I-frames they are transmitted first, followed by datagrams containing parts of P-frames. Finally the datagrams which are related to B-frames are transmitted. Furthermore, with the semantic VQM, the deadline, which defines when a packet is deleted from the queue, becomes frame type dependent. Therefore a weighting factor (w) is added to Equation 3.6, which leads to

$$\hat{L}_p(t) = w(\Phi)\frac{\hat{D}}{t - T_{tx}} \qquad (3.7)$$

for the calculation of the lifetime-index. For the weights the values in Equation 3.8 are considered, which were defined originally in [19]. The Φ denotes the frame type, which is extracted from the considered packet.

$$w(\Phi) = \begin{cases} 1, & \Phi = \text{I Frame} \\ 0.75, & \Phi = \text{P Frame} \\ 0.5, & \Phi = \text{B Frame} \end{cases} \qquad (3.8)$$

This means, that P-frames are dropped 25% earlier than I-frames and B-frames are dropped 50% earlier. Therefore, additional capacity for I- and P-frames is provided by the deletion of packets related to B-frames. Since B-frames are no prediction-source for other frames the errors caused by the removal of the packets will not propagate. The algorithm of the semantic VQM, as it is independently performed for each access point queue containing video data, is illustrated in Figure 3.3: At first all video packets in the queue are sorted according to the type of the related video frame. The packets which are related to I-frames are shifted to the tail of the queue, followed by the packets related to P- and B-frames. Then basically the same iteration over all packets in the queue as in the simple VQM starts. However, w is considered in the calculation of $\hat{L}_p$ in step 2. If the frame specific deadline for a packet is hit ($\hat{L}_p < 1$), it is dropped during step 3. The algorithm ends if the head of the queue is reached. Although this algorithm is performed once per downlink phase in the considered system this is not essential. An more infrequent execution would decrease the calculation-overhead caused by the semantic VQM to the access point, but it would also lower the adaption capabilities of the scheme to the actual queue content.

The two discussed VQM schemes are the building blocks and the yardstick for the VQM scheme proposed in Chapter 5. A newly developed VQM should clearly outperform both of them. This is investigated in the performance study in Chapter 6.
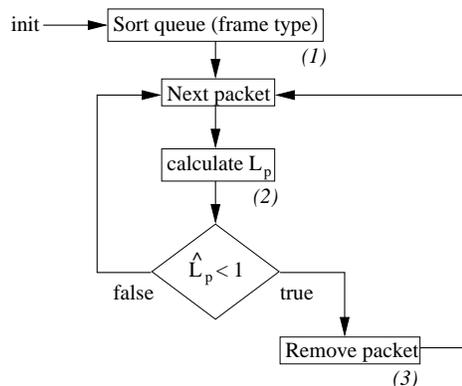
Figure 3.3: Algorithm of the semantic VQM, applied for each access point queue

TKN-04-017

Page 19

Chapter 4

Performance study of dynamic OFDM-FDMA schemes

So far the performance of the dynamic OFDM-FDMA schemes, as discussed in Chapter 3, was investigated for a homogeneous video scenario in [11]. This means, that MPEG-4 coded video streams are send to all WTs in the cell. Taking the transmission of Web-pages via the HTTP and the TCP protocol into account leads to the following two questions:

- How does dynamic OFDM-FDMA performs considering only the transmission of HTTP data?

- How does dynamic OFDM-FDMA performs considering the simultaneous transmission of HTTP data and video streams?

To answer these questions two scenarios are considered in this performance study: In Section 4.4 a homogeneous Web-traffic scenario is investigated, where all J_h WTs in the cell are receiving Web-pages servers located in the Internet. Therefore, the traffic model as explained in Section 2.5 is applied. The heterogeneous scenario, investigated in Section 4.5, provides a closer expression of the Internet traffic characteristics: Here J_h WTs receive Web-pages, while the rest of the WTs in the cell receives MPEG-4 coded VBR video streams. This traffic mixture is interesting due to the totally different claims both parts have to the transmission system:

- The transmission of Web-pages is based on TCP, which provides reliable transmission regardless to the transmission delay caused by Automatic Repeat Request (ARQ) in case of errors. Furthermore TCPs flow control adapts its sending rate in order to avoid of congestion.

- In contrast to TCP the UDP, as applied for the video transmission, uses no flow control. Basically the sender bit-rate depends on the application. If demanded by the sender, as much bandwidth as provided by the lower layers is used. As discussed in Section 3.3 for real-time video streaming the faultless transaction has much lower priority than the fast transmission of the data.

Considering the antagonistic characteristic of the transport layer protocols leads to the assumption, that in cases with high amounts of UDP traffic the bit-rate of TCP will drop significantly. Therefore, the heterogeneous scenario is compared to the homogeneous case.

4.1 Transmission metrics

In this section the metrics, which we will consider in both performance studies of this technical report, are defined. Since what finally counts is the benefit the user perceives at the application layer, we will consider two aspects here: At first the system is investigated at the application layer, which leads to separate metrics for the transmission of HTTP and MPEG-4 video. Secondly we will define metrics which are close to the perception of the user, for example, the Web-page transmission time and the Distortion In Interval (DIV).

4.1.1 HTTP transmission metrics

In order to measure the performance provided by the transmission system to those WTs which receive Web-pages, we consider the metrics as listed in Table 4.1. In [22] Menascé et al. came to the conclusion that latency and throughput are the two most important performance metrics for Web-systems. These two classes are considered in this technical report. The throughput is measured in Bits per second (Bps) on the application layer of each WT separately. To get closer to the users perception the latency of a requested Web-page is monitored by measuring the transmission time of a single page. The faster a Web-page is received and displayed, the more the user is satisfied with the transmission system. [1] Furthermore, the transmission delay of a Web-page should be predictable for the user. This is the fewer the case, the higher the delay varies. This variation is represented by the standard deviation (σ) of the Web-page transmission time.

HTTP Metric	Unit
TCP throughput (bit-rate)	Bits per second [Bps]
Web-page transmission time	Seconds [s]
σ of the Web-page transmission time	Seconds [s]

Table 4.1: Considered HTTP transmission metrics

For the presentation in the performance studies the results obtained for all J_h WTs during the whole simulation time are averaged over J_h. In addition to the mean values the variance of the results was investigated for 95% confidence. Due to their small size (less than 0.5% of the plotted value) no confidence intervals are shown in the diagrams given in the performance studies.

---

[1]Here it is assumed, that the client-program on the application layer e.g. the Web-browser does not cause any delays.

4.1.2 MPEG-4 video transmission metrics

At the end of the transmission process the transmitted video is shown to the user. Then the users subjective impression of the video quality is what finally counts. Therefore, the measurement of video quality must base on the quality perceived by the user. Since the subjective evaluation has to be done by tests, where humans have to watch and to grade the video, which are time intensive and costly, we use a different approach here: By applying the DIV metric, as proposed in [19], we are able to consider the subjective impression in an objective metric. The DIV metric is based on the Mean Opinion Score (MOS), which is a standard for the subjective rating of still pictures [17]. By the use of Table 4.2 [25] the subjective metric MOS can be derived from the objective Peak Signal-to-Noise Ratio (PSNR), which is widely used for the evaluation of picture quality [29]. Therewith, in our system the MOS for a single picture of the decoded video can be obtained by calculating its PSNR value.

PSNR [dB]	MOS grade	Video quality
> 37	5	Excellent
31 − 37	4	Good
25 − 31	3	Fair
20 − 25	2	Poor
< 20	1	Bad

Table 4.2: Possible PSNR to MOS conversion as in [25]

To obtain one value for the quality of a large number of still pictures, as in a video stream, the simplest approach would be to obtain one MOS value per video frame and then calculate the mean. Although this metric is comprehensible it usually gives no adequate reflection of the users impression. For example: If the first seconds of the video are lost during the transmission and the quality of frames received during the following larger time interval is excellent, a good MOS value is obtained on the average, which will clearly not match the subjective user impression. To avoid the false estimation as performed by time-averaging metrics the number of transmitted video frames with a lower MOS grade than before the transmission process is counted for a given time interval. This assumes the video quality to be always excellent in prior to the transmission. Therefore the DIV expresses the grade of distortion caused by the transmission system as perceived by the viewer of the transmitted video. A detailed discussion of the DIV metric is provided in [11].

In this technical report the threshold of 20% lost frames for the time interval of 20 seconds, as both defined in [11], is considered. Results for the DIV higher than the threshold are considered to lead to an unacceptable decrease of the video quality. If results above this bound are shown it is indicated by a horizontal line at the 20% mark.

In addition to the DIV we will show the packet loss rates for the transmitted video streams. This is done separately for each frame type to illustrate the applied algorithms of the investigated VQM. In the presented results the average over the number of video receiving WTs is shown for both video metrics.

4.2 Simulation framework

All results presented in the performance studies were obtained via discrete event simulation. Figure 4.1 shows the structure of the simulated system and illustrates how the modules are interconnected. The system consists of application servers, the access point and the wireless terminals. Each application server module consists basically of a traffic generator and a transport protocol submodule. Depending on the type of the sent traffic this is either a WWW traffic generator, according to the *Malaga* WWW traffic model (Section 2.5) combined with a TCP submodule or the video traffic model and the UDP submodule. The video traffic generator is based on trace-files which were produced using MPEG-4 coded high-motion VBR video streams. From the server modules messages are sent to the queues which are located at the access point. While different types of traffic use different queues, one queue for each WT is assumed. The modules for the queues, the application servers and for the WTs are replicated according to the chosen numbers of video and WWW terminals.

Beneath the queues the basic functions of the access point can be divided into two modules. In the channel module the CNR is generated considering the effects path loss, shadowing and fading. From the resulting CNR values the channel states are derived using the adaptive modulation scheme as described in Section 2.2 while a constant transmission power is assumed. The channel states are then sent to the scheduler module. It contains three submodules where the subcarrier assignments (AS), the allocation of subcarriers (AL) and the VQM are performed. Several submodules can be used, containing e.g. static or dynamic algorithms. According to the results of the submodules for the used subcarriers (AS), the number of subcarriers (AL) or the index of the removed video packets (VQM) the scheduler interacts with the queues. This interaction and calculation process is done cyclic every T_f and is followed by the transmission of the message at the head of each queue. This is done simultaneous to the J WT modules where the messages are received, processed (according to the used transport protocol) and statistics for the measured values are calculated.
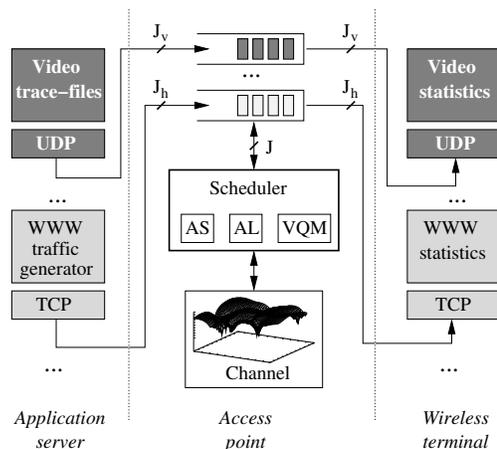
Figure 4.1: Basic structure of the applied discrete event simulator

4.3   Scenario setup and simulation parameters

Within the first performance study we will investigate three possible combinations of the dynamic OFDM-FDMA schemes:

1. Static OFDM-FDMA: static subcarrier assignment and static subcarrier allocation, the spectral position and the number of the subcarriers a single WT (N) obtains is fixed

2. Semi-dynamic OFDM-FDMA: dynamic subcarrier assignment and static subcarrier allocation, the spectral position of a subcarrier is adapted to the channel-states, N is fixed

3. Dynamic OFDM-FDMA: dynamic subcarrier assignment and dynamic subcarrier allocation, additionally N is adapted to the traffic load

As mentioned at the beginning of this chapter we will consider two scenarios for these three combinations of OFDM-FDMA: In the first scenario all J_h WTs in the cell will receive Web-pages using the HTTP and the TCP protocol. The number of WTs is varied from 1 to the maximum number of WTs per cell (J_{max}) as given in Table 4.3. In the second scenario heterogeneous traffic is considered: Here a fixed number of WTs (J_h) receives Web-pages, while a number of WTs (J_v) simultaneously receives MPEG-4 coded VBR video streams. For each scenario and OFDM-FDMA scheme a series is recorded by varying J_v from 0 to $J_{max} - J_h$.

For the system model during the simulation all parameters as given in Chapter 2 are applied. Table 4.3 lists those parameters which are specific for the chosen scenarios. Since the focus of this performance study is not to investigate the VQM schemes, as proposed in Section 3.3 and Chapter 5, the simple VQM is applied for all the simulations of the heterogeneous scenario.

Scenario Parameter	Value
Maximum amount of WTs per cell (J_{max})	48
J_h in the homogeneous setup	varied [1 : 48]
J_h in the heterogeneous setup	12
J_v in the heterogeneous setup	varied [1 : 36]

Table 4.3: Scenario parameters as applied for the OFDM-FDMA performance study

4.4   Results for HTTP traffic streams

In this section the homogeneous case is considered where all the WTs are receiving Web-pages using HTTP. Since with higher values for J_h the limited amount of subcarriers has to be distributed to more and more WTs the average throughput rate per WT slopes. Compared to the static and the semi-dynamic OFDM-FDMA results a clear benefit for the dynamic OFDM-FDMA scheme can be found for the average throughput, shown in Figure 4.2. For both metrics, the performance gain rises slightly with the number of WTs. It is interesting,

that by the use of the dynamic subcarrier assignment compared to the static scheme a higher gain (up to 16% with $J_h = 16$ for the bit-rate) is achieved than with full dynamic OFDM-FDMA compared to the semi-dynamic scheme (up to 5%). Finally, with dynamic OFDM-FDMA for $J_h = 48$ a maximum gain of 30% is achievable for the TCP bit-rate at the HTTP receiving WTs.
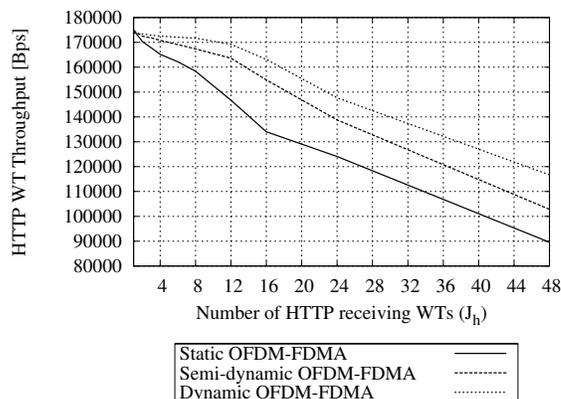
Figure 4.2: Bit-rate for different OFDM-FDMA schemes in a HTTP traffic load scenario

The average transmission time of a single Web-page is shown in Figure 4.3. For all considered OFDM-FDMA schemes it rises with J_h. The comparison of the dynamic OFDM-FDMA system to the static and the semi-dynamic case leads to the conclusion that for all considered numbers of J_h the page transmission time obtained with the dynamic OFDM-FDMA scheme, stays below those results of the other two MAC schemes. Furthermore the performance gain for dynamic OFDM-FDMA rises with J_h. The maximum page transmission time achieved with dynamic OFDM-FDMA for $J_h = 48$ is 140% smaller than the value achieved with the static scheme and 21% smaller than the transmission time achieved with the semi-dynamic scheme. This shows, that even in the case where the dynamic allocation algorithm is able to distribute only a single subcarrier per WT on the average, the statistical multiplex leads to an increase of performance. This is caused by the high variance of the bit-rate due to TCPs flow control and the burstyness of the considered Web-model. For the standard deviation of the page transmission time (Figure 4.4), which expresses the predictability of the Web-page transmission delay, a similar behavior as in Figure 4.3 is shown. Here the values obtained for the static scheme are rising much faster than those for the schemes which employ dynamic subcarrier assignment. Comparing the values obtained for the dynamic case with those of the semi-dynamic OFDM-FDMA scheme illustrates again the effect the exploitation of the statistical multiplex has to the HTTP transmission metrics.
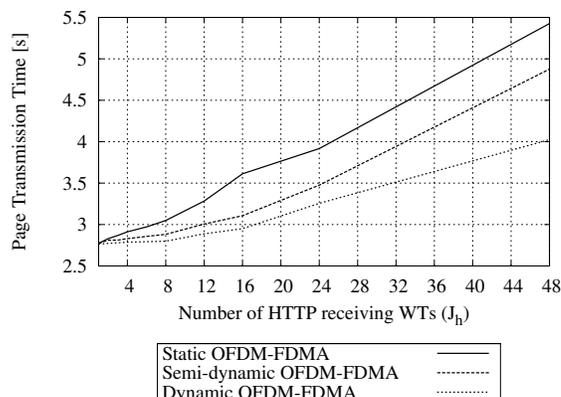
Figure 4.3: Average transmission time per Web-page for different OFDM-FDMA schemes in a HTTP traffic load scenario
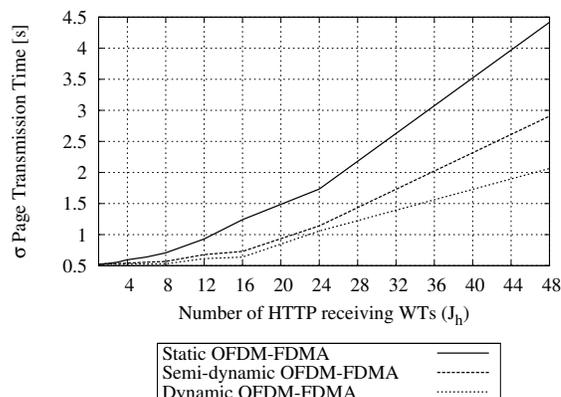
Figure 4.4: Standard deviation of the transmission time per Web-page for different OFDM-FDMA schemes in a HTTP traffic load scenario

4.5 Results for heterogeneous traffic streams

The following results are obtained for the heterogeneous scenario where in addition to the 12 WTs which receive HTTP, a variable amount of video receiving WTs (J_v) is considered (shown on the x-axis). As in the homogeneous case in general the bit-rate as shown in Figure 4.5 falls for all considered OFDM-FDMA schemes due to the limited channel resources. The highest gain for the bit-rate is 80%, achieved for $J_v = 36$. Compared to the scenario with homogeneous traffic schemes, the values which are obtained for dynamic OFDM-FDMA show a totally different behavior for higher numbers of J_v. Since in the heterogeneous case the bit-rate which is achieved for dynamic OFDM-FDMA falls strongly for J_v greater than 28 this leads to the conclusion, that the benefit due to the statistical multiplex is lower with the transmission of many VBR video streams in parallel to the Web-traffic.

In Figure 4.6 the average transmission time of a Web-page per WT is shown. The arrangement of the series, measured for the three different OFDM-FDMA schemes, is similar to the

TKN-04-017
Page 26
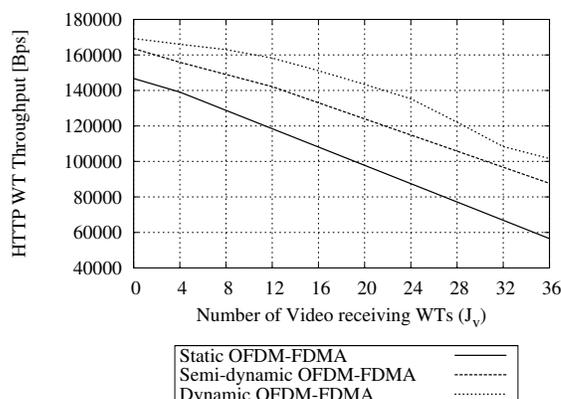
Figure 4.5: Bit-rate of HTTP receiving WTs for different OFDM-FDMA schemes in a heterogeneous traffic load scenario
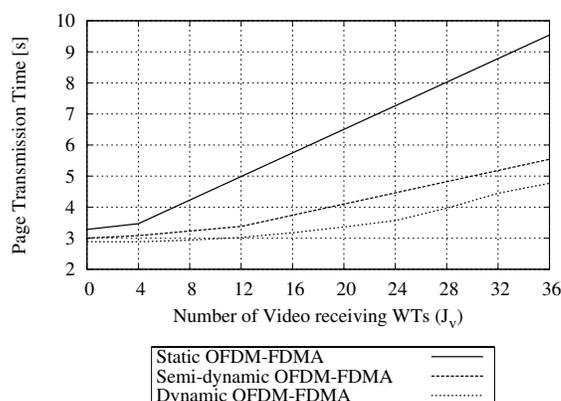
Figure 4.6: Average transmission time per Web-page for different OFDM-FDMA schemes in a heterogeneous traffic load scenario

homogeneous case: On the average the dynamic OFDM-FDMA scheme leads to the fastest transmission of a Web-page followed by the semi-dynamic and by the static scheme. The gain achieved with dynamic allocation in comparison to the dynamic assignment of subcarriers is the highest with 100% for $J_v = 36$. However, compared to the improvement of up to 140% for the same amount of WTs in the cell ($J = 48$) in the homogeneous case the maximum benefit achieved in the heterogeneous scenario is much smaller. The standard deviation of the page transmission time (Figure 4.7) shows a slightly different qualitative behavior than the page transmission time. The dynamic OFDM-FDMA scheme provides a much smaller variation of the page transmission time than the simple FDMA schemes. The gain achieved by the dynamic subcarrier assignment is much higher (up to 215% in comparison to the static scheme) than the additional gain achieved by the dynamic subcarrier allocation (up to 105% in comparison to the semi-dynamic scheme).

The investigation of the DIV metric, which expresses the amount of distortion caused by the transmission system to the video stream, is interesting since it shows the effect the choice

TKN-04-017
Page 27
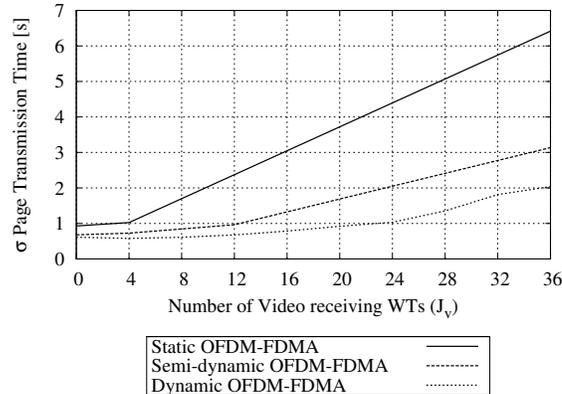
Figure 4.7: Standard deviation of the transmission time per Web-page for different OFDM-FDMA schemes in a heterogeneous traffic load scenario
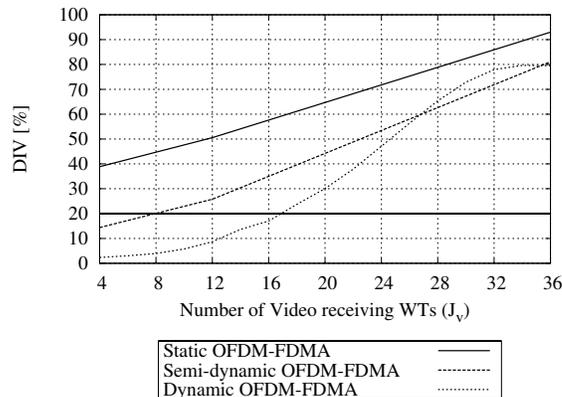
Figure 4.8: DIV for different OFDM-FDMA schemes in a heterogeneous traffic load scenario

of the OFDM-FDMA scheme has to the video quality. In Figure 4.8 the positive impact is significant for the DIV results obtained with dynamic OFDM-FDMA for low numbers of video receiving WTs. For example, for $J_v = 4$ the DIV is 37% lower with the dynamic approach. If J_v rises the DIV also rises for all transmission schemes. The behavior of the DIV values obtained for high J_v with the dynamic OFDM-FDMA scheme is interesting: For more than 32 video receiving WTs the curve flattens until the slope becomes constant for $J_v = 36$. However, this behavior is not relevant for the performance analysis due to the fact that DIV values higher than 20% stand for a very bad video quality. Finally the comparison of the DIV for all the applied OFDM-FDMA schemes leads to the conclusion that for scenarios with small numbers of video receiving WTs, which means small load to the transmission system, dynamic OFDM-FDMA schemes clearly outperform the static case. In scenarios with high traffic load, the distortion of the video quality can not be prevented by dynamic OFDM-FDMA schemes.

The investigation of the HTTP metrics in the heterogeneous case shows a clear benefit if dynamic OFDM-FDMA is employed. However, from the comparison of the HTTP results

obtained for the homogeneous and the heterogeneous scenario it follows, that the benefit due to dynamic OFDM-FDMA sinks faster for rising amounts of video receiving WTs than for rising numbers of HTTP receiving WTs.

Chapter 5

Adaptive video queue management

5.1 Problem description

In Section 3.3 we discussed the *simple VQM* using a FIFO scheduling policy and the *semantic VQM* which adapts its scheduling policy to the semantic information, as extracted from video packets in the access point queue. One important aspect of the semantic VQM is that the removal of semantically less important parts of the video stream, e.g. B-frames, provides additionally capacity to more important parts such as I-frames. This is essential in scenarios with high traffic load, which occurs in one of the following cases:

1. Many receiving WTs: As discussed in Chapter 4 the performance gain achieved with the dynamic allocation scheme drops due to the lack of subcarriers.

2. Data streams with a high bit-rate: The queue length for the affected WTs rises. Due to dynamic allocation these WTs receive higher amounts of subcarriers.

3. Bad channel states: Due to low subcarrier CNR only small amounts of bits can be send during T_s.

Since the resources of the radio channel are limited, congestion may occur which leads to higher transmission delays regardless which type of application layer data is transported. Due to its strict timing boundaries (Section 3.3) then especially real-time video transmission will suffer: The rising transmission delay of the IP datagrams, containing parts of video frames, lead to more and more datagrams which will hit their deadline and are therefore removed by the simple or the semantic VQM. In this chapter we will propose a further stage of the semantic VQM, called *adaptive VQM*. The basic idea of this VQM is to free capacity in the case of high traffic load by the preemptive removal of packets from the video queues. In addition to the consideration of semantic information, which is extracted from the video packets in the queue, the scheduling policy is adapted to the amount of traffic load the system estimates. The objective is to avoid congestion in the access point queues from which all types of application layer traffic may profit. This approach is known under the topic of *adaptive filtering* of video streams [14] and was evaluated for wired networks in [15].

In the following section we will introduce all parts of the adaptive VQM for the considered wireless system and explain how they work together. Furthermore aspects as traffic load

estimation, semantic video filters and the related parameters and algorithms are discussed. In Chapter 6 we will investigate the performance of the adaptive VQM in comparison to the simple and the semantic VQM.

5.2 Optimization approach

The adaptive VQM provides the adaption of the scheduling policy to the traffic load on the access point. To fulfill this tasks basically two new mechanisms are added to the plain semantic VQM: The first has to estimate the traffic load to the monitored system and is called *system load estimator*. The second mechanism, called *semantic video filter* works on the video queue under the objective to lower the bit-rate of the contained video stream while assuring small decreases of the video quality. Therefore, semantic parameters are considered by the filter algorithm. The steps of the adaptive VQM are listed in Table 5.1.

In general the estimation of the traffic load can be treated as simple accounting calculation at the access point, i.e. the amount of free resources is compared to the needed resource capacity:

$$\frac{\text{Bits to transmit}}{\text{Bits transmittable per } T_f}. \tag{5.1}$$

If the result of Equation 5.1 is greater than 1 a lack of resources occurs for the considered T_f. The problem with this approach is to decide which amount of Bits is considered in the upper part of the fraction. Neither the full length of the queue nor the mean number of Bits transmittable per T_f will produce useful results. Therefore, in the adaptive VQM a different approach is used, which is presented in the following discussion of the basic algorithm. As shown in Figure 5.1 in the first step the calculation of the *lifetime-index* (L_p) is done for all video packets in the queue. This is done exactly as shown in Equation 3.7 for the plain semantic VQM. Then the basic algorithm of the semantic VQM (Figure 3.3) is performed, where all packets with an outdated lifetime-index are removed. During this so-called *non-load-adaptive packet removal* the size of all removed packets is totalized independently for each WT to $S_j(t)$. This value is a indirect measure of the needed capacity according to Equation 5.1 since it represents the capacity which should have been provided by the transmission system during the preceding cycles. The S_j is used during step 3 for the estimation of the traffic load. Here it is compared to the free downlink capacity of the MAC frame from the preceding downlink phase ($F_j(t-1)$), as shown in Equation 5.2.

$$R_j(t) = \frac{S_j(t)}{F_j(t-1)} \tag{5.2}$$

Step	VQM Component
1	Video packet lifetime calculator
2	Non-load-adaptive packet removal
3	Traffic load estimator
4	Semantic video filter

Table 5.1: Basic steps of the adaptive VQM

Since the VQM runs prior to the dynamic OFDM-FDMA scheme only the value measured for F_j in the preceding downlink phase can be used for the comparison. The $S_j(t)$ value already expresses the result of the traffic load to the system. If $S_j(t)$ is high, many video packets, or those related to more important frame types of greater size, exceed their deadline. If this is combined with a small estimated downlink capacity ($F_j(t-1)$), then additional filtering of the video queues will lower the probability of congestion and therefore the S_j in the next cycle. If the resulting load-rate $R_j(t)$ exceeds a given threshold α then the access point is considered to be overloaded. To limit the distortion of the video stream due to semantic filtering a second threshold β is specified. The distortion-rate

$$P_j(t) = \frac{S_j(t)}{Q_j(t)} \qquad (5.3)$$

where the size of the removed packets is compared to the actual total queue size ($Q_j(t)$) is calculated once per algorithm cycle (step 3). If P_j exceeds the threshold no semantic filter is executed in order to prevent a further distortion of the video. Note that this stop criterion is not applied during step 2 of the algorithm since only outdated packets are removed during the iteration. Briefly, the semantic filter is only executed during a cycle if the load-rate R_j exceeds α and the distortion-rate stays below β (Figure 3.3).

The objective of the semantic video filter is to free system capacity by lowering the bit-rate of the video streams within the access point queue. For this purpose several types of video filters, e.g. *re-quantization-*, *low-pass-* or *color-to-monochrome-filters* were proposed for the usage with streaming data in wired networks [35]. However, the complexity of the filter is not a problem for distributed systems, the added load might be problematic for the access point. Typically the access point has to serve many WTs and is still in a high-load situation if the filter is performed. Therefore, the much simpler class of *frame-drop* filters is chosen for the adaptive VQM. Frame-drop filters delete video frames or parts of them, i.e. packets, to enable adaption. For wired systems this approach is examined in [15].

The filtering algorithm of the adaptive VQM, considers again the type of the video frame to which the packet is related. As shown in Figure 5.2 it consists of two iterations: In the first iteration the filter inspects the whole queue regarding to the frame types and deletes all packets which are related to B-frames. The algorithm returns if it reaches the head of the queue or the distortion-rate is lower than β. If no B-frame is found during the first iteration, a second iteration starts, where the filter deletes all P-frames in the queue. Although the removal of all packets in the queue may cause drastic distortion effects to the video, this has one benefit: If the queue length rises, typically the risk of congestion rises. Therefore, it is profitable to delete more packets if the queue is longer. Due to the strong impact to the quality of the video (error propagation) no I-frames related data is removed by the algorithm. Thus, the filter provides the explicit, preemptive prioritization of I-frame related packets if high load occurs. Compared to the semantic but reactive strategy this approach should minimize the removal of I-frame related packets due to exceeded deadlines. In order to lower the complexity the filter does make no use of further semantic parameters in the video stream. For example, the second iteration could be modified in order to only remove every second P-frame per GOP. On the one hand, this would provide a finer granularity of the changes to the video quality. On the other hand, this results in more calculation complexity due to the maintenance of a GOP index for each packet. Keeping the algorithm complexity as small as

possible allows the execution of the adaptive VQM every T_f. This is not essential but will provide the most actual information to the load estimator.

The choice of the thresholds α and β is critical for the function of the adaptive VQM. The optimization for a single scenario with $J_h = J_v = 12$ shows acceptable results for $\alpha = 5$ and $\beta = 0.1$. Since these values are used for all scenarios investigated in this technical report they represent not more than heuristics. However, the optimization or the adaption of the two thresholds is not considered here.
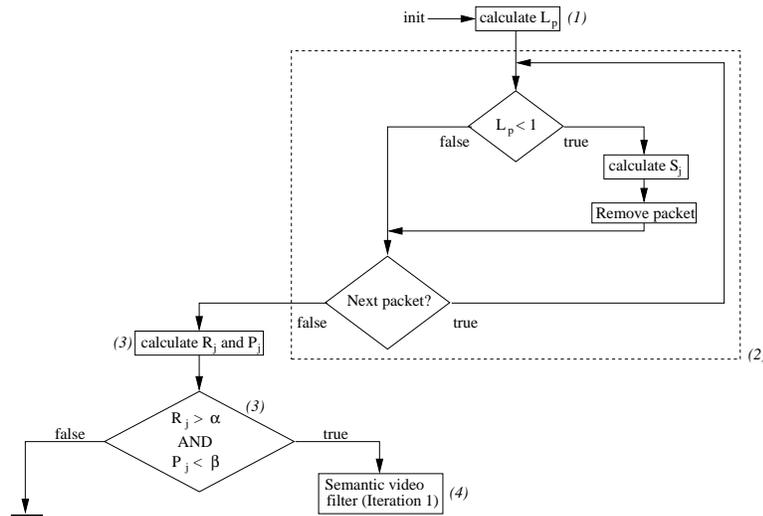
Figure 5.1: Basic algorithm of the adaptive VQM as performed independently for each queue
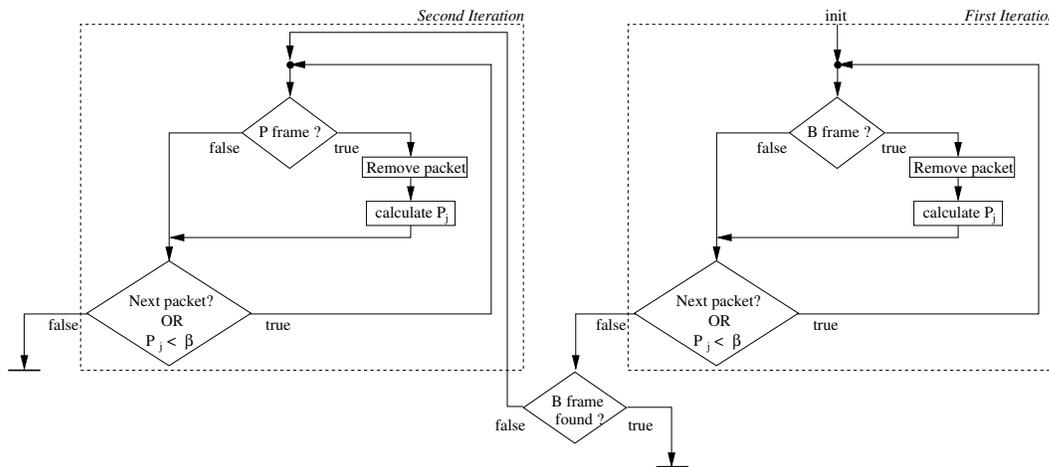
Figure 5.2: Algorithm of the semantic video filter as invoked by the system load estimator independently for each queue

Chapter 6

Performance study of the proposed queue management method

6.1 Scenario setup and simulation parameters

The main topic of this investigation is to analyze the performance of the developed VQM. Since heterogeneous traffic streams are considered, it is also interesting to compare the gain which is achieved for both stream types. A further focus is to analyze how the adaptive VQM adapts to varying traffic load. For the performance study the simulation results obtained with the following three VQM schemes:

1. Simple VQM: the FIFO approach using a non-semantic packet removal strategy

2. Semantic VQM: Adaption of the scheduling policy due to semantic features of the video stream (Section 3.3)

3. Adaptive VQM: Enables adaption to the traffic load, by performing semantic filtering of the video queues if congestion is expected (Chapter 5)

The chosen scenario is basically the same as in Section 4.3. Since the impact of various dynamic OFDM-FDMA scheduling schemes on scenarios with heterogeneous traffic load is examined in Chapter 4 we only apply full dynamic OFDM-FDMA scheme here.

We assume only heterogeneous traffic streams. This means that J_v WTs are receiving VBR video streams in parallel to the transmission of Web-pages to J_h WTs in the cell. In the first performance study (Section 6.2) J_v is varied from 1 to 36. We assume a fixed number of 12 WTs for J_h. Therewith, the traffic load from setup to setup increases due to the higher amount of simultaneous received video streams per cell. This scenario is basically the same as in Section 4.3. A different method is chosen in Section 6.3. Here J_v and J_h are changed simultaneously, while the ratio

$$J_r = \frac{J_v}{J_h} \tag{6.1}$$

is fixed. For varying J_v this leads to a fixed proportion of HTTP receiving WTs per cell. The amount of J_v and J_h rises from setup to setup until the maximum of $J_v + J_h = 48$ wireless terminals per cell is reached.

For all further modeling parameters the values as listed in Chapter 2 are employed. Again, the metrics for the transmission of Web-pages and video streams from Section 4.1 are considered.

6.2   Results for a fixed number of HTTP traffic streams

The bit-rate of the average TCP throughput achieved for all the Web-page receiving WTs in Figure 6.1 shows a clear benefit for the adaptive VQM. While with higher traffic load, due to higher amounts of video receiving WTs in the cell (J_v), the throughput rate strongly decreases for the simple and the semantic VQM the slope is much smaller with the adaptive VQM. For the case $J_v = 36$, where the highest traffic load is produced, the adaptive VQM outperforms the semantic VQM for 24% and the simple VQM for 32% for the bit-rate.
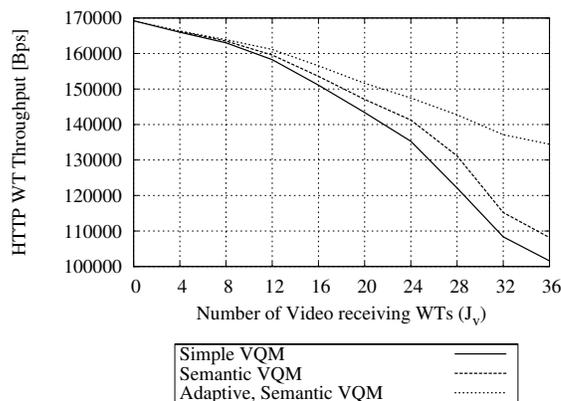
Figure 6.1: Bit-rate of HTTP receiving WTs for different video queue management methods in a heterogeneous traffic load scenario

The average transmission time per Web-page is shown for rising J_v in Figure 6.2. As for the TCP throughput the adaptive VQM outperforms the other VQM schemes. For J_v values greater than 8 the simple VQM leads to the longest page transmission times, followed by the semantic VQM. The highest gain is achieved for the adaptive VQM when the maximum traffic load is simulated as for $J_v = 36$. Compared to the semantic VQM it is 25% lower and 33% lower in comparison to the simple VQM. The plot of the standard deviation of the page transmission time (Figure 6.3) shows the similar quality as the page transmission time. Therefore, in case of higher J_v the standard deviation is the lowest with the adaptive VQM. For the maximum of J_v this leads to the gain for the predictability of 63% in comparison to the semantic and of 83% in comparison to the simple VQM.

With the plots for the HTTP metrics it is shown that the adaptive semantic VQM provides a clear benefit in cases where the Web-page receiving WTs and a high number of video receiving WTs are competing for the limited resources of the transmission system. The high queuing delay in the high load scenarios leads to falling throughput rates and rising transmission times for all the considered VQM. However, with the adaptive VQM the downgrade is much flatter than with the two non-load-adaptive VQM. This leads to the conclusion that the
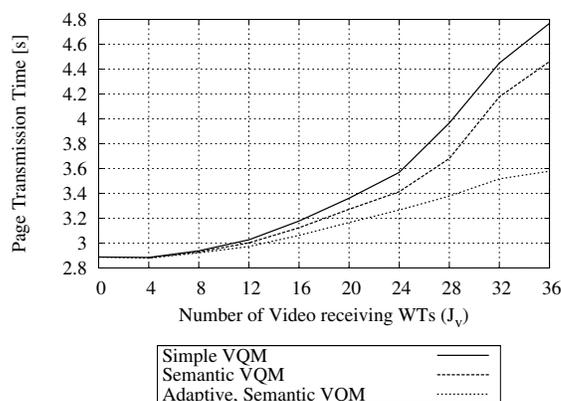
Figure 6.2: Average transmission time per Web-page for different video queue management methods in a heterogeneous traffic load scenario
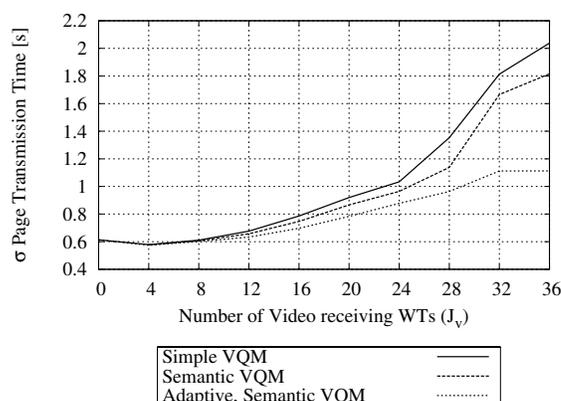
Figure 6.3: Standard deviation of the transmission time per Web-page for different video queue management methods in a heterogeneous traffic load scenario

HTTP receiving WTs clearly profit from the estimation of the traffic load and the adaptive filtering of the video queues.

It is of great interest which impact the additional video filtering performed in the adaptive VQM has to the video quality. Figure 6.4 shows the average DIV, which is a metric for the amount of distortion the transmission systems causes to the video as perceived by the viewer. Therefore, higher DIV values express higher amounts of distorted video frames. Generally for rising amounts of video receiving WTs (J_v) the DIV values rise. Compared to the simple VQM, it is interesting that for J_v smaller than 16 the DIV is slightly higher with the semantic and the adaptive semantic VQM (up to 2% for $J_v = 4$). Since for both schemes the results for J_v greater than 16 are better (lower), compared to the simple VQM, this implies that the non-load-adaptive packet removal strategy itself or the chosen weight parameters, which are common for both of the schemes, have to be optimized. The maximum quality improvement of 13%, compared to the simple VQM, is obtained for $J_v = 36$ with the adaptive VQM. However, improvements in this range do not reach the user since for DIV rates higher than

TKN-04-017
Page 36
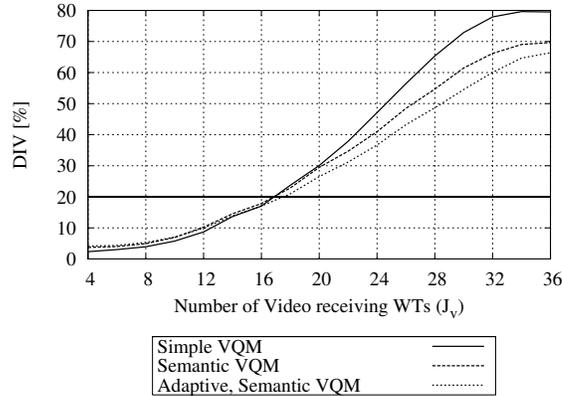
Figure 6.4: DIV for different video queue management methods in a heterogeneous traffic load scenario

20% the quality of the video becomes unacceptable [11].

The video packet loss rates for the different VQM shown in Figure 6.5 illustrate the influence of the semantic video filter as applied in the adaptive VQM. In Figure 6.5(a) the total percentage of the lost video packets, regardless to the video frame type, is shown. Although for $J_v < 26$ the simple VQM leads to the smallest loss rates compared to the other VQM schemes, more errors are caused by loosed packets to the semantically more important I- and P-frames (Figure 6.5(b) and 6.5(c)). This is especially the case with higher numbers of video receiving WTs. Here even the total loss rate of the simple VQM exceeds those of the semantic and the adaptive VQM. For example, for $J_v = 36$, 4% more I- and 6% more P-frame related packets are lost with the simple VQM than with the adaptive VQM. The benefit, which is achieved by the two VQM using a semantic approach, can be explained by comparing the loss rates for the semantically least important B-frames (Figure 6.5(d)). Since here the packet loss rate is up to 10% higher for the semantic approach, it is shown wherefrom the capacity to prioritize the transmission of I- and P-frames is taken. Considering the absolute values shows that the most lost packets are related to B-frames. The DIV gain, which is achieved by the adaption to the traffic load in comparison to the plain semantic VQM, can be explained by the packet loss rates for the scenarios for J_v greater than 24. Compared to the values obtained with the semantic VQM the percentage of the removed I- and P-frames with the adaptive VQM is lower (Figure 6.5(b) and 6.5(c)). Therefore, with high traffic load the preemptive removal of video packets leads to less errors in the semantically more important frames. Due to error propagation a single error in an I- or P-frame affects a high number of frames within the video stream. Thus, also a small benefit, as shown for the loss rates of the I- and P-frames for the adaptive VQM is important to avoid the significant reduction of the video quality.
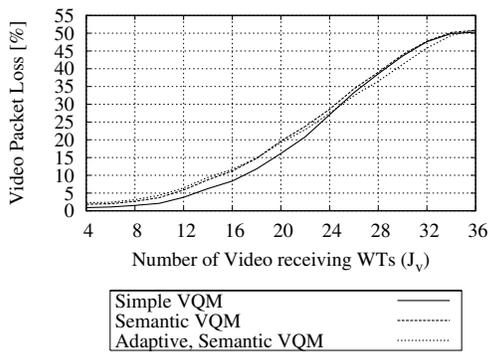
(a) Percentage of total lost video packets
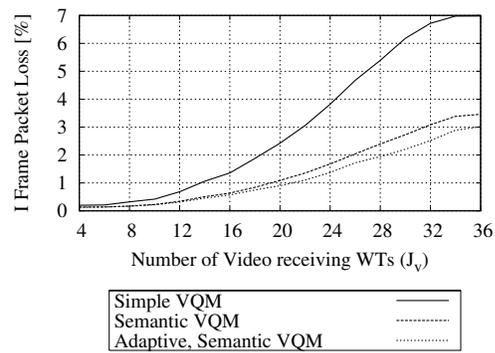
(b) Percentage of lost video packets related to I frames
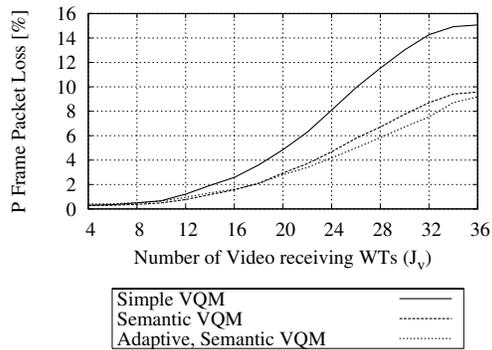
(c) Percentage of lost video packets related to P frames
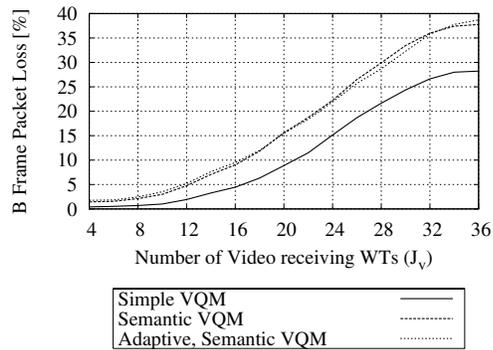
(d) Percentage of lost video packets related to B frames

Figure 6.5: Video packet loss rates for different video queue management methods

6.3   Results for varying numbers of HTTP traffic streams

In this section the combination of the adaptive VQM with dynamic OFDM-FDMA is further examined. In order to analyze the dependency of the VQM from the traffic-load scenario, three different J_r setups with varying J_h and J_v are investigated. The results are shown versus J_v, while the maximum number of J_v depends on J_r (Equation 6.1). Since various setups are investigated, the simulation results are presented in a shorter form. In place of absolute values the gain of the adaptive VQM in comparison to the simple VQM schemes is shown. Further investigation of the adaptive VQM, e.g. the comparison to the semantic VQM is presented in [31]. Although the shown gains may increase for rising J_v it has to be considered that the results for all metrics typically become worse if more video flows are "loaded" to the system. In order to investigate only the gains for acceptable video quality no results are presented for the DIV if the threshold of 20% is exceeded.

In Figure 6.6 the gain achieved by the adaptive VQM versus the simple approach is plotted for the TCP bit-rate of the HTTP receiving WTs. Considering the increase of the results for rising J_v shows clearly, how the gain of the adaptive VQM depends on the scenario, which is predetermined by J_r. The higher the amount of video streams the higher the achieved gain. For example, the comparison of the results at $J_{max} = 48$ for $J_r = 1/4$ and $J_r = 1$ leads to the conclusion that 30% more video streams in the cell result in the additional gain of 21% for the adaptive VQM on the average. For higher amounts of WTs the gain decreases for those scenarios with higher proportions of J_h, starting at $J_v = 8$ for $J_r = 1/4$ and at $J_v = 20$ for $J_r = 1$. This slope is not shown for the highest J_r considered where 80% of the WTs are receiving video streams. Here the gain at J_{max} is with 26% only slightly higher than the gain of 25% for $J_r = 1$.
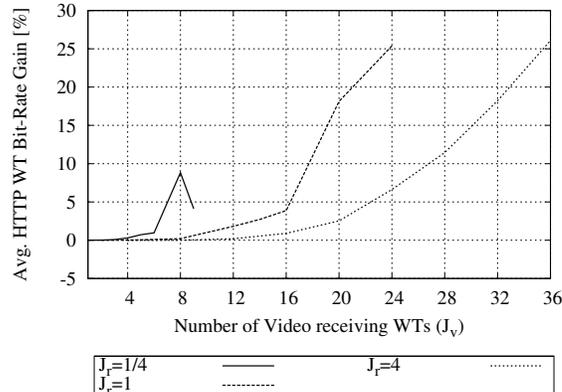
Figure 6.6: HTTP bit-rate gain achieved with adaptive VQM compared to the simple VQM for different J_r

The high gain for the adaptive VQM versus the simple approach for the TCP throughput is clearly reflected by the results for the Web-page transmission time metrics, shown in Figure 6.7 and Figure 6.8. For the mean and the standard deviation of the Web-page transmission time the gain rises for higher J_r and slopes for the two lower J_r if J_{max} is nearly reached. Although the highest gain for the mean of 27% is achieved for $J_r = 4$ the maximum
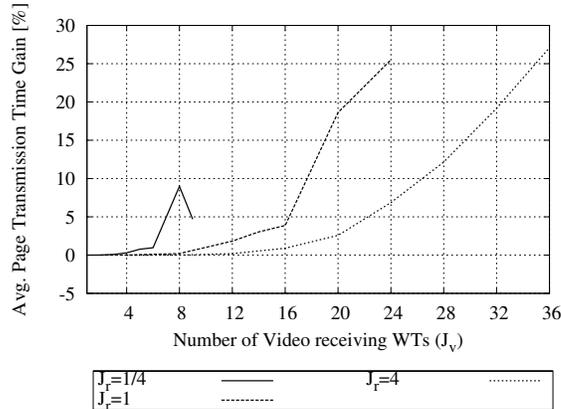
Figure 6.7: Web-page transmission time gain achieved with adaptive VQM compared to the simple VQM for different J_r
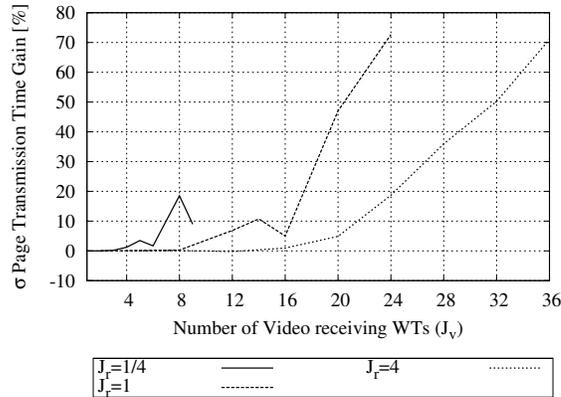
Figure 6.8: Gain for the standard deviation of the Web-page transmission time achieved with adaptive VQM compared to the simple VQM for different J_r

benefit of 73% for the standard deviation of the Web-page transmission time is reached for $J_r = 1$.

The comparison of the DIV, obtained for the adaptive and the simple VQM shows in Figure 6.9 how the combination of semantic weighting and the load-adaptive filtering influences the video quality. As with the investigation for a fixed J_h in Section 6.2, here a quality decrease for low J_v shows up. This decrease is higher for lower J_r and nearly achieves a maximum of 3% for $J_r = 1/4$. However, the absolute values for the DIV, shown exemplarily for $J_r = 1/4$ in Figure 6.10, indicate only a small quality decrease with both semantic VQM, which does not affect the DIV significantly. Since in this part of the performance study no DIV values larger than 20% are considered, basically no gains can be found. However, a very slight gain shows up for higher J_v with $J_r = 4$.

In this section the performance of the adaptive VQM in combination with dynamic OFDM-FDMA is investigated for varying amounts of video and HTTP receiving WTs in the cell. For the analyzed scenarios it is shown that the combination of adaptive VQM and dynamic OFDM-FDMA clearly outperforms the simple VQM. The highest gains are typically

TKN-04-017   Page 40
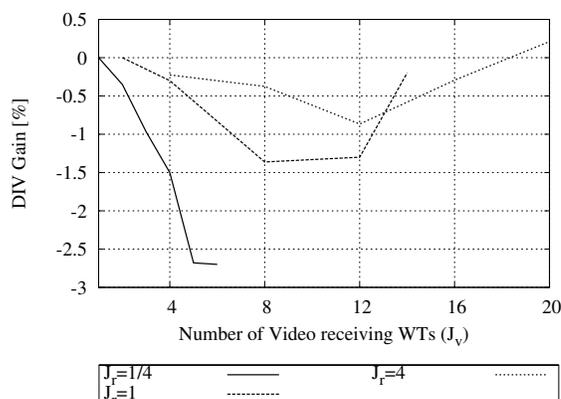
Figure 6.9: DIV gain achieved with adaptive VQM compared to the simple VQM for different J_r
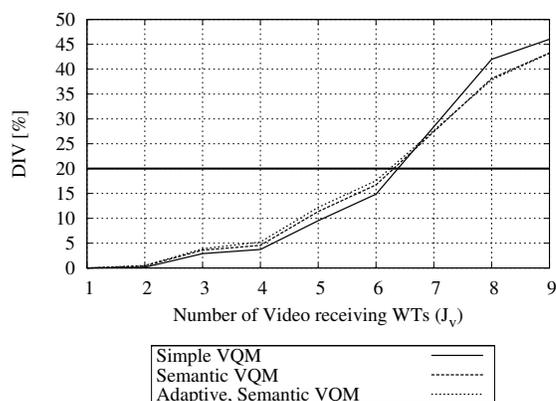
Figure 6.10: DIV for different video queue management methods and dynamic OFDM-FDMA for $J_r = 1/4$

achieved for high amounts of WTs if more than 50% of the heterogeneous traffic is streaming video. For these high-load scenarios the adaptive filtering provides an additional acceleration of the HTTP transmission. The highest benefit is obtained for the standard deviation of the Web-page transmission time, where the simple VQM is outperformed by up to 73%.

In general the benefit is higher, the higher the proportion of video data compared to the Web-traffic is. While this is more critical for $J_r < 1$ in the considered scenarios the VQM is not too sensitive. This is shown by the results obtained for the worst considered case with $J_r = 1/4$, i.e. only 20% of WTs in the cell are receiving video streams. Here the benefit which is achieved for the bit-rate by the adaptive VQM versus the simple scheme decreases for up to 21% in comparison to the scenarios with higher amounts of video traffic. However, even for low proportions of video streams in the queues benefits of up to 19% can be achieved (Figure 6.8).

The influence of the adaptive filtering to the video quality is represented by the DIV rate. As for fixed J_h in Section 6.2 a slight decrease shows up for small numbers of video streams

in the cell. This is exemplarily shown for $J_r = 1/4$ in Figure 6.10, whereby the maximal decrease of 3% results from the employment of both semantic VQM approaches. Since it occurs in a range, where the absolute DIV rate is low, it causes no perceptible influence to the video quality and is clearly justified by the benefit of the adaptive VQM for the HTTP transmission metrics.

TKN-04-017   Page 42

Chapter 7

Conclusions and future work

We studied the performance of the dynamic OFDM-FDMA scheduling schemes for a homogeneous and for a heterogeneous traffic scenario. While in the homogeneous case only the transmission of Web-pages, using the HTTP and the TCP protocol, is simulated, the simultaneous transmission of MPEG-4 coded VBR video streams is considered in the heterogeneous scenario. Three different combinations of dynamic OFDM-FDMA optimization schemes are investigated. Static OFDM-FDMA is the trivial case, where no adaption is performed. Semi-dynamic OFDM-FDMA stands for the dynamic assignment of subcarriers according to the measured channel states. The third case, dynamic OFDM-FDMA, considers the dynamic assignment and the dynamic allocation of subcarriers. The dynamic subcarrier allocation is a statistic multiplexing scheme which enables the adaption to the varying application traffic load. The simulation results show, that the dynamic OFDM-FDMA approach clearly outperforms the static and the semi-dynamic case. In the homogeneous scenario the maximal performance gain, compared to static OFDM-FDMA, is 30% for the TCP bit-rate and 140% on the average for the page transmission time. Although the absolute values are lower with the heterogeneous case dynamic OFDM-FDMA leads to gains of up to 80% for the bit-rate and 100% for the Web-page transmission time. For the video quality it shows, that with dynamic OFDM-FDMA the distortion of video frames could be lowered by up to 37%. However, this is only achievable for low numbers of WTs in the cell.

For all considered OFDM-FDMA schemes the quality of the HTTP and video metrics drops significantly for higher traffic load. Even with dynamic OFDM-FDMA the bit-rate drops for 48% in the homogeneous case if the number of WTs is raised from 1 to 48. This is obvious since the channel resources have to be distributed to more WTs. Although the metric-decrease is the lowest with dynamic OFDM-FDMA additional gains can be achieved by adapting the scheduling policy to the traffic load.

Therefore, in this technical report the adaptive VQM is proposed for heterogeneous traffic scenarios, which is applicable on top of dynamic OFDM-FMDA. It follows the cross-layer optimization approach and employs semantic filtering of video streams in case of high estimated system load. In general, the system load is considered to be high, if higher data-rates are delivered by the application servers to the access point than the wireless channel is able to transfer. Since the data-rates multiply with the number of WTs in the cell, this is especially the case with high amounts of WTs. In the performance study the adaptive VQM is compared to an extended FIFO approach, called simple VQM and to the semantic VQM

proposed in [19]. For high load scenarios it shows up, that simple VQM is outperformed by the load-adaptive filtering for up to 32% for the HTTP bit-rate and for up to 83% for the standard deviation of the Web-page transmission time. As shown in Section 6.3 the gain rises with higher proportions of video traffic in the cell. However, even with lower proportions of video on the over-all traffic load still acceptable results can be achieved.

The drastic distortion of the transmitted video streams in scenarios with high traffic load can not be prevented by the adaptive VQM. However, the additional application of the semantic filter does not cause noticeable quality decreases. Compared to the FIFO approach the largest additional distortion is 2%, which is not perceptible by the user. From the analysis of the results for the packet loss rates it follows, that with the adaptive VQM it is possible to protect the semantically important parts of the video stream. Respectively, the packet loss rate for I-frames is 4% and the loss rate for P-frames is 6% lower than with the plain FIFO approach. This leads to the conclusion, that a selective removal of video packets is more advantageous than the reactive scheduling strategy.

Beside the performance, the clear modular structure of the adaptive semantic VQM has to be considered: The strict separation of the load estimation and the semantic filter enables the adoption for other coding technologies than MPEG-4 and simplifies the development of new filter algorithms.

Due to the achieved performance increase and its generic design the adaptive VQM seems to be worth to be further investigated. The first aspect which shows up is that all the investigated optimization schemes add additional complexity to the system. It has to be analyzed whether the achieved performance gain legitimates the complexity increase. The design of the adaptive VQM already considers the trade-off between calculation complexity and performance gain. With larger execution intervals the calculation overhead can be lowered by taking drawbacks for the adaption into account. Therefore, the second task is to examine how the adaptive VQM performs in this cases. The algorithm makes use of several heuristical chosen parameters. These parameters are far from being optimal. Thus, the third task for future work is to optimize these parameters or find self-optimizing solutions.

TKN-04-017   Page 44

Appendix A

Acronyms

16-QAM 16 Quadrature Amplitude Modulation

64-QAM 64 Quadrature Amplitude Modulation

256-QAM 256 Quadrature Amplitude Modulation

ARQ Automatic Repeat Request

ADA Advanced Dynamic Algorithm

BPSK Binary Phase Shift Keying

CIF Common Intermediate Format

CNR Channel Gain-to-Noise Ratio

DIV Distortion In Interval

DLC Data Link Control

FDMA Frequency Division Multiple Access

FEC Forward Error Correction

FIFO First-In-First-Out

GOP Group of Pictures

HTTP Hypertext Transfer Protocol

HTML Hypertext Markup Language

ICI Intercarrier Interference

IEEE Institute of Electrical and Electronics Engineers, Inc.

IP Internet Protocol

ISI Intersymbol Interference

MAC Medium Access Control

MCM Multi Carrier Modulation

MPEG Moving Pictures Expert Group

MOS Mean Opinion Score

MSS Maximum Segment Size

MTU Maximum Transmission Unit

OFDM Orthogonal Frequency Division Multiplexing

PSNR Peak Signal-to-Noise Ratio

QoS Quality Of Service

QPSK Quadrature Phase Shift Keying

RTT Round-Trip-Time

SEP Symbol Error Probability

SCM Single Carrier Modulation

SNR Signal-to-Noise Ratio

TCP Transmission Control Protocol

TDMA Time Division Multiple Access

UDP User Datagram Protocol

VBR Variable Bit-Rate

VQM Video Queue Management

WTs Wireless Terminals

WWW World-Wide-Web

TKN-04-017
Page 46

Bibliography

[1] A. Aguiar and J. Gross. Wireless channel models. Technical Report TKN-03-007, Telecommunication Networks Group, April 2003.

[2] T. Berners-Lee, R. Fielding, and H. Frystyk. Hypertext transfer protocol – HTTP/1.0. RFC 1945, IETF, May 1996.

[3] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. Tripathi. Enhancing throughput over wireless LANs using channel state dependent packet scheduling. In *IEEE INFOCOM*, 1994.

[4] J. Cavers. *Mobile Channel Characteristics*. Kluwer Academic, 2000.

[5] ETSI. *BRAN; HIPERLAN Type 2; Data Link Control Layer; Part 1: Basic Data Transport Functions*, ts 101-761-1 edition, December 2001.

[6] F. Fitzek, P. Seeling, M. Reisslein, and M. Rossi et al. Investigation of the GoP structure for H.26L video streams. Technical Report acticom-02-004, acticom mobile networks, December 2002.

[7] S. Floyd and T. Henderson. The NewReno modification to TCP's fast recovery algorithm. RFC 2582, IETF, April 1999.

[8] C. Fraleigh, S. Moon, B. Lyles, and C. Cotton et al. Packet-level traffic measurements from the sprint IP backbone. *IEEE Network*, November 2003.

[9] J. Gross, H. Karl, F. Fitzek, and A. Wolisz. Comparison of heuristic and optimal subcarrier assignment algorithms. In *Proc. of the 2003 International Conference on Wireless Networks (ICWN'03)*, June 2003.

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

[11] J. Gross, J. Klaue, H. Karl, and A. Wolisz. Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming. *Computer Communications*, 2004.

[12] J. Gross, S. Valentin, H. Karl, and A. Wolisz. A study of impact of inband signaling and realistic channel knowledge for an example dynamic OFDM-FDMA system. *To appear in European Transactions on Telecommunications*, 2005.

TKN-04-017 Page 47

[13] J. Gross, S. Valentin, and R. Vu. Dynamic OFDM-FDMA systems under realistic assumptions: On the influence of channel knowledge and inband signaling. Technical Report TKN-04-015, Telecommunication Networks Group, Technische Universität Berlin, September 2004.

[14] M. Hemy, U. Hengartner, P. Steenkiste, and T. Gross. MPEG system streams in best-effort networks. In *Proc. Packet Video '99*, New York, USA, April 1999.

[15] M. Hemy, P. Steenkiste, and T. Gross. Evaluation of adaptive filtering of MPEG system streams in IP networks. In *Proc. IDME*, 2000.

[16] IEEE. *Supplement to standard for telecommunications and information exchange between systems – LAN/MAN Specific requirements – Part 11: Wireless LAN MAC and PHY specifications: High speed physical layer in the 5-GHz band*, P802.11a/D7.0 edition, July 1999.

[17] ITU. *ITU-T Recommendation P.930 – Principles of a reference impairment system for video*, August 1996.

[18] V. Jacobson and M. J. Karels. Congestion avoidance and control. In *Computer Communication Review*, volume 18, pages 314–329, August 1988.

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

[20] B. Krongold, K. Ramchandran, and D. Jones. Computationally efficient optimal power alloction algorithms for multicarrier communication systems. In *Proc. of IEEE Int. Conference on Communications (ICC)*, pages 1018–1022, 1998.

[21] T. Le-Ngoc, N. Damji, and Y. Xu. Dynamic resource allocation for multimedia services over OFDM downlink in cellular systems. In *Proc. of Spring Vehicular Technology Conference (VTC Spring)*, 2004.

[22] D. A. Menascé and V. A. F. Almeida. *Capacity Planning For Web Performance*. Prentice Hall, Inc., Upper Saddle River, NJ, USA, first edition, 1998.

[23] B. Moraru, F. Copaciu, G. Lazar, and V. Dobrota. Practical analysis of TCP implementations: Tahoe, Reno, NewReno. In *RoEduNet International Conference*, pages 125–130, June 2003.

[24] K. N. Ngan, C. W. Yap, and K. T. Tan. *Video Coding for Wireless Communication Systems*. Signal Processing and Communications. Marcel Dekker, Inc., 2070 Madison Avenue, NY, USA, 2001.

[25] J.-R. Ohm. *Bildsignalverarbeitung für Multimedia-Systeme*. TU Berlin, 1999.

[26] S. Pietrzyk and G. J. M. Janssen. Multiuser subcarrier allocation for QoS provision in the OFDMA systems. In *Proc. of IEEE Vehicular Technology Conference Fall 2002*, September 2002.

[27] T. Rappaport. *Wireless Communications Principles & Practice.* Prentice Hall, Inc., Upper Saddle River, NJ, USA, 1999.

[28] A. Reyes-Lecuona, E. González-Parada, E. Casilari, and J. C. Casasola et al. A page-oriented WWW traffic model for wireless system simulations. In *Proc. ITC 16*, June 1999.

[29] M. J. Riley and I. E. G. Richardson. *Digital Video Communications.* Artech House, 1997.

[30] W. R. Stevens. *TCP/IP Illustrated Volume 1.* Professional Computing Series. Addison-Wesley Longman, Inc., 1994.

[31] S. Valentin. Scheduling of heterogeneous data streams in the downlink of a dynamic OFDM-FDMA wireless cell. Master's thesis, Telecommunication Networks Group, Technische Universität Berlin, September 2004.

[32] A. Varga. OMNeT++ discrete event simulation system. Software is available on the Internet: `http://www.omnetpp.org` (verified September 7, 2004), June 2003. (Release 2.3).

[33] P. Vary, U. Heute, and W. Hess. *Digitale Sprachsignalverarbeitung.* B. G. Teubner, Stuttgart, Germany, 1998.

[34] K. Wehrle, J. Reber, D. Holzhausen, and V. Boehm et al. The OMNeT++ Internet protocol suite. Software is available on the Internet: `http://www.omnetpp.org` (verified September 7, 2004), December 2001.

[35] N. Yeadon, F. García, D. Hutchison, and D. Shepherd. Continuous media filters for heterogeneous internetworking. In *Proc. of SPIE Multimedia Computing and Networking (MMCN'96)*, San Jose, USA, January 1996.