Efficient Voice Communication in Wireless Packet Networks

Frank Mertz and Peter Vary

Institute of Communication Systems and Data Processing (ind), RWTH Aachen University, 52056 Aachen, Germany
E-Mail: {mertz,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

We present a methodology to optimize the system parameterization for speech and audio transmission in heterogeneous packet networks with wireless access. To this end, we determine the theoretical performance of forward error correction (FEC) schemes on packet level, i.e., the resulting frame loss distribution at the receiver after error correction. The calculation utilizes a statistical model for packet loss distribution which is adaptable for different packet transmission time intervals and packet sizes. The optimal parameterization of the transmission scheme, i.e., the encoding rate, frame length, and amount of redundancy per packet, are then found by optimizing the resulting conversational quality. We utilize the ITU-T E-model which provides a quality prediction in dependence on codec, end-to-end delay, and frame loss distribution. The procedure will be illustrated for realistic transmission scenarios of wireless packet network access technologies.

1 Introduction

Communication networks are developing towards all-IP networks, providing packet transmission based on the IP protocol stack. Such networks allow the flexible realization of various services on the same transmission platform, e.g., speech conversation, audio/video streaming, as well as Internet and email applications. At the same time, a fixed-mobile convergence can be observed, with mobile networks offering the necessary data rates to realize these services also on mobile devices. All such communication networks together can then be seen as a single large network consisting of interconnected high speed core networks and various types of wired and wireless access technologies (DSL, UMTS, WLAN, etc.).

The preferred modus of operation in such a heterogeneous packet network is a transparent end-to-end IP transmission of voice calls and other media streams, i.e., the user terminals negotiate the source codec to use in the initial call setup phase and no transcoding is required. The stream of media packets then traverses different types of transmission channels which may induce varying packet delays and packet losses, e.g., due to bit errors on wireless links. At the receiver, lost frames will be replaced by a packet loss concealment (PLC) algorithm which estimates the missing signal segment, e.g., with extra-/interpolation techniques. Additionally, if the end-to-end transmission channel provides enough capacity, the impact of packet losses at the receiver can be reduced by transmitting redundancy on packet level for the purpose of error correction or assisting the PLC [10, 11, 12]. The redundancy can either be media dependent, e.g., repetition of the most important codec parameters, repeated encoding of a frame at a lower rate, or it can be media independent in the form of forward error correction (FEC) on frame level, i.e., repeated frames, XOR combinations of frames, or FEC frames derived by standard block codes.

In this paper, we will present methodologies for an optimal parameterization of media independent FEC schemes for speech conversation over heterogeneous networks with wireless access. Based on a flexible model of the packet loss channel, introduced in Sec. 2, we will show in Sec. 3 how the resulting frame loss rate at the receiver can be theoretically determined in dependence on the system parameters, i.e., frame length per packet and the parameters of the packet level FEC scheme. These theoretical evaluations will finally be utilized in Sec. 4 to find an optimal system parameterization for achieving the best possible voice conversation quality in different exemplary network scenarios.

2 Modeling Packet Channels

Packet transmission channels usually experience a combination of single packet losses and loss bursts of several consecutive packets. A widely used statistical model for packet loss is the Gilbert-Elliott model, originally utilized to model burst-noise binary channels. We will give a short review of the model and an approach to adapt it to different packet transmission time intervals (TTI), which usually correspond to the codec frame length. Subsequently, we will explain how to adapt the model to describe the loss behavior for packet streams of different packet sizes. This adaptation is necessary for describing wireless channels with bit errors where the loss probability, i.e., the probability of containing residual bit errors, depends on the length of the packet. Both adaptations together will allow us to compare different transmission configurations (i.e., codec frame lengths and data rates, as well as redundancy rates) based on a single base channel model.

2.1 Packet Loss Model

The Gilbert-Elliott model, introduced by Gilbert in [4] and generalized by Elliott in [2], is a 2-state Markov model as defined in Fig. 1. The two states differ in their loss probability, state G having a low loss probability $P_{e,G}$ and state B having a higher loss probability $P_{e,B}$, with $0 \leq P_{e,G} \ll P_{e,B} \leq 1$. For each packet of the transmission, a state transition is made according to the given transition probabilities P_{ij} ($i, j \in \{G,B\}$). The current state determines the probability for a packet loss. We use this generalized model instead of the often used simplification ($P_{e,G}=0$, $P_{e,B}=1$) because it has a better ability to model a large variety of loss distributions. The mean packet loss rate P_{pl} can then be computed as

$$P_{pl} = \frac{P_{BG}}{P_{GB}+P_{BG}} \cdot P_{e,G} + \frac{P_{GB}}{P_{GB}+P_{BG}} \cdot P_{e,B}, \qquad (1)$$

with the fractions denoting the probabilities to be in state G or state B, respectively. For the calculations of FEC capabilities in Sec. 3, the probability of m losses in n consecutive packets is calculated as [2]

$$P(m,n) = \frac{P_{BG}}{P_{GB}+P_{BG}} \cdot G(m,n) + \frac{P_{GB}}{P_{GB}+P_{BG}} \cdot B(m,n), \qquad (2)$$

with the conditional probabilities of m losses in n packets when the channel is first in state G or B, $G(m,n)$ and $B(m,n)$, respectively. These can be derived by recursive calculation:

$$\begin{aligned} G(m,n) = {} & (1-P_{e,G})\,(P_{GG}\cdot G(m,n-1) + P_{GB}\cdot B(m,n-1)) \\ & + P_{e,G}\,(P_{GG}\cdot G(m-1,n-1) + P_{GB}\cdot B(m-1,n-1)), \end{aligned} \qquad (3)$$

with $G(0,1) = 1 - P_{e,G}$, $G(1,1) = P_{e,G}$, $B(0,1) = 1 - P_{e,B}$, and $B(1,1) = P_{e,B}$. $B(m,n)$ is calculated accordingly.

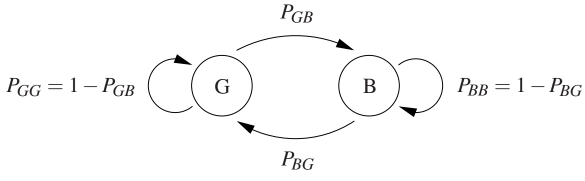

Figure 1: Gilbert-Elliott Model: 2-state Markov model with transition probabilities P_{ij} and different loss probabilities $P_{e,i}$ in state G and B ($i, j \in \{G,B\}$) with $0 \leq P_{e,G} \ll P_{e,B} \leq 1$.

Furthermore, a constraint of being in a certain state at the packet following the considered n packets can be included in this calculation. The respective conditional probabilities will be denoted with subscripts G and B and are calculated in the same way as above, but with different initial terms: $G_G(i,1) = G(i,1) \cdot P_{GG}$ and $B_G(i,1) = B(i,1) \cdot P_{BG}$ for being in state G at the following packet, and $G_B(i,1) = G(i,1) \cdot P_{GB}$ and $B_B(i,1) = B(i,1) \cdot P_{BB}$ for being in state B; $i \in \{G,B\}$.

2.2 Adaptation of the Channel Model

The above channel model does not necessarily have to refer to actual IP packets with protocol headers and payload. The model can rather refer to transmitted data packages in general and describe whether these contain transmission errors or not. The length of these data packages may be only a fraction of the size of a real IP packet. Such a base model can then be adapted to reflect the transmission of packet streams with different time bases, i.e., packet transmission time intervals, and different packet sizes. For this, however, the base model needs to be of a sufficient resolution, i.e., the length of the data packages and the TTI need to be small enough.

2.2.1 Adaptation for Multiples of the TTI

We first assume an increase of the transmission time interval from the original value of the channel model, T'_{TTI}, to an integer multiple, i.e. an increase by the factor $k_t = T_{\mathrm{TTI}}/T'_{\mathrm{TTI}}$. At first assuming no increase in the packet size, the state transition probabilities of the new effective channel model, denoted with the superscript (k_t), can be derived from the original values [2, 9]:

$$P_{GB}^{(k_t)} = \frac{P_{GB}}{P_{GB}+P_{BG}}\{1-(P_{GG}-P_{BG})^{k_t}\}, \tag{4a}$$

$$P_{BG}^{(k_t)} = \frac{P_{BG}}{P_{GB}+P_{BG}}\{1-(P_{GG}-P_{BG})^{k_t}\}. \tag{4b}$$

The error probabilities of the two states are unaffected by the change of the TTI: $P_{e,G}^{(k_t)} = P_{e,G}$, $P_{e,B}^{(k_t)} = P_{e,B}$; and the overall loss rate P_{pl} remains the same. The distribution of loss lengths, however, is changed due to the modified state transition probabilities.

2.2.2 Adaptation for Arbitrary Packet Sizes

We determined an approach for adapting the channel model to account for different packet sizes, which directly correspond to the packet transmission times on channels with constant transmission rates. Such an adaptation is possible if the following constraints are met: First, the new packet transmission time τ_p has to be approximately an integer multiple of the model's base transmission time τ'_p: $k_p \approx \tau_p/\tau'_p$. Second, the base transmission time τ'_p must be equal to the base transmission time interval T'_{TTI}, i.e., the base model must describe the channel at 100 % utilization. An increase of the packet transmission time by the factor k_p may result in a higher loss rate on wireless channels with bit errors. This is determined by the probability to loose any number and combination of consecutive k_p data packages of the original transmission time τ'_p. However, since the loss probabilities of the sequential packets depend on the state transitions of this sequence, the state the channel will be in at the start of the following transmission time interval has to be taken into account. This leads to transition dependent loss probabilities instead of state dependent loss probabilities as considered in the standard Gilbert-Elliott model. The error probabilities of the adapted model depend on both factors k_t and k_p and will therefore be denoted by the superscript (k_t,k_p):

$$P_{e,GG}^{(k_t,k_p)} = \left(\sum_{i=1}^{k_p} G_G(i,k_p)\, P_{GG}^{(k_t-k_p)} + \sum_{i=1}^{k_p} G_B(i,k_p)\, P_{BG}^{(k_t-k_p)}\right) \cdot \frac{1}{P_{GG}^{(k_t)}} \tag{5a}$$

$$P_{e,GB}^{(k_t,k_p)} = \left(\sum_{i=1}^{k_p} G_G(i,k_p)\, P_{GB}^{(k_t-k_p)} + \sum_{i=1}^{k_p} G_B(i,k_p)\, P_{BB}^{(k_t-k_p)}\right) \cdot \frac{1}{P_{GB}^{(k_t)}} \tag{5b}$$

$$P_{e,BG}^{(k_t,k_p)} = \left(\sum_{i=1}^{k_p} B_G(i,k_p)\, P_{GG}^{(k_t-k_p)} + \sum_{i=1}^{k_p} B_B(i,k_p)\, P_{BG}^{(k_t-k_p)}\right) \cdot \frac{1}{P_{BG}^{(k_t)}} \tag{5c}$$

$$P_{e,BB}^{(k_t,k_p)} = \left(\sum_{i=1}^{k_p} B_G(i,k_p)\, P_{GB}^{(k_t-k_p)} + \sum_{i=1}^{k_p} B_B(i,k_p)\, P_{BB}^{(k_t-k_p)}\right) \cdot \frac{1}{P_{BB}^{(k_t)}} \tag{5d}$$

The transition probabilities $P_{ij}^{(k)}$ are calculated according to (4a-b) with the given increase factor. The probabilities of i losses in k_p packets (G_G, G_B, ...) are calculated as explained in Sec. 2.1.

3 FEC on Packet Level

We consider a general (i.e., media independent) scheme for end-to-end error protection by forward error correction (FEC) which is suitable for application in packet based audio transmission.

First, an *(n,k) Reed-Solomon (RS) block code* is considered at packet level. It is applied in parallel for every byte position of k successive speech frames such that $n-k$ parity frames are calculated. These FEC frames are then piggybacked one by one to the following packets containing the following original frames. At the receiver, the positions of errors, i.e. lost packets, can be derived from the sequence numbers in the RTP headers of received packets. Hence, the receiver knows which original and which FEC frames are lost. Then, an erasure correction of up to $n-k$ losses can be performed in a group of n frames. Based on the channel model introduced in Sec. 2, probabilities can be derived which describe the resulting loss distribution after erasure decoding, i.e., frame regeneration, at the receiver. The resulting frame loss rate after correction, P_{fl}, can be calculated, in a similar way as shown in [3] for a simplified channel model, as

$$P_{fl} = \sum_{i=n-k+1}^{n} \sum_{j=i-n+k}^{\min(k,i)} \frac{j}{k} \cdot \big(P_G(j,k)\, G(i-j,n-k) + P_B(j,k)\, B(i-j,n-k)\big). \tag{6}$$

Here, the index of the first sum, i, describes the number of total losses in a block of n frames such that an erasure correction is not possible, i.e., at least $n-k+1$. The index of the second sum, j, is the number of losses in the k original media frames which then contribute to the loss rate. Finally, $i-j$ results to the remaining losses in the $n-k$ parity frames.

For a simple *repetition of frames* with p repetitions per frame, i.e., for $k=1$ and $n=p+1$, (6) simplifies to

$$P_{fl} = P(p+1,p+1). \tag{7}$$

Finally, we consider a FEC scheme of transmitting *XOR combinations* of frames as redundant information piggybacked to following packets. In particular, the XOR combination of the two preceding speech frames are piggybacked as additional FEC information to each packet. In doing so, the information of each frame will be transmitted three times, once as original frame and twice as XOR combination with other frames. In case of packet loss, the lost original frame might then be reconstructed with help of the FEC frames from following packets. Hence, a single frame cannot be recovered if three successive packets are lost. If two successive packets are lost, the next is received, followed by another packet loss, two successive frames cannot be reconstructed. This is reflected in the following residual frame loss rate:

$$\begin{aligned} P_{fl} &= P(3,3) \\ &+ 2\cdot P_G(2,2)\cdot(G_G(0,1)\cdot G(1,1) + G_B(0,1)\cdot B(1,1)) \\ &+ 2\cdot P_B(2,2)\cdot(B_G(0,1)\cdot G(1,1) + B_B(0,1)\cdot B(1,1)). \end{aligned} \tag{8}$$

The resulting mean burst lengths are calculated with the probability of a burst start P_{burst} as $\overline{b} = P_{fl}/P_{\mathrm{burst}}$. The derivations shall not be explained here in detail.

At the expense of a higher delay, all of the above schemes can be made more robust against burst errors by delaying the transmission of the FEC frames by one or two packets, thereby introducing a greater distance to the original frames. The distance in packets between the original frame and the FEC frames will be denoted by dp in the following, with $dp=1$ if the FEC frame is transmitted in the packet directly following the original frame.

AMR mode	PESQ-MOS	MOS-LQO	I_e
12.2 kbit/s	3.893	4.031	5
10.2 kbit/s	3.798	3.921	9
7.95 kbit/s	3.661	3.754	15
7.4 kbit/s	3.640	3.724	16
6.7 kbit/s	3.557	3.616	20
5.9 kbit/s	3.473	3.503	23
5.15 kbit/s	3.370	3.360	27
4.75 kbit/s	3.307	3.270	29

Table 1: Equipment impairment factors I_e (rounded) for the AMR modes obtained with PESQ acc. to the methodology in [7].

4 System Optimization

4.1 Prediction of Quality with the E-Model

Optimization criterion for the parameterization of speech conversation services has to be the perceived quality. For assessing different types of quality impairments, including frame losses and delay, the ITU has standardized the E-model [5], a non-intrusive computational model for speech quality prediction.

The E-model output is the so-called "Rating Factor" R, ranging from 0 (worst) to 100 (best). A value of about 70 describes so-called toll quality. The rating factor is calculated by adding the individual impairment factors and subtracting them from the maximum value 100. Assuming some basic default impairments as defined in [5], the calculation of the rating factor becomes

$$R = 93.2 - I_d - I_{e,eff}\,, \tag{9}$$

with the delay impairment factor I_d, depending on the end-to-end delay, and the equipment impairment factor $I_{e,eff}$, describing codec distortion and frame losses.

The end-to-end delay d describes the total one-way mouth-to-ear delay, including en-/decoding, packetization, and network delays, as well as the de-jitter delay (receiver buffer). Below a total delay of 150 ms there is no quality degradation, above 200 ms the quality of a speech conversation starts to decrease considerably because the interactivity gets affected.

The frame loss dependent equipment impairment factor $I_{e,eff}$ includes the equipment impairment factor for codec distortions I_e and is defined in [5] as

$$I_{e,eff} = I_e + (95 - I_e) \cdot \frac{100 \cdot P_{fl}}{\frac{100 \cdot P_{fl}}{BurstR} + B_{pl}}, \tag{10}$$

with the frame loss rate P_{fl}, a codec specific packet loss robustness factor B_{pl}, and the burst ratio $BurstR$. $BurstR$ is defined as the quotient of the average burst length (number of successive frame losses) on the channel and the theoretical average burst length under random, i.e., independent losses of the same rate.

Provisional planning values for equipment impairment and packet loss robustness factors of different speech codecs are given in [6]. The studies in the following sections require the equipment impairment factors I_e of several AMR encoding modes. However, up to now only the value for the Enhanced Full-Rate speech codec, the highest AMR codec mode, has been standardized. We therefore adopted the proposed methodology from [7] to determine the equipment impairment factors for the other modes with the objective speech quality measure PESQ [8]. With this procedure, we obtained the results listed in Tab. 1. The equipment impairment factor for the 12.2 kbit/s AMR mode exactly resulted in the value already defined in [6].

Also for the burst sensitivity factor B_{pl}, only the value for the 12.2 kbit/s AMR mode is given in [6]. This factor is codec dependent, i.e., it depends on interframe dependencies and the implemented packet loss concealment scheme. We assume that the standard concealment of the AMR codec is used. In lack of standardized values, we assume the same factor $B_{pl} = 10$ for all modes of the AMR codec until more precise values are standardized. We believe this is justifiable because the AMR codec modes have a similar general structure, use the same frame loss concealment, and therefore also show similar effects of error propagation.

P_{pl}	$\bar{b}$	P_{GB}	P_{BG}	$P_{e,G}$	$P_{e,B}$
1.30 %	1.18	0.00559	0.74416	0.00559	0.99999
2.72 %	1.28	0.00729	0.50941	0.01477	0.89371
4.82 %	1.47	0.02286	0.59729	0.01174	0.99993
13.2 %	1.84	0.07797	0.53291	0.00501	1.0

Table 2: Packet loss rates P_{pl}, mean burst lengths $\bar{b}$ and corresponding channel model parameters determined from simulations of a 17.6 kbit/s dedicated packet-switched UMTS channel.

4.2 VoIP on UMTS Packet Channels

In the first scenario we consider a Voice over IP (VoIP) transmission using the AMR codec on a UMTS packet channel. For the fixed frame length of the AMR codec, we will show how to optimize the quality by controlling the encoding rate and the amount of redundancy to add with packet level FEC schemes.

4.2.1 UMTS Channel Model

A dedicated UMTS downlink channel (DTCH) as defined in [1], Sec. 7.1.123, has been simulated on bit level with the Synopsys System Studio Software. The channel had a maximum data rate for the IP packet stream of 17.6 kbit/s, with a transport block size of 360 bit (including 8 bit RLC header) and a TTI of 20 ms. The channel assumes the use of header compression (ROHC), i.e., the IP/UDP/RTP headers are reduced to a 3 byte ROHC header. Channel coding has been performed using a 1/3 Turbo code. If a 16 bit CRC detected residual bit errors after channel decoding, the transport block and the contained IP packet were discarded. Packet loss sequences have been generated for different channel qualities with loss rates of 1-13%. From these loss sequences, the parameters of the Gilbert-Elliott model were determined using the Baum-Welch algorithm [13], an algorithm for maximum likelihood estimation of hidden Markov model parameters, i.e., here the transition and loss probabilities of the Gilbert-Elliott model. The determined parameters, as well as the resulting packet loss rates, P_{pl}, and mean burst lengths, $\bar{b}$, are given in Tab. 2.

4.2.2 Trade-off between Base Quality and Robustness

Assuming a dedicated channel with a fixed transmission data rate, the available data rate may be either fully used for the encoded speech signal, or a lower encoding rate may be chosen which then leaves room for enhancing error robustness by transmission of redundancy. The following studies have therefore been using the 12.2 kbit/s AMR mode for transmission without redundancy, the 6.7 kbit/s mode for all FEC schemes with code rate 1/2, and the 4.75 kbit/s mode for schemes with rode rate 1/3. Slight differences in the resulting packet lengths are negligible.

The E-model rating factor has been calculated as described in Sec. 4.1 based on the used AMR mode and the predicted residual frame loss rate and mean burst length for the considered FEC scheme, which have been determined according to Sec. 3. The results are shown in Fig. 2 for different UMTS channel models, comparing the residual frame loss rate after correction by the respective FEC scheme and the resulting E-model rating factor. A network transmission delay of 100 ms has been assumed.

At 0% packet loss rate, the curves converge to a value determined by the equipment impairment factor of the respective AMR mode and the impairment factor for the delay. The latter includes the additional delay required for the respective FEC scheme. For increasing packet loss rates, the R value of the 12.2 kbit/s AMR mode without FEC decreases quickly since none of the lost frames can be recovered. At low loss rates, however, it is still better than the lower encoding modes with FEC protection because of its higher base quality. At increasing packet loss rates, a simple *repetition* of one frame transmitted in the following packet (distance $dp = 1$) can already reduce the resulting frame loss rate at the receiver and thereby lead to a slower decrease of quality in spite of the lower AMR mode (MR67: REP $p = 1$, $dp = 1$). Since the considered channel does not produce completely independent losses, but burst losses, the resulting loss

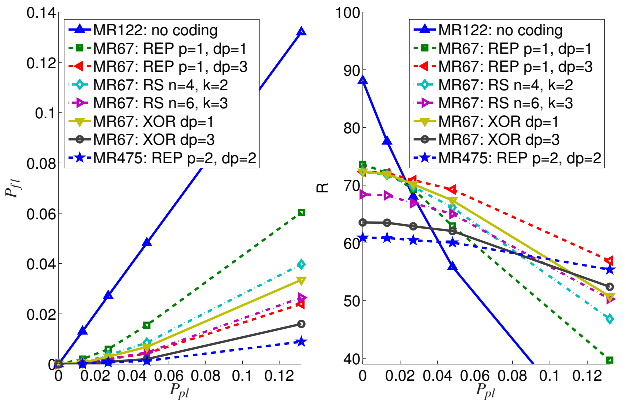

Figure 2: Residual frame loss rate after correction and E-model rating factor for different FEC schemes; AMR on UMTS channels with different packet loss rates and base delay of 100 ms.

rate for this repetition can be considerably lowered when transmitting the repeated frame three packets later (MR67: REP $p=1$, $dp=3$) and thereby breaking some of the loss bursts. The increased delay leads to some quality degradation which, however, is more than compensated by the increased error robustness. The 4.75 kbit/s AMR mode with 2 redundant copies per packet and a transmission distance of 2 packets (MR475: REP $p=2$, $dp=2$) cannot compete with the other schemes at the considered loss rates. Only at very high loss rates, the highly increased robustness against loss can compensate for the low base quality and the increased delay. *Block codes* can be designed flexibly and are efficient in reconstructing missing frames, as can be seen for two exemplary configurations with code rate 1/2 (MR67: RS $n=4$, $k=2$; RS $n=6$, $k=3$). Because of the increased delay, the gain in robustness when using longer block lengths n only leads to better quality at higher loss rates. Not as flexible as the block codes, but nevertheless very efficient at certain rates is the transmission of specific *XOR combinations* of frames as redundant information in following packets. The XOR scheme in this example is of code rate 1/2 (see Sec. 3). Although it achieves the second lowest residual frame loss rate, the XOR scheme with delayed transmission of the FEC frames (MR67: XOR $dp=3$) does not provide the best quality according to the R factor because of its large increase in delay. The repetition of a single frame per packet which is transmitted three packets later leads to the best overall quality for the considered channel model parameters.

4.3 VoIP using PCM on WLAN Channels

In this scenario we consider 64 kbit/s PCM speech transmission over WLAN. We will show how to utilize the proposed adaptable channel model to determine a suitable frame length per packet and the amount of redundancy to add with FEC.

We simulated a WLAN channel with the bit level IEEE 802.11a simulation model from the MathWorks Simulink Communications Blockset at different SNR values. The transmission data rate of the channel was set to a fixed rate of 6 Mbit/s. Based on the simulated error patterns, we determined a channel model with a high resolution ($T'_{\mathrm{TTI}} = \tau'_p = 0.08\,\mathrm{ms}$) which can be adapted for different packet lengths and TTIs according to Sec. 2. Here, no header compression is considered. The required overhead for transmitting the packet headers increases significantly for decreasing frame lengths. The minimum possible frame length will therefore be limited by the available data rate.

The frame loss effect for arbitrary frame lengths cannot be determined with the E-model yet. Therefore, we compare only the resulting frame loss rates for different frame lengths and FEC schemes, shown together with the required total data rate in Fig. 3. In general, a short frame length leads to a lower residual loss rate because the probability of having bit errors in the packet gets less. However, the increased data rate for a lower frame length may instead be invested into the transmission of additional redundancy which will also decrease the effective frame loss rate at the receiver. The FEC choice and amount of redundancy together with a specific frame length will be limited by the tolerable total delay. The curves in Fig. 3 show that a transmission with 10 ms frame length and one redundant frame per packet leads to a considerably lower loss rate than a transmission with 5 ms frame length and no redundancy. Both require a similar total data rate. A further increase of the frame length to 20 ms and a transmission of 2 redundant frames per packet does not decrease the loss rate much further for the considered channel.

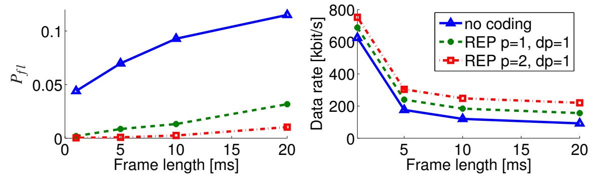

Figure 3: Frame loss rate and data rate (including packet and MAC headers) for different frame lengths and FEC schemes; VoIP using PCM on WLAN channel with SNR=20 dB.

5 Conclusions

We presented a general methodology for analytical determination of the expected quality of a VoIP transmission over heterogeneous packet networks with wireless access. The quality depends on the loss statistics of the end-to-end channel, the source encoding rate, the frame length, the used FEC scheme, and the required delay. The presented approach can be utilized for optimizing a transmission strategy for a given network scenario. It can further be used for adaptation during call or streaming if suitable information for updating the channel model is available. The reliability of the prediction of the expected quality strongly depends on the underlying quality model, here the E-model. The model should therefore always incorporate the newest standardizations.

References

[1] 3GPP. TR 25.993: Typical examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) supported by UTRA, Mar. 2008.

[2] E.O. Elliott. Estimates of error rates for codes on burst-noise channels. *The Bell System Technical Journal*, 42:1977–1997, 1963.

[3] P. Frossard. FEC performance in multimedia streaming. *IEEE Communications Letters*, 5(3):122–124, March 2001.

[4] Edgar N. Gilbert. Capacity of a burst-noise channel. *The Bell System Technical Journal*, 39(5):1253–1265, sep 1960.

[5] ITU-T Rec. G.107. The E-model, a computational model for use in transmission planning, March 2005.

[6] ITU-T Rec. G.113. Transmission impairments due to speech processing, November 2007.

[7] ITU-T Rec. P.834. Methodology for the derivation of equipment impairment factors from instrumental models, July 2002.

[8] ITU-T Rec. P.862. Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, Feb. 2001.

[9] Wenyu Jiang and Henning Schulzrinne. Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss. In *NOSSDAV '02*, NY, USA, 2002. ACM Press.

[10] I. Johansson, T. Frankkila, and P. Synnergren. Bandwidth efficient AMR operation for VoIP. In *IEEE Workshop on Speech Coding*, Tsukuba, Ibaraki, Japan, October 2002.

[11] R. Lefebvre, G.T. Philippe, and R. Salami. A study of design compromises for speech coders in packet networks. In *Proc. of the Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, volume 1, pages 265–8, 17-21 May 2004.

[12] Frank Mertz and Peter Vary. Packet Loss Concealment with Side Information for Voice over IP in Cellular Networks. In *ITG-Fachtagung Sprachkommunikation*, Kiel, Germany, April 2006.

[13] L.R. Welch. Hidden Markov Models and the Baum-Welch Algorithm. *IEEE Information Theory Society Newsletter*, Dec. 2003.