DISTORTION-MINIMIZING NETWORK-AWARE SCHEDULING FOR UMTS VIDEO STREAMING

Olivia Nemethova[*], *Wolfgang Karner*[*], *Claudio Weidmann*[**] *and Markus Rupp*[*]

[*]Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, 1040 Vienna Austria, E-mail: {onemeth, wkarner, mrupp}@nt.tuwien.ac.at

[**]Telecommunications Research Center Vienna (ftw.)
Donau-City-Strasse 1, 1220 Vienna, Austria, E-mail: weidmann@ftw.at

ABSTRACT

In this paper we propose a mechanism for scheduling video streaming data of a single user over a UMTS dedicated channel. The proposed method makes use of link error prediction and schedules the more important data in the time intervals with lowest predicted error probability. The importance of the information is measured by means of the distortion that would be caused by its loss. The distortion is estimated by a novel proposed model based on the size of the lost information in bits and in picture blocks. The presented scheduling improves the PSNR at the decoder by up to 2 dB while preserving rate.

1. INTRODUCTION

The transmission of real-time video services over mobile networks is particularly challenging, especially due to the fact that limited radio resources lead to high link error probability. The redundancy has to be kept low due to system capacity limitations. Furthermore, the impact of a lost packet on distortion is not uniform across the stream.

In UMTS (Universal Mobile Telecommunication System) DCH (Dedicated Channel) downlink direction, scheduling is performed in the Radio Network Controller (RNC) at the Radio Link Control (RLC) protocol layer. The scheduling mechanism is not standardized, but rather implementation-specific. Scheduling of the packets belonging to one user's stream can considerably improve the quality of service at the receiver, especially in case of streams containing informations with different importance. In [1] and [2] (opportunistic) scheduling algorithms are presented which make use of the characteristics of the streamed video data. In [1], the more important parts of the video stream are transmitted prior to the less important ones, in order to ensure more opportunities for retransmissions in case of error; in [2], the priority-based scheduling exploits the diversity gains from the channel variations when having more than one stream. Another approach is [3], where prediction of link errors is used in connection with call admission control and scheduling algorithms in order to avoid overloading the system, thus improving the quality of the services with higher priority. Rate-distortion optimized (RDO) video streaming is proposed in [4]. Here, a model for delay and loss in the Internet is used to perform scheduling at the IP layer. In [5], we proposed a simple and efficient IP packet based mechanism that predicts the transmission intervals with low probability of errors and postpones the transmission of packets containing the intra-predicted frames to these intervals, meanwhile transmitting packets containing inter-predicted frames. In this paper we extend our work and propose scheduling for UMTS at the RLC packet layer that utilizes link error prediction and a novel reference-free distortion estimation. The underlying model considers losses at the RLC PDU (Protocol Data Unit) level, rather than at the frame/slice level; we assume that the video decoder can use the information in a NALU (Network Abstraction Layer Unit, typically encapsulated further in an RTP/UDP/IP packet) up to the first missing RLC PDU as proposed in [6].

The paper is structured as follows: Section 2 presents the model for estimating distortion caused by missing transport blocks. Sections 3 introduces the mechanism for predicting the transmission intervals with low error probability. Section 4 describes the proposed scheduling algorithm. The experimental results are evaluated in Section 5. Section 6 contains conclusions and some final remarks.

2. DISTORTION MODEL

Various distortion models have been proposed in literature so far. In [7], a semi-analytical model is proposed for H.264/AVC considering the impairments resulting from the loss or damage of the entire frame. Another approach can be seen in [8], where the visibility of packet loss is modeled depending on a set of motion and spatial video parameters. The method proposed in [9] combines the analysis-by-synthesis approach with model-based estimation. Full decoding of the damaged frame as well as the knowledge of the undamaged original are required to obtain the resulting distortion; the distortion caused by propagation is estimated by an empirical model. In [10], the distortion over the whole sequence is estimated for both compression and packet loss impairments.

Our proposed scheduling mechanism (cf. Section 4) deals with RLC PDUs. We thus need a distortion model with RLC PDU granularity, instead of the more common IP/RTP/NALU granularity. The task of designing such a model is somewhat simplified by the fact that the decoder will not be able to recover information in PDUs after the first lost PDU of a NALU. Thus, if a NALU is made up of k PDUs, only k cases need to be tested. The impairments caused by the loss of a PDU will propagate from the frame of origin to all frames that reference it. Thus, we subdivide the distortion estimation into two steps:

1. determine the *primary distortion*, i.e. the distortion within the frame caused by a lost PDU,
2. determine the *distortion propagation*, i.e. the distortion in the following frames until the end of the GOP.

Note that the distortion is estimated after error concealment,

and therefore depends on the performance of the latter. The estimation considers the position and amount of lost data, as well as the size of the corresponding picture area. 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.

2.1 Data Acquisition

In order to obtain a data set consisting of measured distortions, several assumptions about channel and encoder settings are necessary. In the considered UMTS architecture, the RLC PDU of size $q = 320$ bits is the smallest data unit allowing CRC error detection. Thus we chose to measure the position and amount of lost data in multiples of q. The size of the missing picture area is measured in multiples of macroblocks (MBs). For temporal error concealment, the zero motion method is taken, as it represents the worst case. Spatial error concealment is used only for the first frame in the sequence. The experiments were all performed using the H.264/AVC JM reference software [11] modified to support RLC segmentation. A set of five QCIF video sequences with various contents and with a frame rate reduced to 7.5 f/s was used to obtain the distortion model. All the sequences were encoded using slices of size limited to 650 bytes, $QP = 26$, GOP size of 40 frames, one reference frame only, no RDO, no rate control, no DP, CAVLC, with remaining settings corresponding to the baseline profile.

The mean square error (MSE) ε between the error-free compressed frame $\mathbf{F}_n^{\text{ori}}$ and the frame $\mathbf{F}_n^{\text{deg}}$ affected by packet loss was used as distortion measure

$$\varepsilon = \frac{1}{N_\text{r} \cdot N_\text{c}} \sum_{i=1}^{N_\text{r}} \sum_{j=1}^{N_\text{c}} \left[\mathbf{F}_n^{\text{deg}}(i,j) - \mathbf{F}_n^{\text{ori}}(i,j) \right]^2, \qquad (1)$$

where N_r and N_c denote the size of the frame in vertical and horizontal direction, respectively.

For each video sequence, each possible position of an RLC PDU error was simulated. There was always only one error per GOP in order to follow the distortion propagation caused by each error separately. Fig. 1 visualizes the data set obtained by these simulations for P frames as a $\text{PSNR} = 10 \cdot \log_{10}\left(255^2/\varepsilon\right)$. The smooth dark background corresponds to points where no data was available. The PSNR is not a totally smooth function of the number of MBs and RLC PDUs. In particular, the left part of the plot and the part with more than 50 MBs show several points of high distortion. These correspond to intra coded MBs (I-MBs).

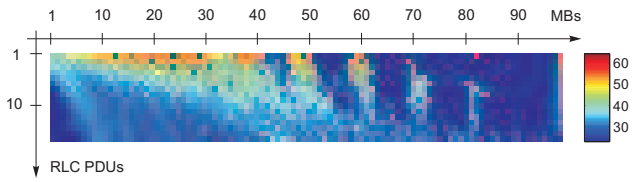

Figure 1: Primary distortion (expressed as PSNR in dB) averaged over all simulated video sequences, over the number of missing MBs and RLC PDUs.

Unfortunately, this effect on distortion cannot be suppressed or compensated, since the number of I-MBs within the lost part of a slice is not known and their frequency depends on the encoder implementation.

2.2 Empirical Distortion Modeling

In order to build a model based on the obtained data set, the estimated primary distortion $\widehat{\varepsilon}_0$ has to be represented as a function of the number of missing RLC packets (PDUs) N and the corresponding number of missing macroblocks M:

$$\widehat{\varepsilon}_0 = f(M,N). \qquad (2)$$

The task is to find an appropriate model function $f(.)$ and its best fitting parameters for the given scenario (i.e. encoder settings/error concealment assumptions).

The distortion grows monotonically with increasing number of missing RLC PDUs. We observed that this growth is better described by a rational rather than exponential function. With increasing number of missing MBs, the distortion decreases. Thus, we propose the distortion model

$$\widehat{\varepsilon}_0 = e^{a \cdot M + \frac{b}{N} + c} \qquad (3)$$

with parameters a, b, and c determined by least square fitting. For the model data set, the following values were obtained: $a = -0.041$, $b = -6.371$, and $c = 4.482$.

In order to model the error propagation in a GOP, all obtained distortion sequences $\{\varepsilon_k\}_{k=1,2\ldots}$, where k is the frame index, were normalized by the corresponding primary distortion ε_0. On average, an exponential decrease was observed:

$$\widehat{\varepsilon}_k = \widehat{\varepsilon}_0 \cdot e^{-s \cdot k}, \qquad (4)$$

where 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, 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 estimation performance of the proposed method was tested for all video sequences in the model data set, as well as for three additional ones. Table 1 summarizes the results in terms of Pearson correlation coefficient ρ for both primary distortion and error propagation model, obtained by decoding of all possible RLC PDU loss error patterns. The performance of the distortion estimation varies for different video sequences. Especially scene cuts and frames with significant

Video sequence	$\rho(\varepsilon_0)$	$\rho(\varepsilon_k)$
1	0.86	0.84
2	0.88	0.81
3	0.85	0.78
4	0.68	0.80
5	0.84	0.77
6	0.87	0.85
7	0.79	0.67
8	0.83	0.73

Table 1: Performance of the distortion estimation for video sequences in the model data set (1–5) and new sequences (6–8).

amount of I-MBs cause lower correlation. However, in the scheduling application absolute distortion values are not as important as the relation between the distortions caused by different error patterns.

3. LINK ERROR MODEL

Based on the analysis of the link error characteristics from UMTS DCH (Dedicated Channel) measurements in live networks [12], we know that there is correlation and thus memory in the link. Furthermore, in [13] the authors have shown that the impact of the UMTS DCH quality based power control — which provides error prediction properties — on the link error characteristics becomes significant especially in static scenarios.

Defining the number of subsequently received erroneous TBs (Transport Blocks) as error burst, the successive error-free TBs as error gap, and erroneous TBs which are separated by less than or equal to 12 (for the 384 kbit/s bearer) error-free TBs as error cluster (Fig. 2), we can describe the UMTS DCH link error process in the following way.

Due to the effects of jointly coded and interleaved TBs within one TTI (Transmission Timing Interval), the UMTS radio link produces error bursts which are separated by short gaps (gap length ≤ 12). These bursts and short gaps, lying within up to three successive TTIs, then form an error cluster. The power control algorithm of the UMTS DCH separates successive clusters by long gaps in order to reach the link quality target (block error rate of 1% in the considered networks) as illustrated in Fig. 2.

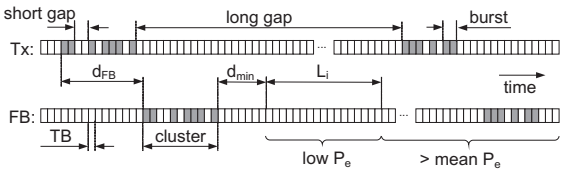

Figure 2: Schematic illustration of the link error characteristics for UMTS DCH in downlink direction.

For a prediction of future link errors, the knowledge of past transmission errors is necessary at the transmitter. That is accomplished in UMTS via the RLC AM (Radio Link Control Acknowledged Mode). Of course, the transmitter receives the RLC AM error feedback information from the receiver after a certain feedback delay d_{FB} which is in the order of three TTIs ($=30$ ms for the measured 384 kbit/s bearer). Thus, error prediction within one error cluster is not possible but the location of the error clusters or in other words the length of the long gaps can be estimated.

We have shown in [13] that the lengths of the long gaps are not geometrically distributed, thus providing memory. Furthermore, it was argued that the long gaps follow a Weibull distribution with shape parameter $a > 1$, meaning the distribution is convex from zero up to the inflection point (mode). This leads to a region with a very small probability of having gaps between the maximum of the short gap lengths ($\cong 12$) and about 300 TBs [13]. Therefore, if within a time interval $d_{\min}$ (larger than the maximum of the short gaps) after the last error there are no error reports, we can assume to be out of an error cluster, within a long error gap and thus at the beginning of an interval with a very low probability for the occurrence of another error cluster, due to the convex distribution of the long gaps.

This capability of predicting the start of intervals with very low error probability has already been used by the authors for cross-layer optimized video streaming (presented in [5]), where a significant gain in quality of the streamed video was obtained. For the scheduling mechanism presented in this work it is not enough to detect the beginning of the intervals with very low error probability at $d_{\min}$ after the last error report, but it is also necessary to estimate the length L_{i} of these intervals.

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. 3 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 with Weibull-distributed gap lengths. We can see from this curve that the error probability is very small right after an error event, staying below the total mean error probability until approximately 2500 TBs. In practice, we may estimate L_{i} via the intersection of this conditional mean curve with the unconditional mean P_{e}.

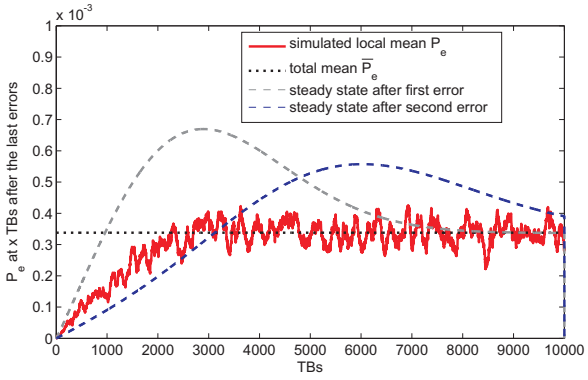

Figure 3: P_{e} for Weibull distributed long gaps (Weibull shape parameter $a = 2.176$, scale parameter $b = 3350$).

To obtain an analytic expression for L_{i} we first compute the steady state error probability $P_{\mathrm{e,ss}}$. Let L be the random variable measuring the time (distance) between two error events. Assuming that the error process is stationary, Kac's lemma [14] implies that

$$P_{\mathrm{e,ss}} = \frac{1}{\mathrm{E}\{L\}}, \tag{5}$$

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

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 steady state error probability $P_{\mathrm{e,ss}}$. Therefore, we approximate the probability $P_{\mathrm{e}}(l)$ of having an error at the lth TB after an error event assuming that the steady state is reached before the third error after the initial error event:

$$\begin{aligned} P_{\mathrm{e}}(l) \quad &\cong P(L_1 = l) + P(L_1 + L_2 = l) \\ &\quad + P_{\mathrm{e,ss}} \cdot P(L_1 + L_2 < l). \end{aligned} \tag{6}$$

Here, L_1 and L_2 denote the gap lengths before the first and the second error, respectively. The result (blue dashed curve in

Fig. 3) 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).

4. SCHEDULING MECHANISM

The proposed scheduling mechanism using link error prediction and distortion estimation is illustrated in Fig. 4.

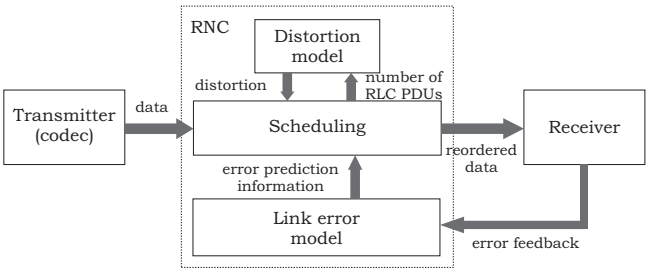

Figure 4: Functional scheme of the proposed scheduling mechanism.

The performance of the scheduler can be controlled by the scheduling buffer length $L_\mathrm{b} \geq L_\mathrm{i}$ and the maximum acceptable delay 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 provides a guiding value for determining the size of d_max.

Based on the RLC feedback from the receiver, the link error model identifies the transmission intervals with low error probability and delivers this information to the scheduler. The distortion model provides an estimation of the *cumulative distortion* $\widehat{\varepsilon}_\Sigma$ that would be caused by the loss of a particular RLC PDU. This is calculated for every RLC PDU in the scheduling buffer as

$$\widehat{\varepsilon}_\Sigma = \sum_{k=n}^{m} \widehat{\varepsilon}_{k-n}, \qquad (7)$$

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.

At each step, if an RLC PDU remained in the scheduling buffer longer than 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 $\widehat{\varepsilon}_\Sigma$ is chosen from the scheduling buffer. Outside the interval L_i, the RLC PDU with the lowest $\widehat{\varepsilon}_\Sigma$ is scheduled. If several RLC PDUs have the same (lowest or highest) value of $\widehat{\varepsilon}_\Sigma$, the oldest among them is scheduled.

5. EXPERIMENTAL RESULTS

We implemented all presented methods using the Joint Model reference software [11]. 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 the same settings as those used for the distortion model (cf. Section 2.1). Its transmission was simulated 1500 times (representing $\sim$3h of video) for each of the considered methods to obtain sufficient statistics. The CRC of the RLC PDUs ($q = 320$ bits) was used to detect errors. Decoding of a video slice was performed up to the first erroneous RLC PDU within the corresponding packet, as proposed in [6]. The following methods were compared:

- **Without scheduling**: decoding using unchanged link error traces obtained from measurements in a live UMTS network,
- **Scheduling limit**: achievable limit under assumptions that the position of errors is known and both d_max and L_b are larger than one entire video sequence,
- **I scheduling**: prioritized scheduling of the I frames as proposed in [5], but assuming the usage of the correctly received data from NALU up to the first missing RLC PDU,
- **RLC scheduling**: configurations of the proposed method for different d_max and L_b.

In this paper we present simulations for combinations $d_\mathrm{max} = 8400$ TBs, $L_\mathrm{b} = 3600$ TBs and $d_\mathrm{max} = 3600$ TBs, $L_\mathrm{b} = 1200$ TBs (denoted further as 8400/3600 and 3600/1200), which corresponds to the maximum delay of 7 and 3 seconds, respectively.

Fig. 5 shows the empirical CDF of frame PSNR for the presented methods. It illustrates well the difference between the I scheduling and the RLC scheduling method. All CDFs approach one at the PSNR $= \infty$ (error-free frames). The I

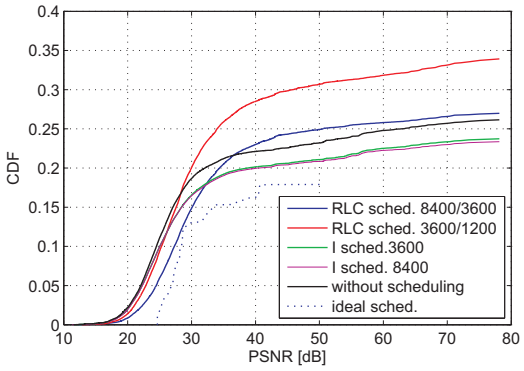

Figure 5: CDF of frame PSNR achieved by presented scheduling methods.

scheduling method globally reduces the number of erroneous frames since it moves the errors from the first frame in GOP to the later ones, making the error propagation chain shorter. The proposed RLC scheduling reduces the number of frames with higher distortion and consequently increases the number of frames with lower distortion. It can even happen that the total number of frames containing an error (even small) increases after applying RLC scheduling. This is caused by burst errors possibly affecting multiple frames, since we do not consider the combinations of errors in time; i.e. if an error occurs followed by another error in the next frame at the same position, then the cumulative distortion is smaller or equal to the sum of the two distortions. However, as we do not have an information about the motion between the two frames, we cannot reliably determine the additional distortion caused by another error. Moreover, considering all possible combinations of two and more errors would increase the complexity of the scheduler considerably.

Fig. 6 depicts the PSNR over the frames of the tested "videoclip" sequence. Again, the difference between the I scheduling and RLC scheduling can be seen. The PSNR at the end of the sequence decreases for the RLC scheduling. This effect is caused by the high predicted distortions for a loss of the first frame. In our simulation we encoded the video sequence 1500 times, thus the first frame follows after the last one again. Since the first frame is concealed only by spatial error concealment, the corresponding distortion caused by a loss of its parts is higher. In Table 2, the

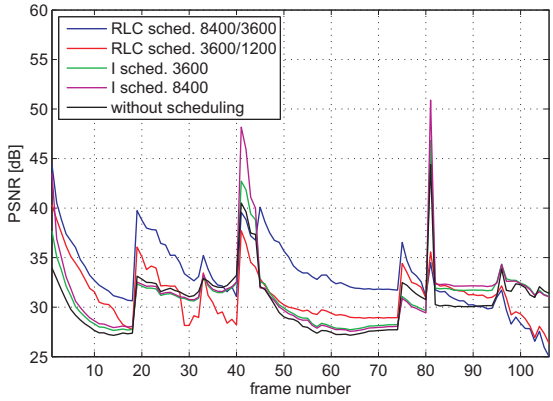

Figure 6: PSNR in time: comparison of presented scheduling methods.

proportion of error-free frames is summarized together with the mean distortion for all presented methods. The I scheduling is not as sensible to smaller $d_{\max}$ as the RLC scheduling.

Method	Error-free frames [%]	Mean PSNR [dB]
Without scheduling	73.86	29.29
I sched. ($d_{\max} = 3600$ TBs)	76.27	30.26
I sched. ($d_{\max} = 8400$ TBs)	76.63	30.32
RLC sched. (3600/1200)	66.07	30.44
RLC sched. (8400/3600)	73.01	32.03
Scheduling limit	82.09	35.53

Table 2: Number of erroneous frames and mean PSNR for the presented scheduling methods.

6. CONCLUSIONS

In this paper we investigate an RLC layer based scheduling for UMTS that makes use of both distortion estimation and link error prediction. We propose a distortion model that allows for estimation of distortion caused by loss of RLC PDUs without decoding the video stream. Based on the prediction of intervals with low error probability, the scheduler prioritizes the RLC PDUs whose loss would lead to highest distortion. We used error traces obtained from live UMTS network measurements for our experiments. The proposed scheduling method gains in average 2 dB compared to common (in-order) scheduling. Since the loss of packets containing I frames usually causes the highest distortions, we compared the proposed scheduler (based on the distortion model) with a scheduler that prioritizes the packets belonging to I frames. The gain achieved by using the distortion model is then still about 1 dB. These gains are significant from the point of view of user perception.

7. ACKNOWLEDGMENT

The authors 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.

This research has been conducted in part within the NEWCOM Network of Excellence in Wireless Communications funded through the EC 6th Framework Program.

Our thanks also go to Carlos Teijeiro Castela who collected the data for distortion modeling in his bachelor thesis.

REFERENCES

[1] S.H. Kang, A. Zakhor, "Packet Scheduling Algorithm for Wireless Video Streaming," Proc. of 12th Packetvideo Workshop 2002, Pittsburgh PA, USA, 2002.

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

[3] P. Koutsakis, "Scheduling and Call Admission Control for Burst-Error Wireless Channels," 10th IEEE Symp. on Computers and Communications (ISCC), 2005.

[4] M. Kalman, B. Girod, "Rate-Distortion Optimized Video Streaming Using Conditional Packet Delay Distributions," IEEE MMSP, Siena, Italy, Sep. 2004.

[5] W. Karner, O. Nemethova, M. Rupp, "Link Error Prediction Based Cross-Layer Scheduling for Video Streaming over UMTS," 15th IST Mobile & Wireless Communications Summit, Myconos, Greece, Jun. 2006.

[6] O. Nemethova, W. Karner, A. Al-Moghrabi, M. Rupp, "Cross-Layer Error Detection for H.264 Video over UMTS," Proc. of Wireless Personal Mobile Communications, Int. Wireless Summit, Aalborg, Denmark, Sep. 2005.

[7] C. Bergeron, C. Lamy-Bergot, "Modelling H.264/AVC sensitivity for error protection in wireless transmissions," Proc. of IEEE MMSP, Oct. 2006.

[8] S. Kanumuri, P.C. Cosman, A.R. Reibman, V.A. Vaysham-payan, "Modeling Packet-Loss Visibility in MPEG-2 Video," IEEE Trans. on Multimedia, vol. 8, no. 2, Apr. 2006.

[9] F. De Vito, D. Quaglia, J.C. De Martin, "Model-based Distortion Estimation For Perceptual Classification of Video Packets," Proc. of IEEE MMSP, Sep. 2004.

[10] K. Stuhlmüller, N. Färber, M. Link, B. Girod, "Analysis of Video Transmission over Lossy Channels," IEEE Jour. on Selected Areas in Communications, vol. 18, no. 6, Jun. 2000.

[11] H.264/AVC Software Coordination, "Joint Model Software", ver. 11.0, available in `http://iphome.hhi.de/suehring/tml/`.

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

[13] W. Karner, O. Nemethova, M. Rupp, "Link Error Prediction in Wireless Communication Systems with Quality Based Power Control," Proc. of IEEE Int. Conf. on Communications (ICC), Glasgow, UK, Jun. 2007.

[14] M. Kac, "On the Notion of Recurrence in Discrete Stochastic Processes," Bulletin of the American Mathematical Society 53 (1947): pp. 1002–1010.