Link Error Analysis and Modeling for Video Streaming Cross-Layer Design in Mobile Communication Networks

Wolfgang Karner, Olivia Nemethova, Philipp Svoboda, and Markus Rupp

Particularly in wireless communications, link errors severely affect the quality of the services due to the high error probability and the specific error characteristics (burst errors) in the radio access part of the network. In this work, we show that thorough analysis and appropriate modeling of radio-link error behavior are essential to evaluate and optimize higher layer protocols and services. They are also the basis for finding network-aware cross-layer processing algorithms which are capable of exploiting the specific properties of the link error statistics, such as predictability. This document presents the analysis of the radio link errors based on measurements in live Universal Mobile Telecommunication System (UMTS) radio access networks as well as new link error models originating from that analysis. It is shown that the knowledge of the specific link error characteristics leads to significant improvements in the quality of streamed video by applying the proposed novel network- and content-aware cross-layer scheduling algorithms. Although based on live UMTS network experience, many of the conclusions in this work are of general validity and are not limited to UMTS only.

Keywords: Link error model, error prediction, UMTS, video streaming, H.264/AVC, cross-layer.

Manuscript received Apr. 6, 2007; revised Sept. 11, 2007.

Wolfgang Karner (email: wkarner@nt.tuwien.ac.at, phone: 43 1 58801 38983), Olivia Nemethova (email: onemeth@nt.tuwien.ac.at), Philipp Svoboda (email: psvoboda@nt.tuwien.ac.at), and Markus Rupp (email: mrupp@nt.tuwien.ac.at) are with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, Vienna, Austria.

I. Introduction

Third generation mobile communication systems like the Universal Mobile Telecommunication System (UMTS) [1] already offer relatively high data rates and great flexibility, making the systems very attractive for many new services. Despite the fact that these services can be delivered to the end users via wireless communication networks, the quality of most of such services (like video streaming) heavily suffers from the error characteristics of the underlying network. In particular, the high error probability and the correlation between single errors (resulting in error bursts), caused by fading effects together with properties of the air interface, lead to significant degradation of the user's perceived quality of service [2]-[4]. Therefore, in order to provide a basis for the optimization of the quality of the services, a thorough understanding of the error characteristics in the radio access part of the mobile communication networks is necessary.

In this work, we analyze the specific error characteristics of the UMTS Terrestrial Radio Access Network (UTRAN) based on measurements in live UMTS networks of three different operators in Vienna, Austria. The investigation of the link errors in various measurement scenarios with several different mobility levels is presented.

For the evaluation of new services and protocols, as well as for the investigation of the performance of novel (cross-layer) methods, the measured traces often provide insufficient statistics; therefore, stochastic models of the observed error characteristics of the link are required. Furthermore, for efficient operation of the proposed cross-layer algorithms

themselves, the link error characteristics have to be available in the form of stochastic models (providing an analytic expression).

There have been several previous studies offering various approaches for modeling erroneous links in telecommunications [5]-[32]. Most of these models are kept as general as possible and try to minimize the computational effort (calculation time for generating an error trace). In contrast to most previous studies (entailing a high number of parameters and complex estimation methods), our models offer good usability with only a small number of parameters, which can easily be determined from measured data. Furthermore, the goal of our models is not minimum computational effort as is offered by models based on Markov chains, but to represent specific details of the measured error characteristics (namely, error predictability) at a high accuracy. In this work, we show that despite their generality, the well-known models are not capable of describing the observed specific error characteristics of the UMTS wireless link properly [33]. Therefore, a new modeling approach is presented [34], [35].

Detailed analysis of the measured link error traces reveals that the quality-based power control mechanism of the UMTS dedicated channel (DCH) induces correlation properties, and thus, memory in the link error characteristics, which in turn offers the prediction of the link errors. In order to utilize radio resources efficiently and at the same time provide the required quality of service (QoS) for the end user, a quality-based (power) control mechanism is required also in new or future mobile communication systems. Thus, despite the fact that our analysis is performed for the UMTS DCH, similar error behavior (predictability) can also be expected for novel and future mobile communication systems if they comprise a quality-based control mechanism.

In order to predict the link behavior, the transmitter has to be informed of past link errors in the forward link via a feedback of the error status of the received data. In case of the UMTS this is accomplished by the radio link control (RLC) acknowledged mode (AM). In this work we show that in spite of the feedback delay in the RLC AM, we can predict intervals of low transmission error probability [36]; furthermore, we can estimate the time variable instantaneous failure rate [33], both at a high accuracy. From the models of the UMTS DCH link error characteristics, simple estimators for future error probability can be derived.

An improvement in the quality of services like video streaming can be achieved by optimizing the service to cope with the special link error characteristics, by adapting the characteristics of the wireless link to meet the requirements of the service, or by installing a cross-layer design to optimally connect the characteristics of the link with the service properties. All the approaches highly depend on consolidated knowledge about the specific error characteristics of the considered link. In this work, we are focusing on cross-layer methods for improving H.264/AVC (Advanced Video Coding) [37] video quality over UMTS DCH.

In [38], we proposed a cross-layer error detection method for H.264/AVC video streaming. The performance evaluation of the proposed design showed that the method is capable of reaching great improvements in video quality, despite having the special burst error characteristics of the UMTS DCH.

Of course, the knowledge of the link error characteristics as well as the special link error models are not only required for performance evaluation of new (cross-layer) methods. The very specific error properties of the link can also be explicitly exploited to improve the end-to-end service quality. In this work, we show that the predictability of the link errors can be used for network-aware cross-layer processing [39] in order to provide substantial gain in video quality with respect to the required data rate. The proposed method transmits additional redundancy information for decoding of the streamed video only in case of expected high link-error probability.

Cross-layer processing algorithms achieve greater improvements in QoS when service characteristics are considered in addition to the awareness of network characteristics. We present a network and data-priority aware cross-layer scheduling algorithm [36] which provides significant quality improvements in the decoded video by rescheduling the transmission of IP packets containing streamed coded video data according to their priority level. High priority packets (containing parts of I video frames) are transmitted in time intervals with lower predicted error probability.

Considering more priority levels for smaller data packets, namely, radio link control protocol data units (RLC PDUs), and setting the priority according to the distortion the loss of the packet would cause, results in even higher video quality. We developed a network-aware, distortion-based scheduling algorithm [40] which is video-content sensitive and capable of exploiting the predictability of the instantaneous link failure rate.

The remainder of this paper is organized as follows. In section II, the protocol stack for streaming video data over the packet switched (PS) domain of UMTS is introduced. Section III presents the measurement-based analysis of the link error characteristics, also containing the measurement setup. Models capable of describing these special link characteristics are presented, and the specific properties of the errors in the UMTS radio link – like predictability – are analyzed. In section IV, novel cross-layer processing methods are described which rely on network awareness and exploit the memory in the link (predictability of link errors). To evaluate the proposed cross-

layer algorithms, the results regarding improvement of video quality are presented. Section V provides a summary and conclusions.

II. Protocol Architecture for Video over UMTS

After segmenting the video frames into smaller parts (slices) and encoding them, the following packetization procedure must be performed.

As shown in the schematic illustration in Fig. 1, the encoded video slices or equivalently the network abstraction layer units (NALUs) containing the slice data are encapsulated into real time protocol (RTP) packets with a header of 12 bytes. Each RTP packet then is transmitted within the UDP protocol which adds a header of 8 bytes. If no segmentation is needed, the UDP packets are further packed within IP packets with a header size of 20 bytes for IPv4 and 40 bytes for IPv6. These IP packets are then transmitted over the UTRAN [41], where at the RLC [42] layer they are segmented into RLC PDUs. In the case of a bearer with a data rate $\leq$ 384 kbps, a 320 bit (40 bytes) payload within the RLC packets is usually used [43].

For packet switched bearers the RLC of the UTRAN can work in AM offering RLC packet retransmissions or in unacknowledged mode (UM) enabling only error detection but no feedback. The header size for the RLC AM packet is 16 bits, whereas the RLC UM packet header contains 8 bits. After the mapping of the packets onto the transport channel performed by the medium access control (MAC) layer [44], the RLC payload in connection with the RLC and MAC headers becomes a transport block (TB) [45]. Then, after attaching cyclic redundancy check (CRC) bits to the TBs in the physical layer, these are segmented/concatenated into code blocks and the bit stream is encoded by a channel code [46], [47]. For packet oriented applications, turbo coding is usually used with a coding rate of 1/3, which can further be punctured to match the rate with the physical resources.

Before mapping the bit stream onto the physical channels for transmission, a first interleaving over one transmission timing interval (TTI), radio frame segmentation, transport channel multiplexing, physical channel segmentation, and a second interleaving (over one radio frame) is performed.

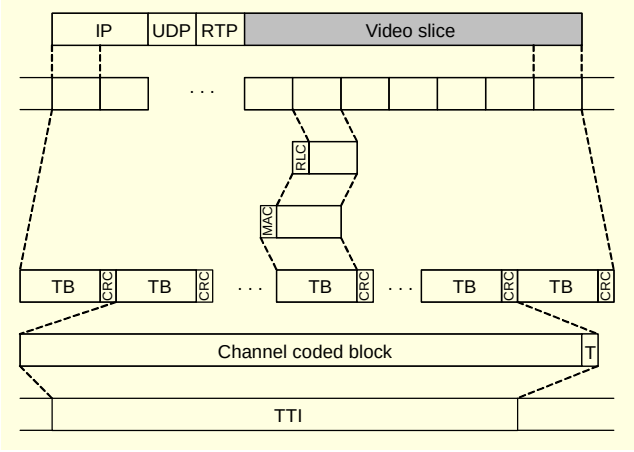

Fig. 1. Schematic illustration of a packetization example for the transmission of a video slice over UMTS.

III. Measurement-Based Analysis and Modeling of Link Layer Error Characteristics

The link error analysis and the resulting modeling were based on measured link layer error traces. The measurements have been realized in live UMTS networks of three different operators in the city center of Vienna, Austria. Due to the similarity of the measured results from the three networks [35], the measured error traces from only one operator's live network are considered in the following presentation.

1. Measurement Setup

For the measurements, a UDP data stream with bit rates of up to 360 kbps (372 kbps including UDP/IP overhead) in DL was sent from a PC over the UMTS network to a notebook using a UMTS terminal as a modem via a USB connection. In Fig. 2, a schematic illustration of the setup for the measurements in the live networks is given. Additionally, we performed measurements in a reference network (a separate network for acceptance testing). For the Uu interface (the wireless link in UTMS) in this case, we used a cable connection between the node-B antenna connector and a shielding box where the mobile was enclosed to avoid interference, multipath propagation, and fading effects.

In order to be capable of tracing the internal measurements of the mobiles, various WCDMA TEMS mobiles[1)] logging, were used as terminals in connection with TEMS Investigation software by Ericsson [48]. In this paper, we refer to the mobiles

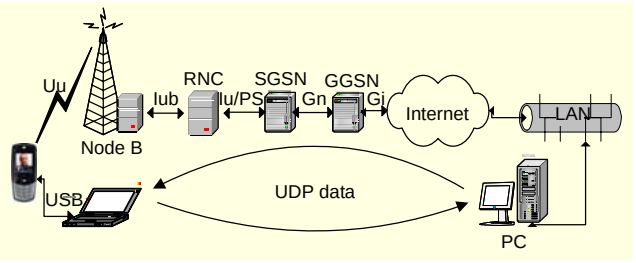

Fig. 2. Scheme of the measurement setup.

1) Modified UMTS mobiles for TEMS data logging, offered by Ericsson. The mobiles are modified in a way that they provide internal measurements via the interface to the notebook.

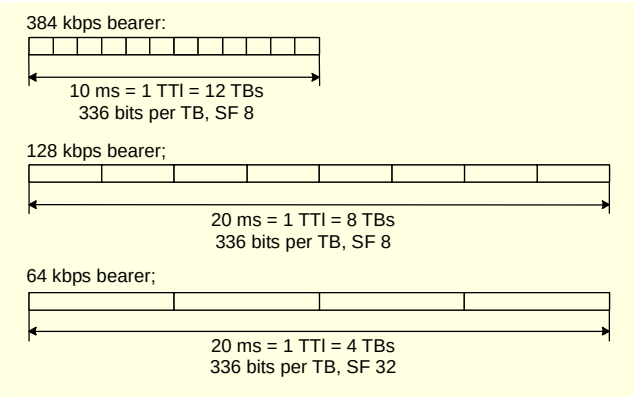

Fig. 3. Illustration of bearer parameters.

used as mobile 1 (Motorola A835), mobile 2 (Motorola E1000), mobile 3 (Sony Ericsson Z1010), and mobile 4 (Sony Ericsson V800).

After parsing the export files of the TEMS Investigation software tool, various parameters of the UMTS system can be analyzed. One of these is the CRC information of the received TBs which is used for the analysis of the UMTS link error characteristics.

In the considered UMTS networks, turbo coding with a coding rate of 1/3 was used for the DCH in DL with a TB size of 336 bits, consisting of 320 bits for the RLC payload and 16 bits for the RLC AM header. Although the RLC AM with its error detection and feedback mechanism was used, the link error analysis was performed without consideration of retransmissions, thus, offering independence in selecting RLC AM/UM and the RLC AM parameter adjustment (discard timer, maximum number of retransmissions). The settings of the spreading factor (SF), TTI, and the number of TBs jointly coded and transmitted per TTI are shown in a schematic illustration in Fig. 3 for the three bearers which were used in our measurements.

Another very important parameter of the UTRAN for evaluating the error characteristics of the DCH is the block error ratio (BLER) quality target value for the outer loop power control (OLPC) which was set to 1% in the networks used for the measurements. As a consequence, the OLPC tries to adjust the signal to interference ratio (SIR) target for the inner loop power control (ILPC) in a way that the required BLER quality (1% in our case) is satisfied. We will show in section III.2 that this quality target is missed significantly in all the scenarios.

For the analysis of the UMTS DCH link error characteristics, we have considered several scenarios with different mobility characteristics, which we will refer to as "static," "small-scale movements," "walking indoors," "tramway," "car-city," "car-highway" and "reference." The measurements for the static case were taken in an office in the city center of Vienna, Austria, with the UMTS terminal located on a table in a typical office environment. Due to few movements of persons or other objects around the mobile station, there were only small variations in the channel. The measurement of small-scale movements was performed by a person sitting at the table and randomly tilting and moving the UMTS mobile with his hands. In the walking indoors scenario, the measurement results were obtained from a person walking around inside the building. The rest of the scenarios are outdoor scenarios with the measurements taken in a tram going around the city center of Vienna (tramway) and going by car either on a street in Vienna (car-city) with moderate speed up to 50 kmph or on a highway with higher speeds up to 100 kmph (car-highway). For the reference scenario, the mobile was enclosed within a shielding box to avoid interference, multipath propagation and fading effects, with the direct signal from the node-B antenna connector fed into the box via a planar antenna coupler inside the box.

In order to get sufficient statistics of the measured link error characteristics, we recorded several data traces with a length of about one hour each in all of the mentioned scenarios.

2. Link Error Analysis

When data is transmitted over wireless mobile communication systems, most of the errors originate from the radio link, which also becomes the main or only error source in the transmission chain as long as the frequency spectrum is a scarce resource and there is interference limitation in the system. Since the highest error probability is found in the radio link, we focus our error analysis primarily on the wireless access part of the UMTS network.

As already mentioned in section II, the UMTS standard specifies an error check in the radio link in the form of a CRC per TB in the physical layer [45]. This information is perfectly suitable to use in our link error analysis, as all the important error sources within the radio link (fading, interference, multipath propagation, power control, as well as channel coding and interleaving effects) are included in the resulting

Table 1. Probability of TB and TTI errors, 384 kbps bearer.

Scenario	$\overline{P}_e$(TB)	$\overline{P}_e$(TTI)
Static	2.66×10^{-3}	4.72×10^{-3}
Small-scale movements	2.22×10^{-2}	2.34×10^{-2}
Walking indoors	1.98×10^{-2}	2.44×10^{-2}
Tramway	1.44×10^{-2}	1.70×10^{-2}
Car-city	2.06×10^{-2}	2.48×10^{-2}
Car-highway	2.34×10^{-2}	2.63×10^{-2}

link error statistics.

The first parameter considered for the analysis of the link error characteristics is the total mean TB error probability $\overline{P}_e(\mathrm{TB})$ (over the complete measured trace) presented in Table 1 for the different scenarios, measured with 'mobile 1' and a 384 kbps PS bearer. Similar results were obtained with the other mobiles and bearers.

In all the scenarios with mobility, the TB error probability is around 2×10^{-2}, whereas in the static case, the error probability is smaller by one order of magnitude. Due to these results, we can conclude that the transmit power control (TPC) algorithm of the UMTS is not capable of adjusting the required BLER target value (1%); moreover, we can see that there are roughly only two different link error characteristics in terms of total link error probability: static and mobile (regardless of which kind of mobility). An interesting and new conclusion, contradicting common opinion among researchers, is that the small-scale movements taken from a person just sitting and moving the mobile with his hands result in the same error probability as the results taken when traveling by car. We assume that this is due to the spatial small-scale fading in the indoor scenario caused by multipath propagation. Together with the small-scale movements (changing the position of the mobile station by hand) the mobile is affected by fading in time. Another reason may be the antenna characteristics of the mobile station in connection with the small-scale movements (tilting and turning the mobile station) again resulting in a time variable channel. The resulting fading of received code power (or equally the changing of SIR) in time, together with the non-optimality of the power control mechanism leads to high error probability in the small-scale movement scenario.

When comparing $\overline{P}_e(\mathrm{TB})$ in Table 1 with the total TTI error probability $\overline{P}_e(\mathrm{TTI})$ (from the same measurements) we note that the results are almost equal, especially in the scenarios with mobility. With 12 TBs per TTI in case of the 384 kbps bearer and the assumption of a memoryless binary symmetric channel (BSC), one would expect a much higher total TTI error probability (such as 15.9% in the tramway scenario). Therefore, we can conclude that there is a high correlation between the error states of the received TBs within one TTI leading to bursty TB error behavior of the UMTS DCH.

The analysis of the number of erroneous TBs within erroneously received TTIs presented in Fig. 4 for the 384 kbps bearer measured with mobile 1 leads to the same conclusions – especially for the scenarios with mobility where the probability of having all (twelve) TBs within one erroneously received TTI is between 0.8 and 0.9. On the other hand, the probability of having only one out of twelve TBs received erroneously within one TTI increases the less movement disturbs the transmission. We observe probabilities of up to 0.7 with the mobile in the shielding box of the reference network where there is no interference, no multipath propagation, and no fading in the propagation environment.

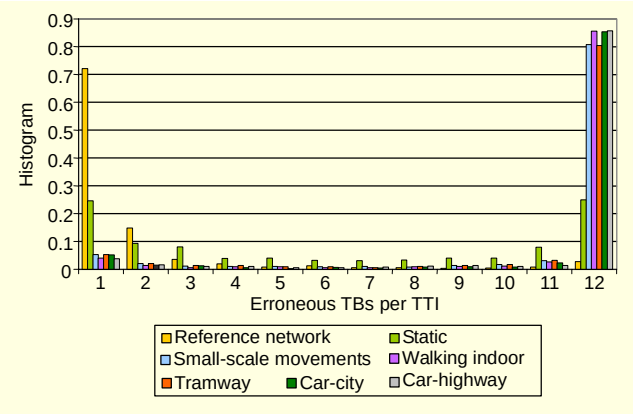

Fig. 4. Number of erroneous TBs within erroneously received TTIs (384 kbps bearer, mobile 1).

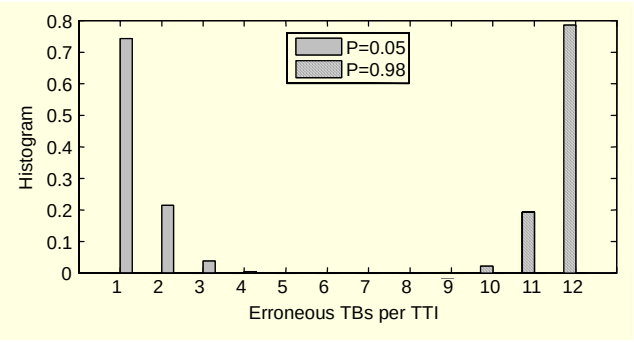

Fig. 5. Theoretic analysis of the number of erroneous TBs within one TTI: binomial distributions with 0.05 and 0.98 TB error probability.

This leads to the conclusion that the small variations in transmit power caused by the UMTS DCH TPC lead to a very small TB error probability within one TTI (meaning within one jointly turbo-coded and interleaved data block), whereas the deep fades of the scenarios with movement heavily influence the channel and thus lead to a high TB error probability within an erroneous TTI. This conclusion is verified by using low (0.05) and high (0.98) error probability for Bernoulli experiments, resulting in comparable distributions (binomial) of the number of erroneous TBs per TTI (Fig. 5) for the reference network and the case with movement, respectively. The static case can be regarded as a mixture of both cases. These results can be used for refinement of the link error models, especially in dynamic scenarios as presented in section III.4.

For further analysis of the correlation properties of the link errors outside one TTI (one jointly turbo-coded and interleaved data block) we gathered the statistics of the number of subsequent error-free TTIs, what we call the *TTI-gaplength,*

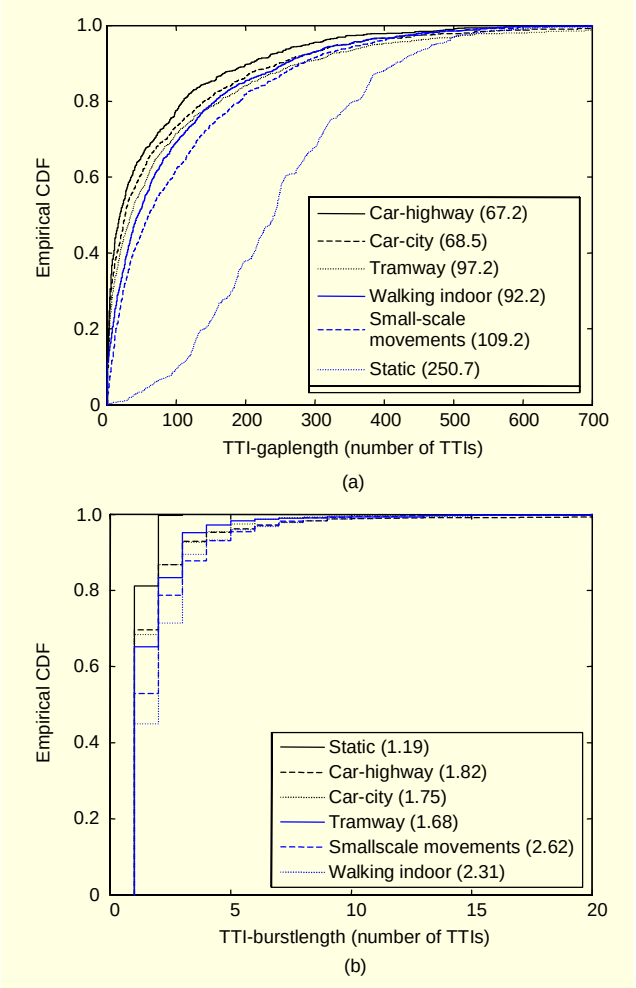

Fig. 6. Statistics of TTI-gaplengths and TTI-burstlengths (mean values in the legends) with a 384 kbps bearer using mobile 1.

while the number of subsequent erroneous TTIs is called *TTI-burstlength*. According to [49], an error burst is defined as a group of successive units (bits, packets, or TBs, and TTIs in our case) in which two successive erroneous units are always separated by less than a given number L_c of correct units. For our work, L_c is equal to zero for bursts.

From the presentation of the TTI gaplengths and TTI-burstlengths in Fig. 6, we observe that the sequences of error-free TTIs are up to 700 and there are up to ten TTIs subsequently erroneously received. Again, we recognize that the statistics for the scenarios with movement are similar while the static scenario shows a considerably different distribution of the TTI-gaplengths and much shorter TTI-burstlengths of up to only two. The reason is that in the scenarios with movement, the influence of fading-effects dominates the link error characteristics, while in the static scenario, the impact of the OLPC algorithm becomes significant.

Due to the high probability of having all TBs within one TTI

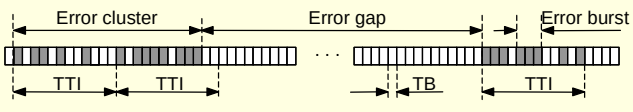

Fig. 7. Schematic illustration of TB error bursts, gaps, and clusters.

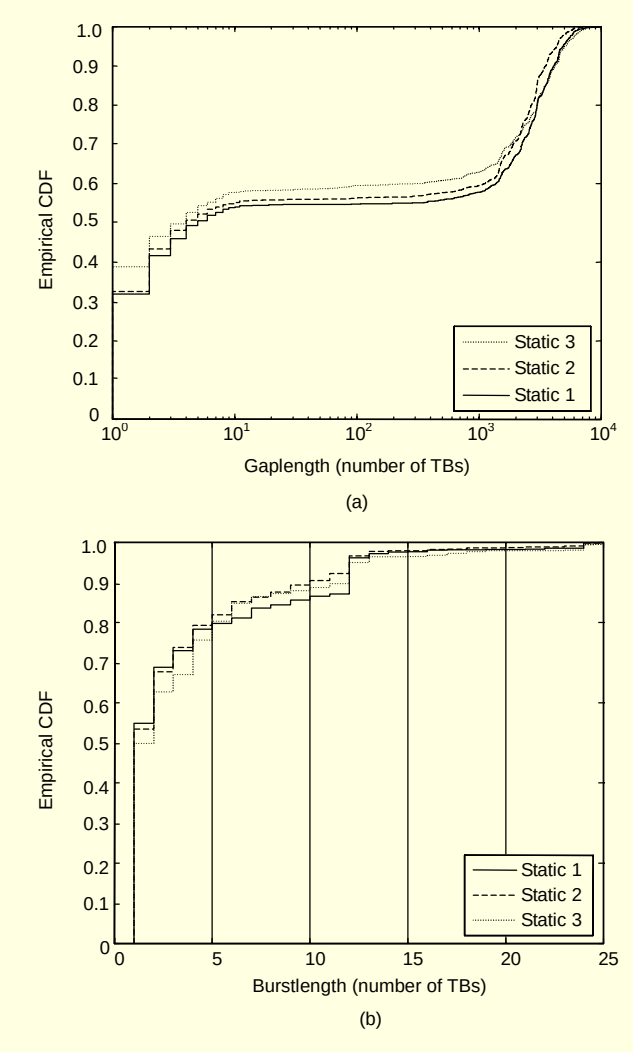

Fig. 8. Statistics of TB gaplengths and burstlengths with a 384 kbps bearer for the static scenario using mobile 1.

erroneously received for the scenarios with movement (as shown in Fig. 4) the gap and burst analysis in terms of TTIs would be sufficient, whereas in the static case the granularity of the analysis has to be refined to TB level in order to catch correlation properties of TB errors within one TTI.

From now on, we define the number of error-free subsequently received TBs as error gap with the according *gaplength* and the *burstlength* as the number of TBs subsequently received with error–again, according to [49], with L_c equal to zero. Additionally, we define an error cluster as a group of erroneously received TBs, if, for L_c equal to the

number of TBs per TTI, the TB error bursts are separated by $\leq L_c$ error-free TBs. In Fig. 7, we present a schematic illustration of the TB error bursts, gaps, and error clusters.

Figure 8 shows the results of the analysis of the gaplengths and burstlengths (in number of TBs) for the static scenario measured with a 384 kbps bearer and mobile 1 at three different locations on different dates and at different times of day in order to have different propagation, interference, and network load situations [34]. We observe that there is only a little difference between the statistics.

Furthermore, in measurements with additional DCH and high-speed downlink packet access (HSDPA) users introduced to create intra- and inter-cell interference, we found out that additional interference in the DL does not change the measured statistics as long as the UMTS call admission control is working properly to keep the required link power below the power limit.

We can see in Fig. 8 that there are short gaps of $\leq$ 12 TBs, which are the gaps within one or two successive TTIs. The statistics also include long gaps from 200 to 7,000 TBs. The burstlengths are between one and 25, with two higher steps at 12 and 24 TBs caused by the probability of having all TBs erroneously received within one TTI (there are 12 TBs within one TTI in case of the 384 kbps bearer) which is still not negligible even in the static case.

Of course, the turbo coding and the interleaver algorithms of the UMTS radio access affect the link error characteristics, especially within one radio frame, TTI, or coding block. However, the measurement-based analysis of the positions of the erroneous TBs within one TTI has shown that all possible TB locations exhibit almost the same error probability (see Fig. 9). Similar results are observed from the analysis of the output of turbo code simulations in [50].

The uniform distribution of the positions of the erroneous TBs within one TTI, together with the high probability of not all the TBs within one TTI being received with error, results in short error bursts (separated by short gaps) within one and even

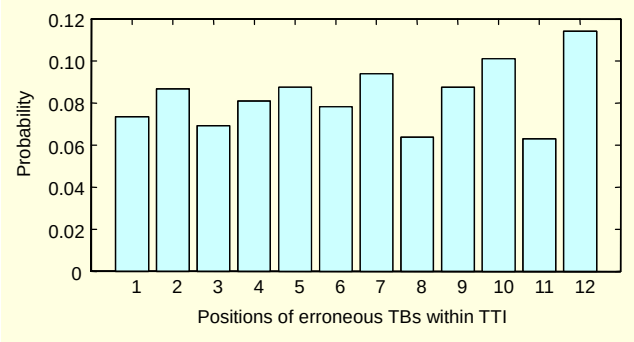

Fig. 9. Probabilities of TB error positions for static scenario using mobile 1 with a 384 kbps bearer.

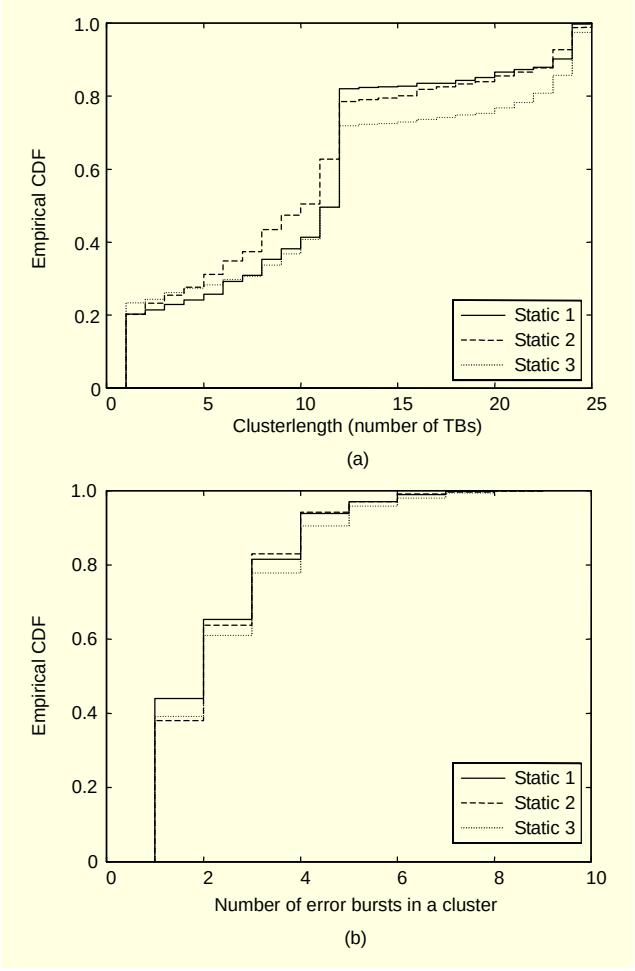

Fig. 10. Statistics of TB clusterlengths and number of error bursts per cluster with a 384 kbps bearer for static scenario using mobile 1.

two successive TTIs. Therefore, in order to analyze the error properties within one or more successive TTIs (meaning the characteristics of and within an error cluster), it is also interesting to present the statistics of the clusterlength and the number of error bursts in a cluster as shown in Fig. 10 for $L_c = 12$. There, we show three static measurements which have mean clusterlengths of 10.84, 11.19, and 14.97; mean numbers of erroneous TBs per error cluster of 7.97, 7.74, and 9.81; and mean numbers of bursts per cluster of 2.19, 2.23, and 2.38, respectively. We observe that the clusterlength is $\leq$ 25 in most cases, and we do have one to eight error bursts within one such error cluster. Since the influence of the quality-based power control algorithm on the link error characteristics becomes significant in the static scenarios, the error clusters are separated by long error gaps.

In order to use radio resources in cellular wireless communication systems efficiently, the minimum transmit power for each link has to be found while meeting the quality

target of the link. In the UMTS this task is performed by the OLPC algorithm [51], a slow (100 Hz) closed-loop control mechanism which sets the target SIR for the fast ILPC in order to converge to a required link quality given by the network (1% in our work). The ILPC then regulates the link power level with a rate of 1,500 Hz to meet the target SIR.

Within the 3rd Generation Partnership Project (3GPP) the OLPC is not entirely specified. A common way to implement the control procedure is to use CRC information feedback of the transmitted TBs together with a sawtooth algorithm as described in [1] and [52]. If a TB is received in error, the target SIR value is increased by $K \cdot \Delta$, whereas in error-free transmission, it is reduced by Δ, where Δ is the step size in dB. If K is chosen to be $1/(\mathrm{BLER}_{\mathrm{target}} - 1)$, the algorithm tries to keep the BLER always less than or equal to $\mathrm{BLER}_{\mathrm{target}}=1/(K+1)$. Thus, by selecting a proper value for K, a quality level for the wireless link can be set.

Figure 11 shows a schematic illustration of this OLPC algorithm for the case when an error occurs in the link with a probability of one if the $\mathrm{SIR}_{\mathrm{target}}$ and the SIR at the receiver fall below a certain threshold. Obviously, in this special case, there is an erroneous TB after K error-free TBs.

As we have seen, in the UMTS DCH, error clusters are generated due to joint interleaving and coding over one TTI

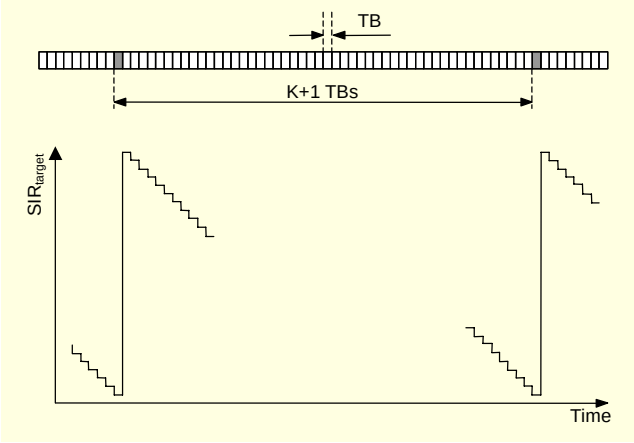

Fig. 11. Quality-based power control (ideal).

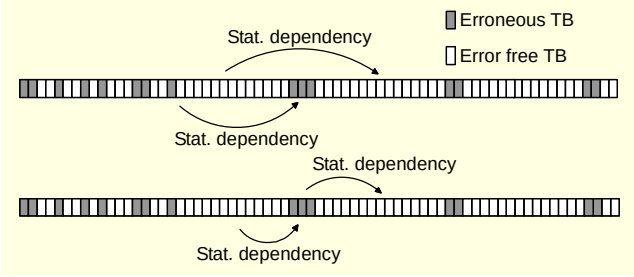

Fig. 12. Schematic illustration of correlation between error gaps and bursts.

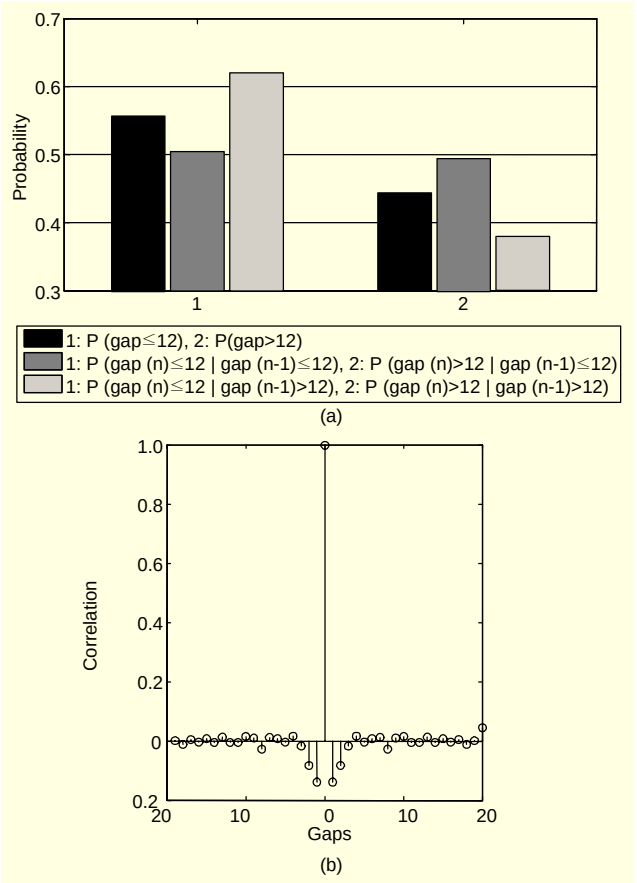

Fig. 13. (a) Statistical dependency between successive gaps and (b) autocorrelation of gaplengths. 384 kbps bearer, static 2 scenario, mobile 1.

and also due to the stochastic nature of the channel. The power control algorithm, which still tries to meet the quality target for the link, increases the target SIR by $K \cdot \Delta$ times the number of errors in the cluster. This leads to a multiple of K error-free TBs (long gap) between two error bursts.

Having several error bursts separated by short gaps within an error cluster and the occurrence of error clusters following long gaps means a correlation between successive TBs and statistical dependencies between neighboring gaps and bursts as indicated in the schematic illustration in Fig. 12.

The validation of this conclusion is given in Fig. 13(a), where we see that the probability of having short ($\leq$ 12 TBs) and long ($>$ 12 TBs) gaplengths does depend on previous gaplengths. Similarly, regarding the autocorrelation function of the sequence of gaplengths in Fig. 13(b), the memory in the channel outreaches two successive gaps. Similar results can be shown for bursts and for the dependencies between gaps and bursts.

For many applications and for the analysis of codes, another important measure of the link error characteristics is the statistic of the error weight probability $P(m,n)$, that is, the probability of

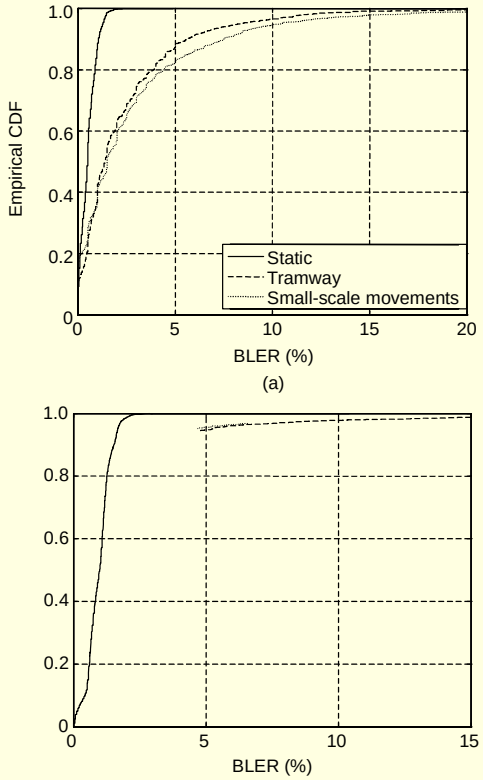

Fig. 14. Comparison of DL BLER (%) for the 384 kbps measured with mobile 2 (a) and mobile 3 (b) and for the 128 kbps (c) and 64 kbps (d) bearers measured with mobile 2.

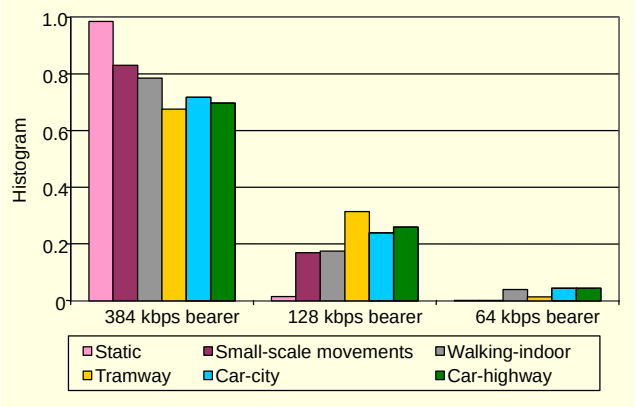

Fig. 15. Bearer usage with dynamic bearer type switching using mobile 1.

384 kbps bearer is around 75%, whereas in the static case it is almost 100%. Of course, the bearer type switching depends on the network coverage situation and on the quality of the receiver in the mobile. In [53], we presented a detailed analysis of dynamic bearer type switching based on measurements in live UMTS networks.

4. Modeling of Link Layer Error Characteristics

To provide sufficient statistics, simulations with statistical link models are often required instead of directly performing measurements or using simulations with measured traces.

A. Link Error Models – A Literature Survey

Beginning in the 1960s with the work of Gilbert [5], researchers understood that a memoryless BSC is not capable of describing the bursty nature of communication channels. Gilbert's model is a two-state Markov model with one "good" and one "bad" state, where no errors occur in the good state, but in the bad state, the error probability takes some value > 0. Due to the straightforwardness of the parameter estimation, a simplified version of Gilbert's model has often been used in subsequent studies, where the error probability in the bad state is one. Then, the model changes from a hidden Markov model (HMM) to a two-state Markov chain. Thus, the two parameters for this simplified Gilbert model can be calculated directly from the measured error trace by using the mean error gaplength and the mean burstlength or the mean error probability.

Enhancements to Gilbert's model can be found in the work of Elliott [6], where errors can occur also in the good state. This model, also known as Gilbert-Elliott channel (GEC), overcomes the limitation of Gilbert's model in having geometric distributions of burstlengths only. Beside being an HMM, it is also non-renewal, which means, for example that the current burstlength is not statistically independent of the previous burstlength. This, of course, brings opportunities for channel modeling but also complicates parameter estimation. For Gilbert's model, parameter estimation is possible by using trigram statistics or curve fitting, whereas the parameters for non-renewal HMMs like the GEC have to be estimated, as for example using the Baum-Welsh algorithm [7].

Also, in the 1960s, it was proposed [8]-[11] to use renewal processes to model the error characteristics of communication links with the suggestion in [8] to use independent Pareto ($f(t \mid a) = a \cdot t^{a},\ t \geq 1,\ 0 < a < 1$) distributions for the intervals between successive errors.

Further enhancements to Gilbert's model were published by Fritchman [12], proposing partitioned Markov chains with several error-free and error states. The well-known Fritchman model is a simplified version with one error state only. Being a renewal model, the error-free run (gaplength) distribution uniquely specifies the model; thus, the model parameters can be found via curve-fitting. Each state of the Fritchman model represents a memoryless BSC; therefore, Fritchman's model is limited to polygeometric distributions of the gaplengths. This led to the proposal of slowly spreading Markov chains for error modeling in [13].

These previous models (see [14] for a survey) have attracted a great deal of attention from researchers who have used them or tried to improve them. In [15] and [16], methods for the parameter estimation of Gilbert's and Fritchman's models are presented, while in [17], [18], [19] and [20] HMMs in general are addressed. New models to describe special error characteristics of (wireless) communication links have been developed up to the present, including a three-state model for Poisson distributed burstlengths [21], a two-state model with segmented exponential distributed burst- and gaplengths [22], two- and three-state Markov models [23] and a finite state Markov channel (FSMC) [24]. Other recently published bipartite models [25] are similar to Fritchman's model but they partition the burst- and gaplength statistics in subintervals. A bit error model for IEEE 802.11b uses chaotic maps [26]. In [27] a two-state process is used, in which the gaps and bursts are generated from measured distributions. A run length model (RLM) is presented in [28], generating the error trace with a two-state renewal process with a mixture of geometric distributions which then is shown to be equal to a four-state Markov model. A three-state RLM is presented in [29]. Another RLM which uses half-normal distributions for burst- and gaplengths is shown in [30] for modeling the packet errors in digital video broadcasting – handheld (DVB-H).

B. Link Error Models for the UMTS DCH

Most of the new models have been proposed to describe special properties of the links instead of taking general HMMs.

In contrast to general models which need a high number of parameters and complex parameter estimation methods, the new models are required to offer good usability, meaning a small number of parameters which can easily be determined and adapted to changing system configurations. Furthermore, the goal in designing a new model is to represent the specific details of the measured error characteristics with high accuracy, whereas a small computational effort to generate an error trace has become less significant.

Often, as in our case, the specific link error properties seen in the measurement-based analysis already lead to a particular approach for designing a link error model.

Of course, we consider the well known modeling approaches in our work, but we demonstrate that these models are not appropriate for describing the error statistics of the UMTS DCH.

As the link error characteristics of the UMTS DCH may change when adjusting different parameters in the UTRAN (like different TF, TTI-length, SF, and so on) our goal is to present a universal UMTS DCH modeling approach resulting from the specific link error characteristics of the UMTS DCH seen from the analysis in section III.2. We particularly focus on the characteristics of the 384 kbps UMTS DCH PS bearer in DL which currently represents the DCH PS bearer with the highest throughput available in the considered networks.

When modeling the UMTS DCH link error characteristics, we have to consider the two cases static and movement only, as already concluded. Because the probability of having all TBs erroneously received within one TTI (given this TTI is erroneous) in the scenarios with movement is very high (≈ 0.9), we suggest modeling the link error characteristics in the movement case based on TTI granularity.

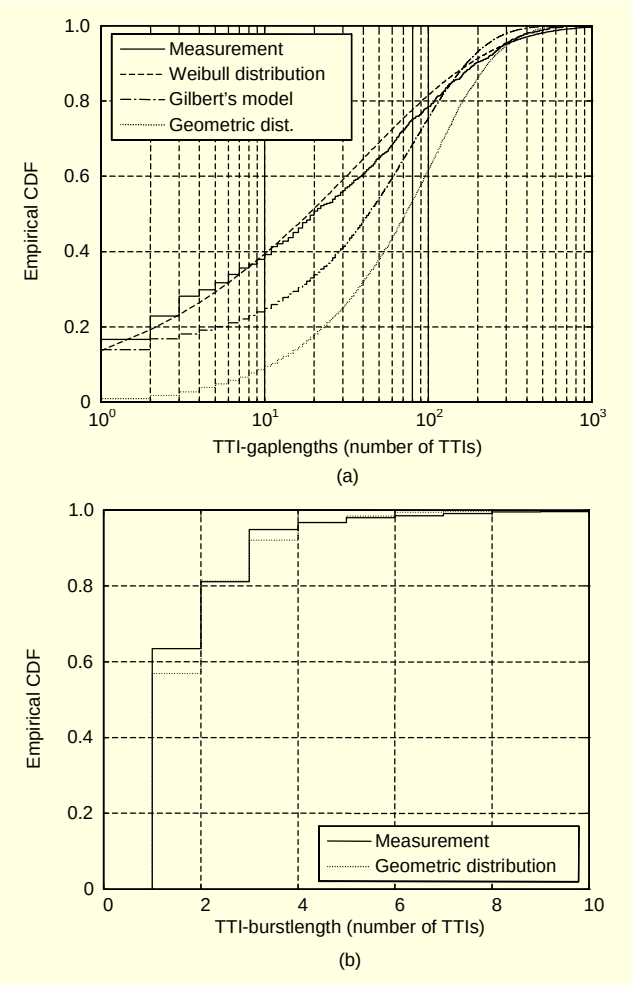

Fig. 16. Comparison of empirical CDFs of TTI-gaplengths and TTI-burstlengths. Measurements vs. two-state Markov chain, Gilbert's model and Weibull distribution. 384 kbps bearer, car-city scenario, mobile 1.

C. UMTS DCH Link Error Modeling – Movement Case

We have shown the existence of error bursts (up to ten TTIs long); therefore, the simplest way of modeling the link errors via a memoryless BSC is not possible. One slightly more sophisticated modeling approach is to extend the single-state model of the memoryless BSC (Bernoulli trials) to a two-state Markov chain, where in one state, error-free TTIs are generated, and in the other state, erroneous TTIs are generated, as in the simplified Gilbert model [5] with an error probability of one in the bad state. This two-state Markov chain is uniquely specified by two parameters, namely, the packet loss rate and the average TTI-burstlength (or the average TTI-gaplength), which can be directly determined from the measured error trace as the model is not an HMM.

As shown in Fig. 16, the geometric distribution of TTI-burstlengths, as generated by this simplified Gilbert model (two-state Markov chain), fits quite well to the measured empirical CDF of the lengths of TTI bursts (here shown for the car-city scenario). When comparing the simulated statistics (geometric distribution) of TTI-gaplengths to the measured statistics, it can be observed that the two-state Markov chain is not capable of representing the correct TTI-gaplength distribution.

Enhancing the model from the simplified version of Gilbert's model to Gilbert's HMM with three parameters (q=0.6137, P=0.00139, and h=0.2439 for the car-city scenario, estimated by using trigram-statistics as proposed in [5]) does not bring sufficient accuracy. The form of the CDF is not even met. Further improvements within the same modeling strategy could be reached via Fritchman's partitioned Markov chains with more than three states (>> 4 parameters). As we want to keep the number of model parameters as small as possible, and because a Weibull distribution perfectly fits the measured TTI-gaplength statistics (see Fig. 16), a change in the strategy towards other renewal processes like semi-Markov models or

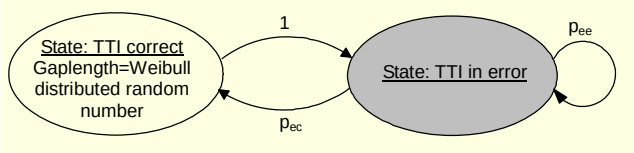

Fig. 17. Half semi-Markov model for the UMTS DCH movement case.

RLMs seems appropriate. The two-parameter Weibull [54] PDF and CDF are given by

$$f(x \mid a,b) = ba^{-b}x^{b-1}e^{-(\frac{x}{a})^b}, \quad (1)$$

$$F(x \mid a,b) = 1 - e^{-(\frac{x}{a})^b}, \quad (2)$$

where a and b are scale and shape parameters, respectively.

In order to be able to generate the TTI-gaplengths via a Weibull distribution and the TTI-burstlengths in a discrete sense with a Markov chain (geometric distribution), a half semi-Markov model with two states can be used (see Fig. 17).

The state in which error-free TTIs are generated is formed according to a renewal process with Weibull-distributed gaplengths, whereas erroneous TTIs are produced one by one in the discrete state. After generating an error gap, a single erroneous TTI is generated which may again be followed by a state change according to the transition probability p_{ec}=1-p_{ee}. Therefore, three parameters are required for defining the model: two parameters for the Weibull distribution (shape parameter a=0.5927, scale parameter b=43.4) and p_{ee}=0.43 (values for the car-city scenario, mobile 1), which are determined by curvefitting to the TTI-gaplength statistics and direct calculation from the mean TTI-burstlength, respectively. Refinements of the model are possible by adding the probabilities of having x TBs erroneously received within one TTI (For 11 parameters, see Fig. 4.) as presented in [35] or by making use of the fact that these values follow a binomial distribution (only one parameter required) as shown in Fig. 5.

D. UMTS DCH Link Error Modeling – Static Case

For modeling the link error characteristics of the UMTS DCH in the static scenario, a model based on TTI granularity is not sufficient due to the high probability of receiving less than all TBs in error within one erroneous TTI. Therefore, the goal of a first modeling approach is to generate a sequence of TBs with the correct statistics of gaplengths and burstlengths.

In cases of the memoryless channel and the simplified Gilbert model, the produced gaplengths follow a geometric distribution [55]. As we have shown in section III.2 (see Fig. 8), there are two main areas with high probability of gaplengths. Obviously, these statistics of gaplengths cannot be met via a single geometric distribution; therefore, the memoryless

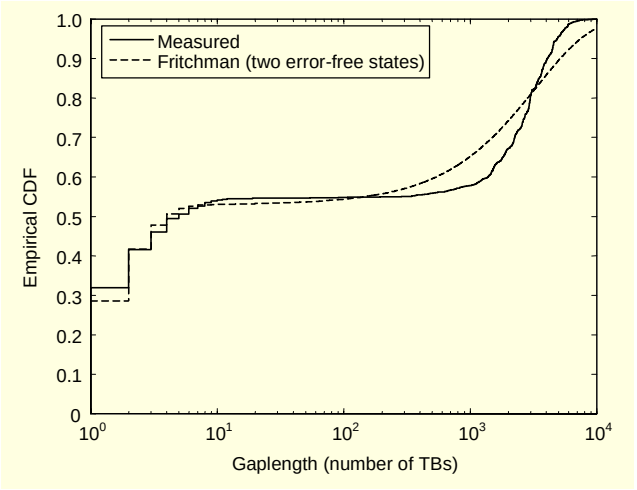

Fig. 18. Gaplength distribution, measured vs. Fritchman's model (two error-free states), 384 kbps bearer, static scenario, mobile 1.

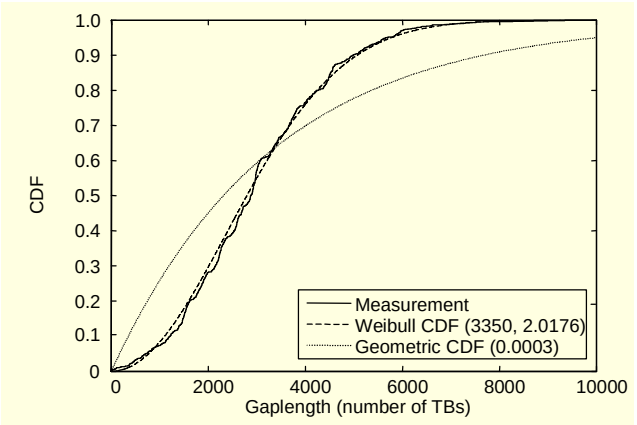

Fig. 19. Distribution of long gaps, measured data vs. Weibull and geometric distributions. 384 kbps bearer, static scenario, mobile 1.

channel and the simplified Gilbert model are not adequate tools in this case.

When using Fritchman's partitioned Markov chains with two error-free states and following the proposal of [15] (estimating the model parameters by separate curve fitting for different parts of the distribution function), the measured distribution of gaplengths is met as shown in Fig. 18.

One of the error-free states is responsible for the distribution of the short gaps, whereas the second state builds the form of the distribution for the long gaps. It can be shown that the fitting for the short gaps can be improved by adding additional error-free states to the model, but further states do not add accuracy to the fitting of the distribution for the long gaps. The reason can be seen in Fig. 19, where the empirical CDFs of the measured data and a CDF of a geometric distribution are shown in linear scale. Note that in Fig. 19 only the CDFs of the

long gaplengths (> 12 TBs) are plotted.

We observe that the curve of the measured data is convex between zero and the inflection point (mode). The geometric distribution, on the other hand, is concave all over its support. Therefore, fitting a geometric distribution (as formed by one state of Fritchman's partitioned Markov chains) to the measured distribution is not possible. Furthermore, we conclude that the best fit is already reached with only one geometric distribution (meaning only one error-free state in Fritchman's model for the long gaps) because every additional geometric distribution in a linear combination with only positive weighting factors adds concavity to the curve, and thus, increases the fitting error. Note that the polygeometric CDF is given by

$$F(x \mid N)_{polygeometric} = 1 - \sum_{i=1}^{N} \mu_i \lambda_i^x, \quad (3)$$

with the constraints of $0<\mu_i<1$ and $0<\lambda_i<1$ for Fritchman's partitioned Markov chains (μ_i and λ_i correspond to the transition probabilities of the Markov chain) and N is the number of error-free states. Thus, all models producing polygeometric distributions of gaplengths like Gilbert's model, Fritchman's model (with any number of states N), or the GEC cannot be used for modeling UMTS DCH link error characteristics in static scenarios.

As the best fit of Fritchman's model to the measured distribution of the long gaplengths is reached with one error state and due to the memoryless property (constant failure rate) of the geometric distribution, Fritchman's model is also incapable of describing the predictive nature of the measured link error characteristics (see section III.5 for detail). The geometric distributions of gaplengths provided by Gilbert's model and by the GEC lead to the same conclusions.

As we can see in Fig. 19, a Weibull distribution with scale parameter a=3350 and shape parameter b=2.0176 perfectly fits the measured distribution of the long gaps. Furthermore, a Weibull distribution (a=1.2, b=0.7) can be used to describe the distribution of the short gaps as shown in Fig. 20(a), and another Weibull distribution (a=1.1, b=0.55) with two additional steps at 12 and 24 TBs represents the statistics of burstlengths for the 384 kbps bearer (see Fig. 20(b)).

Following the explained method of describing adequate statistics for gap- and burstlengths we arrive at a two-state model (two-state alternating Weibull renewal process) as shown in Fig. 21, where in one state, correct TBs are generated, and in the other state, erroneous TBs are generated. After each calculation of either burst- or gaplength, the state is changed. In the correct state, the number of subsequent error-free TBs (gaplength) is calculated via a two-part Weibull distributed random number with a probability of p_{sg}=0.55 to produce short gaps. The number of subsequent erroneous TBs (burstlength)

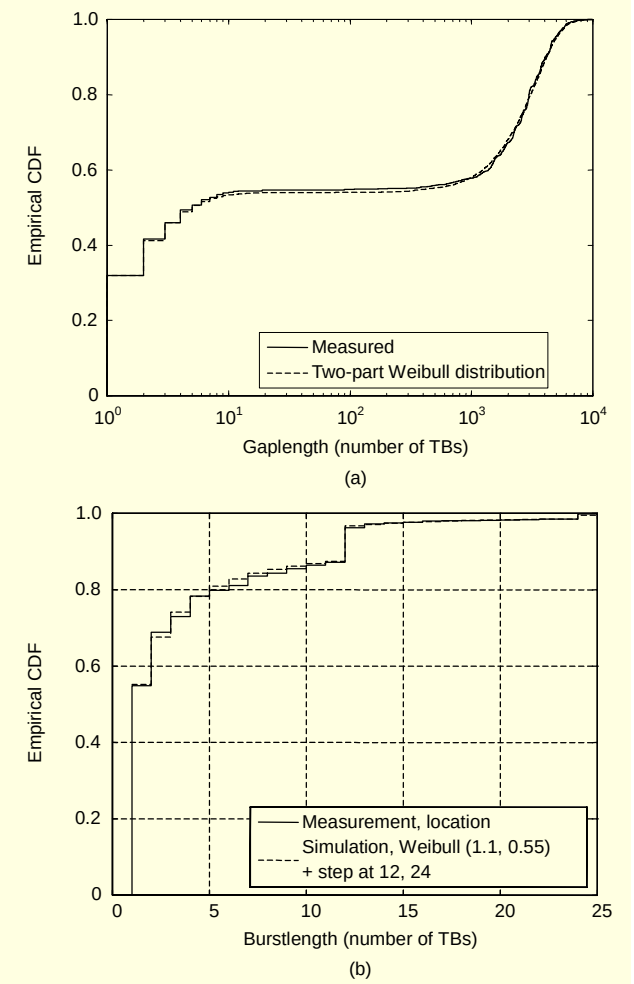

Fig. 20. Comparison of empirical CDFs of (a) gapplength and (b) burstlength. Measurements vs. two-state Weibull renewal process. 384 kbps bearer, static scenario, mobile 1.

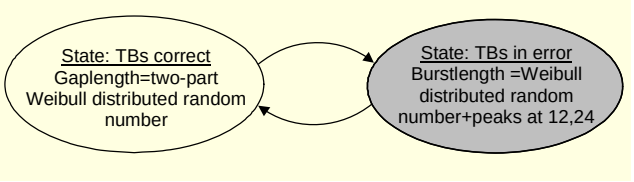

Fig. 21. Two-state model for the UMTS DCH static scenario. Two-state alternating Weibull renewal process (static model 1).

in the error state is also calculated via a Weibull distributed random number (with a probability p_{Wb}) but with additional burstlengths of 12 and 24 with probabilities of p_{12}=0.09 and p_{24}=0.01, respectively. For bearers other than the 384 kbps bearer or for different TBs, these peaks are found at other burstlengths (namely, at 8 and 16 TBs for the 128 kbps bearer used in this work).

This two-state model can be equivalently presented with five

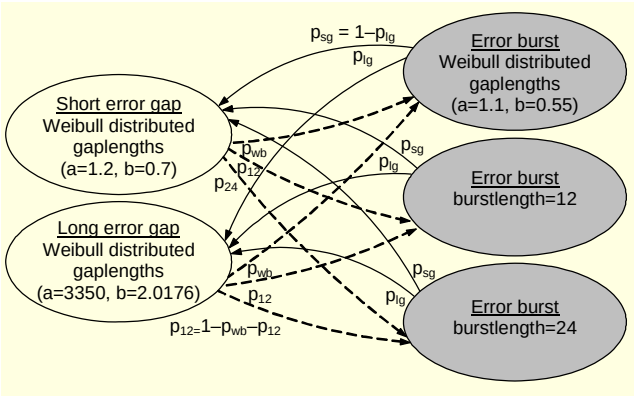

Fig. 22. Equivalent five-state model illustration.

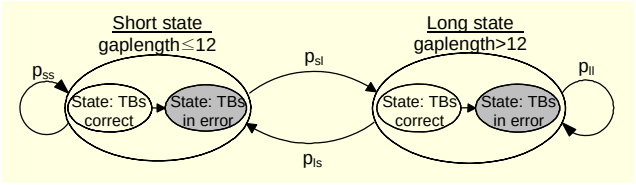

Fig. 23. Non-renewal two-layer Markov model (Markov modulated Weibull renewal process).

states as shown in Fig. 22, which clearly shows that for modeling the error characteristics of the UMTS DCH for the static scenario in this way, a total number of nine parameters is required. These include four parameters for the two Weibull distributions and one parameter for the separation between the distributions of the correct states (error gaps). Additionally, two parameters are required for the Weibull distribution of the error bursts, and two parameters are necessary to select burstlengths of either 12 or 24 or according to the Weibull distribution.

From Fig. 22, we can conclude that this model represents a renewal process, meaning that, for example, the probability of producing a short (long) gap does not depend on the length of the last error burst, and it is also independent of previous gaplengths. On the other hand, as shown in section III.2, there is statistical dependency between subsequent gaps, subsequent bursts, and also between a gap and the following burst. Therefore, in order to capture these correlation properties, a non-renewal model like an HMM is required. Instead of taking a general HMM, we again make use of the special error characteristics of the UMTS DCH to keep the model as simple as possible. Nonetheless, the model describes a major part of the dependencies between gaps and bursts by utilizing conditional probabilities of having long and short gaps and the conditional probabilities for burstlengths of 12 and 24.

Thus, we propose the following non-renewal two-layer model, where dependency between gaps and bursts is introduced via the upper layer two-state Markov chain, which

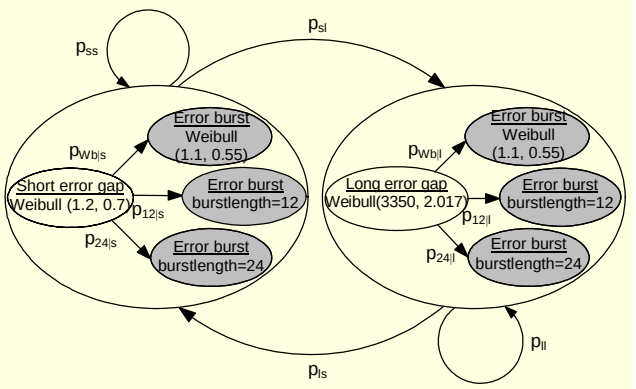

Fig. 24. Detailed illustration of the non-renewal two-layer Markov model (static-model 2).

modulates the parameters of the lower layer process (see Fig. 23). This model also represents an HMM, since from the model output (erroneous/error-free TB) the current state of the model cannot be determined.

The model has one "short state" (gaps with ≤ 12 TBs) and one "long state" (gaps with > 12 TBs), and the corresponding transition probabilities are the conditional probabilities out of section III.2 (according to Fig. 13). After entering a certain state, the gaplength is calculated via a Weibull distributed random number for either the small or the large gaplengths according to the separation in section III.2 and depending on the current upper layer state. Then, while staying in the same state of the upper-layer Markov chain, a burst is calculated with different probabilities of either creating a Weibull-distributed burstlength or a burst with a length of 12 or 24, again, depending on the actual upper layer state. Then, the state of the upper layer Markov chain possibly changes, and the procedure begins again with calculation of the gaplength. Thus, via the upper layer Markov chain, the model introduces dependency between subsequent gaps and also between a gap and the following burst. In Fig. 24, a detailed illustration of this two-layer model is presented. We also conclude that the following 11 parameters are required for a complete definition of the model.

Two parameters are necessary for the determination of the transition probabilities of the upper layer Markov chain (p_{ss}=1-p_{sl}=0.53, p_{ls}=1-p_{ll}=0.56). Two additional parameters are used for the Weibull distributions of each of the correct states, and only two parameters more for the Weibull distribution of both erroneous states as the Weibull distributions are the same for the erroneous states of both upper layer states.

The probability of generating Weibull distributed burstlengths depends on the current state of the upper layer Markov chain in case of the erroneous states. For these probabilities ($p_{Wb|s}$=0.99, $p_{Wb|l}$=0.78) and the separation between a burstlength of 12 and 24 ($p_{12|x}/p_{24|x}$=9), three

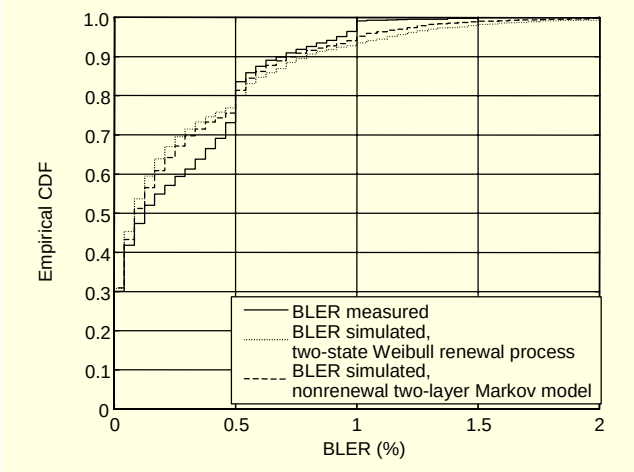

Fig. 25. Comparison of BLERs: measurement vs. models, static 1, mobile 1, 384 kbps bearer.

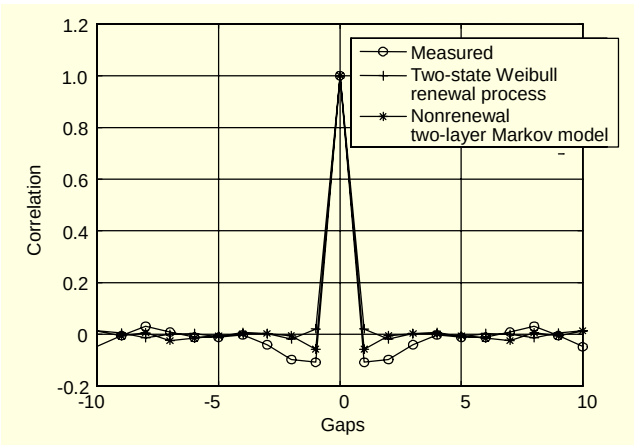

Fig. 26. Comparison of autocorrelation functions: measurement vs. models, static 2, mobile 1, 384 kbps bearer.

additional parameters are required ($p_{\mathrm{Wb}|x} + p_{12|x} + p_{24|x} = 1$).

Despite the fact that this model is a type of HMM, all the parameters can directly be determined from a measured link error trace without the necessity of difficult parameter estimation algorithms.

The two-state Weibull renewal process (static model 1) and the non-renewal two-layer Markov model (static model 2) produce correct total TB error probabilities of 0.0027 and 0.0026, respectively. Also, the mean clusterlengths (10.50, 10.35), the mean numbers of erroneous TBs per error cluster (7.90, 7.83), and the mean numbers of error bursts per cluster (2.17, 2.18) are very close to the measured results for the scenario static 1. Furthermore, as presented in Fig. 25, both models are capable of generating an error trace with an appropriate BLER. However, there is just a little improvement from the non-renewal two-layer Markov model (static model 2) over the two-state Weibull renewal process (static model 1) for the case of the static 1 scenario.

The potential of introducing correlation between subsequent gaps and bursts of the non-renewal two-layer Markov model becomes evident when analyzing the autocorrelation functions of the sequence of gaps (see Fig. 26) for the scenario static 2, which includes a higher grade of dependency between the gaps. While the two-state Weibull renewal process does not produce any dependency between successive gaps, we observe in Fig. 26 that the non-renewal two-layer Markov model captures the correlation properties between neighboring gaps.

5. Analysis of Link Error Predictability in the UTRAN

Detailed knowledge about the link error characteristics is necessary for the optimization of services (such as their codecs or protocols) for use over wireless mobile communication links [2]-[4]. Link error statistics can also be used for performance evaluation of new (cross-layer) mechanisms [38], as for example, for streaming video data transmission in the UMTS network.

Statistics of link error characteristics can be used to optimize the end-to-end transmission chain in a static sense and even dynamically and adaptively to predict future link errors based on past error occurrence and to use this predictability together with cross-layer optimization procedures as will be discussed in greater detail.

The analysis of the link error characteristics of the UMTS DCH in DL shows that there is correlation between the error states of subsequently transmitted TBs (leading to error bursts) and also between successive bursts and gaps resulting in error clusters. Generally, these error correlation properties and the memory in the channel can be used to predict future link errors.

To be able to predict link errors, the transmitter has to be aware of past link errors in the forward link. That is, there has to be a feedback link for the error status of the received data from the receiver back to the transmitter. For the UMTS DCH, this is accomplished by utilizing the error feedback of the RLC AM. For delay sensitive services, we can adjust the maximum number of allowed retransmissions for the RLC AM to zero or equivalently adjust the discard timer properly. The RLC AM in the considered live UMTS networks uses 16 bits of CRC information for each TB to detect transmission errors.

In case there was no or only very small (< 1 ms) feedback delay, the single TB errors within an error cluster could be predicted at the transmitter. Unfortunately, the feedback delay d_{FB} in the UMTS RLC AM ($d_{\mathrm{FB}} \geq 30$ ms also in case of 10 ms TTIs) is in the order of the cluster size; therefore, no error prediction within an error cluster is possible.

For this reason, we focus on predicting error clusters rather than single error events. Moreover, in a first approach, we detect the points in time with the least probability for the occurrence

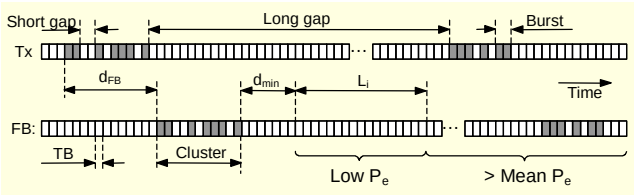

Fig. 27. Schematic illustration of the interval with low link error probability.

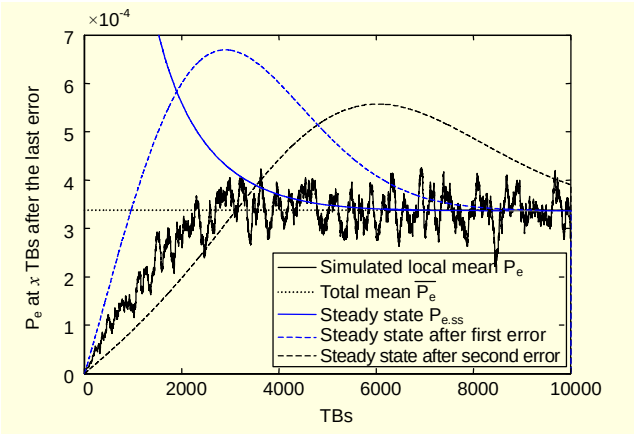

Fig. 28. P_e for Weibull distributed long gaps (Weibull shape parameter a=2.176, scale parameter b=3,350).

of an error cluster.

From the statistics of the gaplengths in the static scenario (see Fig. 8), we know that there are short gaps between the error bursts within an error cluster and long gaps between successive error clusters. We also know that there is a region in the CDF between the short and long gaps with a very low probability for the occurrence of a gap. This is because the CDF of the long gaps is convex between zero and the inflection point (mode). Moreover, long error gaps in the static case are not geometrically distributed (Weibull distribution with shape parameter $a \neq 1$), resulting in memory in the link error characteristics. These properties can be used to detect intervals with low link error probability as shown in Fig. 27.

In case there is no error report via the feedback information within a certain time interval d_{min}, which has to be larger than the longest of the short gaps, we can be sure that we are outside of an error cluster and the current transmission takes place within one of the long error gaps. Due to the convex distribution of the long gaps, there is an interval with very low transmission error probability, and we can detect its start at the minimum delay d_{min} after the last error. Of course, the total detection delay also includes the feedback delay d_{FB}.

To determine the length L_i of the intervals with lower error probability, we make use of a theoretic analysis with long gaps separating single error events only. In Fig. 28, the simulated error probability P_e at a point x TBs after an error event (in practice corresponding to the occurrence of an error cluster) is presented for Weibull-distributed gaplengths. We conclude from this curve that the error probability is very small right after an error event, staying below the total mean error probability for approximately 2,500 TBs. In practice, we may estimate L_i via the intersection of this conditional mean curve with the unconditional mean $\overline{P_e}$.

To obtain an analytic expression for L_i, we first compute the steady state error probability $P_{e,ss}$. Let L be the random variable measuring the time (distance in number of packets) between two error events. Assuming that the error process is stationary, Kac's lemma [56] implies that

$$P_{e,ss} = \frac{1}{\mathrm{E}\{L\}}, \tag{4}$$

where $\mathrm{E}\{L\} = \sum_{i=1}^{\infty} i \cdot P\{L = i\}$ is the average recurrence time corresponding to the average gaplength.

It turns out that the stationarity assumption holds over longer ranges, but not in immediate succession to an error event, where the measured error probability is actually lower than the total mean error probability $\overline{P_e}$, and the steady state error probability $P_{e,ss}$ tends towards one. Therefore, we approximate the probability $P_e(l)$ of having an error at the l-th TB after an error event, assuming that the steady state is reached before the third error after the initial error event:

$$\begin{aligned} P_e(l) \approx\ & P(L_1 = l) + P(L_1 + L_2 = l) \\ & + P_{e,ss} \cdot P(L_1 + L_2 < l). \end{aligned} \tag{5}$$

Here, L_1 and L_2 denote the gaplengths before the first and second errors, respectively. The result (black dashed curve in Fig. 28) shows that taking the intersection between this curve and the total mean as the end of L_i gives a good approximation, whereas the assumption of reaching the steady state after the first error leads to an underestimation of L_i (grey dashed curve).

For some cross-layer methods (see section IV.1), it is not sufficient to detect intervals with lower error probability; rather, an estimation of the instantaneous probability of transmission errors is required. As previously mentioned, due to the feedback delay, an exact error prediction within the error clusters is not possible, but the location of the error clusters can be predicted. Therefore, we estimate the instantaneous link error probability by using the expected failure rate resulting from the long gaps between successive error clusters.

Throughout this paper, the following notation is used to describe the error process. An erroneously received TB is indicated by 1, while 0 means error-free transmission. A positive integer in the exponent determines the number of consecutive erroneous or error-free TBs. For example, the

sequence '000001' can be written as '$0^5$1'. A gap with length m is defined as the number of 0s between two 1s and

$$p_M(m) = P(0^m 1 \mid 1) \tag{6}$$

for all positive integers m is the probability mass function (PMF) of the gaplengths. The conditional probability P(B|A) means the probability of sequence B, following sequence A. By definition,

$$\sum_{m=0}^{\infty} P(0^m 1 \mid 1) = 1 \tag{7}$$

and $P(0^0 1|1)$=0, as gaps with length zero are not considered as gaplengths. The CDF of the gaplengths is then defined as

$$F_M(m) = P\{M \le m\} = \sum_{k=0}^{m} P(0^k 1 \mid 1). \tag{8}$$

The conditional link error probability $P(1|10^m)$ (the error probability conditioned to the number of error-free TBs since the last error) can be expressed by

$$P(1 \mid 10^m) = \frac{P(10^m 1)}{P(10^m)} = \frac{P(0^m 1 \mid 1) \cdot P(1)}{P(0^m \mid 1) \cdot P(1)} = \frac{P(0^m 1 \mid 1)}{P(0^m \mid 1)}, \tag{9}$$

where $P(0^m|1)$ denotes the probability of having a gaplength of at least length m and can be written in terms of the complementary cumulative distribution function (CCDF) and PMF of the gaplengths:

$$P(0^m \mid 1) = \sum_{k=m}^{\infty} P(0^k 1 \mid 1) = 1 - F_M(m) + p_M(m). \tag{10}$$

Thus, the conditional link error probability $P(1|10^m)$ can be expressed as

$$P(1 \mid 10^m) = \frac{p_M(m)}{1 - F_M(m) + p_M(m)} \approx \frac{p_M(m)}{1 - F_M(m)}, \tag{11}$$

also presenting the approximation via the expected failure rate [55].

In section III.4 we show that the long link error gaps can perfectly be fitted via a Weibull distribution [33], [34]. Thus, after inserting the Weibull PMF and the Weibull CCDF in (11), we obtain the estimated conditional error probability

$$\hat{P}(1 \mid 10^m)_{Weibull} = \frac{b a^{-b} m^{b-1} e^{-(\frac{m}{a})^b}}{e^{-(\frac{m}{a})^b}} = b a^{-b} m^{b-1}. \tag{12}$$

We estimate $\hat{P}(1|10^m)$ by using b = 2.0176, corresponding to the statistics of the long gaps measured in [34]. With $b \approx 2$, Weibull distribution becomes Rayleigh, and the estimated expected failure rate can be expressed by the much simpler term

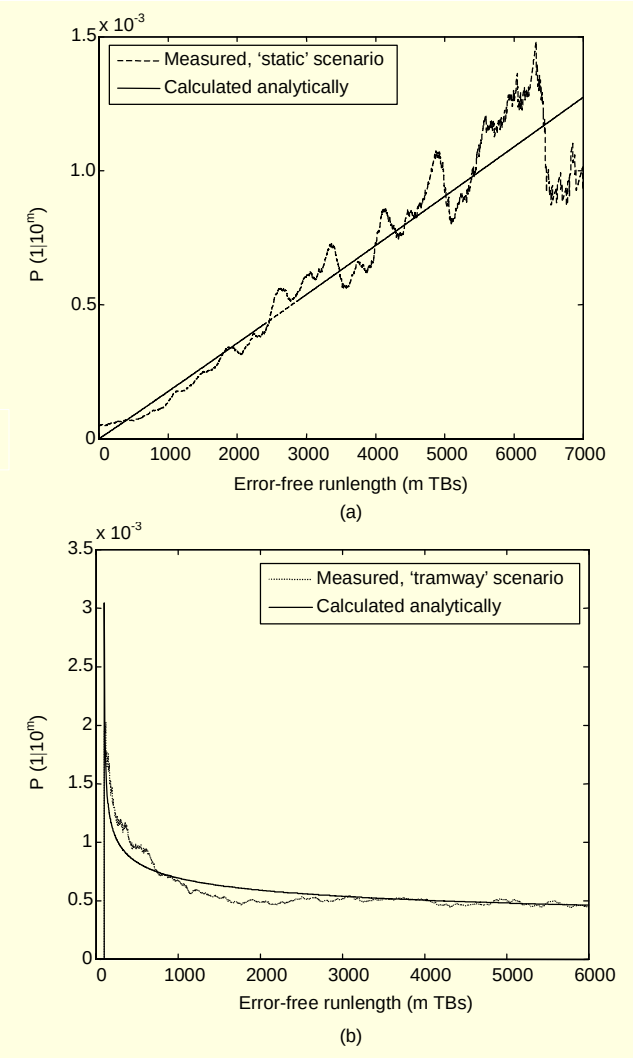

Fig. 29. Conditional link error probability, 384 kbps bearer, static scenario (a), tramway scenario (b), mobile 1.

$$\hat{P}(1 \mid 10^m) = \frac{2m}{a^2}, \tag{13}$$

with a=3350. Thus, it increases linearly with the error-free run length m (the number of error-free TBs received so far after the last error). It can be shown that (11) leads to an approximately linear increase in the considered region of ≤10,000 TBs even without the mentioned approximation.

The measured conditional link error probability $P(1|10^m)$ from the static scenario can be seen in Fig. 29(a). The linear estimator for the failure rate perfectly predicts the transmission error probability based on the number of error-free TBs since the last error (error-free runlength). In [33], we present further analysis of the influence of the feedback delay d_{FB} and the quality of the error prediction due to the modeling error of the gaplength statistics.

The performance of such an error estimator can be seen in

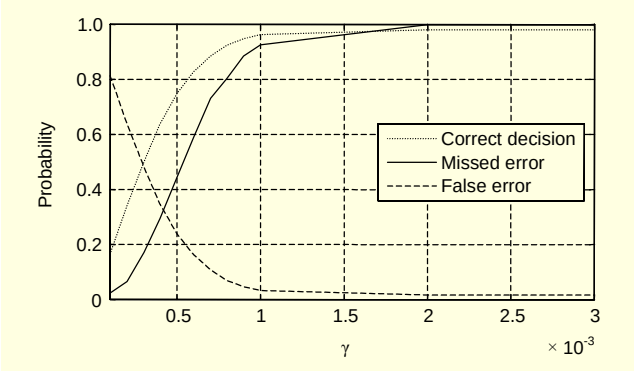

Fig. 30. Prediction performance.

Fig. 30 for different decision thresholds γ and the static scenario.

This figure was obtained by applying the following hard decision: if $\hat{P}(1|10^m) \geq \gamma$ an error is predicted. The probability of false error P_f is the conditional probability that $\hat{P}(1|10^m) \geq \gamma$ if no error occurred. The probability of missed error P_m is the conditional probability that $\hat{P}(1|10^m) < \gamma$ if an error occurred. The probability of a correct decision $P_c = 1 - [(1-P_{err})\cdot P_f + P_{err}\cdot P_m]$ is the probability that an error was predicted and occurred plus the probability that an error was not predicted and did not occur; P_{err} denotes the probability of a link error.

In most applications (cross-layer processing algorithms), both P_m and low P_f are desired. Therefore, thresholds around the crossing of P_m and P_f are relevant.

IV. Cross-Layer Optimization for Video Streaming over UMTS

The 3GPP standardizing UMTS has approved the inclusion of H.264/AVC as an optional feature in Release 6 of its packet oriented mobile multimedia telephony [57] and streaming services [58] specifications. Although H.264/AVC is designed for robust delivery of streamed video over error prone mobile communication channels, the link errors severely affect the video quality. In the following we present how the specific link error characteristics, especially the predictability of link errors, can be exploited in cross-layer processing methods to optimize the transmission of H.264/AVC video streams.

Whereas traditional (opportunistic) scheduling algorithms, such as those presented in [59] and [60], concentrate on the multiuser diversity gain from channel variations, on the multiservice gain, or on the gain from multiple streams, we optimize the video quality of a single video stream for one user. Instead of having a shared wireless channel, our cross-layer processing algorithms focus on the UMTS DCH (dedicated to a single stream of one user only) and the exploitation of the error characteristics of this channel.

Moreover, our cross-layer scheduling algorithms consider the priority of streamed video data packets. The finally presented network and video-content aware cross-layer scheduling method explicitly defines the data priority according to the video content.

Note, the proposed in-stream scheduling algorithms can also be included in multiuser, multiservice, and multistream (opportunistic) schedulers of new or future mobile communication systems with shared channels in order to further increase the QoS or to save radio resources.

1. Network Aware Cross-Layer Processing

As already shown in section II, the coded video data is transmitted by using the UDP protocol, providing a 16 bit checksum for the error control mechanism. The UDP packets containing parts of video slices are further segmented into smaller RLC PDUs (320 bits) in the UTRAN. These RLC PDUs together with the 16 bit CRC form TBs, again providing a possibility for error detection.

In case of a link error (leading to one erroneous TB), the whole UDP packet is usually discarded due to a wrong 16 bit

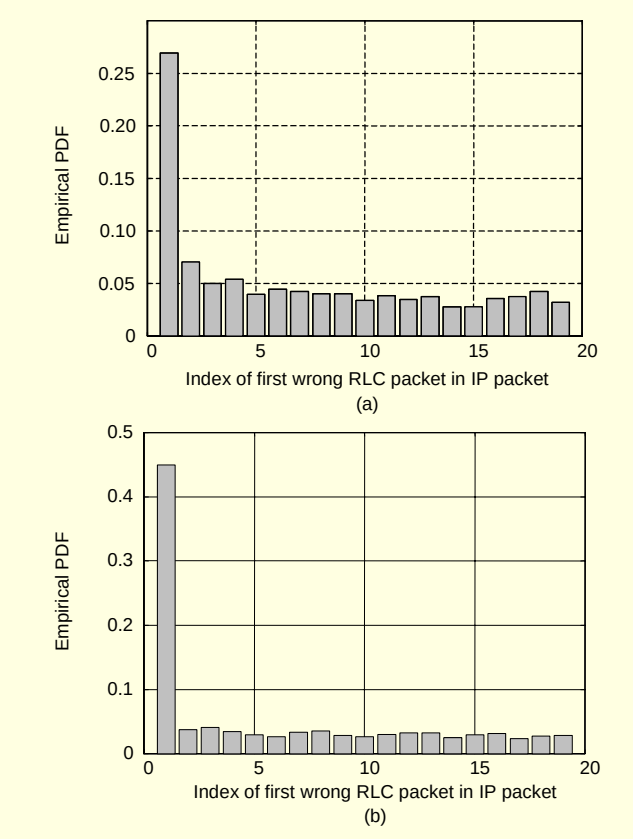

Fig. 31. Distribution of the index of the first erroneous RLC packet (TB) within IP packet: (a) static scenario and (b) movement scenario. Mobile 1.

UDP checksum, leading to significant quality degradation in the streamed video. In [38], we propose and evaluate a cross-layer error detection mechanism which allows for a utilization of the correct parts of the video slice within the erroneous UDP packet.

Without any change of standards, and without adding data overhead, the correct parts of the UDP packet until the first erroneous TB can be utilized for decoding the streamed video data by simply forwarding the error information from layer two to the higher layers (video decoder) including UDP Lite [61], providing the possibility of forwarding the erroneous UDP data packets, and CUDP [62], where error location information is also forwarded between the protocol layers.

From the results of the analysis with the link error characteristics of the UMTS DCH shown in Fig. 31, we can conclude that in more than 73% or 65% of cases, this method provides gain in video quality in the static scenario and movement scenario, respectively. Obviously, the shorter error bursts in the static scenario lead to a lower probability of receiving the first RLC PDU within an erroneous UDP packet in error, which increases the efficiency of the proposed method.

To evaluate the improvement of the end-to-end video quality, we use the peak signal-to-noise ratio of the luminance component (Y-PSNR) given for the n-th luminance frame Y_n by

$$\text{Y-PSNR}(n) = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}(n)} \quad \text{and} \tag{14}$$

$$\text{MSE}(n) = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} [Y_n(i,j) - F_n(i,j)]^2, \tag{15}$$

where MSE(n) denotes the mean square error of the n-th luminance frame Y_n compared to the luminance frame F_n of the reference sequence. The resolution of the frame is $N \times M$, and indexes i and j address particular luminance values within the frame. As a reference sequence, we used the non-compressed original (non-degraded) sequence.

In Fig. 32, we show that with the proposed method, up to 15 dB gain in video quality (Y-PSNR averaged over erroneous frames) can be reached when performing spatial error concealment for the missing parts of the erroneous slices (in case of an I frame). For this analysis, the Foreman sequence was encoded with a quantization parameter (QP) of 28, and the I frame frequency was set to 20. Video slices of a maximum of 700 bytes have been used because this value provides a good trade-off between resulting slice error probability and data overhead [63].

The proposed method still gains up to 6 dB when using temporal error concealment methods like simple copying from the same position in the last frame (in case of a P frame). The Y-PSNR for Fig. 32 was calculated comparing the degraded/concealed sequence to the original non-compressed one. More detailed information about the simulation setup is presented in [38].

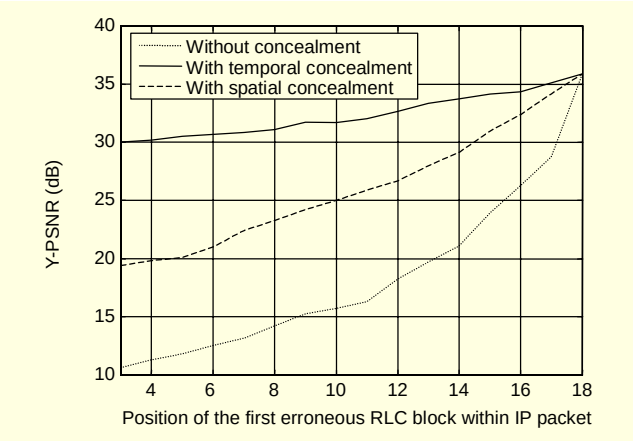

Fig. 32. Y-PSNR vs. index of first erroneous RLC packet within the UDP packet.

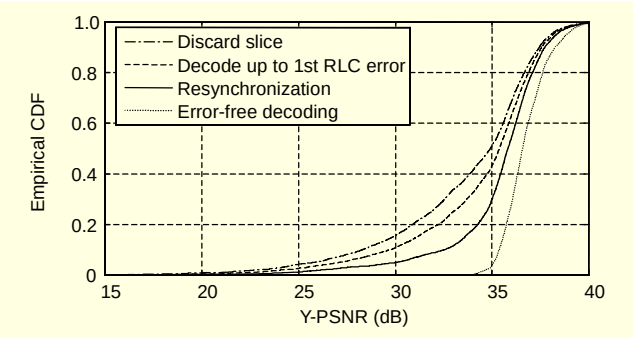

Fig. 33. Quality of video reconstruction for different error handling mechanisms, soccer match video sequence, link error trace from static scenario, mobile 1.

The gain of the proposed method (decoding the video up to the first erroneous RLC PDU within the UDP packet) for a soccer match video sequence (QP=26) simulated over a measured link error trace is presented in Fig. 33. Although there is a low overall probability of a TB error of 0.266% (resulting in a UDP/IP packet error rate of 0.888%) there is considerable improvement in comparison to the usual practice of discarding the whole erroneous UDP packet (video slice).

The utilization of the correct parts within one erroneous UDP packet after an erroneous RLC PDU of course would provide further increase in the video quality. Due to the variable length coding (VLC) in H.264/AVC, this can only be accomplished either by changing the encoding process or by transmitting additional side information. The reason is that after a bit error, context adaptive VLC (CAVLC) may easily desynchronize,

making the correct distinction between neighboring codewords impossible. The decoding of such a desynchronized stream may result in considerable visual impairments or may become even impossible due to non-existing code-words or too many/few bits left for decoding. To overcome this problem, a reversible VLC is proposed in [64], resulting in a higher data rate due to lower compression gain. Another possibility is to transmit additional synchronization information either within the VLC stream [65] or outside of the stream as we proposed in [38]. The resynchronization information can be advantageously represented as the position of the start of the first macroblock (MB) contained in the RLC PDU. For RLC PDUs with 320 bits and fixed length encoding, 9 bits per MB position indicator (MBPI) are required [39].

Figure 33 shows that the gain from decoding all error free parts of the UDP packet (achievable with the additional resynchronization information) compared to the decoding until the first error occurrence is still more than 5 dB in Y-PSNR for some erroneous frames, which is significant from the user experienced quality point of view.

Of course, sending the side information necessary for the VLC resynchronization requires a higher rate. In order to decrease the required rate for sending the resynchronization information, we can make use of the link error predictability from section III.5 in a proposed redundancy control mechanism [39].

Since the MBPI information is used by the decoder only if an error in the radio interface occurs, we can reduce the required rate by sending the side information packets only if a high error probability is predicted. The decision about low/high error probability is based on a single threshold γ in the estimated expected failure ratio $\hat{P}(1|10^m)$ according to (13), where the error probability is estimated via the instantaneous error-free run distribution. It turns out that $\gamma=2\times10^{-4}$ leads to good results.

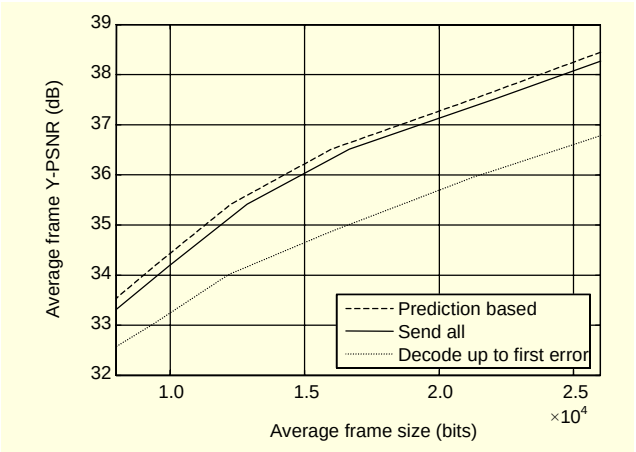

Fig. 34. Average Y-PSNR over the rate. The rate is expressed as average size of a frame including side information in bits.

Figure 34 illustrates the average quality at the decoder normalized by the required rate.

There is an improvement of approximately 0.25 dB by using the error prediction as compared to the full resynchronization (sending all MBPIs). This is a considerable improvement, since the averaging was performed over the entire video sequence, containing error-free packets and having a small TB error probability (0.266%).

We also can conclude from Fig. 34 that the proposed methods (full resynchronization and resynchronization based on link error prediction) provide the possibility of significantly reducing the rate at the radio interface while keeping the same quality at the decoder.

2. Network and Data-Priority Aware Cross-Layer Scheduling

Currently, there is great effort in research to optimize services like video streaming for transmission over wireless communication networks with their higher error probability and bursty link error characteristics. The introduction and continuous improvement of H.264/AVC is the best example. Mobile communication links are also being optimized, as their properties are adapted to meet the requirements of the services for providing better QoS. For example, in [66], a truncated power control is introduced to improve the video quality and ensure efficient transmission of the video stream.

Another approach is to apply a cross-layer algorithm which connects the service properties with the network properties in an optimum way by making use of the continuous information flow from higher layers (data priority information from application layer) and from the channel or the network (channel quality status information from link layer or physical layer). In [67] and [68], (opportunistic) scheduling algorithms are presented which make use of the characteristics of the streamed video data. In [67], the more important parts of the video stream are transmitted prior to the less important ones to ensure more opportunities for retransmissions in case of an error. On the other hand, the priority-based scheduling presented in [68] exploits the diversity gains embedded in the channel variations when there is more than one stream. The gain from having a multi-user and multi-service scenario is also exploited in another approach in [69], where prediction of the link errors is used in connection with call admission control and scheduling algorithms to avoid the system being overloaded and thus improving the quality of the services with higher priority.

We focus on the transmission of a single video stream within one channel which is dedicated to this particular stream. We use

the ability to predict the link errors of the UMTS DCH to improve scheduling of the individual layer-three packets of the video stream with different priority levels. Unlike in [66], where transmission is stopped in times of bad channel quality, our scheduling algorithm makes use of all the available bandwidth, but delays the packets with higher priority to a position where least error probability is predicted [36]. Our method allows the improvement of the quality of streamed video without reducing the quality of other services as occurs in [69].

The start of intervals with low error probability can be predicted, as presented in section III.5. We have shown that after the occurrence of an error cluster the transmission takes place within one of the long gaps; thus, we can predict a very low error probability beginning at d_{min} after the last error of the cluster.

In Fig. 36, we present the packet error probability of layer-three packets with a size of 720 bytes (= 18 TBs), transmitted with delays d_r to the end of d_{min} = 37 TBs as illustrated in Fig. 35. The packet size of 720 bytes was selected as it represents a typical size for video streaming layer three packets (including overhead) for an adjusted maximum slice size of ≈ 700 bytes in the video codec setting, which in turn provides a good trade-off between slice error probability and data overhead. Furthermore, 720 bytes can be divided by 320 bits (RLC PDU size). The simulations for Fig. 36 were performed using the novel UMTS DCH link error model (static model 2).

Figure 36 shows that right at d_{min}, the packet error probability reaches its minimum (< 0.1%), exceeding the total mean (over the whole trace) of 0.925% at d_r ≈ 3,000 TBs.

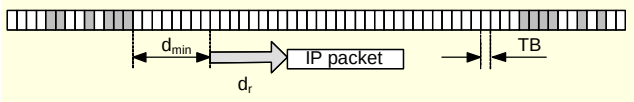

Fig. 35. Illustration of the simulation procedure for deriving Fig. 36.

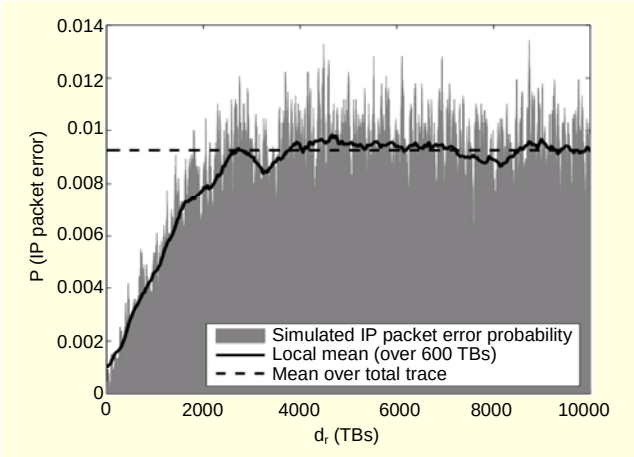

Fig. 36. Packet error probability vs. relative position d_r to the end of d_{min}.

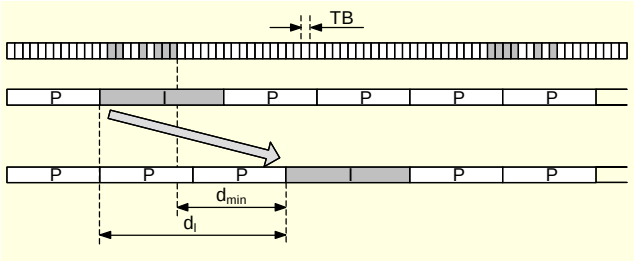

Fig. 37. Schematic illustration of the proposed network and data-priority aware cross-layer scheduling algorithm.

The goal of the proposed scheduling algorithm is to transmit the packets with high priorities at the time instants for which low error probability is predicted. In a video stream, the I frames are more important than the P and B frames since they refresh the stream. If an error occurs in an I frame, it propagates over the whole group of pictures (GOP) up to the next I frame in the worst case. To prevent possible error propagation beyond an I frame, spatial error concealment is usually used for I frames, having lower performance than the temporal error concealment which is used for P frames. In this work, we give layer-three packets containing I frames higher priority than packets containing P frames (no B frames are used).

The proposed scheduling algorithm is illustrated schematically in Fig. 37, where in the upper part, the time series of the transmitted TBs is shown with the erroneous TBs marked. The scheduling algorithm tries to map the I packets (layer-three packets containing parts of I frames) onto TBs which are to be transmitted within the time intervals in which low error probability is predicted. This is accomplished by buffering the incoming I packets and delaying their transmission to a point after the next occurrence of an error cluster. After detecting d_{min} consecutive error-free TBs following an error report (performed by resetting the counter if there is an erroneous TB within d_{min}), that is, after the expected end of the error cluster, all the I packets within the buffer are transmitted. The rest of the TBs (RLC PDUs) are filled with the incoming P packets (layer-three packets containing parts of P frames), which may be transmitted in times of high expected link error probability as they are less important for decoding the video stream.

This method causes additional transmission delay (d_l) for the I packets, and thus, also for the I frames. However, in [36], we have shown that the maximum resulting transmission delay for the I frames is about 6 seconds, which does not cause any deterioration in the quality of the received video as it remains within the storage capacity of common sizes of playout buffers (pre-roll buffers) at the receiving terminals. Current video streaming applications usually use 5 to 20 second playout buffers to cope with the channel fluctuations and the inherent

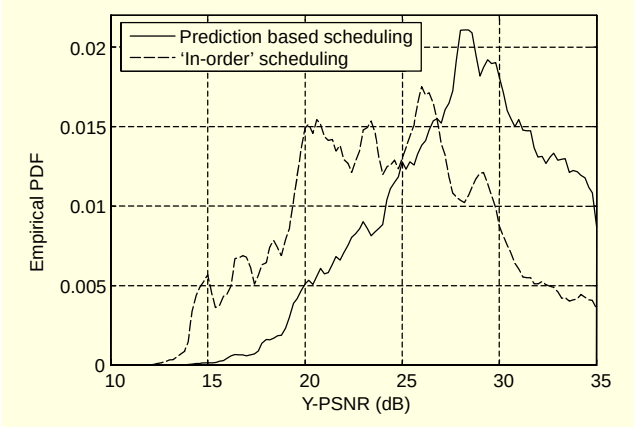

Fig. 38. Gain of network and data-priority aware cross-layer scheduling. Empirical PDFs of lower Y-PSNRs per frame.

variable-bit-rate nature of coded video sequences [67]. However, if the maximum allowed I frame transmission delay (≤ playout buffer size) is exceeded, the concerned I frames are transmitted immediately. We have also shown in [36] that with a limitation of the playout buffer to 3 seconds, our method still can be used successfully in 85 % of cases.

The resulting gain in video quality can be seen in Fig. 38, where the empirical PDFs of Y-PSNR values per frame are presented. The results have been generated in simulations with the well-known Foreman video sequence in QCIF resolution (144×176 pixel) encoded with I and P frames only, in slicing mode two with 700 bytes per slice, 15 frames per second, and an I frame frequency of 20. With a quantization parameter of 25, a video stream with an average bit rate of 300 kbps was obtained.

The empirical PDFs in Fig. 38 show that the presented cross-layer scheduling method reduces the number of frames with lower Y-PSNR values, whereas the number of frames with higher Y-PSNR values increases, thus, resulting in a shift of approximately 5 dB to better video quality. Note, Fig. 38 only contains the lower Y-PSNR range. In the higher range, there is a peak for the Y-PSNR values corresponding to the error-free frames.

For the results of Fig. 38, the decoder used temporal error concealment for the inter-predicted frames (P frames) and spatial error concealment for the intra-coded frames (I frames). To see the benefit of the proposed scheduling algorithm without assuming a particular error concealment method, we evaluated the reduction in the number of erroneous I and P frames, which was up to 83.4% and 4.2%, respectively.

The proposed method does not require changes in the standard method of encoding the video stream and can be applied in UMTS networks without changes to 3GPP specifications. The only required modification is to forward both the content of the layer-three packets (I or P) and the error status of the received TBs in the protocol stack. At the scheduler (transmitting side), the information about the error status of the TBs at the receiving terminal is needed. This feedback information can be acquired by using the RLC AM mode with the maximum number of retransmissions set to zero.

Of course, the feedback delay of RLC AM (≈ 30 ms if one TTI=10 ms) imposes delayed detection and less efficient utilization of the intervals with low error probability. However, a delay of approximately 36 TBs only causes negligible loss in the efficiency of the method.

3. Network and Video-Content Aware Cross-Layer Scheduling

The proposed cross-layer processing method presented in the previous section exploits the specific link error characteristics and the different priority levels of streamed data packets, where the priority of the data packets is set according to their importance for the video decoding process but regardless of the particular video content.

In addition to the unequal importance of I and P frames for the decoding process, the efficiency of the coding and the performance of the error concealment methods heavily depend on the video content. For example, in scenes with movement, packet loss could result in severe degradation in video quality due to spreading of the error over the picture, and temporal error concealment after the loss of a packet at a scene cut is not capable of reproducing the subsequent video frames properly.

Therefore, in order to minimize the distortion in the decoded video stream caused by corrupted data packets, we propose a network-aware cross-layer scheduling method [40], which treats the individual data packets according to their priority which is based on the distortion their loss would cause. In this way, the scheduler considers coding characteristics as well as video content properties.

Rate-distortion optimized (RDO) video streaming was proposed in [70], where a model for delay and loss in the Internet is used to perform scheduling at the IP layer. In our work, the distortion model required to define the priority levels of transmitted data packets considers losses at the RLC PDU level as presented in [40]. The impairments caused by the loss of a PDU are modeled as *primary distortion* (which is the distortion in the frame of the lost PDU) and *distortion propagation* (the propagated distortion in the subsequent frames until the end of the GOP), including error concealment. The model assumes that the video decoder can use the information in an NALU, typically encapsulated further in an RTP/UDP/IP packet, up to the first missing RLC PDU as proposed in section IV.1. The distortion estimation in the model considers the position and amount of lost data, as well as the

size of the corresponding picture area (affected by the loss of the single PDU), thus, introducing awareness of video content like movement, scene cuts, and so on.

The primary distortion $\hat{\varepsilon}_0$ is represented as a function of the number of affected RLC packets (PDUs) N and the corresponding number of affected macroblocks M:

$$\hat{\varepsilon}_0 = e^{a \cdot M + \frac{b}{N} + c} \tag{16}$$

with parameters a, b, and c determined by least squares fitting.

In the model data set, the following values were obtained: a=–0.041, b=–6.371 and c=4.482. The number of lost/damaged MBs required by the model can be obtained without fully decoding the video as a proportion of the MBs per NALU packet (if it is fixed) corresponding to the proportion of lost RLC PDUs, by reading of the entropy encoded stream (without decoding the video) or by extra signaling.

The error propagation is modeled as an exponential decrease of the primary distortion:

$$\hat{\varepsilon}_k = \hat{\varepsilon}_0 \cdot e^{-s \cdot k}, \tag{17}$$

where k is the frame index and s determines the speed of decay, which depends strongly on the nature of the motion in the sequence. Since we assume that only M and N are available for the scheduler, the error propagation has to be estimated by an average over all sequences in the model data set. The decay thus obtained is $s = 0.08$.

The proposed network and video-content aware cross-layer scheduling mechanism then makes use of the estimation of the *cumulative distortion* $\hat{\varepsilon}_\Sigma$ that would be caused by the loss of a particular RLC PDU. This is calculated for every RLC PDU in the scheduling buffer as

$$\hat{\varepsilon}_\Sigma = \sum_{k=1}^{m} \hat{\varepsilon}_{k-n}, \tag{18}$$

where n denotes the number of the frame where the RLC PDU under consideration is located, and m is the number of frames in the GOP.

Concerning network awareness, this scheduling algorithm again makes use of the predictability of intervals with low error probability. The estimated length L_i of these intervals (calculated according to (5)) is also required.

At each step, if an RLC PDU remains in the scheduling buffer longer than the maximum acceptable delay d_{max}, then it is scheduled immediately. Otherwise, at each step within the time interval L_i, the RLC PDU associated with the highest estimated cumulative distortion $\hat{\varepsilon}_\Sigma$ is chosen from the scheduling buffer. Outside the interval L_i, the RLC PDU with the lowest $\hat{\varepsilon}_\Sigma$ is scheduled. If several RLC PDUs have the same (lowest or highest) value of $\hat{\varepsilon}_\Sigma$, the oldest among them is scheduled.

The performance of the scheduler can be controlled by the scheduling buffer length $L_b \geq L_i$ and d_{max}. The delay d_{max} depends on the requirements of the service. It should be chosen smaller than or equal to the application's play-out buffer length. Moreover, the maximum gap length (see Fig. 8) provides a guiding value for determining the size of d_{max}.

For evaluation of the gain of the proposed scheduler, we chose the music video clip sequence "Videoclip" as a test sequence, since it contains a variety of different scenes separated by cuts and gradual transitions. The sequence was encoded with 7.5 frames per second, a GOP size of 40 frames, and a QP of 26. The size of the slices was limited to 650 bytes, and it had no RDO, no rate control, no DP, and only one reference frame. CAVLC was used with the remaining settings corresponding to the baseline profile. The transmission of the sequence over a measured link error trace of the static scenario was simulated 1,500 times (representing approximately 3 h of video).

Figure 39 shows the empirical CDFs of frame PSNR values of the content-aware scheduling method compared to the data-priority aware algorithm as proposed in the previous section and to the transmission without a special scheduling method (with in-order scheduling). Note, all CDFs approach one at PSNR = ∞ (error-free frames), and Fig. 39 only represents the relevant part of the CDF for this analysis.

Figure 39 shows that the data-priority aware scheduling method globally reduces the number of erroneous frames since it moves the errors from the first frame (I frame) in the GOP to the later frames (P/B frames), making the error propagation chain shorter. The proposed content aware scheduling, on the other hand, reduces the number of frames with higher distortion and consequently increases the number of frames with lower distortion. As shown in Fig. 39, the total number of frames containing even a small error can increase after applying the content aware scheduling algorithm. This is caused by a possible

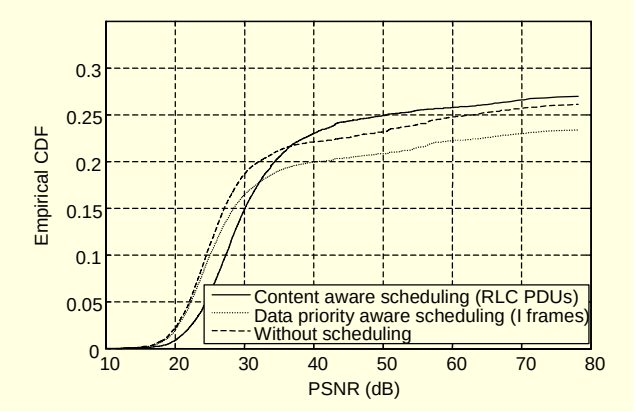

Fig. 39. CDF of frame PSNR: video-content aware scheduling vs. data-priority aware scheduling for videoclip sequence.

spreading of (TB) burst errors onto multiple frames in order to reach minimum distortion in the video. However, due to the different mechanisms of the two scheduling methods, we cannot compare their performance by Fig. 39; therefore, the mean PSNR values have to be used instead.

The resulting mean PSNR for transmission without scheduling was 29.29 dB. With data-priority aware scheduling, (I frame scheduling) 30.32 dB was reached; with video-content aware scheduling, 32.03 dB was reached.

From these results, we conclude that considering the specific link error characteristics (predictability of intervals with lower error probability) within cross-layer algorithms provides up to 2.7 dB in quality improvement, which is significant from the users' perceived quality point of view. These improvements are reached with simultaneous awareness of the importance of the streamed data packets (UDP/IP) for the decoding process (offering 1 dB) and with consideration of the priority of the transmitted data packets (RLC PDUs) with respect to the video content, resulting in 1.7 dB more. These improvements are considerable as the simulated link error rate was very small (0.26% based on the measurements in the live UMTS networks), which, therefore, offers small improvement margins only.

V. Summary and Conclusions

In this work, we presented a link error analysis of the UMTS DCH based on measurements in live networks with different levels of mobility. We showed that due to the link error characteristics, basically two scenarios have to be distinguished: static and dynamic (regardless of which kind of mobility). The analysis showed that the quality-based power control mechanism induces correlation to the link error characteristics, resulting in error bursts, error clusters, and the predictability of the link errors.

We showed that due to the feedback delay, the errors within the bursts or clusters cannot be predicted. We presented methods to predict intervals with low error probability and the instantaneous failure rate, which rely on the fact that the lengths of the error-free intervals between the error clusters are not geometrically distributed due to the properties of the quality-based power control mechanism. Therefore, in novel or future mobile communication systems, we can expect the possibility of link error prediction if the systems comprise a quality-based control mechanism.

We demonstrated that well known error models are not capable of properly describing the measured specific error characteristics (especially error predictability) of the UMTS DCH. Therefore, we presented a new modeling approach based on semi-Markov models and a (Markov modulated) Weibull renewal processes, providing good usability and high accuracy with low complexity. Because the basic idea behind this modeling approach is to consider the specific properties of the underlying network, it can also be applied to other new or future mobile communication systems.

It is important to consider the specific link error characteristics for the performance evaluation of new services, (cross-layer) designs or higher layer protocols. We showed that a thorough understanding of the specific link error characteristics also becomes significant for the development of cross-layer algorithm which explicitly exploit the particular properties of the underlying link.

We presented a novel cross-layer processing design for H.264/AVC video streaming over UMTS. A considerable gain in video quality was reached introducing network awareness by making use of the capability to predict instantaneous failure rates. Redundancy information for decoding the streamed video data is sent only when high error probability is predicted.

Further improvements in video quality were achieved by adding awareness of the priority of the encoded video data packets. The presented network and data-priority aware scheduler transmit packets with high importance for the decoding process in time intervals where low error probability is expected.

Finally, we presented a novel network-aware scheduler which considers the priority of the layer-two packets of the encoded video stream according to the distortion they would cause in case of their loss. Thus, via a distortion model, the scheduler additionally is aware of the video content, such as scene cuts or movement. Up to 2.7 dB gain in video quality was achieved.

Acknowledgements

We thank Mobilkom Austria AG for technical and financial support of this work. The views expressed in this paper are those of the authors and do not necessarily reflect the views within Mobilkom Austria AG .

References

[1] H. Holma and A. Toskala, *WCDMA for UMTS, Radio Access For Third Generation Mobile Communications*, John Wiley & Sons, Ltd., 2004.

[2] Y.J. Liang, J.G. Apostolopoulos, and B. Girod, "Analysis of Packet Loss for Compressed Video: Does Burst-Length Matter?" *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, 2003, pp. 684-687.

[3] W. Karner, O. Nemethova, and M. Rupp, "The Impact of Link Error Modelling on the Quality of Streamed Video in Wireless Networks," *3rd IEEE Int. Symp. on Wireless Comm. Systems*

(ISWCS), Valencia, Spain, Sept. 2006.

[4] M. Zorzi and R.R. Rao, "Perspectives on the Impact of Error Statistics on Protocols for Wireless Networks," *IEEE Personal Comm.*, vol. 6, pp. 32-40, Oct. 1999.

[5] E.N. Gilbert, "Capacity of a Burst-Noise Channel," *Bell Systems Technical Journal*, vol. 39, Sept. 1960, pp. 1253-1265.

[6] E.O. Elliott, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell Systems Technical Journal*, vol. 42, Sept. 1963, pp. 1977-1997.

[7] L.R. Rabiner and B.H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, vol. 3, Jan. 1986, pp. 4–16.

[8] J.M. Berger and B. Mandelbrot, "A New Model for Error Clustering in Telephone Circuits," *IBM Journal*, July 1963.

[9] S.M. Sussman, "Analysis of the Pareto Model for Error Statistics on Telephone Circuits," *IEEE Trans. Comm. Systems*, vol. 11, June 1963, pp. 213-221.

[10] E.O. Elliott, "A Model for the Switched Telephone Network for Data Communications," *Bell Systems Technical Journal, Tech. J.*, vol. 44, Jan. 1965, pp. 89-119.

[11] B. Mandelbrot, "Self-Similar Error Clusters in Communication Systems and the Concept of Conditional Stationarity," *IEEE Trans. Comm.*, vol. 13, no. 1, Mar. 1965, pp. 71-90.

[12] B.D. Fritchman, "A Binary Channel Characterization Using Partitioned Markov Chains," *IEEE Trans. Information Theory*, vol. 13, no. 2, Apr. 1967, pp. 221-227.

[13] J-P.A. Adoul, B.D. Fritchman, and L.N. Kanal, "A Critical Statistic for Channels with Memory," *IEEE Trans. Information Theory*, vol. 18, no. 1, Jan. 1972.

[14] L.N. Kanal and A.R.K. Sastry, "Models for Channels with Memory and Their Applications to Error Control," *Proc. IEEE*, vol. 66, no. 7, July 1978.

[15] J-Y. Chouinard, M. Lecours, and G.Y. Delisle, "Estimation of Gilbert's and Fritchman's Models Parameters Using the Gradient Method for Digital Mobile Radio Channels," *IEEE Trans. Vehicular Technology*, vol. 37, no. 3, Aug. 1988.

[16] C. Pimentel and F. Blake, "Modeling Burst Channels Using Partitioned Fritchman's Markov Models," *IEEE Trans. Vehicular Technology*, vol. 47, no. 3, Aug. 1998.

[17] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. the IEEE*, vol. 77, no. 2, Feb. 1989.

[18] W. Turin and M.M. Sondhi, "Modeling Error Sources in Digital Channels," *IEEE Journal on Sel. Areas in Comm.*, vol. 11, no. 3, Apr. 1993.

[19] S. Sivaprakasam and K.S. Shanmugan, "An Equivalent Markov Model for Burst Errors in Digital Channels," *IEEE Trans. Comm.*, vol. 43, no. 2/3/4, 1995.

[20] W. Turin, *Digital Transmission Systems: Performance Analysis and Modelling*, McGraw-Hill, New York, 1999.

[21] R.P. Aldridge and M. Ghanbari, "Bursty Error Model for Digital Transmission Channels," *IEE Electronic Letters*, vol. 31, no. 25, Dec. 1995.

[22] G.T. Nguyen and B. Noble, "A Trace-Based Approach for Modeling Wireless Channel Behavior," *Proc. the 1996 Winter Simulation Conf.*, 1996.

[23] V. Tralli and M. Zorzi, "Markov Models for the Physical Layer Block Error Process in a WCDMA Cellular System," *Proc. IEEE GLOBECOM*, vol. 2, Nov. 2002, pp. 1925-1929.

[24] H.S. Wang and N. Moayeri, "Finite-State Markov Channel — A Useful Model for Radio Communication Channels," *IEEE Trans. On Vehicular Tech.*, vol. 44, no. 1, Feb. 1995.

[25] A. Willig, "A New Class Of Packet- And Bit-Level Models For Wireless Channels," *Proc. 13th IEEE Int. Symp. Personal, Indoor and Mobile Radio Comm.*, vol. 5, Sept. 2002, pp. 2434-2440.

[26] A. Köpke, A. Willig, and H. Karl, "Chaotic Maps as Parsimonious Bit Error Models of Wireless Channels," *Proc. IEEE INFOCOM*, vol. 1, Apr. 2003, pp. 513-523.

[27] A. Konrad, B.Y. Zhao, A.D. Joseph, and R. Ludwig, "A Markov-Based Channel Model Algorithm for Wireless Networks," *Wireless Networks*, vol. 9, 2003, pp. 189-199.

[28] J. McDougall, J. Joseph, Y. Yi, and S. Miller, "An Improved Channel Model for Mobile and Ad Hoc Network Simulations," *Proc. Int. Conf. on Comm., Internet, and Information Technology* (*CIIT*), St. Thomas, Virgin Islands, USA, Nov. 2004.

[29] J. McDougall, Y. Yi, and S. Miller, "A Statistical Approach to Developing Channel Models for Network Simulations," *IEEE Wireless Comm. and Networking Conf. (WCNC)*, vol. 3, Mar. 2004, pp. 1660-1665.

[30] J. Poikonen, "Half-Normal Run Length Packet Channel Models Applied in DVB-H Simulations," *Proc. 3rd IEEE Int. Symp. on Wireless Comm. Systems (ISWCS)*, Valencia, Spain, Sept. 2006.

[31] M. Azimi, P. Nasiopoulos, and R.K. Ward, "Offline and Online Identification of Hidden Semi-Markov Models," *IEEE Trans. Signal Processing*, vol. 53, no. 8, Aug. 2005.

[32] C.X. Wang, W. Xu, "Packet-Level Error Models for Digital Wireless Channels," *Proc. IEEE Int. Conf. Comm.* (*ICC*), vol. 4, May 2005, pp. 2184-2189.

[33] W. Karner, O. Nemethova, and M. Rupp, "Link Error Prediction in Wireless Communication Systems with Quality Based Power Control," *Proc. IEEE Int. Conf. Comm. (ICC)*, Glasgow, Scotland, June 2007.

[34] W. Karner and M. Rupp, "Measurement-Based Analysis and Modelling of UMTS DCH Error Characteristics for Static Scenarios," *Proc. 8th Int. Symp. DSP and Comm. Systems (DSPCS)*, Sunshine Coast, Australia, Dec. 2005.

[35] W. Karner, P. Svoboda, and M. Rupp, "A UMTS DL DCH Error Model Based on Measurements in Live Networks," *Proc. 12th Int. Conf. Telecomm. (ICT)*, Capetown, South Africa, May 2005.

[36] W. Karner, O. Nemethova, P. Svoboda, and M. Rupp, "Link Error

Prediction Based Cross-Layer Scheduling for Video Streaming over UMTS," *Proc. 15th IST Mobile & Wireless Comm. Summit*, Myconos, Greece, June 2006.

[37] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003, pp. 560-576.

[38] O. Nemethova, W. Karner, A. Al-Moghrabi, and M. Rupp, "Cross-Layer Error Detection for H.264 Video over UMTS," *Proc. Wireless Personal Multimedia Comm. (WPMC)*, Aalborg, Denmark, Sept. 2005.

[39] O. Nemethova, W. Karner, and M. Rupp, "Error Prediction Based Redundancy Control for Robust Transmission of Video over Wireless Links," *Proc. IEEE Int'l Conf. Comm. (ICC)*, Glasgow, UK, June 2007.

[40] O. Nemethova, W. Karner, C. Weidmann, and M. Rupp, "Distortion-Minimizing Network-Aware Scheduling for UMTS Video Streaming," invited paper at *EUSIPCO*, Poznan, Poland, Sept. 2007.

[41] 3GPP TS 25.301, "Radio Interface Protocol Architecture," v.6.4.0, Sept. 2005.

[42] 3GPP TS 25.322, "Radio Link Control (RLC) Protocol Specification," v.6.9.0, Oct. 2006.

[43] 3GPP TR 25.993, "Typical Examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) Supported by Universal Terrestrial Radio Access (UTRA)," v.4.2.0, Sept. 2006.

[44] 3GPP TS 25.321, "Medium Access Control (MAC) Protocol Specification," v.6.10.0, Sept. 2006.

[45] 3GPP TS 25.302, "Services Provided by the Physical Layer," v.4.8.0, Sept. 2003.

[46] 3GPP TS 25.212, "Multiplexing and Channel Coding (FDD)," v.6.9.0, Oct. 2006.

[47] 3GPP TR 25.944, "Channel Coding and Multiplexing Examples," v.4.1.0, June 2001.

[48] http://www.ericsson.com/solutions/tems/

[49] ITU-T Rec. M.60, 3008; ITU-T Rec. Q.9, 0222.

[50] F. Navratil, "Fehlerkorrektur im Physical Layer des UMTS," Master's thesis (in German), Institute of Comm. and Radio-Frequency Engineering, Vienna University of Technology, Austria, Nov. 2001.

[51] 3GPP TS 25.101, "User Equipment (UE) Radio Transmission and Reception (FDD)," v.6.12.0, June 2006.

[52] A. Sampath, P.S. Kumar, and J.M. Holtzman, "On Setting Reverse Link Target SIR in a CDMA System," *Proc. 47th IEEE Vehicular Technology Conf.*, vol. 2, 1997, pp. 929-933.

[53] W. Karner, O. Nemethova, and M. Rupp, "A Measurement Based Model for UMTS DL DCH Dynamic Bearer Type Switching," *Proc. 1st IEEE Int. Symp. Wireless Pervasive Computing* (*ISWPC*), Phuket, Thailand, Jan. 2006.

[54] D.N.P. Murthy, M. Xie, and R. Jiang, *Weibull Models,* John Wiley & Sons, Ltd., 2004.

[55] A. Papoulis and P.S. Unnikrishna, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2002.

[56] M. Kac, "On the Notion of Recurrence in Discrete Stochastic Processes," *Bulletin of the American Mathematical Society,* vol. 53, 1947, pp. 1002-1010.

[57] 3GPP, TSG Services and System Aspects, "Packet Switched Conversational Multimedia Applications; Default Codecs (Rel.6)," ver.6.4.0.

[58] 3GPP, TSG Services and System Aspects, "Transparent End-to-End Packet-Switched Streaming Service (PSS); Protocols and Codecs (Rel. 6)," ver. 6.8.0.

[59] J. Tang, L. Zhang, and C.K. Siew, "An Opportunistic Scheduling Algorithm for MPEG Video Over Shared Wireless Downlink," *IEEE ICC*, 2006.

[60] A. Farrokh and V. Krishnamurthy, "Opportunistic Scheduling for Streaming Multimedia Users in High-Speed Downlink Packet Access (HSDPA)," *IEEE Trans. Multimedia*, vol. 8, no. 4, Aug. 2006.

[61] M. Welzl, "Passing Corrupt Data across Network Layers: An Overview of Recent Developments and Issues," *Eurasip Journal on Applied Signal Processing*, 2005, pp. 242-247.

[62] H. Zheng and J. Boyce, "An Improved UDP Protocol for Video Transmission Over Internet-to-Wireless Networks," *IEEE Trans. Multimedia*, vol. 3, no. 3, Sept. 2001.

[63] 3GPP TR 26.937, "Transparent End-to-End Packet Switched Streaming Service (PSS); RTP Usage Model," v.6.0.0, Mar. 2004.

[64] Y. Takishima, M. Wada, and H. Murakami, "Reversible Variable Length Codes," *IEEE Trans. Comm*., vol. 42, no. 2/3/4, 1994.

[65] A. Kiely, S. Dolinar, M. Klimesh, and A. Matache, "Error Containment in Compressed Data Using Sync Markers," *Proc. Int. Conf. on Information Theory*, June 2000.

[66] E. Cianca, F.H.P. Fitzek, M. DeSanctis, M. Bonanno, R. Prasad, and M. Ruggieri, "Improving Performance for Streaming Video Services over CDMA-Based Wireless Networks," *Proc. Int. Symp. Wireless Personal Multimedia Comm*. (WPMC 2004), Padova, Italy, Sept. 2004.

[67] S.H. Kang and A. Zakhor, "Packet Scheduling Algorithm for Wireless Video Streaming," *Proc. 12th Intl. Packetvideo Workshop (PV)*, Pittsburgh PA, USA, 2002.

[68] R.S. Tupelly, J. Zhang, and E.K.P. Chong, "Opportunistic Scheduling for Streaming Video in Wireless Networks," *Proc. Conf. Information Sciences and Systems*, Johns Hopkins University, Baltimore, MD, Mar. 2003.

[69] P. Koutsakis, "Scheduling and Call Admission Control for Burst-Error Wireless Channels," *Proc. 10th IEEE Symp. Computers and Comm. (ISCC)*, 2005.

[70] M. Kalman and B. Girod, "Rate-Distortion Optimized Video Streaming Using Conditional Packet Delay Distributions," *Proc. IEEE MMSP*, Siena, Italy, Sept. 2004.

Wolfgang Karner received his Dipl.-Ing. (MS) degree in electrical engineering with a specialization in communications and information technology from Vienna University of Technology, Austria in 2003 (with distinction). He is currently a research assistant and is working towards his Dr.techn. (PhD) degree at the Institute of Communications and Radio Frequency Engineering, Vienna University of Technology, Austria. His research interests include performance evaluation, modeling, simulation and cross-layer optimization of wireless communication systems with the focus on radio resource optimization in UMTS (DCH, HSDPA, HSUPA), and the transmission of multimedia services.

Olivia Nemethova received her BS and MS degrees from Slovak University of Technology in Bratislava in 1999 and 2001, respectively, both in informatics and telecommunications. She received her PhD in electrical engineering from Vienna University of Technology in 2007. From 2001 until 2003, she was with Siemens as a system engineer. She worked on UMTS standardization within 3GPP TSG RAN2 as a Siemens delegate. In parallel, she worked within an international property rights management team responsible for evaluation of IPRs regarding radio access networks. In 2003, she joined the Institute of Communications and Radio-Frequency Engineering at Vienna University of Technology as a research and teaching assistant. Her current research interests include error resilient transmission of multimedia over wireless networks, video processing and mobile communications.

Philipp Svoboda received his Dipl.-Ing. (MS) degree in electrical engineering in 2004 from the Vienna University of Technology, Austria. He is currently working towards his PhD at the Vienna University of Technology on traffic generation in mobile networks. He has worked in the METAWIN (Measurement and Traffic Analysis in Wireless Networks) project focusing his research on user behavior in the GPRS and UMTS core network of Mobilkom Austria AG. At the moment, he is involved in the DARWIN (Data Analysis and Reporting in Wireless Networks) project extending his work from METAWIN towards anomaly detection. In his spare time, he is investigating traffic models for online games and measuring service performance in 3G networks. His current research interests include traffic generation, statistical analysis of IP level information, and modelling of new services for cellular mobile networks.

Markus Rupp received his Dipl.-Ing. degree in 1988 at the University of Saarbruecken, Germany and his Dr.-Ing. degree in 1993 at the Technische Universitaet Darmstadt, Germany, where he worked with Eberhardt Haensler on designing new algorithms for acoustical and electrical echo compensation. From November 1993 until July 1995, he had a postdoctoral position at the University of Santa Barbara, California with Sanjit Mitra, where he worked with Ali H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 until August 2001, he was a member of the Technical Staff in the Wireless Technology Research Department of Bell-Labs at Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11 and UMTS. He is presently a full professor for Digital Signal Processing in Mobile Communications at the Technical University of Vienna. He was associate editor of IEEE Transactions on Signal Processing from 2002-2005, is currently associate editor of JASP EURASIP Journal of Applied Signal Processing, JES EURASIP Journal on Embedded Systems, Research Letters in Signal Processing, Research Letters in Communications, and is elected AdCom member of EURASIP. He authored and co-authored more than 250 papers and patents on adaptive filtering, wireless communications, and rapid prototyping, as well as automatic design methods.