Efficient Audio Communication over Heterogeneous Packet Networks with Wireless Access

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors der
Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Ingenieur
Frank Mertz
aus Hamburg

Berichter: Universitätsprofessor Dr.-Ing. Peter Vary
Universitätsprofessor Dr.-Ing. Klaus Wehrle

Tag der mündlichen Prüfung: 22.12.2010

AACHENER BEITRÄGE ZU DIGITALEN NACHRICHTENSYSTEMEN

Herausgeber:

Prof. Dr.-Ing. Peter Vary
Institut für Nachrichtengeräte und Datenverarbeitung
Rheinisch-Westfälische Technische Hochschule Aachen
Muffeter Weg 3a
52074 Aachen
Tel.: 0241-80 26 956
Fax.: 0241-80 22 186

Bibliografische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar

1. Auflage Aachen:
Wissenschaftsverlag Mainz in Aachen
(Aachener Beiträge zu digitalen Nachrichtensystemen, Band 28)
ISSN 1437-6768
ISBN 3-86130-654-9

Wissenschaftsverlag Mainz
Süsterfeldstr. 83, 52072 Aachen
Tel.: 02 41 / 2 39 48 oder 02 41 / 87 34 34
Fax: 02 41 / 87 55 77
www.Verlag-Mainz.de

Herstellung: Druckerei Mainz GmbH,
Süsterfeldstr. 83, 52072 Aachen
Tel.: 02 41 / 87 34 34; Fax: 02 41 / 87 55 77
www.Druckservice-Aachen.de

Gedruckt auf chlorfrei gebleichtem Papier

"D 82 (Diss. RWTH Aachen University, 2010)"

Dedication

This dissertation is dedicated in loving memory to my mother Wiebke Mertz. Her encouragement, support, and love have made it possible for me to arrive at this stage.

Acknowledgements

The research for this thesis has been done during my time as research assistant at the *Institute of Communication Systems and Data Processing (IND)* at the *Rheinisch-Westfälische Technische Hochschule Aachen (RWTH Aachen University)*.

First, I would like to express my sincere gratitude to my supervisor Prof. Dr.-Ing. Peter Vary for his continuous support and the valuable discussions we had over the years. I also would like to thank Prof. Dr.-Ing. Klaus Wehrle for showing interest in the thesis and for being the second reader.

Furthermore, I would like to thank all my colleagues at the IND for a very supporting and pleasant working environment. In particular, my gratitude goes to Dipl.-Math. Annika Böttcher, Dipl.-Ing. Tobias Breddermann, Dr.-Ing. Gerald Enzner, Dipl.-Ing. Bernd Geiser, Dr.-Ing. Hauke Krüger, Dipl.-Ing. Heiner Löllmann, and Dipl.-Ing. Birgit Schotsch for many inspiring discussions. Their proof-reading of various parts of the manuscript resulted in valuable improvements. Many thanks also to my students who contributed to this work.

Last but not least, I am deeply grateful to my parents and my friends for their support and encouragement over the years. Dear Yiya, thank you for your love, encouragement, and understanding.

Aachen, January 2011 Frank Mertz

Abstract

Communication networks are transforming towards all-IP networks with different fixed-line and wireless access technologies (e.g., DSL, UMTS, WLAN). One of the biggest challenges in such heterogeneous packet networks for the realization of real-time services (e.g., Voice over IP, music, and video streaming) is dealing with transmission impairments which include variable packet transmission delays and packet losses. Current systems utilize different techniques to combat these impairments, such as a *Jitter Buffer* to compensate the variance in transmission delay and reduce jitter based losses, *Forward Error Correction* (FEC) to recover lost frames, and *Packet Loss Concealment* (PLC) to estimate unrecoverable frame losses. An issue which has not yet been sufficiently addressed is how to choose the best suitable methods and parameterize them optimally for a given scenario of application and transmission channel.

This dissertation develops a new methodology to analyze and determine the most suitable FEC method for a given transmission scenario. This approach aims at the best possible transmission quality by considering both signal quality and end-to-end delay. The main results include:

- A new channel model for packet losses which is based on the Gilbert-Elliott model and can be adapted for different packet sizes and transmission intervals;
- Analytical determination of residual frame loss probabilities after erasure correction for different FEC methods based upon the new channel model;
- Fair comparison of methods with different packet sizes and transmission intervals given the adaptability of the new channel model;
- Derivation of optimal system parameters by applying the newly developed methodologies to real-life scenarios of speech and music transmission in UMTS, WLAN, and heterogeneous packet networks.

Furthermore, new PLC techniques for state-of-art speech codecs are developed, which are particularly suitable for wireless transmission channels. In particular, the estimation of lost codec parameters is improved by transmission of low-rate side information in following packets. By utilizing steganographic methods, this side information can be transmitted as hidden bit stream in the encoded bits of the following frame, thereby requiring no additional data rate.

Contents

Abbreviations & Mathematical Notation

List of Abbreviations

3GPP	Third Generation Partnership Project
AAC	Advanced Audio Coding
ACELP	Algebraic Code Excited Linear Prediction
ACK	Acknowledgment
ADPCM	Adaptive Differential Pulse Code Modulation
ADSL	Asynchronous Digital Subscriber Line
ADU	Application Data Unit
AMR	Adaptive Multi-Rate (speech codec)
AMR-WB	Adaptive Multi-Rate Wideband (speech codec)
ARQ	Automatic Repeat reQuest
ATA	Analog Telephone Adapter
BFI	Bad Frame Education
BPSK	Binary Phase-Shift Keying
CCK	Complementary Code Keying
CDF	Cumulative Distribution Function
CELP	Code Excited Linear Prediction
CN	Core Network
CRC	Cyclic Redundancy Check
DPCM	Differential Pulse Code Modulation
DBPSK	Differential BPSK
DSL	Digital Subscriber Line
DSSS	Direct-Sequence Spread Spectrum
DTX	Discontinuous Transmission
ETSI	European Telecommunications Standards Institute
FEC	Forward Error Correction
FMC	Fixed Mobile Convergence
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat reQuest
HMM	Hidden Markov Model
HSPA	High Speed Packet Access

HSDPA	High Speed Downlink Packet Access
HSUPA	High Speed Uplink Packet Access
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
iLBC	Internet Low Bit Rate Codec
IMS	IP Multimedia Subsystem
ISCD	Iterative Source Channel Decoding
ISF	Immittance Spectral Frequencies
ITU	International Telecommunications Union
IP	Internet Protocol
IPv4	Internet Protocol, Version 4
IPv6	Internet Protocol, Version 6
ISP	Internet Service Provider
LAN	Local Area Network
LBG	Linde-Buzo-Gray algorithm
LLC	Logical Link Control
LP	Linear Prediction
LSF	Line Spectral Frequencies
LTE	Long Term Evolution
LTP	Long Term Prediction
MAC	Medium Access Control
MDC	Multiple Description Coding
MDCT	Modified Discrete Cosine Transform
MOS	Mean Opinion Score
MPDU	MAC Protocol Data Unit
MPEG	Moving Picture Experts Group
MPLS	Multi Protocol Label Switching
OFDM	Orthogonal Frequency-Division Multiplexing
PB	Piggibacked Transmission
PCM	Pulse Code Modulation
PDCP	Packet Data Convergence Protocol
PDF	Probability Density Function
PDU	Protocol Data Unit
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
PLCP	Physical Layer Convergence Protocol
POTS	Plain Old Telephony Service
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase-Shift Keying
RAB	Radio Access Bearer
REP	Repetition Code
RLC	Radio Link Control

ROHC	RObust Header Compression
RS	Reed-Solomon Code
RTCP	RTP Control Protocol, Real-time Transport Control Protocol
RTP	Real-time Transport Protocol
RTX	Retransmission
SD	Spectral Distance
SDH	Synchronous Digital Hierarchy
SDP	Session Description Protocol
SEP	Separate Transmission
SIP	Session Initiation Protocol
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UED	Unequal Error Detection
UEP	Unequal Error Protection
ULP	Uneven Level Protection
UMTS	Universal Mobile Telecommunications System
VoIP	Voice over Internet Protocol
VoWLAN	Voice over Wireless LAN
VPN	Virtual Private Network
WAN	Wide Area Network
WLAN	Wireless LAN
XOR	Exclusive Disjunction Code

List of Principal Symbols

Channel Model

G	channel model state with low(er) loss probability
B	channel model state with high(er) loss probability
$P_{\mathrm{t},XY}$	transition probability from state X to state Y; $X, Y \in \{\mathrm{G}, \mathrm{B}\}$
$P_{\mathrm{e},X}$	loss probability in state X; $X \in \{\mathrm{G}, \mathrm{B}\}$
$P_{\mathrm{s},X}$	probability of being in state X; $X \in \{\mathrm{G}, \mathrm{B}\}$
k_{t}	factor between adapted channel model and base model for the packet transmission time interval
k_{p}	factor between adapted channel model and base model for the packet transmission time
$P_{\mathrm{e},XY}^{(k_{\mathrm{t}},k_{\mathrm{p}})}$	transition dependent loss probability; transition from state X to Y; $X, Y \in \{\mathrm{G}, \mathrm{B}\}$; channel model has been adapted with factors k_{t} (for T_{TTI}) and k_{p} (for τ_{p})
$P_{\mathrm{t},XY}^{(k_{\mathrm{t}})}$	transition probability from state X to state Y with k_{t} intermediate steps; $X, Y \in \{\mathrm{G}, \mathrm{B}\}$
l_{JB}	available delay budget for the considered packet at the receiver

$P_{l,J}$	probability of loss due to jitter

Probabilities of Frame and Packet Loss

b	burst length (in frames or packets)
g	gap length, i.e., error-free length (in frames or packets)
$\overline{b}$	mean burst length (in frames or packets)
$\overline{g}$	mean gap length (in frames or packets)
P_{c}	conditional error probability
P_{e}	error probability
P_{fl}	frame loss probability
P_{pl}	packet loss probability
P_{u}	unconditional error probability
$P_{\mathrm{b}}(b)$	probability that a burst has the length b
$P_{\mathrm{b,s}}$	probability of occurrence for a burst start
$P_{\mathrm{bl}}(b)$	probability of occurrence for a burst of length b
$P_{\mathrm{g}}(g)$	probability that a gap has the length g
P_{gap}	probability of occurrence for a gap start
$P_{\mathrm{gl}}(g)$	probability of occurrence for a gap of length g
$P(m,n)$	probability of m losses in n successive packets
$P_{XY}(m,n)$	probability of m losses in n successive packets; channel is in state X at first packet and in state Y at $n+1$-th packet; $X, Y \in \{\mathrm{G}, \mathrm{B}\}$
ρ	loss pattern of a group of successive packets; example: $\rho = \{1\, x^2\, 0\}^2$ denotes a pattern of a single lost packet, followed by two packets which are received or lost, followed by a received packet; the pattern occurs twice
$P^{\mathrm{pat}}(\rho)$	probability of occurrence for loss pattern ρ
N_{rtx}	maximum number of transmission attempts for each packet (ARQ)
$P_{\mathrm{rtx}}(n)$	probability of requiring n transmission attempts for a given packet when at most N_{rtx} attempts are allowed (ARQ)
$\overline{n}_{\mathrm{rtx}}$	average number of transmission attempts per packet (ARQ)

Packet Structure and Transmission

L_{h}	size of packet headers (MAC, IP, UDP, RTP) [bit]
L_{plh}	size of RTP payload header for each frame in packet [bit]
L_{p}	size of packet incl. MAC, IP, UDP, RTP headers [bit]
L_{pl}	size of packet payload [bit]
L_{ACK}	size of an acknowledgment packet [bit]

N_f	number of codec frames per packet
r_c	code rate of the FEC scheme
$r_\mathrm{c,p}$	packet code rate of the FEC or retransmission scheme
d_il	interleaver depth
l_il	interleaver length
R_c	encoding rate of the media codec [bit/s]
R_p	packet data rate [bit/s]
$R_\mathrm{p,0}$	packet data rate without FEC or retransmission [bit/s]
$R_\mathrm{p,max}$	maximum packet data rate available on the channel [bit/s]
R_ch	transmission rate on the channel [bit/s]
T_c	fixed frame length of the codec [ms]
T_f	frame length per packet [ms]
T_la	look-ahead of the media codec [ms]
T_TTI	transmission time interval between packets [ms]
τ_p	packet transmission time, i.e., the time the channel is occupied [ms]; note: this is not the transmission delay on the channel
τ_ACK	transmission time of an acknowledgment packet [ms]
δ_p	channel access delay for the retransmission at the sender [ms]
δ_ACK	channel access delay for the acknowledgment at the receiver [ms]
D	end-to-end delay including encoding, packetization, transmission, buffering, and processing delays [ms]
D_max	maximum delay tolerated by the application [ms]
D_s	delay introduced at the sender [ms]
D_r	delay introduced at the receiver [ms]
D_enc	delay introduced by the media codec, i.e., framing delay and possibly look-ahead [ms]
$D_\mathrm{s,fec}$	delay introduced by the FEC scheme at the sender [ms]
$D_\mathrm{r,fec}$	delay introduced by the FEC scheme at the receiver [ms]
D_fec	delay introduced by the FEC scheme [ms]
D_proc	total processing delay [ms]
$D_\mathrm{s,proc}$	processing delay introduced at the sender [ms]
$D_\mathrm{r,proc}$	processing delay introduced at the receiver [ms]
D_buf	length of the receiver buffer [ms]
D_tx	packet transmission delay on the channel [ms]
d_p	distance (packets) between repeated frames for the repetition code (REP)

d_r	number of frame lengths to wait for future packets at the receiver to be used for erasure decoding (XOR code)
m_r	number of frame lengths to consider from the past to be used for erasure decoding (XOR code)

Packet Loss Concealment

$A(z)$	short-term prediction filter in a CELP based speech codec
T_0	pitch lag in a CELP based speech codec
g_a	adaptive codebook gain in a CELP based speech codec
g_c	fixed codebook gain in a CELP based speech codec
$e(n - T_0)$	adaptive codebook vector in a CELP based speech codec
$\hat{s}(n)$	synthesized speech signal in a CELP based speech codec
$\mathbf{q}(n)$	LSF/ISF vector of frame n
$\mathbf{r}(n)$	mean removed residual LSF/ISF vector of frame n
$\overline{\mathbf{q}}$	constant mean LSF/ISF vector (expectation)
f_p	vector of prediction factors for the LSF/ISF coefficients
$\hat{\mathbf{q}}(n)$	estimated LSF/ISF vector for frame n
α_n	weighting factor for LSF/ISF interpolation
$\mathbf{e}_q(n)$	quantized estimation error for LSF/ISF vector
$\hat{e}_{T_0}$	quantized estimation error for pitch lag

1

Introduction

The infrastructure of communication networks is currently undergoing a technological revolution, changing from the traditional circuit-switched technology towards packet-switched transmission using the *Internet Protocol* (IP) suite. The adoption of packet-switched transmission provides high flexibility in delivering various services like telephony, music and video streaming, email, instant messaging, web browsing, etc. The further advantage of such a network structure is that it uses standardized and cost-efficient components and can be easily extended. Nevertheless, such a way of transmission also involves a significant amount of signaling overhead due to the packet headers and leads to variable packet transmission delays and packet losses. These properties pose challenges on the realization of real-time services like voice communication and music streaming. The objective of this work is to develop strategies to optimize realizations of such services so that users can have the best experience under given constraints of the application and the network.

The development towards *all-IP* networks is progressing quickly. Many providers of classical circuit-switched telephone services, both fixed-line and mobile, have already migrated their core network infrastructures to IP-based transmission technology. While circuit-switched connections are still partly used for the last mile to the customer (especially for telephone services), this technology is likely to phase out over the next decade. In contrast to the incumbent operators of fixed-line telephone services, other service providers already deliver *Voice over IP* (VoIP) (i.e. telephone services) directly to the customer via their packet-switched DSL[1] or cable access networks. At the same time, a *fixed-mobile convergence* (FMC) can be observed, with mobile communication standards (e.g., UMTS[2]/HSPA[3] and the upcoming LTE[4]) providing increasing transmission rates. With high transmission rates, mobile access networks will eventually be able to facilitate most services that current fixed-line DSL connections provide, with the additional great advantage of having

[1] DSL – Digital Subscriber Line
[2] UMTS – Universal Mobile Communications System
[3] HSPA – High Speed Packet Access
[4] LTE – Long Term Evolution

access to the services anytime and anywhere. Such developments will eventually lead to a world of highly flexible communication networks, with core networks based on standardized IP transmission technology. These networks can be easily interconnected and therefore provide high flexibility of delivering a multitude of services. Attached to these core networks, a range of fixed and wireless access technologies will offer consumers the possibility to access the same personalized applications and services from different devices and locations, even when roaming.

The operators of such flexible *"all-IP"* networks also face new challenges. A wide range of third-party application and service providers may offer a variety of IP based services to the consumers, many of them free of charge as they can be easily financed, e.g., via advertisements. Consider, for example, VoIP applications like Skype, and Sipgate, IP-based services such as the map, email, and office applications of Google, video platforms like Youtube, fast-growing social network communities like Facebook, MySpace and StudiVZ, or micro-blogging services such as Twitter, just to name a few. The multitude of these available services poses the threat of downgrading the network operators to plain "bit-pipes", i.e., simple distributors of data packets. For the operators, a possible way out of this dilemma is to provide different degrees of service quality at different charges, and to offer their own services with a guaranteed quality to the customers. Consequently, the general trend for providers is to become integrated service providers, which offer the consumers a whole range of communication and entertainment services in a bundled form, i.e., telephone services, high speed Internet access, and television channels, all realized on the common IP based network infrastructure.

Interconnected core networks with different fixed-line and wireless access technologies that provide an end-to-end IP transmission channel to the aforementioned applications face new technological problems. The two most challenging problems are packet loss and packet transmission delay. An additional problem is the general limitation of the available transmission rate, especially when wireless access networks are considered. Each application, on the other hand, has specific demands on their perceived quality by the users, which is determined, to a large extent, by the amount of transmission errors, the efficiency of its implemented error concealment algorithms, and the end-to-end delay of the signal transmission. These problems and demands of networks and applications form a set of constraints when one tries to find the optimal transmission parameters. The optimization criterion has to be the best possible user experience, i.e., the optimal perceived quality of music signals, video streams, or conversational exchanges in the case of a voice call.

This work investigates and develops algorithms and models for the optimization of voice communication and music streaming in current packet-switched networks. The main focus is on wireless packet transmission channels (especially Wireless LAN and UMTS), possibly interconnected to a heterogeneous wide area network (WAN) with varying packet transmission delays. The contributions of the work can be classified into two areas: first, a new concept for an optimal parameterization of the transmission which includes coding, packetization and packet level error pro-

tection; second, an improved algorithm for the concealment of errors that cannot be recovered.

In the first part of this work, three chapters are devoted to develop a strategy for an optimal system design, which examines ways to determine a suitable choice of transmission parameters and error protection schemes with the goal to achieve the best possible user impression. Chapter 3 develops a flexible channel model which models commonly observed packet loss distributions and can be adapted to different packet sizes and transmission time intervals. The flexible adaptability of the model is essential for the theoretical optimization of the system. Chapter 4 discusses different schemes of packet level error protection, i.e., *forward error correction* (FEC) codes and retransmission. Based on the channel model introduced in Chapter 3, the rate and distribution of residual frame losses after error correction are derived theoretically for each of these schemes. Chapter 5 applies the theoretical results of the preceding chapters to concrete optimization problems in real-life systems (i.e., voice conversation and music streaming on UMTS, WLAN, and heterogeneous packet channels).

The second part of this work deals with unrecoverable transmission errors. Data rate and delay constraints of typical networks and their applications usually prohibit the utilization of error protection schemes that are effective enough to correct all errors. Hence, some residual errors in the form of frame losses will remain. *Packet Loss Concealment* (PLC) algorithms are therefore important for constructing a suitable replacement signal for the lost frames at the receiver's end. In Chapter 6, a new packet loss concealment approach is developed for CELP-based speech codecs which is particularly suited for wireless transmission channels. This approach generates low-rate side information for each frame at the sender and transmits together with the following frame. The side information is then used at the receiver to assist the packet loss concealment. A further contribution of the chapter is a new way of transmitting the side information as steganographic bitstream hidden in the original codec bitstream.

Finally, Chapter 7 summarizes the findings of the preceding chapters.

2

Packet Based Speech and Audio Transmission

Packet-switched multimedia transmission is based on a complex system of various components at transmitter and receiver, involving several transmission protocols from different layers of the network architecture. In order to achieve the optimal perceived quality, many system parameters need to be jointly optimized depending on the type of application and the characteristics of the network. An overview of the general system structure of speech or music transmission in packet-switched networks is given in Section 2.1. In the following, the different system components and protocols that are involved are described in more detail, concentrating on those aspects which are important for the consideration of an optimal system parameterization. Protocols for session initiation and control, as well as *Quality of Service* (QoS) control are described briefly in Section 2.2. The description of the media transmission itself will start with the media codecs for compressing the signal to transmit (Section 2.3), proceed with the packetization of the media frames including all protocols down to the network layer (Section 2.4), and continue to the aspects of packet transmission through different kinds of wired and wireless networks (Section 2.5). The impairments that the transmission may experience in different networks are separately discussed in Section 2.6. Finally, an overview on different possible techniques for combating the effect of packet losses is given in Section 2.7. These techniques comprise sender-driven *Forward Error Correction* (FEC) schemes, the use of a receiver buffer, as well as receiver-based *Packet Loss Concealment* (PLC) approaches. In most systems, a combination of these measures is applied. The optimal choice and parameterization of these techniques is the central focus of this work and will be discussed in the subsequent chapters. The optimization requires the assessment of the resulting quality as perceived by a user of the speech or music transmission service. Standardized means for a formal prediction of this quality are discussed in Appendix H.

2.1 System Description

The transmission of speech and audio signals over IP channels involves several signal processing algorithms and transmission protocols on the different layers of the network architecture. Figure 2.1 gives an overview of the different system components, including a classification into logical layers according to two common reference models, the TCP/IP Model, also referred to as Internet Protocol Suite [Braden 1989], and the Open Systems Interconnection (OSI) Reference Model [ITU-T Rec. X.200 1994; ISO/IEC 7498-1:1994 1994]. The layers of the TCP/IP model are not as strictly defined as those of the ISO OSI model, however, they roughly encompass the functionality of the ISO OSI layers as shown on the right side of the figure[1].

Within the transmitter, the signal is first segmented into frames, encoded (compressed) if necessary, and optionally redundant data is generated for end-to-end error protection, e.g., FEC. For streaming applications from a media server, the encoding process usually has been applied beforehand and the encoded media frames are stored internally. The media and FEC frames are collected in the transmit buffer and subsequently assigned to the payloads of RTP (Real-time Transport Protocol) PDUs (Protocol Data Units). RTP is considered as *application layer* protocol in the Internet Protocol Suite, as it provides the means to be adapted to different applications of multimedia transmission while keeping the underlying transport and network layers application independent. The ISO OSI model assigns RTP to the *session layer*.

The transport and network layer protocols, UDP (User Datagram Protocol) and IP (Internet Protocol), add their own headers with necessary information, e.g., for addressing, and thereby form the IP packets to transmit over the network. The packets are in general not disassembled during transmission, since the IP and UDP protocols provide an end-to-end transmission independent of the actual underlying network. The lower network layers, i.e., data link and physical layer in the OSI model, are network dependent and specific for the considered transmission channel. In a heterogeneous network, the IP packets are therefore handed to different link layer protocols along the transmission path over, e.g., Ethernet, WLAN, or UMTS transmission channels.

The link layer protocols are responsible for the adaptation to the physical layer. They assign the IP packets to the channel specific framing format, segmenting or concatenating the packets in this process if necessary. In some mobile network standards, a further protocol at the top of the link layer may implement a compression of the considerably large IP, UDP and RTP headers before transmission. In UMTS the *Packet Data Conversion Protocol* (PDCP) implements such header compression according to [Bormann et al. 2001] (cf. Section 2.4.2).

The transmission on the physical channel is usually protected against transmission errors using channel coding techniques. Nevertheless, the channel decoder

[1] Strictly speaking, the transport layer of the TCP/IP model includes some part of the OSI session layer, and the link layer some part of the OSI network layer.

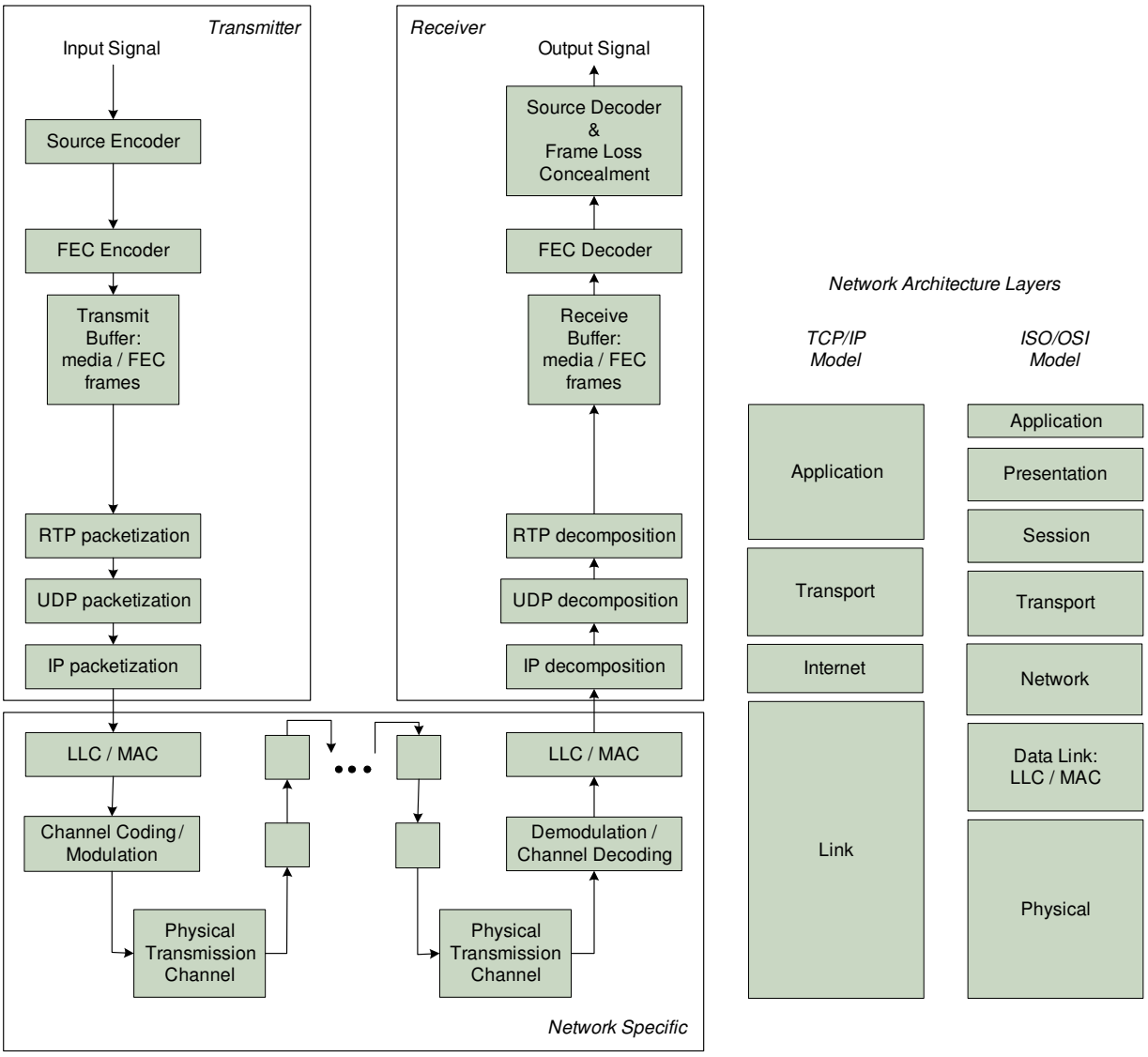


Figure 2.1: System overview of the transmission of speech, music, or other multimedia signals in packet networks using the IP Protocol Suite. Network architecture layer classification according to TCP/IP model and ISO OSI model shown on the right.

may not be able to correct all transmission errors. Residual errors are usually detected by *Cyclic Redundancy Check* (CRC) checksums which are implemented at the link layer and the affected PDUs (protocol data units) and hence the affected IP packets are discarded. Some link layers, e.g., UMTS HSPA (high speed packet access) channels, implement a fast retransmission technique for PDUs which still contain residual bit errors after channel decoding. The link layer usually ensures to deliver only error-free packets to the upper network layer, which therefore will observe some losses of complete IP packets. In heterogeneous packet networks, the application has usually no influence on the parameterization of link and physical layer. Additional error protection measures might therefore be employed on frame level as, e.g, FEC.

Besides transmission errors, the packets also experience a transmission delay. In packet networks, this transmission delay may exhibit a high degree of variation from its minimum value, mainly due to congestion and buffering at network nodes. The result is a so-called jitter in the interarrival times of subsequent periodically transmitted packets of a *Voice over IP* (VoIP) or music streaming application. Depending on the tolerable end-to-end delay, some packets may have to be discarded at the receiver if they arrive after their scheduled playout time.

The receiver of a multimedia IP packet stream decomposes the received packets and extracts the contained media frames as well as possibly any further redundant information, e.g., FEC frames. The frames are first placed in a receiver buffer before they are further processed at the appropriate time for playout. Depending on its length, this buffer is able to compensate for some of the variation in the packet's interarrival times. The maximum possible length of the buffer depends on the application constraints regarding the maximum tolerable end-to-end delay. If a frame has not arrived by the scheduled playout time because of packet loss or delay, redundant information, if available, may be used to recover the missing frame. With a joint parameterization of receiver buffer length and packet level FEC schemes, an optimal trade-off between loss rate and delay can be achieved. The frames are then decoded and the synthesized signal is played out. If a missing frame cannot be reconstructed from redundant encodings or transmissions, it is estimated by suitable PLC routines.

2.2 Session Setup and Quality of Service Control

Besides the actual transmission protocols for the media data, further protocols are required for session initiation and control, as discussed in Section 2.2.1. In some networks, the QoS of specific applications and transmission streams may be guaranteed by the protocols described in Section 2.2.2.

2.2.1 Signaling Protocols for Managing a Session of Multimedia Packet Transmission

Several signaling protocols have been standardized for packet transmission applications. These protocols control the connection setup and disconnect procedure, the negotiation of the media codec, and the session itself, e.g., by sending streaming commands like start, pause, stop, etc. The ITU has standardized a protocol framework for IP based multimedia conferencing sessions in [ITU-T Rec. H.323 2006], defining a set of protocols to use for data and control transmission. Whereas, the standardization organization for Internet protocols, IETF (Internet Engineering Task Force), has standardized the *Session Initiation Protocol* (SIP) [Rosenberg et al. 2002], a simple and flexible text-based protocol which can be easily extended with further functionalities. It utilizes the *Session Description Protocol* (SDP) [Handley et al. 2006] for negotiating the parameters of the media transmission, e.g., the codec type, etc. SIP is the chosen protocol for the *IP Multimedia Subsystem* (IMS) [3GPP TS 23.228], a standardized framework initiated by 3GPP for offering a variety of IP based applications and service in mobile communication networks.

Signaling protocols are not the focus of this work. For the considered applications, it is therefore assumed that suitable signaling protocols are employed to perform session initiation and parameter negotiation/exchange if necessary.

2.2.2 Traffic Management and Quality of Service

For providing QoS in an efficient and scalable manner in IP networks, several architectures have been proposed and standardized by the IETF: *Integrated Services* (IntServ) [Braden et al. 1994], using the *Resource ReSerVation Protocol* (RSVP) [Braden et al. 1997]; *Differential Services* (DiffServ) [Blake et al. 1998],[Grossman 2002]; and *Multiprotocol Label Switching* (MPLS) [Rosen et al. 2001]. In current core networks of service providers, the IP/MPLS transmission technology is increasingly applied, providing a high flexibility in traffic management and the setup of virtual private networks (VPN). For QoS control in such networks, the *DiffServ* protocol is used. The routers sort the incoming packets into different queues according to their priority and perform an according scheduling algorithm. In current systems, this protocol is favored above the more complex *IntServ* protocol which reserves the required bandwidth for a service at all routers along the path through the network. In spite of the employment of these protocols in providers' core networks, a multimedia transmission in the public Internet or in a heterogeneous network is usually still a best-effort transmission of the IP/UDP/RTP packets which have to compete with other applications.

2.3 Speech and Audio Codecs for Packet Networks

This section gives an overview on speech and audio coding standards from the perspective of applications using packet transmission. It will focus on the spe-

cific demands and constraints of such packet transmission scenarios and how they influence the choice of a suitable codec.

2.3.1 Demands and Constraints for Packet Transmission

The choice of an appropriate speech or audio codec in general depends on a) the demands of the respective application, i.e., the desired audio bandwidth, quality, and end-to-end delay; b) the constraints of the device, i.e., microphone/speaker characteristics, available computational complexity; and on c) the constraints of the network, i.e., transmission delay, available bit rate, amount and characteristics of transmission errors and hence the required error robustness of the codec.

The flexible packet transmission technology provides possibilities to easily overcome the traditional narrowband communication quality (audio bandwidth below 4000 Hz) by employing speech codecs which encode speech signals with a wider audio bandwidth, e.g., so-called wideband (50-7000 Hz), super-wideband (50-12000 Hz), or fullband (20-20000 Hz) quality.

Most of the current VoIP applications employ standardized speech codecs, e.g., PCM, if a fairly high data rate is available, or standards with a higher compression rate like ITU G.729 or 3GPP AMR. The latter standards, which have been developed for mobile communication systems, achieve a low data rate at the expense of a high sensitivity towards packet losses, i.e., the loss of complete frames, especially due to a considerable amount of error propagation. Some applications therefore rather employ proprietary codecs which have been designed for a higher robustness against packet losses, e.g., iLBC and SILK in the popular "Skype" IP telephony software. Only recently, increasing efforts have been made in ITU and 3GPP for standardizing speech codecs which are specifically designed for VoIP applications in packet networks. These standards provide data rate scalability by allowing to discard less important parts of the bitstream (layered coding schemes, cf. Section 2.3.2.4). Furthermore, they contain measures to enhance the error robustness in case of packet loss. An overview on the recently standardized ITU-T Codecs G.711.1, G.718, G.719, and G.729.1 is given in [Cox et al. 2009]. In the following section, various speech and audio coding standards will be reviewed and compared regarding their suitability for packet based applications.

2.3.2 Speech and Audio Coding Standards and their Suitability for Packet Transmission

The digital representation of an audio signal's waveform, e.g., of speech or music, usually requires a high data rate. For the efficient transmission of such signals, source coding techniques are applied which reduce the date rate through signal compression. The reduction is achieved by utilizing signal properties, models of the sound production process (e.g., human speech production) and the perception process of the auditory system, i.e., the human ear, in order to remove redundancy and irrelevance from the source signal. Common codecs either encode the waveform

directly or employ linear prediction (LP) utilizing periodicities in the signal and derive a set of parameters for the considered signal segment which can be used at the receiver to reproduce the signal with the desired accuracy. Codecs which encode speech signals with a wider audio bandwidth often combine linear prediction with transform coding techniques, e.g., the ITU-T G.729.1 codec. The existing codecs can be classified into various coding schemes which differ in their suitability for packet-based applications and networks.

2.3.2.1 Waveform Speech Codecs

Waveform codecs directly transmit the signal's waveform by encoding each sample separately. A data rate reduction can only be achieved by using companding techniques and/or differential transmission of samples. Examples of speech codecs from this category are PCM (pulse-code modulation) [ITU-T Rec. G.711 1988] at 64 kbit/s for narrowband speech, and ADPCM (adaptive differential pulse-code modulation), e.g., [ITU-T Rec. G.726 1990] at 16–40 kbit/s for narrowband, or [ITU-T Rec. G.722 1988] at 48–64 kbit/s for wideband speech. In case of transmission errors, e.g, frame losses, these codecs have an advantage over codecs which involve prediction techniques, because no error propagation results into the signal segment following a loss. The loss itself is usually replaced with an periodic extrapolation of the preceding signal segment by repeating the previous pitch period (see, e.g., [Gunduzhan and Momtahan 2001]).

2.3.2.2 Hybrid Speech Coding Schemes

The class of speech coding schemes most widely applied, especially in cellular networks, is the class of so-called hybrid codecs. These codecs employ parametric coding schemes with an additional transmission of the quantized error signal, achieving a high compression rate and still providing a good quality of the speech signal including a fairly high naturalness. The dominant technology in recent years has been the CELP (code excited linear prediction) coding principle, first introduced in [Schroeder and Atal 1985].

The CELP encoding principle is based on the source-filter model of speech production and applies linear prediction to determine the coefficients of the vocal tract filter. The excitation of this filter is then formed using an adaptive codebook for the periodic contribution, and a fixed codebook of pulse vectors for the innovation in the signal. The parameters are determined for frames of usually 20 ms length, each divided into four sub-frames. The determination of the parameters follows an *Analysis-by-Synthesis* principle, i.e., optimizing the decoded signal during the encoding process. The vector quantization of the excitation contributions for each sub-frame perform a closed-loop search for the optimal entries. The codebook search is done in a "perceptually weighted domain" in order to adjust the resulting quantization error to the current spectral envelope of the signal. Such a *noise shaping* results in a better perceptual quality than a white noise error.

Prominent examples of CELP codecs are: the *Enhanced Full Rate* (EFR) codec [ETSI] with 12.2 kbit/s, which is the standard codec in cellular GSM networks; the

Adaptive Multi Rate (AMR) codec [3GPP TS 26.090], the standard codec for UMTS networks with code rates from 4.75 kbit/s to 12.2 kbit/s; its wideband version AMR-WB [3GPP TS 26.190] encoding speech signals with a wider audio bandwidth of up to 7 kHz at bit rates from 6.6 kbit/s to 23.85 kbit/s; and the ITU-T G.729 speech codec with 8 kbit/s [ITU-T Rec. G.729 1996a].

The high encoding efficiency of these codecs is gained at the expense of a fairly high sensitivity to transmission errors, especially frame losses. Due to the predictive encoding of parameters and the codec's inherent structure, particularly the adaptive codebook, the loss and estimation of a frame leads to a considerable error propagation into following frames. Within IETF, the *internet Low Bit Rate Codec* (iLBC) [Andersen et al. 2004] has been standardized which avoids the typical frame interdependencies of CELP based codecs and thereby increases the loss robustness. However, this advantage is achieved at the expense of a higher data rate of 13.33–15.2 kbit/s for the same base quality as 8–12 kbit/s codecs. iLBC is one of the codecs implemented in the popular Skype Voice over IP software. Currently, however, the standard codec used in Skype is the so-called SILK speech codec which provides scalability in several dimensions including wideband and super-wideband audio quality. It has been developed by Skype and published as Internet-Draft at the IETF in [Vos et al. 2010]. SILK uses a fixed frame length of 20 ms and can be operated at different operating points which can be switched any time during operation on a frame-by-frame basis. This gives this codec a high flexibility to adapt to the available transmission bandwidth as well as changing channel characteristics. The operating point of the codec is defined by the following parameters: sampling rate (8-24 kHz), packet rate (1-5 frames per packet), bitrate (6-40 kbit/s), packet loss resilience (amount of inter-frame dependencies), use of in-band forward error correction (additional low bitrate encoding of onsets or transients added to a subsequent packet), complexity (high, medium, low), and use of discontinuous transmission (DTX). The codec claims to achieve good quality at around 1 bit/sample, and transparent quality for most material at 1.5 bits/sample.

2.3.2.3 Multiple Description Coding

Multiple Description Coding (MDC) techniques aim for a higher robustness against transmission errors by generating several independent descriptions for a media frame which are then transmitted separately. The reception of all descriptions enables the receiver to reconstruct the signal at its optimal quality. The reception of a subset or only a single description is nevertheless sufficient to regenerate the signal, although at a somewhat lower quality. Particularly designed MDC schemes usually require a higher data rate for achieving the same base quality as single description coding schemes because of the loss in coding efficiency. To achieve the error robustness, the descriptions have to be transmitted on independent transmission paths, e.g., on separate transmission channels if available. Otherwise, nearly independent transmission can be achieved by a transmission at different time points on the same channel, which, however, increases the end-to-end delay. In packet-based applications, the descriptions are usually transmitted in separate packets on the same

channel, causing further transmission overhead due to the required packet headers, unless descriptions of successive frames are transmitted together in a packet. For a discussion of practical design examples, see, e.g., [Chen and Chen 1997; Goyal and Kovacevic 1998; Kubin and Kleijn 1999; Jiang and Ortega 2000; Pradhan et al. 2004; Puri et al. 2005]. MDC schemes do not necessarily require a larger transmission bandwidth. Alternative approaches utilize the interleaved subsets of samples or parameters as multiple descriptions (see, e.g., [Jayant and Christensen 1981; Martin et al. 2001]). In the *FlexCode* project within the European Commission's Sixth Framework Programme "Information Society Technologies" [FlexCode 2009], a highly flexible system of generic source and channel coders has been developed [Bruhn et al. 2008; Schmalen et al. 2010]. Different channel coding concepts are applied to different sets of source coding parameters (model parameters and transform coefficients) and an instantaneous adaptation of the source and channel coding rates to varying channel conditions is supported.

Multiple description schemes may further be utilized for data rate scalability, if a specific description of each frame may be dropped by the network without reducing the quality of the signal too much. However, this is usually better fulfilled for *embedded coding schemes* as described in the following section.

2.3.2.4 Embedded (Layered) Coding Schemes

Embedded coding schemes generate a single base layer allowing the receiver to decode the signal at a base quality. In addition, a certain amount of enhancement layers are generated which provide a further refinement of the signal and thereby improve the resulting signal quality. As for MDC schemes, the network may omit parts of the packet stream if the transmission path to a receiver has a limited capacity. However, in contrast to MDC schemes each layer is dependent on the layers lower than itself, i.e., the layers are hierarchical. Hence, at least the base layer needs to be received to be able to decode the signal. Main application scenario for these codecs are point-to-multipoint or even multipoint-to-multipoint connections, e.g., a speech or video conference application, where the participants are connected via different access technologies with different transmission rate capabilities. The flexibility of adapting the transmission rate within the network requires some intelligence within the network to be able to decide where to strip the packets of parts of its content.

Recently standardized layered coding schemes include the ITU-T codecs G.729.1, G.711.1, and G.718. The two former codecs include the respective ITU-T codecs G.729 and G.711 as their base layer. All of these codecs are scalable in bit rate and audio bandwidth. The lower layers are based on the CELP coding technique and deliver telephone band speech quality. Higher layers usually employ transform coding techniques like MDCT (modified discrete cosine transform) and provide good audio quality of up 7 kHz.

2.3.2.5 Coding Schemes for Audio Signals

Audio signals require a different encoding approach than speech signals. The more complex and variable signal structure cannot be described by a production model as in speech coding. Instead, a model of the auditory system is utilized for compression, describing the frequency dependent listening sensitivity as well as masking effects. Most audio codecs therefore employ frequency based coding schemes utilizing such psychoacoustic models to achieve the compression of the audio signal. The signal is first windowed and converted from time- to frequency domain, e.g., using the MDCT. In the frequency-domain, the signal is then quantized with frequency dependent accuracy based on the psychoacoustic model and afterwards encoded. Prominent examples of audio coding schemes are MPEG-1 Audio Layer 3 [ISO/IEC 11172-3:1993 1993], more commonly referred to as MP3, and Advanced Audio Coding (AAC) [ISO/IEC 13818-7:2006 2006].

For these audio codecs, the error propagation in case of packet losses is limited to the length of overlapping windows. Most concealment schemes rely on frame repetition or sub-band extrapolation techniques in order to estimate the missing signal segment.

2.3.3 Speech and Audio Codecs Considered in This Work

The investigations and developments of this work will be applied exemplary to some of the presented coding schemes which are of considerable importance for actual applications (PCM, AMR and AMR-WB for speech signals, and MP3 for music signals). The uncompressed PCM speech codec is widely used in VoIP systems employed in company LANs. It provides PSTN (*Public Switched Telephone Network*) quality and does not require transcoding at the gateways to PSTN. In spite of the relatively high data rate it can also be used over Wireless LAN connections. With considerably lower encoding rates, the AMR and AMR-WB codecs are intended for cellular communication networks. They are widely implemented on mobile devices and may therefore be used for VoIP services as well. Finally, the MP3 coding scheme is widely used and will be considered for music streaming applications.

2.4 Packet Structure for IP based Multimedia Transmission

In packet networks that are based on the Internet Protocol (IP) Suite, the transmission of multimedia signals utilizes a specific set of these protocols. The encoded media frames are first processed by the *Real-time Transport Protocol* (RTP), which packetizes the frames according to a standardized codec specific payload format and further includes information on the sequential order of the packets. On the transport layer, the connection-less *User Datagram Protocol* (UDP) is used for multimedia applications, and finally the network (Internet) layer *Internet Protocol* (IP) provides the end-to-end transmission as for all packets in the network. The properties and

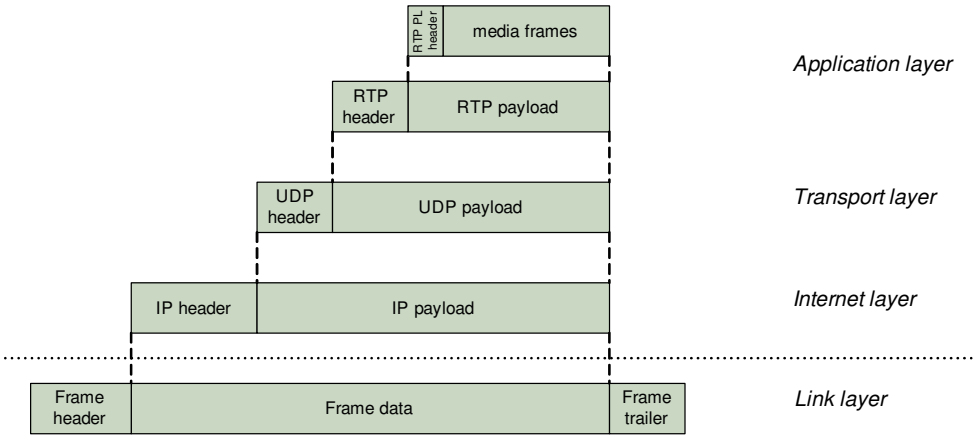


Figure 2.2: Packetization of media frames

functionalities of these protocols are shortly reviewed in Appendix A. The packets are assembled as explained in the following section, including a short overview on the specific RTP payload formats for the considered speech and audio codecs. The considerable overhead of packet headers can be reduced on wireless links that support the standardized *RObust Header Compression* (ROHC) algorithm [Bormann et al. 2001] as explained in Section 2.4.2.

2.4.1 Packetization of Media Frames and Assembly of IP Packets

The packetization of the encoded media frames with the IP/UDP/RTP protocols is visualized in Figure 2.2. At the application layer, the encoded media frames are assembled according to the respective RTP payload format. Depending on the specific media codec and the requirements of the application, one or several frames are packed together in each packet. Additional payload header(s) may be attached (RTP PL header) and this payload data together with the RTP header then forms the RTP packet which is handed to the transport layer. On transport and network layer, the respective protocols, UDP and IP, attach their headers with the required information regarding packet length, addressing, etc. The so formed IP packet is then transmitted end-to-end, i.e., without intermediate decomposition, utilizing the respective link layer protocols and underlying physical channels of the network paths it traverses. The link layer protocols of different networks usually attach further specific headers, e.g., a MAC header, and possibly a trailer, e.g., a CRC checksum. The link layer protocols of those networks which are relevant for this work, i.e., WLAN, and UMTS, are described in Section 2.5.

2.4.1.1 RTP Payload Formats for Important Speech & Audio Codecs

In the following, a short overview is given on the standardized RTP payload formats for the different speech and audio codecs considered in this work.

The payload format for the sample based speech encoding scheme PCM is defined in [Schulzrinne and Casner 2003, Sections 4.3 and 4.5.14]. It provides a free

choice of the segment length, i.e., the number of samples, to transmit in each packet. The length of the segment is implicitly given in the packet length field of the IP header and a particular payload header is therefore not required. The timestamp field of the RTP header provides an accurate positioning of the received data on the time-line, so that a change of the segment length within a transmission session is possible.

AMR and AMR-WB encoded speech signals are packetized according to the payload format defined in [Sjoberg et al. 2007]. The format allows to transmit an arbitrary number of frames in a packet which may be encoded with different modes of the multi-rate codecs. A specific payload header defines the frame type of each contained frame and optionally contains a separate CRC checksum for each frame. The payload format is designed to optionally support *Unequal Error Protection* (UEP) and *Unequal Error Detection* (UED) in order to take full advantage of the bit error robustness of the AMR and AMR-WB speech codecs. The application of such schemes is discussed in Chapter 2.7.5.

As exemplary audio codec, the widely used MP3 encoding scheme is considered. Although the file format stores encoded MP3 blocks of a constant length, these blocks may actually contain information of more than one audio frame. The so-called bit reservoir at the end of a block may be used by other than the current frame if their length did not completely fit into the previous MP3 blocks. An RTP payload format for a robust transmission of MP3 encoded frames has been standardized in [Finlayson 2008]. Here, the bit reservoirs are resolved and all information belonging to a single audio frame is collected and transmitted in a single packet leading to packets of varying length. As consequence, the loss of a packet only affects a single audio frame. A more detailed description of this procedure is given in Appendix B.

2.4.1.2 RTP Payload Formats for Forward Error Correction

The application of FEC requires the standardization of specific payload formats unless already included in a codec's main payload format definition. Several payload formats for forward error correction have been standardized by IETF. In the approach defined in [Perkins et al. 1997], redundant encodings of a frame can be transmitted piggybacked to the following packet. The encoding can be either a copy of the original frame or a lower rate version generated by a different codec or a lower encoding mode of a multi-rate codec. An additional payload header contains timestamp offsets and block lengths of the contained secondary encodings.

[Li 2007] defines a payload format for generic forward error correction which supports a wide range of media independent FEC configurations using parity codes and also supports unequal error protection. The parity blocks are generated as bitwise XOR of data blocks. The overall concept is termed in the standard as *Uneven Level Protection* (ULP), since payload data are protected by one or more protection levels. The configuration is signaled in-band, i.e., adaptation during a session is possible. The FEC data, i.e., the parity blocks, are either transmitted as separate packet stream with own RTP and RTP payload headers as defined in [Li

2007], or they may be treated as secondary encodings and piggybacked to following packets as defined in [Perkins et al. 1997] and explained above.

Finally, [Rey et al. 2006] defines a payload format for RTP retransmissions, an effective packet loss recovery technique for real-time applications with relaxed delay bounds. In this approach, the retransmitted RTP packets are sent in a separate stream from the original RTP stream.

2.4.1.3 Calculation of Packet Size and Packet Data Rate

On wireless transmission channels, the experienced loss distribution depends on the frequency and size of the transmitted packets. The size of the media packets depends on the encoding rate of the utilized media codec, R_c, and the frame length transmitted in each packet, T_f. The frame length T_f is defined here as the actual length of the new encoded signal segment that is transmitted in a single packet. For sample-based coding schemes like PCM, the frame length is arbitrary and can assume any integer multiple of the sample time. For frame-based codecs, the frame length T_f can — according to the previous definition — assume any integer multiple N_f of the codec's own frame length T_c, i.e.,

$$T_\mathrm{f} = N_f \cdot T_\mathrm{c}. \tag{2.1}$$

The size of the media data per packet, L_f, is then calculated from the encoding rate R_c and frame length T_f as

$$L_\mathrm{f} = R_\mathrm{c} \cdot T_\mathrm{f}. \tag{2.2}$$

The RTP payload format may define an additional payload header which contains necessary information for decomposing the payload and extracting the encoded frames. Furthermore, the payload header may contain additional information for each contained codec frame, e.g., an identifier for the encoding rate of multi-rate codecs or a CRC checksum of the frame. The codec dependent payload header therefore may consist of a common part which is attached once, $L_\mathrm{plh}^{(c)}$, and a frame dependent part which is attached for each contained frame, $L_\mathrm{plh}^{(f)}$. The length of the payload header is then given as

$$L_\mathrm{plh} = L_\mathrm{plh}^{(c)} + N_f \cdot L_\mathrm{plh}^{(f)}. \tag{2.3}$$

The total length of the IP packet, L_p, is given as the sum of the IP/UDP/RTP protocol headers, L_h, the payload header L_plh, and the size of the payload, i.e., the encoded media frames, L_f:

$$L_\mathrm{p} = L_\mathrm{h} + L_\mathrm{plh} + L_\mathrm{f} = L_\mathrm{h} + L_\mathrm{plh} + R_\mathrm{c} \cdot T_\mathrm{f}. \tag{2.4}$$

The frame length per packet also defines the frequency of the packet transmission. Hence, the IP packet data rate results to

$$R_\mathrm{p} = \frac{L_\mathrm{p}}{T_\mathrm{f}} = \frac{L_\mathrm{h} + L_\mathrm{plh} + R_\mathrm{c} \cdot T_\mathrm{f}}{T_\mathrm{f}} = \frac{L_\mathrm{h} + L_\mathrm{plh}}{T_\mathrm{f}} + R_\mathrm{c}. \tag{2.5}$$

The total size of the protocol headers, L_h, depends on the utilized IP version. The IP header amounts to 20 byte for IPv4 and 40 byte for IPv6. With the 12 byte RTP header and the 8 byte UDP header, this results in a header size of $L_\mathrm{h} = 40$ byte (IPv4/UDP/RTP) or $L_\mathrm{h} = 60$ byte (IPv6/UDP/RTP), respectively.

From the given byte numbers it becomes clear that the overhead in data rate introduced by the various headers is enormous compared to the fairly small speech data rate of some common codecs. For instance, when sending a single speech frame per packet, encoded with the highest AMR bit rate of 12.2 kbit/s, a 40 byte header (IPv4/UDP/RTP) is necessary to transport 32 byte of speech data including payload header. The overhead may be reduced by sending more speech frames in each RTP packet, but this leads to an increase of the latency and also reduces the frame erasure robustness, because in this case the loss of one packet would instantly mean a loss of several successive speech frames. To avoid such a burst loss, the frames can be interleaved before packetization, however, only at the expense of a further significant increase of the end-to-end delay. The header overhead becomes particularly relevant when there is a wireless link within the transmission route, e.g., the link between a base station and an UMTS hand-held. In this case, header compression algorithms are absolutely necessary to reduce the protocol overhead, as described in the following section.

An overview of the resulting IP packet sizes and data rates for different speech and audio codecs is given in Appendix C. The tables in the appendix also show the impact of the header overhead and the possible reduction through header compression. Furthermore, the transmission of several frames per packet and the transmission of redundancy for forward error correction are considered, which are the focus of the investigations in Chapter 4 of this work.

2.4.2 Robust Header Compression for Wireless Links

In packet switched speech transmission, the size of packet headers compared to the speech payload is very high, as shown in the previous section. Particularly, in transmissions over radio channels, this enormous overhead leads to a very low bandwidth efficiency. Therefore, the IETF has standardized an algorithm to compress RTP/UDP/IP headers for transmission over wireless links: Robust Header Compression (ROHC [Bormann et al. 2001][2]).

The header compression algorithm is based on the significant redundancy between header fields of consecutive packets belonging to the same packet stream. The header size can be significantly reduced by sending full static information (e.g. IP addresses, UDP ports) only initially, and utilizing dependencies and predictability to reduce the size of the dynamic fields in the compressed header (e.g., RTP sequence number and timestamp). A dynamic field like the UDP/UDP-Lite checksum

[2] Further profiles have been defined in [Jonsson and Pelletier 2004] for IP only compression and in [Pelletier 2005] for UDP Lite. Corrections and clarifications have been published in [Jonsson et al. 2007]. A second version of the ROHC protocol (ROHCv2) has been standardized in [Pelletier and Sandlund 2008], introducing some simplifications and enhancements. ROHCv2 introduces own header formats and does not obsolete ROHC.

that changes irregularly with every single packet has still to be transmitted completely within the compressed header. The compression scheme is robust in dealing with packet loss as well as residual errors in a received packet by optional use of a CRC in the compressed header. The algorithm compresses an RTP/UDP/IPv6 header down to a minimal size of 3 bytes without ROHC header CRC or 4 byte with ROHC header CRC, both including the 2 byte UDP checksum which is mandatory for IPv6.

Different modes of operation are defined for the ROHC header compression scheme in [Bormann et al. 2001]: the unidirectional, the bidirectional optimistic, and the bidirectional reliable mode. The modes differ in the utilization of a feedback channel for acknowledgments and error recovery if the contexts of compressor and decompressor get out of sync due to transmission errors. The *unidirectional* mode has to be used in network scenarios where a feedback channel is unavailable or undesirable. In this mode, the compressor will frequently transmit the full packet headers to ensure a possible resynchronization of the compressor and decompressor states in case of intermediate transmission errors. The *bidirectional optimistic* mode utilizes a feedback channel for sending error recovery requests from the decompressor to the compressor. This mode aims for a high compression efficiency while making only sparse use of the feedback channel. It is therefore applicable only for channels with low loss rates. Finally, the *bidirectional reliable* mode ensures a higher robustness against transmission errors by a more intensive use of the feedback channel. This mode will be assumed for application in the different scenarios of speech and music transmission over wireless packet channels in Chapter 5.

2.5 Packet Transmission over Wired and Wireless Networks

The IP packets, which are assembled according to the description in the previous section, are transmitted end-to-end through the network, i.e., they are not decomposed and reassembled in between. This aspect guarantees the service flexibility of packet networks and relieves the network nodes itself of otherwise required intelligence and processing power to decompose and analyze the packet contents.

The transmission of the IP packets is managed by link layer protocols and the underlying physical layer of the respective network or transmission link. In the following, a short overview is given on the different types of packet-based networks, considering fixed, wireless, and mobile networks. The investigations and developments conducted in this work are mainly focusing on wireless packet transmission channels like Wireless LAN and UMTS packet-switched (PS) channels. An overview on these transmission standards is given in Appendix D.

In today's communication infrastructure, the different communication networks are all interconnected and can be seen as a single large heterogeneous network with different access technologies. According to the OECD [OECD], the broadband subscriptions worldwide as of December 2008 can be classified by technology into

60% DSL, 28% cable modem, 10% fiber & LAN, and 2% other technologies, e.g., satellite or WiMAX [IEEE Std 802.16 2004]. These numbers do not yet include the increasing use of wireless broadband access through Wireless LAN access points at public spots and third generation cellular networks (UMTS). These additional access technologies are used when not at home and are therefore not included in above statistics.

The preferred modus of operation for voice communication and media streaming applications in such a heterogeneous network is a transparent end-to-end IP transmission. The clients negotiate the media codec to use in the call initiation phase (cf. Section 2.2.1) and no transcoding is required in between.

The characteristic of an effective end-to-end transmission channel in such a heterogeneous network will depend on the chosen access technology, the number of network transitions, as well as the quality of the involved transmission channels. The characteristics of the networks' transmission channels may vary greatly in terms of packet loss distribution and packet delay variation (jitter).

2.5.1 Local Area Networks (LAN)

Sufficiently designed local area networks (LANs), e.g., corporate networks, provide enough capacity to prevent congestion and packet losses. The distances between the connected clients is usually short, the end-to-end delay is therefore rather low and the variation is also not significant. Because of its limited size, the network can be monitored fairly easily and adapted or enhanced when necessary. The same network infrastructure can be used for data and voice traffic and the employment of IP telephony in such a controlled network can be expected to deliver a quality comparable to PSTN. Calls are routed to the PSTN or circuit-switched mobile networks (e.g., GSM) via gateways.

2.5.2 Broadband Access Networks

Broadband access service providers offer different DSL (digital subscriber line) technologies or access via cable TV with a wide range of access data rates. The core or backbone networks of such providers are in general dimensioned with enough capacity to prevent packet losses and excessive packet delays. Within its own core network, a broadband service provider may further provide a guaranteed quality of service for specific applications, e.g., VoIP telephony, or for specific users. Quality control is achieved through technologies like MPLS as discussed in Section 2.2.2.

However, for applications and connections which cross the boundaries of the own core network, network transitions and different intermediate network types between the user and the other end-point of the transmission may lead to packet losses and considerable variation in packet delays. Additionally, many users of fixed broadband access networks use a Wireless LAN router at home to connect their notebooks and other devices with the flexibility of moving around. This rapid spread of wireless Internet access also has some drawbacks. The increasing number of

WLAN access points in private homes leads to an increasing amount of overlapping channels when several base stations use the same transmission frequency. The result is an increasing amount of distortions and therefore also transmission failures, collisions, and less channel available times.

2.5.3 Wireless and Mobile Access Networks

Besides the wireless access through WLAN routers at home, an increasing number of WLAN access points in public places, hotels, cafés, and restaurants, etc., provides roaming users the ability to use IP services with their mobile devices. Device manufacturers are designing a multitude of capable smartphones and ultra-portable notebooks and mobile network providers offer transmission flat rates and IP services. The available data rates in mobile networks are increasing further through the development and deployment of new network technologies like UMTS-HSPA (High Speed Packet Access) and UMTS-LTE (Long Term Evolution). Hence, communication networks are quickly proceeding in their development towards the goal of providing "access anytime and anywhere" to the telephone and IP services that previously were only available via Cable or DSL.

The loss characteristic of a wireless packet transmission channel is in general a "bursty" packet loss distribution, i.e., there are periods of low loss densities and periods of higher loss densities. A mobile user may be moving and thereby experience a further variation of the error characteristics, i.e., phases of varying channel quality, caused by fading and shadowing effects, as well as inter-symbol and multi-user interference. The system may therefore employ an adaptive power control and also adapt the employed modulation scheme for higher error robustness, resulting in a variable transmission rate, as, e.g., in the WLAN standard [IEEE Std 802.11 2007].

2.5.4 Public Internet

The transmission characteristics of an Internet connection mainly depend on the available capacity and the amount of competing traffic. Packet losses may occur due to buffer overflows at network nodes in case of congestion, or due to network failures. The public Internet does in general not inflict high packet loss rates if the particular access network (DSL, WLAN, etc.) is not considered. However, it experiences packet delays with a high degree of variation which in turn may lead to packet losses at the receiver if a packet arrives after its scheduled playout time. The transmission characteristic is in general dependent on the distance between the end points (e.g., national versus international connections), the specific world region, i.e., how good the network is developed (e.g., first world versus third world countries), and the considered time of day (e.g., peak hours at lunch time or in the evening in contrast to hours of low usage, e.g., very early in the morning).

Furthermore, the network capacities are constantly enhanced worldwide by installing new and more fiber cables, as well as employing new technologies for transmitting higher data rates, e.g., by using new modulation schemes. The increased

capacity has no significant effect on the minimum transmission times since that mainly depends on the distance, but it may reduce jitter as it avoids congestion. Transmission times are reduced in parts of the world by changing the employed technology. In the recent years, countries with satellite links, e.g., Argentina and China, have moved to land lines, reducing the minimum *Round Trip Times* (RTT) from more than 400 ms to much lower values. Currently, fiber optic connections are planned for parts of Africa.

Various efforts have been made to characterize the loss and delay characteristics of the Internet, e.g., [Bolot 1993], [Bolot et al. 1995], [Bolot 1995], [Yajnik et al. 1999], [Karam and Tobagi 2001], [Sun and Ifeachor 2004]. Because of its heterogeneity and constant development, however, it is not possible to define a single explicit standard model of the Internet. The yearly "ICFA SCIC Network Monitoring Report"[3] provides statistics of various international Internet connections based on frequent ping measurements. According to the report from January 2009 [Cottrell 2009], minimum RTTs between the USA and Europe are 80-200 ms, and between the USA and China 200-250 ms. Packet loss rates have decreased over the last years, leading to a yearly average of below 1% between the USA and Europe as well as Asia. However, connections to Africa still experience around 7% loss rates and round trip times of up to and partly exceeding 400 ms.

2.6 Transmission Impairments in Packet Networks

The transmission of multimedia content over packet-switched networks of various types is subject to transmission errors and delay. However, the characteristics of these impairments are specific to the packet transmission process and shall be discussed in the following.

2.6.1 Packet Forming and Transmission Delay

In packet-switched networks, several parts of the transmission chain of sender, network, and receiver contribute to the resulting end-to-end delay of the signal transmission.

Delay introduced in the sender

For real-time speech communication, the minimum delay contribution of the *sender* consists of an algorithmic delay and a processing delay. The algorithmic delay of a speech codec is the frame length T_f plus a possible lookahead T_la, $D_\mathrm{enc} = T_\mathrm{f} + T_\mathrm{la}$, i.e., the length of the signal segment that needs to be collected before encoding. Furthermore, if FEC or frame interleaving is used before packetization, an additional algorithmic delay, $D_\mathrm{s,fec}$, needs to be considered. The processing delay, $D_\mathrm{s,proc}$, comprises all delays from the computational processing, e.g., for encoding, packetization, and transmission. This results in a delay of $D_\mathrm{s} = D_\mathrm{enc} + D_\mathrm{s,fec} +$

[3] http://www.slac.stanford.edu/xorg/icfa/scic-netmon/

$D_{\mathrm{s,proc}}$. For streaming applications, all frames are read from a previously encoded file and are therefore instantaneously available. Therefore, the delay contribution of the sender is limited to the processing delays of packetization and transmission, i.e., $D_{\mathrm{s}} = D_{\mathrm{s,proc}}$.

Delay introduced in the network

The transmission delay of packets in a packet-switched network consists of a fixed and a variable part. The *fixed part* comprises the transmission and propagation time over network links as well as transit delays (processing times) in network components. The size of this fixed transmission delay D_{tx} depends on the number of links, i.e., node connections, the packets have to traverse, the distance as well as the types and transmission rates of these links.

A further *variable part* of the transmission delay is caused by queuing delays in network elements, medium access protocols regulating access to, e.g., a shared wireless channel, possible retransmissions in case of transmission failures, among others. This *packet delay variation* is sometimes also referred to as jitter and has a minimum value of zero. If the packets take different transmission paths through the network, e.g., when a network node becomes unavailable/fails, or when it gets congested, then the fixed transmission time becomes variant in that it differs for each considered end-to-end transmission path. The most common example is the public Internet, where there is no control over the paths the packets take. In such a transmission scenario, the fastest transmission path then sets the fixed minimum end-to-end delay and the alternative slower transmission paths add to the delay variation. The resulting contribution of the variable delay component to the total end-to-end delay is determined by the length of the receiver buffer, i.e., the amount of additional time a packet is given to arrive before it is considered lost.

Depending on the considered application, the end-to-end delay and its variation may have a severe effect on the resulting service quality. For conversational services, an increasing delay will negatively affect the interactivity of the communication partners. Furthermore, the effect of other impairments, e.g., echo, may increase significantly with an increasing transmission delay of the signal. Therefore, the end-to-end delay must not exceed a certain application dependent maximum. The actual effect of delay on the conversation quality can be assessed with the ITU-T E-model as described in Appendix H.3.

Delay introduced in the receiver

The delay introduced at the receiver is determined by the length of the receiver (jitter) buffer, D_{buf}, defined here as the total time between the earliest possible time of reception (i.e., after sender and minimum transmission delay) and the playout of the signal. The buffer shall compensate for some of the transmission delay variation. Its length has to be at least long enough to cover the necessary time to wait for FEC data or interleaved frames to arrive before error correction and decoding are possible, $D_{\mathrm{r,fec}}$, as well as any processing delays, $D_{\mathrm{r,proc}}$. Hence, the receiver delay equals the buffer length, $D_{\mathrm{r}} = D_{\mathrm{buf}}$, with $D_{\mathrm{buf}} \geq D_{\mathrm{r,fec}} + D_{\mathrm{r,proc}}$. This definition

makes the buffer delay independent from the actual transmission settings, e.g., the utilized FEC scheme.

End-to-End Delay

The different delay components of sender, network, and receiver together determine the resulting end-to-end delay of the signal transmission. For reasons of simplicity, all processing delays are neglected in this work, i.e., $D_{\mathrm{s,proc}} = D_{\mathrm{r,proc}} = 0$. Then, the end-to-end delay of a conversational application is calculated as

$$D = D_{\mathrm{enc}} + D_{\mathrm{s,fec}} + D_{\mathrm{tx}} + D_{\mathrm{buf}}. \tag{2.6}$$

For streaming applications, the calculations simplifies to

$$D = D_{\mathrm{tx}} + D_{\mathrm{buf}}. \tag{2.7}$$

2.6.2 Packet Losses

In the transmission of packets on fixed and wireless packet networks, packets may get lost due to several reasons. In case of congestion in a part of the network, packets may have to be dropped at network nodes because of overflowing queues or buffers. If network components should fail completely, all packets still stored in their queues will be lost. Furthermore, impairments on the transmission channel, especially on wireless channels, will cause packet losses. Besides these "direct" losses, packets may also get lost "indirectly" by being discarded at the receiver when arriving too late.

Packet losses lead to the loss of the contained media frames. Some of these frames might be recovered through implemented FEC schemes. However, due to delay and data rate constraints, it may not be feasible to design these FEC schemes so that they can recover all losses. The general approaches of combating packet loss and recovering and estimating lost frames are discussed in Section 2.7. The optimal parameterization of forward error correction schemes and improved approaches for the concealment of lost frames are the central focus of this work.

2.6.3 Bit Errors on Wireless Transmission Channels

On wireless transmission channels, path loss, shadowing, and small scale fading lead to distortions in the received signal. Adaptive modulation schemes and advanced channel coding principles are incorporated to limit these effects but residual bit errors will remain in some transmission blocks after channel decoding. These errors are usually detected by a checksum implemented on the link layer. Some wireless transmission systems, e.g., HSPA, will initiate a retransmission of erroneous transmission blocks, otherwise the block has to be discarded. Depending on the considered link layer, these transmission blocks may contain a single IP packet, several IP packets, or also fractions of IP packets. If residual errors are detected, the complete transmission block is usually discarded leading to the loss of all IP

packets which have at least a fraction in this transmission block. In general, only error free IP packets are handed up to the network layer. Especially the packet headers, added by the transmission protocols, must not contain any bit errors to prevent an uncontrolled behavior, e.g., mis-routing or misinterpretation of packets. Some media codecs might tolerate some residual bit errors in a less important part of their bitstream which facilitates the use of *Unequal Error Detection* (UED) techniques on higher layers as further discussed in Chapter 2.7.5.

2.7 Techniques for Combating Packet Loss

The transmission of speech and audio signals over heterogeneous packet networks experiences packet losses and thereby losses of media frames. Packet losses occur due to transmission errors, network failures, or extensively delayed packets, as described in the previous section. Packet-based communication systems therefore need to implement techniques to combat the loss of signal frames, e.g., a receiver (jitter) buffer, forward error correction, and retransmission to limit the frame loss rate, as well as packet loss concealment techniques to mitigate the effect on the resulting quality. Extensive overviews of different strategies can be found in the literature, e.g., in [Perkins et al. 1998], [Wah et al. 2000], or [Lefebvre et al. 2004]. The following summary adopts the classification from [Perkins et al. 1998] into sender-driven repair methods and receiver-based error concealment schemes. It will be extended by a class of sender-assisted concealment approaches which basically reflect a combination of these approaches.

The design of robust packet based transmission systems for speech and audio applications is the central topic of this work. The following chapters of this work develop new approaches for optimizing redundant transmission schemes and improving receiver-based and sender-assisted packet loss concealment techniques.

2.7.1 Sender-driven Packet Loss Recovery

Sender-driven approaches limit the impact of packet loss by adding redundancy to following packets. These redundancy adding schemes can be further classified into media dependent and media independent approaches.

General Forward Error Correction (FEC) schemes, e.g., frame repetition or various types of block codes, belong to the *media independent* approaches and have been proposed for packet transmission in [Bolot et al. 1999], [Rosenberg et al. 2000], [Frossard 2001], [Jiang and Schulzrinne 2002], [Li 2007], among others. Such media independent error recovery schemes for packet transmission will be considered in Chapter 4 of this work, where new approaches for an optimal parameterization are developed based on a model of the transmission channel. An example for a *media dependent* approach is the additional transmission of a low-bit-rate version of each frame, so-called LBR (low bit-rate redundancy), i.e., the same frame encoded with a different codec or a different codec mode in case of multi-rate codecs. [Jiang

and Schulzrinne 2002] compared the two different concepts, transmission of redundancy derived with FEC and transmission of LBR. In their studies, FEC performed better than LBR. Other media dependent approaches only transmit a selective and/or partial redundancy in following packets. Here, *partial redundancy*, denotes the redundant transmission of only a subset with the most important parameters of a frame. *Selective redundancy*, on the other hand, is defined as the redundant transmission of information on only the most sensitive frames in the signal, i.e., transitional segments of the speech or music signal whose loss cannot be concealed easily and therefore would lead to a considerable impairment. These approaches therefore limit the additionally required data rate on the channel. Possible schemes have been proposed, e.g., in [Johansson et al. 2002],[Tosun and Kabal 2005].

Media independent FEC schemes can also be used for a partial protection of a frame's parameters. So-called *Unequal Error Protection* (UEP) schemes protect only the most important parameters or the most sensitive bits of these parameters. In case of packet losses, the remaining information which cannot be recovered will be estimated. The objective of these schemes is a graceful degradation of the quality with increasing losses in transmission systems. For example, [Horn et al. 1999] proposed such a scheme of UEP across packets for the application with scalable video coding.

The different techniques of sender-based loss recovery schemes that have been proposed in the literature differ in their requirements on additional data rate, increased delay, as well as computational complexity. The choice and parameterization therefore strongly depends on the demands of the considered application and the constraints of the network, possibly limiting the achievable error correction capabilities. These approaches therefore generally still require an receiver based approach for cases where the lost information cannot be completely recovered.

Finally, packet retransmission schemes, e.g., *automatic repeat request* (ARQ) techniques, might be employed if the round-trip time across the channel is low or the application tolerates a high delay. A more detailed discussion on the utilization of packet retransmission versus forward error correction is given in the following.

Forward Error Correction versus Retransmission

The error robustness of IP transmission on wireless packet channels can be increased by media independent error recovery schemes like *Forward Error Correction* (FEC), *Automatic Repeat reQuest* (ARQ), or hybrid schemes. ARQ mechanisms are closed-loop mechanisms based on the retransmission of packets which were not received at the destination or contain errors. FEC mechanisms on the other hand are open-loop mechanisms based on the transmission of redundant information so that some loss in the original information can be recovered at the receiver.

In a pure ARQ scheme, transmission errors are detected and for the corrupted or missing data block a retransmission is requested via a feedback channel. Because this retransmission may again be corrupted, several retransmission attempts may be required depending on the channel's error distribution. The maximum delay can therefore not be guaranteed when using ARQ. To limit the delay, the number

of retransmission attempts may be limited if the receiver can cope with lost data segments, e.g., by using concealment techniques. ARQ protocols further need to deal with the possibility that the feedback information about the reception or loss of a packet may get corrupted or lost itself. Standard ARQ protocols, e.g., *Selective Repeat ARQ*, are defined in [Comroe and Costello 1984], [Fairhurst and Wood 2002]. The acknowledged mode of the *Radio Link Control* (RLC) protocol in UMTS, for example, provides retransmissions of packets which are not acknowledged as correctly received. And also in the Wireless LAN standards, retransmission of packets is defined for unicast transmission streams.

When the network does not provide a feedback channel or a feedback would be too slow, e.g., the end-to-end transmission in a heterogeneous network, or more extreme, in satellite or deep space communication, ARQ cannot be used. In these cases, FEC has to be used for increasing the loss robustness. Another scenario in which ARQ is often also not feasible is the multicast streaming of music or video, especially when the multicast group size increases. The transmission channels to the receivers are assumed to be independent and therefore experience different loss patterns. This may lead to many retransmissions as a retransmission will already be initiated if only one of the receivers requests for it. Instead, FEC's open loop control is inherently suited to support large multicast groups.

The potential of FEC mechanisms depends to a large extent on the characteristics of the packet loss process in the network, favoring a small average number of consecutive losses. However, FEC increases the transmission data rate and thereby may lead itself to an increase of the loss rate on the channel. Furthermore, FEC is computationally demanding, so it is desirable to choose the simplest FEC code that still matches the requirements of the current transmission system.

Finally, hybrid schemes use a FEC code for error detection and correction and incorporate ARQ only if the correction fails, thereby leading to considerably less retransmissions than in pure ARQ systems. Furthermore, each transmission attempt may use a different encoding of the data which can then be combined at the receiver to yield an error free packet, even if all single transmissions have been corrupted. Such a *code combining* is, e.g., used in the hybrid ARQ (HARQ) scheme of UMTS HSPA (High Speed Packet Access) channels. However, the application of these hybrid schemes on a end-to-end basis in a heterogeneous network would still cause too much delay. As in the HARQ scheme of HSPA, these methods are therefore usually applied directly in layer 2 or layer 1 of a wireless transmission link and not on an end-to-end basis.

2.7.2 Receiver-based Packet Loss Concealment

Receiver-based concealment approaches aim at concealing the effect of a frame loss at the receiver. In the literature, such techniques are synonymously referred to

as *packet loss concealment* (PLC), *frame loss concealment*, or *frame erasure concealment*[4]. In circuit-switched transmission systems, a missing frame is usually recovered by parameter extrapolation (e.g., [3GPP TS 26.091]) or periodic pitch replication (e.g., [Gunduzhan and Momtahan 2001]). In packet networks, however, the use of a receiver buffer for compensating delay variations, as discussed in Section 2.7.4, may often lead to situations in which the packet following a lost one has already been received. The received future frames may then be utilized to conceal the missing frame which facilitates the use of more reliable parameter and waveform interpolation techniques instead of a simple extrapolation. See, e.g., [Johansson et al. 2002; Mertz et al. 2003; Fingscheidt and Perez 2002; Wang and Gibson 2001].

However, the performance of PLC algorithms has its limit. For speech signals, they are able to perform adequately up to loss lengths of 50-60 ms, i.e., when the loss starts to cover whole phonemes (cf., e.g., [Tobagi 2004]). Higher loss lengths cannot be concealed anymore without further information and a reduction of the naturalness and intelligibility of the speech may result. PLC algorithms therefore start to slowly mute the signal if the lost segment gets too long. Besides on the length of the lost segment, the performance strongly depends on the actual signal structure. For example, a lost segment within a vowel can be much easier reconstructed as a transitional segment from an unvoiced to a voiced sound.

In Chapter 6, a more detailed overview on state-of-the-art receiver-based PLC algorithms is given and a new approach is developed. This approach is based on a low complex approximation of lost CELP codec parameters which is dependent on the current signal structure around the lost segment.

2.7.3 Sender-assisted Packet Loss Concealment

Approaches which do not aim at completely recovering lost frames or parameters, but rather transmit specific additional information which can be utilized by the receiver's concealment routine, do not clearly belong to either of the previous classes. Therefore, a further class is introduced here which will be referred to as *sender-assisted* concealment approaches. In the literature, several approaches have been proposed which can be classified in this group, e.g., [Johansson et al. 2002; Lefebvre et al. 2004; Agiomyrgiannakis and Stylianou 2005; Tosun and Kabal 2005].

A suitable sender-assisted approach for the AMR speech codec is developed in Section 6.4 of this work (see also [Mertz and Vary 2006]), which generates low rate side information to assist the concealment. This information can be transmitted as hidden bit stream in the codec parameters as explained in Section 6.5 (cf. [Mertz and Vary 2008]).

The use of sender-assisted approaches is particularly aimed at medium or low rate codecs and transmission channels with a low available data rate, i.e., especially

[4] The term *frame erasure concealment* is mostly used in circuit-switched mobile communication systems like GSM, where some frames are rendered unusable by residual bit errors in the important bits, as detected by a physical layer CRC.

mobile application scenarios. Here, a low rate side information can achieve a considerable improvement of the quality and robustness of the transmission. On channels with higher available data rates, the use of FEC schemes becomes feasible which is able to recover complete frames and considerably reduce the resulting frame loss rate itself.

2.7.4 Receiver Buffer for Compensation of Jitter

The receiver in a packet-based transmission system incorporates a buffer to compensate delay variations which is therefore also called jitter buffer. This buffer holds the still encoded frames which have been extracted from the payload of the IP/UDP/RTP packets and delays them before further processing. If the system implements forward error correction schemes on packet level, or if other redundant information is transmitted in following packets, the buffer also needs to be large enough so that the redundant information can still be exploited when it arrives.

The length of the jitter buffer significantly contributes to the end-to-end delay of the transmission as given in (2.6) and (2.7). Hence, the maximum length of the buffer, $D_{\mathrm{buf,max}}$, depends on the delay demands of the application, i.e., the maximum tolerable end-to-end delay of the signal, $D_{\max}$:

$$D_{\mathrm{buf,max}} = D_{\max} - D_{\mathrm{s}} - D_{\mathrm{tx}}. \tag{2.8}$$

The length of the buffer will determine the amount of jitter that can be compensated: the longer the buffer, the more delay variation may exist, i.e., the more time each single packet has to arrive before it has to be considered lost. Therefore, the parameterization and possibly adaptation of the jitter buffer length needs to find an optimal trade-off between delay and loss rate.

For scenarios with tight constraints on the end-to-end delay, e.g., voice communication scenarios (telephony and video conferencing) with typically a maximum of 100-150 ms for high quality or up to 300 ms if constraints are a bit relaxed (cf. [ITU-T Rec. G.114 2003]), the tolerable delay may not be large enough to facilitate a length of the jitter buffer that would be necessary to keep the packet loss rate at an acceptable level. Therefore, techniques for a highly adaptive control of the jitter buffer length have been proposed in the literature, e.g., [Ramjee et al. 1994], [Moon et al. 1998], [Liang et al. 2001], [Sun and Ifeachor 2004], [Tobagi 2004]. These schemes try to follow the delay profile of the transmission closely and constantly adapt the buffer's length accordingly.

Packets that arrive too late are considered lost for the decoding process. However, [Gournay et al. 2003] proposed to still utilize these packets arriving just after their playout time to correct the states of the decoder, i.e., the parameter memories, in order to limit the error propagation.

This work will not consider adaptive receiver buffers, but a fixed length chosen to achieve the best trade-off between delay and loss rate (cf. Section 5.6). In Section 4.3 it will be shown that forward error correction on packet level can be used to recover delayed frames which otherwise would have to be considered lost. The application

of FEC thereby allows for a shorter buffer length and hence a shorter end-to-end delay.

2.7.5 Utilization of Packets with Residual Bit Errors

In currently employed packet transmission systems, bit errors anywhere in a transport block are detected by physical layer checksums and all erroneous transport blocks (protocol data units - PDUs) and the contained packets are discarded. This system behavior guarantees independence of the packets' contents and facilitates an end-to-end IP transmission without the need for intermediate decomposing and evaluation of packets. The result is a flexible network structure without considerable computational complexity at network nodes. The disadvantage is that also packets are discarded which could still be utilized at the receiver.

While packet headers have to be absolutely error free before the packets are forwarded in the protocol stack, this does not necessarily apply to the complete packet payload. The bits within a frame of encoded multimedia signals like speech, music, or video can be classified into groups of different importance for the resulting signal quality. These groups therefore exhibit different sensitivities towards bit errors. In the following, two concepts will be discussed for utilizing packets with residual bit errors at the receiver.

2.7.5.1 Discussion of Concepts

A first concept to utilize packets with residual bit errors is to discard only packets with errors in the packet headers and the most important part of the payload, while tolerating bit errors in unimportant payload parts. This can be achieved by applying *unequal error detection* (UED) on packet level using the UDP-Lite protocol, which has been standardized for this purpose by IETF in [Larzon et al. 2004]. As a modification of UDP, UDP-Lite provides a partial checksum coverage on the packet, covering at least the packet headers. This approach requires some system changes across several layers so that erroneous packets are handed from the physical layer up to the transport layer where they are evaluated. Note that the decomposition of the transport blocks into IP packets may already fail due to errors in the bits defining the packet boundaries. In this case, the packets are irrecoverable.

A second concept tries to utilize the complete payload by considering additional reliability information from the channel decoder. Any information on the likelihood of bit errors or even the reliability of every single bit might be exploited at the receiver, e.g., by applying *soft decision source decoding* (SDSD) techniques [Fingscheidt 1998; Fingscheidt et al. 1998]. Besides the payload data, it is also possible to extend this approach to the various packet headers. The reliability information would be utilized piecewise at the respective protocol layers. This approach requires further system changes compared to the first concept. Besides the possibly erroneous packets, the reliability information needs to be handed as meta data through the complete protocol stack from the physical layer upwards to the application.

This might only be feasible if the physical layer and the application, i.e., the media decoder, are located in the same device, e.g., a mobile phone using downlink transmission. Considering the uplink direction of a mobile access network, the IP packets are forwarded from the receiving base station through the provider's backbone and possibly via another wireless link down to a receiving terminal. Along this transmission path there is usually not enough capacity to transmit additional meta information of considerable bit rate. Furthermore, the second wireless link produces bit errors itself which would also affect the meta data. The evaluation of the reliability information from a wireless uplink transmission therefore needs to be already done at the receiving network entity after the radio transmission. Hence, a complete decoding of the media frames and subsequent re-encoding and packetization is required, contradicting the concept of end-to-end IP transmission without media dependent processing within the network nodes. This aspect and the considerable complexity overhead make this approach rather unfeasible, at least for the uplink direction of wireless channels.

Further concepts of utilizing bit level reliability information have been introduced for circuit-switched mobile networks. These approaches apply an iterative process of source and channel decoding (so called *Iterative Source-Channel Decoding* (ISCD) [Adrat 2003]) or additionally taking the demodulation into this loop (so-called *Turbo DeCodulation* [Clevorn 2006]). These approaches exchange extrinsic information between the different steps. The joint utilization of all available information is able to deliver considerable performance improvements. An application for packet transmission, however, would again require a complete decoding and re-encoding of the signal at the end of the wireless link in uplink direction.

2.7.5.2 Unequal Error Detection for Packet-Switched Channels

The approach with the least impact on the concept of an end-to-end IP transmission is the utilization of unequal error protection and detection. The application of UDP-Lite for unequal error detection in VoIP transmissions over UMTS has been investigated in [Mertz et al. 2005]. The simulation results show that the achievable performance gain (in terms of less discarded IP packets and therefore better speech quality) depends on the distribution of the residual bit errors after channel decoding and thus on the choice of the channel coding scheme: the application of UED can improve the quality for channels with convolutional channel coding but not for channels with Turbo coding. The different results were shown to be due to the fact that Turbo coding is more effective in correcting bit errors, which results in either error-free packets or packets with a high amount of residual errors. Convolutional coding, on the other hand, can correct fewer errors and leaves a certain amount of packets with only few bit errors, which can then benefit from the UED method, especially when header compression is applied to reduce the sensitive header size.

3

Model of the Packet Transmission Channel

The optimal design and parameterization of media transmission systems and services requires a sufficiently accurate model of the transmission channel and its effects on the transmitted data. Such effects can be transmission errors (bit errors or erasures of longer contiguous data blocks), fixed and variable transmission delays, as well as linear or non-linear signal distortions, e.g., echo. For the channel model in this work, only transmission errors and delay effects are considered.

Demands on the Channel Model

Transmission channels differ in their characteristics of transmission errors and transmission delays. A wireless GSM link, for example, can be described in detail with a model of residual bit errors and a fixed transmission delay. Internet connections, on the other hand, are characterized by IP packet losses and variable delays. The choice of a suitable channel model does however not only depend on the characteristics of the physical channel, but also on the characteristics of the considered application. For the description of the channel's effects on a specific application or service, a certain level of abstraction from the underlying physical channel is required. This abstraction refers to two different properties of the model, the assumed size of the transmitted data blocks and the timing of these data blocks, i.e., how often they are transmitted. The channel model may therefore rather describe snapshots of the channel at regular time points.

For packet-based speech or music transmission services, an abstraction level is required that describes the transmission of packets of a certain size at regular time intervals. For a specific packet transmission scenario, the desired channel model should describe the distribution of packet losses, irrespective of their origin of loss, i.e., whether they occur through transmission errors or delayed packets. The structure of the channel model itself may therefore be completely independent

from the underlying physical channel, i.e., whether it is a UMTS radio channel, a Wireless LAN link, or an end-to-end IP connection in the public Internet.

Packet size and transmission time interval of a packet stream of speech or music signals depend on the chosen system parameters, such as codec choice, frame length, frames per packet, and a possibly employed forward error correction scheme. For the optimal parameterization of packet-based speech and music transmission services as considered in this work, different parameter sets have to be compared which consequently differ in packet size and/or packet transmission time interval.

Depending on the transmission channel and the applied channel coding techniques, the probability of packet loss may depend on the size of the packets. On wireless packet channels, residual bit errors in transmission blocks are usually detected on the physical layer by use of a *Cyclic Redundancy Check* (CRC). In this case, smaller packets may have a lower residual bit error probability than larger ones. All erroneous packets are discarded and therefore lost.

This leads to an additional demand on the channel model. The level of abstraction needs to be adjustable in order to allow a comparison of different system parameterizations. In other words, the channel model needs to describe the transmission effects with a high resolution that can be downsampled for different application settings. For a sufficient description of the channel, the model does not necessarily need to reflect single bit errors, it can consider transmitted data blocks of a certain size and then determine whether such a block contains errors or not. The size of these data blocks should be the greatest common divisor of the considered packet lengths or smaller.

To summarize, a suitable channel model for the packet-based transmission of speech and music signals over heterogeneous networks needs to reflect packet losses and it needs to be flexible, i.e., adaptable to different packet sizes and packet transmission time intervals. Last but not least, the model has to be manageable, i.e., the mathematical description needs to be compact enough and has to allow the calculation of probabilities of specific loss patterns with feasible complexity.

A suitable model which meets the above demands is the generalized Gilbert-Elliott model [Elliott 1963], a two-state hidden Markov model with different loss probabilities in each state. Compared to its more widely applied simplification with pure 'loss' and 'no loss' states, it provides a greater flexibility in modeling various packet loss distributions of heterogeneous networks. The comparison of both model variants for different simulated wireless channels will show that the simplified version is indeed in certain cases not sufficient to describe the resulting loss distribution.

Chapter Outline

The Gilbert-Elliott model will be shortly reviewed in Section 3.1 and compared to alternative proposals from the literature. Section 3.2 then describes how this base model can be adapted to different packet transmission time intervals (TTI), which usually correspond to the frame length of the utilized speech or audio codec.

Subsequently, a novel procedure is developed for adapting the model to describe the loss behavior for packet streams of different packet sizes. Both adaptations together will provide the necessary basis for comparing different transmission configurations (i.e., codec frame lengths and data rates, as well as redundancy rates) on a given channel.

In the considered applications and services, variable transmission delays of packets may lead to packet losses if the total end-to-end delay budget of a packet is exceeded. Delay variations of packet networks can be modeled with the Weibull distribution (cf. [Sun and Ifeachor 2004]) and the additional packet loss probability can then be incorporated into the Gilbert-Elliott packet loss model according to Section 3.3.

The correct parameterization of the channel model requires detailed knowledge of the transmission channel which can be either gained by real life measurements or by appropriate system simulations. The measured or simulated error patterns are then used to train the parameters of the channel model as explained in Appendix E.1. Simulation results for WLAN and UMTS packet transmission channels are presented in Appendix E.2 and the respective channel models are developed. For a transmission scenario in heterogeneous networks, several trained models may be available for the different parts of the end-to-end transmission channel. An effective end-to-end channel model can be determined by concatenating these models as derived in Appendix G.

The proposed statistical packet loss model and its adaptations are able to reflect the observed behavior of fixed and wireless packet transmission channels and can be utilized for the development of robust transmission schemes and packet loss concealment algorithms. The probabilities of specific loss patterns which are needed for the determination of the forward error correction capabilities in Chapter 4 are derived in Section 3.4.

3.1 Modeling Packet Loss: Gilbert(-Elliott) Models and Alternatives

The loss process on a packet network is often of bursty nature, i.e., it shows a mixture of short and longer loss bursts (cf. Section 2.5). In this work, a burst is defined as the loss of one or more successive packets or data blocks. In the following, several statistical packet loss models are described which are commonly used in the literature to model packet loss distributions. The models will be discussed, first, with respect to their ability to model various loss distributions of different burstiness, and second, with respect to the availability of a manageable mathematical description for calculating probabilities of specific loss events. These probabilities are needed for the prediction of forward error correction capabilities in Chapter 4.

The considered channel models are used to describe packet losses and not single bit errors within packets. However, the models do not necessarily refer to real IP packets containing protocol headers and a defined payload. They can also describe

transmitted data blocks of a specific length which may be only a fraction of the size of a real IP packet. The model then describes whether this fraction is lost, e.g., because it contains residual bit errors. In Section 3.2 it will be explained how such a base model can be adapted to reflect a transmitted packet stream with different packet sizes and different transmission time intervals between successive packets.

3.1.1 Notation for Describing Packet Loss Distributions

The following notation will be used in describing the burst length distributions and loss probabilities of the considered channel models:

b	–	length of a loss burst, i.e., number of consecutively lost packets; assuming that directly preceding and following packets are received
$P_{\mathrm{b,s}}$	–	probability of a burst start, i.e., the occurrence of a packet loss after a received packet
$P_{\mathrm{b}}(b)$	–	probability that a single loss burst is of length b, i.e., that a loss event consists of b successively lost packets
$P_{\mathrm{bl}}(b)$	–	probability of occurrence of a loss burst with length b, i.e., of b successively lost packets
$\overline{b}$	–	mean burst length
P_{fl}	–	mean packet loss probability

The distribution of burst lengths can be sufficiently described by the probability of a burst start $P_{\mathrm{b,s}}$ and the probability of a single burst having length b, denoted by $P_{\mathrm{b}}(b)$. Combining both, the probability of occurrence of a burst with length b is calculated as

$$P_{\mathrm{bl}}(b) = P_{\mathrm{b,s}} \cdot P_{\mathrm{b}}(b). \tag{3.1}$$

The expected mean burst length $\overline{b}$ can be either calculated from $P_{\mathrm{b}}(b)$ as

$$\overline{b} = \sum_{b=1}^{\infty} b\, P_{\mathrm{b}}(b), \tag{3.2}$$

or, if $P_{\mathrm{b,s}}$ and the expected mean packet loss probability P_{fl} are known, as

$$\overline{b} = \frac{P_{\mathrm{fl}}}{P_{\mathrm{b,s}}}. \tag{3.3}$$

The calculation of the probabilities $P_{\mathrm{b,s}}$, $P_{\mathrm{b}}(b)$, and P_{fl} depends on the considered model and is described in the respective following sections[1].

[1] The calculations of these probabilities are in general known in the literature for the considered models. They shall be reviewed here in the respective notation of this work before a novel approach for adaptation of the model to different packet sizes is introduced.

The channel models itself are defined by their respective model parameters. All considered models consist of one or multiple states with different transition probabilities between the states. Depending on the model, not all transitions are allowed. Each state has a different probability of packet loss. The following parameters are used to describe the different channel models:

- $P_{\mathrm{t},ij}$ – state transition probability from state i to state j
- $P_{\mathrm{e},i}$ – probability of a packet loss in state i
- $P_{\mathrm{s},i}$ – probability that the channel is currently in state i

3.1.2 The Bernoulli Model for Independent Packet Losses

The Bernoulli model describes a process of independent packet losses of a specified packet loss probability $P_{\mathrm{fl}} = P_{\mathrm{e}}$. This model is therefore insufficient in describing the burst behavior of packet losses which are observed, e.g., for wireless links. However, it may still suffice for other transmission channels, e.g., local area networks (LANs).

For the Bernoulli model, the probability of a burst start is determined with the specified loss probability P_{e} as the probability of occurrence of a received packet (no loss), followed by a lost packet:

$$P_{\mathrm{b,s}} = (1 - P_{\mathrm{e}}) \cdot P_{\mathrm{e}}. \tag{3.4}$$

The distribution of burst lengths, i.e., the probability that a burst is of length b, is given as the probability of $b-1$ successive losses after assuming a burst start, followed by a single received packet determining the end of the burst:

$$P_{\mathrm{b}}(b) = P_{\mathrm{e}}^{b-1} \cdot (1 - P_{\mathrm{e}}). \tag{3.5}$$

The probability of occurrence of a burst with length b is then calculated as

$$P_{\mathrm{bl}}(b) = P_{\mathrm{b,s}} \cdot P_{\mathrm{b}}(b) = (1 - P_{\mathrm{e}})^2 \cdot P_{\mathrm{e}}^b. \tag{3.6}$$

Finally, the mean burst length of a Bernoulli packet loss process is given as

$$\begin{aligned} \overline{b} &= \sum_{b=1}^{\infty} b \cdot P_{\mathrm{b}}(b) = \sum_{b=1}^{\infty} b \cdot P_{\mathrm{e}}^{b-1} \cdot (1 - P_{\mathrm{e}}) = \frac{1 - P_{\mathrm{e}}}{P_{\mathrm{e}}} \cdot \sum_{b=1}^{\infty} b \cdot P_{\mathrm{e}}^b \\ &= \frac{1 - P_{\mathrm{e}}}{P_{\mathrm{e}}} \cdot \frac{P_{\mathrm{e}}}{(1 - P_{\mathrm{e}})^2} = \frac{1}{1 - P_{\mathrm{e}}}. \end{aligned} \tag{3.7}$$

The single parameter of the Bernoulli model, the error (loss) probability P_{e}, can be determined easily from given measurements, i.e., from recorded loss traces.

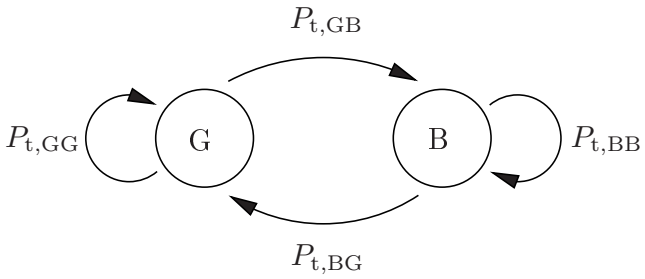


Figure 3.1: Gilbert(-Elliot) Model: 2-state model with specified transition probabilities $P_{t,ij}$ and different loss probabilities $P_{e,i}$ in state G and B ($i, j \in \{G, B\}$) with $P_{e,G} \ll P_{e,B}$

3.1.3 Gilbert(-Elliott) Models for Bursty Packet Losses

The Gilbert(-Elliott) model [Gilbert 1960; Elliott 1963] is a 2-state Markov model as defined in Fig. 3.1. Originally utilized to model burst-noise binary channels, it is also widely used to describe bursty packet loss distributions. The two states of the model differ in their loss probability. State G (good) has a low loss probability $P_{e,G}$ and state B (bad) has a higher loss probability $P_{e,B}$. For each transmitted packet, a state transition is made according to the given transition probabilities $P_{t,ij}$ with $i, j \in \{G, B\}$. The new state then determines the probability that the considered packet is lost. With these two states the model is able to reflect a loss behavior which consists of periods with high loss densities and periods with no losses or low loss densities.

In the literature, the Gilbert(-Elliott) model is used in several forms differing in their degree of simplification:

Gilbert model In the original model proposed by Gilbert in [Gilbert 1960], the loss probability in state G is set to $P_{e,G} = 0$, i.e., G becomes a lossless state. The loss probability in state B, however, is variable in the range $0 < P_{e,B} \leq 1$, i.e., not every packet is necessarily lost in state B.

Simplified Gilbert model In a simplified version of the Gilbert model, the loss probabilities in states G and B are set to $P_{e,G} = 0$ and $P_{e,B} = 1$, respectively. Thereby, the two states are becoming a lossless and a loss state, and the loss distribution is only controlled by the state transition probabilities. Due to its simplicity, this variant is most often found in the literature for packet loss modeling.

Generalized Gilbert model or Gilbert-Elliott model The highest flexibility for modeling different loss distributions is achieved by allowing arbitrary values for the loss probabilities of the two states according to $0 \leq P_{e,G} \ll P_{e,B} \leq 1$, as proposed by Elliott in [Elliott 1963].

The properties and limitations of these variants are discussed in more detail in the following, starting with the most general form, the Gilbert-Elliott model.

3.1.3.1 Generalized Gilbert Model (Gilbert-Elliott Model)

A generalized form of the Gilbert model has been proposed by Elliott in [Elliott 1963], which later on has often been referred to as Gilbert-Elliott model. The error probability in the good state G is now allowed to be greater than zero, i.e., $0 \leq P_{\mathrm{e,G}} \ll P_{\mathrm{e,B}} \leq 1$. The Gilbert-Elliott model is usually parameterized such that state G produces rare, single packet losses, whereas state B generates loss bursts.

The probabilities of being in state G or B can be determined from the state transition probabilities:

$$P_{\mathrm{s,G}} = \frac{P_{\mathrm{t,BG}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}} \quad \text{and} \quad P_{\mathrm{s,B}} = \frac{P_{\mathrm{t,GB}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}}. \tag{3.8}$$

With the transition probabilities and the loss probabilities of the states, the mean packet loss rate P_{fl} can be computed by considering the occurrence and error probabilities of both states:

$$P_{\mathrm{fl}} = P_{\mathrm{s,G}} \cdot P_{\mathrm{e,G}} + P_{\mathrm{s,B}} \cdot P_{\mathrm{e,B}} = \frac{P_{\mathrm{t,BG}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}} \cdot P_{\mathrm{e,G}} + \frac{P_{\mathrm{t,GB}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}} \cdot P_{\mathrm{e,B}}. \tag{3.9}$$

The probability of a burst start is then given as

$$P_{\mathrm{b,s}} = P_{\mathrm{s,G}} \cdot (1 - P_{\mathrm{e,G}}) \cdot (P_{\mathrm{t,GG}}\, P_{\mathrm{e,G}} + P_{\mathrm{t,GB}}\, P_{\mathrm{e,B}}) \tag{3.10}$$

$$+ P_{\mathrm{s,B}} \cdot (1 - P_{\mathrm{e,B}}) \cdot (P_{\mathrm{t,BG}}\, P_{\mathrm{e,G}} + P_{\mathrm{t,BB}}\, P_{\mathrm{e,B}}). \tag{3.11}$$

Given (3.9) and (3.11), the average length of a loss burst can be calculated as

$$\overline{b} = \frac{P_{\mathrm{fl}}}{P_{\mathrm{b,s}}}. \tag{3.12}$$

The calculation of the burst length distribution, i.e., the probability $P_{\mathrm{b}}(b)$, is more complex than for the other two variants, because losses may occur in both states. Therefore, a burst possibly includes several state changes. The derivation of the distribution will be given in Section 4.1.4.

The model parameters, i.e., the state transition and error probabilities, can be determined from recorded loss traces using the training procedure described in Appendix E.1.

3.1.3.2 Gilbert Model

The original Gilbert model, proposed by Gilbert in [Gilbert 1960] to model a burst-noise binary channel, sets the loss probability in the good state to zero ($P_{\mathrm{e,G}} = 0$). For an erroneous channel, the error probability in the bad state is greater than zero, but may be smaller than one ($0 < P_{\mathrm{e,B}} \leq 1$).

To simulate burst noise, the states G and B must tend to persist. Therefore, the transition probabilities $P_{\mathrm{t,GB}}$ and $P_{\mathrm{t,BG}}$ need to be small. The run lengths of the state sequences G and B have geometric distributions with means $1/P_{\mathrm{t,GB}}$ for the G-runs and $1/P_{\mathrm{t,BG}}$ for the B-runs. If it can be assumed that burst events are

statistically independent of each other, their distance in time will have a geometric distribution, and therefore the model's geometric distribution of G-runs can be justified [Gilbert 1960]. However, a geometric distribution of B-runs is only justified by mathematical simplicity and may not result from real measurements. If $P_{\mathrm{e,B}} < 1$, the distribution of burst lengths does not directly correspond to that of the B-runs, because not all packets in state B are lost. Hence, the distribution of burst lengths can to a certain extent be shaped by an appropriate choice of the loss probability $P_{\mathrm{e,B}}$.

For the Gilbert model, the mean packet loss rate defined in (3.9) simplifies to

$$P_{\mathrm{fl}} = P_{\mathrm{s,B}} \cdot P_{\mathrm{e,B}} = \frac{P_{\mathrm{t,GB}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}} \cdot P_{\mathrm{e,B}}. \tag{3.13}$$

The probability of a burst start, i.e., the probability of receiving a packet and losing the next, is given as

$$P_{\mathrm{b,s}} = P_{\mathrm{s,G}} \cdot P_{\mathrm{t,GB}} \cdot P_{\mathrm{e,B}} + P_{\mathrm{s,B}} \cdot (1 - P_{\mathrm{e,B}}) \cdot P_{\mathrm{t,BB}} \cdot P_{\mathrm{e,B}}. \tag{3.14}$$

For the calculation of the burst length distribution a burst start is assumed, i.e., a lost packet in state B. For a burst length b, the channel then needs to remain in state B for $b-1$ packets, all of which have to be lost. Finally, the channel either changes to state G or it remains in state B and receives the next packet, which ends the burst:

$$P_{\mathrm{b}}(b) = P_{\mathrm{t,BB}}^{b-1} \cdot P_{\mathrm{e,B}}^{b-1} \cdot (P_{\mathrm{t,BG}} + P_{\mathrm{t,BB}}(1 - P_{\mathrm{e,B}})). \tag{3.15}$$

The probability of occurrence of a burst of length b is then given by:

$$P_{\mathrm{bl}}(b) = P_{\mathrm{b,s}} \cdot P_{\mathrm{b}}(b). \tag{3.16}$$

Finally, the mean burst length can be computed as

$$\overline{b} = \frac{P_{\mathrm{fl}}}{P_{\mathrm{b,s}}}. \tag{3.17}$$

If $P_{\mathrm{e,B}} < 1$, it is impossible to reconstruct the state sequence directly from a given error sequence. Instead, a training algorithm needs to be employed to estimate the model parameters (cf. Appendix E.1).

3.1.3.3 Simplified Gilbert Model

The variant most often seen in the literature is a simplified version of the original Gilbert model, where the loss probability in state B is fixed to $P_{\mathrm{e,B}} = 1$. Therefore, a packet is never lost in state G, because $P_{\mathrm{e,G}} = 0$, and it is always lost in state B, since $P_{\mathrm{e,B}} = 1$.

The mean packet loss rate P_{fl} of the simplified Gilbert model, which is sometimes referred to as the unconditional loss rate P_{u} in the literature, now becomes

$$P_{\mathrm{fl}} = P_{\mathrm{u}} = \frac{P_{\mathrm{t,GB}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}}. \tag{3.18}$$

The probability of transition from the loss state to itself is then referred to as conditional loss probability P_c:

$$P_c = 1 - P_{t,BG} = P_{t,BB}. \tag{3.19}$$

As in the original Gilbert model, the residence times for states G and B are geometrically distributed. For this model, however, also the burst and gap lengths are geometrically distributed, as they directly correspond to the state sequence. The distribution of burst lengths, i.e., the probability of a given burst having a length of b successive packets, is given by the probability of staying in the loss state for $b-1$ times and then changing to the no-loss state, which ends the burst:

$$P_b(b) = P_{t,BB}^{b-1} P_{t,BG} = (1 - P_{t,BG})^{b-1} P_{t,BG}. \tag{3.20}$$

The probability of a burst start, i.e., being in state G and then changing to state B, computes to

$$P_{b,s} = P_{s,G} \cdot P_{t,GB} = \frac{P_{t,BG}}{P_{t,GB} + P_{t,BG}} \cdot P_{t,GB}. \tag{3.21}$$

The probability of occurrence of a burst of length b is then given by:

$$P_{bl}(b) = P_{b,s} \cdot P_b(b) = \frac{P_{t,BG}}{P_{t,GB} + P_{t,BG}} \cdot P_{t,GB} \cdot P_{t,BB}^{b-1} \cdot P_{t,BG}. \tag{3.22}$$

Finally, the mean burst length can be computed in the same way as in (3.7):

$$\overline{b} = \sum_{n=1}^{\infty} n \cdot P_b(n) = \frac{1}{P_{t,BG}}. \tag{3.23}$$

The simplified Gilbert model is the mathematically least complex of the variants. The state sequence of this model is always reconstructible from a given error sequence, because the state alone determines whether a packet is received or lost. This allows an easy determination of the parameters from a given measurement trace. However, the simplified Gilbert model is restricted to a geometrical distribution of the burst lengths which does not necessarily correspond to real network scenarios.

A special case of the simplified Gilbert model results when $P_{t,GB} + P_{t,BG} = 1$ applies. The model then turns into a Bernoulli model with independent losses of rate $P_{fl} = P_{t,GB}$.

3.1.4 Alternative Channel Model: 4-State Markov Model

Some alternative channel models have also been proposed in the literature for modeling packet loss distributions. Compared to the Gilbert(-Elliott) models, Markov models with a higher order than two may provide a higher accuracy in modeling a

measured loss distribution. A widely used example of a four-state Markov model is defined and used in [ITU-T Rec. G.1020 2006; Clark 2001; ETSI TIPHON TS 101 329-5 Annex E] and shown in Figure 3.2. This model consists of two 2-state models representing periods of high loss rate (burst period) and periods with no or single packet losses (gap periods), defined by the following four different states:

1: packet is received successfully within a gap period
2: packet is received successfully within a burst period
3: packet is lost within a burst period
4: packet is lost within a gap period (isolated packet loss)

The definition of bursts and gaps for this model differs from that applied in the previous section. In this case, not all packets might be lost in a burst and not all are necessarily received in a gap period.

For the derivation of the model parameters from the measured loss traces, a minimum gap length $G_{\min}$ has to be defined. This parameter defines the minimum number of consecutive packets that need to be received after a lost packet in the burst period, such that these received packets are assigned to a gap period. A gap ends as soon as there is a loss of at least 2 successive packets. The model parameters and the state sequence can then be determined directly from the loss traces as explained in [ITU-T Rec. G.1020 2006].

Like the generalized Gilbert-Elliott model (Section 3.1.3.1), the 4-state Markov model is also able to model a large variety of packet loss distributions. The calculation of probabilities of specific loss patterns, however, is considerably more complex than for the Gilbert-Elliott model. The simulations of wireless channels in Appendix E.2 will show that the generalized Gilbert-Elliott model is sufficient in modeling the observed loss behavior. It has therefore been chosen for the studies of this work.

3.2 Extended Gilbert-Elliott Model Considering Various Transmission Time Intervals and Packet Sizes

One of the central aspects of this work is the optimal parameterization of packet-based speech and music transmission systems in heterogeneous network scenarios with wireless access. For this purpose, several packet level forward error correction (FEC) schemes will be analyzed in Chapter 4 with respect to their error correction capabilities. The application of forward error correction on packet level results in additional packets for the FEC data or in increased packet sizes if the FEC data is piggybacked to the original media packets. Hence, the transmission time interval and the packet size vary among different considered methods. For most packet transmission channels, the experienced distribution of packet losses is influenced by these two parameters. On wireless channels, e.g., the probability of a packet

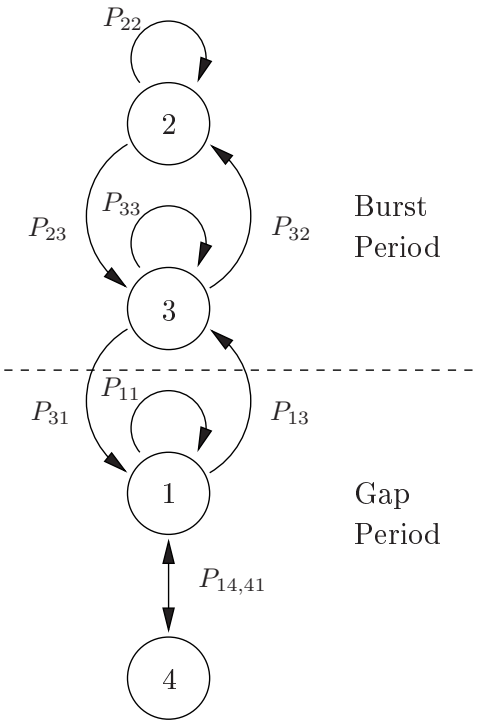


Figure 3.2: 4-state Markov Model: 4-state model with specified transition probabilities P_{ij} $(i, j \in \{1, 2, 3, 4\})$. Packet is received in states 1 and 2; packet is lost in states 3 and 4.

loss usually increases with an increasing packet size. Channels with bursty loss behavior cause longer burst losses, i.e., higher number of successively lost packets, if the transmission time interval between the packets is short. Thus, each considered scheme requires a specifically tailored channel model reflecting the appropriate transmission time interval and packet size. A separate measurement or simulation and subsequent training for all possible scenarios would be far too complex.

Therefore, a flexible base channel model is required which can be adapted to different packet sizes and transmission time intervals. The generalized Gilbert-Elliott model is a suitable basis of such a flexible model. The following sections will review the procedure of adapting the transmission time interval of this model and develop a novel approach for adapting the model to different packet sizes. The adaptation will be subject to the constraint that the base model is of sufficient resolution. The resulting extended Gilbert-Elliott model will be used as a basis for the prediction of residual loss distributions in connection with different FEC schemes in Chapter 4.

3.2.1 Model Adaptation for Multiples of the Transmission Time Interval

First, an increase of the transmission time interval from the original value of the channel model T'_{TTI} to a new value of T_{TTI} is assumed, i.e., an increase by the factor

$$k_{\mathrm{t}} = \frac{T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}}. \tag{3.24}$$

The new transmission time interval needs to be an integer multiple of the original value, such that the factor k_t becomes an integer. The adaptation of the channel model parameters is only possible if this constraint is met. At first, no increase in the packet size is assumed. Then, the state transition probabilities of the new effective channel model, denoted with the superscript (k_t), can be derived from the original values as follows [Elliott 1963]:

$$P_{\mathrm{t,GB}}^{(k_\mathrm{t})} = \frac{P_{\mathrm{t,GB}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}} \cdot \left(1 - (P_{\mathrm{t,GG}} - P_{\mathrm{t,BG}})^{k_\mathrm{t}}\right) \tag{3.25a}$$

$$P_{\mathrm{t,BG}}^{(k_\mathrm{t})} = \frac{P_{\mathrm{t,BG}}}{P_{\mathrm{t,GB}} + P_{\mathrm{t,BG}}} \cdot \left(1 - (P_{\mathrm{t,GG}} - P_{\mathrm{t,BG}})^{k_\mathrm{t}}\right) \tag{3.25b}$$

$$P_{\mathrm{t,GG}}^{(k_\mathrm{t})} = 1 - P_{\mathrm{t,GB}}^{(k_\mathrm{t})} \tag{3.25c}$$

$$P_{\mathrm{t,BB}}^{(k_\mathrm{t})} = 1 - P_{\mathrm{t,BG}}^{(k_\mathrm{t})} \tag{3.25d}$$

The error probabilities of the two states are unaffected by the change of the transmission time interval:

$$P_{\mathrm{e,G}}^{(k_\mathrm{t})} = P_{\mathrm{e,G}} \tag{3.26a}$$

$$P_{\mathrm{e,B}}^{(k_\mathrm{t})} = P_{\mathrm{e,B}} \tag{3.26b}$$

The same applies to the probabilities of being in either of the states $P_{\mathrm{s,G}}$ and $P_{\mathrm{s,B}}$, and therefore, the overall loss rate P_fl of the channel model remains unchanged as well. What changes is the distribution of loss lengths due to the modified state transition probabilities.

3.2.2 Model Adaptation for Arbitrary Packet Sizes

In the following, a novel approach for adapting the channel model to account for different packet sizes is developed. On channels with a constant transmission rate[2], the packet size is directly proportional to the packet transmission time τ_p, which is defined here as the quotient of packet size L_p and transmission rate on the channel R_ch:

$$\tau_\mathrm{p} = \frac{L_\mathrm{p}}{R_\mathrm{ch}}. \tag{3.27}$$

The packet transmission time should not be confused with the end-to-end delay of a packet transmission.

[2] A constant transmission rate is assumed for the channels considered in this work. In case of heterogeneous networks with several transmission links, the transmission rate of the effective channel is determined by the link which has the lowest transmission rate.

The adaptation of the model for different packet sizes 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:

$$k_\mathrm{p} \approx \tau_\mathrm{p}/\tau'_\mathrm{p}. \tag{3.28}$$

Second, the base transmission time τ'_p needs to be equal to the base transmission time interval T'_TTI, i.e., the base model has to describe the channel at 100 % utilization:

$$\tau'_\mathrm{p} \stackrel{!}{=} T'_\mathrm{TTI}. \tag{3.29}$$

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 lose 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 is 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 are therefore denoted by the superscript $(k_\mathrm{t}, k_\mathrm{p})$:

$$P_{\mathrm{e},XY}^{(k_\mathrm{t},k_\mathrm{p})} = \frac{1}{P_{\mathrm{t},XY}^{(k_\mathrm{t})}} \sum_{Z\in\{\mathrm{G,B}\}} \sum_{i=1}^{k_\mathrm{p}} P_{XZ}(i,k_\mathrm{p})\; P_{\mathrm{t},ZY}^{(k_\mathrm{t}-k_\mathrm{p})} \tag{3.30}$$

Here, the error probability $P_{\mathrm{e},XY}^{(k_\mathrm{t},k_\mathrm{p})}$ for a transition from state X to Y, with $X, Y \in \{\mathrm{G,B}\}$, is calculated using the following terms:

- $\sum_{i=1}^{k_\mathrm{p}} P_{XZ}(i,k_\mathrm{p})$ is the probability of at least 1 and up to k_p errors in k_p successive packets according to the original channel model; the channel is in state X at the first packet and in state Z at the $(k_\mathrm{p}+1)$th packet, with $X, Z \in \{\mathrm{G,B}\}$.
- $P_{\mathrm{t},ZY}^{(k_\mathrm{t}-k_\mathrm{p})}$ is the probability of transition from state Z at the $(k_\mathrm{p}+1)$th packet to state Y at the first packet of the following new transmission time interval T_TTI, with $Z, Y \in \{\mathrm{G,B}\}$.
- The sum over $Z \in \{\mathrm{G,B}\}$ covers both possible intermediate states at the $(k_\mathrm{p}+1)$th packet.
- $P_{\mathrm{t},XY}^{(k_\mathrm{t})}$ is the probability of transition from state X to Y for two successive packets at the distance of the new transmission time interval T_TTI.

The transition probabilities $P_{\mathrm{t},XY}^{(k)}$ are calculated according to (3.25a-d) with the given increase factor. The probabilities of i losses in k_p packets, $P_{XY}(i,k_\mathrm{p})$, are calculated as will be explained in Section 3.4.

3.2.3 Examples for the Channel Model Adaptation

Table 3.1 shows some exemplary cases for increasing the transmission time interval (from first to second row), for increasing the packet transmission time (from second to third row) or for applying both at the same time (from first to fourth row). The loss behavior of the resulting packet stream can be derived from the extended Gilbert-Elliott channel model using (3.25a-d) and (3.30) with the according values of k_t and k_p. Note that the base model for the adaptations describes the channel at 100% utilization as required, i.e., the packet transmission time interval equals the packet transmission time of a single packet.

k_t	k_p	Packet Transmission Time Line								
1	1	p1	p2	p3	p4	p5	p6	p7	p8	...
2	1	p1		p2		p3		p4		...
2	2	p1		p2		p3		p4		...
3	2	p1			p2			p3		...

Table 3.1: Examples of channel model adaptation for increasing transmission time interval and/or packet transmission time. Packet sizes and hence transmission times are indicated by gray background.

3.3 Modeling Varying Transmission Delay (Jitter)

In the transmission over packet networks, not all packets necessarily experience the same transmission delay. There are several reasons for such delay variations, also called jitter, which have been already discussed in Section 2.6.1.

Several distributions have been used in the literature to model the varying transmission delay. The choice of the best suitable distribution depends on the considered network and application. Measurements of [Sun and Ifeachor 2004] have shown that in general the Weibull distribution leads to a better fit for VoIP traces than Pareto and exponential distributions, where the latter is included as a special case in the Weibull distribution. The Weibull distribution is therefore considered for modeling delay variation in this work. In general, the choice of a suitable distribution may vary for different considered channels and applications. Since the general methodology described in this work is not limited to the Weibull distribution, it can also be used with other distributions for specific network and application characteristics if required.

In time sensitive applications, delay variations may lead to packet losses if the delay of a packet exceeds its scheduled playout time. The probability of packet loss caused by delay depends on the available delay budget of the packet. After a short review of the Weibull distribution, the calculation of this loss probability and its integration into the packet loss channel model will be explained.

3.3.1 The Weibull Distribution for Modeling Jitter

The Weibull distribution is defined by its probability density function (PDF)

$$f_X(x) = \gamma\,\alpha^{-\gamma}\,(x-\mu)^{\gamma-1}\,e^{-\left(\frac{x-\mu}{\alpha}\right)^{\gamma}}, \tag{3.31}$$

or its cumulative distribution function (CDF)

$$F_X(x) = P(X \leq x) = 1 - e^{-\left(\frac{x-\mu}{\alpha}\right)^{\gamma}}. \tag{3.32}$$

For describing the delay variation of packet transmission, the parameters of the distribution have the following meanings: The end-to-end delay of a packet is denoted by x. The parameter μ describes the minimum transmission delay of each packet, which depends on the considered channel, the distance between sender and receiver, etc. Delay values below this minimum delay are not possible, i.e., $f_X(x) = 0$ and $F_X(x) = 0$ if $x < \mu$. Finally, α and γ are the general scale and shape parameters of the Weibull distribution and therefore determine the actual form of the distribution depending on the network's delay characteristic. The Weibull distribution contains the exponential distribution as special case for $\gamma = 1$.

Example for VoIP traces in the Internet

The example shown in Figure 3.3 has been taken from [Sun and Ifeachor 2004]. The figure shows the PDF and CDF of the delay variation derived from measurements of VoIP traces from Plymouth, UK, to Beijing, China. For these traces, the parameters of the Weibull distribution have been determined as $\mu = 116\,\text{ms}$, $\alpha = 15.9$, $\gamma = 0.4451$.

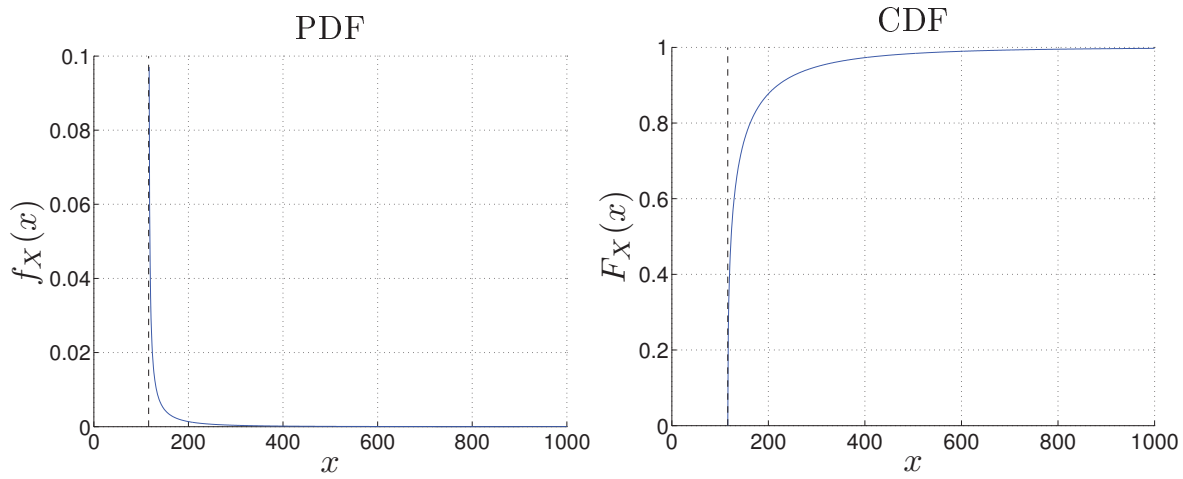


Figure 3.3: Weibull distribution modeling the measured delay variation of VoIP traces from Plymouth, UK, to Beijing, China, taken from [Sun and Ifeachor 2004]. Weibull parameters: $\mu = 116\,\text{ms}$, $\alpha = 15.9$, $\gamma = 0.4451$

3.3.2 Packet Loss Due to Jitter Depending on Receiver Buffer Length

Besides losses caused by transmission errors and buffer overflows, packets can also be lost due to jitter. In this case, the packet is still arriving at the receiver, however, it arrives to late to be considered. The amount of losses due to jitter depends on the chosen jitter buffer length l_{JB} at the receiver. The larger the buffer, the more time does a packet have to arrive. However, for most applications, especially for conversational services like Voice over IP or video telephony, the tolerable end-to-end delay is limited. Therefore, the length of the jitter buffer is also limited, since it determines the resulting end-to-end delay of the service. Depending on the jitter buffer length available at the receiver, l_{JB}, the probability of packet loss due to jitter, i.e. the probability of a packet arriving too late, is given by

$$P_{l,J} = P(\tau_{\mathrm{p}} > l_{\mathrm{JB}}) = 1 - F_X(l_{\mathrm{JB}}). \tag{3.33}$$

3.3.3 Incorporating Jitter Losses into the Gilbert-Elliott Packet Loss Model

The Gilbert-Elliott model described in Section 3.1.3 models end-to-end packet losses which include jitter based losses only if the model is trained for a specific transmission scenario with a fixed jitter buffer length at the receiver. In this case, however, the model cannot be used for optimizing the jitter buffer length itself. Hence, an alternative approach is needed for incorporating jitter based losses into the model while retaining the jitter buffer length as a parameter. Such an approach shall be developed in the following.

The Gilbert-Elliott channel model is trained without consideration of losses due to jitter, i.e., a receiver jitter buffer of indefinite length is assumed. The packet losses due to jitter can afterwards be incorporated into the model by modification of the error probabilities, which then become dependent on the chosen jitter buffer length l_{JB}. The new loss probabilities for the two states are defined by the events of packet loss due to transmission errors and packet loss due to delay. These two events are not disjoint, i.e., a packet could be delayed so that it would have to be regarded lost and at the same time it may also be affected by transmission errors. In the following, it is assumed that these events are statistically independent. Hence, the loss probabilities of the two states in the Gilbert-Elliott model become

$$P'_{\mathrm{e,G}} = P_{\mathrm{e,G}} + P_{l,J} - P_{\mathrm{e,G}} \cdot P_{l,J}, \tag{3.34a}$$

$$P'_{\mathrm{e,B}} = P_{\mathrm{e,B}} + P_{l,J} - P_{\mathrm{e,B}} \cdot P_{l,J}. \tag{3.34b}$$

The resulting total loss rate then becomes

$$P'_{\mathrm{e}} = P_{\mathrm{s,G}} \cdot P'_{\mathrm{e,G}} + P_{\mathrm{s,B}} \cdot P'_{\mathrm{e,B}}. \tag{3.35}$$

The assumption of statistical independence implies that transmission errors and delay variations are of independent origin. In case of network congestion, delay and loss due to buffer overflows may not be completely independent. However, this dependency is not considered in this work.

3.4 Deriving Probabilities of Loss Patterns from Model Parameters

For evaluating the expected quality of multimedia transmission in packet networks, the probabilities of occurrence of specific loss patterns are required. Particularly in conjunction with forward error correction schemes, the detailed knowledge of such probabilities is necessary to predict the residual loss distribution after error correction. In the following, the calculation of these probabilities is explained for the generalized Gilbert-Elliott model from Section 3.1.3 following the derivation in [Elliott 1963]. Subsequently, the probabilities are derived for the extended model introduced in Section 3.2, thereby providing the means for a reliable comparison of packet streams with different packet sizes and transmission time intervals. The derived probabilities are utilized in the evaluation of different forward error correction schemes as will be discussed in Chapter 4.

3.4.1 Generalized Gilbert-Elliott Model

Consider a group of n successive packets. The probability of m losses in these n consecutive packets will be denoted as $P_{XY}(m,n)$, with the two subscripts indicating the channel state X at the first packet and state Y at the packet directly following the group of packets, i.e., the $(n+1)$-th packet. X and Y can assume either one of the model states, G and B, i.e., $X, Y \in \{\mathrm{G}, \mathrm{B}\}$.

The probability of m losses in n consecutive packets with arbitrary channel states at the first and following packets can be calculated according to [Elliott 1963] as

$$\begin{aligned} P(m,n) &= P_{\mathrm{s,G}}\big(P_{\mathrm{GG}}(m,n) + P_{\mathrm{GB}}(m,n)\big) + P_{\mathrm{s,B}}\big(P_{\mathrm{BG}}(m,n) + P_{\mathrm{BB}}(m,n)\big) \\ &= \sum_{\substack{X,Y \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P_{XY}(m,n) \end{aligned} \tag{3.36}$$

with the probabilities to be in a certain state, $P_{\mathrm{s},X}$, $X \in \{\mathrm{G}, \mathrm{B}\}$, calculated as given in (3.8), and the following conditional probabilities:

- $P_{\mathrm{GG}}(m,n)$: Pr(m losses in n packets; in G at $n+1$-th packet | start in G),
- $P_{\mathrm{GB}}(m,n)$: Pr(m losses in n packets; in B at $n+1$-th packet | start in G),
- $P_{\mathrm{BG}}(m,n)$: Pr(m losses in n packets; in G at $n+1$-th packet | start in B),
- $P_{\mathrm{BB}}(m,n)$: Pr(m losses in n packets; in B at $n+1$-th packet | start in B).

The probabilities $P_{XY}(m,n)$, $X, Y \in \{\mathrm{G}, \mathrm{B}\}$, can be derived by recursive calculation:

$$\begin{aligned} P_{XY}(m,n) = {} & (1 - P_{\mathrm{e},X})\,(P_{\mathrm{t},X\mathrm{G}} \cdot P_{\mathrm{GY}}(m, n-1) + P_{\mathrm{t},X\mathrm{B}} \cdot P_{\mathrm{BY}}(m, n-1)) \\ & + P_{\mathrm{e},X}\,(P_{\mathrm{t},X\mathrm{G}} \cdot P_{\mathrm{GY}}(m-1, n-1) + P_{\mathrm{t},X\mathrm{B}} \cdot P_{\mathrm{BY}}(m-1, n-1)), \end{aligned} \tag{3.37}$$

where the respective terms stand for the following probabilities:

- $(1 - P_{\mathrm{e},X})\, P_{\mathrm{t},X\mathrm{G}}\, P_{\mathrm{GY}}(m, n-1)$: no loss in state X, transition to state G, m losses occur in the following $n-1$ packets, channel is in state Y at following packet
- $(1 - P_{\mathrm{e},X})\, P_{\mathrm{t},X\mathrm{B}}\, P_{\mathrm{BY}}(m, n-1)$: no loss in state X, transition to state B, m losses occur in the following $n-1$ packets, channel is in state Y at following packet
- $P_{\mathrm{e},X}\, P_{\mathrm{t},X\mathrm{G}}\, P_{\mathrm{GY}}(m-1, n-1)$: loss in state X, transition to state G, $m-1$ losses occur in the following $n-1$ packets, channel is in state Y at following packet
- $P_{\mathrm{e},X}\, P_{\mathrm{t},X\mathrm{B}}\, P_{\mathrm{BY}}(m-1, n-1)$: loss in state X, transition to state B, $m-1$ losses occur in the following $n-1$ packets, channel is in state Y at following packet

The initial terms of the recursions in (3.37) are given by

$$P_{XY}(0,1) = (1 - P_{\mathrm{e},X})\, P_{\mathrm{t},XY}, \tag{3.38a}$$

$$P_{XY}(1,1) = P_{\mathrm{e},X}\, P_{\mathrm{t},XY}, \tag{3.38b}$$

$$\sum_{Y \in \{\mathrm{G},\mathrm{B}\}} P_{XY}(0,1) = 1 - P_{\mathrm{e},X}, \tag{3.38c}$$

$$\sum_{Y \in \{\mathrm{G},\mathrm{B}\}} P_{XY}(1,1) = P_{\mathrm{e},X}, \tag{3.38d}$$

and $P_{XY}(m,n) = 0$ for $m < 0$ and $m > n$; $X, Y \in \{\mathrm{G}, \mathrm{B}\}$.

3.4.2 Extended Gilbert-Elliott Model

The error and transition probabilities of the extended channel model are marked by a superscript (k_t) or $(k_\mathrm{t}, k_\mathrm{p})$, as introduced in Section 3.2, reflecting the increase factors of the transmission time interval, k_t, and the packet size or packet transmission time, k_p, respectively. These superscripts will also be applied to the probabilities describing loss distributions, e.g., the probabilities of occurrence of m losses in n successive packets as derived in the following. When applied to the calculation of residual loss probabilities for forward error correction schemes in Chapter 4, the superscripts will be omitted to facilitate better readability.

As derived in Section 3.2.2, the error probabilities, i.e., the loss probabilities of a packet in each state, become dependent on the transition probabilities when considering an increased packet size. Therefore, the probabilities describing the event of m packet losses in a group of n consecutive packets have to be modified accordingly by replacing the respective error probabilities in (3.36) and (3.37) with

their transition dependent forms from (3.30) resulting in the following probabilities for the extended Gilbert-Elliott model:

$$\begin{aligned} P^{(k_\mathrm{t},k_\mathrm{p})}(m,n) =& P_{\mathrm{s,G}}\big(P_{\mathrm{GG}}^{(k_\mathrm{t},k_\mathrm{p})}(m,n) + P_{\mathrm{GB}}^{(k_\mathrm{t},k_\mathrm{p})}(m,n)\big) \\ &+ P_{\mathrm{s,B}}\big(P_{\mathrm{BG}}^{(k_\mathrm{t},k_\mathrm{p})}(m,n) + P_{\mathrm{BB}}^{(k_\mathrm{t},k_\mathrm{p})}(m,n)\big) \\ =& \sum_{\substack{X,Y \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(m,n) \end{aligned} \tag{3.39}$$

for arbitrary channel states at the first and following packet. The probabilities $P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(m,n)$ are calculated recursively according to

$$\begin{aligned} P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(m,n) =& (1 - P_{\mathrm{e},X\mathrm{G}}^{(k_\mathrm{t},k_\mathrm{p})})\, P_{\mathrm{t},X\mathrm{G}}^{(k_\mathrm{t})}\, P_{\mathrm{G}Y}^{(k_\mathrm{t},k_\mathrm{p})}(m,n-1) \\ &+ (1 - P_{\mathrm{e},X\mathrm{B}}^{(k_\mathrm{t},k_\mathrm{p})})\, P_{\mathrm{t},X\mathrm{B}}^{(k_\mathrm{t})}\, P_{\mathrm{B}Y}^{(k_\mathrm{t},k_\mathrm{p})}(m,n-1) \\ &+ P_{\mathrm{e},X\mathrm{G}}^{(k_\mathrm{t},k_\mathrm{p})}\, P_{\mathrm{t},X\mathrm{G}}^{(k_\mathrm{t})}\, P_{\mathrm{G}Y}^{(k_\mathrm{t},k_\mathrm{p})}(m-1,n-1) \\ &+ P_{\mathrm{e},X\mathrm{B}}^{(k_\mathrm{t},k_\mathrm{p})}\, P_{\mathrm{t},X\mathrm{B}}^{(k_\mathrm{t})}\, P_{\mathrm{B}Y}^{(k_\mathrm{t},k_\mathrm{p})}(m-1,n-1), \end{aligned} \tag{3.40}$$

with the following initial terms:

$$P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(0,1) = (1 - P_{\mathrm{e},XY}^{(k_\mathrm{t},k_\mathrm{p})})\, P_{\mathrm{t},XY}^{(k_\mathrm{t})}, \tag{3.41a}$$

$$P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(1,1) = P_{\mathrm{e},XY}^{(k_\mathrm{t},k_\mathrm{p})}\, P_{\mathrm{t},XY}^{(k_\mathrm{t})}, \tag{3.41b}$$

$$\begin{aligned} \sum_{Y\in\{\mathrm{G,B}\}} P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(1,1) &= P_{\mathrm{e},X\mathrm{G}}^{(k_\mathrm{t},k_\mathrm{p})}\, P_{\mathrm{t},X\mathrm{G}}^{(k_\mathrm{t})} + P_{\mathrm{e},X\mathrm{B}}^{(k_\mathrm{t},k_\mathrm{p})}\, P_{\mathrm{t},X\mathrm{B}}^{(k_\mathrm{t})} \\ &= \sum_{i=1}^{k_\mathrm{p}} \Big(P_{X\mathrm{G}}(i,k_\mathrm{p}) + P_{X\mathrm{B}}(i,k_\mathrm{p})\Big), \end{aligned} \tag{3.41c}$$

$$\sum_{Y\in\{\mathrm{G,B}\}} P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(0,1) = 1 - \sum_{Y\in\{\mathrm{G,B}\}} P_{XY}^{(k_\mathrm{t},k_\mathrm{p})}(1,1). \tag{3.41d}$$

4

Analysis of Forward Error Correction Capabilities on Packet Transmission Channels

In the transmission of speech and audio over heterogeneous IP networks, the packets possibly cross several transmission links between sender and receiver. These links may all have different properties in terms of available data rate, transmission delay, delay variation (jitter), and packet loss. Wireless links are prone to interference, fading effects, and noise, resulting in bit errors within the received signal which lead to packet losses if not corrected. Channel coding schemes are therefore usually employed at the physical layer in order to protect the transmitted bits and minimize the resulting block error rate and thereby the resulting packet loss rate. A more detailed discussion on the properties of different transmission channels and the means employed for physical layer error protection is given in Section 2.5.

In general, an application has no way of influencing the parameters of intermediate transmission links, e.g., the code rate of physical layer error protection. A way to control the end-to-end quality of packet-based transmissions of speech or other multimedia signals like music or video is to implement an error protection scheme on a higher system level, i.e., on the packet level. In the terminology of the Open Systems Interconnection (OSI) Reference Model [ITU-T Rec. X.200 1994; ISO/IEC 7498-1:1994 1994], the end-to-end protection is applied on the application layer, where also media encoding and RTP packetization is located.

Both, application and transmission network set certain constraints on the design of suitable end-to-end forward error correction (FEC) schemes: The available data rate on the channel limits the maximum amount of redundancy that can be added, and the delay constraints of the application restrict the choice of an applicable FEC scheme and its parameterization. Furthermore, the computational complexity may have to be considered when developing solutions for specific products with limited processing power and battery capacity, e.g., mobile devices. Under these

constraints, the design of the end-to-end FEC scheme will have to be optimized for the expected channel behavior, i.e., the nature and distribution of losses, i.e., their frequency and burstiness.

The application of media independent error protection schemes for packet-based transmission of speech, music, or video frames has been widely considered, implemented, and described in the literature. Furthermore, standardization efforts have been made in the IETF to define appropriate RTP payload formats, e.g., in [Li 2007] (cf. Section 2.4.1). However, a still not sufficiently answered question is how to optimally utilize the available techniques, i.e., which scheme and which parameterization should be used in a specific situation. The optimization requires knowledge about the error protection capabilities of each considered scheme. More specific, the residual loss distribution of frames at the receiver after error correction has to be determined.

In this chapter, theoretical calculations of the correction capabilities of different forward error correction schemes will be developed. The derived formulas will then subsequently be applied in Chapter 5 to real-life speech and audio transmission scenarios over various networks, demonstrating their applicability for determining the best suitable system parameterization.

The considerations in this chapter will focus on media independent schemes for end-to-end error protection by FEC which are suitable for application in packet-based audio transmission. The concepts, e.g., standard block codes or XOR combinations, will be applied on packet level to complete media frames, i.e., the FEC codes are applied in parallel for every bit or byte position of two or more successive media frames such that a certain number of parity frames is generated, depending on the code rate of the FEC code. These parity or FEC frames can then be transmitted either in separate FEC packets or more efficiently by piggybacking[1] the FEC frames one by one to the following packets containing further original media frames. A separate transmission of the FEC frames may be required, e.g., if backwards compatibility has to be ensured. At the receiver, the positions of errors, i.e., lost frames, 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 a certain number of losses in a certain group of frames can be performed, depending on the error correction capabilities of the code. Based on the adaptable channel model introduced in Chapter 3, probabilities are derived for a set of common FEC schemes in Section 4.1 which describe the resulting loss distribution after erasure decoding, i.e., frame regeneration, at the receiver. The calculations include the adaptation of the channel model to the specific properties of the considered FEC scheme, i.e., the resulting packet size and transmission time interval. Thereby, the capabilities of different FEC schemes can be reliably compared for a given transmission channel, as discussed in Chapter 5.

[1] *Piggybacking* in packet data transmission denotes the transmission of additional information, i.e., here the FEC frames, within the same packet payload as some other information, i.e., here the media frames. Hence, the FEC frames are *piggybacked* to the media frames.

An alternative approach to systematically adding redundancy by applying forward error correction is to use retransmission in case of packet loss as discussed in Section 4.2. Such a retransmission can be initiated by the receiver using standard Automatic Repeat-reQuest (ARQ) protocols [Comroe and Costello 1984; Fairhurst and Wood 2002], e.g., *Selective Repeat ARQ*. However, for real-time services in heterogeneous networks an end-to-end based ARQ scheme would usually cause too much delay. Retransmission may nevertheless be used efficiently if it is directly applied on the physical or data link layer of a wireless transmission system. A prominent example is the Hybrid ARQ (HARQ) technique applied on UMTS-HSPA channels. A more detailed discussion on the conditions under which FEC or ARQ schemes are preferred is given in Section 2.7.1. The remaining frame loss rate and distribution after a limited number of transmission attempts are derived in Section 4.2. The performance and data rate efficiency of retransmission schemes is then compared to redundancy adding forward error correction schemes in Section 5.2.2.

Finally, in Section 4.3 the evaluation of FEC capabilities will be discussed for channels with varying transmission delay (jitter), where FEC frames can also be utilized to recover frames from delayed packets. The calculation of the loss probability of a specific frame after possible regeneration from other media and FEC frames has to consider different delay budgets for each involved packet. The application of FEC schemes in such transmission scenarios may allow for a reduction of the receiver buffer length and thereby the end-to-end delay.

4.1 Theoretical Determination of Residual Losses for Different FEC Schemes

In this section, different forward error correction (FEC) schemes for application on packet level are discussed. The focus is on block codes (e.g., Reed-Solomon codes), exclusive disjunction, i.e., XOR combinations of frames, as well as on a computationally very simple frame repetition scheme. All these schemes are applied in existing services of multimedia IP transmission. In the literature, several approaches have been made to derive the theoretical performance of some of these schemes on a packet loss channel. The residual frame loss rate and mean burst length when applying a Reed-Solomon code have been calculated in [Frossard 2001], assuming that the FEC frames were transmitted separately from the original frames in specific FEC packets. [Jiang and Schulzrinne 2002] showed a similar calculation for piggybacking the FEC frames to the following media frame packets. Both used the simplified Gilbert model in which the states directly determine the reception or loss of a packet ($P_{\mathrm{e,G}} = 0$ and $P_{\mathrm{e,B}} = 1$, cf. Section 3.1.3.3). In Appendix E.2 it has been shown that this simplified model is not sufficient in describing the effects of packet loss on wireless packet channels like UMTS and WLAN, especially when considering a flexible model that should be adapted for different transmission time intervals and packet sizes. Although a change of the transmission time interval has

been considered in [Jiang and Schulzrinne 2002], the influence of the packet size has been neglected.

The following sections consider the different FEC schemes mentioned above. Depending on the variable parameters of each code, the packet sizes and transmission time intervals of each scheme are determined, considering both a separate and piggybacked transmission of the FEC frames. Furthermore, the required data rate and the resulting delay will be computed. Finally, the residual loss probabilities and burst lengths after erasure correction will be derived. In contrast to [Frossard 2001] and [Jiang and Schulzrinne 2002], the derivation of these properties is based on the extended Gilbert-Elliott model as introduced in Section 3.2, thereby correctly including the dependencies on the packet transmission time interval and the packet size. This allows for a realistic and fair comparison of the different schemes and their individual parameterizations on wireless packet channels. The determination of the optimal scheme and parameterization is discussed for various channels and application scenarios in Chapter 5.

4.1.1 Interleaved Transmission of Media Frames

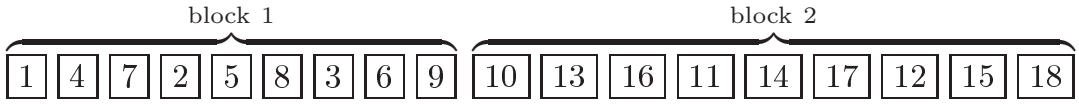
In general, wireless transmission channels, such as Wireless LAN, do not produce statistically independent packet losses, but rather bursts of packet losses (cf. analysis of channel measurements in Appendix E.2). A well-known concept to cope with bursty error distributions is to employ an interleaver which breaks up long error bursts into several shorter ones. Interleavers are applied, e.g., to the bit stream in mobile communication systems or to the data stored on digital storage media like CD/DVD.

Since the errors in the considered transmission systems consist of packet losses and therefore of complete media frames (assuming the transmission of one or several media and/or FEC frames per packet), the interleaver has to operate on packet level. Interleaving on packet level is achieved by reordering the frames prior to transmission, such that the transmission time point of successive frames is drawn further apart. The resulting loss distribution after deinterleaving, i.e., reordering at the receiver, consists of mainly short loss lengths (i.e. single frames) which can be recovered more efficiently by the used frame level FEC schemes. If no FEC can be applied, shorter losses are also more easily concealed by the packet loss concealment routine at the receiver (cf. Chapter 6). However, the objective of an interleaver can only be achieved at the expense of an increased end-to-end delay.

In the following, a general block interleaver is considered which is defined by two parameters, the interleaver depth d_{il} and the interleaver length l_{il}. Consider the following example with $d_{\mathrm{il}} = l_{\mathrm{il}} = 3$. Two successive blocks of $d_{\mathrm{il}} \cdot l_{\mathrm{il}} = 9$ frames are shown, represented by consecutively numbered squares:

l_{il} columns			l_{il} columns			
1	4	7	10	13	16	d_{il} rows
2	5	8	11	14	17	
3	6	9	12	15	18	

The encoded media frames are written column-wise into the interleaver matrix at the sender. Once a block is complete, i.e., the matrix is filled, the frames are read out row-wise and sent as payload in individual packets, resulting in the following transmission order:



This procedure is repeated for every block of $d_{\mathrm{il}} \cdot l_{\mathrm{il}}$ frames. The receiver buffers the received frames before play-out for reconstructing the correct order. Consequently, a block interleaver causes a delay of about $2\, d_{\mathrm{il}}\, l_{\mathrm{il}}$ times the length of a frame[2]. The interleaver length l_{il} determines the distance in the interleaved sequence between two originally consecutive frames, while the interleaver depth d_{il} determines the original distance between two consecutive frames from the interleaved sequence. The appropriate choice of the interleaver parameters d_{il} and l_{il} depends on the considered FEC scheme and the expected packet loss distribution on the channel, as well as a possible delay constraint.

The effect of an interleaver which spreads successive frames by the length l_{il} can be approximated through an adaptation of the channel model as described in Section 3.2 by increasing the transmission time interval of the original packet stream, T_{TTI}, to a new effective transmission time interval $T^{\mathrm{eff}}_{\mathrm{TTI}} = l_{\mathrm{il}} \cdot T_{\mathrm{TTI}}$. Assuming that the channel model is based on an arbitrary transmission time interval T'_{TTI} which not necessarily has to equal the transmission time interval of the considered packet stream, T_{TTI}, the adaptation factor is calculated as

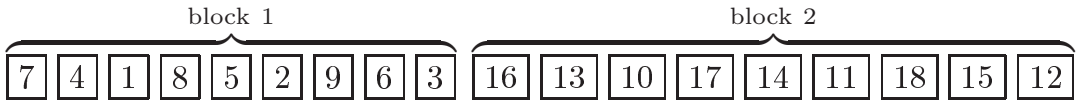
$$k_{\mathrm{t}} = \frac{T^{\mathrm{eff}}_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}}. \tag{4.1}$$

The interleaving of frames does not increase the size of the packets, i.e., only possibly differing packet sizes of the channel model and the packet stream need to be considered in the adaptation factor $k_{\mathrm{p}} = \tau_{\mathrm{p}}/\tau'_{\mathrm{p}}$. If the transmission time interval and the packet size of the channel model already equal those of the considered packet stream, i.e., $T'_{\mathrm{TTI}} = T_{\mathrm{TTI}}$ and $\tau'_{\mathrm{p}} = \tau_{\mathrm{p}}$, the adjustment of the interleaver is done with $k_{\mathrm{t}} = l_{\mathrm{il}}$ and $k_{\mathrm{p}} = 1$.

For such a modification of the channel model an ideal behavior of the interleaver is assumed, i.e., that two originally successive frames are always spaced l_{il} packets apart in the interleaved sequence. However, at the boundaries of successive blocks, the described block interleaver does not increase the distance of the last frame of the previous block and the first frame of the following block. This block boundary effect can have a significant impact on the performance of the interleaver when it occurs too often, i.e., when d_{il} and l_{il} are small. For these cases it is beneficial to use a slightly modified interleaver, where the rows of the matrix in the example above

[2] An efficient implementation can reduce the delay to a minimum of $2(d_{\mathrm{il}} - 1)(l_{\mathrm{il}} - 1)$ times the length of a frame, since the transmission can already start before the matrix is completely filled.

are transmitted in reverse order[3], resulting in the following transmitted sequence of frames:



4.1.2 Overview of Considered FEC Schemes

A large variety of possible forward error correction techniques may be applied on packet level for end-to-end protection on packet-switched channels. The forward error correction schemes considered in this work are a low complex repetition of frames, the exclusive disjunction of frames (i.e. XOR combination), and a flexible block code, e.g., implemented as Reed-Solomon code. For each of these schemes, the FEC frames may either be transmitted in separate FEC packets or they may be piggybacked to the original media packets.

Figures 4.1 and 4.2 visualize each scheme and explain the packetization of media and FEC frames as well as the timing of each packet transmission. The successive media frames are denoted by rectangles marked with the letters w, x, y, z, a, b, c, d, or with $a_1, a_2, \ldots, b_1, b_2, \ldots$ for the block codes shown in Figure 4.2. FEC frames can be copies of original frames and are then marked like the original itself. If FEC frames are the result of an XOR combination of two frames, they are marked with a '+' symbol in between the letters of the contributing frames, e.g., $a + b$. Finally, FEC frames which are derived as parity frames from block codes are denoted by a line over the letter of the according block, e.g., $\overline{a}_1, \overline{a}_2, \ldots$. Depending on the considered scheme, several media and FEC frames together or each frame by itself form a packet's payload. The packet headers are indicated by black squares.

Depending on its parameterization and the chosen transmission strategy, each FEC scheme may lead to different packet sizes and time points of packet transmission, as visualized in the figures, and consequently to different packet data rates and end-to-end delays. For every FEC scheme, the formulas of the transmission time interval T_{TTI} and the packet transmission time τ_{p} are given in Table 4.1, together with the packet data rate R_{p} and the end-to-end delay D that the application of each scheme requires.

The packet transmission time interval, T_{TTI}, is the time distance between the transmission of successive media or FEC packets. When the FEC frames are transmitted in separate FEC packets, they still have to be sent within the original length of a frame, T_{f}, preferably regularly spaced in this interval as shown for the different FEC schemes in Figure 4.1. In the piggybacked transmission scenarios, the packets are transmitted at the original transmission time interval which equals the frame length in a packet, T_{f}.

The packet transmission time, τ_{p}, is defined as the time that a fixed-rate transmission channel is occupied with the transmission of a packet. It is determined by

[3] In this case, an efficient implementation can only reduce the delay to a minimum of $2d_{\mathrm{il}}(l_{\mathrm{il}}-1)$ times the length of a frame.

the size of the packet headers L_{h} (possibly including layer 2 headers, cf. Section 2.5), the size of a media/FEC frame L_{f}, the number of frames transmitted in a packet, and finally the transmission rate R_{ch} on the channel. Hence, for the transmission of media and FEC frames in separate packets, all packets have the same size and packet transmission time as the original media packets. For the piggybacked transmission schemes, the packet size depends on the considered scheme and parameterization. The time it takes to transmit a single packet may be considerably smaller than the transmission time interval, e.g., on high data rate channels. In Figure 4.1 and 4.2, the relations of header size to payload size, as well as packet transmission time to transmission time interval are only exemplary and do not necessarily reflect realistic proportions. In real applications, these relations depend on the chosen media codec, the channel data rate, and other factors.

The packet data rate, R_{p}, is determined by packet size and frequency, i.e., the number of packets transmitted per time interval. The minimum end-to-end delay of the media signal, D, consists of a contribution from the sender, D_s, the transmission delay of the network, D_{tx}, and a contribution from the receiver, D_r, as discussed in detail in Section 2.6.1. For the delay calculations in this section, a real-time speech communication is assumed. For reasons of simplicity, the processing delays of sender and receiver are neglected, i.e., $D_{\mathrm{s,proc}} = D_{\mathrm{r,proc}} = 0$, and the employed speech codec is assumed to have no lookahead, i.e., $T_{\mathrm{la}} = 0$. The delay contribution of the sender then consists only of the algorithmic delay for collecting the media frame(s) before encoding, optional interleaving, and transmission. The speech codec is further assumed to have a constant bit rate and frame length T_{f}.[4] Depending on the utilized FEC scheme and interleaver settings, several frames may need to be collected before the first packet is transmitted which contributes to the sender delay D_s. For streaming applications, the delay contribution of the sender, D_s, is always zero as all frames can be assumed to be already encoded and readily available from a storage medium. Finally, the receiver buffer length D_r needs to be large enough so that the FEC frames in following packets can be utilized to recover a lost media frame.

A more detailed discussion of each scheme and its parameterization is given in the following sections.

No Forward Error Correction

If no forward error correction is applied, i.e., no redundancy is added, the transmitted frames are not protected against packet losses. The loss distribution, however, can be influenced by an interleaved transmission of the frames. This will not reduce the final loss rate, but it will reduce the mean burst length of losses in the signal. Through the use of an interleaver, longer loss bursts are broken into shorter losses, which are easier to conceal by *packet loss concealment* (PLC) algorithms (cf. Chapter 6). In the following, a quadratic block interleaver is considered with a variable interleaver length l_{il} and interleaver depth $d_{\mathrm{il}} = l_{\mathrm{il}}$.

[4] The frame length may be the chosen length of a codec with arbitrary but fixed frame length or a multiple of the codec's fixed frame length as defined in Section 2.4.1.

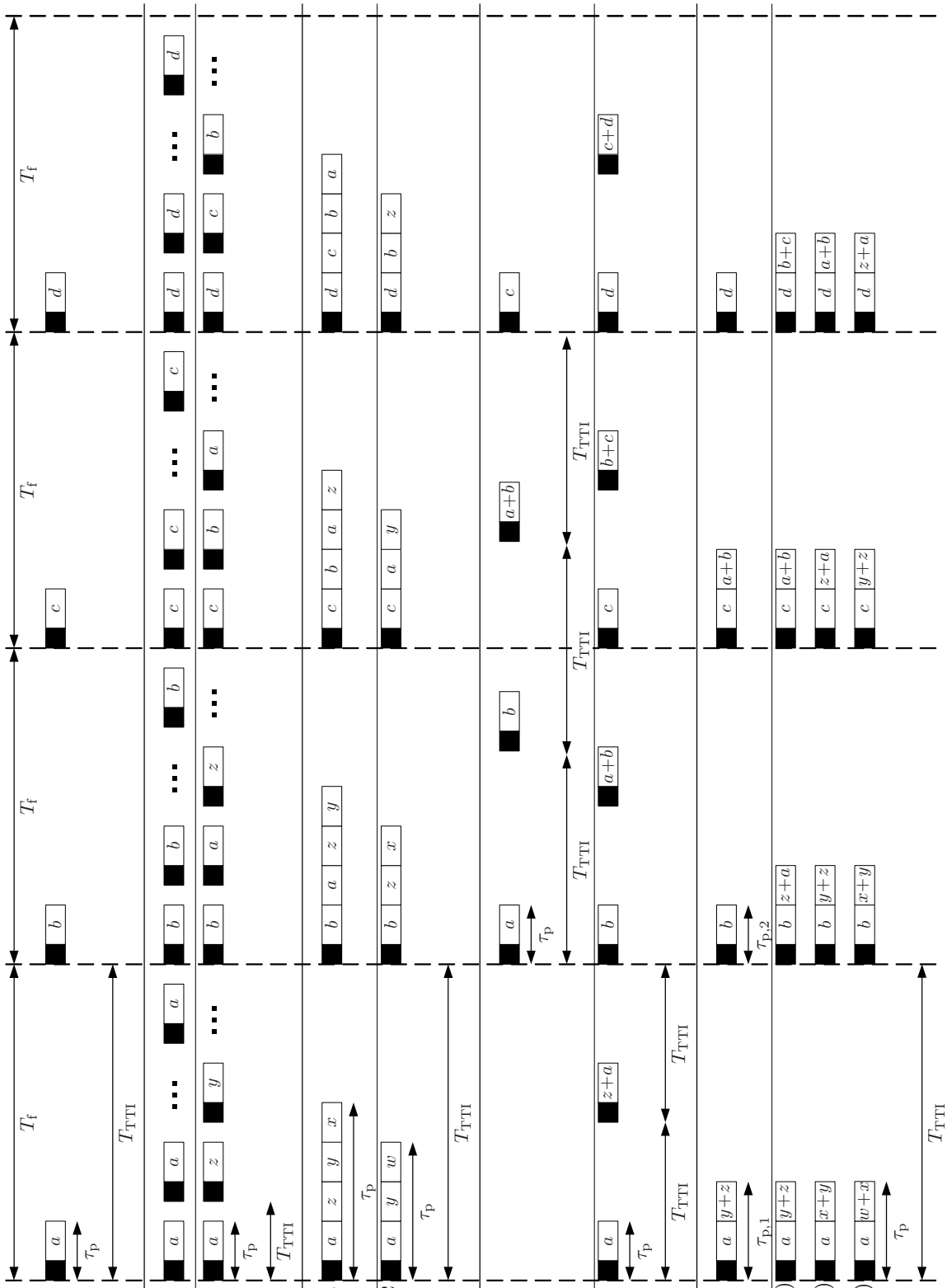


Figure 4.1: Forward Error Correction (FEC) Schemes: Repetition Code (REP) and XOR Combinations (XOR); transmission of FEC frames in separate packets or piggybacked (PB) to original media packets. Sequence of media frames: $w, x, y, z, a, b, c, d, \ldots$

Scheme	T_{TTI}	τ_{p}	R_{p}	$D = D_s + D_{\mathrm{tx}} + D_r$	
				$l_{\mathrm{il}} = 1$	$l_{\mathrm{il}} > 1$
no FEC	T_{f}	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = T_{\mathrm{f}}$ $D_r = 0$ $D = T_{\mathrm{f}} + D_{\mathrm{tx}}$	$d_{\mathrm{il}} = l_{\mathrm{il}}$ $D_s = D_r = l_{\mathrm{il}}\, d_{\mathrm{il}}\, T_{\mathrm{TTI}}$ $D = 2\, l_{\mathrm{il}}^2\, T_{\mathrm{f}} + D_{\mathrm{tx}}$
REP Scheme 1)	$\frac{T_{\mathrm{f}}}{p+1}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}}\,(p+1)$	$D_s = T_{\mathrm{f}}$ $D_r = p\, T_{\mathrm{TTI}}$ $D = \left(1 + \frac{p}{p+1}\right) T_{\mathrm{f}} + D_{\mathrm{tx}}$	$d_{\mathrm{il}} = 2\,(p+1)$ $D_s = D_r = l_{\mathrm{il}}\, d_{\mathrm{il}}\, T_{\mathrm{TTI}}$ $D = 4\, l_{\mathrm{il}}\, T_{\mathrm{f}} + D_{\mathrm{tx}}$
REP Scheme 2)	$\frac{T_{\mathrm{f}}}{p+1}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}}\,(p+1)$	$D_s = T_{\mathrm{f}}$ $D_r = p\, T_{\mathrm{f}} + p\, T_{\mathrm{TTI}}$ $D = \left(p + 2 - \frac{1}{p+1}\right) T_{\mathrm{f}} + D_{\mathrm{tx}}$	N/A
REP-PB	T_{f}	$\frac{L_{\mathrm{h}}+(p+1)\, L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+(p+1)\, L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = T_{\mathrm{f}}$ $D_r = p\, d_p\, T_{\mathrm{TTI}}$ $D = (p\, d_p + 1)\, T_{\mathrm{f}} + D_{\mathrm{tx}}$	N/A
XOR Scheme 1)	$\frac{2}{3}\, T_{\mathrm{f}}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{3}{2}\, \frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = 2\, T_{\mathrm{f}}$ $D_r = 2\, T_{\mathrm{TTI}}$ $D = \frac{10}{3}\, T_{\mathrm{f}} + D_{\mathrm{tx}}$	$d_{\mathrm{il}} = 6$ $D_s = D_r = l_{\mathrm{il}}\, d_{\mathrm{il}}\, T_{\mathrm{TTI}}$ $D = 8\, l_{\mathrm{il}}\, T_{\mathrm{f}} + D_{\mathrm{tx}}$
XOR Scheme 2)	$\frac{1}{2}\, T_{\mathrm{f}}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$2\, \frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = T_{\mathrm{f}}$ $D_r = d_r\, T_{\mathrm{f}}$ $D = (1 + d_r)\, T_{\mathrm{f}} + D_{\mathrm{tx}}$	$d_{\mathrm{il}} = 10$ $D_s = D_r = l_{\mathrm{il}}\, d_{\mathrm{il}}\, T_{\mathrm{TTI}}$ $D = 10\, l_{\mathrm{il}}\, T_{\mathrm{f}} + D_{\mathrm{tx}}$
XOR-PB Scheme 1)	T_{f}	$\frac{L_{\mathrm{h}}+\frac{3}{2}L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+\frac{3}{2}L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = T_{\mathrm{f}}$ $D_r = 2\, T_{\mathrm{TTI}}$ $D = 3\, T_{\mathrm{f}} + D_{\mathrm{tx}}$	N/A
XOR-PB Scheme 2)	T_{f}	$\frac{L_{\mathrm{h}}+2L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+2L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = T_{\mathrm{f}}$ $D_r = d_r\, T_{\mathrm{f}}$ $D = (1 + d_r)\, T_{\mathrm{f}} + D_{\mathrm{tx}}$	N/A
RS	$\frac{k}{n}\, T_{\mathrm{f}}$	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{n}{k}\, \frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = k\, T_{\mathrm{f}}$ $D_r = (n-1)\, T_{\mathrm{TTI}}$ $D = \left(2\, k - \frac{k}{n}\right) T_{\mathrm{f}} + D_{\mathrm{tx}}$	$d_{\mathrm{il}} = n$ $D_s = D_r = l_{\mathrm{il}}\, d_{\mathrm{il}}\, T_{\mathrm{TTI}}$ $D = 2\, l_{\mathrm{il}}\, k\, T_{\mathrm{f}} + D_{\mathrm{tx}}$
RS-PB	T_{f}	$\frac{L_{\mathrm{h}}+\frac{n}{k}L_{\mathrm{f}}}{R_{\mathrm{ch}}}$	$\frac{L_{\mathrm{h}}+\frac{n}{k}L_{\mathrm{f}}}{T_{\mathrm{f}}}$	$D_s = T_{\mathrm{f}}$ $D_r = (n-1)\, T_{\mathrm{TTI}}$ $D = n\, T_{\mathrm{f}} + D_{\mathrm{tx}}$	$d_{\mathrm{il}} = \left\lfloor \frac{n}{k} \right\rfloor \cdot k$ $D_s = D_r = d_{\mathrm{il}}\, l_{\mathrm{il}}\, T_{\mathrm{TTI}}$ $D = 2\, d_{\mathrm{il}}\, l_{\mathrm{il}}\, T_{\mathrm{f}} + D_{\mathrm{tx}}$

Table 4.1: Transmission time interval T_{TTI}, packet transmission time τ_{p}, packet data rate R_{p}, and end-to-end delay D for all considered FEC schemes.

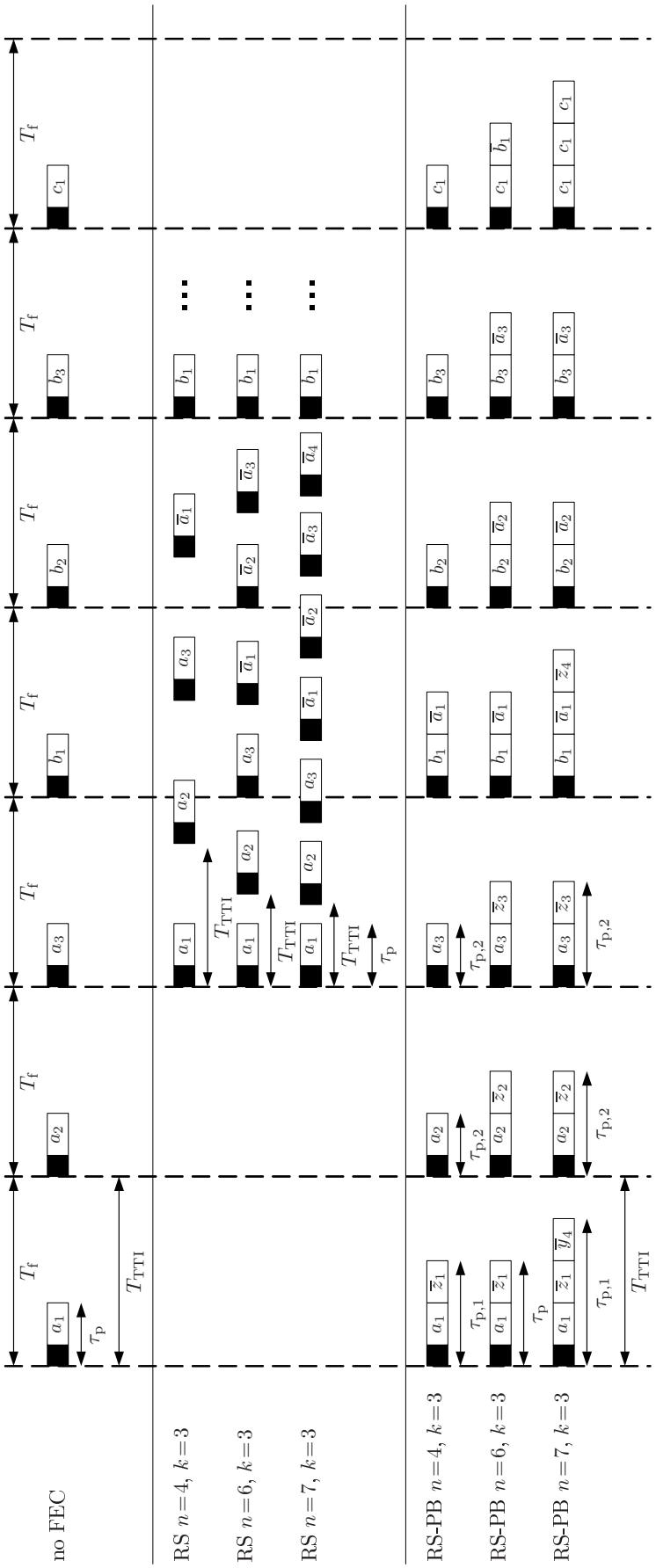


Figure 4.2: Forward Error Correction (FEC) Schemes: Reed-Solomon Block Codes (RS); transmission of FEC frames in separate packets or piggybacked (PB) to original media packets. Sequence of media frames: $y_1, y_2, y_3, z_1, z_2, z_3, a_1, a_2, a_3, b_1, b_2, b_3, \ldots$

Repetition Code

The simplest and computationally least expensive way of adding redundancy is to use a repetition code, i.e., to transmit several identical copies of a media frame in different packets. The receiver needs to receive one of these copies, otherwise the frame is lost. However, the repetition code leads to a high increase of the transmission data rate and is in this aspect the most inefficient way of adding redundancy. The number of repetitions of each media frame is given by the parameter p, i.e., including the original copy every frame is sent $p+1$ times. The repetitions are either sent in separate packets of their own or piggybacked to following packets containing other media frames (originals and repetitions).

When transmitting repetitions of media frames in separate packets, the order of transmission can be set to one of the following two possibilities. In *REP Scheme 1)*, the original frame and its repetitions are all sent consecutively in separate packets within the original transmission time interval, which amounts to the frame length T_f. An optional block interleaver for this scheme should use a fixed interleaver depth $d_\mathrm{il} = 2(p+1)$ such that the copies of two successive original frames are spaced apart by another packet. A suitable interleaver length l_il then determines the distance of the different copies of a frame and has to be chosen depending on the desired capability and the burst characteristics of the channel. Another approach, as shown in *REP Scheme 2)*, is to transmit within a frame length T_f first the original frame, then the copy of the previous frame, followed by a copy of the previous to the previous, and so on. Thus, the distance between the copies of a frame is increased at the expense of a higher delay, which should lead to a higher robustness against burst errors. This results in an interleaved transmission of the frames, although different from the above block interleaver. No further block interleaver is therefore considered for this scheme.

The repetitions can also be piggybacked to the following packets containing originals and repetitions of other media frames (*REP-PB*). Here, the distance parameter d_p controls to which packet a repeated frame is attached, i.e., how many packets lie between each copy of a frame. Packet n then contains the following frame numbers: $n, n-d_p, n-2\,d_p, \ldots, n-p\,d_p$. Examples are given in Figure 4.1. The use of a distance $d_p > 1$ has a similar effect compared to the use of a block interleaver before transmitting the packets, because it spreads the copies of a single frame further apart in time and is therefore able to recover from longer burst losses. At the same time, a significantly lower delay is required compared to a block interleaver with the same performance. Therefore, a further block interleaver is not considered.

Exclusive Disjunction (XOR) Schemes

XOR schemes are more rate efficient and require a slightly higher computational complexity than the repetition scheme above, but it is still low compared to the block codes discussed below. In the XOR schemes, several media frames are bitwise added modulo 2 to generate additional FEC frames which can be used at the receiver

to regenerate lost frames. In the considered scenarios the frames have all the same length. If the involved media frames were of different lengths, the shorter ones would have to be padded with zeros before the XOR operation such that all involved frames have the same length.[5] The erasure correction capabilities of simple XOR schemes are limited and they are not as flexible as other schemes in terms of code rate and block length (cf. Reed-Solomon codes below). Nevertheless, XOR schemes are still sufficient in many cases and have the advantage of a low computational complexity.

Two exemplary schemes of generating XOR combinations of two successive media frames are considered here. The schemes differ in the amount of redundancy, i.e., in the resulting data rate. As for the repetition code, the FEC frames may be either transmitted as separate packets or piggybacked to subsequent media packets. The first XOR code (*XOR Scheme 1)* and *XOR-PB Scheme 1)*) generates an FEC frame for every group of two successive media frames, resulting in a code rate of $r_\mathrm{c} = 2/3$. The second code (*XOR Scheme 2)* and *XOR-PB Schemes 2a,b,c)*) generates a new FEC frame after each media frame as XOR combination of the two previous original frames, thereby transmitting information on a specific frame additionally as part of two FEC frames. The code rate then results to $r_\mathrm{c} = 1/2$. For the piggybacked transmission of this scheme, three different delays for attaching the FEC data are considered (see *XOR-PB Schemes 2a,b,c)* in Figure 4.1). At the expense of a slightly increased overall delay, a higher robustness against packet loss bursts can be achieved when the FEC data is further separated from the original media frame.

The receiver needs to wait for future packets with media and FEC frames that can be used to reconstruct a previously lost frame. Theoretically, the reception of an original frame may allow the iterative reconstruction of frames a long distance in the past, if all intermediate FEC frames have been received. In practical applications, the delay is limited and the receiver can only wait for a certain number d_r of frame lengths T_f for future packets to arrive. A choice of $d_r = 4$ for *XOR Scheme 2)* (separate transmission of FEC frames) has shown to be the best compromise between delay and performance. For the piggybacked *XOR-PB Scheme 2a)*, a choice of $d_r^{\mathrm{2a)}} = 4$ leads to the best compromise. To include the same amount of information for reconstruction in *XOR-PB Schemes 2b) and 2c)*, d_r has to be chosen as $d_r^{\mathrm{2b)}} = d_r^{\mathrm{2a)}} + 1 = 5$ and $d_r^{\mathrm{2c)}} = d_r^{\mathrm{2a)}} + 2 = 6$.

When transmitting the FEC data in separate packets, the dimensioning of an optional block interleaver uses a fixed interleaver depth $d_\mathrm{il} = 6$ for *XOR Scheme 1)* and $d_\mathrm{il} = 10$ for *XOR Scheme 2)* to achieve a sufficient spacing of the frames. A suitable interleaver length l_il has to be chosen depending on the desired capability and the burst characteristics of the channel. Again, no interleaving is considered for the piggybacked transmission schemes.

[5] If the information on the original lengths of the frames is not already part of the frame data, it has to be communicated to the receiver, e.g., using respective payload header fields in the according payload format, e.g., [Li 2007].

Block Codes

Block codes, e.g., Reed-Solomon (RS) codes, provide a flexible way of applying forward error correction to the packet stream with arbitrary amounts of redundancy. However, they are also computationally more complex when compared to the schemes described in the previous sections. A systematic (n, k)-block code adds $n-k$ parity frames to each group of k media frames, resulting in a total of n code frames describing the k media frames. When applying a Reed-Solomon Code, the code usually operates on symbols of 8 bit length, i.e., on complete bytes. The calculation of the parity information is then performed in parallel for each byte position of the k media frames, resulting in the according bytes of the $n-k$ parity frames. If the k media frames of one encoding block are of different lengths, the shorter ones first have to be extended with bytes of zeros such that all frames have the same length. The padding bytes are only needed for the encoding process and do not have to be transmitted. The parity frames therefore always have the length of the longest frame of the current group of k contributing frames.

Block codes can detect and correct a certain number of errors within an encoded block, depending on the chosen parameterization of the code. In the considered application of packet transmission, errors need not be detected, but are known at the receiver in form of erasure information, i.e., the position of lost packets and the contained frames. The receiver can therefore perform an erasure correction (recovery) of up to $n-k$ lost frames within a group of n encoded frames.

The transmission of the $n-k$ FEC frames, either in separate FEC packets or piggybacked to the original media packets, is visualized in Figure 4.2. The string of media packets is segmented into groups of k packets, denoted by a_i, b_i, and so forth, with $i = 1, \ldots, k$. The parity (FEC) frames which are calculated for each group of media frames are denoted by a line above the letter, e.g., for group a_i the FEC frames are denoted by $\overline{a}_j$ with $j = 1, \ldots, n-k$.

In case media and FEC frames are transmitted in separate packets (*RS* schemes), the transmission of an encoded block of packets, i.e., the media and FEC packets of a group, can start when all media packets of this group are available and the FEC packets have been generated. Figure 4.2 shows the transmission time points of the packets for several exemplary combinations of n and k. An optional interleaver uses a fixed interleaver depth $d_{\mathrm{il}} = n$ and a suitable interleaver length l_{il} depending on the desired capability and the burst characteristics of the channel.

The piggybacked transmission of FEC frames (*RS-PB* schemes) is visualized in Figure 4.2 using the same n, k combinations as for the separate transmission. For an optional block interleaver, the following choice of the interleaver depth d_{il} assures that the distance between two successive parts of an encoded block is always at least the given interleaver length l_{il}:

$$d_{\mathrm{il}} = \left\lceil \frac{n}{k} \right\rceil \cdot k. \tag{4.2}$$

4.1.3 Data Rate and Delay Constraints

The packet data rate R_p of the transmission is constrained by the capacity of the channel or in case of multiplexed transmission of several packet streams, by the capacity of the logical channel of a single stream. The data rate R_p is therefore limited and cannot exceed the maximum available data rate $R_\mathrm{p,max}$, i.e.,

$$R_\mathrm{p} \leq R_\mathrm{p,max}. \tag{4.3}$$

This constraint limits the possible code rate, i.e., the amount of redundancy that can be added by any FEC scheme, and therefore limits the parameterization of the different schemes. It also influences the choice of the codec, i.e., the encoding rate, and the frame length to transmit in each packet, because the latter determines the resulting header overhead. The choice of an optimal frame length and the determination of an optimal trade-off between encoding rate and added redundancy is discussed in Section 5.2 of the following chapter.

The transmission of multimedia frames usually requires a low end-to-end signal delay. This constraint of a maximum delay D_max is more or less tight depending on the considered application, e.g., conversational or streaming services:

$$D \leq D_\mathrm{max}. \tag{4.4}$$

The delay constraint limits the parameterization of the FEC schemes and particularly the possible interleaver length l_il, depending on the frame length T_f of the codec.

The specific limitations of parameters depending on the maximum available data rate $R_\mathrm{p,max}$ and the maximum tolerable delay D_max are listed in Table 4.2 for all considered FEC schemes. Most parameter constraints can be directly derived from the according formulas of the packet data rate R_p and end-to-end delay D given for each scheme in Table 4.1. An exemplary case shall be considered in more detail: For the block code with separate transmission of media and FEC packets (RS), the delay D must not exceed the given maximum D_max:

$$\left(2\,k - \frac{k}{n}\right) T_\mathrm{f} + D_\mathrm{tx} \leq D_\mathrm{max}. \tag{4.5}$$

Furthermore, the data rate R_p must not exceed the available maximum data rate $R_\mathrm{p,max}$:

$$\frac{n}{k}\,\frac{L_\mathrm{h} + L_\mathrm{f}}{T_\mathrm{f}} \leq R_\mathrm{p,max}. \tag{4.6}$$

Hence, the code rate $r_\mathrm{c} = k/n$ to use must be larger than a minimum value, i.e.,

$$\frac{k}{n} \geq \frac{L_\mathrm{h} + L_\mathrm{f}}{R_\mathrm{p,max}\,T_\mathrm{f}}. \tag{4.7}$$

Assuming the minimum possible code rate according to (4.7), this value can be substituted in (4.5) to finally derive a maximum value for the length of an information block k:

$$k \leq \frac{(D_{\max} - D_{\mathrm{tx}})\, R_{\mathrm{p,max}} + L_{\mathrm{h}} + L_{\mathrm{f}}}{2\, R_{\mathrm{p,max}}\, T_{\mathrm{f}}}. \tag{4.8}$$

4.1.4 No Forward Error Correction

For the calculation of the frame loss distribution in case of no forward error correction, but optional interleaving, the given channel model (with possibly differing T'_{TTI} and τ'_{p}) needs to be adjusted to the actual transmission time interval and packet size as described in Section 3.2. The effect of an interleaver can be included according to Section 4.1.1, finally resulting in the following adjustment factors for the transmission time interval and packet size, k_{t} and k_{p}, respectively:

$$k_{\mathrm{t}} = \frac{T^{\mathrm{eff}}_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{f}}}{T'_{\mathrm{TTI}}} \quad \wedge \quad k_{\mathrm{p}} = \frac{\tau_{\mathrm{p}}}{\tau'_{\mathrm{p}}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}}}{R_{\mathrm{ch}}\, \tau'_{\mathrm{p}}}. \tag{4.9}$$

In the following, all probabilities are assumed to be derived from the appropriately adjusted channel model. The superscripts $(k_{\mathrm{t}}, k_{\mathrm{p}})$ indicating the adjustment factors of the channel model are omitted in order to facilitate readability.

The frame loss probability P_{fl} at the receiver is the probability of losing a single packet, since no error correction is possible:

$$P_{\mathrm{fl}} = P(1,1). \tag{4.10}$$

The term $P(m,n)$ denotes the probability of losing m out of n consecutive packets which is calculated from the adjusted channel model as explained in Section 3.4. Here, the different possibilities of channel states at the first and at the $(n+1)$-th packet are included with their respective probabilities.

A frame loss burst is defined in this work as the event of loosing a certain number of consecutive frames. Hence, the loss of a single frame also constitutes a burst – a burst of length 1 – if the preceding and following frames are received. A burst start is defined as the event of losing a media frame after receiving the preceding media frame. Considering no forward error correction, each packet contains a single media frame. Hence, the probability of a burst start $P_{\mathrm{b,s}}$ is given as the probability to first receive a packet and then to lose the next packet. This probability can be determined with the general probability of losing m out of n packets, $P_{XY}(m,n)$, assuming that the channel is in state X at the first packet and in state Y at the $(n+1)$-th packet:

$$P_{\mathrm{b,s}} = \sum_{\substack{X,Y \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P_{XY}(0,1)\big(P_{Y\mathrm{G}}(1,1) + P_{Y\mathrm{B}}(1,1)\big). \tag{4.11}$$

Scheme	$\boldsymbol{R_{\mathrm{p}} \leq R_{\mathrm{p,max}}}$	$\boldsymbol{D \leq D_{\max}}$
no FEC	$\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}} \leq R_{\mathrm{p,max}}$	$l_{\mathrm{il}} \leq \sqrt{\frac{D_{\max}-D_{\mathrm{tx}}}{2\,T_{\mathrm{f}}}}$
REP Scheme 1)	$p \leq \frac{R_{\mathrm{p,max}}\,T_{\mathrm{f}}}{L_{\mathrm{h}}+L_{\mathrm{f}}} - 1$	$p \leq \frac{D_{\max}-D_{\mathrm{tx}}-T_{\mathrm{f}}}{2\,T_{\mathrm{f}}+D_{\mathrm{tx}}-D_{\max}}$, if $D_{\max} \leq 2\,T_{\mathrm{f}} + D_{\mathrm{tx}}$; $l_{\mathrm{il}} \leq \frac{D_{\max}-D_{\mathrm{tx}}}{4\,T_{\mathrm{f}}}$
REP Scheme 2)	$p \leq \frac{R_{\mathrm{p,max}}\,T_{\mathrm{f}}}{L_{\mathrm{h}}+L_{\mathrm{f}}} - 1$	$p \leq \frac{a-1}{2} + \sqrt{\left(\frac{a-1}{2}\right)^2 + a + 1}$, with $a = \frac{D_{\max}-D_{\mathrm{tx}}}{T_{\mathrm{f}}} - 2$; $p \lesssim \frac{D_{\max}-D_{\mathrm{tx}}}{T_{\mathrm{f}}} - 2$, if $D_{\max}$ is large
REP-PB	$p \leq \frac{R_{\mathrm{p,max}}\,T_{\mathrm{f}}-L_{\mathrm{h}}-L_{\mathrm{f}}}{L_{\mathrm{f}}}$	$p \cdot d_p \leq \frac{D_{\max}-D_{\mathrm{tx}}-T_{\mathrm{f}}}{T_{\mathrm{f}}}$
XOR Scheme 1)	$\frac{3}{2}\,\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}} \leq R_{\mathrm{p,max}}$	$l_{\mathrm{il}} \leq \frac{D_{\max}-D_{\mathrm{tx}}}{8\,T_{\mathrm{f}}}$
XOR Scheme 2)	$2\,\frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{T_{\mathrm{f}}} \leq R_{\mathrm{p,max}}$	$d_r \leq \frac{D_{\max}-D_{\mathrm{tx}}-T_{\mathrm{f}}}{T_{\mathrm{f}}}$ $l_{\mathrm{il}} \leq \frac{D_{\max}-D_{\mathrm{tx}}}{10\,T_{\mathrm{f}}}$
XOR-PB Scheme 1)	$\frac{L_{\mathrm{h}}+\frac{3}{2}\,L_{\mathrm{f}}}{T_{\mathrm{f}}} \leq R_{\mathrm{p,max}}$	$3\,T_{\mathrm{f}} + D_{\mathrm{tx}} \leq D_{\max}$
XOR-PB Scheme 2)	$\frac{L_{\mathrm{h}}+2\,L_{\mathrm{f}}}{T_{\mathrm{f}}} \leq R_{\mathrm{p,max}}$	$d_r \leq \frac{D_{\max}-D_{\mathrm{tx}}-T_{\mathrm{f}}}{T_{\mathrm{f}}}$
RS	$\frac{k}{n} \geq \frac{L_{\mathrm{h}}+L_{\mathrm{f}}}{R_{\mathrm{p,max}}\,T_{\mathrm{f}}}$	$k \leq \frac{(D_{\max}-D_{\mathrm{tx}})\,R_{\mathrm{p,max}}+L_{\mathrm{h}}+L_{\mathrm{f}}}{2\,R_{\mathrm{p,max}}\,T_{\mathrm{f}}}$ $l_{\mathrm{il}} \leq \frac{D_{\max}-D_{\mathrm{tx}}}{2\,k\,T_{\mathrm{f}}}$
RS-PB	$\frac{k}{n} \geq \frac{L_{\mathrm{f}}}{R_{\mathrm{p,max}}\,T_{\mathrm{f}}-L_{\mathrm{h}}}$	$n \leq \frac{D_{\max}-D_{\mathrm{tx}}}{T_{\mathrm{f}}}$ $l_{\mathrm{il}} \leq \frac{D_{\max}-D_{\mathrm{tx}}}{2\,d_{\mathrm{il}}\,T_{\mathrm{f}}}$

Table 4.2: Parameter limitations for different FEC schemes depending on maximum available data rate, $R_{\mathrm{p,max}}$, and maximum tolerable end-to-end delay, $D_{\max}$.

With P_{fl} and $P_{\mathrm{b,s}}$, the mean burst length $\overline{b}$ at the receiver, i.e., the mean number of consecutively lost frames after deinterleaving, is then calculated as

$$\overline{b} = \frac{P_{\mathrm{fl}}}{P_{\mathrm{b,s}}}. \tag{4.12}$$

4.1.5 Repetition Code

When transmitting p redundant copies of each frame, the receiver experiences a frame loss only if all $p+1$ copies of a particular frame are lost. The calculation of the probability of a frame loss at the receiver and the mean burst length, i.e., the average number of consecutively lost frames, has to take into account whether the repeated frames are transmitted in separate packets or piggybacked to the following original packets.

4.1.5.1 Separate Transmission of FEC Frames

For the calculation of the respective loss probabilities, a given channel model with differing transmission time interval T'_{TTI} and packet transmission time τ'_{p} needs to be adjusted as described in Section 3.2 using the following factors, depending on the utilized scheme (cf. Table 4.1):

REP Scheme 1)

$$k_{\mathrm{t}}^{1)} = \frac{T_{\mathrm{TTI}}^{\mathrm{eff},1)}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{f}}}{(p+1)\, T'_{\mathrm{TTI}}} \tag{4.13a}$$

$$k_{\mathrm{p}}^{1)} = \frac{\tau_{\mathrm{p}}}{\tau'_{\mathrm{p}}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}}}{R_{\mathrm{ch}}\, \tau'_{\mathrm{p}}} \tag{4.13b}$$

REP Scheme 2)

$$k_{\mathrm{t}}^{2)} = \frac{T_{\mathrm{TTI}}^{\mathrm{eff},2)}}{T'_{\mathrm{TTI}}} = \frac{(p+2)\, T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{p+2}{p+1} \cdot \frac{T_{\mathrm{f}}}{T'_{\mathrm{TTI}}} \tag{4.14a}$$

$$k_{\mathrm{p}}^{2)} = \frac{\tau_{\mathrm{p}}}{\tau'_{\mathrm{p}}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}}}{R_{\mathrm{ch}}\, \tau'_{\mathrm{p}}} \tag{4.14b}$$

For the adjustment of the transmission time interval with factor k_{t}, the new effective transmission time interval $T_{\mathrm{TTI}}^{\mathrm{eff}}$ is set to the distance between the copies of a frame. For *REP Scheme 1)*, the effect of a block interleaver is considered by a multiplication with the interleaver length l_{il} as explained in Section 4.1.1, resulting in the new effective transmission time interval $T_{\mathrm{TTI}}^{\mathrm{eff},1)} = l_{\mathrm{il}}\, T_{\mathrm{TTI}}$ with $T_{\mathrm{TTI}} = T_{\mathrm{f}}/(p+1)$ from Table 4.1. In *REP Scheme 2)*, the copies of a frame are transmitted with a distance of $p+2$ packet transmission time intervals, resulting in the effective transmission time interval $T_{\mathrm{TTI}}^{\mathrm{eff},2)} = (p+2)\, T_{\mathrm{TTI}}$.

Since every frame is sent $p+1$ times in independent packets, the probability of losing a frame, P_{fl}, is the probability of losing $p+1$ successive packets in the appropriately adjusted channel model (*REP Scheme 1)* and *REP Scheme 2)*):

$$P_{\mathrm{fl}} = P(p+1, p+1). \tag{4.15}$$

A burst is defined as the loss of one or several successive frames while the preceding and following frame of the burst are received. Thus, the probability of a burst start $P_{\mathrm{b,s}}$ is the probability of receiving a frame and losing the following frame. For *REP Scheme 1)* it is given as the probability of receiving at least one of the copies of a frame, i.e., losing up to p of the $p+1$ packets containing this frame, and then losing all copies of the following frame, i.e., all $p+1$ packets containing these frames:

$$P_{\mathrm{b,s}}^{1)} = \sum_{\substack{X,Y \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X} \sum_{i=0}^{p} P_{XY}(i, p+1) \sum_{Z\in\{\mathrm{G,B}\}} P_{YZ}(p+1, p+1). \tag{4.16}$$

For *REP Scheme 2)*, a slightly different approach is used: The probability of a burst start, i.e., the probability of losing all copies of the current frame and receiving at least one copy of the previous frame, can also be calculated by subtracting the probability of the event of losing all copies of the preceding and of the current frame from the probability of losing all copies of the current frame, not taking into account whether the copies of the previous frame are lost. The latter probability is given by P_{fl} in (4.15), calculated from the adjusted channel model with $k_{\mathrm{t}}^{2)}$ and $k_{\mathrm{p}}^{2)}$ as given in (4.14). For the calculation of the probability of losing all copies of the previous and current frame, the implicit interleaving of the packets has to be taken into account. This event is therefore equivalent to the occurrence of a specific loss pattern which shall be denoted as $\{1\, x^p\, 1\}^{p+1}$ according to the following new notation.

Notation for specific loss patterns: This new notation shall be explained with the example above, $\{1\, x^p\, 1\}^{p+1}$. In this pattern notation, x stands for a packet which is either received or lost, 0 stands for a received packet, and 1 for a lost packet. An exponent denotes that a specific part of a pattern or the whole pattern itself occurs a certain number of times in direct sequence. Hence, the pattern $\{1\, x^p\, 1\}^{p+1}$ denotes a $(p+1)$-fold occurrence of the pattern $\{1\, x^p\, 1\}$, which itself consists of a lost packet, followed by p packets that are each either lost or received, followed by another lost packet. The respective probabilities of the (repeated) occurrence of loss patterns are calculated according to the formulas given in Appendix F and shall be denoted by $P^{\mathrm{pat}}(\{..\}^{\cdot\cdot})$. Note: The superscript "pat" differentiates the probability of occurrence of a specific loss pattern from the probability of m losses in n packets which is denoted by $P(m, n)$.

Following this notation, the probability of the event of losing all copies of the previous and current frame, i.e., the probability of occurrence of the loss pattern

$\{1\,x^p\,1\}^{p+1}$, is given as $P^{\text{pat}}(\{1\,x^p\,1\}^{p+1})$, calculated as explained in Appendix F. Note that the involved probabilities in the calculation of $P^{\text{pat}}(\{1\,x^p\,1\}^{p+1})$ need to be derived from a differently adjusted channel model with $k_\text{t} = T_\text{TTI}/T'_\text{TTI} = T_\text{f}/((p+1)\,T'_\text{TTI})$ and $k_\text{p} = k_\text{p}^{2)}$. The probability of a burst start for *REP Scheme 2)* is then given as

$$P_{\text{b,s}}^{2)} = P_\text{fl} - P^{\text{pat}}(\{1\,x^p\,1\}^{p+1}). \tag{4.17}$$

Finally, the mean length of a burst $\overline{b}$ is calculated from the computed probabilities P_fl and $P_\text{b,s}$ of the respective scheme as in (4.12).

4.1.5.2 Piggybacked Transmission of FEC Frames

For the calculation of the loss probabilities in case of piggybacked transmission, the given channel model needs to be adjusted to the transmission time interval and packet size of this FEC scheme, especially considering the increased packet size due to the piggybacked transmission of the frame repetitions. For values of the distance parameter $d_p > 1$, the copies of a frame are transmitted further than one packet apart. This will be considered in the probability calculation of the loss patterns and is therefore not considered in the adaptation of the channel model. Hence, the effective transmission time interval in the adapted model equals the transmission time interval of the scheme, i.e., $T_\text{TTI}^\text{eff} = T_\text{TTI} = T_\text{f}$. The adaptation is done as described in Section 3.2 using the following factors:

$$k_\text{t} = \frac{T_\text{TTI}^\text{eff}}{T'_\text{TTI}} = \frac{T_\text{TTI}}{T'_\text{TTI}} = \frac{T_\text{f}}{T'_\text{TTI}} \quad \wedge \quad k_\text{p} = \frac{\tau_\text{p}}{\tau'_\text{p}} = \frac{L_\text{h} + (p+1)\,L_\text{f}}{R_\text{ch}\,\tau'_\text{p}}. \tag{4.18}$$

For $d_p = 1$, every frame is sent $p+1$ times in successive packets. The probability of losing a frame is given by the probability of losing $p+1$ successive packets:

$$P_\text{fl} = P(p+1, p+1) \tag{4.19}$$

For $d_p > 1$, the probability of losing a frame is the probability of losing all packets containing a copy of this frame, which are spaced d_p packets apart. The probability of the pattern of $d_p - 1$ successive packets (either received or lost) followed by a lost packet, i.e. $\{x^{d_p-1}\,1\}$, can be calculated for different start and end states as shown in Appendix F. The mean loss rate P_fl is then determined[6] by the probability of a $(p+1)$-fold occurrence of pattern $\{x^{d_p-1}\,1\}$:

$$P_\text{fl} = P^{\text{pat}}(\{x^{d_p-1}\,1\}^{p+1}). \tag{4.20}$$

The probability of a burst start also depends on the value of d_p. For $d_p = 1$, the probability of a burst start is given as the probability to first receive a packet

[6] The loss probability P_fl for $d_p > 1$ can also be calculated according to (4.19) using an appropriately downsampled channel model with $k_\text{t} = (d_p\,T_\text{f})/T'_\text{TTI}$.

(containing the current frame and the copies of the p previous frames) and then lose the next frame, i.e. lose the following $p+1$ successive packets:

$$P_{\mathrm{b,s}} = \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P_{XY}(0,1)\, P_{YZ}(p+1,p+1). \tag{4.21}$$

For $d_p = 2$, the probability of a burst start, i.e., the probability of receiving the previous and losing the current frame, is calculated by subtracting the probability of losing the previous and current frame from the probability of losing the current frame independently of whether the previous one is received or lost (cf. calculation in (4.17)):

$$P_{\mathrm{b,s}} = P_{\mathrm{fl}} - P(2(p+1), 2(p+1)) = P^{\mathrm{pat}}(\{x\,1\}^{p+1}) - P(2(p+1), 2(p+1)). \tag{4.22}$$

For $d_p > 2$ the probability of a burst start computes as the probability of losing all $p+1$ copies of two successive frames, $P^{\mathrm{pat}}(\{x^{d_p-2}\,1^2\}^{p+1})$, subtracted from the probability of losing all copies of the second frame, no matter whether the first is lost or not, $P^{\mathrm{pat}}(\{x^{d_p-1}\,1\}^{p+1})$:

$$P_{\mathrm{b,s}} = P^{\mathrm{pat}}(\{x^{d_p-1}\,1\}^{p+1}) - P^{\mathrm{pat}}(\{x^{d_p-2}\,1^2\}^{p+1}). \tag{4.23}$$

Finally, the mean length of a burst $\overline{b}$ is calculated from the computed probabilities P_{fl} and $P_{\mathrm{b,s}}$ as in (4.12).

4.1.6 Exclusive Disjunction (XOR) Codes

The residual frame loss rate and mean burst length at the receiver after error correction, i.e., recovering of lost frames from received XOR combinations, has to take into account whether the FEC frames have been transmitted separately or piggybacked to original packets.

4.1.6.1 Separate Transmission of FEC Frames

The residual frame loss rate and mean burst length for *XOR Scheme 1)* can be calculated as for a Reed-Solomon code with $n = 3$ and $k = 2$ (cf. Section 4.1.7.1).

For the calculation of the respective loss probabilities of *XOR Scheme 2)*, the given channel model needs to be adjusted to the transmission time interval and packet size of the FEC scheme as described in Section 3.2 using the following factors:

$$k_{\mathrm{t}} = \frac{T^{\mathrm{eff}}_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{f}}}{2\, T'_{\mathrm{TTI}}} \quad \wedge \quad k_{\mathrm{p}} = \frac{\tau_{\mathrm{p}}}{\tau'_{\mathrm{p}}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}}}{R_{\mathrm{ch}}\, \tau'_{\mathrm{p}}}. \tag{4.24}$$

The calculation of the residual frame loss probability P_{fl} for *XOR Scheme 2)* is explained with the help of Table 4.3, depicting patterns of lost packets. Assume the loss of the packet containing the original frame d. This frame cannot be recovered from other media and FEC packets if the following packet loss constellations occur:

Packet stream (media frames & XOR combinations)	a	z + a	b	a + b	c	b + c	d	c + d	e	d + e	f	e + f	g	f + g
i) Constellations of packet loss patterns leading to loss of frame d without possibility of reconstruction from XOR combination $c+d$														
$\{1\,1\}$	x	x	x	x	x	x	1	1						
$\{1\,1\,1\,0\}$	x	x	x	x	1	1	1	0						
$\{1\,1\,\{1\,0\}^2\}$	x	x	1	1	1	0	1	0						
$\{1\,1\,\{1\,0\}^3\}$	1	1	1	0	1	0	1	0						
ii) Constellations of packet loss patterns not allowing a reconstruction of frame d from the XOR combination $d+e$														
$\{x\,1\}$									x	1	x	x	x	x
$\{1\,0\,x\,1\}$									1	0	x	1	x	x
$\{\{1\,0\}^2\,x\,1\}$									1	0	1	0	x	1
$\{\{1\,0\}^3\}$									1	0	1	0	1	0

Table 4.3: *XOR Scheme 2)*: Constellations of packet loss patterns leading to an unrecoverable loss of frame d (1 indicating a lost packet, 0 indicating a received packet, and x indicating that this packet is either lost or received). The complete set of constellations leading to this event is derived by combining each of the patterns from i) with each pattern from ii). For this example, $m_r = 3$ and $d_r = 4$ are assumed as explained in the text.

i) See case i) in Table 4.3: If the packet containing FEC frame $c+d$ is lost, the packets before d are irrelevant for recovering of frame d. If $c+d$ is received, frame c must not have been recovered. Similarly, frame c has not been recovered if a) all packets containing this frame or an XOR combination of it with another frame have been lost or if b) FEC frame $b+c$ has been received and frame b has not been received or recovered. This can be extended into the whole past of the packet stream. For the calculation of the probability, however, only about $m_r = 10$ steps are necessary until the contribution can be neglected.

ii) See case ii) in Table 4.3: If the packet containing FEC frame $d+e$ is lost, frame e and all following packets are irrelevant for recovering of d. If $d+e$ is received, frame e must be lost and unrecoverable from future frames. Frame e cannot be recovered if FEC frame $e+f$ is lost or received and frame f cannot be recovered. This can be extended further into the future and is limited by the parameter d_r defined in Section 4.1.2, i.e., by the tolerated delay.

Frame d is irrecoverably lost if one of the packet loss constellations from i) occurs together with a constellation from ii). Hence, the probability of a frame loss at the receiver can be calculated as the sum of the probabilities of all possible combinations of loss patterns from i) and ii):

$$P_{\mathrm{fl}}^{2)} = \sum_{i=0}^{m_r} \Big(\sum_{j=0}^{d_r-2} P^{\mathrm{pat}}(\{1\,1\,\{1\,0\}^{i+j}\,x\,1\}) + P^{\mathrm{pat}}(\{1\,1\,\{1\,0\}^{i+d_r-2}\,1\,0\}) \Big). \quad (4.25)$$

In the calculation, the packets within m_r frame lengths in the past and d_r frame lengths in the future of the respective packet are considered.

The calculation of the probability of a burst start is done in a similar way and results to:

$$P_{\mathrm{b,s}}^{2)} = \sum_{i=0}^{m_r} \Big(\sum_{j=0}^{d_r-2} P^{\mathrm{pat}}(\{0\,x\,\{1\,0\}^i\,1\,1\,\{1\,0\}^j\,x\,1\}) + P^{\mathrm{pat}}(\{0\,x\,\{1\,0\}^i\,1\,1\,\{1\,0\}^{d_r-2}\,1\,0\}) \Big). \tag{4.26}$$

The mean length of a burst $\overline{b}$ is finally calculated from the computed probabilities P_{fl} and $P_{\mathrm{b,s}}$ as in (4.12).

4.1.6.2 Piggybacked Transmission of FEC Frames

The residual frame loss rate and mean burst length for *XOR-PB Scheme 1)* can be calculated as for a Reed-Solomon code with $n = 3$ and $k = 2$ (cf. Section 4.1.7.2).

For the calculation of the respective loss probabilities of *XOR-PB Scheme 2)*, the given channel model needs to be adjusted to the packet transmission time interval and packet size of this FEC scheme as described in Section 3.2 using the following factors:

$$k_{\mathrm{t}} = \frac{T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{T_{\mathrm{f}}}{T'_{\mathrm{TTI}}} \quad \wedge \quad k_{\mathrm{p}} = \frac{\tau_{\mathrm{p}}}{\tau'_{\mathrm{p}}} = \frac{L_{\mathrm{h}} + 2L_{\mathrm{f}}}{R_{\mathrm{ch}}\,\tau'_{\mathrm{p}}}. \tag{4.27}$$

For *XOR-PB Scheme 2a)*, the loss patterns leading to the loss of a frame and to a burst start are shown in Table 4.4. A single frame loss results if three successive packets are lost. If two successive packets are lost, the next is received, and the following lost, two frames will be lost. These cases determine the frame loss rate:

$$P_{\mathrm{fl}}^{2a)} = P(3,3) + 2\,P^{\mathrm{pat}}(\{1^2\ 0\ 1\}). \tag{4.28}$$

The probability of a burst start is given as the probability of receiving a packet and then either losing the following three packets or losing two packets, followed by a received packet, and finally losing another packet:

$$P_{\mathrm{b,s}}^{2a)} = P^{\mathrm{pat}}(\{0\ 1^3\}) + P^{\mathrm{pat}}(\{0\ 1^2\ 0\ 1\}). \tag{4.29}$$

The probabilities of occurrence of the given patterns of lost and received packets, $\{1^2\ 0\ 1\}$, $\{0\ 1^3\}$, and $\{0\ 1^2\ 0\ 1\}$, are calculated as explained in Appendix F.

For *XOR-PB Scheme 2b)*, the residual frame loss rate is calculated by considering all possible loss patterns in a group of 6 successive packets (cf. Table 4.5):

$$P_{\mathrm{fl}}^{2b)} = P^{\mathrm{pat}}(\{1\ x\ 1^2\}) + 2\,P^{\mathrm{pat}}(\{1^3\ 0\ 1\}) + 3\,P^{\mathrm{pat}}(\{1^3\ 0\ 0\ 1\}). \tag{4.30}$$

The probability of a burst start results to

$$\begin{aligned} P_{\mathrm{b,s}}^{2b)} = \quad & P^{\mathrm{pat}}(\{0\ 1\ x\ 1^2\}) + P^{\mathrm{pat}}(\{0\ 1^3\ 0\ 1\}) + P^{\mathrm{pat}}(\{0\ 1^3\ 0^2\ 1\}) \\ & + P^{\mathrm{pat}}(\{0\ 1^2\ 0\ 1^2\}). \end{aligned} \tag{4.31}$$

For *XOR-PB Scheme 2c)*, finally, the residual frame loss rate is calculated by considering all possible loss patterns in a group of 8 successive packets (cf. Table 4.6):

$$\begin{aligned} P_{\mathrm{fl}}^{2c)} = \quad & P^{\mathrm{pat}}(\{1\ x^2\ 1^2\}) + 2\,P^{\mathrm{pat}}(\{1^2\ x\ 1\ 0\ 1\}) + 3\,P^{\mathrm{pat}}(\{1^4\ 0^2\ 1\}) \\ & + 4\,P^{\mathrm{pat}}(\{1^4\ 0^3\ 1\}) \end{aligned} \tag{4.32}$$

And the probability of a burst start is computed as

$$\begin{aligned} P_{\mathrm{b,s}}^{2c)} = \quad & P^{\mathrm{pat}}(\{0\ 1\ x^2\ 1^2\}) + P^{\mathrm{pat}}(\{0\ 1^2\ x\ 0\ 1^2\}) + P^{\mathrm{pat}}(\{0\ 1^3\ 0^2\ 1^2\}) \\ & + P^{\mathrm{pat}}(\{0\ 1^2\ x\ 1\ 0\ 1\}) + P^{\mathrm{pat}}(\{0\ 1^3\ 0\ 1\ 0\ 1\}) + P^{\mathrm{pat}}(\{0\ 1^4\ 0^2\ 1\}) \\ & + P^{\mathrm{pat}}(\{0\ 1^4\ 0^3\ 1\}) \end{aligned} \tag{4.33}$$

For all schemes, the resulting average burst length is calculated from the respective probabilities of a loss and a burst start as in (4.12).

Packet Stream	a $y+z$	b $z+a$	c $a+b$	d $b+c$	e $c+d$
loss of b	x	1	1	1	x
loss of b and c	x	1	1	0	1
loss of a and b	1	1	0	1	x
burst start	0	1	1	1	x
at b	0	1	1	0	1

Table 4.4: *XOR-PB Scheme 2a)*: Packet loss patterns leading to frame loss and burst start; patterns consist of received packets (0), lost packets (1), and arbitrary packets (x).

Packet Stream	z $w+x$	a $x+y$	b $y+z$	c $z+a$	d $a+b$	e $b+c$	f $c+d$	g $d+e$
loss of b	x	x	1	x	1	1	x	x
loss of b and c	x	x	1	1	1	0	1	x
loss of b, c, d	x	x	1	1	1	0	0	1
	x	0	1	x	1	1	x	x
burst start	x	0	1	1	1	0	1	x
at b	x	0	1	1	1	0	0	1
	0	1	1	0	1	1	x	x

Table 4.5: *XOR-PB Scheme 2b)*: Packet loss patterns leading to frame loss and burst start; patterns consist of received packets (0), lost packets (1), and arbitrary packets (x).

Packet Stream	y $u+v$	z $v+w$	a $w+x$	b $x+y$	c $y+z$	d $z+a$	e $a+b$	f $b+c$	g $c+d$	h $d+e$	i $e+f$
loss of b	x	x	x	1	x	x	1	1	x	x	x
loss of b,c	x	x	x	1	1	x	1	0	1	x	x
loss of $b-d$	x	x	x	1	1	1	1	0	0	1	x
loss of $b-e$	x	x	x	1	1	1	1	0	0	0	1
	x	x	0	1	x	x	1	1	x	x	x
	x	x	0	1	1	x	1	0	1	x	x
burst start	x	x	0	1	1	1	1	0	0	1	x
at b	x	x	0	1	1	1	1	0	0	0	1
	x	0	1	1	x	0	1	1	x	x	x
	x	0	1	1	1	0	1	0	1	x	x
	0	1	1	1	0	0	1	1	x	x	x

Table 4.6: *XOR-PB Scheme 2c)*: Packet loss patterns leading to frame loss and burst start; patterns consist of received packets (0), lost packets (1), and arbitrary packets (x).

4.1.7 Block Codes (e.g., Reed-Solomon Codes)

Block codes, e.g., Reed-Solomon codes, are able to recover up to $n-k$ frames in an encoding group of n frames by erasure correction. For the derivation of the residual frame loss rate and mean burst length at the receiver after erasure correction it has to be taken into account whether the FEC frames have been transmitted separately or piggybacked to original packets. For erasure correction scenarios, systematic codes are preferred over non-systematic codes, because the receiver can then still utilize any received information frames even if an erasure correction of other frames is not possible. Therefore, a systematic block code is assumed in the following.

4.1.7.1 Separate Transmission of FEC Frames

For the calculation of the respective loss probabilities, the given channel model needs to be adjusted to the packet transmission time interval and packet size of the *RS* scheme (cf. Table 4.1) as described in Section 3.2 using the following factors:

$$k_\mathrm{t} = \frac{T_\mathrm{TTI}^\mathrm{eff}}{T'_\mathrm{TTI}} = \frac{l_\mathrm{il}\, T_\mathrm{TTI}}{T'_\mathrm{TTI}} = \frac{l_\mathrm{il}\, k\, T_\mathrm{f}}{n\, T'_\mathrm{TTI}} \quad \wedge \quad k_\mathrm{p} = \frac{\tau_\mathrm{p}}{\tau'_\mathrm{p}} = \frac{L_\mathrm{h} + L_\mathrm{f}}{R_\mathrm{ch}\, \tau'_\mathrm{p}}. \tag{4.34}$$

The resulting frame loss rate after possible corrections of up to $n-k$ packet losses in the group of n packets is calculated by taking all possible loss distributions in this group into account which lead to the unrecoverable loss of one or several media frames:

$$P_\mathrm{fl} = \sum_{i=n-k+1}^{n} \sum_{j=i-n+k}^{\min(k,i)} \frac{j}{k} \sum_{\substack{X,Y \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X} P_{XY}(j,k) \Big(P_{Y\mathrm{G}}(i-j, n-k) + P_{Y\mathrm{B}}(i-j, n-k) \Big) \tag{4.35}$$

with the indices of the two sums given as

- i: number of losses in a block of n packets; no erasure correction possible if the number of losses is at least $n-k+1$
- j: number of losses in the k original media packets; only these losses contribute to the loss rate
- $i-j$: remaining losses in the $n-k$ parity packets

In (4.35), the packet losses in the k packets containing the original media frames and those in the $n-k$ packets containing the parity (FEC) frames have to be differentiated. Therefore, both possible states the channel may be in at the first of the FEC packets have to be considered.

The probability of a burst start needs to be calculated separately for each possible position within the group of k media frames (cf. [Frossard 2001] for a similar

derivation considering the simplified Gilbert model). Let $P_{\mathrm{b,s}}$ denote the probability of a burst start at the j-th frame of a group of k media frames. For a start of a loss burst, the previous frame needs to be received or corrected and the current frame needs to be lost and unrecoverable. Hence, for a burst start at the first frame of a block, i.e., for $j = 1$, the following probabilities are required:

- $P_{\mathrm{X}}^{\mathrm{cor,prev}}$: probability that the packet with the last media frame (not a parity frame) of the previous block is lost but can be recovered, and that the channel is in state X, $X \in \{\mathrm{G}, \mathrm{B}\}$, at the first packet of the current block
- $P_{\mathrm{X}}^{\mathrm{rec,prev}}$: probability that the packet with the last media frame (not a parity frame) of the previous block is received, and that the channel is in state X, $X \in \{\mathrm{G}, \mathrm{B}\}$, at the first packet of the current block

For the special case of $k = 1$, $P_{\mathrm{X}}^{\mathrm{cor,prev}}$ describes the probability that the last and only media packet of the previous block is lost and that a maximum of $n-2$ of the following $n-1$ parity packets are lost as well, so that it can still be recovered. $P_{\mathrm{X}}^{\mathrm{cor,prev}}$ further includes the probability of the channel being in state X at the packet directly following the last parity packet, i.e., the single media packet of the current block, with $X \in \{\mathrm{G}, \mathrm{B}\}$:

$$P_{\mathrm{X}}^{\mathrm{cor,prev}} = \sum_{\substack{V,W \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},V}\, P_{VW}(1,1) \sum_{i=0}^{n-2} P_{WX}(i, n-1). \tag{4.36}$$

The probability $P_{\mathrm{X}}^{\mathrm{rec,prev}}$, i.e., the probability that the last media packet of the previous block is received and the channel is in state $X \in \{\mathrm{G}, \mathrm{B}\}$ at the single media packet of the current block, computes to

$$P_{\mathrm{X}}^{\mathrm{rec,prev}} = \sum_{\substack{V,W \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},V}\, P_{VW}(0,1) \sum_{i=0}^{n-1} P_{WX}(i, n-1). \tag{4.37}$$

For $k > 1$, the probability that the last media packet of the previous block is lost but corrected and that the channel is in state $X \in \{\mathrm{G}, \mathrm{B}\}$ at the packet containing the first media frame of the current block, $P_{\mathrm{X}}^{\mathrm{cor,prev}}$, can be derived as

$$P_{\mathrm{X}}^{\mathrm{cor,prev}} = \sum_{e=0}^{n-k-1} \sum_{b=0}^{\min(k-1,e)} \sum_{\substack{U,V,W \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},U}\, P_{UV}(b, k-1)\, P_{VW}(1,1)\, P_{WX}(e-b, n-k). \tag{4.38}$$

Here, the number of additional losses e must not exceed $n-k-1$ so that the last media frame from the k-th packet is still corrected. The e packet losses are

distributed over the packets preceding and following the lost k-th packet, such that there are b losses in the preceding $k-1$ and $e-b$ losses in the following $n-k$ packets.

The probability that the last media packet of the previous block is received and that the channel is in state $X \in \{\mathrm{G},\mathrm{B}\}$ at the packet containing the first media frame of the current block, $P_{\mathrm{X}}^{\mathrm{rec,prev}}$, computes to

$$P_{\mathrm{X}}^{\mathrm{rec,prev}} = \sum_{\substack{V,W \\ \in\{\mathrm{G},\mathrm{B}\}}} P_{\mathrm{s},V}\, P_{VW}(0,1) \sum_{i=0}^{n-k} P_{WX}(i, n-k). \tag{4.39}$$

Finally, the probabilities are summed up, considering both possible channel states $X \in \{\mathrm{G},\mathrm{B}\}$ for the packet containing the first media frame of the current block:

$$P_{\mathrm{X}}^{\mathrm{prev}} = P_{\mathrm{X}}^{\mathrm{cor,prev}} + P_{\mathrm{X}}^{\mathrm{rec,prev}}. \tag{4.40}$$

The probability of a burst start at the first media packet of a block, $P_{\mathrm{b,s}}(1)$, is now given as the probability that the last media packet of the previous block is received or corrected (as computed above), the first media packet is lost, and enough following packets are lost so that it cannot be recovered, i.e., at least $n-k$ losses occur in the following $n-1$ packets:

$$P_{\mathrm{b,s}}(1) = \frac{1}{k} \sum_{X\in\{\mathrm{G},\mathrm{B}\}} P_{\mathrm{X}}^{\mathrm{prev}} \sum_{Y\in\{\mathrm{G},\mathrm{B}\}} P_{XY}(1,1) \sum_{i=n-k}^{n-1} \Big(P_{Y\mathrm{G}}(i, n-1) + P_{Y\mathrm{B}}(i, n-1) \Big). \tag{4.41}$$

The probability of a burst start at the second media packet of a block, $P_{\mathrm{b,s}}(2)$, is given as the probability to receive the first packet of this block, lose the second, and then have at least $n-k$ further losses in the following $n-2$ packets so that it cannot be corrected:

$$P_{\mathrm{b,s}}(2) = \frac{1}{k} \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G},\mathrm{B}\}}} P_{\mathrm{s},X} P_{XY}(0,1) P_{YZ}(1,1) \sum_{i=n-k}^{n-2} \Big(P_{Z\mathrm{G}}(i, n-2) + P_{Z\mathrm{B}}(i, n-2) \Big). \tag{4.42}$$

Finally, the probability of a burst start at the j-th media packet of a block with $2 < j \leq k$ can be computed considering the necessary number of losses in the preceding and following packets of the block. Let e denote the number of further errors so that the burst start is not correctable. These e losses are distributed over the $j-2$ previous and $n-j$ following packets: $e-b$ losses in the first $j-2$ packets; the $(j-1)$-th packet is received and the j-th packet is lost, together marking the

start of the burst; further b losses in the $n-j$ following packets. The probability of a burst start at the j-th media packet of a block, $P_{\mathrm{b,s}}(j)$, is then calculated as:

$$P_{\mathrm{b,s}}(j) = \frac{1}{k} \sum_{e=n-k}^{n-2} \sum_{b=\max(0,e-(j-2))}^{\min(e,n-j)} \sum_{\substack{W,X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},W}\, P_{WX}(e-b, j-2) \ldots$$

$$\ldots P_{XY}(0,1) P_{YZ}(1,1) \Big(P_{Z\mathrm{G}}(b, n-j) + P_{Z\mathrm{B}}(b, n-j) \Big). \quad (4.43)$$

The overall probability of a burst start can then be calculated as the sum of the probabilities of a burst start at the j-th packet:

$$P_{\mathrm{b,s}} = \sum_{i=1}^{k} P_{\mathrm{b,s}}(i). \quad (4.44)$$

Finally, the mean length of a burst is calculated from the probabilities P_{fl} in (4.35) and $P_{\mathrm{b,s}}$ in (4.44) as in (4.12).

4.1.7.2 Piggybacked Transmission of FEC Frames

For the calculation of the respective loss probabilities, the given channel model needs to be adjusted to the packet transmission time interval and packet size of the *RS-PB* scheme as described in Section 3.2 using the following factors:

$$k_{\mathrm{t}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{TTI}}}{T'_{\mathrm{TTI}}} = \frac{l_{\mathrm{il}}\, T_{\mathrm{f}}}{T'_{\mathrm{TTI}}} \quad \wedge \quad k_{\mathrm{p}} = \frac{\tau_{\mathrm{p}}}{\tau'_{\mathrm{p}}} = \frac{L_{\mathrm{h}} + \frac{n}{k} L_{\mathrm{f}}}{R_{\mathrm{ch}}\, \tau'_{\mathrm{p}}}. \quad (4.45)$$

The resulting frame loss rate after possible corrections of up to $n-k$ frame losses in the group of n frames can be calculated in the same way as for the separate transmission of the FEC frames (cf. (4.35)):

$$P_{\mathrm{fl}} = \sum_{i=n-k+1}^{n} \sum_{j=i-n+k}^{\min(k,i)} \frac{j}{k} \sum_{\substack{X,Y \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X} P_{XY}(j,k) \Big(P_{Y\mathrm{G}}(i-j, n-k) + P_{Y\mathrm{B}}(i-j, n-k) \Big). \quad (4.46)$$

However, it is important to note that although (4.35) and (4.46) are the same, the calculation is based on differently adjusted channel models because of differing transmission time intervals and packet sizes. The frame loss rates for the cases of separate and piggybacked transmission of the FEC frames will therefore in general not be the same on a given channel.

The calculation of the probability of a burst start for the piggybacked transmission of the FEC frames differs from that for the separate transmission, but the approach is similar. Again, the probability of a burst start has to be calculated

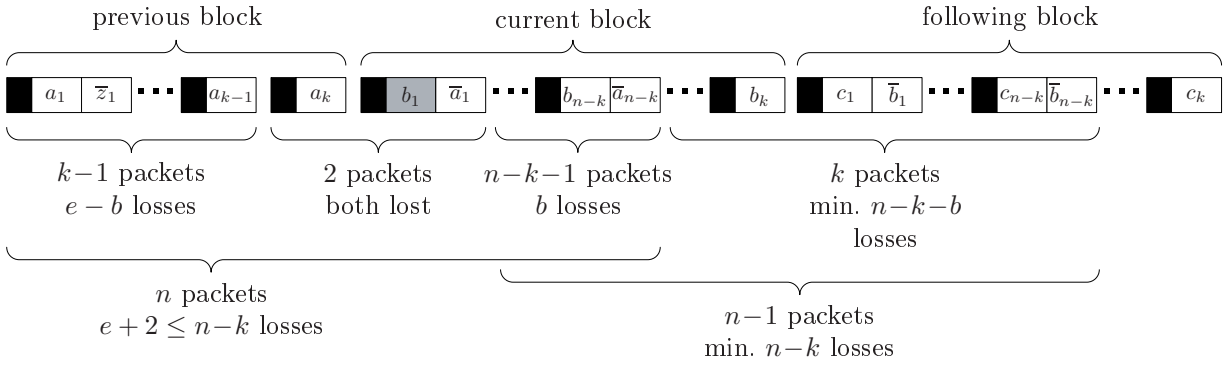


Figure 4.3: Block code, piggybacked transmission of FEC frames: Illustration for the calculation of the probability of a burst start at first frame, b_1, of current block.

separately for every position j in the group of k media frames. In the piggybacked transmission scheme, each group or block consists of k packets which contain the original k media frames and possibly further FEC frames. The $n-k$ FEC frames of a block are subsequently piggybacked to the packets of the following block(s).

For a burst start at the first frame of a block of k media frames, i.e., for $j=1$, the first packet of the block, which contains this frame, is lost. The previous frame needs to be available and it has to be distinguished whether it has been received with the last packet of the previous block or whether that packet has been lost and the frame can be recovered utilizing FEC data. In case the previous packet has been received, there need to be at least $n-k$ losses in the $n-1$ packets following the lost packet so that the loss of the first frame cannot be corrected. The probability of this event is calculated as

$$P_{\mathrm{b,s}}^{(A)}(1) = \frac{1}{k} \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X} P_{XY}(0,1) P_{YZ}(1,1) \sum_{i=n-k}^{n-1} \Big(P_{Z\mathrm{G}}(i,n-1) + P_{Z\mathrm{B}}(i,n-1) \Big). \tag{4.47}$$

If $k>1$ and $n-k>1$, there is another possible event which causes a burst start at the first frame of an encoding block: The last packet of the previous block is lost, but the contained media frame can be recovered, and the first packet of the current block is lost and the media frame cannot be recovered. Because of the piggybacked transmission of the FEC frames, a possible recovering of these two frames involves partly the same packets. The derivation of the probability of this event is explained with the help of the illustration given in Figure 4.3[7]. The packets containing the last frame of the previous block, a_k, and the first frame of the current block, b_1, are assumed lost. For a burst start at frame b_1, frame a_k must be recoverable utilizing other media frames together with the parity frames $\overline{a}_i$, and frame b_1 must not be recoverable from the other media frames and the parity frames $\overline{b}_i$, $1 \leq i \leq n-k$.

[7]Without loss of generality, it has been assumed in this example that $n < 2k-1$.

Since $\overline{a}_1$ is lost together with b_1, the $k-1$ packets preceding the packet with frame a_k and the $n-k-1$ packets following that with frames b_1 and $\overline{a}_1$ must together not contain more than $n-k-2$ losses, such that a_k can be recovered. This number of losses is expressed by e, which is split into $e-b$ losses among the preceding $k-1$ packets and b losses among the following packets. At the same time, the $n-1$ packets following the lost packet with frame b_1 must contain at least $n-k$ more losses such that b_1 cannot be recovered. Hence, depending on the number of losses in the $n-k-1$ packets following that of b_1, i.e., depending on b, the last k packets with media and FEC frames of the current block need to contain at least $n-k-b$ losses.

Incorporating these considerations, the probability of the considered event, i.e., that the last packet of the previous block is lost while its media frame can be recovered and that the first packet of the current block is lost while its media frame cannot be recovered, is derived with the following formula:

$$P_{\mathrm{b,s}}^{(B)}(1) = \sum_{e=0}^{n-k-2} \sum_{b=\max(0,e-(k-1),n-2\,k)}^{\min(e,n-k-1)} \sum_{\substack{W,X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},W}\, P_{WX}(e-b,k-1)\ldots$$

$$\ldots P_{XY}(2,2)\, P_{YZ}(b,n-k-1) \sum_{i=n-k-b}^{k} \Big(P_{Z\mathrm{G}}(i,k) + P_{Z\mathrm{B}}(i,k)\Big). \quad (4.48)$$

The calculation is taking the different states that the channel might be in at certain intermediate packets into account. If $k=1$ or $n-k=1$, this probability is set to zero, i.e., $P_{\mathrm{b,s}}^{(B)}(1)=0$.

The total probability of a burst start at the first frame of an encoding block is then given by combining the probabilities of the two events from (4.47) and (4.48):

$$P_{\mathrm{b,s}}(1) = P_{\mathrm{b,s}}^{(A)}(1) + P_{\mathrm{b,s}}^{(B)}(1). \quad (4.49)$$

For the occurrence of a burst start at the second frame of a block, the first packet needs to be received, the second lost, and then there need to be at least $n-k$ further losses in the following $n-2$ packets so that the losses cannot be recovered, resulting in the same equation as for the separate transmission (cf. (4.42)):

$$P_{\mathrm{b,s}}(2) = \frac{1}{k} \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X} P_{XY}(0,1) P_{YZ}(1,1) \sum_{i=n-k}^{n-2} \Big(P_{Z\mathrm{G}}(i,n-2) + P_{Z\mathrm{B}}(i,n-2)\Big). \quad (4.50)$$

The probability of a burst start at the j-th media frame of a block ($2 < j \leq k$)

can also be computed as for the separate transmission of FEC packets in (4.43):

$$P_{\mathrm{b,s}}(j) = \frac{1}{k} \sum_{e=n-k}^{n-2} \sum_{b=\max(0,e-(j-2))}^{\min(e,n-j)} \sum_{\substack{W,X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},W}\, P_{WX}(e-b, j-2) \ldots$$

$$\ldots P_{XY}(0,1)\, P_{YZ}(1,1) \Big(P_{Z\mathrm{G}}(b, n-j) + P_{Z\mathrm{B}}(b, n-j) \Big). \quad (4.51)$$

It is again important to note that compared to the corresponding probabilities for the separate transmission of the FEC frames, the calculations in (4.50) and (4.51) are based on a differently adjusted channel model to account for the different transmission time interval and packet size for the piggybacked transmission.

The overall probability of a burst start can then be calculated as the sum of the probabilities of a burst start at the j-th packet:

$$P_{\mathrm{b,s}} = \sum_{i=1}^{k} P_{\mathrm{b,s}}(i). \quad (4.52)$$

Finally, the mean length of a burst is calculated from the probabilities P_{fl} in (4.46) and $P_{\mathrm{b,s}}$ in (4.52) as in (4.12).

4.2 Theoretical Determination of Residual Losses after Retransmission

On wireless links with short transmission delays and an available feedback channel it may be beneficial to implement retransmissions upon request instead of generally adding redundancy with forward error correction schemes, as discussed in Section 2.7.1. In the following, the properties of such a retransmission scheme will be derived, i.e., delay, data rate, as well as the residual loss rate are determined in dependence on a maximum number of transmission attempts. The maximum number of transmissions for each packet including the first attempt shall be given by the parameter N_{rtx}. This parameter depends on the transmission rate of the channel as well as further data rate and delay constraints of network and application.

The retransmission scheme will be considered for high speed wireless links with short delay, on which all retransmission attempts can be executed before the next media packet needs to be transmitted. Therefore, a simple *Stop & Wait ARQ* algorithm will be assumed for the derivation. In this ARQ variant, the sender waits for a feedback from the receiver after each transmitted packet. If a negative acknowledgment is received, i.e., the packet has not been correctly received by the receiver, the sender initiates a retransmission of the last packet until the transmission attempt counter for this packet reaches the given maximum. Upon reception of a positive acknowledgment or when the maximum number of attempts is reached, the sender

stops the retransmission, resets the counter, and continues with the transmission of the following packets. On channels with a high transmission delay, other ARQ schemes would lead to a better performance which continue transmitting media packets while still waiting for the feedback information on previous packets.

Consider the example in Figure 4.4. Four successive packets of a media stream, marked as a, b, c, and d, are transmitted over a wireless link with retransmission capabilities and a maximum of $N_{\mathrm{rtx}} = 4$ transmission attempts. After each packet transmission, the receiver sends back a short feedback packet including a positive or negative acknowledgment. In this example, packet a has to be transmitted three times before it is received correctly, packet b is already correctly received after the first attempt, and packet d needs one retransmission. Packet c is transmitted four times, i.e., the maximum number of attempts. If it is still not received correctly in the last attempt, the packet will be considered lost and not retransmitted again. The negative acknowledgment transmitted in the given example indicates that packet c is finally lost.

Transmission Time Interval and Packet Transmission Time

The transmission time interval between two media packets containing different consecutive media frames, T_{TTI}, equals the frame length T_{f}, i.e., $T_{\mathrm{TTI}} = T_{\mathrm{f}}$. In case retransmissions are required for a packet, the transmission time interval between two successive transmission attempts of the same packet, $T_{\mathrm{TTI}}^{\mathrm{rtx}}$, depends on the time it needs to receive the feedback from the receiving side of the wireless link:

$$T_{\mathrm{TTI}}^{\mathrm{rtx}} = \tau_{\mathrm{p}} + \delta_{\mathrm{ACK}} + \tau_{\mathrm{ACK}} + \delta_p, \tag{4.53}$$

with the transmission time of the media packet τ_{p}, the channel access delay δ_{ACK} for the feedback transmission at the receiver, the transmission time of the feedback packet τ_{ACK}, and finally the channel access delay δ_p for the retransmission attempt. Since no redundancy is added to the media packets itself, the transmission time of each packet remains as:

$$\tau_{\mathrm{p}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}}}{R_{\mathrm{ch}}}. \tag{4.54}$$

It is assumed that the receiver will transmit feedback information for each packet. The transmission time of a feedback packet with a length of L_{ACK} bit including headers, which contains the positive or negative acknowledgment of successfully receiving the previously transmitted media packet, is given as:

$$\tau_{\mathrm{ACK}} = \frac{L_{\mathrm{ACK}}}{R_{\mathrm{ch}}}. \tag{4.55}$$

The possibility of losing the feedback packet itself, which in turn will lead to a retransmission of the media packet whether needed or not, is neglected in the considerations. It is assumed that the loss probability of a feedback packet is considerably smaller than that of a media packet. On WLAN channels, e.g., the feedback packets

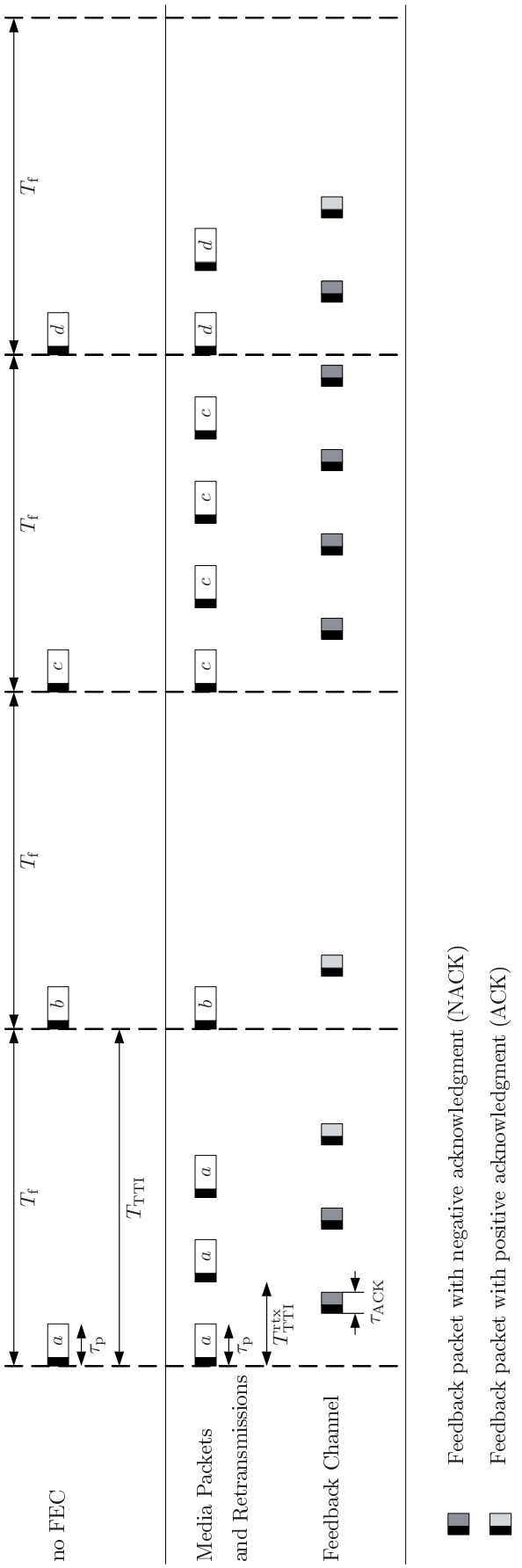


Figure 4.4: Retransmission of not acknowledged packets. In this example, a maximum number of transmission attempts $N_\mathrm{rtx} = 4$ is assumed.

are usually transmitted at a lower channel transmission rate and are therefore less error-prone.

The use of a *Stop & Wait ARQ scheme* sets a constraint on the maximum number of transmission attempts. To avoid buffer overflow and real-time problems, all transmission attempts must be executed within the transmission time interval of the original packets, i.e., within the frame length T_f of the contained media:

$$N_\mathrm{rtx}\, T_\mathrm{TTI}^\mathrm{rtx} \overset{!}{\leq} T_\mathrm{f}. \tag{4.56}$$

For the calculation of loss probabilities for a given channel model, the model needs to be adapted to the transmission time interval and packet transmission time (packet size), of the retransmission scheme derived in (4.53) and (4.54). The adaptation is done as described in Section 3.2 using the following factors:

$$k_\mathrm{t} = \frac{T_\mathrm{TTI}^\mathrm{rtx}}{T'_\mathrm{TTI}} = \frac{\tau_\mathrm{p} + \delta_\mathrm{ACK} + \tau_\mathrm{ACK} + \delta_p}{T'_\mathrm{TTI}}, \tag{4.57}$$

$$k_\mathrm{p} = \frac{\tau_\mathrm{p}}{\tau'_\mathrm{p}}. \tag{4.58}$$

In the following, all respective probabilities are assumed to be derived from the appropriately resampled channel model. The superscripts indicating the resampling of the channel model are omitted to facilitate readability.

Packet Data Rate

Since every packet may require a different number of transmission attempts, only an average of the expected total packet data rate can be derived. With the channel model adapted to the transmission time interval between retransmissions as given above, the probability of requiring n transmission attempts for a given packet can be calculated as:

$$P_\mathrm{rtx}(n) = \begin{cases} P(0,1) & ; \quad n = 1 \\ \displaystyle\sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P_{XY}(n-1,n-1)\, P_{YZ}(0,1) & ; \quad 1 < n < N_\mathrm{rtx} \\ P(N_\mathrm{rtx}-1, N_\mathrm{rtx}-1) & ; \quad n = N_\mathrm{rtx} \end{cases}, \tag{4.59}$$

with the probabilities of m packet losses in n consecutive packets, $P_{XY}(m,n)$, considering start and end states, $X, Y \in \{\mathrm{G},\mathrm{B}\}$, as introduced in Section 3.4. If a packet is transmitted N_rtx times, the respective probability $P_\mathrm{rtx}(N_\mathrm{rtx})$ does not give an indication whether it is received correctly at the final attempt. This will be considered later when determining the residual loss probability. With the probabilities P_rtx, the average number of transmission attempts per packet is given as:

$$\overline{n}_\mathrm{rtx} = \sum_{n=1}^{N_\mathrm{rtx}} n\, P_\mathrm{rtx}(n). \tag{4.60}$$

The average packet data rate, not considering the rate required for the feedback transmissions, then results to:

$$R_{\mathrm{p}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}}}{T_{\mathrm{f}}} \cdot \overline{n}_{\mathrm{rtx}}. \tag{4.61}$$

Including the rate required for the feedback transmissions, the average total data rate results to:

$$R'_{\mathrm{p}} = \frac{L_{\mathrm{h}} + L_{\mathrm{f}} + L_{\mathrm{ACK}}}{T_{\mathrm{f}}} \cdot \overline{n}_{\mathrm{rtx}}. \tag{4.62}$$

End-To-End Delay

The end-to-end delay for the retransmission scheme depends on the maximum allowed number of retransmissions N_{rtx}. Assuming a voice conversation scenario, it consists of one frame length at the sender side (frame has to be collected before transmission), the transmission time over the channel D_{tx} (including the packet transmission time τ_{p} and any further propagation and network delays), and finally $N_{\mathrm{rtx}} - 1$ transmission time intervals at the receiver to wait for the retransmissions:

$$\begin{aligned} &D_s = T_{\mathrm{f}} \quad \wedge \quad D_r = (N_{\mathrm{rtx}} - 1)\, T_{\mathrm{TTI}}^{\mathrm{rtx}} \\ &\Rightarrow D = D_s + D_{\mathrm{tx}} + D_r = T_{\mathrm{f}} + (N_{\mathrm{rtx}} - 1)\, T_{\mathrm{TTI}}^{\mathrm{rtx}} + D_{\mathrm{tx}}. \end{aligned} \tag{4.63}$$

Constrained Data Rate and Delay

The channel transmission rate R_{ch} sets a limit on the maximum number of transmission attempts via (4.56), resulting in the following condition:

$$N_{\mathrm{rtx}} \leq \frac{T_{\mathrm{f}}}{T_{\mathrm{TTI}}^{\mathrm{rtx}}} = \frac{T_{\mathrm{f}}}{\tau_{\mathrm{p}} + \delta_{\mathrm{ACK}} + \tau_{\mathrm{ACK}} + \delta_p} = \frac{T_{\mathrm{f}}}{\frac{L_{\mathrm{h}} + L_{\mathrm{f}} + L_{\mathrm{ACK}}}{R_{\mathrm{ch}}} + \delta_{\mathrm{ACK}} + \delta_p}. \tag{4.64}$$

If the packet data rate on the channel must not exceed a given value $R_{\mathrm{p,max}}$ which is below the actual transmission rate of the channel R_{ch}, i.e., $R_{\mathrm{p,max}} < R_{\mathrm{ch}}$, the average number of transmission attempts for every packet is further limited according to

$$\overline{n}_{\mathrm{rtx}} \leq \frac{R_{\mathrm{p,max}}\, T_{\mathrm{f}}}{L_{\mathrm{h}} + L_{\mathrm{f}}}, \tag{4.65}$$

and therefore, the maximum number of transmissions N_{rtx} is constrained according to (4.60), depending on the loss distribution of the channel which is reflected in the probabilities $P_{\mathrm{rtx}}(n)$.

A maximum tolerable end-to-end delay $D \leq D_{\max}$ also sets a constraint on the maximum number of transmissions:

$$N_{\mathrm{rtx}} \leq \frac{D_{\max} - D_{\mathrm{tx}} - T_{\mathrm{f}}}{T_{\mathrm{TTI}}^{\mathrm{rtx}}} + 1. \tag{4.66}$$

Residual Frame Loss Probability and Distribution

For each packet, a maximum of N_{rtx} transmission attempts is made, i.e., the packet is retransmitted several times if it is not received correctly. Only if the last transmission attempt is still not successful, the packet and the media frame it contains are finally lost. The probability of a frame loss is therefore determined as the probability to lose all N_{rtx} transmitted packets:

$$P_{\mathrm{fl}} = P(N_{\mathrm{rtx}}, N_{\mathrm{rtx}}). \tag{4.67}$$

This equation also applies if the condition of (4.56) is not met.

The probability of a burst start is given as the probability to first receive a frame, i.e., not to lose all possible N_{rtx} transmissions of that frame, and then to lose the next frame, i.e., lose all N_{rtx} transmissions of that frame. In this calculation, the channel state transitions in the time between the last transmission attempt of a frame and the first attempt of the following frame have to be considered. This time can be divided into s transmission time intervals $T_{\mathrm{TTI}}^{\mathrm{rtx}}$, which is rounded to the next nearest integer according to

$$s = \left\lfloor \frac{T_{\mathrm{f}} - N_{\mathrm{rtx}}\, T_{\mathrm{TTI}}^{\mathrm{rtx}}}{T_{\mathrm{TTI}}^{\mathrm{rtx}}} + \frac{1}{2} \right\rfloor. \tag{4.68}$$

For this consideration, condition (4.56) has to be met. Then, the probability of a burst start is calculated from the probability of not losing all N_{rtx} packets with the previous frame, the s channel transitions until the transmission of the first packet with the following frame, and the probability of losing all N_{rtx} packets with that frame:

$$P_{\mathrm{b,s}} = \sum_{\substack{W,X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},W} \cdot \left[\sum_{i=0}^{N_{\mathrm{rtx}}-1} P_{WX}(i, N_{\mathrm{rtx}}) \right] \cdot \left[\sum_{i=0}^{s} P_{XY}(i, s) \right] \cdot P_{YZ}(N_{\mathrm{rtx}}, N_{\mathrm{rtx}}). \tag{4.69}$$

Finally, the mean length of a burst is calculated as

$$\overline{b} = \frac{P_{\mathrm{fl}}}{P_{\mathrm{b,s}}}. \tag{4.70}$$

4.3 Forward Error Correction on Channels with Varying Transmission Delay (Jitter)

In heterogeneous packet-switched networks with delay variations (jitter) in the transmission of packets, forward error correction (FEC) and jitter buffer management need to be considered together and have to be optimized jointly. An applied

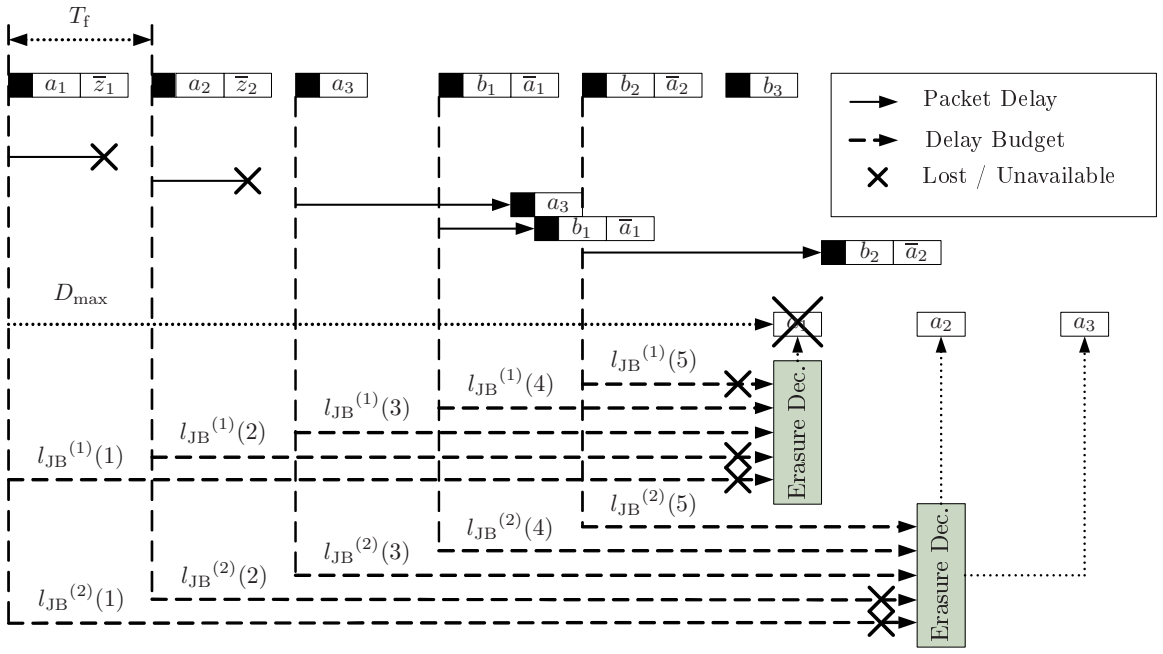


Figure 4.5: Example of using Forward Error Correction (FEC) on channels with packet losses and jitter. The packets of an FEC block have different effective jitter buffer lengths.

FEC scheme may not only recover lost frames but also regenerate frames from delayed packets and thereby allow for a reduction of the jitter buffer length at the receiver.

For a stream of media packets without additional FEC, each packet has the same delay budget, i.e., the same tolerated end-to-end delay, and the probability of a packet loss due to jitter can be calculated as derived in Section 3.3. When FEC is applied to a group of frames and the generated parity frames are either transmitted in separate packets or attached (piggybacked) to following packets, the assumption of the same delay budget does not hold anymore. The receiver needs to wait for all packets containing frames from an encoding block before it can start attempting to recover the possibly lost (or delayed) first frame of the block by erasure decoding. At this point in time, the frames of the encoding block have different delay budgets, i.e., a different maximum allowed delay at the receiver depending on the point in time they have been sent. Furthermore, each frame may be necessary to recover any of the other frames in the block. The delay budget of a frame therefore also depends on the location of the lost frame within the encoding block.

An example is given in Figure 4.5. Here, a (5,3)-block code is applied to a group of three successive media frames, resulting in two additional FEC frames which are piggybacked to the following packets. The first two packets containing the first two original frames, a_1 and a_2, are lost, and the next three packets experience a variable delay. At the time point that a_1 needs to be decoded at the receiver, which is defined by the tolerated end-to-end delay D_{max}, only two of the five packets containing the media and FEC frames of block a have been received. Therefore, the erasure decoder at the receiver cannot recover the first frame a_1. At the time

the second frame needs to be decoded, the delayed packet has finally arrived and the erasure decoder can recover frame a_2. The packet with the third frame a_3 has been received successfully, it does not have to be recovered.

Assume a maximum tolerated delay for each frame of $D_{\max}$. When the receiver considers frame j in an encoding block of n frames, which have been transmitted in successive packets, the preceding frames $i = 1, \ldots, j-1$ will have a delay budget larger than $D_{\max}$, and the following frames $i = j+1, \ldots, n$ will have a smaller delay budget. Hence, consider each packet's effective delay budget ${l_{\mathrm{JB}}}^{(j)}(i)$ at the correction attempt for frame j:

$$ {l_{\mathrm{JB}}}^{(j)}(i) = D_{\max} - (i - 1 - j + 1)\, T_{\mathrm{f}} = D_{\max} - (i - j)\, T_{\mathrm{f}}, \tag{4.71} $$

with $1 \leq i \leq n$, $1 \leq j \leq k$. Then, the probability that frame i of an encoding block is not available at the time frame j needs to be recovered is calculated as

$$ P_{l,J}^{(j)}(i) = 1 - F_X({l_{\mathrm{JB}}}^{(j)}(i)), \tag{4.72} $$

with the cumulative distribution function (CDF) of the transmission delay distribution $F_X(x)$ (cf. Section 3.3). Hence, (4.72) describes a time dependent function of the probability of packet loss due to jitter. These jitter loss probabilities $P_{l,J}^{(j)}(i)$ can be combined with the Gilbert-Elliott model for packet loss according to (3.34b) in Section 3.3.3. Subsequently, the probabilities of specific residual loss patterns, e.g., m losses in an encoding block of n frames, can be determined as described in Section 3.4. In this calculation, special attention has to be given to the dependency of $P_{l,J}$ on both the position of the considered frame within the encoding block, j, and the position of the other frames of the block, i, such that in every step of the calculation, the correct probability is used. Finally, the overall loss rate after erasure correction on a channel with packet losses and jitter can be determined according to the derivation in Section 4.1, utilizing the position dependent probabilities of loss patterns.

5

System Optimization for Speech and Audio Transmission over Packet Networks

The optimization of the perceived quality for audio transmission in heterogeneous packet networks involves several constraints which depend on the particular application and the present network characteristics. This chapter shows how the developed framework of channel modeling and theoretical determination of residual frame loss distributions, which has been developed in the previous chapters, can be applied to the design of actual applications.

The flexibility of the channel model introduced in Chapter 3, which can be adapted to different transmission time intervals and packet sizes, provides the base for a fair comparison of different parameterizations. Utilizing the appropriately adapted channel model, residual frame loss rate, burst length, as well as resulting delay and required data rate are determined according to Chapter 4. A closed-form mathematical solution yielding a single optimal parameter set cannot be derived because of the necessary channel model adaptation for each parameterization. Therefore, each transmission scheme and parameterization needs to be calculated separately. However, the constraints of delay and data rate usually reduce the required number of cases to compare considerably.

The following sections analyze important optimization problems for common scenarios of packet-based speech and music transmission and determine the optimal parameter settings. In Section 5.1, the general optimization criteria and constraints are described in dependence on different application demands and network characteristics, and the variable parameters for the optimization are reviewed. Section 5.2 discusses general important questions regarding the choice of system parameters which arise in various scenarios. In Sections 5.3 to 5.6, finally, four specific scenarios will be addressed which have a strong relevance for current communication

services: 1) music streaming over Wireless LAN, 2) IP telephony using uncompressed speech (PCM) on Wireless LAN; 3) IP telephony (Voice over IP) on dedicated UMTS packet channels using the AMR speech codec; and 4) IP telephony in networks with a high variance of the packet transmission delay, e.g., long distance connections over the public Internet. Under the given constraints for data rate and delay, the optimal system parameterization will be determined for each scenario, e.g., regarding the frame length to transmit in each packet and the application of forward error correction or retransmission schemes.

5.1 Optimization Problem: Criteria, Parameters, and Constraints

The optimal parameterization of a system requires a clear definition of a) the criterion to optimize, b) the involved system parameters and their effect on the optimization criterion, and c) possible constraints on these parameters, whether set by specific demands on the system or by certain physical realization constraints.

5.1.1 Optimization Criterion

Optimization criterion for the parameterization of multimedia transmission services, i.e., speech, music, or video, has to be the perceived quality at the receiver. What determines this quality depends on the individual media type and application, as well as on the type of impairments on the transmission channel and how they are dealt with, i.e., the employed error protection and concealment technique. The quality of a voice call, for example, is determined by the quality of the speech signal and the quality of the conversation, i.e., the interactivity between the conversation partners which may be affected by too much signal delay. This quality can be determined from other system parameters using the ITU-T E-Model as explained in Appendix H.3.

For a scenario of music streaming, the quality is determined solely by the resulting signal quality, i.e., depending on distortions by packet loss and the limits of employed frame loss concealment algorithms. Signal delay has no impairment effect in streaming applications unless it exceeds a value of about 1-2 seconds. Therefore, if only a single audio codec with a fixed concealment scheme is considered, e.g., the MP3 audio codec, the resulting packet loss rate and burstiness may be used as quality criterion.

5.1.2 Variable Parameters in Packet-Based Multimedia Transmission

The general structure of packet-based multimedia transmission has been comprehensively described in Chapter 2. The variable parameters of the transmission of speech and music signals shall be shortly reviewed and summarized in this section as they will be the focus of the following optimization approaches.

5.1.2.1 Media codec and packetization of media frames

The first parameter is the choice of the codec itself. As described in Section 2.3, a variety of standardized codecs are available which differ with respect to encoding rate, speech/audio quality, and sensitivity to transmission errors like frame erasures. The choice of the encoding rate will have an influence on the base quality that can be achieved in case of an error free transmission. The relation between encoding rate and quality for different codecs is in general not a linear function but can nevertheless be taken as qualitative indication. If a multi-rate codec is considered, e.g., the AMR and AMR Wideband speech codecs, the quality increases monotonically with the increasing encoding rate.

The second parameter is the length of the new signal segment which is transmitted in each packet. Some codecs operate with an arbitrary frame length (e.g., PCM), most codecs, however, have a fixed frame length of, e.g., 20 or 30 ms. The segment length per packet can then be an integer multiple of this frame length, i.e., the variable parameter in this case is the number of frames placed in each packet. The segment length per packet contributes directly to the end-to-end delay of the transmission. This delay component occurs at the sender side because the speech segment needs to be collected before encoding and packetization. On the other hand, the segment length per packet also influences the resulting data rate by determining the required amount of packet header overhead. The choice of the segment length per packet may therefore be limited if the application tolerates only a specific maximum end-to-end delay or data rate, or if an increasing end-to-end delay leads to a decreasing quality. Furthermore, the packet size may have an influence on the probability of loss, especially on wireless packet channels. Finally, the segment length determines the minimum length of a lost segment in case of unrecoverable packet loss. Both the length and frequency of losses determine the possible quality that can be achieved by packet loss concealment algorithms at the receiver.

5.1.2.2 Forward Error Correction (FEC)

Forward error correction techniques on application level can be used to protect the transmission against some packet losses and recover a certain number of lost frames at the receiver as discussed in detail in Chapter 4. Each FEC scheme has its own set of variable parameters which determine the code rate, i.e., the amount of redundancy to transmit, and thereby the error correction capabilities. Furthermore, an interleaved transmission of the media and FEC frames can be chosen, which enhances the robustness against burst losses at the expense of a higher end-to-end delay. The choice of the FEC parameters is usually limited by data rate and delay constraints of the application and network. Finally, the choice of transmitting the FEC frames either in separate packets or piggybacked to packets with other media frames is a further variable in the optimization process if not decided by other constraints, e.g., a limited data rate. This choice will influence the packet size and transmission time interval and therefore also the packet loss distribution on wireless channels.

5.1.2.3 Receiver Buffer and Frame Loss Concealment

The receiver contains a buffer for compensating variances in the packet transmission delay through the network and to collect all packets containing the frames of an encoding block if FEC is used. The length of this buffer is a variable parameter which has direct influence on the resulting frame loss rate and on the resulting delay. The choice of this buffer length has to find the optimal compromise between these two impairment effects. Lost or delayed frames which cannot be recovered by the FEC erasure correction are replaced by an estimation from the packet loss concealment routine. This routine is not standardized and can therefore be implemented differently by each system developer. However, different frame erasure concealment algorithms will not be considered as variable parameter. It is assumed that the system contains a fixed routine which should be the best available or affordable one regarding its computational complexity.

5.1.3 Application Demands: Audio Quality and Delay

Different applications or services of packet-based multimedia transmission have different demands on the network resources as well as on quality and end-to-end delay. The required data rate on the transmission channel depends on the type of signal to transmit, e.g., video, music, or speech signals, on the frame length per packet, and on further applied error protection schemes. Furthermore, each application demands a certain quality of service in terms of signal quality, error robustness, and end-to-end delay. These differences may lead to different strategies for an optimal parameterization of the transmission.

5.1.3.1 Speech Conversation (IP telephony, Voice over IP)

In speech conversations, i.e., telephone applications, some signal distortions are tolerated by the users if the artifacts are not too extreme and the frequency of their occurrence is not too high. Of higher importance than the actual signal quality is the intelligibility of the speech and the possibility of interaction without too much delay, e.g., for interrupting with a question, signaling understanding, or showing emotional reaction. An end-to-end (mouth-to-ear) delay of up to 150 ms is generally considered unnoticeable. According to [ITU-T Rec. G.114 2003], one-way delays exceeding 400 ms are unacceptable. However, for low bit rate codecs the maximum tolerable delay is usually considerably lower and rather amounts to values around 300 ms [D. De Vleeschauwer and Petit 2000].

5.1.3.2 Music Streaming

In streaming applications, the playback of the media signal (video, music, or speech) begins shortly after starting the transmission, usually with a moderate delay for buffering at the receiver. For streaming applications, the end-to-end delay of the transmission is far less important than for conversational speech services. The user will tolerate a short delay of up to 1–2 seconds before the playback starts. The signal quality of the playback, on the other hand, is of much greater importance

and the user will not tolerate noticeable signal dropouts or impairments. Hence, packet losses must not lead to the loss of long signal segments which cannot be well concealed. The available delay budget can therefore be utilized for adding redundancy (forward error correction), the retransmission of lost packets, a large receiver buffer for jitter compensation, or a combination of these. The joint objective of these measures is to reduce the frame loss rate to a minimum.

5.1.4 Network Constraints: Transmission Delay, Errors, and Capacity

The transmission network will in general set certain constraints for the applications. The available data rate on the channel is usually limited, depending on the type of network and possibly concurring applications and services. Furthermore, the transmission through the network requires a certain time with possible variation (jitter). Together with any additional delay of the application itself, e.g., for framing, encoding, and forward error correction, and with the length of the receiver buffer, the total end-to-end delay for a specific application scenario results, which might have an impact on the resulting quality. Furthermore, the transmission through the network might be subject to transmission errors resulting in packet losses.

In a heterogeneous network consisting of several parts with different transmission characteristics, e.g., different wired and wireless transmission channels, the end-to-end delay results as the sum of all contributing channels together, while the data rate constraint is determined by that intermediate transmission link which provides the minimal data rate along the transmission path.

In the process of finding the optimal parameterization, these constraints will limit the choice of transmission parameters like codec rate, frame length, forward error correction scheme, etc. The network constraints may also lead to the conclusion that a specific application cannot be realized on a given network such that it meets its quality demands.

5.2 General Questions and Considerations

The following sections will discuss general questions regarding the choice of transmission scheme and parameterization which arise when designing an application for packet-based transmission of multimedia signals like speech or music. Most of these questions cannot be answered universally. The solution, i.e., the optimal choice of a specific parameter or transmission scheme, rather depends on the specific application and network characteristics under consideration. The general approaches discussed in this section are therefore applied to various realistic scenarios in Sections 5.3–5.6 of this chapter.

5.2.1 Choice of Frame Length per Packet

In the packet-based transmission of speech or general audio signals, the amount of new data to transmit in each packet is to some degree flexible, as explained in Section 2.4.1. The signal segment length defined here as frame length per packet, T_f, is only constrained by the frame length of the respective codec, T_c, and can theoretically assume an arbitrary multiple of this frame length. Some waveform based codecs, e.g., PCM for speech signals [ITU-T Rec. G.711 1988], even support an arbitrary frame length down to the duration of just a few samples. Most other codecs, however, have frame lengths of 5–30 ms.

Effect on Data Rate and Signal Delay

In practice, the choice of the frame length per packet is limited by delay constraints of the application and data rate constraints of the transmission channel. A short frame length leads to a higher packet rate and thereby to a higher overhead of packet header information which has to be transmitted. If no forward error correction is considered, the required IP packet data rate depends on the frame length per packet, T_f, and the codec's encoding rate, R_c, as derived in (2.5), i.e.,

$$R_\mathrm{p} = \frac{L_\mathrm{h} + L_\mathrm{plh}}{T_\mathrm{f}} + R_\mathrm{c}, \tag{5.1}$$

with the size of all protocol headers from RTP down to layer 2 of the considered network, L_h, and the codec dependent RTP payload header size L_plh. As can be seen, the header overhead is inversely proportional to the frame length transmitted per packet. The larger the frame length per packet is, the lower is the additionally required data rate for packet headers, i.e., the more rate efficient the packet transmission becomes. If the transmission channel provides a maximum transmission rate of $R_\mathrm{p,max}$, the packet data rate in (5.1) must not exceed this value, i.e., $R_\mathrm{p} \leq R_\mathrm{p,max}$. Hence, the frame length needs to be larger than a certain minimum:

$$T_\mathrm{f} \geq \frac{L_\mathrm{h} + L_\mathrm{plh}}{R_\mathrm{p,max} - R_\mathrm{c}}. \tag{5.2}$$

Note that the considered media codec itself may have a fixed frame length T_c, so that the frame length per packet, T_f, can only assume multiples of this minimum length.

In a conversational scenario, e.g., a *Voice over IP* call, a signal segment with frame length T_f needs to be collected first, before it is encoded, packetized, and transmitted. The end-to-end delay, D, therefore includes the frame length as delay contribution at the sender in addition to the packet transmission delay through the network, D_tx, and the receiver buffer length, D_buf, as defined in (2.6). The use of a large frame length therefore leads to an increase of the end-to-end delay which might have a negative effect on the conversational quality if the transmission delay

in the network is already high. If the total end-to-end delay is limited to $D_{\max}$, the frame length must be shorter than a maximum value according to

$$T_{\mathrm{f}} \leq D_{\max} - D_{\mathrm{tx}} - D_{\mathrm{buf}}. \tag{5.3}$$

For streaming applications, all frames can be assumed to be already encoded and readily available from a storage medium. The end-to-end delay then consists only of the packet transmission delay, as defined in (2.7). The choice of the frame length is in this case not constrained by the maximum tolerated delay.

Choice of Frame Length in Conjunction with FEC and Retransmission

An additional factor to be considered in the optimization of the frame length per packet is the possible application of forward error correction or retransmission, as discussed in Chapter 4. Since the packet data rate decreases with increasing frame length, the frame length can be deliberately increased to provide capacity for the application of FEC at the same data rate. With an arbitrary FEC scheme of code rate r_{c} and *separate* transmission of the FEC frames, the packet data rate computes to

$$R_{\mathrm{p}} = \frac{1}{r_{\mathrm{c}}} \frac{L_{\mathrm{p}}}{T_{\mathrm{f}}} = \frac{1}{r_{\mathrm{c}}} \frac{L_{\mathrm{h}} + L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}}{T_{\mathrm{f}}} = \frac{1}{r_{\mathrm{c}}} \left(\frac{L_{\mathrm{h}} + L_{\mathrm{plh}}}{T_{\mathrm{f}}} + R_{\mathrm{c}} \right). \tag{5.4}$$

Hence, for a maximum data rate $R_{\mathrm{p,max}}$ and a desired code rate r_{c} of the FEC scheme, the frame length needs to be increased to a new minimum according to

$$T_{\mathrm{f}} \geq \frac{L_{\mathrm{h}} + L_{\mathrm{plh}}}{r_{\mathrm{c}}\, R_{\mathrm{p,max}} - R_{\mathrm{c}}}. \tag{5.5}$$

For a *piggybacked* transmission of the FEC frames in the original media packets, the packet data rate results to

$$R_{\mathrm{p}} = \frac{L_{\mathrm{p}}}{T_{\mathrm{f}}} = \frac{L_{\mathrm{h}} + \frac{1}{r_{\mathrm{c}}} \left(L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}} \right)}{T_{\mathrm{f}}} = \frac{L_{\mathrm{h}} + \frac{1}{r_{\mathrm{c}}}\, L_{\mathrm{plh}}}{T_{\mathrm{f}}} + \frac{R_{\mathrm{c}}}{r_{\mathrm{c}}}. \tag{5.6}$$

The new minimum frame length can then be calculated as

$$T_{\mathrm{f}} \geq \frac{r_{\mathrm{c}}\, L_{\mathrm{h}} + L_{\mathrm{plh}}}{r_{\mathrm{c}}\, R_{\mathrm{p,max}} - R_{\mathrm{c}}}. \tag{5.7}$$

When applying FEC, the increase of the frame length is again constrained by the delay demands of the application. The maximum frame length depends on the specific end-to-end delay for the considered FEC scheme, as derived in Chapter 4 for a variety of common FEC schemes (cf. Table 4.1).

For retransmission schemes as discussed in Section 4.2, the choice of the frame length has an influence on the parameterization of the system. An increase of the frame length leads to a higher possible number of transmission attempts, because the header overhead in the transmission is reduced (cf. (4.64)). Depending on

the desired maximum number of transmission attempts, the frame length needs to assume a minimum value according to (4.56) so that a transmission of all attempts is possible within the time length of the frame length:

$$T_\mathrm{f} \geq N_\mathrm{rtx}\, T_\mathrm{TTI}^\mathrm{rtx}, \tag{5.8}$$

and

$$T_\mathrm{TTI}^\mathrm{rtx} = \tau_\mathrm{p} + \delta_\mathrm{ACK} + \tau_\mathrm{ACK} + \delta_p, \tag{5.9}$$

with

$$\tau_\mathrm{p} = \frac{L_\mathrm{h} + R_\mathrm{c}\, T_\mathrm{f}}{R_\mathrm{ch}}. \tag{5.10}$$

Hence, the minimum frame length results as

$$T_\mathrm{f} \geq N_\mathrm{rtx}\, (\delta_\mathrm{ACK} + \tau_\mathrm{ACK} + \delta_p)\, \frac{R_\mathrm{ch}}{R_\mathrm{ch} - N_\mathrm{rtx}\, R_\mathrm{c}}. \tag{5.11}$$

Interrelation between Frame Length and Packet Loss Probability

The frame length per packet directly determines the size of the packets and the frequency with which they are transmitted (cf. values of transmission time interval T_TTI and packet transmission time τ_p in Table 4.1, Section 4.1.2). The choice of the frame length therefore also needs to take the error characteristics on the transmission channel into account. Wireless transmission channels, for example, usually experience a strong interdependence between packet size and loss probability. Larger packets, i.e., packets with longer frame lengths, are more likely to contain residual errors and to be discarded. Short frame lengths, on the other side, lead to smaller packet sizes, but a higher frequency of packet transmission. Channels with burst errors might therefore cause losses of several successive packets if they are transmitted shortly after each other. The resulting frame loss rate and distribution at the receiver therefore needs to be considered for each application and network scenario separately. It can be determined for an appropriately adapted channel model according to the formulas derived in Chapter 4.

Quality Optimization

The resulting quality of a packet transmission service of media signals depends on the base quality of the applied source codec, the error characteristic of the transmission, i.e., the rate and distribution of frame losses and the length of the lost segments. For conversational applications such as *Voice over IP* telephony also the end-to-end signal delay. The influence of delay and loss rate on the resulting quality can be assessed with the standardized E-Model, as discussed in Appendix H.3. However, the dependence between quality and different loss lengths has not yet been sufficiently resolved in in the E-Model. In the scenarios discussed later in this chapter, the effect of the frame length is therefore assessed by curves of loss rates and mean burst lengths only.

Conclusion

To summarize, the maximum packet data rate and the maximum end-to-end delay, i.e., the constraints of application and network, define the boundaries for the frame length per packet, T_f. The optimal choice of T_f within these boundaries depends on codec constraints, the expected channel characteristic, and the application's priorities regarding data rate and delay. Especially on transmission channels where the loss probability of a packet depends on the actual size of the packet, a careful choice of the length is required. In Section 5.4, the choice of the frame length will be discussed for the specific scenario of a PCM speech conversation in Wireless LAN (Voice over WLAN, VoWLAN).

5.2.2 Forward Error Correction versus Retransmission

As discussed in Section 2.7.1, the use of retransmission techniques (ARQ schemes) is not feasible for delay sensitive applications if the transmission delay on the considered channel is too high. In these cases the additional two-way delay of sending back the retransmission request and retransmitting the original packet normally exceeds the application's tolerated maximum delay. An exemplary scenario is a *Voice over IP* call over the public Internet. A further scenario where retransmissions are usually not applied, is a multicast transmission of speech, audio, or video signals to a group of receivers. The reason here is the high number of required retransmissions for an increasing number of receivers. Assuming nearly independent channel characteristics for each receiver, the probability that a single packet has not been received by at least one of the receivers becomes fairly high, leading to a considerable increase of the average number of necessary transmission attempts. The following analysis is based on the simplifying assumption that all considered receivers experience the same channel quality, i.e., that the same channel model applies. In a real-life scenario, the calculation needs to focus on that group of receivers which experience the worst channel quality, i.e., the receivers at the boundary of the transmission range, because this group determines the number of necessary transmission attempts.

The probability of requiring n transmission attempts for a single packet when sending to a multicast group of N_r receivers (for which the same channel model applies) can be derived from the probability of needing n transmission attempts for a single arbitrary receiver, $P_\mathrm{rtx}(n)$, $1 \leq n \leq N_\mathrm{rtx}$, as defined in (4.59) in Section 4.2. First, the probability that such a single receiver requires n or less transmission attempts is derived as the sum of the probabilities that it requires exactly i transmission attempts, with $i = 1, \ldots, n$:

$$P^{(1)}_\mathrm{rtx,cum}(n) = \sum_{i=1}^{n} P_\mathrm{rtx}(i). \tag{5.12}$$

The probability that all N_r receivers together require n or less transmission attempts is then calculated by raising this probability to the power of N_r:

$$P_{\mathrm{rtx,cum}}^{(N_r)}(n) = \left(P_{\mathrm{rtx,cum}}^{(1)}(n)\right)^{N_r} = \left(\sum_{i=1}^{n} P_{\mathrm{rtx}}(i)\right)^{N_r}. \tag{5.13}$$

Finally, the probability of requiring exactly n transmission attempts for a single packet in a multicast scenario with N_r receivers is given as the probability that at least one of the receivers requires n transmission attempts and no receiver requires more than that. This probability can be derived by taking the probability of n or less transmission attempts for all receivers, $P_{\mathrm{rtx,cum}}^{(N_r)}(n)$, and subtracting from it the probability of $n-1$ or less transmission attempts for all receivers, $P_{\mathrm{rtx,cum}}^{(N_r)}(n-1)$:

$$\begin{aligned} P_{\mathrm{rtx}}^{(N_r)}(n) &= P_{\mathrm{rtx,cum}}^{(N_r)}(n) - P_{\mathrm{rtx,cum}}^{(N_r)}(n-1) \\ &= \left(\sum_{i=1}^{n} P_{\mathrm{rtx}}(i)\right)^{N_r} - \left(\sum_{i=1}^{n-1} P_{\mathrm{rtx}}(i)\right)^{N_r}, \end{aligned} \tag{5.14}$$

with $n \geq 2$, and $P_{\mathrm{rtx}}^{(N_r)}(1) = P_{\mathrm{rtx,cum}}^{(N_r)}(1) = \left(P_{\mathrm{rtx}}(1)\right)^{N_r}$. With the probabilities $P_{\mathrm{rtx}}^{(N_r)}(n)$, the average number of transmission attempts per packet in a multicast transmission scenario with N_r receivers which experience the same channel quality is then calculated as:

$$\overline{n}_{\mathrm{rtx}}^{(N_r)} = \sum_{n=1}^{N_{\mathrm{rtx}}} n\, P_{\mathrm{rtx}}^{(N_r)}(n). \tag{5.15}$$

The average number of transmission attempts $\overline{n}_{\mathrm{rtx}}$ and therefore the resulting packet data rate R_{p} increases with an increasing number of receivers N_r. At the same time, the higher packet data rate may also be utilized by a FEC scheme with a suitable code rate r_{c}, at some point possibly exceeding the error correction capabilities of the retransmission scheme. The following derivation provides a comparison of both approaches regarding their data rate efficiency. Furthermore, formulas are derived which determine the possible FEC code rate r_{c} which results in the same packet data rate as required by the retransmission scheme. Both separate and piggybacked transmission of FEC frames are considered.

First, a generally applicable packet level code rate $r_{\mathrm{c,p}}$ shall be defined as the ratio between the packet data rate without FEC or retransmission, $R_{\mathrm{p,0}}$, and the packet data rate with the respective scheme, R_{p}, i.e,

$$r_{\mathrm{c,p}} = \frac{R_{\mathrm{p,0}}}{R_{\mathrm{p}}}. \tag{5.16}$$

For retransmission schemes, this packet level code rate results as the reciprocal value of the average number of transmission attempts per packet, $\overline{n}_{\mathrm{rtx}}^{(N_r)}$:

$$r_{\mathrm{c,p}}^{(\mathrm{rtx})} = \frac{1}{\overline{n}_{\mathrm{rtx}}^{(N_r)}}. \tag{5.17}$$

For FEC schemes, the calculation of $r_{\mathrm{c,p}}$ differs depending on whether the FEC frames are transmitted separately or piggybacked. For a separate transmission, the packet data rate results as

$$R_{\mathrm{p}}^{(\mathrm{fec,sep})} = \frac{1}{r_{\mathrm{c}}} \frac{L_{\mathrm{h}} + L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}}{T_{\mathrm{f}}}, \tag{5.18}$$

with the code rate of the FEC scheme, r_{c}. Hence, the packet level code rate for separate transmission of FEC frames equals the code rate of the FEC scheme, i.e.,

$$r_{\mathrm{c,p}}^{(\mathrm{fec,sep})} = r_{\mathrm{c}}. \tag{5.19}$$

For a piggybacked transmission, the packet data rate computes to

$$R_{\mathrm{p}}^{(\mathrm{fec,pb})} = \frac{L_{\mathrm{h}} + \frac{1}{r_{\mathrm{c}}} \left(L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}\right)}{T_{\mathrm{f}}}. \tag{5.20}$$

The packet level code rate is then calculated as

$$r_{\mathrm{c,p}}^{(\mathrm{fec,pb})} = \frac{L_{\mathrm{h}} + L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}}{L_{\mathrm{h}} + \frac{1}{r_{\mathrm{c}}} \left(L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}\right)}. \tag{5.21}$$

The data rate efficiency of retransmission and FEC schemes can now be directly compared from the respective packet level code rates $r_{\mathrm{c,p}}$. An increasing number of receivers in a retransmission scenario leads to a higher number of required transmission attempts as derived above. The use of a FEC scheme instead of retransmission becomes more data rate efficient once the packet level code rate of the retransmission scheme gets below the fixed packet level code rate of the considered FEC scheme, i.e., $r_{\mathrm{c,p}}^{\mathrm{rtx}} < r_{\mathrm{c,p}}^{\mathrm{fec}}$. For separate transmission, this condition results to

$$r_{\mathrm{c}} > \frac{1}{\overline{n}_{\mathrm{rtx}}^{(N_r)}}. \tag{5.22}$$

For piggybacked transmission, the FEC scheme is more data rate efficient if

$$r_{\mathrm{c}} > \frac{L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}}{\overline{n}_{\mathrm{rtx}}^{(N_r)} \left(L_{\mathrm{h}} + L_{\mathrm{plh}} + R_{\mathrm{c}}\, T_{\mathrm{f}}\right) - L_{\mathrm{h}}}. \tag{5.23}$$

The increase of the number of transmissions with increasing number of receivers shall be visualized on two exemplary scenarios. The first scenario is a multicast streaming of MP3 encoded music signals over WLAN, as will be discussed in more detail in Section 5.3. In Figure 5.1, the average required number of transmission attempts $\overline{n}_{\mathrm{rtx}}^{(N_r)}$ is plotted against the number of receivers N_r for different channel SNRs . For low channel SNRs, the required number of transmission attempts increases steeply especially for numbers of receivers below 20, e.g., for SNR = 10 dB and $N_r = 10$ reaching a value of about $\overline{n}_{\mathrm{rtx}}^{(10)} = 4.5$. Hence, the use of forward error correction instead of retransmission would be preferable in this scenario as long as

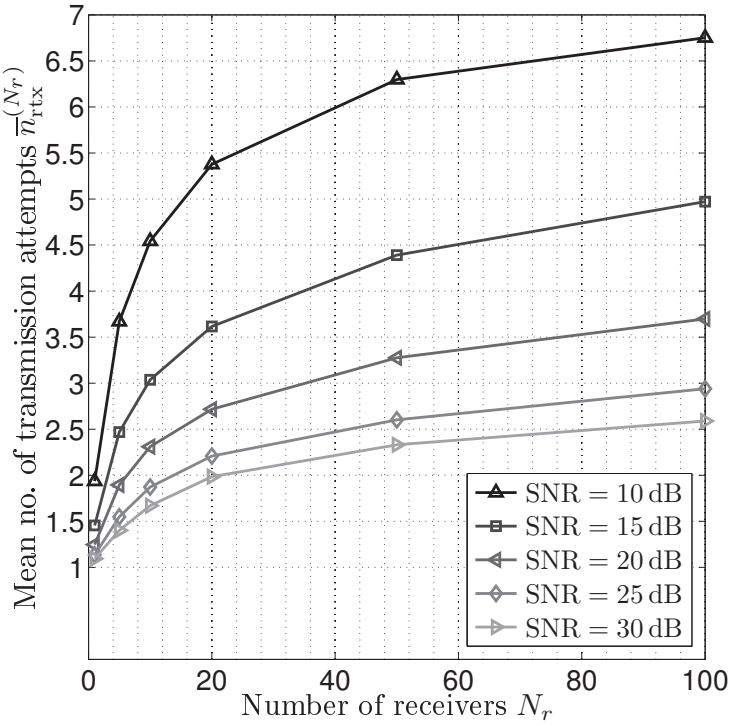


Figure 5.1: Multicast scenario of MP3 music streaming on a WLAN channel (no header compression): Average number of transmission attempts in dependence on the number of receivers; results for different channel SNRs.

a code rate on packet level $r_{\mathrm{c,p}} > 1/\overline{n}_{\mathrm{rtx}}^{(10)} = 1/4.5$ is sufficient for achieving at least a similar performance as the retransmission scheme. With (5.22) and (5.23), this results to a code rate of the FEC scheme of $r_{\mathrm{c}} > 1/4.5$ for separate and $r_{\mathrm{c}} > 1/5.07$ for piggybacked transmission, considering the given scenario. It is assumed that the higher delay requirement due to the application of FEC is not of significance for streaming applications.

The second scenario differs from the first only in the type of media considered. Instead of music, a PCM encoded speech signal shall be transmitted with an arbitrary frame length T_{f} in each packet. Figure 5.2 shows the relation between transmission attempts and number of receivers in three graphs for three different channel SNRs. Each graph shows the curves for different frame lengths T_{f}. The number of transmission attempts again increases steeply with an increasing number of receivers in the range below $N_r = 20$. The slope of the curves becomes less steep for larger N_r. Besides the strong dependence on the channel SNR, the curves for the different frame lengths T_{f} also indicate a considerable dependency between the frame length and the number of transmission attempts, as will be discussed in more detail in Section 5.4.

5.2.3 Choice of Forward Error Correction Scheme

In scenarios where the end-to-end delay over the network is too high for applying retransmission, the use of packet based forward error correction (FEC) is an effective

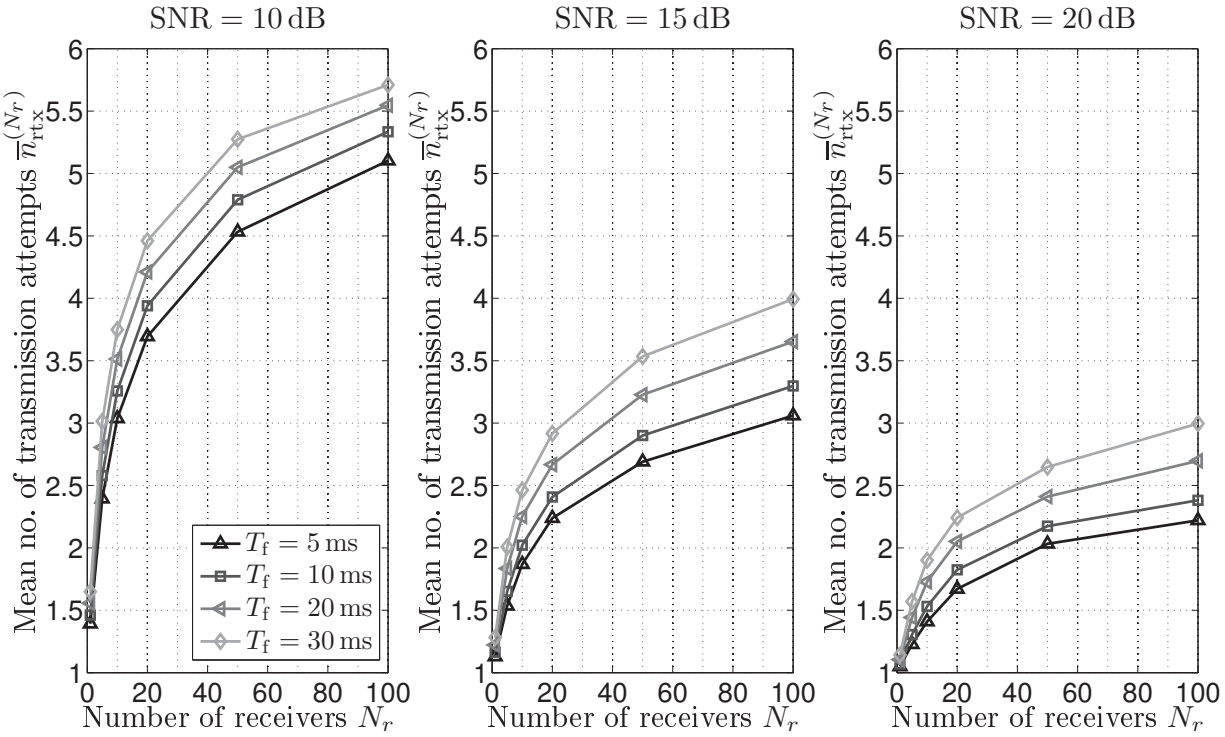


Figure 5.2: Multicast scenario of PCM speech streaming on a WLAN channel (with ROHC header compression): Average number of transmission attempts in dependence on the number of receivers; results for different channel SNRs and frame lengths per packet, T_f.

means to assure the required quality of an application. The transmitted media frames are protected against packet losses on an end-to-end basis across the possibly extremely heterogeneous network path which may consist of several independent sub-networks. As a precondition, the additional transmission capacity for the FEC data needs to be available across the network or the transmission capacity of the media frames needs to be reduced, e.g., by increasing the frame length per packet as discussed above. Alternatively, the application may also reduce the encoding rate of the utilized codec, either by changing the codec itself, or for multi-rate codecs such as the AMR and AMR Wideband codecs by using a different encoding mode.

A variety of FEC schemes commonly applied for end-to-end error protection in packet transmission of media signals has been analyzed in Chapter 4. The error correction capabilities of each scheme have been derived in dependence of its parameterization and based on a statistical model of the transmission channel. In the derivation, the specific properties of the FEC scheme regarding packet sizes and transmission time intervals and their influence on the loss characteristics of the channel have been taken into account by adapting the channel model according to the procedure introduced in Section 3.2.

The choice of the optimal FEC scheme to apply in a given scenario depends on both the application demands and the network constraints as discussed above. Applications with a high demand for a limited end-to-end delay, e.g., *Voice over IP*, can only apply a limited range of FEC schemes, especially when the transmission delay through the network is already fairly high. Possible FEC schemes which do not require too much additional delay, are repetition and XOR codes with no or a very limited interleaving of the frames. The application of FEC for *Voice over IP* applications and its optimal parameterization are discussed for the use of PCM encoded speech over WLAN in Section 5.4 and for AMR encoded speech over UMTS packet channels in Section 5.5.

Streaming applications, on the other hand, are not delay sensitive and may apply more flexible FEC schemes like block codes, possibly with large block lengths and interleaving if necessary. The application of streaming MP3 music files over a WLAN network and the optimal FEC parameterization are discussed in Section 5.3.

5.2.4 Separate vs. Piggybacked Transmission of FEC Data

Besides the actual choice of the FEC scheme, another decision has to be made in the process of setting up a packet transmission session, and that is how to packetize the original media frames and the redundant FEC frames. In general, there are two alternatives: either the redundant data is sent as separate FEC packets or it is piggybacked to the original media packets. This decision influences the resulting data rate and delay of the transmission, as well as the packet size and transmission time interval between successive packets. Depending on the channel, the latter two may have an impact on the experienced packet loss rate and distribution.

Obviously, piggybacking the FEC information to the original data packets is the more data rate efficient approach since it only increases the payload data rate. If the

FEC data is sent as separate packet stream, the packet headers of these packets will add to the total required data rate. However, the size of each packet is smaller when using separate transmission which might lead to fewer losses on channels where the loss probability depends on the packet size.

Because of the differences in data rate and delay, the choice of the transmission schemes may be determined by the delay and data rate constraints of the considered application and network. If the constraints allow the use of either scheme, the decision will depend on the loss characteristics of the transmission channel and the specific FEC scheme under consideration, as discussed for each application scenario later in the chapter.

5.2.5 Forward Error Correction to Reduce Jitter Buffer Length

The packet transmission of media signals over networks with varying packet transmission delays (jitter) requires the use of a receiver buffer (jitter buffer) to compensate for this variation. Depending on the delay constraints of the application, the length of the buffer is chosen to achieve a compromise between the resulting end-to-end signal delay and the residual packet loss rate due to late arrivals.

If FEC is applied, the redundant FEC frames can be used at the receiver to recover lost frames as well as frames from extensively delayed packets which have not been received yet. In dependence on the experienced delay distribution, the application of FEC may therefore facilitate a reduction of the jitter buffer length at the receiver and thereby achieve a reduction of the overall end-to-end delay. However, the benefits of applying FEC can only be utilized if the variation of the transmission delay is at least nearly time independent, i.e., if not too many successive packets are all experiencing a large delay. Such long delay bursts may occur if several successive packets are buffered in the network due to congestion and then released in quick succession once the congestion is resolved. These frames could only be recovered when using a large block length of the FEC code or interleaving which again introduces delay. Alternative approaches, such as an adaptive control of the jitter buffer length at the receiver, may therefore be more suitable in these cases.

The determination of the residual frame loss rate and the average length when applying FEC on channels with variable transmission delay have been derived in Section 4.3. The joint optimization of the receiver's jitter buffer length and the parameterization of the utilized FEC scheme is discussed in detail in Section 5.6 for the exemplary scenario of a VoIP transmission over the public Internet, which experiences a high variance in the packet transmission delay.

5.2.6 Adaptation of System Parameterization for Changing Channel Characteristics

The channel model introduced in Chapter 3 is able to describe the characteristics of diverse transmission channels, including variations with phases of higher and lower loss rates, reflected in the two states of the Gilbert-Elliott model. However, the model does not describe long-term variations of the rate and distribution of losses in a network. Such long time variations, e.g., differences depending on the time of day (morning vs. evening, etc.), can be considered by training several models and choosing the right one for the current situation.

If no suitable model is available or if the network characteristics are expected to undergo considerable variations during an active session of transmission, a live adaptation of the channel model is possible under some conditions. Any adaptation requires a communication from the receiver back to the sender, because the experienced loss distribution is known at the receiver while the parameterization of the transmission scheme has to be done at the sender. For such an adaptation, different scenarios are possible which differ in the distribution of the computational effort between sender and receiver and in the amount of feedback information that needs to be communicated.

Based on the observed loss statistics, the receiver may adapt the channel model itself and communicate the new model parameters to the sender, which requires only a very low data rate. The sender itself will then adapt the transmission parameters to the new channel model. This procedure requires a certain amount of computational complexity at the receiver to (re-)train the model parameters. If this complexity is not available, the receiver can alternatively communicate the the observed loss statistics to the sender, e.g., collected over a certain time and then transmitted in a single feedback packet.

The main problem of the channel model adaptation besides the computational complexity of the parameter training is the fact that the measurement of the channel characteristic is always based on the current transmission time interval and packet size of the transmission, i.e., on the current parameterization of codec, frame length, and FEC scheme. However, the optimization requires a high resolution of the model, especially for wireless channels. Therefore, an adaptation of the model might be unfeasible unless additional capacity can be utilized for high resolution measurements.

5.3 Multicast Music Streaming on Wireless LAN

The transmission of multimedia signals underlies specific media and application dependent demands as discussed in Section 5.1.3. The first scenario that shall be considered is a transmission of music signals as multicast stream over a Wireless LAN network to several receivers. The audio frames may, e.g., be encoded by the popular MPEG-1 Audio Layer 3 digital audio encoding standard, which is more commonly referred to as MP3 [ISO/IEC 11172-3:1993 1993]. This scenario differs

from a multicast streaming application that transmits a stream from the Internet to a receiver with a WLAN access. In such a scenario, the WLAN link could use retransmissions to limit or even prevent losses on this part of the transmission path. In the considered multicast streaming scenario, however, no retransmissions of packets shall be considered as these would occur too often (cf. Section 5.2.2), especially for a high number of receivers. Instead, a forward error correction (FEC) scheme shall be applied for protecting the transmitted MP3 frames. The quality criterion is the perceived audio signal quality, because the end-to-end delay is of minor importance for a streaming scenario. Since this quality is directly linked to the experienced frame loss rate, the minimization of the frame loss rate P_{fl} will be the objective of the optimization. For a high quality impression of the listeners, the rate of frame losses shall be close to zero, i.e., only occasional frame losses of preferably short length are tolerated which can be effectively concealed at the receiver.

For the considered scenario, the Gilbert-Elliott channel model of the 6 Mbit/s IEEE 802.11a channel is considered, which has been derived in Appendix E.2.2 for various signal-to-noise ratios (SNR). Based on this channel model, the frame loss rates after error correction, P_{fl}, and the mean burst length in frames, $\overline{b}$, are determined for a subset of FEC schemes and parameterizations as derived in Chapter 4, i.e., according to the following equations:

FEC scheme	P_{fl}	$P_{\mathrm{b,s}}$	$\overline{b}$
no FEC	(4.10)	(4.11)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
REP-PB, $p = 1$, $d_p = 1$	(4.19)	(4.21)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
REP Scheme 1), $p = 1, 2, 3$	(4.15)	(4.16)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR Scheme 2)	(4.25)	(4.26)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
RS (n, k)	(4.35)	(4.44)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$

The results are shown together with the required data rate R_{p} (including the IP/UDP/RTP resp. ROHC packet headers) and the resulting delay (without channel propagation delay and jitter buffer component) in Figures 5.3 and 5.4 in dependence on the channel SNR. Data rate and delay are both independent of the channel quality since no retransmissions are considered. They are calculated as given in Section 4.1.2, Table 4.1. The available data rate on the channel may be limited if the channel needs to be shared with other transmission streams from various users. Such a limit has to be taken into account in the choice of the FEC scheme to apply. Because of a strong dependence between the loss rate and the packet size on WLAN channels, a transmission of the FEC data in separate packets is considered. The negative effect of the larger packet size in the piggybacked transmission can be observed in Figure 5.3 by comparing the curves for the repetition code with one repetition per frame (*REP-PB,* $p = 1$, $d_p = 1$ for piggybacked and *REP Scheme 1),* $p = 1$ for separate transmission).

The curves show that a single repetition of each frame is not sufficient to achieve a very low residual frame loss rate, unless the channel conditions become very

good, i.e., for SNR $> 30\,\text{dB}$. However, for this range, a block code with code rate $r_\text{c} = 3/4$ and an appropriate block length is sufficient and more data rate efficient ($RS\,(n,k) = (12,9)$). In the range of 25-30 dB, a block code with rate $r_\text{c} = 1/2$ and a minimum block length of $n = 4$ is able to reduce the frame loss rate to almost 0%. The same performance is achieved in this range for the XOR Scheme 2), which also has a code rate of $r_\text{c} = 1/2$. An SNR of 20-25 dB requires a higher block length of the RS code of at least $n = 8$ with still the same code rate $r_\text{c} = 1/2$. For lower SNR values, the packet loss rate increases significantly and therefore requires an adaptation of the code rate to $r_\text{c} = 1/4$. For 15-20 dB, a block length of $n = 8$ is sufficient, while a higher block length of $n = 12$ is required for SNR values of 10-15 dB.

Considering the code rates of the FEC schemes required for different channel SNRs, the potential benefit of applying retransmission instead of FEC can be determined with the help of Figure 5.1 in Section 5.2.2. For SNR = 10 dB, the application of FEC becomes more data rate efficient if the number of receivers N_r in the multicast group exceeds 7, because the average number of transmission attempts $\overline{n}_\text{rtx}$ exceeds $1/r_\text{c} = 4$. Accordingly, the number of receivers needs to be larger than 6 for SNR = 20 dB and larger than 14 for SNR = 25 dB such that $\overline{n}_\text{rtx} > 1/r_\text{c} = 2$. For SNR = 30 dB, the use of a FEC code with $r_\text{c} = 3/4$ is more efficient than retransmission for $N_r > 6$.

5.4 Voice over IP on Wireless LAN (VoWLAN)

In contrast to the streaming of music signals as discussed above, the scenario of a voice conversation has much tighter delay constraints. This influences the optimal parameterization of the packet transmission when considering error prone channels like WLAN. The current section discusses the system parameterization regarding frame length per packet, packet retransmission, and forward error correction for WLAN channels and heterogeneous wide area networks (WAN) with wireless access.

Voice over IP applications within *local area networks* (LAN) of high capacity usually do not apply speech codecs with high compression rates, but rather use PCM speech [ITU-T Rec. G.711 1988] with an encoding rate of $R_\text{c} = 64\text{kbit/s}$ in order to provide the same quality as the so-called *plain old telephony service* (POTS) in *public switched telephone networks* (PSTN). The use of PCM also avoids the otherwise required transcoding at gateways to PSTN. Furthermore, PCM speech allows an arbitrary choice of the frame length to transmit in each packet and therefore provides an additional degree of freedom compared to other coding standards with fixed frame lengths of, e.g., 20 ms. Some users in a LAN environment, e.g., company networks, are usually connected via WLAN access, e.g., with their laptops or smartphones. Furthermore, an increasing number of WLAN access points in public places, airports, trains, hotels, cafés, and restaurants, etc., provides roaming users the ability to use VoIP services with their mobile devices if tolerated by the network

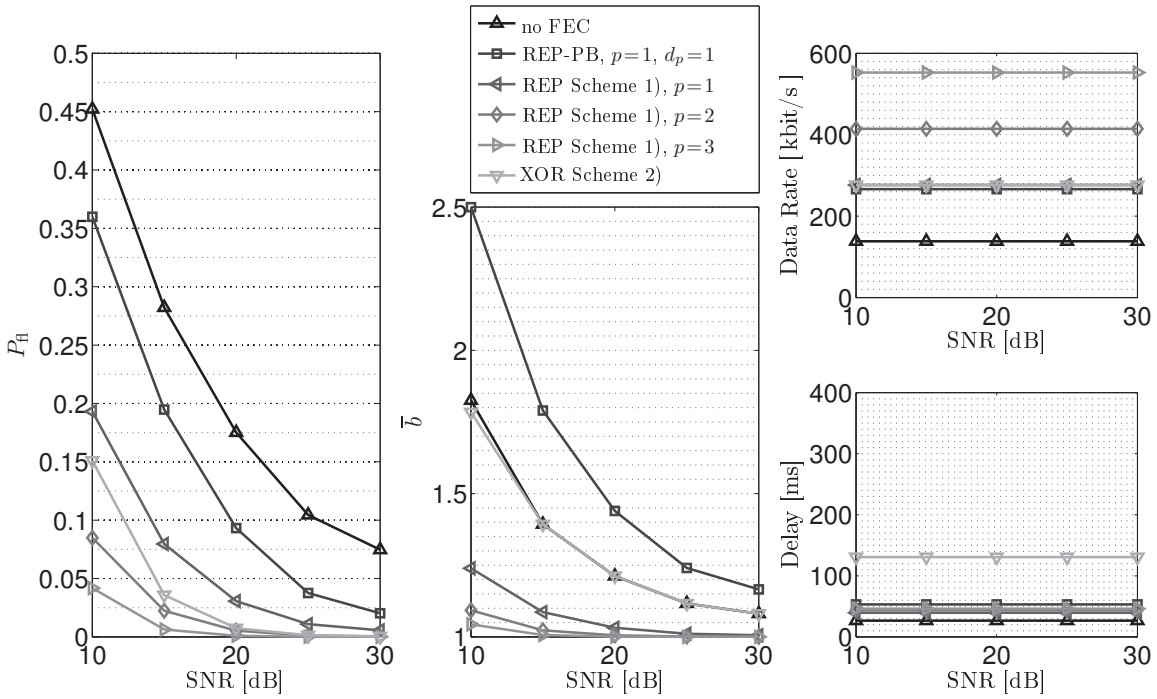


Figure 5.3: MP3 Multicast Streaming on WLAN channel with different SNR; use of ROHC header compression: Frame loss rate P_{fl}, mean burst length $\overline{b}$, data rate, and delay for different FEC schemes (repetition and XOR codes).

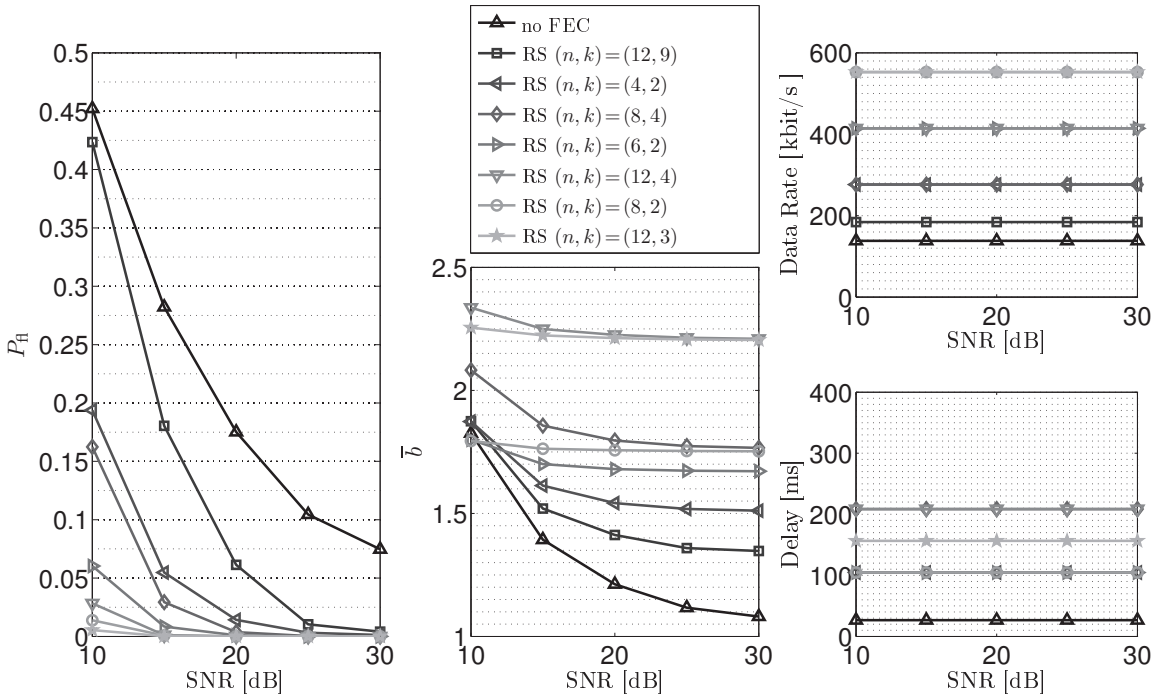


Figure 5.4: MP3 Multicast Streaming on WLAN channel with different SNR; use of ROHC header compression: Frame loss rate P_{fl}, mean burst length $\overline{b}$, data rate, and delay for different FEC schemes (block codes, e.g., Reed-Solomon (RS)).

operator. The application of *Voice over WLAN* (*VoWLAN*) is therefore an increasingly important scenario. While properly designed local area networks usually do not experience considerable packet delays and losses, Wireless LAN channels normally suffer from a considerable amount of bit errors and, as a result, from packet losses. Furthermore, a competition for channel access may lead to variable packet delays in case of high network load. Finally, if the voice call is further routed over a connected wide area network, the packets experience a considerable additional delay with a possibly high variation. These network characteristics have to be taken into account when parameterizing the VoIP application.

The following studies are based on the channel model of the exemplary WLAN channel with 6 Mbit/s derived in Appendix E.2.2. The model has a high resolution with $T'_{\mathrm{TTI}} = \tau'_{\mathrm{p}} = 0.08\,\mathrm{ms}$, so that it can be adapted to various packet transmission time intervals and packet sizes, as needed for a comparison of different transmission schemes and frame lengths per packet. Three different scenarios are considered in the following sections. Section 5.4.1 focuses on a VoIP call over a WLAN channel utilizing retransmissions. The average necessary number of transmission attempts and the optimal frame length to transmit in each packet are determined. In Section 5.4.2, a transmission of redundancy with forward error correction schemes is considered instead of retransmissions. This scenario applies to streaming applications where retransmission would be inefficient for a high number of receivers. Finally, in Section 5.4.3 the scenario of a VoIP call over a heterogeneous network is discussed where both participants are connected via WLAN to an unspecified intermediate network which is assumed to introduce additional packet losses. The focus of the investigation for this scenario is whether to apply an end-to-end error protection using FEC schemes, or to rely on retransmissions on the wireless links only.

5.4.1 Optimal Frame Length with Layer 2 Retransmissions

For the the first scenario of a VoIP call on Wireless LAN using PCM speech, no additional FEC mechanisms are considered. Instead, layer 2 retransmissions are initiated with a maximum number of transmission attempts N_{rtx} for each packet. The required value of N_{rtx} and the optimal frame length to be transmitted in each packet, T_{f}, shall be determined for different channel SNRs. It is assumed that there are not many further applications competing for channel access or that the WLAN standard variant for QoS Enhancements [IEEE Std 802.11e 2005] is employed, which guarantees a high priority for the retransmissions. Otherwise, a considerable increase of the delay might result if each retransmission has to compete for channel access.

The transmission of PCM speech with an arbitrary frame length T_{f} per packet and no FEC results in a packet transmission time interval equal to the frame length, i.e., $T_{\mathrm{TTI}} = T_{\mathrm{f}}$. The size of each packet depends on the frame length and the

encoding rate of the utilized speech codec. For the considered scenario the packet size therefore computes to

$$L_\mathrm{p} = L_\mathrm{h} + L_\mathrm{plh} + L_\mathrm{f} = L_\mathrm{h} + L_\mathrm{plh} + R_\mathrm{c}\, T_\mathrm{f}, \tag{5.24}$$

with the encoding rate $R_\mathrm{c} = 64\mathrm{kbit/s}$ and a total header size of $L_\mathrm{h} = 68\,\mathrm{byte}$, consisting of the WLAN MAC header and CRC (28 byte) and the IP/UDP/RTP headers (40 byte). For the application of ROHC header compression, the IP/UDP/RTP headers are reduced to the 3 byte ROHC header and the total header size becomes $L_\mathrm{h} = 31\,\mathrm{byte}$. Hence, for a frame length of, e.g., $T_\mathrm{f} = 20\,\mathrm{ms}$, the total packet size results to $L_\mathrm{p} = 228\,\mathrm{byte}$ without and $L_\mathrm{p} = 191\,\mathrm{byte}$ with header compression. The packet transmission time, i.e., the ratio between the packet size and the channel transmission rate, here considered as $R_\mathrm{ch} = 6\,\mathrm{Mbit/s}$, results to

$$\tau_\mathrm{p} = \frac{L_\mathrm{p}}{R_\mathrm{ch}} = \frac{L_\mathrm{h} + R_\mathrm{c}\, T_\mathrm{f}}{R_\mathrm{ch}}. \tag{5.25}$$

For the use of retransmissions upon request, the packet transmission time interval between the single transmission attempts has been derived in Section 4.2, (4.53), as

$$T_\mathrm{TTI}^\mathrm{rtx} = \tau_\mathrm{p} + \delta_\mathrm{ACK} + \tau_\mathrm{ACK} + \delta_p. \tag{5.26}$$

The size of an acknowledgment packet in the WLAN standard, including its own specific WLAN header and CRC, amounts to $L_\mathrm{ACK} = 14\,\mathrm{byte}$, resulting in a transmission time of $\tau_\mathrm{ACK} = L_\mathrm{ACK}/R_\mathrm{ch} = 18.7\,\mu\mathrm{s}$. The channel access delays δ_ACK and δ_p are neglected in this scenario.

For each considered frame length T_f, the channel model of the WLAN channel has been adapted to the correct packet transmission time interval $T_\mathrm{TTI}^\mathrm{rtx}$ and the packet transmission time τ_p as given above (cf. Section 4.2 for a more detailed explanation). According to the derivations in Chapter 4, the expected frame loss rate P_fl (4.67), the mean burst length $\overline{b}$ (4.70), as well as the resulting packet data rate (4.61) and delay (4.63) (without channel propagation delay and jitter buffer component) have been calculated for each frame length and maximum number of transmission attempts. The results are plotted against the frame length T_f in Figure 5.5 and 5.6 for 10 dB and 20 dB channel SNR, respectively. No header compression has been assumed and the data rate is given for the case of a single receiver.

The curves show a clear dependency between packet size and residual packet loss rate, with a shorter packet size leading to a considerably lower loss probability P_fl. This advantage, however, comes at the expense of an increase in the packet data rate. As discussed in Section 5.2.1, the required data rate for the packet headers is inversely proportional to the frame length, leading to a steep rise for shorter frame lengths.

The channel with 10 dB SNR shows an extremely high loss rate, which, however, can be considerably reduced by applying retransmission mechanisms. A limitation

to a maximum of $N_{\mathrm{rtx}} = 5$ transmission attempts is sufficient to reduce the frame loss rate to around 1%, with only a slight increase for larger frame lengths. In general, the dependency between loss probability and frame length is more and more reduced when increasing the maximum number of transmission attempts. At an SNR of 20 dB, a maximum of $N_{\mathrm{rtx}} = 3$ or $N_{\mathrm{rtx}} = 4$ transmission attempts is already sufficient to achieve an extremely low residual frame loss rate.

The end-to-end delay is mainly dependent on the chosen frame length, because the retransmission attempts are made in quick succession on the wireless link, which itself is considered to have only a low transmission delay. An effect of the delay on the quality is therefore not expected for the given range of frame lengths.

The application of header compression can reduce the data rate significantly for short frame lengths. However, there is still a steep increase, because the WLAN MAC header of 28 byte is not compressed in the standard ROHC header compression scheme (cf. Section 2.4.2). The loss behavior is very similar to the case of no header compression. Only a slight overall decrease in the loss rate can be expected for each curve which can be attributed to the slightly smaller packet sizes.

The results show that a layer 2 retransmission scheme is very effective in guaranteeing a low residual loss rate over a wide range of channel qualities, while requiring only slightly increased data rate. The end-to-end delay is not increased considerably if competition for channel access can be avoided or retransmissions are prioritized. The otherwise strong dependency of the loss rate on the utilized frame length per packet is significantly reduced when choosing an appropriate maximum number of transmission attempts. Therefore, a moderate size of the frame length around 20 ms is a reasonable choice which limits the overall packet data rate.

In the considered scenario, the channel needs to have an SNR of around 20 dB and the frame length needs to be below 10 ms to allow a transmission with a still acceptable packet loss rate if no retransmissions can be used. On channels where the packet loss rate increases with increasing packet length, the frame length should be chosen as small as the data rate allows in order to minimize the residual loss rate.

5.4.2 Joint Optimization of Frame Length and Forward Error Correction

In contrast to the conversational application discussed in the previous section, the current section assumes a streaming application where PCM speech signals are transmitted over WLAN to a multicast receiver group. The same general settings shall be considered as in the previous section, i.e., the use of PCM speech with an arbitrary frame length T_{f} per packet and the 6 Mbit/s WLAN channel modeled in Appendix E.2.2.

The streaming of media signals to a multicast receiver group does usually not allow the use of retransmissions, because a large number of receivers significantly increases the required number of retransmissions as discussed in Section 5.2.2. Instead, error robustness can be achieved through the employment of appropriately

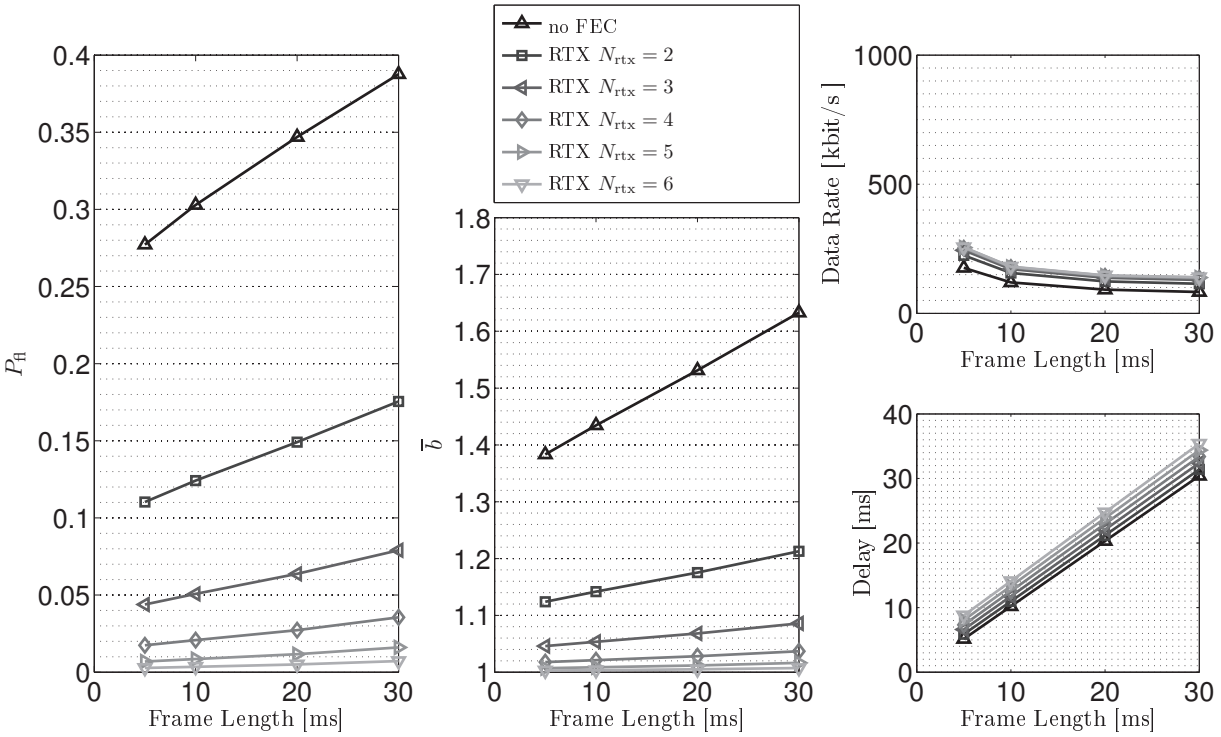


Figure 5.5: VoIP using PCM on WLAN channel with SNR=10 dB; no header compression: Frame loss rate P_{fl}, mean burst length $\overline{b}$, data rate and delay for different frame lengths and max. number of layer 2 transmission attempts N_{rtx}

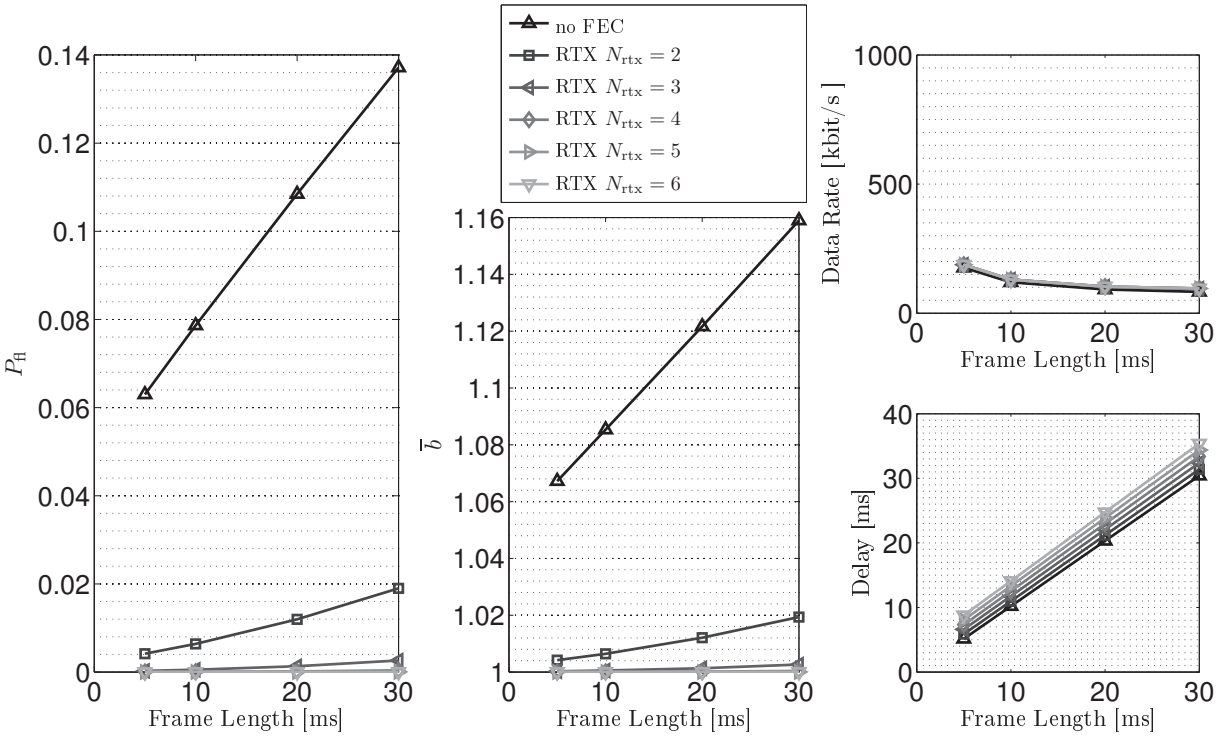


Figure 5.6: VoIP using PCM on WLAN channel with SNR=20 dB; no header compression: Frame loss rate P_{fl}, mean burst length $\overline{b}$, data rate and delay for different frame lengths and max. number of layer 2 transmission attempts N_{rtx}

parameterized packet level FEC schemes which are able to recover a certain amount of frame losses. The flexible channel model introduced in Chapter 3 facilitates a realistic comparison of different frame lengths as well as different FEC codes and their parameterizations through an appropriate adaptation of transmission time interval and packet size.

In Figure 5.7, 5.8, and 5.9, the effects of different FEC schemes with some exemplary parameterizations are shown for the WLAN channel assuming a relatively low SNR of 10 dB, i.e., a high error rate. The frame loss rate P_{fl}, the mean burst length $\overline{b}$, the required packet data rate R_{p}, and the resulting delay (without channel propagation delay and jitter buffer component) are plotted against the frame length per packet, T_{f}. The values are calculated according to the following equations derived in Chapter 4:

FEC scheme	P_{fl}	$P_{\mathrm{b,s}}$	$\overline{b}$
no FEC	(4.10)	(4.11)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
REP-PB, $p = 1, 2, 3$, $d_p = 1$	(4.19)	(4.21)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
REP Scheme 1), $p = 1, 2, 3$	(4.15)	(4.16)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
REP Scheme 2), $p = 1, 2, 3$	(4.15)	(4.17)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR-PB Scheme 1) = RS-PB $(3, 2)$	(4.46)	(4.52)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR-PB Scheme 2a)	(4.28)	(4.29)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR-PB Scheme 2b)	(4.30)	(4.31)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR-PB Scheme 2c)	(4.32)	(4.33)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR Scheme 1) = RS $(3, 2)$	(4.35)	(4.44)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
XOR Scheme 2)	(4.25)	(4.26)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
RS-PB (n, k)	(4.46)	(4.52)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$
RS (n, k)	(4.35)	(4.44)	$\overline{b} = P_{\mathrm{fl}}/P_{\mathrm{b,s}}$

First, consider the variants of the repetition scheme depicted in Figure 5.7. For the same number of repetitions, the separate transmission of the repeated frames (*REP Schemes 1) and 2), w/o and with implicit interleaving of the frames, cf. Section 4.1.5.1*) leads to a lower frame loss rate than the piggybacked transmission (*REP-PB*). This is a clear indication for the dependence of packet loss rate on packet size on the considered channel. For a decreasing frame length, the loss rate and mean burst length for the two transmission variants slowly converge. In general, the mean burst length is lower for the separate transmission than for the piggybacked transmission of the FEC frames, when compared at the same frame length. For frame lengths of 5 ms and larger, the two schemes with separate transmission show the same performance in terms of loss rate and average burst length. Towards lower frame lengths, however, a stronger increase of the frame loss rate can be observed for *REP Scheme 1)*. The reason for this behavior lies in the loss characteristic of the channel, producing rather short burst phases, i.e., phases with increased errors. Hence, only packets that are transmitted with very short transmission time intervals, as given for frame lengths below 5 ms, will experience burst losses of successive packets. *REP Scheme 2)* then leads to a better performance because of

the interleaved transmission. However, except for extremely short frame lengths, a smaller frame length per packet generally leads to a lower frame loss rate at the receiver. The high packet loss rate on the WLAN channel with an SNR of only 10 dB requires a code rate of $r_c = 1/3$ of the repetition scheme, i.e., $p = 2$ repetitions for each frame, to achieve a residual loss rate below 5% which can be effectively concealed by the receiver's packet loss concealment algorithm. If transmitted in separate FEC packets (*REP Scheme 1) and 2)*), the frame length should be chosen around 20 ms to limit the required data rate. The piggybacked transmission of the repetitions (*REP-PB*) requires a shorter frame length of around 5 ms to achieve a comparable residual loss rate. Because of the shorter frame length, the required data rate becomes also comparable to the separate transmission scheme. The cost of applying such a repetition scheme on the given channel is a data rate of almost three times the rate of the packet stream without FEC.

The curves in Figure 5.8 for an alternative FEC scheme based on XOR combinations of frames show a similar behavior regarding the correlation between frame length and frame loss rate, as well as the different effects of utilizing a separate versus a piggybacked transmission of the FEC frames. For the considered WLAN channel with an SNR of 10 dB, the XOR schemes with a code rate of $r_c = 1/2$, *XOR-PB Scheme 2a,b,c)* and *XOR Scheme 2)*, are able to reduce the residual frame loss rate to below 5% for an appropriately chosen frame length per packet. As for the repetition code, the XOR code with separate transmission of the FEC frames (*XOR Scheme 2)*) is the most efficient variant, allowing a frame length of 10-15 ms per packet. The piggybacked scheme has to use a shorter frame length of about 5 ms to achieve a comparable residual frame loss rate. The variant *XOR-PB Scheme 2c)*, which involves a certain degree of interleaving, shows the best performance among the piggybacked schemes, as the resulting delay does not affect the quality of a streaming application. The schemes with a lower code rate of $r_c = 2/3$, *XOR-PB Scheme 1)* and *XOR Scheme 1)*, cannot reduce the high original frame loss rate to an acceptable amount. Similar to the repetition code, the frame length must not be chosen too small to avoid being affected by the channel's burst errors.

Finally, the application of systematic block codes, e.g., Reed-Solomon codes, is analyzed in Figure 5.9. This FEC scheme also requires a code rate around $r_c = 1/2$ and a relatively low frame length in order to achieve an acceptable residual frame loss rate on the given WLAN channel. At the expense of a higher end-to-end delay, a larger block length n leads to better results at this code rate as it is able to correct a higher number of consecutive losses. The increase in the delay is not of significance for the considered streaming application scenario.

The choice of the optimal scheme to apply on the considered WLAN channel with a high error rate will depend on the constraint of the data rate and the loss tolerance of the codec. For example, a target residual frame loss rate of around 3% can be achieved with a frame length of $T_f = 5$ ms, RS code with $(n, k) = (8, 4)$ and piggybacked transmission, resulting in a data rate of 181 kbit/s and a delay of 40 ms. With separate transmission of this FEC code, the same loss rate can be achieved

at the same total data rate by choosing a frame length of $T_\mathrm{f} = 10\,\mathrm{ms}$. However, the delay is increased to 80 ms.

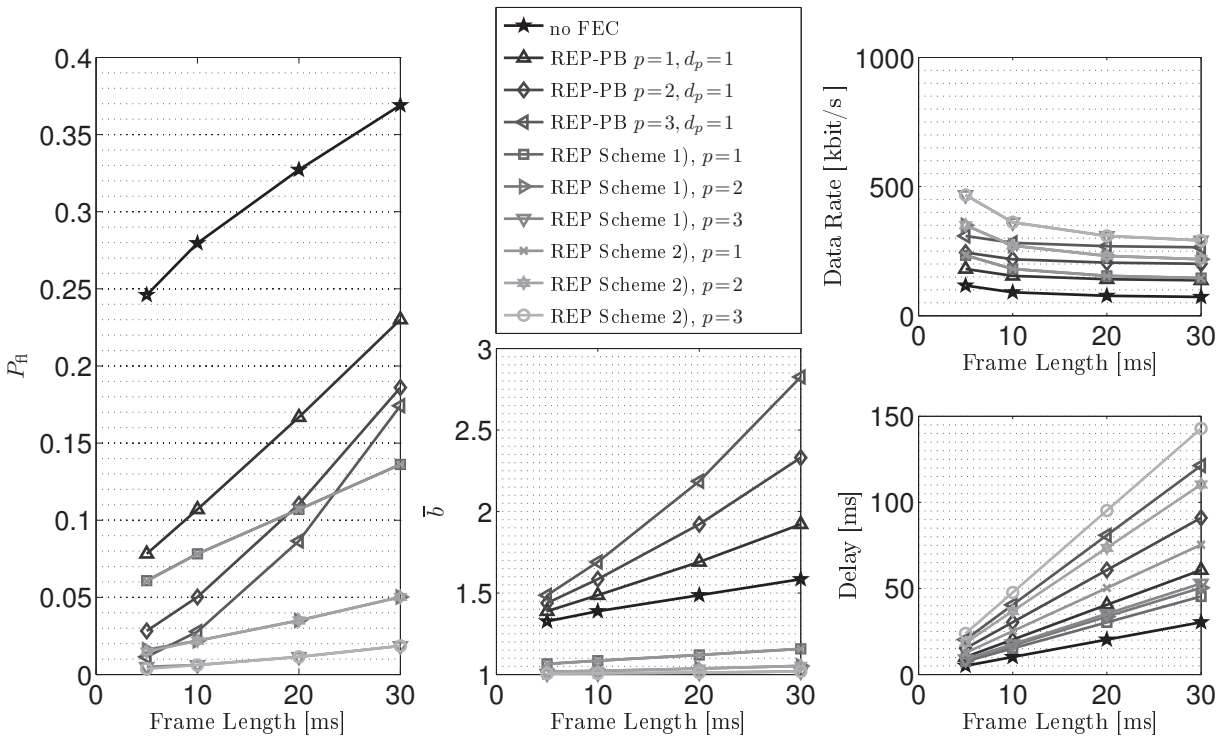


Figure 5.7: VoIP using PCM on WLAN channel with SNR=10 dB; with ROHC header compression: Frame loss rate P_fl, mean burst length $\overline{b}$, data rate and delay for different frame lengths and FEC schemes (repetition codes).

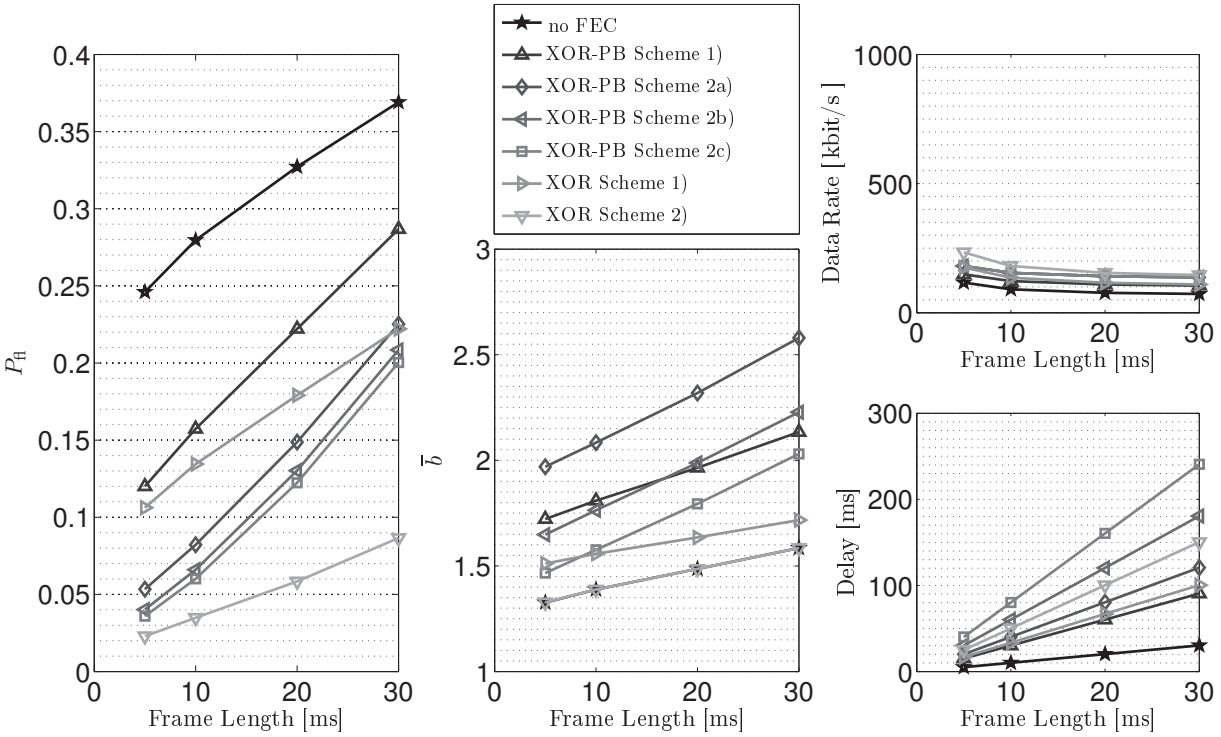


Figure 5.8: VoIP using PCM on WLAN channel with SNR=10 dB; with ROHC header compression: Frame loss rate P_{fl}, mean burst length $\overline{b}$, data rate and delay for different frame lengths and FEC schemes (XOR codes).

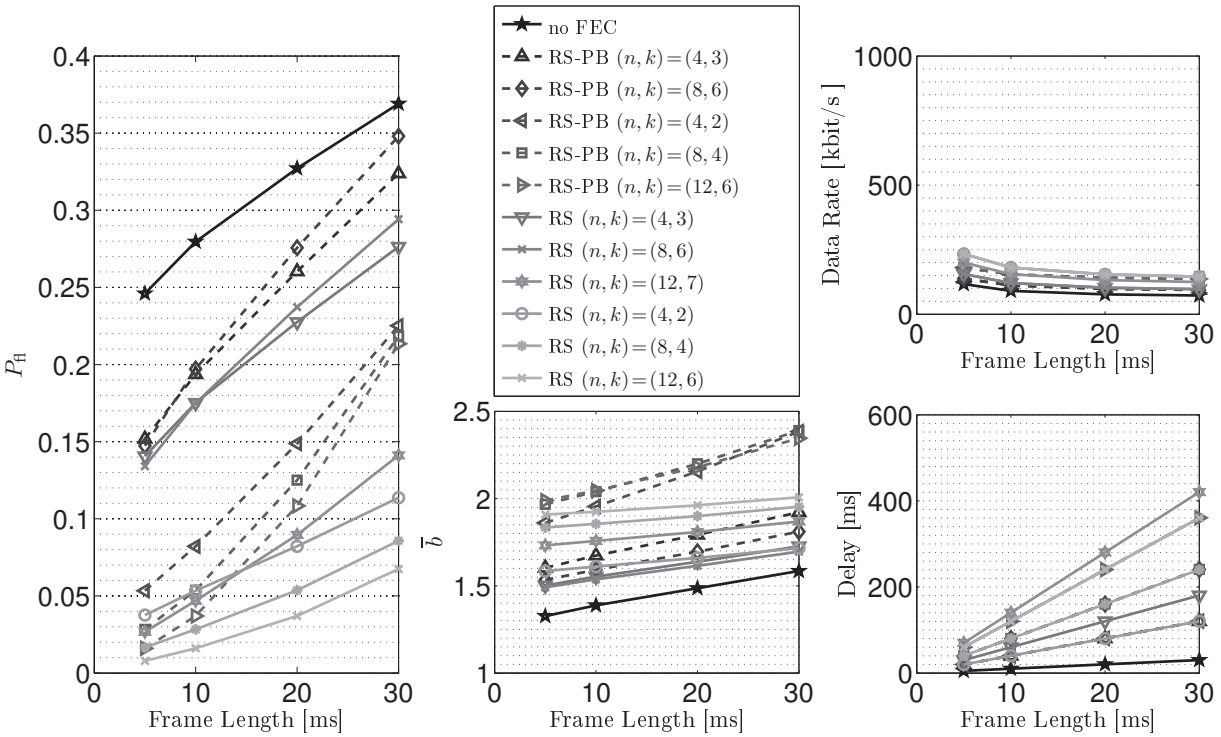


Figure 5.9: VoIP using PCM on WLAN channel with SNR=10 dB; with ROHC header compression: Frame loss rate P_{fl}, mean burst length $\overline{b}$, data rate and delay for different frame lengths and FEC schemes (RS block codes).

5.4.3 Optimal Parameterization for Heterogeneous Networks with WLAN Access

In this section, a VoIP connection with PCM speech over a heterogeneous network shall be considered. Two WLAN access networks are assumed that are connected through a not further specified core network which itself may consist of a group of connected networks. The core network is assumed to generate independent packet losses of a certain rate which are independent of the packet size and transmission time interval as it can be assumed for high capacity networks. The transmission delay through the core network is assumed as 100 ms without a considerable amount of variation (jitter)[1]. The use of header compression (ROHC) is assumed on each wireless link, but not in the core network.

Objective is a reliable end-to-end transmission which is robust enough to cope with transmission errors, i.e., packet losses, on the different network parts, i.e., the WLAN links and the core network. To achieve this robustness, several transmission strategies are possible which are compared and discussed in the following:

A) The use of retransmissions on the WLAN links without further error protection in the core network;
B) the application of end-to-end FEC schemes and additional retransmissions on the WLAN links; and
C) the application of end-to-end FEC schemes without using retransmissions on the WLAN links.

Unicast transmissions on Wireless LAN normally utilize packet retransmissions on layer 2 in case of transmission errors. The maximum number of transmission attempts for real-time services is limited by the time constraint that the last transmission attempt needs to be completed before the packet with the following media frame needs to be transmitted (cf. (4.56) in Section 4.2). This constraint guarantees the real-time functionality of the transmission even if every packet requires the maximum number of transmission attempts. The additional delay caused by the retransmission is therefore approximately one frame length. The effectiveness of the retransmission scheme in Wireless LAN has been shown in Section 5.4.1. The first approach in the considered heterogeneous network scenario is therefore to rely on the retransmission mechanisms on both WLAN links and to add no additional redundancy on packet level (approach *A)*). However, this approach does not provide error protection against losses in the core network which then directly result in frame losses at the receiver. Effective frame loss concealment algorithms can only be applied at the receiver as long as rate and length of the losses stay low.

If the loss rate in the core network increases, additional error protection mechanisms have to be employed. FEC schemes applied on packet level provide end-to-end protection against packet losses anywhere on the transmission path. In the considered network, these schemes might be optimized to compensate for the core network losses only (approach *B)*), relying on the retransmissions on the WLAN

[1] The packet transmission on networks with high variation in the transmission delay will be discussed in Section 5.6.

links to compensate for losses on the wireless channels. Then, the applied FEC scheme influences the retransmissions. The higher number of packets in case of separate transmission may reduce the maximum number of transmission attempts and the increased packet sizes for piggybacked transmission may lead to a higher number of erroneous packets and therefore retransmission requests. Alternatively, the FEC scheme can be parameterized for the expected loss characteristic of the end-to-end transmission channel including the wireless links (approach *C)*). In this case, no further retransmissions are executed on the WLAN links. The required code rate of the FEC scheme is expected to be considerably higher in this approach than for approach *B)* to be able to compensate the additional and possibly bursty losses on the wireless channels.

The results of the different approaches discussed above are shown in Figure 5.10 for an assumed packet loss rate of 6% in the core network and in Figure 5.11 for a packet loss rate of 12%. The SNR on both WLAN links is assumed to be 15 dB. Residual frame loss rate, average burst length, required data rate and resulting end-to-end delay are shown in dependence on the frame length per packet. The plotted data rate is the data rate required on the wireless channels, as a possible constraint may be rather in effect there than in the core network. The curves of frame loss rates P_{fl} have been calculated for the different approaches as follows:

Approach A)

The frame loss probability on the WLAN access channel, $P_{\mathrm{fl}}^{\mathrm{WLAN}}$, is calculated assuming fast retransmissions with a maximum number of $N_{\mathrm{rtx}} = 6$ transmission attempts according to (4.67). In this calculation, the channel model derived for the WLAN channel with an SNR of 15 dB (cf. Section E.2.2) is applied. For reasons of simplicity, the same model is assumed for both WLAN links. No forward error correction is applied and a single frame is transmitted in each packet. In the core network (CN) this leads to a fixed frame loss rate of $P_{\mathrm{fl}}^{\mathrm{CN}} = 6\%$ or $P_{\mathrm{fl}}^{\mathrm{CN}} = 12\%$, depending on the considered packet loss rate.

The packet losses on the three transmission paths (WLAN, CN, WLAN) can be assumed as statistically independent. Hence, the end-to-end frame loss probability is calculated from the three individual probabilities $P_{\mathrm{fl}}^{1)}$, $P_{\mathrm{fl}}^{2)}$, and $P_{\mathrm{fl}}^{3)}$ as

$$P_{\mathrm{fl}} = P_{\mathrm{fl}}^{1)} + P_{\mathrm{fl}}^{2)} + P_{\mathrm{fl}}^{3)} - P_{\mathrm{fl}}^{1)} \cdot P_{\mathrm{fl}}^{2)} - P_{\mathrm{fl}}^{1)} \cdot P_{\mathrm{fl}}^{3)} - P_{\mathrm{fl}}^{2)} \cdot P_{\mathrm{fl}}^{3)} + P_{\mathrm{fl}}^{1)} \cdot P_{\mathrm{fl}}^{2)} \cdot P_{\mathrm{fl}}^{3)}. \quad (5.27)$$

With $P_{\mathrm{fl}}^{1)} = P_{\mathrm{fl}}^{3)} = P_{\mathrm{fl}}^{\mathrm{WLAN}}$ and $P_{\mathrm{fl}}^{2)} = P_{\mathrm{fl}}^{\mathrm{CN}}$ this results to

$$P_{\mathrm{fl}} = 2\,P_{\mathrm{fl}}^{\mathrm{WLAN}} + P_{\mathrm{fl}}^{\mathrm{CN}} - 2\,P_{\mathrm{fl}}^{\mathrm{WLAN}}\,P_{\mathrm{fl}}^{\mathrm{CN}} - P_{\mathrm{fl}}^{\mathrm{WLAN}}\,P_{\mathrm{fl}}^{\mathrm{WLAN}} + P_{\mathrm{fl}}^{\mathrm{WLAN}}\,P_{\mathrm{fl}}^{\mathrm{CN}}\,P_{\mathrm{fl}}^{\mathrm{WLAN}}. \quad (5.28)$$

Approach B)

For approach B), an effective end-to-end channel model needs to be determined in dependence on the considered FEC scheme. This end-to-end model needs to incorporate the effects of the fast retransmissions on each WLAN access channel as derived in the following.

The choice of the applied end-to-end FEC scheme itself and the choice of piggybacked or separate transmission of the FEC frames determine the packet transmission time interval $T_{\mathrm{TTI}}^{\mathrm{FEC}}$, i.e., the interval between successive packets generated at the transmitter, and the packet transmission time $\tau_{\mathrm{p}}^{\mathrm{FEC}}$, which depends on the packet size. The end-to-end channel model will be based on these parameters. Due to the consideration of retransmissions on the WLAN channels, the end-to-end model will result as Extended Gilbert-Elliott model with transition dependent loss probabilities, as newly introduced in Section 3.2.

The resulting model of a single WLAN access channel incorporating the effects of retransmissions is derived as follows. First, the base channel model of the WLAN channel is adapted to the parameters of the considered end-to-end FEC scheme, i.e., the according transmission time interval $T_{\mathrm{TTI}}^{\mathrm{FEC}}$ and packet transmission time $\tau_{\mathrm{p}}^{\mathrm{FEC}}$. The adaptation of the model is done as explained in Section 3.2 using the following factors:

$$k_{\mathrm{t}}^{\mathrm{FEC}} = \frac{T_{\mathrm{TTI}}^{\mathrm{FEC}}}{T'_{\mathrm{TTI}}}, \tag{5.29}$$

$$k_{\mathrm{p}}^{\mathrm{FEC}} = \frac{\tau_{\mathrm{p}}^{\mathrm{FEC}}}{\tau'_{\mathrm{p}}}. \tag{5.30}$$

T'_{TTI} and τ'_{p} are the transmission time interval and packet transmission time of the WLAN base model. The state transition probabilities $P_{\mathrm{t},XY}^{(k_{\mathrm{t}}^{\mathrm{FEC}})}$ of the resulting WLAN model result as given in (3.25) and are independent from the use of retransmissions.

For the determination of the transition dependent loss probabilities of the resulting WLAN model, the use of retransmissions on the WLAN channel needs to be considered. To this end, the base channel model of the WLAN channel is adapted to the transmission time interval of the retransmissions, $T_{\mathrm{TTI}}^{\mathrm{rtx}}$, as given in (4.53), and the packet transmission time $\tau_{\mathrm{p}}^{\mathrm{FEC}}$ of the considered end-to-end FEC scheme. Hence, the adaptation uses the following factors:

$$k_{\mathrm{t}}^{\mathrm{rtx}} = \frac{T_{\mathrm{TTI}}^{\mathrm{rtx}}}{T'_{\mathrm{TTI}}} = \frac{\tau_{\mathrm{p}} + \delta_{\mathrm{ACK}} + \tau_{\mathrm{ACK}} + \delta_p}{T'_{\mathrm{TTI}}}, \tag{5.31}$$

$$k_{\mathrm{p}}^{\mathrm{FEC}} = \frac{\tau_{\mathrm{p}}^{\mathrm{FEC}}}{\tau'_{\mathrm{p}}}. \tag{5.32}$$

The probabilities of losing m of n successive packets, $P_{XY}^{(k_{\mathrm{t}}^{\mathrm{rtx}},k_{\mathrm{p}}^{\mathrm{FEC}})}(m,n)$, are then calculated according to (3.40). A packet is lost if all N_{rtx} transmission attempts fail.

The probability of this event results to $P_{XY}^{(k_\mathrm{t}^\mathrm{rtx},k_\mathrm{p}^\mathrm{FEC})}(N_\mathrm{rtx},N_\mathrm{rtx})$, with the channel state at the first transmission attempt, X, and the state at the theoretical $(N_\mathrm{rtx}+1)$-st attempt, Y. For the calculation of the transition dependent loss probabilities, the remaining interval between the last transmission attempt of the current packet and the first transmission attempt of the following packet needs to be taken into account. This interval of length $T_\mathrm{TTI}-N_\mathrm{rtx}\,T_\mathrm{TTI}^\mathrm{rtx}$ can approximately be divided into s intervals of the retransmission interval $T_\mathrm{TTI}^\mathrm{rtx}$, with

$$s=\left\lfloor\frac{T_\mathrm{TTI}-N_\mathrm{rtx}\,T_\mathrm{TTI}^\mathrm{rtx}}{T_\mathrm{TTI}^\mathrm{rtx}}+\frac{1}{2}\right\rfloor. \tag{5.33}$$

The transition dependent loss probabilities are finally calculated as

$$P_{\mathrm{e},XY}^{(k_\mathrm{t}^\mathrm{FEC},k_\mathrm{p}^\mathrm{FEC})}=\sum_{Z\in\{\mathrm{G,B}\}}P_{XZ}^{(k_\mathrm{t}^\mathrm{rtx},k_\mathrm{p}^\mathrm{FEC})}(N_\mathrm{rtx},N_\mathrm{rtx})\sum_{i=0}^{s}P_{ZY}^{(k_\mathrm{t}^\mathrm{rtx},k_\mathrm{p}^\mathrm{FEC})}(i,s), \tag{5.34}$$

with the last sum denoting the state transition probability across the s intervals of the retransmission interval $T_\mathrm{TTI}^\mathrm{rtx}$.

The derived transition probabilities $P_{\mathrm{t},XY}^{(k_\mathrm{t}^\mathrm{FEC})}$ and transition dependent loss probabilities $P_{\mathrm{e},XY}^{(k_\mathrm{t}^\mathrm{FEC},k_\mathrm{p}^\mathrm{FEC})}$ together form the effective channel model for a WLAN access channel using retransmissions. For the end-to-end channel model of both WLAN access channels and the intermediate core network, the Extended Gilbert-Elliott models of the WLAN channels — derived as explained above — are concatenated as explained in Appendix G.5. Assuming statistically independent losses in the core network, the loss probability of the core network is finally incorporated as explained in Appendix G.3.

With the resulting end-to-end channel model, the residual frame loss rates for the different considered end-to-end FEC schemes are calculated as derived in the respective subsection of Section 4.1, e.g., according to (4.15) for REP Scheme 1) and (4.19) for REP-PB.

Approach C)

For the determination of the end-to-end frame loss rate in approach C), again an effective end-to-end channel model is determined from the single models of the WLAN access channels and the core network. However, since no retransmissions are considered on the WLAN channels, the determination of the end-to-end channel model is more straightforward for approach C) than for approach B) as discussed above.

The channel models are first adapted to the respective transmission time interval and packet size for the considered end-to-end FEC scheme according to Section 3.2. The extended Gilbert-Elliott models of both WLAN access channels are then combined into a single model as explained in Appendix G.5, before the Bernoulli channel model of the core network is incorporated according to Appendix G.3. The resulting

end-to-end channel model for the respective transmission time interval and packet size is then utilized in the determination of the residual frame loss rate for the considered FEC scheme as developed in Section 4.1 (i.e., (4.15) for REP Scheme 1) and (4.19) for REP-PB).

Discussion of results

Without retransmission and other error protection schemes, the connected transmission channels lead to a high overall loss rate (cf. curve labeled 'no FEC'). The horizontal curve for the use of retransmissions without further FEC, approach A), clearly shows the effectiveness of the retransmission such that only the packet losses in the core network remain. These, however, cannot be compensated in this approach. At the expense of a considerably higher data rate, the end-to-end FEC schemes applied with retransmissions turned off on the WLAN links, approach C), can recover losses on all network parts leading to a tolerable residual frame loss rate. However, a code with low code rate is required, e.g., a repetition code with $r_c = 1/3$, to achieve a loss rate below 5% for frame lengths lower than 15 ms. The optimal result is achieved by approach B), i.e., the utilization of layer 2 retransmissions on the WLAN channels together with the application of an end-to-end FEC scheme which is parameterized to compensate for the losses in the core network. For the considered scenario, a single repetition of each frame, piggybacked to the following packet, is sufficient for packet loss rates of 6–12 % in the network. If the losses in the core network are of bursty nature, a FEC scheme with a certain degree of interleaving should be considered. For lower loss rates in the core network, the application of end-to-end FEC schemes may not be necessary if they can be effectively concealed at the receiver.

An argument against the application of end-to-end FEC schemes is the increase of data rate in the core network which may also increase the congestion and thereby contribute itself to a further increase of the loss rate. If such a behavior has to be expected, a different speech codec could be used which is able to adapt its encoding rate to provide capacity for the FEC frames. A suitable codec is, e.g., the AMR codec which will be discussed on UMTS channels and a public Internet connection in the following sections.

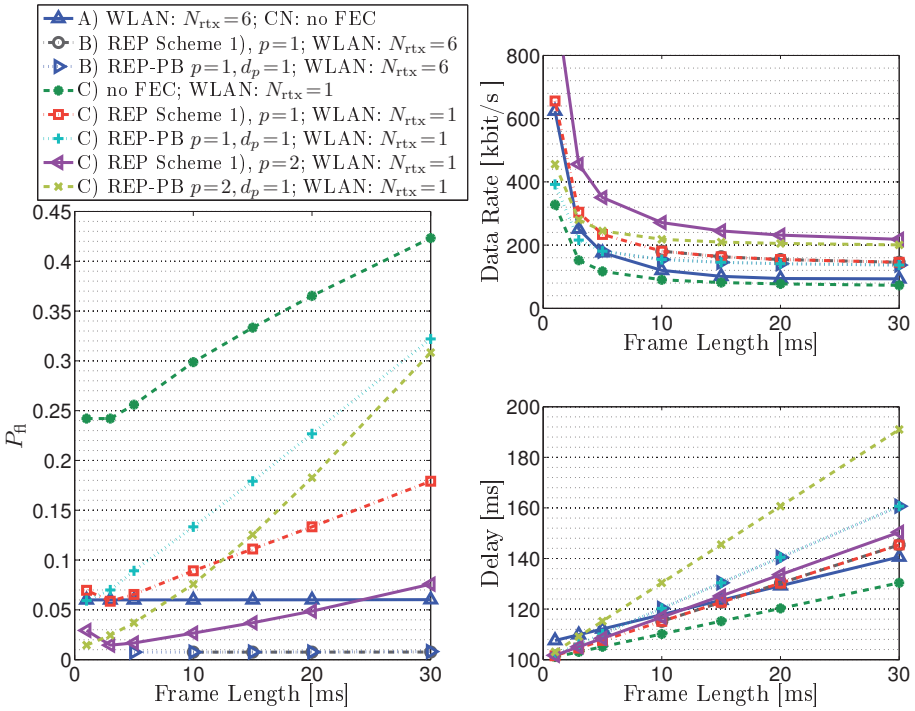


Figure 5.10: VoIP using PCM in heterogeneous network: WLAN uplink channel with SNR=15 dB, IP network with 6% packet loss, WLAN downlink channel with SNR=15 dB; ROHC header compression on WLAN links: Frame loss rate, data rate, and delay for different frame lengths and transmission strategies A), B), C) as explained in the text.

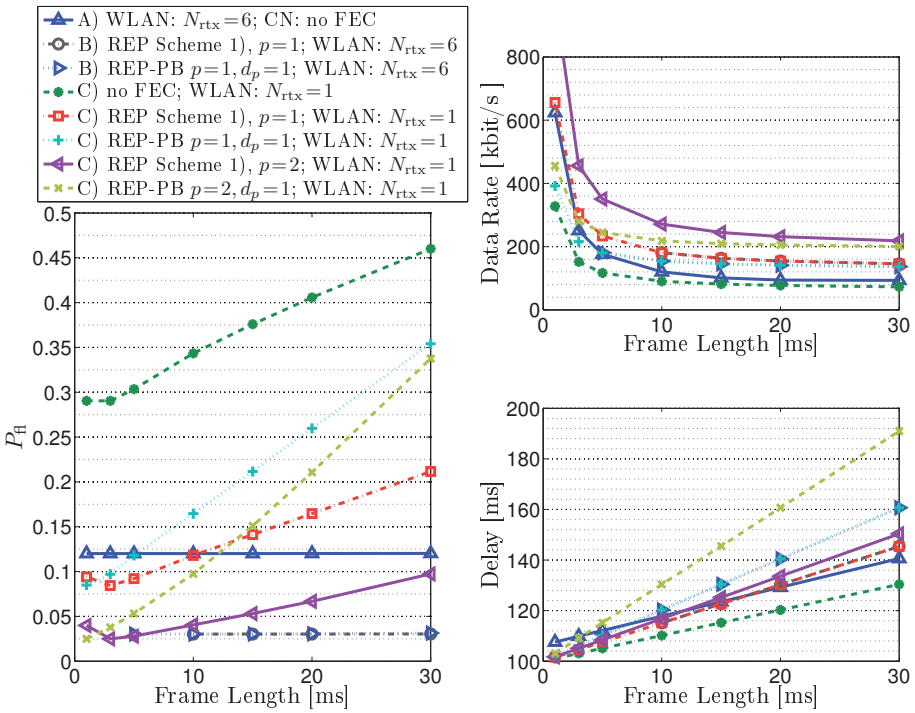


Figure 5.11: VoIP using PCM in heterogeneous network: WLAN uplink channel with SNR=15 dB, IP network with 12% packet loss, WLAN downlink channel with SNR=15 dB; ROHC header compression on WLAN links: Frame loss rate, data rate, and delay for different frame lengths and transmission strategies A), B), C) as explained in the text.

5.5 Voice over IP on UMTS Packet Channels

In this scenario, a Voice over IP (VoIP) transmission on a UMTS packet-switched (PS) channel is considered. The AMR speech codec is used, which is also defined as the standard codec for circuit-switched UMTS channels and therefore already implemented in most devices. For the fixed frame length of the multi-rate AMR codec, the quality shall be optimized by controlling the encoding rate and the amount of redundancy to add with packet level FEC schemes. The potential of the AMR codec to achieve a robust quality through adaptation of source encoding rate and channel coding rate in dependence on the current channel quality is currently not utilized in circuit-switched systems. The following discussion will demonstrate that the application of the AMR codec in VoIP systems may utilize this adaptability as intended.

Assuming a dedicated PS 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 the error robustness by transmission of redundancy. The following studies therefore use 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. All FEC frames will be piggybacked to the speech packets. The different schemes then require approximately the same total data rate. Slight differences in the resulting packet lengths are negligible for the channel model behavior.

For the prediction of the resulting conversational quality, the E-model rating factor is determined as described in Appendix H.3. The calculation is based on the employed AMR mode, the predicted residual frame loss rate and mean burst length for the considered FEC scheme, as well as the resulting end-to-end delay. The loss and delay predictions are determined according to the derivations in Chapter 4 and are based on the UMTS channel model from Appendix E.2.1. In particular, the residual frame loss rate is calculated according to (4.10) for no FEC, (4.19) for repetition with piggybacked transmission (REP-PB), (4.46) for bock codes (RS-PB), and (4.28), (4.32) for XOR-PB schemes. The resulting delay for the considered FEC scheme is calculated as listed in Table 4.1.

The results are shown in Fig. 5.12 for different packet loss rates P_{pl} on the UMTS channel. The graph compares the residual frame loss rate after correction by the respective FEC scheme (left) and the resulting E-model rating factor (right). 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 which depends on the utilized FEC scheme. For increasing packet loss rates on the channel, the E-model rating factor R 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 and the lower delay. At increasing packet loss rates, a simple *repetition* of one frame transmitted in the following

packet (distance $d_\mathrm{p} = 1$) 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 (6.7, REP $p=1$, $d_\mathrm{p}=1$). Since the considered channel does not produce completely independent losses, but rather burst losses, the resulting loss rate for this repetition can be considerably lowered when transmitting the repeated frame three packets later (6.7, REP $p=1$, $d_\mathrm{p}=3$) 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 (4.75, REP $p=2$, $d_\mathrm{p}=2$) cannot compete with the other schemes at the considered loss rates. Only at very high loss rates, the considerably increased robustness against loss can compensate for the low base quality and the increased delay. *Block codes* are flexible in design (code rate, block lengths) and are efficient in reconstructing missing frames, as can be seen for two exemplary configurations with code rate 1/2 (6.7, RS $n=4$, $k=2$; 6.7, 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 (cf. Section 4.1.6.2). Although it achieves the second lowest residual frame loss rate, the XOR scheme with delayed transmission of the FEC frames (6.7, XOR Scheme 2c)) 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 UMTS channel if the rate of packet loss exceeds 2 %. At lower loss rates, no FEC is required and the transmission rate should be completely utilized for the highest AMR encoding mode.

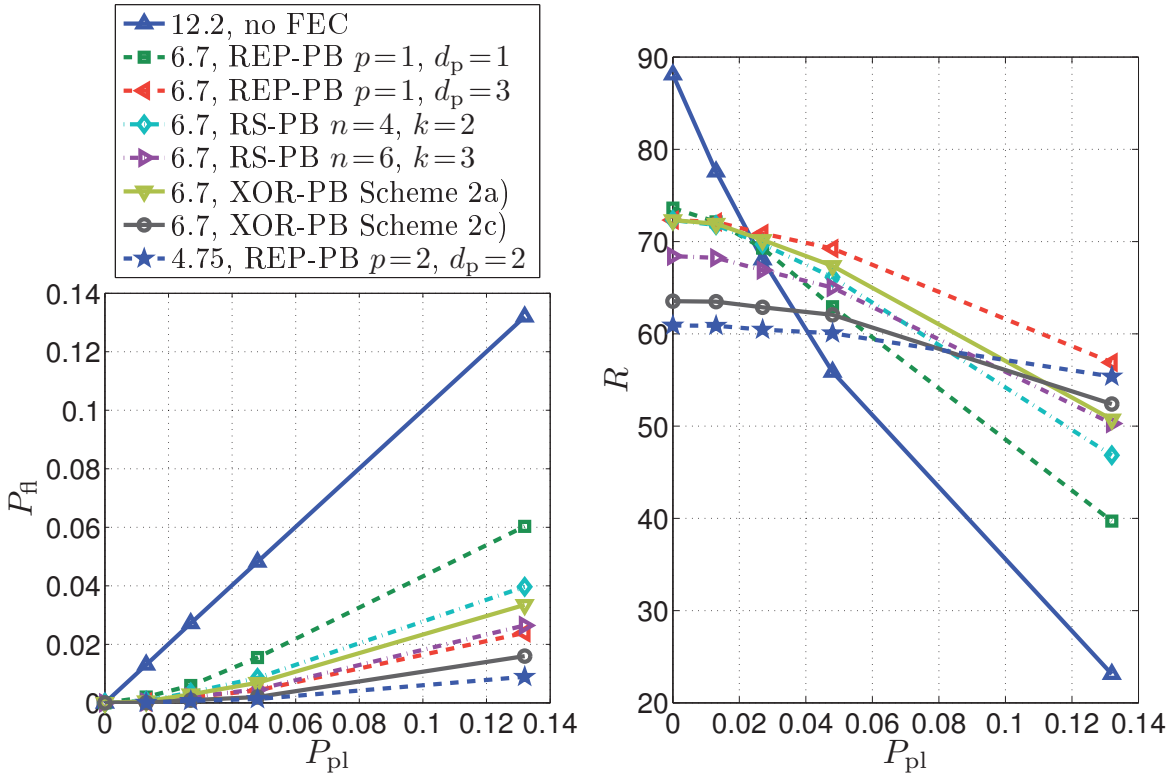


Figure 5.12: VoIP using AMR codec on UMTS channels with different packet loss rates and base delay of 100 ms: Residual frame loss rate after correction and E-model rating factor for different AMR encoding rates and FEC schemes.

5.6 Voice over IP on IP Channels with Varying Transmission Delays

In packet transmission networks, the packets may experience varying end-to-end delays as explained in Section 2.6.1. Especially for long-distance connections in the public Internet, the delay variations (jitter) may get fairly large. To compensate for this variation, the receiver can extend the length of the receiver buffer and thereby reduce the amount of packets that will have to be disregarded due to delay. However, although a lower packet loss rate increases the quality of the signal, the conversational quality of a *Voice over IP* call also depends on the total end-to-end delay. Hence, the receiver buffer cannot be extended arbitrarily.

5.6.1 Optimal Choice of Jitter Buffer Length

The receiver in networks with considerable variation in transmission delay needs to find the optimal trade-off between the end-to-end delay of the transmission and the resulting packet loss rate by controlling the length of the jitter buffer. The relation between jitter buffer length and packet loss rate for a given channel model is given in (3.34b) and (3.35) in Section 3.3.3. The buffer length then directly determines the resulting end-to-end delay. Finally, the influence of packet loss and delay on the resulting quality can be jointly assessed by the rating factor R of the ITU-T E-Model as defined in [ITU-T Rec. G.107 2005] and introduced in Appendix H.3.

The following scenario shall be considered as an example: A *Voice over IP* call using the AMR speech codec (12.2 kbit/s mode) is transmitted over the public Internet and experiences the delay and jitter characteristics from the example given in Section 3.3.1 with the following parameters of the Weibull distribution: $\mu = 116$ ms, $\alpha = 15.9$, $\gamma = 0.4451$. This example has been taken from [Sun and Ifeachor 2004], where the optimal jitter buffer length has been determined in the same way. Further packet losses are assumed at different loss rates, which are not caused by the varying delay but other factors such as congestion or transmission errors on wireless links in the access networks. The resulting frame loss rate including delay inflicted losses, calculated as given in (3.35), and the according E-Model rating factor are given in Figure 5.13 in dependence on the end-to-end delay. The end-to-end delay D includes the chosen jitter buffer length as defined in (2.6). The different curves describe different additional loss rates of 0% to 6%. For an increasing buffer length, the resulting frame loss rate approaches the network loss rate as the variation of the transmission delay becomes more and more compensated. However, an increasing delay is also an increasing amount of impairment to the conversational quality. Each curve of the E-Model rating factor R shows a point of optimum in which the best compromise is achieved between loss rate and delay.

In [Sun and Ifeachor 2004], an adaptive jitter buffer has been proposed which is able to adjust the playout delay to "spikes" in the transmission delay, i.e., sudden groups of packets which have a significantly higher delay than the rest. This adaptation increases the overall quality as it reduces the packet losses while preventing

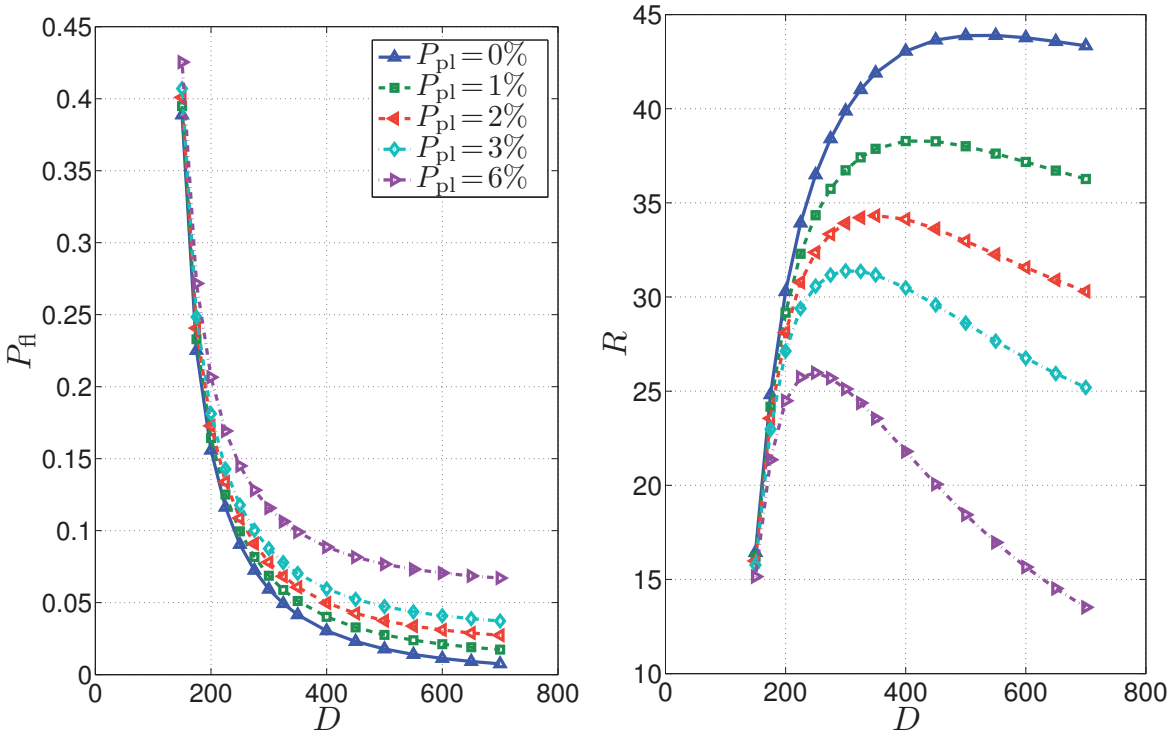


Figure 5.13: VoIP using AMR codec (12.2 kbit/s mode) on IP channel with jitter (Weibull distribution: $\mu = 116$ ms, $\alpha = 15.9$, $\gamma = 0.4451$) and additional packet losses of different rates P_{pl}: Residual frame loss rate and E-model rating factor for different end-to-end delays D which include the receiver buffer length D_{buf}.

the impairment of a constantly high playout delay. Nevertheless, the transmission channel of the given example does not allow for a high quality VoIP transmission. The delay required for compensating the jitter inflicted losses is too high, resulting in a poor conversational quality. Especially if additional network losses occur, the quality does not reach an acceptable level.

5.6.2 Joint Optimization of Jitter Buffer and Forward Error Correction

The previous example has shown that the quality of a VoIP connection over an IP network with a high one-way delay and considerable amount of jitter is strongly limited through the resulting end-to-end delay and packet losses. A possible solution for achieving an acceptable quality in such severe network conditions is the use of forward error correction (FEC) schemes. When adding redundancy, the source codec rate should be reduced in order to keep the overall packet data rate constant. This prevents a possible worsening of the transmission characteristics as it might occur when increasing the data rate in a network with a considerable amount of load. Assuming the same channel characteristics as in the previous section, the following combinations of AMR codec modes and FEC schemes are considered, which result in comparable packet data rates:

- AMR 12.2 kbit/s; no FEC
- AMR 6.7 kbit/s; repetition of one frame piggybacked to following packet (REP-PB, $p = 1$, $d_p = 1$)
- AMR 6.7 kbit/s; Reed-Solomon (n, k) block code of rate 1/2; FEC frames piggybacked to following packets (RS-PB, $(n, k) = (4, 2)$, $(6, 3)$, or $(8, 4)$)
- AMR 4.75 kbit/s, repetition of two frames piggybacked to following packets (REP-PB, $p = 2$, $d_p = 1$)
- AMR 4.75 kbit/s; Reed-Solomon (n, k) block code of rate 1/3; FEC frames piggybacked to following packets (RS-PB, $(n, k) = (6, 2)$, or $(9, 3)$)

In contrast to the studies in Section 5.5, no interleaving or delayed transmission of the FEC frames are considered in this case because the losses in the channel model considered here do not show a bursty behavior. The resulting frame loss rate depends on the considered FEC scheme and the end-to-end delay which itself depends on the chosen receiver jitter buffer length. For the calculation of the loss rate, the probability of jitter based losses is incorporated into the channel model according to (3.34b) and (3.35) using the position dependent loss probability derived in (4.72). As explained in Section 4.3, the dependency of $P_{l,J}$ on both the position of the considered frame within the encoding block and the position of the other frames of the block needs to be taken into account in calculating the loss probabilities for the considered FEC scheme.

The resulting frame loss rate and E-model rating factor in dependence on the end-to-end delay D including the jitter buffer length are shown in Figures 5.14, 5.15, and 5.16 for 0%, 3%, and 6% additional network losses, respectively. The curves show that piggybacked FEC data is able to recover parts of jitter as well as network losses, leading to a considerable increase of the resulting quality. The narrow peaks of the curves, however, also show that a careful dimensioning of the receiver buffer length is essential for achieving the optimal performance.

For no additional packet losses (Figure 5.14), the optimal performance is achieved for the Reed-Solomon block code with $(n, k) = (6, 3)$ which is able to recover the loss of 3 frames within a group of 6 successive packets. This scheme requires a buffer length of 250 ms. A shorter block length of $(n, k) = (4, 2)$ results in an only slightly lower rating factor R which it already achieves at a buffer length of 225 ms. A code with the same rate but larger block length of $(n, k) = (8, 4)$, on the other hand, leads to a lower quality in spite of its higher correction capabilities because of the increased delay at the receiver. The transmission of more redundancy at the expense of a lower base quality does not lead to an improved quality in this case as evident from the curves for the AMR 4.75 kbit/s mode with different FEC schemes of rate 1/3.

The maximum achievable quality decreases only slightly with increasing network loss rate (cf. Figure 5.15 and 5.16) because the optimal FEC scheme is capable of recovering most losses of both types, delay and network based. For a higher network

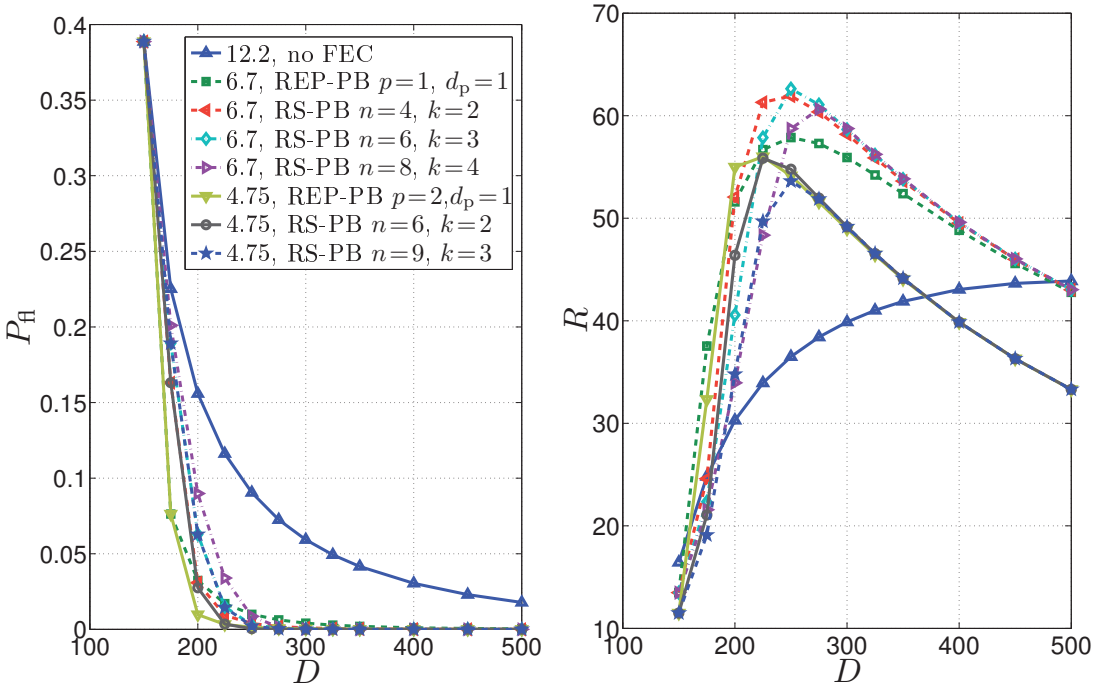


Figure 5.14: VoIP using AMR codec (12.2 kbit/s mode) on IP channel with jitter (Weibull distribution: $\mu = 116$ ms, $\alpha = 15.9$, $\gamma = 0.4451$), no additional packet losses: Residual frame loss rate after correction and E-model rating factor for different FEC schemes and end-to-end delays D which include the receiver buffer length D_{buf}.

loss rate, a larger block length of the RS code should be used. For example, at 6% network losses, the $(n, k) = (6, 3)$ code leads to clearly better results than the $(n, k) = (4, 2)$ code (cf. Figure 5.16).

5.7 Conclusions

This chapter has discussed the system optimization problem for a range of common application and network scenarios and determined exemplary parameter settings. The choice of various system parameters, such as the codec to utilize, the frame length to transmit in each packet, as well as the parameterization of FEC and retransmission schemes have been discussed in detail. To find the optimal parameter settings, the channel model and analytical results of previous chapters have been applied to specific relevant application and network scenarios, e.g., VoIP or music streaming on UMTS, WLAN and the Internet for exemplary channel conditions.

It has been shown that the optimization needs to consider the specific demands and constraints of the given scenario in order to achieve the optimal quality in the most efficient way considering the available resources. The framework which has been developed in this work has shown to provide the necessary base for the optimization of these complex transmission systems.

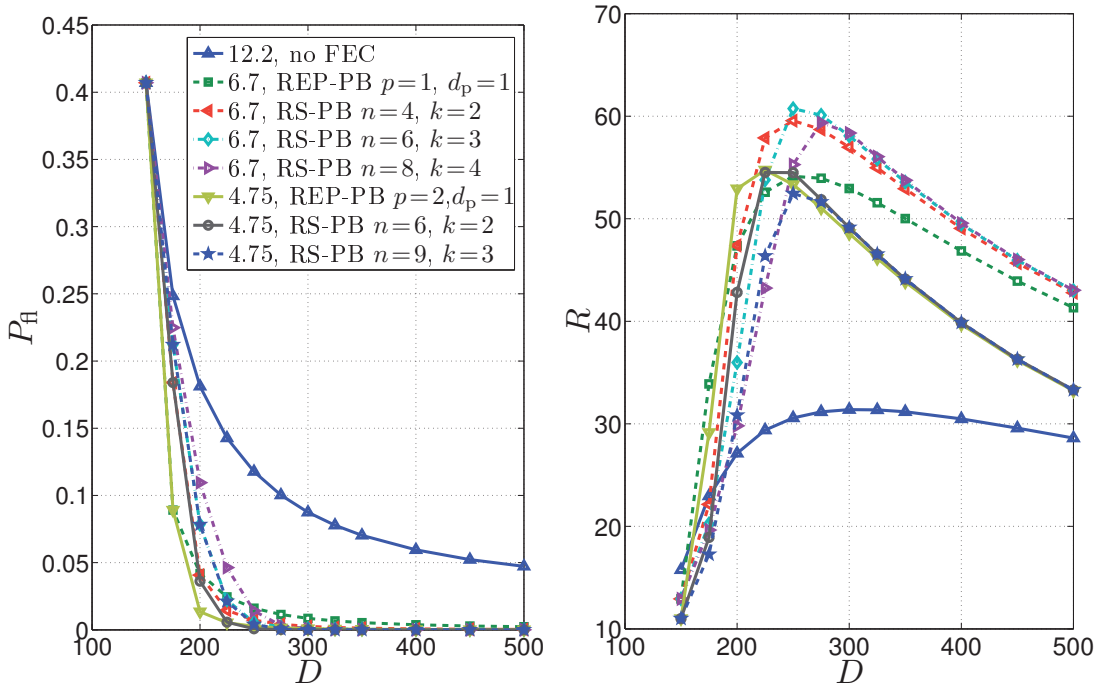


Figure 5.15: VoIP using AMR codec (12.2 kbit/s mode) on IP channel with jitter (Weibull distribution: $\mu = 116$ ms, $\alpha = 15.9$, $\gamma = 0.4451$), additional packet losses with rate $P_{\mathrm{pl}} = 3\%$: Residual frame loss rate after correction and E-model rating factor for different FEC schemes and end-to-end delays D which include the receiver buffer length D_{buf}.

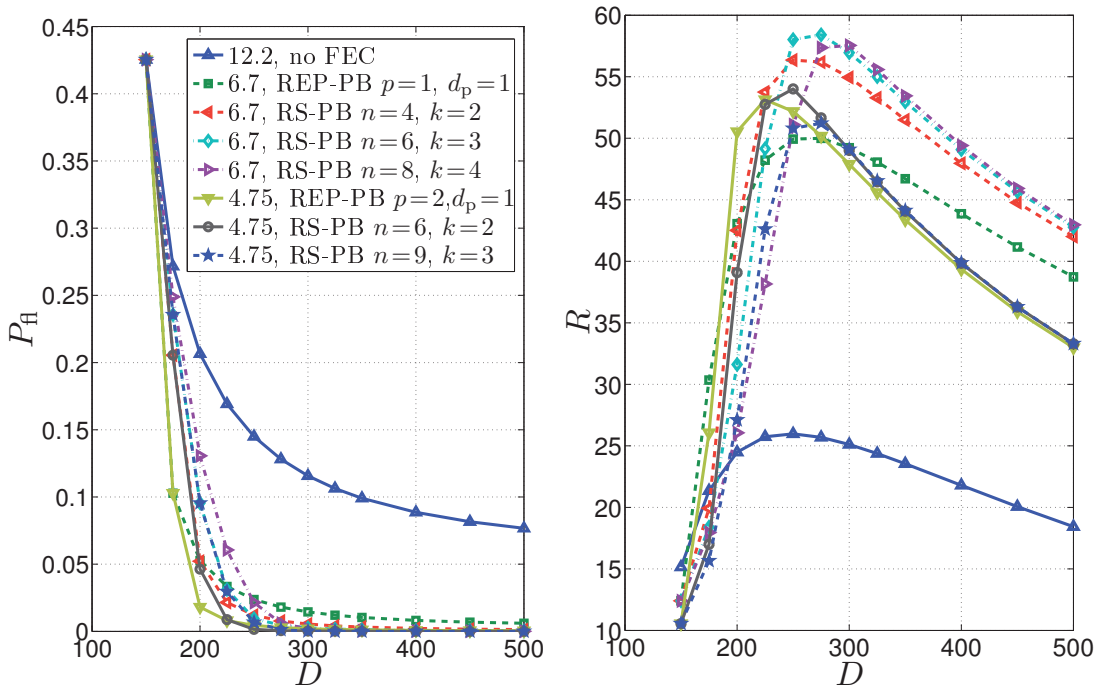


Figure 5.16: VoIP using AMR codec (12.2 kbit/s mode) on IP channel with jitter (Weibull distribution: $\mu = 116$ ms, $\alpha = 15.9$, $\gamma = 0.4451$), additional packet losses with rate $P_{\mathrm{pl}} = 6\%$: Residual frame loss rate after correction and E-model rating factor for different FEC schemes and end-to-end delays D which include the receiver buffer length D_{buf}.

6

Packet Loss Concealment

The transmission of speech and audio signals over possibly heterogeneous packet-switched communication networks experiences packet losses and thereby losses of media frames, caused by various sources of impairment as described in Section 2.6.2. With current protocols, packets are either received error-free or otherwise lost completely[1]. The length of the signal segment which is lost in case of a packet loss depends on the number of speech/audio frames that are transmitted in each packet and on the frame length itself. A certain amount of frame losses may be compensated by forward error correction schemes or retransmissions as discussed in Chapter 4. In real-life scenarios, the utilization of such methods is however limited by delay constraints of the application and data rate constraints of the network. The receiver will therefore still experience frame losses under bad network conditions. Hence, packet loss concealment, or more specific, frame loss concealment algorithms have to be implemented in the receiver to generate a suitable replacement for lost signal frames which cannot be recovered.

The current chapter will discuss approaches for the concealment of lost speech frames which have been encoded by speech codecs based on the CELP (code excited linear prediction) encoding principle [Schroeder and Atal 1985]. This coding scheme is most widely used in state-of-the-art speech codecs for mobile communication networks, e.g., the *Adaptive Multi-Rate* (AMR) codec [3GPP TS 26.090], as well as for the base layers of recently standardized scalable codecs for heterogeneous packet networks, e.g., [ITU-T Rec. G.729.1 2007] and [ITU-T Rec. G.718 2008]. First, an overview is given on the different concepts of packet loss concealment that are implemented in current systems and proposed in the literature. In the following, two new concepts are developed which improve the performance of standard concealment techniques implemented in CELP based speech codecs.

The first approach classifies the signal structure of the preceding and following segments and applies particularly tailored extra- and interpolation techniques based on the current voicing state of the signal. In the second approach, the transmission of low rate side information in a following packet is proposed which assists the

[1] See Chapter 2.7.5 for a discussion of utilizing packets with residual bit errors in the payload.

decoder in the concealment of a lost frame. It will be shown that the side information can be transmitted as hidden bit stream in the codec's parameters and therefore does not require any additional date rate.

6.1 State-of-the-Art in Packet Loss Concealment

Measures for combating packet loss in packet-based transmission systems may be classified into *sender-driven*, *receiver-based*, and *sender-assisted* approaches as explained in Section 2.7. While sender-driven approaches have been discussed in detail in Chapters 4 and 5, the current chapter focuses on *packet loss concealment* (PLC), i.e., the concealment of lost and unrecoverable signal frames at the receiver, possibly assisted by specific side information which is derived at the sender and transmitted in a following packet.

The receiver-based concealment of lost signal segments of multimedia signals utilizes a-priori knowledge on the signal characteristics[2] together with information from already received segments preceding or following the loss (if available) to construct an appropriate estimation as replacement for the lost segment. The concealment algorithms for speech, music, and video signals will therefore differ from each other in specific details, but also have basic principles in common. The estimation process itself further differs in dependence on the utilized coding scheme, e.g., for speech or music signals between waveform and parametric or hybrid encoding schemes.

In this chapter, the focus will be on speech signals. Speech signals can be assumed to be piecewise stationary (short-term stationarity) and to often exhibit a high degree of periodicity. This property is utilized by source coders to achieve a high compression rate and thus a low resulting bit rate. The periodicity and stationarity of speech is also the essential basis for applying an effective concealment of lost segments.

Packet loss concealment algorithms for *waveform codecs* use methods of waveform substitution and are usually based on an extrapolation of the previous signal segment by pitch period repetition as, e.g., defined in [ITU-T Rec. G.711 Appendix I 1999] for A/μ-law encoded PCM speech. Overlap-add techniques are applied to smooth the transition between the extrapolated signal and the following received signal segment. For an improved quality, it has also been proposed to determine and extrapolate the spectral envelope of the preceding segment, and to subsequently perform the repetition of the pitch structure in the LP residual domain [see Gunduzhan and Momtahan 2001]. The latter approach is also applied for the wideband coding standard ITU-T G.722 as explained in [Kövesi and Ragot 2008; ITU-T Rec. G.722 Appendix IV 2008]. Further approaches include, e.g., waveform extrapolation in sub-bands [Clüver and Noll 1996], or time scale modification techniques [Sanneck et al. 1996]. Finally, [Shetty and Gibson 2006] propose a PLC algorithm for

[2]Depending on the signal type (speech, music, video), this a-priori knowledge may include models of the speech production process or of auditory and visual perception, as well as statistical information, e.g., probability distributions of signals and their features.

ITU-T G.722 which is assisted by a transmission of low rate side information piggybacked to the following packet. This side information includes the encoder states and the correct pitch value of the previous frame. The use of this side information is claimed to improve the concealment as well as to limit error propagation.

Packet loss concealment algorithms for frame-based *parametric* or *hybrid speech codecs* perform the concealment usually in the parameter domain. The most prominent representative of the hybrid codec family is the CELP (code excited linear prediction) encoding principle [Schroeder and Atal 1985]. These codecs decompose the speech signal into its spectral envelope and a mixture of properly scaled periodic and noise-like excitation signals, described by a set of parameters. The concealment is usually based on an extrapolation of the parameters, as implemented in most conventional concealment algorithms. If future frames are already available in the receiver buffer, they can be utilized for interpolative approaches [see, e.g., Wang and Gibson 2001; Johansson et al. 2002; Mertz et al. 2003].

The CELP codec structure, which has originally been developed for circuit-switched mobile networks, is very sensitive to frame losses. To achieve a trade-off between encoding efficiency and error propagation, the latter has not been completely removed. The predictive encoding of parameters and in particular the application of a long-term prediction lead to a considerable inter-frame dependency which still results in a strong effect of error propagation. An inappropriately estimated frame might therefore effect the quality of several frames following the loss. Compared to waveform codecs, this reduced error robustness is the prize payed for a considerably lower bit rate. Alternative encoding approaches that have been designed particularly for application in networks with packet losses avoid these inter-frame dependencies, e.g., the Internet Low Bit Rate Codec (iLBC) [Andersen et al. 2004]. The higher robustness against packet loss, however, is achieved at the expense of an increased bit rate and/or a reduced base quality of the decoded signal.

Several techniques have been proposed in the literature to reduce the error propagation of conventional CELP based codecs. For example, [Chibani et al. 2005] presented an approach that constrains the contribution of the adaptive codebook — mainly responsible for the error propagation — by limiting its gain. Thereby, the innovative codebook is forced to partially model the pitch excitation. According to the authors, this modification of the encoder does not lead to a significant reduction in the base quality. The introduced periodicity in the innovation can subsequently also be utilized at the receiver to improve the estimation of a lost frame, [Chibani et al. 2006]. Another approach proposed in [Gournay et al. 2003] makes use of late arriving packets which already had been regarded as lost and therefore concealed. The information in these late packets is still evaluated and used to correct the decoder states in order to limit the error propagation. A considerable improvement with this approach is therefore only expected if this scenario occurs with a certain frequency, e.g., if the receiver buffer has to be very short because of tight delay constraints.

There have been several proposals in the literature for applying statistical estimation techniques to determine the parameters of a lost frame [see, e.g., Martin et

al. 2001; Murthi et al. 2006; Rodbro et al. 2006; Lahouti et al. 2007]. Such methods are claimed to achieve a higher accuracy of the estimated parameters, however, at the expense of a very high computational complexity. A cost-benefit consideration may therefore not approve the use in current systems, and indeed, such approaches are not yet applied in standardized speech codecs. Most proposals have concentrated on the signal's spectral envelope. In [Agiomyrgiannakis and Stylianou 2005], the estimation is further assisted by side information. While among all parameters the spectral envelope is certainly very important for the quality of the speech sound, the other parameters have a significant influence themselves and must not be neglected in the concealment. The different significance of the parameters for the resulting quality is investigated in Section 6.2, also revealing a dependence on the current signal structure. Thus, an integrated approach is required which involves all parameters and which adapts to the instantaneous signal characteristics.

In Section 6.3, a voicing dependent concealment approach is developed based on the author's proposal in [Mertz et al. 2003], considering different concealment techniques for voiced, unvoiced and transitional speech segments. Alternative concealment approaches utilizing a classification of the signal structure can be found, e.g., in [Jelinek and Salami 2007; Vaillancourt et al. 2007]. Other approaches consider signal classification based on Hidden Markov Models (HMM) and subsequently employ statistical estimation techniques of parameters based on class dependent probability distributions [see, e.g., Murthi et al. 2006; Rodbro et al. 2006].

In contrast to proposals for transmitting selective or partial redundancy, e.g., copies of certain important parameters for specific sensitive frames (cf. Section 2.7), several approaches have been proposed in the literature which derive and transmit specific side information for the receiver's concealment routine. These techniques are categorized as *sender-assisted packet loss concealment* in this work. This side information is generally transmitted at a low bit rate (below 1 kbit/s), since these approaches are targeted at heterogeneous packet networks including wireless transmission links. Section 6.4 of this work will introduce low complex parameter estimation techniques which are supported by low rate side information, based on the author's proposals in [Mertz and Vary 2006]. These techniques can be applied to any standardized CELP codec. Section 6.5 will then show, by the example of the AMR codec, how this side information can be embedded as hidden steganographic bit stream in the codec bits, thereby requiring no additional bit rate and achieving backward compatibility with legacy systems [Geiser et al. 2008].

Concealment strategies utilizing signal classification and side information are increasingly incorporated in new speech codecs. The recently standardized scalable codecs introduced in Section 2.3.2.4 (VMR-WB; ITU-T G.729.1, G.711.1, and G.718) were especially designed for the application in packet networks and already include such measures for a higher robustness against packet loss. They incorporate efficient packet loss concealment algorithms considering signal classification and the concealment is further assisted with low rate side information as part of each frame's bit stream.

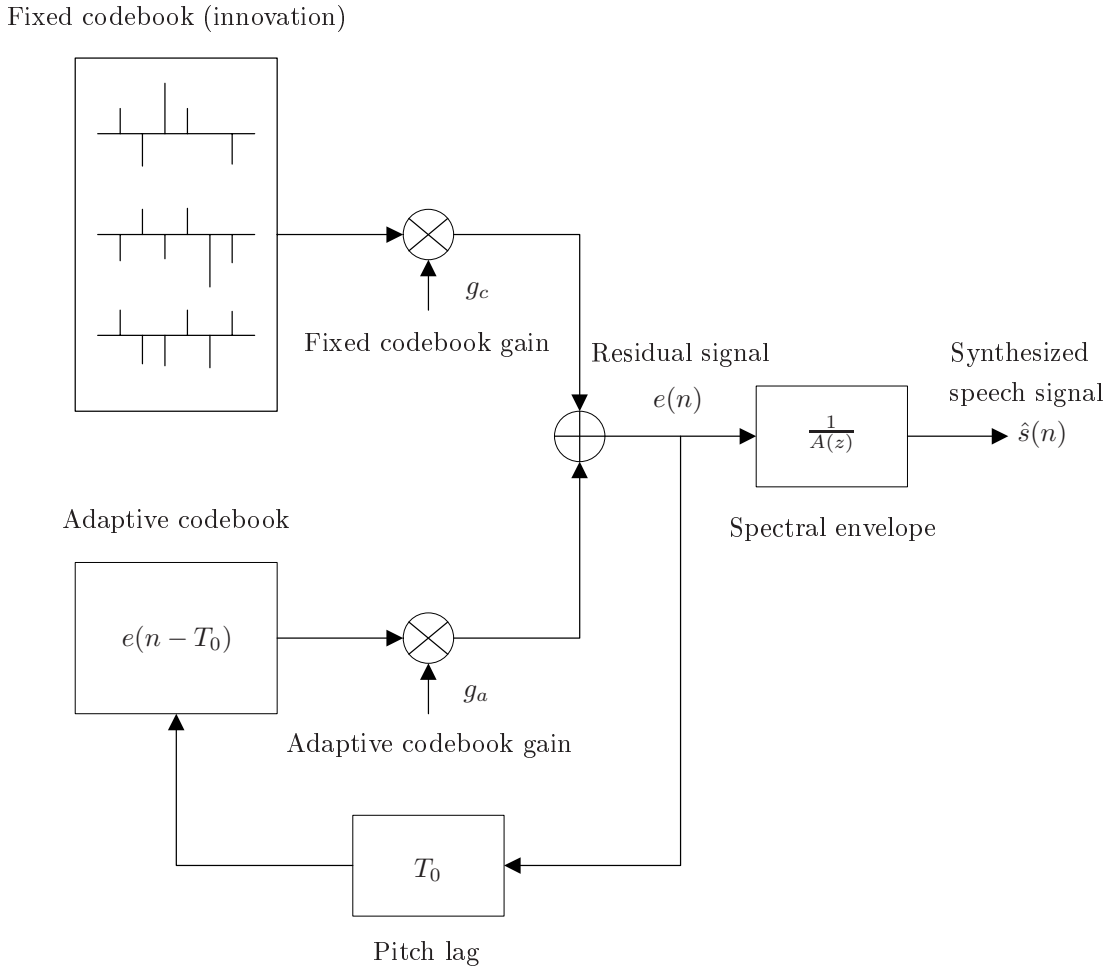


Figure 6.1: CELP decoder structure: Frame-wise synthesis of the speech signal from the received parameters of each frame: spectral envelope, pitch lag, fixed codebook entry, and fixed and adaptive codebook gains.

6.2 CELP Codec Parameters and their Significance for Quality

Typically, CELP codecs compute the following parameters for each frame of the speech signal, which has normally a length of 10-30 ms and is often 20 ms long. After transmission, the decoder resynthesizes the speech signal from the received parameters as shown in Figure 6.1:

1. **Spectral envelope**: coefficients of a short-term prediction filter $A(z)$, usually encoded as line spectral frequencies (LSF) and calculated once or twice per frame; the speech signal is filtered with the analysis filter to remove the spectral envelope; all other parameters are derived from this residual signal.

2. **Pitch lag** T_0: coefficient of a long-term prediction filter, expressing the current periodicity of the residual signal; calculated for every sub-frame of, e.g.,

5 ms; usually realized as an index into an adaptive codebook which contains the preceding quantized residual signal.

3. **Innovation sequence**: entry from a fixed codebook of noise-like and usually sparsely filled pulse vectors; models the error of short- and long-term prediction; chosen for every sub-frame of about 5 ms.
4. **Codebook gains**: calculated for every sub-frame of, e.g., 5 ms
   a. Gain g_a of the contribution from the long-term prediction, i.e., the adaptive codebook entry.
   b. Gain g_c of the contributing innovation signal, i.e., the fixed codebook entry.

The parameters have a different significance for the resulting speech quality and therefore they also have a different sensitivity to transmission errors or loss. In mobile communication systems like GSM, the bitstream of encoded parameters is therefore sorted into different classes of sensitivity for which subsequently different grades of error protection are applied (so-called *unequal error protection*). However, there is a significant difference between the sensitivity to bit errors and the sensitivity to loss. The former is determined by the influence of bit errors on the resulting quality, i.e., when a distorted parameter is used for speech synthesis. The sensitivity to loss, on the other hand, describes the influence on the quality in case a parameter has to be estimated because it is lost or severely distorted and therefore discarded. Hence, this sensitivity describes how well a parameter can be estimated from the information at hand, e.g., preceding and following frames, and/or particularly transmitted side information. Knowledge about the parameters' variation in significance is important for the development of suitable frame loss concealment methods. When considering the transmission of side information to assist the receiver based concealment, such knowledge about the influence of each parameter will help in deciding which amount of side information (i.e., bit rate) is needed for each parameter.

The parameters' sensitivity to loss will also depend on the current signal structure. The CELP encoding principle is based on short-term and long-term prediction utilizing the short-time stationarity of speech as well as the periodicity of voiced speech signals. The following evaluation of the sensitivity will therefore also consider different voicing transitions of the speech signal separately. For this purpose, the voicing state of the preceding and the following frame have been determined according to the method defined later in Section 6.3.1, leading to four different possible transitions: unvoiced-unvoiced (u-u), unvoiced-voiced (u-v), voiced-unvoiced (v-u), and voiced-voiced (v-v). Speech pauses will normally be considered as unvoiced as the signal will only exhibit some low background noise.

In the following, the significance of CELP codec parameters and their sensitivity to loss shall be assessed from a set of measurements and simulations, using the standardized AMR codec [3GPP TS 26.090] as example. As data set, the speech files from [ITU-T Rec. P.834 2002] have been taken.

6.2.1 PESQ Measurements of the Separate Concealment of AMR Codec Parameters

In the simulations described below, some artificial loss scenarios have been assumed which normally do not occur in practical systems but serve well for the comparison of the parameters' loss sensitivity. In the simulations, the set of lost parameters have been varied, i.e., in a specific simulation not all parameters of a frame are considered as lost. Lost parameters were always estimated by the standard concealment method for the AMR codec defined in [3GPP TS 26.091]. The quality of the synthesized speech signals has been assessed by the objective quality measure PESQ (Perceptual Evaluation of Speech Quality) [ITU-T Rec. P.862 2001] (cf. Appendix H.2). Two sets of simulations have been carried out:

a) loss of all but one parameter: all parameters of lost frames are concealed with the standard method, except for one parameter which is set to the correct/original value;

b) loss off only one parameter: all parameters of lost frames are set to correct/original values, except for one parameter which is concealed with the standard method.

For both sets, the different voicing transitions from the frame preceding the loss to the one following the loss[3] have been considered separately. Therefore, random frame losses have been generated for each speech file, but only those frames have been considered lost which belong to the considered voicing category. The mean MOS-LQO[4] values for the two simulation sets a) and b), as computed by the PESQ algorithm, are given in Table 6.1 and Table 6.2, respectively. For the different voicing transitions, different loss rates result as shown in the tables. The reason lies in the distribution of voicing transitions in a speech file.

The simulation results for random losses over *all voicing transitions* show that a loss of the codebook gains has the most severe effect on the resulting speech quality, while the loss of the other parameters seem to have a lesser and — when compared to each other — similar effect. The high loss sensitivity of the gains can also be observed when the losses only occurred in one type of the different voicing transitions, the similarity of the other parameters, however, does not hold anymore.

For a loss within an unvoiced speech signal (u-u transition), the loss of the pitch lags naturally shows the least influence on the quality because of the lack of periodicity in such a signal. Also, the correct choice of the fixed codebook entry seems not to be of the highest importance, although the "estimation" (random entry) leads to a slight quality decrease. A stronger influence can be seen in the LSF coefficients which describe the signal's spectral envelope.

The influence of the loss of parameters is in general much higher within voiced signal segments (v-v transition). Here, the regeneration of the signal's periodicity

[3] The classification has been done according to Section 6.3.1.

[4] The MOS-LQO value computed by the PESQ algorithm is an estimation of the *Mean Opinion Score* (MOS) that would result from a listening test. MOS-LQO stands for *Mean Opinion Score–Listening Quality Objective*.

Retained Parameter	MOS-LQO for different voicing transitions				
	$P_{\mathrm{fl}}=9.8\%$ random voicing transitions	$P_{\mathrm{fl}}=10\%$ only u-u frames lost	$P_{\mathrm{fl}}=3.2\%$ only u-v frames lost	$P_{\mathrm{fl}}=3.0\%$ only v-u frames lost	$P_{\mathrm{fl}}=8.6\%$ only v-v frames lost
LSF (spectral env.)	2.77	3.20	3.19	3.58	2.80
T_0 (pitch lags)	2.61	3.06	3.02	3.47	2.75
g_a, g_c (CB gains)	2.84	3.52	3.20	3.65	2.75
fixed CB entry	2.61	3.09	3.01	3.51	2.72
no parameter	2.52	3.04	2.96	3.44	2.61
LSF & gains	3.26	3.84	3.54	3.85	3.09
T_0 & gains	3.00	3.55	3.33	3.69	2.98
fixed CB & gains	3.04	3.64	3.34	3.74	2.94

Table 6.1: Mean MOS-LQO for simulation set a); AMR 12.2 kbit/s; given parameter is kept as correct/original, others are concealed by the standard method; different frame loss rates and restriction of losses to certain voicing transitions. The PESQ MOS-LQO value for error free transmission (i.e. only reflecting the codec distortion) resulted to 4.03.

Estimated Parameter	MOS-LQO for different voicing transitions				
	$P_{\mathrm{fl}}=9.8\%$ random voicing transitions	$P_{\mathrm{fl}}=10\%$ only u-u frames lost	$P_{\mathrm{fl}}=3.2\%$ only u-v frames lost	$P_{\mathrm{fl}}=3.0\%$ only v-u frames lost	$P_{\mathrm{fl}}=8.6\%$ only v-v frames lost
LSF (spectral env.)	3.47	3.73	3.63	3.84	3.52
T_0 (pitch lags)	3.52	3.95	3.69	3.95	3.35
g_a, g_c (CB gains)	3.10	3.27	3.37	3.70	3.29
fixed CB entry	3.51	3.88	3.72	3.89	3.39
all parameters	2.52	3.04	2.96	3.44	2.61

Table 6.2: Mean MOS-LQO for simulation set b); AMR 12.2 kbit/s; given parameter is concealed by standard method, others are kept as correct/original; different frame loss rates and restriction of losses to certain voicing transitions. The PESQ MOS-LQO value for error free transmission (i.e. only reflecting the codec distortion) resulted to 4.03.

is essential, leading to quality degradation if the pitch lags and gains are incorrectly estimated. The fixed codebook entry is of similar significance because it describes the variation of the signal from the long-term prediction. Slightly less important, but still significant is a sufficient estimation of the signal's spectral envelope.

A loss of parameters at the transition from an unvoiced to a voiced speech segment (u-v transition) shows the highest impact on the resulting quality of the speech signal. At such a transition, the starting voiced speech cannot be predicted yet, and the combination of basically all parameters is necessary to re-synthesize the signal with a good quality. Finally, transitions from voiced to unvoiced signal segments (v-u transition) are the ones least sensitive to loss for any of the parameters, especially when considering the ends of talk spurts.

6.2.2 Conclusions

The results of this study show that the parameters of the AMR codec have different sensitivities to loss and that this sensitivity varies with the current voicing state or transition of the speech signal. In the simulations, the lost parameters have been concealed with the standard concealment technique proposed for the AMR codec in [3GPP TS 26.091]. It is based on parameter extrapolation, assumes no knowledge about future frames, and has therefore shown to be limited in its performance. The estimation of parameters can be improved if further information is available at the receiver. In the following sections, two approaches will be developed and discussed which a) utilize future frames and consider particular suited concealment techniques for the different voicing transitions, and b) utilize additionally transmitted side information to further improve the concealment of the parameters.

6.3 Voicing Controlled Packet Loss Concealment for CELP Encoded Speech Signals

In this section, a novel voicing controlled method for frame loss concealment is developed and applied to the Adaptive Multi-Rate (AMR) speech codec. The inherent ACELP codec structure is very sensitive to frame losses, as the predictive encoding of parameters like the LSF coefficients can lead to error propagation in case of transmission errors. Additionally, an inaccurate estimation of a lost frame will affect the decoding of following frames due to incorrect adaptive codebook entries. A high quality concealment routine is therefore essential to maintain an acceptable quality under adverse network conditions.

In case of frame loss, CELP codecs usually employ a concealment unit which tries to extrapolate the previous signal structure and at the same time gradually lowers the amplitude of the signal resulting in a completely muted signal if several consecutive frames are lost. Such extrapolation/muting based concealment approaches have been developed for circuit-switched cellular networks, where frames may have to be discarded if residual bit errors are detected in the important payload bits. In these systems, the receiver's concealment routine can only utilize the preceding signal for its estimation. In packet-switched networks, on the other hand, receiver buffers (so-called jitter buffers) are required for compensating varying transmission delays. If, in case of frame losses, the frame following the lost frames has already been received, it may be utilized by the frame loss concealment routine. The utilization of frames succeeding a loss for interpolation of parameters has already been discussed, e.g., in [Wang and Gibson 2001; Fingscheidt and Perez 2002] for LSF vectors, and in [Johansson et al. 2002] for pitch lags. However, the proposed methods were applied regardless of the current signal structure. The studies in Section 6.2 have shown that the achievable quality of the concealment routine highly depends on the properties of the lost signal segment. Therefore, a voicing controlled loss concealment is developed that depends on the voicing state of the speech frame

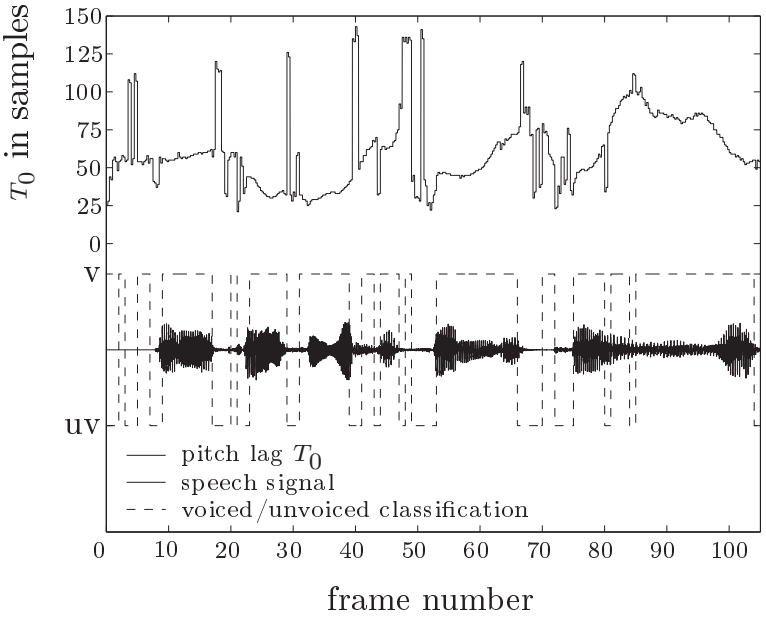


Figure 6.2: Illustration of voiced/unvoiced classification. The pitch lags T_0 (upper solid line) are used to classify the frames (20 ms) of the given speech signal (lower solid line) into voiced (v) and unvoiced (uv) segments (illustrated by the dashed line).

preceding and following the lost frames, i.e., whether the lost frames lie within a voiced, an unvoiced, or a transitional speech segment. The basic concept has been introduced by the author in [Mertz et al. 2003] and will be described in detail in the following.

6.3.1 Voicing Classification

The choice of an appropriate concealment method strongly depends on the current periodicity of the speech signal around the lost frames. After removing the spectral envelope, CELP based speech codecs employ *long-term prediction* (LTP) to achieve their compression ratio, i.e., a previous signal segment is used as prediction for the current subframe. Subsequently determined parameters then just encode the prediction error. For the long-term prediction, the pitch lag T_0 is determined for each subframe. This parameter reflects the period length of the (here assumed) periodic signal and it is the reciprocal value of the fundamental frequency F_0. The curve of the parameter T_0 is therefore a good measure for the periodicity in the signal and will be used for a classification into voiced and unvoiced speech frames. As shown in Figure 6.2, the parameter T_0 undergoes only slight variations in voiced regions of the speech, whereas it has an unpredictable and rather random behavior in unvoiced segments. Here, the T_0 curve is showing great value differences between successive subframes, which results from the missing periodicity in unvoiced speech segments. Thus, the absolute values of the T_0 differences between consecutive subframes can

be used as an indication for voiced/unvoiced speech. A decision function

$$V(n) = \sum_{i=1}^{3} |T_0^{(i)}(n) - T_0^{(i+1)}(n)| \tag{6.1}$$

is computed for each frame n, with $T_0^{(i)}(n)$ the pitch lag of subframe i ($i \in \{1, \ldots, 4\}$). With an appropriate threshold V_{th} for $V(n)$, a classification in voiced and unvoiced frames is done as follows:

$$\text{Frame is } \begin{cases} \text{voiced} & \text{for } V(n) \leq V_{th} \\ \text{unvoiced} & \text{for } V(n) > V_{th} \end{cases} . \tag{6.2}$$

For narrowband speech (sampling frequency 8 kHz) a threshold of $V_{th} = 10$ proved to be suitable for a reliable classification.

The classification may sometimes detect an unvoiced speech frame within a voiced region, as can be observed, e.g., for frame 82 in Figure 6.2. In this case the speech structure undergoes a significant change within a voiced sound, expressed in a jump of the T_0 parameter that causes $V(n)$ to exceed the threshold. Since the voicing controlled choice of an appropriate concealment method will be based on the periodicity of the speech signal, these cases have not to be considered as misclassification, but they in fact support the concealment of lost frames in that speech segment.

The classification as defined above requires a very low computational complexity when used with CELP codecs, because the pitch lags T_0 are already available from the received parameters of a frame. Furthermore, a classification of a received frame after a loss is possible because the pitch lags are not encoded predictively across frames and no other measures are involved which might require the decoded signal (e.g., the zero-crossing rate).

6.3.2 Parameter Estimation Depending on Voicing Transition

The concealment methods described in this section utilize received frames on both sides of the lost frames. Depending on the voicing classification of both the preceding and the following speech frame, a particular concealment method is chosen for each different voicing transition. The concealment is based on the codec parameters, LSF coefficients (spectral envelope), pitch lag, gain factors, and innovation vector (i.e. fixed codebook entry), which are estimated by extra- and interpolation techniques.

The LSF coefficients will be linearly interpolated as proposed in [Fingscheidt and Perez 2002]. This method is briefly reviewed in the following section, before the voicing controlled estimation of the remaining parameters is discussed.

6.3.2.1 LSF Interpolation

In CELP codecs, the LSF coefficients describing the spectral envelope of the signal are usually encoded predictively. In general, moving average (MA) prediction is employed to predict the LSF coefficients of the current frame. The difference between the computed and predicted coefficients, i.e., the residual LSF vector, is then quantized using a vector quantizer. The MA prediction guarantees a limited error propagation in case of frame loss since the re-calculation of the LSF coefficients at the receiver only depends on the received residual vectors. The order of the MA predictor is chosen as a trade-off between maximizing the prediction gain and limiting the error propagation. For example, the ITU G.729 codec [ITU-T Rec. G.729 1996b], which is still used in many Voice over IP systems, employs a 4th order moving average (MA) prediction. With a frame size of 10 ms, the error propagation can reach up to 40 ms for a single lost frame. In the AMR codec [3GPP TS 26.090], which is based on 20 ms frames, only a first order moving average (MA) prediction is used. At the encoder, the mean removed residual LSF vector $\mathbf{r}(n)$ of frame n is calculated from the absolute LSF vector $\mathbf{q}(n)$ as

$$\mathbf{r}(n) = \mathbf{q}(n) - \overline{\mathbf{q}} - f_p \cdot \mathbf{r}(n-1) \tag{6.3}$$

with a constant mean (expectation) LSF vector $\overline{\mathbf{q}}$ and a fixed vector of prediction factors, f_p, for the coefficients. [5] The decoder then recalculates the absolute LSF coefficients from the received residual coefficients of the current and the previous frame. Therefore, error propagation is limited to one frame only in this case.

The extrapolation/muting based concealment unit of the AMR codec, which is standardized in [3GPP TS 26.091] and recommended for use in the cellular GSM system, extrapolates the LSF values from the last correctly received frame and slightly shifts them towards the mean LSF vector. Instead, if a future frame has already been received, the LSF coefficients of a lost frame may be estimated by linear interpolation, as already investigated in [Fingscheidt and Perez 2002]. Prior to interpolation, the absolute LSF coefficients of the lost frames are estimated as in the standard concealment method. This first estimation then provides the basis for decoding the LSF coefficients of the frame following the lost segment. A superior performance of the linear interpolation compared to the standard concealment for LSF coefficients has been shown in [Fingscheidt and Perez 2002] using the spectral distortion measure. This procedure for estimating lost LSF coefficients is therefore applied for all voicing transitions in the proposed voicing controlled concealment approach.

6.3.2.2 Parameter Estimation: Transition *voiced-voiced*

Within a region of voiced speech the signal exhibits a strong periodicity. Expressing this periodicity, the pitch lag normally follows a fairly smooth curve with only

[5] The 12.2 kbit/s mode of the narrowband AMR codec calculates two LSF coefficient sets per frame, i.e., each set for 10 ms of speech. Here, the prediction factor is chosen as $f_p = 0.65$. The wideband AMR-WB codec calculates one set of coefficients for every 20 ms frame and employs a prediction factor $f_p = 1/3$.

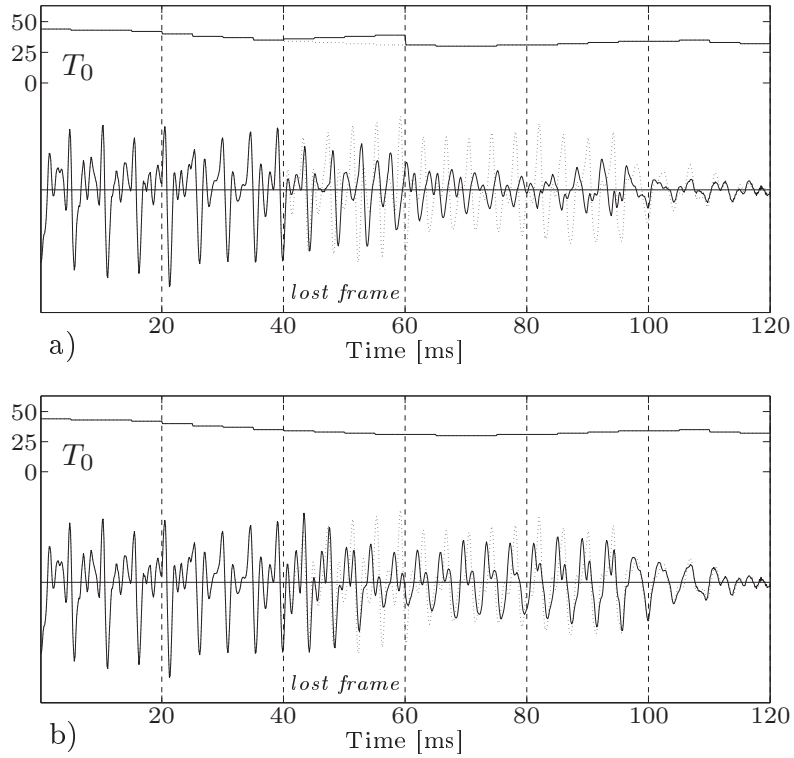


Figure 6.3: Frame loss at voiced-voiced transition: decoded speech signals and respective pitch lags (solid line - signal with concealed frame loss; dotted line - error-free signal) a) standard concealment b) voicing controlled concealment; AMR codec (12.2 kbit/s mode)

small variations from subframe to subframe. In CELP codecs, the pitch lag T_0 is usually calculated every subframe of 5 ms and the parameters T_0 of successive frames are usually quantized independently from each other. Predictive or differential encoding of T_0 is only done within a frame. In the AMR codec, e.g., the pitch lag of the second and fourth subframes are differentially quantized based on the first and third subframes, respectively. Therefore, the missing T_0 values can be linearly interpolated between the fourth subframe of the last received frame and the first subframe of the first received frame behind the loss. In voiced segments, this interpolation leads to a more precise estimation of the original T_0 curve than in case of the conventional error concealment unit, where the missing T_0 values are extrapolated from the last received frame by repeating the value of the fourth subframe. In some codecs, the repeated T_0 values are incremented by 1 for each successively lost subframe in order to introduce some degree of variation and avoid an unnatural periodicity. This repetition/incrementation and the interpolation approach for the pitch lag are shown in the upper curves of Figure 6.3 a) and b).

In conventional error concealment units, the *codebook gains* are both extrapolated from the previous frame and attenuated to prevent possible artifacts, which can be clearly seen in the signal from Figure 6.3. Simulations and auditive comparison have shown that this attenuation of the signal amplitude is not necessary if the missing frames lie within a voiced region of speech, at least not until the

lost segment gets too large. The missing segment can be estimated fairly well with interpolative means, a fluctuation in the signal amplitude is rather perceived as distortion. Therefore, the gain factors of both the adaptive and the fixed codebook are linearly interpolated at voiced-voiced transitions, avoiding an unnecessary fluctuation in the signal amplitude (see Figure 6.3 b)).

For the *innovation vector*, a random entry is chosen from the fixed codebook as in the standard concealment.

Figure 6.3 visualizes the performance of the proposed method (6.3 b)) compared to the standard concealment (6.3 a)) on a single frame loss. With the interpolation approach the signal obviously resembles more closely that of the error-free case than with the standard extrapolative concealment. Simulation results that verify the improvements gained by the proposed concealment method will be presented and discussed in Section 6.3.3.

6.3.2.3 Parameter Estimation: Transition *voiced-unvoiced*

When estimating speech parameters of lost frames at transitions of voiced to unvoiced speech, the *pitch lag* T_0 must not be interpolated to avoid an unnatural change in the fundamental frequency. This would occur when the first pitch lag in the following unvoiced speech frame strongly differs from that of the preceding voiced frame. Better results can be accomplished by extrapolating the pitch lag from the preceding voiced speech frame and thereby achieving a continuation of the periodic signal. The contribution of the *fixed codebook* is set to a random codebook entry as in the conventional concealment unit. Both codebook gains usually experience a considerable change at voiced-unvoiced transitions, the adaptive gain becoming small and the fixed gain increasing. To mitigate possible artifacts at such a transition, which might occur if the energy of the random sequence gets too high, both *codebook gains* are treated as in the standard concealment, i.e., extrapolated and attenuated. This results in a better subjective speech quality, even if in simulations the PESQ MOS-LQO value has been higher when interpolating the codebook gains.

6.3.2.4 Parameter Estimation: Transition *unvoiced-voiced*

The transition from unvoiced to voiced speech is the most difficult position for the concealment of a lost frame. The main reason is that these transitional frames carry the important information on how to build up the periodicity of the beginning voiced segment. In this respect the innovation vector (fixed codebook) is essential, because the long-term prediction cannot contribute substantially yet. Again, a linear interpolation of the *pitch lags* is not advisable because of a possibly large difference between the pitch lag of the beginning voiced frame and the more random values of the preceding unvoiced frame. Also, the repetition of the T_0 value from the preceding unvoiced segment into the starting voiced segment, as it is done in conventional concealment routines, should be avoided to prevent unnatural periodicities based on the almost random pitch lag of unvoiced speech. Therefore, a linear extrapolation technique is employed that estimates the values of the lost frames

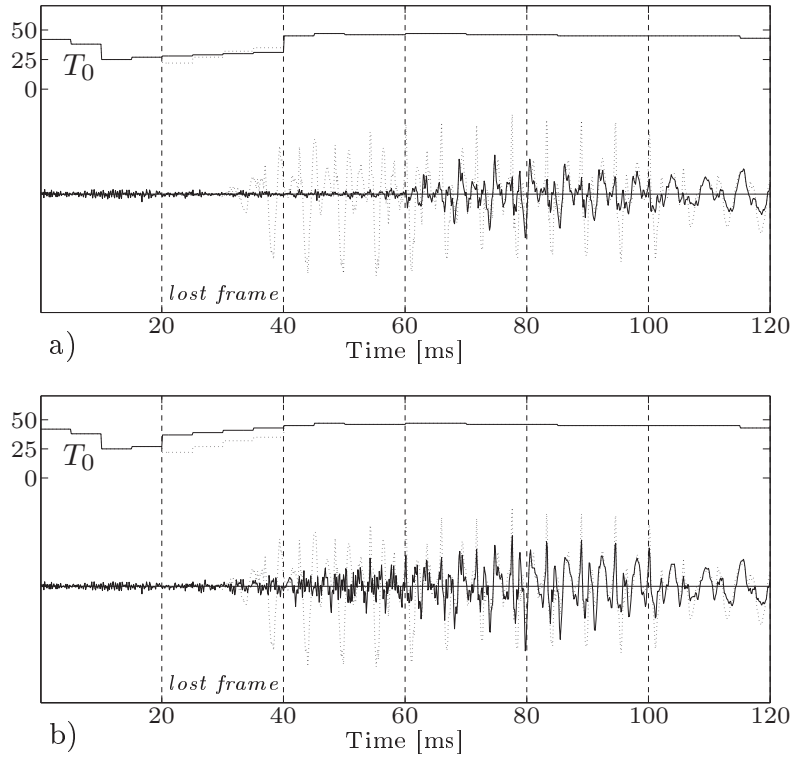


Figure 6.4: Frame loss at unvoiced-voiced transition: decoded speech signals and respective pitch lags (solid line - signal with concealed frame loss; dotted line - error-free signal) a) standard concealment b) voicing controlled concealment with gain interpolation; AMR codec (12.2 kbit/s mode)

from the pitch lags of the following frame. This will help to build up the periodicity of the voiced speech. The contribution of the *fixed codebook* is again set to a random codebook entry. When interpolating the *codebook gains*, the form and periodicity of the voiced signal is reached faster, as can be seen in Figure 6.4 b). However, for the subjective quality it is better to use an extrapolation and attenuation of the gains to mitigate possible artifacts at the transitions.

6.3.2.5 Parameter Estimation: Transition *unvoiced-unvoiced*

The loss of a frame inside an unvoiced speech segment is rather uncritical for the resulting speech quality. The noise-like nature of unvoiced speech allows to estimate a missing frame fairly easily by a spectrally shaped random noise sequence. Because of the missing periodicity in unvoiced speech, the *pitch lag* shows a random behavior. It should be avoided to produce any unnatural periodicity, which might result from repeating a previous pitch lag as done in the standard concealment. Therefore, the T_0 parameter set (4 for each frame) is repeated blockwise from the previous frame to preserve the random behavior. A random fixed codebook entry is chosen for the *innovation vector*, the basis for the noise signal which will be spectrally shaped by the interpolated spectral envelope (LSF coefficients). Both *codebook gains* are linearly interpolated to avoid disturbing amplitude fluctuations.

6.3.3 Performance Results

For an evaluation of the overall performance of the proposed voicing controlled concealment technique several simulations have been run on a speech file of 2 minutes length. Random frame loss has been introduced of about 2%, 3%, and 5%, respectively, occurring at random voicing transitions. The results in Table 6.3 clearly show the superior performance of the proposed technique over the conventional concealment technique. It also performs consistently better than a linear interpolation technique that does not distinguish between the voicing states. The tendencies in the measured PESQ MOS-LQO values have been confirmed by subjective impressions in informal listening tests. However, for transitional segments (v-u and u-v) the subjective quality is slightly better when using extrapolation and attenuation of the codebook gains instead of interpolation in order to mitigate possible artifacts, even if the PESQ MOS-LQO value is smaller.

	frame loss rate		
concealment method	2 %	3 %	5 %
standard concealment	3.538	3.393	3.101
linear interpolation	3.682	3.571	3.392
voicing controlled concealment	3.716	3.623	3.439
voicing controlled concealment & gain muting at u-v and v-u	3.698	3.597	3.360

Table 6.3: PESQ comparison for different concealment methods and channel conditions; AMR codec (12.2kbit/s mode); signal length 2 minutes.

6.4 Improved Packet Loss Concealment by Transmission of Low Rate Side Information

In the previous section, a frame loss concealment scheme has been developed which automatically adopts the estimation of lost codec parameters to the current voicing state of the speech signal. The concept proposed in the current section goes a step further: The optimal concealment strategy for the different codec parameters is already determined at the sender, separately for each frame, and transmitted as low rate side information together with the following frame. This *sender-assisted* approach for robust packet-based speech transmission is therefore an in-between solution between the bit rate intensive sender-driven approaches discussed in Chapter 4 and pure receiver-based approaches like the concept developed in Section 6.3. It is therefore particularly suited for wireless networks with limited transmission bit rates.

The details of this concept, first introduced by the author in [Mertz and Vary 2006] and respective simulation results will be described in the following sections. Finally, Section 6.4.4 will discuss different means of transmitting this side information and a new and elegant approach from a collaborative work [Geiser et al. 2008] will be presented in Section 6.5.

6.4.1 Sender-assisted PLC Approach

The concept of sender-assisted packet loss concealment is to transmit selected side information in succeeding packets in order to assist the receiver's concealment routine in estimating the speech parameters of lost frames. Two *types of side information* are considered in this work:

1. side information on which estimation technique to use for error concealment in the receiver (e.g., extrapolation, interpolation),

2. side information to further improve the estimation (e.g., the quantized estimation error).

The appropriate estimation technique for a parameter depends on the current signal characteristics. While the choice can be based on the voicing state of adjacent frames as proposed in Section 6.3, explicit side information on which particular estimation technique to use for a specific speech frame can further improve the estimation. Due to varying transmission delays, VoIP applications require the use of a receiver buffer (jitter buffer), which facilitates the use of already received succeeding frames in case of packet loss. Therefore, the transmission of side information on previous frames in a later packet does not necessarily require any further additional delay.

Different estimation techniques have been developed and investigated for the parameters of CELP based speech codecs. From simulations with the AMR and AMR Wideband codecs, those methods have been identified for each parameter, which are able to estimate the parameter reliably for different signal characteristics. These sets of estimation techniques are included in the receiver. From each set, the optimal technique for the respective parameter of an individual frame is determined at the sender and transmitted as side information (type 1). Additionally, the sender can calculate the estimation error, i.e., the difference between the original and the estimated parameter in case of loss. This estimation error can be (coarsely) quantized and transmitted as additional information to further improve the concealment (type 2). These types of side information require a significantly lower additional bit rate than other approaches that transmit redundant copies of speech parameters or even complete frames.

6.4.2 Side Information and Concealment Methods for CELP Codec Parameters

The focus of the studies on transmitting side information have been speech codecs based on the CELP principle. The algorithms have been implemented for the AMR and AMR Wideband (AMR-WB) codecs, but can be easily transferred to other CELP based codecs. The following description considers the AMR-WB codec parameters as an example. Simulation results for both the AMR and AMR-WB codecs will be given in Section 6.4.3.

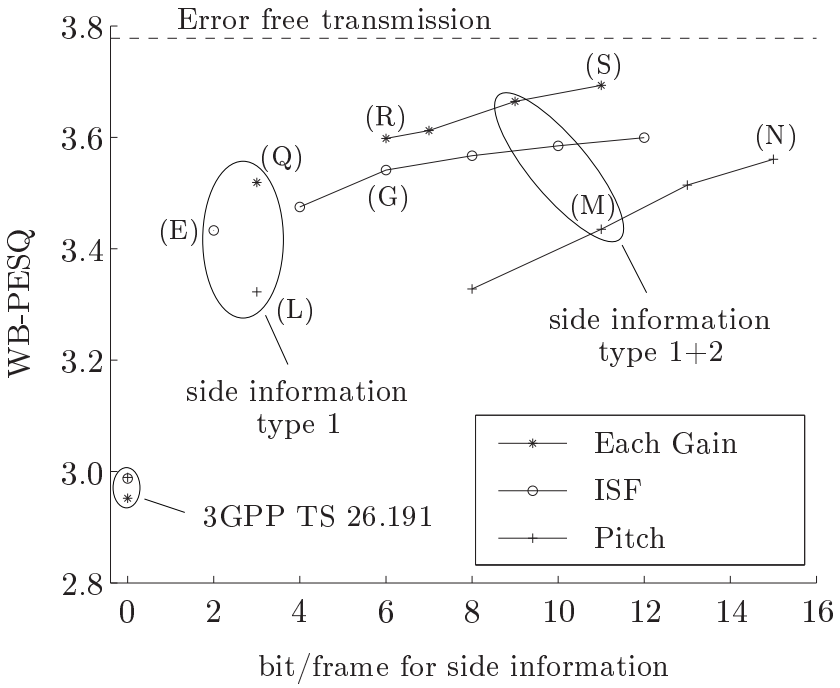


Figure 6.5: WB-PESQ measurements for different sender-assisted concealment approaches at 10 % single frame losses: Each AMR-WB parameter (23.05 kbit/s mode) considered separately, respective other parameters received correctly. Letters in brackets indicate methods from Tables 6.4-6.7 which are explained in detail in the following sections.

For the simulations presented in the following sections, the test files of the TIMIT database [John S. Garofolo 1993] have been used (sampling frequency: 16 kHz). For each speaker, several short files were combined in a single file, resulting in 168 files, each of about 8 to 12 sec length. The files were encoded by the Adaptive Multi-Rate Wideband codec with 23.05 kbit/s. A different set of files from the database has been used for training the vector quantizers of estimation errors, i.e., the difference between original and estimated parameter in case of loss. The vector quantizers have been trained with the LBG algorithm [Linde et al. 1980]. The performance of the different concealment methods will be assessed with the following quality measures, calculated from the estimated speech frames: the parameter SNR (pSNR) for pitch and gain parameters and the mean spectral distance ($\overline{\text{SD}}$) for the spectral envelope. Furthermore, the impact on the resulting speech quality will be discussed with Wideband PESQ [ITU-T Rec. P.862.2 2005] measurements for 10% single and double frame losses in the speech files. In these studies, the different parameters of the AMR-WB speech codec are considered separately and the respective other parameters have been assumed as received correctly in order to focus only on the influence of the considered parameter. The results for single frame losses for the different parameters of the AMR-WB codec have been combined in Figure 6.5 and will be discussed parameter by parameter in the following sections.

6.4.2.1 Spectral Envelope

The AMR-WB codec [3GPP TS 26.190] uses *immittance spectral frequencies* (ISF) as representation for the LP (linear prediction) coefficients to describe the spectral envelope of a speech frame n of 20 ms length. The ISF representation is practically identical to the more often used LSF (line spectral frequencies) representation, except that the last vector coefficient of the ISFs is set to the last filter coefficient itself without transforming it into the LSF domain. For transmission, a first order moving average (MA) prediction of the ISF vector is applied, i.e., the ISF vector at the decoder is calculated from the transmitted residual vectors of the current and previous frame. Hence, the encoder calculates the residual $\mathbf{r}(n)$ from the mean removed ISF vector $\mathbf{q}(n)$ (of dimension 16) according to:

$$\mathbf{r}(n) = \mathbf{q}(n) - \overline{\mathbf{q}} - f_p \cdot \mathbf{r}(n-1) \tag{6.4}$$

with a constant mean (expectation) ISF vector $\overline{\mathbf{q}}$ and a prediction factor $f_p = 1/3$ for the first order moving average (MA) prediction. The receiver recomputes the ISF vector $\mathbf{q}(n)$ of frame n from the transmitted residuals of the current and previous frame, $\mathbf{r}(n)$ and $\mathbf{r}(n-1)$. Therefore, a frame loss always leads to an error propagation of one frame. In the following, $\mathbf{r}$ and $\mathbf{q}$ always denote the (residual) ISF vectors after quantization. Table 6.4 shows the quality that has been achieved by various concealment approaches (**methods A-F**) for single and double frame losses in the simulation setup explained above. In the standard concealment approach (**method A**) [3GPP TS 26.191], the lost ISF vectors are estimated by repeating the previous ISF vector $\mathbf{q}(n-1)$ and shifting it slightly towards the mean ISF vector $\overline{\mathbf{q}}$. However, if the frame following a loss is available, it can be utilized for the concealment. A statistical estimation approach for the spectral envelope parameters has been studied in [Agiomyrgiannakis and Stylianou 2005]. It utilizes Gaussian Mixture Models (GMM) to model the distribution and transmits some side information to assist the receiver in the estimation. However, the spectral envelope considered in [Agiomyrgiannakis and Stylianou 2005] was not predictively coded, and the GMM based estimation is rather complex. Therefore, a sender-assisted interpolation method is developed for the current approach which has a considerable lower computational complexity and is particularly suited to the predictive encoding of LSF/ISF parameters in standard CELP codecs.

Single Frame Losses

Assuming a loss of a single frame of index n, the following interpolation function is proposed to estimate the lost ISF vector:

$$\hat{\mathbf{q}}(n) = \alpha_n \, \mathbf{q}(n-1) + (1-\alpha_n) \, \mathbf{q}(n+1) \tag{6.5}$$

with parameter $\alpha_n \in [0,1]$ which will be transmitted as side information together with frame $n+1$. The interpolation parameter α_n determines the weighting between the previous and the following ISF vector. With this side information, the

concealment routine is informed whether the estimated spectral envelope should be closer to that of the previous or that of the following frame and to which degree. A linear interpolation of the ISF vectors is achieved for $\alpha_n = 0.5$ (**method C**). Note that the following ISF vector $\mathbf{q}(n+1)$ is not explicitly known at the receiver, only the received residual $\mathbf{r}(n+1)$. In [Fingscheidt and Perez 2002] it has therefore been proposed to first extrapolate the ISF as in the standard approach, then calculate $\mathbf{q}(n+1)$, and finally linearly interpolate $\mathbf{q}(n-1)$ and $\mathbf{q}(n+1)$ (**method B**).

Here, an alternative approach is developed which instead results in a closed-form mathematical solution to (6.5). First, $\mathbf{q}(n-1)$ and $\mathbf{q}(n+1)$ are substituted according to (6.4):

$$\hat{\mathbf{q}}(n) = \overline{\mathbf{q}} + \alpha_n \left(f_p \mathbf{r}(n-2) + \mathbf{r}(n-1)\right) + (1-\alpha_n)\left(f_p \mathbf{r}(n) + \mathbf{r}(n+1)\right). \quad (6.6)$$

Next, the unknown residual $\mathbf{r}(n)$ in (6.6) is substituted by its estimation $\hat{\mathbf{r}}(n)$, which is determined from $\hat{\mathbf{q}}(n)$ according to (6.4). Finally, solving for $\hat{\mathbf{q}}(n)$ yields:

$$\hat{\mathbf{q}}(n) = \overline{\mathbf{q}} + a(\alpha_n)\,\mathbf{r}(n-2) + b(\alpha_n)\,\mathbf{r}(n-1) + c(\alpha_n)\,\mathbf{r}(n+1) \quad (6.7)$$

with the factors $a(\alpha_n)$, $b(\alpha_n)$, and $c(\alpha_n)$ determined in dependence on the chosen interpolation factor α_n:

$$a(\alpha_n) = \frac{\alpha_n \cdot f_p}{1 - f_p + \alpha_n \cdot f_p}; \quad b(\alpha_n) = \frac{\alpha_n \cdot f_p^2 - f_p^2 + \alpha_n}{1 - f_p + \alpha_n \cdot f_p}; \quad c(\alpha_n) = \frac{1-\alpha_n}{1 - f_p + \alpha_n \cdot f_p}. \quad (6.8)$$

For the proposed sender-assisted (SA) approach of packet loss concealment, simulations with different sets containing 2, 4, and 8 values of the interpolation factor α_n, i.e., quantizations with 1–3 bit, have been carried out (**methods D, E, F**). For each frame, the optimal α_n has been determined as the one that minimizes the spectral distance SD (in dB) between the spectra belonging to the interpolated vector $\hat{\mathbf{q}}(n)$ and the original vector $\mathbf{q}(n)$, defined in squared form as

$$\mathrm{SD}^2 = \frac{20^2}{2\pi} \int_{-\pi}^{\pi} \left(\log_{10}\left[\frac{1}{|A(e^{j\omega})|}\right] - \log_{10}\left[\frac{1}{|\hat{A}(e^{j\omega})|}\right] \right)^2 d\omega. \quad (6.9)$$

The spectral distance can be calculated with low complexity from the respective cepstral coefficients c_k and $\hat{c}_k$, which in turn can be derived directly from the predictor coefficients a_k [Hagen 1994]:

$$\mathrm{SD}^2 = 2 \cdot 10^2 \cdot (\log_{10} e)^2 \sum_{k=1}^{\infty} [c_k - \hat{c}_k]^2. \quad (6.10)$$

Using 4 values for α_n proves to be the best trade-off regarding quality improvement and additional bit rate (see Table 6.4), i.e., 2 additional bits per frame have to be transmitted (**method E**), leading to a data rate of 100 bit/s. For this case, the

	Concealment Method	$\overline{\text{SD}}$ [dB]	WB-PESQ
single frame losses, 10 % frame loss rate			
A	3GPP TS 26.191 [3GPP TS 26.191]	4.24	2.99
B	Extra-/Interpolation [Fingscheidt and Perez 2002]	3.56	3.26
C	Linear Interpolation: $\alpha_n = 0.5$	3.62	3.30
D	SA, 2 values of α_n: 0.5;0.8	3.23	3.36
E	SA, 4 values of α_n: 0.3;0.5;0.7;0.9	3.03	3.43
F	SA, 8 values of α_n: 0.3;0.4;...;1.0	2.99	3.44
double frame losses, 10 % frame loss rate			
A2	3GPP TS 26.191 [3GPP TS 26.191]	5.25	2.75
B2	Extra-/Interpolation [Fingscheidt and Perez 2002]	4.13	3.16
C2	Linear Interpolation: $\alpha_i = 0.5$	4.14	3.16
E2	SA, 4 values of α_i: 0.3;0.5;0.7;0.9	3.60	3.30

Table 6.4: ISF estimation for the AMR-WB codec (23.05 kbit/s mode): Performance of different concealment approaches for single and double frame losses. A, B, ..., F denote the different approaches for single frame losses described in the text. A2, ..., E2 denote the respective variants for double frame losses.

mean spectral distance $\overline{\text{SD}}$ can be improved by 1.2 dB compared to the standard approach and by about 0.6 dB compared to the linear interpolation **methods B and C**. The noticeable gain in speech quality can be seen from the given Wideband PESQ values for 10% single losses. Note, that for this comparison all other parameters, i.e., pitch lag, gain factors, and fixed codebook entry have been set to their original values, such that only the influence of the spectral envelope is shown. The quality can be further improved by transmitting the quantized estimation error vector $\mathbf{e}_q(n) = \mathbf{q}(n) - \hat{\mathbf{q}}(n)$ (side information type 2). The results are depicted in Figure 6.5 for several bit rates and show a further noticeable improvement for, e.g., 4 additional bits/frame, i.e., a total side information for the ISF of 6 bit/frame, i.e., a data rate of 300 bit/s (marked with **(G)** in Figure 6.5).

Double Frame Losses

If the transmission channel is expected to cause a significant amount of losses of two consecutive speech frames, the proposed sender-assisted concealment approach can be adapted to this case. Assuming a loss of frames n and $n+1$, the following interpolation functions according to (6.5) are used at the receiver:

$$\hat{\mathbf{q}}(n) = \alpha_n \, \mathbf{q}(n-1) + (1-\alpha_n) \, \mathbf{q}(n+1) \tag{6.11a}$$

$$\hat{\mathbf{q}}(n+1) = \alpha_{n+1} \, \mathbf{q}(n) + (1-\alpha_{n+1}) \, \mathbf{q}(n+2) \tag{6.11b}$$

with α_n and α_{n+1} being the side information on the optimal interpolation factors for the individual frames n and $n+1$, which are determined at the transmitter as before, i.e., optimized for single losses. To be able to utilize this information at the

receiver in case of double frame losses, each packet has to transmit two α_i per frame, i.e., the necessary data rate for the side information is increased. In the example of double frame losses, α_n and α_{n+1} are then received together with frame $n+2$. To obtain a closed mathematical solution to (6.11a) and (6.11b), the unknown terms $\mathbf{q}(n+1)$ and $\mathbf{q}(n)$ are substituted by their respective estimations $\hat{\mathbf{q}}(n+1)$ and $\hat{\mathbf{q}}(n)$. Utilizing (6.4) as in the case of single losses, the equations can now be solved for $\hat{\mathbf{q}}(n)$ and $\hat{\mathbf{q}}(n+1)$, finally yielding the following interpolation functions which only depend on the received residual values from before and after the lost frames:

$$\hat{\mathbf{q}}(n) = \overline{\mathbf{q}} + a_1(\alpha_n, \alpha_{n+1})\,\mathbf{r}(n-2) + b_1(\alpha_n, \alpha_{n+1})\,\mathbf{r}(n-1) + c_1(\alpha_n, \alpha_{n+1})\,\mathbf{r}(n+2) \quad (6.12a)$$

$$\hat{\mathbf{q}}(n+1) = \overline{\mathbf{q}} + a_2(\alpha_n, \alpha_{n+1})\,\mathbf{r}(n-2) + b_2(\alpha_n, \alpha_{n+1})\,\mathbf{r}(n-1) + c_2(\alpha_n, \alpha_{n+1})\,\mathbf{r}(n+2) \quad (6.12b)$$

with

$$\begin{aligned}
a_1(\alpha_n, \alpha_{n+1}) &= (2\,\alpha_n + \alpha_n\,\alpha_{n+1}) \,/\, d \\
b_1(\alpha_n, \alpha_{n+1}) &= (17\,\alpha_n - \alpha_{n+1} + 1 + 10\,\alpha_n\,\alpha_{n+1}) \,/\, (3\,d) \\
c_1(\alpha_n, \alpha_{n+1}) &= 9\,(1-\alpha_n)\,(1-\alpha_{n+1}) \,/\, d \\
a_2(\alpha_n, \alpha_{n+1}) &= (10\,\alpha_n\,\alpha_{n+1} - \alpha_n) \,/\, (3\,d) \\
b_2(\alpha_n, \alpha_{n+1}) &= (30\,\alpha_n\,\alpha_{n+1} - \alpha_{n+1} + 1 - 3\,\alpha_n) \,/\, (3\,d) \\
c_2(\alpha_n, \alpha_{n+1}) &= 9\,(1-\alpha_{n+1}) \,/\, d
\end{aligned} \quad (6.13)$$

and

$$d = 7 - 7\,\alpha_{n+1} - \alpha_n + 10\,\alpha_n\,\alpha_{n+1}. \quad (6.14)$$

The simulation results for 10 % double frame losses in Table 6.4 (**methods A2-E2**) show that the use of side information that has been determined at the sender assuming single frame losses (**method E2**) still yields a considerable improvement over [3GPP TS 26.191] (**method A2**). If available, the quantized estimation errors $\mathbf{e}_q(n)$ and $\mathbf{e}_q(n+1)$ can also be utilized for double frame losses to improve the result further, as can be seen from the declining curves of the mean spectral distance between replaced and original (quantized) spectral envelope in Figure 6.6. Note that for double losses the data rate of the side information is increased because for each frame the side information on the two preceding frames has to be transmitted.

A larger number of consecutively lost frames should not be concealed with interpolative methods, but rather using the standard concealment based on extrapolation and muting.

6.4.2.2 Pitch Lag

The pitch lag parameter T_0 describes the long term prediction in CELP based speech codecs and is an indication for the periodicity of the current speech frame. In the AMR and AMR-WB codecs a pitch lag is determined for every subframe, resulting in a set of four T_0 values per frame.

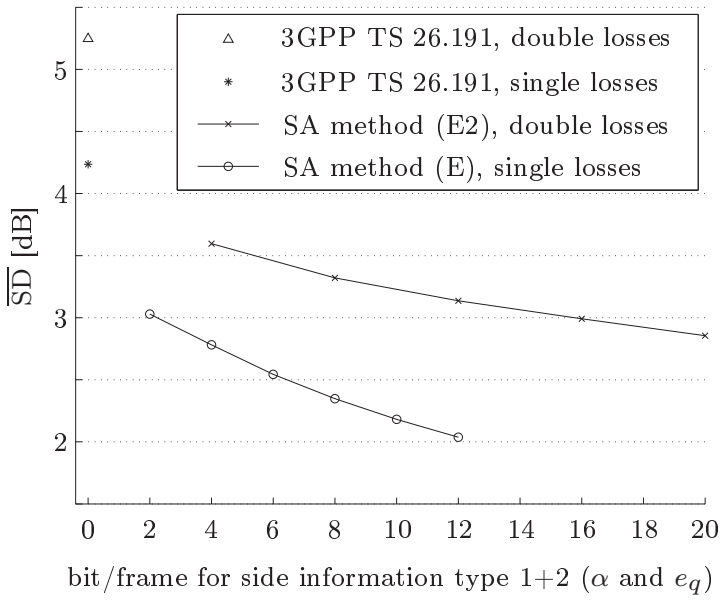


Figure 6.6: ISF estimation for the AMR-WB codec (23.05 kbit/s mode): Mean spectral distance $\overline{\mathrm{SD}}$ between estimation and original quantized spectral envelope for method E and different amount of additional side information for quantizing the estimation error at 10 % single or double frame losses, respectively. Side information derived for each frame: 2 bit for α_n and 2, 4, ..., 10 bit to quantize $e_q(n)$. Method E: Side information is transmitted for previous frame; Method E2: Side information is transmitted for previous two frames.

Single Frame Losses

In the standard concealment method of the AMR-WB codec [3GPP TS 26.191], the pitch lags of a lost speech frame are estimated in dependence on the lag and gain histories, which consist of the preceding five sub-frame values. If the minimum gain in this history is sufficiently large and the difference between maximum and minimum pitch lag is sufficiently small, which indicates a voiced speech segment, the previous pitch lag is just repeated. The lag is also repeated if only the last two gains are sufficiently large, which indicates the beginning of a voiced segment. In any other case, a lag is chosen which is lying in the range of the history, weighted towards bigger lags, and a random variation is added. In Section 6.3, a signal dependent approach has been proposed which automatically decides between interpolation and extrapolation techniques for the lost pitch parameters based on the voicing state of the signal. For the sender-assisted (SA) approach, some side information will be transmitted which assists the receiver's concealment routine in choosing the optimal estimation strategy for a specific frame. The following estimation techniques are considered for the pitch lags $T_0^{(i)}(n)$ of the sub-frames $i \in \{1, \ldots, 4\}$ of a lost frame n:

a) standard concealment according to [3GPP TS 26.191]

b) linear interpolation according to

$$\hat{T}_0^{(i)}(n) = \frac{5-i}{5} T_0^{(4)}(n-1) + \frac{i}{5} T_0^{(1)}(n+1) \;;\; i = 1, \ldots, 4 \tag{6.15}$$

c) a replacement approach, where the first N sub-frames will be estimated by the preceding value and the remaining sub-frames by the first pitch value of the following frame:

$$\hat{T}_0^{(i)}(n) = \begin{cases} T_0^{(4)}(n-1) & \text{for } i = 1, \ldots, N \\ T_0^{(1)}(n+1) & \text{for } i = N+1, \ldots, 4. \end{cases} \tag{6.16}$$

The choice of $N \in [0, 4]$ is included in the side information.

The estimation technique that produces the smallest mean-square error for the pitch lags of a frame is chosen and its index is transmitted to the receiver as side information in a succeeding packet. The 7 different methods, i.e., standard concealment, linear interpolation, and the replacement approach with 5 variants of N, can be described with an index of 3 bit. As in the standard concealment, only integer pitch lags will be considered, the fractional part of the pitch values will therefore be set to zero.

The parameter SNR and WB-PESQ values resulting from the different concealment approaches (**methods I-L**) for the pitch lags are listed in Table 6.5. Method L uses all of the above mentioned estimation techniques and therefore requires 3 additional bits per frame. It leads to a considerable improvement over the standard concealment (**method I**) [3GPP TS 26.191].

In voiced segments, where a correct pitch lag is crucial, the estimation using the method indicated by the side information is often correct, sometimes only differing by a small value. Further improvement can be achieved by transmitting a correction $\hat{e}_{T_0} \in \{-3, -2, -1, 0, 1, 2, 3, 4\}$ for the pitch lag of each subframe, encoded with 3 bit each. If the correct value cannot be achieved by the adjustment, no correction is used ($\hat{e}_{T_0} = 0$). The quantized estimation error can be transmitted with an additional 12 bit/frame (i.e., 600 bit/s), requiring a total of 750 bit/s for the pitch lag side information.

Double Frame Losses

As for the ISF parameter, the side information on the estimation method for the pitch lags has to be sent redundantly if it shall be utilized in case of losing two consecutive frames. The results in Table 6.5 show only a small improvement over the standard approach (**method I2**) when utilizing the side information which has been determined at the sender assuming single frame losses (**method L2**). An improvement in case of double frame losses would therefore require more specific side information in addition to that for single losses, which would increase the total data rate.

	Concealment Method	pSNR [dB]	WB-PESQ
single frame losses, 10 % frame loss rate			
I	3GPP TS 26.191 [3GPP TS 26.191]	6.50	2.99
J	Linear Interpolation	7.08	3.06
K	SA, 2 techniques (1 bit): I, J	8.52	3.14
L	SA, 7 techniques (3 bit): I, J, replacement ($N \in [0, 4]$)	9.95	3.32
double frame losses, 10 % frame loss rate			
I2	3GPP TS 26.191 [3GPP TS 26.191]	6.36	2.91
J2	Linear Interpolation	6.97	2.99
L2	SA, 7 techniques (3 bit): I, J, replacement ($N \in [0, 4]$)	7.88	3.13

Table 6.5: Pitch estimation for the AMR-WB codec (23.05 kbit/s mode): Performance of different concealment approaches for single and double frame losses: standard concealment, linear interpolation, and sender-assisted (SA) approach. I, J, K, and L denote the approaches for single frame losses as described in the text. I2, J2, and L2 denote the respective variants for double frame losses.

6.4.2.3 Adaptive and Fixed Codebook Gains

The adaptive and fixed codebook gains of CELP based codecs determine the contribution of long term prediction (adaptive codebook) and innovation (fixed codebook) to the excitation signal before synthesis filtering. In the AMR-WB codec, the gain of the fixed codebook g_c is transmitted in form of a correction factor $\gamma = g_c / g_c'$ between the fixed codebook gain g_c and its prediction g_c' which depends on the previous sub-frames and the energy of the chosen fixed codebook entry. It is jointly vector quantized with the gain of the adaptive codebook g_a.

Single Frame Losses

The achievable quality of different concealment methods (**methods O-Q**) is shown in Table 6.6. In the standard concealment method (**method O**), the gains of adaptive and fixed codebook are estimated by attenuated values from the previous sub-frames. The more consecutive frames are lost, the more attenuation is used, until the signal is completely muted after 6 frames. While signal muting is necessary in cases of several consecutive frame losses, for short losses of 1-2 frames this attenuation leads to noticeable and unnecessary amplitude fluctuations in voiced speech segments. Therefore, a voicing dependent concealment has been proposed in Section 6.3, which interpolates the gains in case of voiced-voiced and unvoiced-unvoiced transitions, and only attenuates them in transient cases, i.e., voiced-unvoiced or unvoiced-voiced transitions. For the sender-assisted (SA) concealment approach (**method P**), several extrapolation and interpolation techniques are considered and the optimal choice for a specific frame is transmitted as side information in

the following packet. The following estimation techniques for the adaptive codebook gains $g_a^{(i)}(n)$ of the $i = 1, \ldots, 4$ sub-frames of a lost frame n have shown to be sufficient in covering most variations:

a) linear interpolation according to

$$\hat{g}_a^{(i)}(n) = \frac{5-i}{5} g_a^{(4)}(n-1) + \frac{i}{5} g_a^{(1)}(n+1) \;;\; i = 1, \ldots, 4 \tag{6.17}$$

b) sub-frame replacement by previous or following gains:

$$\hat{g}_a^{(i)}(n) = \begin{cases} g_a^{(4)}(n-1) & \text{for } i = 1, \ldots, N_1 \\ g_a^{(1)}(n+1) & \text{for } i = N_1+1, \ldots, 4 \end{cases} \tag{6.18}$$

with $N_1 \in [0, 4]$

c) replacement by the mean gains of the previous or following frame:

$$\hat{g}_a^{(i)}(n) = \begin{cases} \frac{1}{4}\sum_{j=1}^{4} g_a^{(j)}(n-1) & \text{for } i = 1, \ldots, N_2 \\ \frac{1}{4}\sum_{j=1}^{4} g_a^{(j)}(n+1) & \text{for } i = N_2+1, \ldots, 4 \end{cases} \tag{6.19}$$

with $N_2 \in [0, 4]$

Even with a restriction of N_1 to 2 values $\{0, 4\}$ (**method Q**), already a considerable improvement over [3GPP TS 26.191] (**method O**) is achieved. The decision for the optimal estimation technique for the gain of the adaptive codebook is therefore transmitted with 3 bit/frame. Further improvement requires at least 3 more bits/frame to transmit the quantized estimation error vector for the 4 gains of a frame (Figure 6.5, R).

In order to prevent possible artifacts due to the estimated fixed codebook contribution, the fixed codebook gain is estimated as in the standard method, i.e., attenuated, and no side information is transmitted for this parameter.

Double Frame Losses

If transmitted redundantly, the side information can still be utilized in case of double frame losses with only slight modifications at the receiver. The results shown in Table 6.6 still indicate a considerable improvement (**method Q2**) compared to the standard concealment approach (**method O2**) in case of double frame losses. For longer loss lengths, however, the gain factor should be attenuated as in the standard concealment in order to mitigate possible artifacts.

Concealment Method		pSNR [dB]		WB-PESQ
		g_a	γ	
single frame losses, 10 % frame loss rate				
O	3GPP TS 26.191 [3GPP TS 26.191]	N/A	N/A	2.95
P	SA, 11 techniques (4 bit)	11.49	4.67	3.53
Q	SA, 8 techniques (3 bit)	11.27	4.57	3.52
double frame losses, 10 % frame loss rate				
O2	3GPP TS 26.191 [3GPP TS 26.191]	N/A	N/A	2.48
Q2	SA, 8 techniques (3 bit)	8.91	3.63	3.19

Table 6.6: Gain estimation for the AMR-WB codec (23.05 kbit/s mode): Performance of different concealment approaches for single and double frame losses: standard concealment and sender-assisted (SA) approach. O, P, and Q denote the approaches for single frame losses as described in the text. O2 and Q2 denote the respective variants for double frame losses.

6.4.2.4 Fixed Codebook (Innovation)

For the proposed sender-assisted concealment, the contribution of the fixed codebook, also called innovation sequence, is treated as in the standard AMR-WB concealment approach [3GPP TS 26.191], which estimates a missing code vector by a random sequence. Although this vector is usually not the correct one, it still serves well as noise signal for synthesizing unvoiced speech segments. However, the fixed codebook also contributes significantly at transitions and voiced onsets. In order to prevent artifacts due to an incorrect fixed codebook contribution, the fixed codebook gain is attenuated as described above.

Since the innovation vector is a random-like signal, significant improvements could only be achieved by transmitting redundant copies of the complete innovation in following packets. However, the required data rate can be limited to some extent by only transmitting information on important frames, e.g., frames at unvoiced-voiced transitions, as proposed in [Tosun and Kabal 2005].

6.4.3 Performance Results

The approaches for the different codec parameters have been finally combined to evaluate the overall quality improvement by the proposed sender-assisted packet loss concealment approach. Table 6.7 shows the Wideband PESQ [ITU-T Rec. P.862.2 2005] results for different bit rates of additional side information, considering 10% single or double frame losses. In comparison to the standard approach [3GPP TS 26.191], a considerable quality improvement is already achieved by transmitting which estimation techniques to use for concealment (11 bit/frame, i.e. 550 bit/s). The quality can be further increased by transmitting quantized estimation errors of the different parameters.

Results for the narrowband AMR codec and a comparison to the voicing controlled and solely receiver based approach from Section 6.3 are shown in Table 6.8.

Concealment Method (see Table 6.4-6.6, Figure 6.5)	WB-PESQ
single frame losses, 10 % frame loss rate	
3GPP TS 26.191 [3GPP TS 26.191]	2.174
SA, 11 bit/frame (E, L, Q)	2.529
SA, 29 bit/frame (G, M, R)	2.609
SA, 47 bit/frame (H, N, S)	2.677
double frame losses, 10 % frame loss rate	
3GPP TS 26.191 [3GPP TS 26.191]	1.927
SA, 2·11 bit/frame (E2, L2, Q2)	2.277

Table 6.7: Quality of sender-assisted (SA) estimation approach for the AMR-WB codec (23.05 kbit/s mode), combining ISF, pitch, and gain parameter estimation for single and double frame losses. Letters in brackets, e.g., (E,L,Q), refer to points in Figure 6.5.

All simulations have been using the 12.2 kbit/s mode of the AMR codec and the mean PESQ MOS-LQO values have been derived as quality measure.

The columns of the table show different simulation scenarios. Each scenario is characterized by a certain frame loss rate and possibly a restriction of these losses to a certain voicing transition:

Scenario A Frame loss rate $P_{\text{fl}} = 9.8\%$; the frame losses are completely random and therefore also occur at random voicing transitions.

Scenario B Frame loss rate $P_{\text{fl}} = 10.0\%$; the frame losses occur only at unvoiced to unvoiced (u-u) voicing transitions of the speech signal

Scenario C Frame loss rate $P_{\text{fl}} = 3.2\%$; the frame losses occur only at unvoiced to unvoiced (u-u) voicing transitions of the speech signal

Scenario D Frame loss rate $P_{\text{fl}} = 3.0\%$; the frame losses occur only at unvoiced to unvoiced (u-u) voicing transitions of the speech signal

Scenario E Frame loss rate $P_{\text{fl}} = 8.6\%$; the frame losses occur only at unvoiced to unvoiced (u-u) voicing transitions of the speech signal

The rows of the table show for each concealment method the results for a particular codec parameter, i.e., the given codec parameter (LSF, T_0, gains, fixed CB entry) is estimated by the respective concealment method, the other codec parameters are kept as correct/original. The rows with "all parameters" show the result for estimating all parameters of the lost frames, i.e., these rows reflect the final concealment quality of the respective method.

From the results of the three concealment methods for the general behavior (Scenario A, "all parameters"), a clear improvement from the standard extrapolative concealment (MOS-LQO=2.52), to the voicing controlled approach (MOS-LQO=2.74) and further to the sender-assisted concealment (MOS-LQO=2.95) can be observed

Estimated Parameter	PESQ MOS-LQO for different scenarios A-E				
	A: $P_{\text{fl}}=9.8\%$ random voicing transitions	**B:** $P_{\text{fl}}=10\%$ only u-u frames lost	**C:** $P_{\text{fl}}=3.2\%$ only u-v frames lost	**D:** $P_{\text{fl}}=3.0\%$ only v-u frames lost	**E:** $P_{\text{fl}}=8.6\%$ only v-v frames lost
Standard concealment [3GPP TS 26.091]					
LSF (spectral env.)	3.47	3.73	3.63	3.84	3.52
T_0 (pitch lags)	3.52	3.95	3.69	3.95	3.35
g_a, g_c (CB gains)	3.10	3.27	3.37	3.70	3.29
fixed CB entry	3.51	3.88	3.72	3.89	3.39
all parameters	**2.52**	**3.04**	**2.96**	**3.44**	**2.61**
Voicing dependent concealment from Section 6.3					
LSF (spectral env.)	3.55	3.75	3.68	3.86	3.64
T_0 (pitch lags)	3.64	3.94	3.78	3.93	3.50
g_a, g_c (CB gains)	3.44	3.81	3.37	3.70	3.78
all parameters	**2.74**	**3.39**	**3.00**	**3.47**	**2.84**
Concealment with side information from Section 6.4, 11 bit/frame					
LSF (spectral env.)	3.67	3.80	3.74	3.92	3.72
T_0 (pitch lags)	3.77	3.96	3.86	3.96	3.68
g_a, g_c (CB gains)	3.60	3.64	3.74	3.89	3.73
all parameters	**2.95**	**3.37**	**3.28**	**3.65**	**3.00**

Table 6.8: Comparison of concealment approaches for the AMR codec, 12.2 kbit/s mode: Mean MOS-LQO for different simulation scenarios (columns), i.e., different frame loss rates and restriction of losses to certain voicing transitions. The given codec parameter (LSF, T_0, gains, fixed CB entry) is estimated by the respective concealment method, the other codec parameters are kept as correct/original; "all parameters" indicates that all parameters are estimated, i.e., these rows reflect the final concealment quality of the respective method. The variation of voicing transition and estimated codec parameter facilitates an evaluation of the concealment quality for the respective combinations. The PESQ MOS-LQO value for error free transmission (i.e. only reflecting the codec distortion) resulted to 4.03.

which proves the validity of the proposed approaches. The improvement in PESQ values also correspond to the subjective listening impression which confirms the relevant improvement of the signal quality.

The variation of voicing transition and estimated codec parameter further facilitates an evaluation of the concealment quality for each possible combination. The receiver based voicing dependent concealment approach developed in Section 6.3 shows a significant improvement at unvoiced-unvoiced (u-u, Scenario B) and at voiced-voiced (v-v, Scenario E) transitions. The improvement at u-u transitions is mainly due to a better gain estimation and slight improvement of the LSF estimation. For the improvement at v-v transitions the improved estimation of the LSF coefficients, pitch lags, and gains is responsible. The sender-assisted approach with transmission of low-rate side information, which has been developed in Section 6.4, shows a further improvement of the quality at v-v transitions and also a significant improvement at unvoiced-voiced (u-v, Scenario C) and voiced-unvoiced (v-u, Scenario D) transitions at which the voicing dependent approach is only slightly better than the standard concealment. While the improvement at u-v transitions is mainly due to a better estimation of the pitch lags, the improvement at v-u transitions results from a better gain estimation. At v-v transitions, the improvement of the sender-assisted approach is due to a further improved estimation of LSF, pitch lag, and gain parameters.

For short loss lengths of one or two consecutive frames, the developed approaches are able to considerably improve the frame loss concealment for CELP based speech codecs compared to conventional implementations. In case of longer loss lengths, i.e., more than two consecutive frames, the concealment should revert to conventional approaches, i.e., an extrapolation of the signal with subsequent attenuation of the signal amplitude until final muting of the signal after about 120 ms. If such loss lengths are expected to occur frequently, further means of error protection, e.g., a redundant transmission of complete frames should be considered (cf. Chapter 4).

6.4.4 Approaches for Side Information Transmission

There are several possibilities to transmit a side information bitstream in a packet transmission scenario:

a) as additional bits piggybacked to the packets containing the original encoded frames,

b) as separate packet stream, or

c) as a hidden bitstream within the bits of the encoded parameters.

Alternative a), the piggybacked transmission of the side information within the same packet, requires an appropriate RTP payload format definition. All receivers will need to be able to decompose this payload format, even if they might not have implemented the utilization of the side information itself.

The transmission of the side information as separate packet stream, alternative b), only requires a payload format definition for these specific side information packages, the original packet stream is not affected. This approach is therefore backwards compatible, because receivers which do not support this enhancement can just ignore these packets. The drawback of this approach is a significant increase in data rate because of the additionally required packet headers. This contradicts the original intentions of transmitting low rate side information for applications using wireless transmission links with limited capacity.

The most elegant way of transmitting the side information is alternative c). In a collaborative work [Geiser et al. 2008] we proposed to use a watermarking technique presented in [Geiser and Vary 2008] to transmit the side information hidden within the original encoded bit stream. This approach avoids the disadvantages of alternatives a) and b). It requires no additional data rate and is fully backwards compatible. Receivers which do not support the utilization of the side information would decode the signal without detecting — and therefore ignoring — the hidden information. This approach leads to only minor and unnoticeable degradations in the original speech signal as shown in [Geiser and Vary 2008]. Details of an implementation of this concept as well as simulation results will be discussed in Section 6.5.

6.5 Steganographic Transmission of Side Information for Packet Loss Concealment

The transmission of suitable side information for assisting the receiver's packet loss concealment routine can increase the error resilience of Voice over IP applications and other packet-based multimedia transmission systems, as discussed in Section 6.4. Different techniques for the transmission of such side information have been discussed in Section 6.4.4. Which technique to employ in an actual system depends on several factors. Besides technical constraints of codecs and protocols, the choice will be mainly determined by the considered network scenario, i.e., which transmission links are involved, how heterogeneous the group of receivers is, etc. The present section presents an approach to employ a steganographic technique for hiding the side information bit stream in the original bitstream of an ACELP encoded speech signal. This approach from a collaborative work has been introduced in [Geiser et al. 2008]. It provides a solution which requires no additional bit rate and is backwards compatible to legacy systems.

6.5.1 System Concept

The concept of transmitting side information via a *steganographic channel* within the bitstream of the employed speech codec provides several advantages over the conventional (separate) transmission of side information:

1. No additional bit rate is required, the side information bitstream is completely hidden within the codec bitstream.
2. Standard payload formats can be used for the media packets, no particular payload format is required for piggybacking the side information.
3. The system is fully backwards compatible, i.e., the side information bitstream remains undetected by "legacy" terminals and is therefore ignored.
4. With this technique, the side information can also be transmitted over conventional circuit-switched parts of the transmission chain, e.g. GSM or UMTS. As prerequisite it has to be ensured that no transcoding or tandeming operations occur during transmission. Such a decoding and and re-encoding of the speech signal would destroy the side information that is "hidden" in the bitstream.

The steganographic transmission of side information for packet loss concealment is particularly useful in transmission scenarios in heterogeneous networks, e.g., for a call from an IP phone to a GSM cellphone. Figure 6.7 depicts an overall transmission chain in such a heterogeneous network. The transmission is degraded by *both* packet losses in the packet network and residual bit errors in the circuit-switched cellular network. With the proposed approach the side information can be transmitted from end-to-end through different network types. The receiver is then able to utilize the information for the concealment of frames which have either been lost through packet loss or discarded in circuit-switched cellular networks because of residual bit errors.

In the following sections, an exemplary implementation of this technique for the AMR codec at a bit rate of 12.2 kbit/s is described. The respective data hiding technique for the AMR codec, which is shortly reviewed in Section 6.5.2, provides the transmission of a steganographic bitstream of 2 kbit/s. For the application on channels which may leave residual bit errors in parts of the codec bitstream, the hidden bitstream needs to be protected against such bit errors which limits the maximum available rate that is left for the side information itself. A suitable channel coding scheme for the side information will be introduced in Section 6.5.3. Finally, the entire system is evaluated for different rates of side information which has been derived according to Section 6.4. The simulation results are presented and discussed in Section 6.5.4.

6.5.2 Data Hiding Scheme for ACELP Codecs

For the steganographic transmission of the side information, the ACELP (algebraic CELP) data hiding mechanism from [Geiser and Vary 2008] is employed, which allows to hide steganographic data with 2 kbit/s, i.e., 40 bit/frame in the bitstream of the 12.2 kbit/s mode of the AMR codec [3GPP TS 26.090].

In order to maintain a high quality of the decoded speech, the steganographic bits are embedded in less important parts of the encoded bitstream, i.e., in the

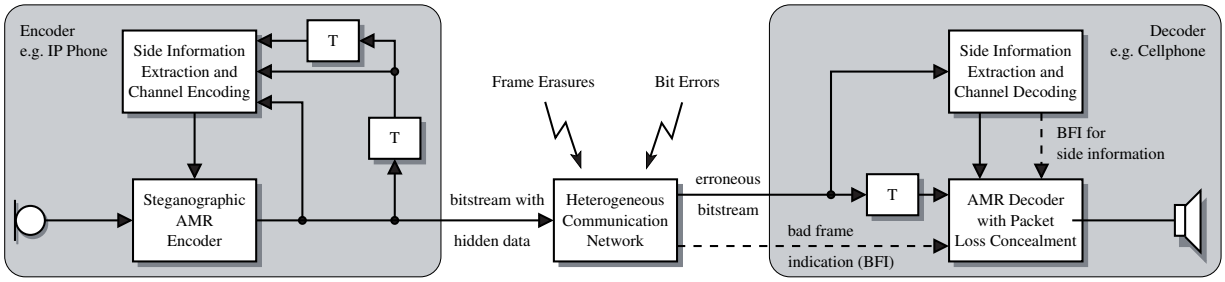


Figure 6.7: System concept: Transmission of side information for packet loss concealment as hidden data within the bitstream of a speech codec (e.g. AMR). Transmission channel: heterogeneous communication network with frames erasures (e.g. due to packet losses or bit errors) and residual bit errors in the codec bits. Exemplary application scenario: voice call from IP phone to GSM cellphone.

fixed codebook (FCB) contribution of the AMR codec. The impact of the hidden bits on the speech quality is minimized by a *joint* implementation of the speech encoding and data hiding operations, cf. [Vary and Geiser 2007]. The key to this "ACELP steganography" is a modified search strategy for the ACELP codebook. The "message" m that shall be embedded into a 5 ms subframe is given as a 10 bit binary sequence, which is further split into five sub-messages with two bits each. To enable the transmission of $N = 10$ steganographic bits, the ACELP codebook (or fixed codebook, FCB) is partitioned into $M = 2^{10}$ *sub-codebooks* that uniquely identify the selected message m. The search for the optimal codebook entry for the subframe is then restricted to the sub-codebook determined by the current message, i.e., 10 bits of side information. Each of the sub-codebooks has a comparable size to the part that is searched in the original codebook. With 4 subframes per frame of 20 ms, a total 40 bit of side information can be transmitted in each frame. A more detailed explanation of this steganographic FCB technique can be found in [Geiser and Vary 2008].

6.5.2.1 Impact of Data Hiding on Speech Quality

It appears plausible that the embedding of 2 kbit/s into a 12.2 kbit/s bitstream ($\approx 16\%$ of the codec rate) should have some impact on the quality of the coded speech. And indeed, an objective analysis with the PESQ speech quality measurement tool [ITU-T Rec. P.862 2001] seems to confirm that there is a certain quality impairment (see results below). Yet, subjective listening tests do not indicate a clear preference of the listeners. In [Geiser and Vary 2008], an ABX experiment has been conducted with 11 experienced listeners (quiet environment, diotic presentation, multiple playback allowed, four trials per sample). The options A and B have been randomly assigned to "standard AMR speech" and "speech with 2 kbit/s of hidden data". Only 162 out of 264 votings, i.e., 61%, correctly identified X as either A or B, indicating that the impairment due to the data hiding process is almost unnoticeable.

6.5.3 Impact of Bit Errors and Channel Coding

If the (modified) AMR bitstream is transmitted over a network which tolerates bit errors in less important bits of the codec bitstream (unequal error protection), e.g., a circuit-switched GSM network or a wireless packet network employing UDP-Lite (see, e.g., [Mertz et al. 2005]), such residual bit errors will inevitably lead to errors in the extracted steganographic bitstream at the receiver. Moreover, the side information bitstream exhibits an increased sensitivity to bit errors, resulting from the combination of two AMR bits in the data hiding procedure from [Geiser and Vary 2008]. Furthermore, the side information has a strong impact on the quality of the speech signal when applied for frame loss concealment. Errors in the side information may lead to clearly audible distortions when, e.g., wrong concealment techniques are applied. Hence, a dedicated error protection for the side information bits is required.

Depending on the data rate of the side information and how many steganographic bits per frame are remaining, the error protection can either be a simple cyclic redundancy check (CRC) for error detection, or a more sophisticated channel code for error correction together with an additional CRC for detecting residual errors. The implemented channel coding scheme follows this two-step approach. The main component is a suitably shortened BCH block code to protect the side information bits in each speech frame. Furthermore, before BCH encoding, a CRC is added to the side information bits to detect residual errors which could not be corrected or detected by the BCH decoder. Suitable (n, k)-BCH codes with codeword length $n = 40$ exist for an information length of $k \in \{1, 7, 13, 16, 22, 28, 34\}$ [Lin and Costello 2004]. For example, in case of a side information length of $l_{\mathrm{SI}} = 8\,\mathrm{bit}$, the $(40, 13)$-code would allow to allocate $k - l_{\mathrm{SI}} = (13 - 8)\,\mathrm{bit} = 5\,\mathrm{bit}$ to the CRC. For $l_{\mathrm{SI}} = 26$, the $(40, 28)$-code offers the best error correction capabilities, but there is only room for 2 CRC bits.

The receiver extracts 40 bit of encoded side information per frame from the received AMR bitstream. Then, the BCH decoder, using the Berlekamp-Massey algorithm [Berlekamp 1968; Massey 1969; Henkel 1989], attempts to correct any bit errors in the side information bitstream. In case the Berlekamp algorithm fails to decode the message, the flag $\mathrm{BFI}_{\mathrm{BCH}}$ is set. After BCH decoding, the CRC is recomputed and compared with the received CRC bits. The flag $\mathrm{BFI}_{\mathrm{CRC}}$ is set if the CRC fails. In addition to the decoded side information bits, the decoder outputs a bad frame indicator (BFI) which is related to the side information bits. To minimize the "false acceptance rate", i.e., side information which is considered to be error free in spite of residual errors, this BFI is computed by a logical OR of both flags $\mathrm{BFI}_{\mathrm{BCH}}$ and $\mathrm{BFI}_{\mathrm{CRC}}$.

6.5.4 Steganographic PLC in a Packet Network with Circuit-Switched GSM Access

The proposed transmission of side information for packet loss concealment using a steganographic channel within the codec bitstream will be evaluated by the following

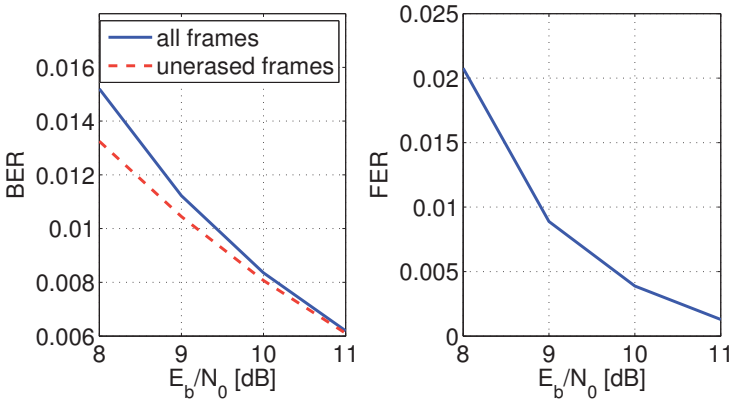


Figure 6.8: GSM bit level simulation: Residual bit error rate (BER) — in all frames and in the unerased frames only — and frame erasure rate (FER) on channels with different E_b/N_0 ratios.

simulations. Two scenarios are considered:

a) The speech signal is transmitted from a Voice over IP (VoIP) terminal over a packet-based network with a certain packet loss rate — wireline (e.g. DSL, LAN) or wireless (e.g. WLAN, UMTS) — to another VoIP terminal.

b) The speech signal is transmitted from the VoIP terminal over a packet network to a gateway where the encoded speech frames are extracted from the IP packets. In a transcoding-free operation, i.e., without de- and re-encoding, the encoded frames are then handed to a cellular GSM network for transmission over a circuit-switched channel to a mobile terminal.

6.5.4.1 GSM Bit Level Simulations

For scenario b), the GSM channel has been simulated with a bit level reference implementation in the Synopsys System Studio Software [Synopsys 2007], assuming a transmission frequency of 900 MHz and a user terminal moving with 50 km/h. The "typical urban" channel model has been chosen with an additional AWGN impairment with different E_b/N_0 ratios. In GSM, the speech codec bits are sorted into classes of sensitivity before channel coding and transmission. A concept of *unequal error protection* is applied in the channel coding process, applying a convolutional channel code to the more important bits and leaving the less important bits unprotected. The class of most important bits is further protected by a CRC for error detection. If residual bit errors are detected in the most important bits, the frame is marked unusable by setting a BFI (bad frame indication) flag. The usable frames might therefore still contain residual bit errors in the lesser important bits, i.e., also affecting the hidden side information bitstream. BFI and bit error patterns have been generated for E_b/N_0 ratios from 8–11 dB. The resulting frame erasure and residual bit error rates are shown in Figure 6.8.

6.5.4.2 Steganographic VoIP simulations

The simulations were made with a speech database of 8 different English speakers (4 male, 4 female). There were 192 files in total, each about 8 s long, i.e., about 25 min of speech. Each file was encoded by the 12.2 kbit/s mode of the AMR speech codec. In this process, the side information for the concealment was generated according to two different setups, utilizing the general approach described in Section 6.4 with the following specific settings:

i) The optimal estimation technique for each parameter is transmitted as side information (SI type 1) with 8 bit/frame which results in a side information bitstream of 400 bit/s (without channel coding):

   - LSF coefficients: Method E from Table 6.4 ($\alpha_n \in \{0.3; 0.5; 0.7; 0.9\}$); 2 bit/frame
   - Pitch lag: Method L from Table 6.5; 3 bit/frame
   - Adaptive codebook gain: Method Q from Table 6.6; 3 bit/frame
   - Fixed codebook gain: no side information transmitted in this scenario

ii) In addition to the optimal estimation technique (SI type 1), also the quantized estimation errors for the pitch lag and adaptive codebook gains are transmitted (SI type 2). Together with the 8 bit/frame for SI type 1 this results in a total of 26 bit/frame, i.e., a side information bitstream of 1.3 kbit/s (without channel coding):

   - Pitch lag: Transmission of correction $\hat{e}_{T_0} \in \{-3, -2, -1, 0, 1, 2, 3, 4\}$ for each subframe ($\hat{e}_{T_0} = 0$ if correct value cannot be achieved by the adjustment); 3 bit/subframe, i.e., 12 bit/frame
   - Adaptive codebook gain: Transmission of the quantized estimation error vector for the 4 gains of a frame; 6 bit/frame

The side information has been protected by the channel coding scheme presented in Section 6.5.3, and the resulting 40 bit/frame (2 kbit/s bitstream) have been embedded into the codec bitstream as explained in Section 6.5.2. For both scenarios a) and b), random packet losses were introduced with a packet loss rate of 0 %, 3 %, or 6 %, respectively. For scenario b), the bit error patterns generated from the GSM bit level simulations were applied and those frames were marked unusable which had bit errors in the class of most sensitive codec bits (i.e., for which the BFI from the GSM channel decoder was set). The received speech frames, i.e., those not lost in the packet transmission and not marked unusable because of corrupted sensitive bits, have then been decoded by the AMR decoder. Lost and unusable frames have been estimated by packet loss concealment. The concealment has utilized the side information extracted from the following frame if available and not corrupted, otherwise the standard concealment routine of the AMR codec has been used. The resulting speech quality has been determined by the PESQ speech quality measurement tool [ITU-T Rec. P.862 2001] and subjective evaluations.

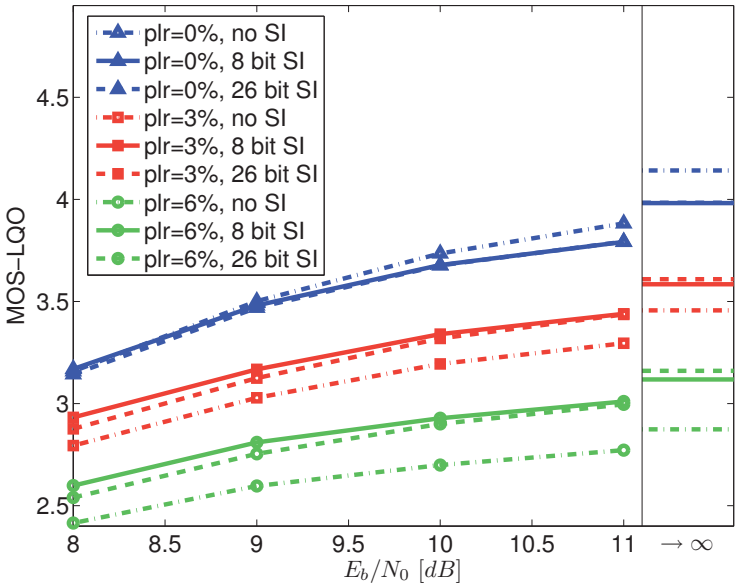


Figure 6.9: MOS-LQO values as measured by PESQ for different amounts of side information on heterogeneous channels with packet loss (IP network) and bit errors (GSM channel); AMR codec with 12.2 kbit/s; SI setups i) with 8 bit/frame and ii) with 26 bit/frame; GSM channel: E_b/N_0 from 8 to 11 dB; IP network: packet loss rates of 0%, 3%, and 6%. For comparison: horizontal lines show values for channels with packet losses only, no bit errors.

6.5.4.3 Simulation Results

The simulation results depicted in Figure 6.9 show the estimated MOS-LQO values measured by PESQ for different side information setups (indicated by different line styles) in dependence on the quality of the GSM channel (given as E_b/N_0 in dB). In addition, different packet loss rates on the packet transmission channel have been considered (plr $= 0\%, 3\%,$ and 6%), distinguished by different line colors and markers in Figure 6.9. The values obtained on a channel with packet losses only and no bit errors (i.e., scenario a)) are shown as horizontal lines without markers. These lines therefore serve as upper bounds for the results of scenario b).

The upper horizontal lines (blue) show the base quality achieved on a channel without bit errors and without packet losses. The MOS-LQO values determined by PESQ show a quality loss for the cases of embedding side information, although the subjective impression reveals no audible difference as discussed in Section 6.5.2.1. There is no dependence on the amount of side information because the side information and error protection bits together always result in a total of 2 kbit/s. For increasing packet loss rates, as shown by the red (middle) and green (lower) horizontal lines, a considerable quality increase by the utilization of side information can be observed which is also confirmed by auditory impression. The higher the loss rate, the higher the quality gain has been. The use of a higher side information rate (26 bit/frame) leads to further quality improvement over the use of 8 bit/frame.

The increase of quality due to utilization of side information can also be consistently shown when considering a GSM bit error channel in addition to the packet

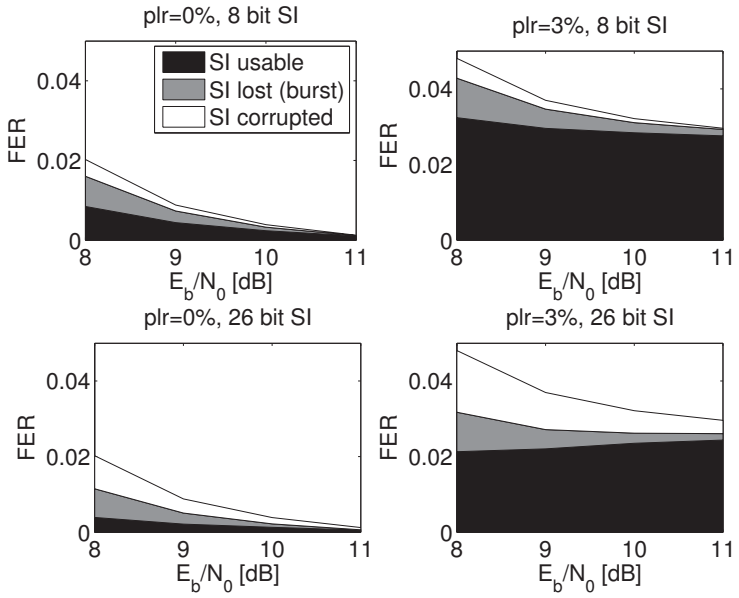


Figure 6.10: Total frame erasure rates (FER) including IP and GSM channel and percentage of usable side information and unusable side information for these frames; AMR codec with 12.2 kbit/s; different E_b/N_0 on the GSM channel; additional packet loss rate of 0% (left column) and 3% (right column); comparing different amount of side information: 8 bit SI (upper graphs) and 26 bit SI (lower graphs).

losses. This is shown by the respective curves (red and square marker for 3% packet loss rate, green and circle marker for 6% packet loss rate). However, here a dependence on the amount of side information can be observed. The higher side information rate, which had shown a quality improvement for the pure packet loss channels, now leads to a lower quality than the low side information rate. The reason for this behavior is the higher sensitivity to residual bit errors, because in case of a higher side information rate less bits are available for the BCH-code and therefore less errors can be corrected. This is confirmed by the higher percentage of losses for which the side information from the following frame is unusable because of uncorrectable bit errors, as shown in Figure 6.10 (compare the upper to the lower graphs).

When the packet loss rate approaches 0% (blue curves, triangle markers in Figure 6.9), the transmission is mainly affected by bit errors on the GSM channel. Here, the benefit of utilizing side information decreases because the side information has been unusable for the majority of losses. The frames have been unusable either because of burst losses, i.e., loss of two or more successive frames, or because of uncorrectable bit errors, as shown in Figure 6.10, left column.

6.6 Conclusions

This work focuses on packet transmission in heterogeneous network scenarios which include wireless access with limited data rates. The studies on packet loss concealment presented in this chapter therefore concentrated on speech codecs based on

the CELP (code excited linear prediction) principle, the state-of-the-art in speech coding for mobile communication systems.

In Section 6.3, it has been shown that a voicing controlled choice of concealment methods that have been in particular designed for each voicing transition consistently improves the performance of standard extrapolation based concealment units. The proposed method utilizes parameter extra- and interpolation and avoids attenuating the signal's amplitude in stationary voiced or unvoiced signal segments when the estimation of a lost frame achieves a high quality.

For a further increase in robustness against packet loss, a new *sender-assisted* packet loss concealment concept has been introduced in Section 6.4 which is based on the transmission of side information to improve the concealment of lost frames at the receiver. Two types of side information have been considered, first, information on what estimation technique is optimal for each codec parameter of a specific frame, and second, a coarse quantization of the respective estimation error. Further advantages of the presented approach are the low necessary bit rate for the side information and the relatively low computational complexity of the employed estimation methods that are based on parameter extra- and interpolation. The low bit rate for the side information makes the approach particularly suitable for wireless transmission scenarios.

Finally, a new approach for the transmission of such low bit rate side information has been presented in Section 6.5. In this approach, the side information is communicated via a steganographic channel within the bitstream of the employed speech codec. In (wireless) packet-switched networks, this approach improves the robustness of the codec against packet losses without requiring additional bit rate. Furthermore, a transparent end-to-end transmission of such side information even over adjacent circuit-switched networks (GSM, UMTS) is possible as long as there is no transcoding involved. However, a sufficient protection of the side information bits by channel coding techniques is required if the end-to-end transmission may leave residual bit errors in the codec bitstream. The proposed concept is inherently backwards compatible because "legacy" terminals without support for the assisted packet loss concealment will not detect and therefore ignore the hidden information while the impact on speech quality remains negligible.

7

Summary

Communication networks are developing towards all-IP networks with flexible core networks of high capacity and different fixed-line and wireless access technologies. These all-IP networks utilize common standardized transmission and signaling protocols which facilitates the development of diverse end-to-end applications and services, including speech, music, and video transmission. The available data rates in the access part of these networks are increasing with the development of new DSL and mobile network technologies. In mobile access networks, the development towards UMTS/HSPA and LTE provides an almost ubiquitous wireless access with data rates from a few hundred kbit/s up to several Mbit/s in future systems. As a consequence, the convergence of fixed and mobile applications and services provides the users with access to services from different devices, at any time and anywhere.

An essential component that enables such convergence of networks and services is the packet-switched transmission technology, which, however, poses new technical problems for the realization of multimedia transmission services. The main problems are packet losses and variable packet transmission delays. Current systems employ several means to overcome these problems. As for packet losses, error protection schemes are applied to recover lost frames at the receiver and thereby reduce the resulting frame loss rate. As for transmission delays, a receiver buffer is usually employed to compensate for variable delays at the expense of an increased end-to-end delay. The optimal parameterization of both components depends on the demands of the application and is also constrained by the properties of the relevant end-to-end IP channel. For frames that are lost or delayed and cannot be recovered by the aforementioned means, a packet loss concealment algorithm at the receiver is usually applied to generate a suitable replacement signal.

The main objective of this work is to explore approaches of how to optimize speech and music transmission on packet-switched networks. The strategy is to determine the optimal choice of transmission parameters and forward error correction (FEC) schemes by utilizing a flexible packet-loss model and taking the expected quality for the users into consideration. Note that the optimization for real-life systems is subject to delay and data rate constraints, which unavoidably lead to

some residual frame losses, especially on wireless packet channels. For such frame losses, an improved packet loss concealment algorithm has been developed, with its main application on the standard speech coding principle, i.e., CELP-based speech codecs. The algorithm derives and transmits side information to assist the receiver-based concealment, thereby improving conventional approaches. Furthermore, a new concept is introduced which transmits the side information as hidden steganographic bitstream within the original encoded bitstream of the speech codec, making the transmission efficient with no data-rate cost. For further details, readers are referred to the following summary, structured in analogy to the main chapters.

Channel Model for Heterogeneous Packet Networks (Chapter 3)

The optimization of the parameters for a packet-switched transmission of speech or music signals requires a reliable model of the transmission channel's error characteristics. Therefore, suitable models have been investigated in Chapter 3 for wireless transmission channels. It has been found that depending on the size and frequency of the transmitted packets, the same application may have to deal with different degrees of packet loss. Larger packets are more likely to have residual bit errors and therefore have to be discarded. Smaller packets are less likely to have residual bit errors, but other problems arise. Their shorter frame length leads to a shorter transmission time interval, which means more packets are transmitted per time unit. This results in a considerable amount of additional packet headers, leading to an increase of the overall data rate. Such an inter-dependency of the size and frequency of the transmitted packets therefore needs to be taken into account when different packetization and FEC schemes are compared in the optimization process (see results in Chapter 5). Channel models have been evaluated in this chapter with the goal to find one that can be adapted to different packet sizes and transmission time intervals. The generalized Gilbert-Elliott model proves to be a suitable base model and novel formulas are derived for adapting this base model to different packet sizes. The resulting extended Gilbert-Elliott model provides the basis for the analytical determination of error correction capabilities in Chapter 4 and establishes the comparability of different techniques. As a prerequisite for the model adaptations, the resolution of the base model has to be high enough.

Analysis of Forward Error Correction Capabilities on Packet Level (Chapter 4)

In heterogeneous packet networks, applications usually have no specific control on the channel coding algorithms that are applied on the physical layer of the different transmission links (e.g., wireless access channels). To control the quality of packet-based speech or music transmission in such networks, the implementation of an additional error protection scheme at the application level is therefore important and has been examined in this chapter.

A set of commonly applied FEC schemes has been investigated in Chapter 4: Reed-Solomon (RS) codes, exclusive disjunction (XOR) of frames, and frame repetition. For each FEC scheme, the residual frame loss rate and distribution after erasure correction has been derived analytically. Their resulting probability functions were found to be dependent on the parameters of the channel model (including the appropriate adaptation to packet size and transmission time interval), the frame length of the media codec, the parameters of the FEC scheme itself, and the packetization strategy. Two different packetization strategies were examined for the transmission of the FEC frames: the transmission of the FEC frames in separate packets (i.e., as independent packet stream) and the transmission of the FEC frames piggybacked to the packets with the original media frames. These theoretical considerations on the error correction capabilities facilitate a fair comparison of different error protection strategies and provide the basis for an optimal system parameterization, as discussed in Chapter 5 for several applications in real-life systems.

System Optimization for Speech and Music Transmission in Packet Networks (Chapter 5)

In real-life systems and applications, the optimal parameterization of a packet-based transmission of speech or music signals depends on various demands and constraints from the relevant application and network scenario, which include the allowed frame loss rate and delay, as well as the available data rate and experienced loss characteristic on the end-to-end transmission channel. In Chapter 5, the adaptable channel model introduced in Chapter 3 and the theoretical evaluation of different FEC schemes and parameterizations in Chapter 4 have been applied to practical applications and transmission scenarios. In particular, four scenarios have been considered, including music streaming and voice conversation on Wireless LAN, UMTS, and heterogeneous packet channels. The results of these scenarios are summarized in the following.

Multicast Music Streaming on WLAN channels

For the multicast streaming of music signals, the application of FEC was found to be generally more data rate efficient than retransmission schemes (except for a very small number of receivers). The optimal choice and parameterization of the FEC scheme itself depends on the channel quality. Since the end-to-end delay is not crucial for a streaming application, around 2 seconds are tolerable, a systematic Reed-Solomon block code with a suitable code rate and block length offers the best data rate efficiency. For example, an (8,4)-code for SNRs of 20-25 dB and an (8,2)-code for SNRs of 15-20 dB are able to achieve a residual loss rate of close to 0% on the considered WLAN channel. Hence, the weak delay constraint for streaming applications can be exploited to achieve a low residual loss rate and a high signal quality.

Voice over IP on WLAN channels

Conversational applications like Voice over IP (VoIP) demand a much lower end-to-end delay not exceeding 300 ms. Too high delay would hinder the interactivity between the conversation partners and therefore directly affect the conversational quality. It was found that if VoIP is transmitted over WLAN, the automatic retransmission in layer 2 is very efficient in recovering lost packets. Such retransmission, however, increases signal delay; the quality of service enhancements according to the IEEE 802.11e standard should therefore be applied. This guarantees a higher priority of the retransmissions on the shared channel. The optimal number of transmission attempts for a single packet has been further examined. It was found to depend on both the channel SNR and the frame length per packet. For an SNR of 20 dB on the considered WLAN channel and a frame length per packet of 5 ms, a residual frame loss rate of around 6% results if no retransmissions are allowed. For larger frame lengths of 20-30 ms the loss rate quickly increases to 11-14%. However, two retransmission attempts after a packet loss will already decrease the frame loss rate to almost 0% for all considered frame lengths of 5-30 ms. An additional application of FEC for end-to-end protection is only necessary if the packets are transmitted over a core network with considerable packet losses. In such a scenario, the optimal approach is to design the FEC scheme for the expected losses in the core network only and to utilize the fast retransmissions on the wireless access links.

Voice over IP on UMTS packet channels

The transmission of VoIP services on UMTS packet channels, which have a lower transmission rate than WLAN channels, requires the use of a speech codec for data rate compression, e.g., the Adaptive Multi Rate (AMR) codec. Additionally, header compression is necessary for such channels. The constraint of the overall transmission capacity does not allow much room for the application of additional forward error correction. The strategy adopted in this chapter is therefore to reduce the speech encoding rate further to make room for error protection. To this end, a multi-rate speech codec like AMR has been employed which (by choosing a mode with lower encoding rate) provides the capacity to transmit further FEC frames while keeping packet size and the total required data rate constant. Depending on the channel characteristics, an optimal trade-off can be achieved between the base signal quality (as determined by the encoding mode) and the error robustness (as provided by the FEC scheme). It became clear that with the conversational quality as the optimization criterion, the base signal quality, the residual frame loss distribution, and the delay increase due to FEC have to be jointly considered to determine the optimal setting for the current channel characteristics. If the packet loss rate on the considered UMTS channel exceeds 2%, the best overall quality is achieved by using the 6.7 kbit/s mode of the AMR codec and transmitting a repetition of each frame piggybacked three packets later. At lower loss rates, no FEC is required and the transmission rate should be completely utilized for the highest AMR encoding mode of 12.2 kbit/s.

The considered UMTS packet channel has been specified by 3GPP explicitly for packet-switched voice services and utilizes Turbo coding. Turbo codes are in general mainly used for high-data-rate services, because their performance increases with the length of the applied interleaver. Nevertheless, studies show that turbo codes still offer some modest gains with respect to convolutional codes with a frame size as low as 100 bits (see, e.g., discussion in [Lee et al. 2000]). The interleaver of the Turbo code in the UMTS standard scrambles the bits of a single transmission block and does not involve preceding or following blocks. The interleaver therefore does not introduce further delay, which facilitates the use of Turbo coding for Voice over IP transmission in UMTS.

Voice over IP on channels with variable packet transmission delays

The scenario of transmitting a VoIP call over a packet network with considerable variation in the packet transmission delay has been discussed. It has been shown that the application of Forward Error Correction (FEC), though reducing the encoding rate of the utilized codec as discussed above, can recover not only frames which have been discarded due to bit errors, but also frames which are delayed. This property has been used to reduce the size of the receiver buffer and consequently the end-to-end delay. The optimal combination of FEC scheme and receiver buffer length has been investigated which leads to the best achievable conversational quality for a specific channel model.

Packet Loss Concealment with Side Information (Chapter 6)

The data rate and end-to-end delay constraints of an application and the relevant network scenario pose constraints on the application of packet-level FEC schemes. Since these FEC schemes are not able to recover all possible frame losses, an efficient packet loss concealment algorithm is needed at the receiver. In this chapter, packet loss concealment algorithms have been developed for CELP-based speech codecs.

The novel approach presented in Section 6.3 adopts the estimation of lost codec parameters to the current voicing state of the speech signal. It is shown that this approach performs significantly better than standard approaches which are mainly based on extrapolation and subsequent muting of the speech signal. A further improvement in Section 6.4 is the transmission of side information to assist the receiver's packet loss concealment. The side information essentially consists of the optimal concealment methods for the codec parameters of each frame suggested at the sender, which can be transmitted with a low additional data rate of 400-1300 bit/s. This sender-assisted approach for packet loss concealment can therefore be classified as a solution between the bit rate intensive sender-driven approaches using FEC (as discussed in Chapter 4) and the receiver-based approaches which do not require additional data rate (as developed in Section 6.3). It is therefore particularly suited for wireless networks with limited transmission rates.

In Section 6.5, an algorithm for transmitting the side information has been introduced which does not require any additional bit rate. This approach utilizes

the method from [Geiser and Vary 2008] for the steganographic transmission of information within the original encoded bits of the AMR speech codec. The side information for the packet loss concealment is then transmitted as hidden bit stream. With this technique, the side information can also be transmitted over conventional circuit-switched parts of the transmission chain, e.g. GSM or UMTS, and utilized if supported by the receiver. A robustness against possible bit errors on such channels can be achieved by applying a suitable channel code to the side information bits, e.g., a BCH block code for error correction and possibly additional CRC bits for error detection. The parameterization of the channel code depends on the available data rate within the hidden bitstream of 2 kbit/s, i.e., the rate of the side information (400-1300 bit/s).

A

IP, UDP, and RTP Protocols

The following sections describe the protocols from the Internet Protocol Suite which are involved in the actual transmission of time-sensitive multimedia signals.

A.1 IP - Internet Protocol

The Internet Protocol is the network layer[1] protocol of the Internet and other networks that are based on the Internet Protocol Suite. Its main functions are the addressing of the packets and a possible segmentation of the datagrams to transmit. Respective header fields therefore contain the IP addresses of source and destination as well as length information and flags to indicate fragmented datagrams (cf. Figure A.1).

The standard protocols from the Internet protocol stack were originally not developed for the transmission of real-time data streams, e.g., for audio or video applications. The Internet Protocol itself does therefore not contain any means for assuring the QoS of a transmission. Nevertheless, the protocol headers of both IP versions, the still most widely used version 4 (IPv4 [Postel 1981a]) and the newer and currently actively deployed version 6 (IPv6 [Deering and Hinden 1998]) provide header fields that can be used for indicating the service class of the transmitted content. This field may be used by traffic management protocols as discussed in Section 2.2.2.

The size of the IP header which is attached to each packet amounts to 40 byte in case of IPv4 and 60 byte for the newer version IPv6. The main motivation behind the development of IPv6 has been the shortage of available IPv4 addresses, leading to an enlargement of the address space, the main contributor to the increase of the header size.

[1] The network layer according to the general ISO OSI reference model is termed Internet layer in the TCP/IP model; see discussion in Section 2.1.

A.2 UDP - User Datagram Protocol

One layer above the network layer, the transport layer of the Internet protocol stack provides two different protocols, the *Transport Control Protocol* (TCP) [Postel 1981b] and the *User Datagram Protocol* (UDP) [Postel 1980].

The connection-oriented TCP numbers the packets and requests a repeated transmission of lost packets. For real-time applications with a limited end-to-end delay, however, there is usually no time for retransmission in case of a packet loss. Therefore, the connection-less UDP is used in such applications, which operates a best-effort transmission of data packets. No guarantee is given whether the packets arrive in correct order or that they arrive at all. Compared to TCP, UDP is a much slimmer protocol, basically just providing the addressing of the target application via port numbers, and a checksum for error detection (cf. Figure A.2). Computation and evaluation of the UDP checksum is optional for IPv4 nodes, but mandatory for IPv6 nodes in the network. UDP does not provide sequence numbers, can therefore not detect missing packets at the receiver and does not initiate packet retransmissions. The ability to reassemble the speech data in correct order at the receiver is not given by UDP and therefore needs to be provided by an application-level protocol, the *Real-time Transport Protocol* (RTP), described in the following section.

The UDP checksum is calculated over the entire protocol data unit, i.e., over both UDP header and payload, plus some additional fields from the IP header. Any detected bit errors in the packet, irrespective of their location, would make the receiver discard the whole packet. A modified version of UDP has therefore been standardized which is more flexible in the ways of error detection within the packet. The *UDP-Lite* protocol [Larzon et al. 2004] is able to deal with partly damaged packets by providing a checksum of variable coverage. This might become relevant when transmitting UDP/IP packets in radio networks that cause bit errors within packets. UDP Lite allows to protect the header fields together with only the most important bits from the payload with a checksum calculated over a variable part of the packet. Only when errors occur in this important bit group, the packets have to be discarded, otherwise they might still be useful for a decoder, knowing that the lesser important bits may contain errors. This *unequal error detection* approach is analyzed and discussed in [Mertz et al. 2005] for the example of a VoIP transmission on UMTS packet channels.

A.3 RTP - Real-Time Transport Protocol

The *Real-time Transport Protocol* (RTP) [Schulzrinne et al. 2003] is an application level protocol that has been developed for the transmission of real-time data streams over the Internet. It provides packet numbering and timestamps to insure correct reordering of packets at the receiver, detection of lost packets, as well as the synchronization of, e.g., parallel video/audio packet streams (cf. Figure A.3).

The RTP header further indicates the type of the packet's payload. For every payload type, i.e., for every media codec, a specific payload format needs to be standardized which defines exactly how to arrange the media frames to form the RTP payload. Depending on the payload format, several frames may be transmitted in each packet, possibly of different encoding rates. The order of the frames may be interleaved and some payload formats allow the transmission of redundancy to enhance the robustness against packet losses, either media dependent or independent. Some payload formats attach a further RTP payload header to signal necessary information for the decomposition to the receiver. Payload formats relevant for this work are described in more detail in Section 2.4.1.1 and Section 2.4.1.2.

A.4 RTCP - RTP Control Protocol

The RTP standard also specifies the *RTP Control Protocol* (RTCP), which provides the transmission of feedback information on the quality of the current transmission. RTCP periodically transmits sender and receiver reports, containing information on the amount of packets sent/received, the number of packets lost and an estimate of the inter-arrival jitter between successively received packets. From timestamp fields an estimate of the current round-trip time can be calculated. This feedback information can be utilized by the sender to choose an appropriate transmission scheme that is suitable for the current channel conditions, and the receiver can use the information to adapt its jitter buffer length.

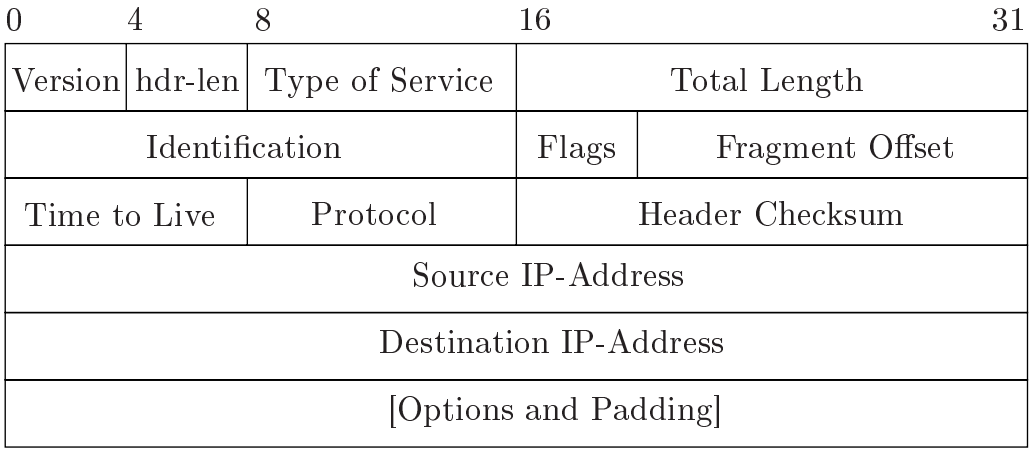


Figure A.1: IP Header

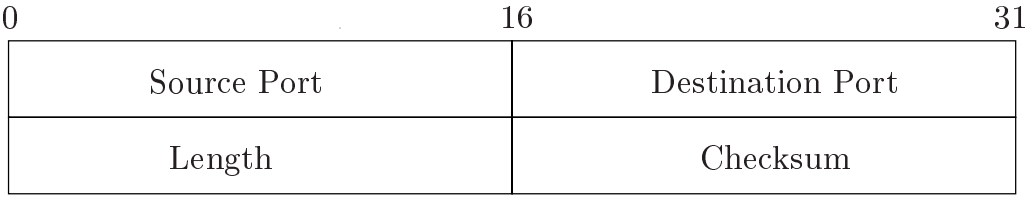


Figure A.2: UDP Header

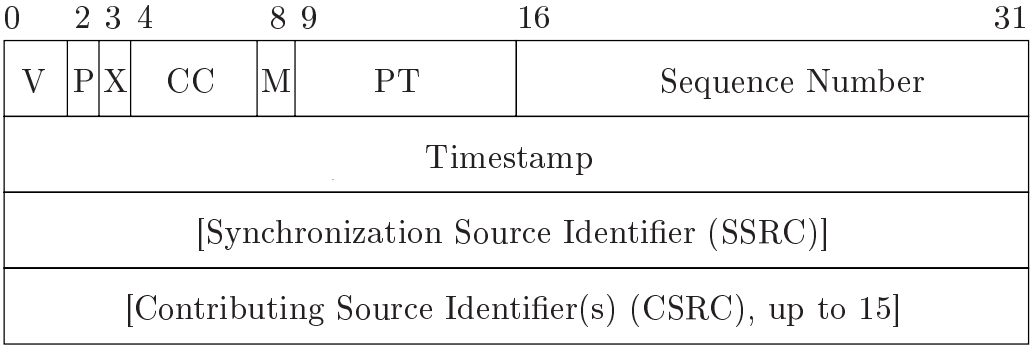


Figure A.3: RTP Header; V: version, P: padding octets, X: header extension, CC: CSRC count, M: marker bit, PT: payload type

B

RTP Payload Format for MP3 Music Signals

B.1 MP3 Bit Stream Format

The MPEG-1 Audio Layer 3 [ISO/IEC 11172-3:1993 1993] audio coding standard, also referred to as MP3, defines the following bit stream format. It consists of so-called MP3 blocks of a fixed length B, each having an additional MP3 block header, as shown in Figure B.1. The block length depends on the chosen fixed encoding rate of the MP3 audio codec, e.g., 128 or 192 kbit/s. Each block contains the encoded bits of an audio frame, however, some frames do not use all available bits in a block. The remaining bits then form a so-called bit reservoir. This bit reservoir will be used by the following frame before the actual bits of the bock are filled. If a bit reservoir is available, a single frame may then also use more than the available bits in the block as, e.g., in block 4 and 5 of Figure B.1. Each block header contains a pointer to the position of the first bit belonging to the current block, which can point to the data part of the current block or to a bit reservoir of a preceding block, such that the receiver is able to extract the bits of the frames.

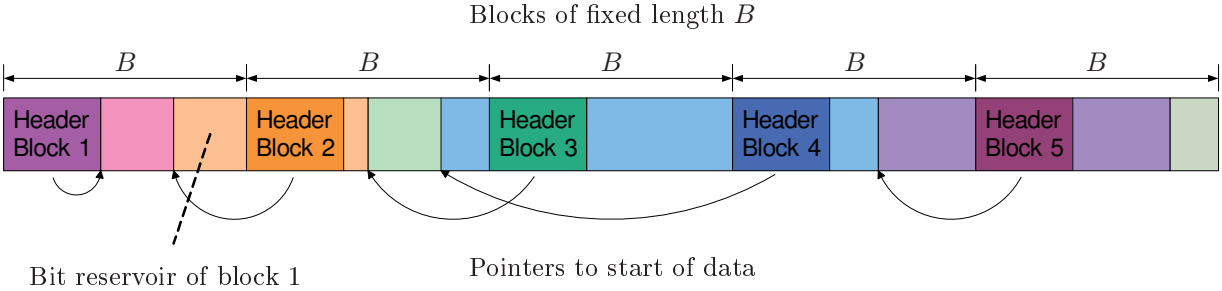


Figure B.1: MP3 Bit Stream Format

B.2 MP3 Frame Packetization

Due to the use of bit reservoirs, as explained above, the blocks of the MP3 bit stream are not independent from each other. If each of these blocks would be transmitted in a single IP packet, this would lead to error propagation in case of packet loss, as the lost block may also contain information of other frames in its bit reservoir. Therefore, an RTP payload format has been standardized in [Finlayson 2008] which first resorts the bit stream in order to make the packets independent from each other. The resorting operation separates the bits of the frames, i.e., the bit reservoirs are resolved such that every IP packet will only contain the bits of a single frame. The resorting of the bit stream results in a higher loss tolerance of the transmission and is therefore used in the considerations and developments of this work. This, however, results in packets of different lengths, which has to be taken into account when applying forward error correction techniques as described in Chapter 4.

The RTP payload format [Finlayson 2008] defines *application data units* (ADUs) which consist of a 2-byte ADU description (containing size and continuation flag), followed by the original 4-byte MPEG header of the current frame. Optionally, a 2-byte CRC may be added. The ADU header is followed by the original side information structure from the MP3 frame and finally, the complete encoded audio data of the MP3 frame including parts from bit reservoirs of previous packets.

The resulting IP packet will have an average length L_p depending on the MP3 encoding bit rate: IPv4, UDP and RTP headers require 20+8+12=40 byte. The following ADU requires 2 byte for the header and an average $\frac{R \cdot T_f}{8}$ byte for an average data rate R and frame length $T_f = 26\,\mathrm{ms}$. This results in $L_\mathrm{p} \simeq 458\,\mathrm{byte}$ for R=128 kbit/s, or $L_\mathrm{p} \simeq 666\,\mathrm{byte}$ for R=192 kbit/s.

The receiver will either have to regenerate the original bit stream which is then decoded by a standard implementation of an MP3 decoder, or the receiver has to contain a modified decoder which is able to decode the resorted MP3 bit stream.

Figure B.2: IETF RFC 3119/5219: A More Loss-Tolerant RTP Payload Format for MP3 Audio

C

Overview of Packet Sizes

Table C.1 derives packet sizes L_p and packet data rates R_p for different codecs, frame lengths T_f or number of frames per packet N_f, and code rates of packet level FEC schemes r_c. The packet size depends to a large extent on the utilized IP protocols, i.e., whether version 4 or 6 is used and whether header compression is applied. L_p and R_p are given for the following header header sizes:

- IPv4/UDP/RTP: $L_\mathrm{h} = 40\,\mathrm{byte}$
- IPv6/UDP/RTP: $L_\mathrm{h} = 60\,\mathrm{byte}$
- ROHC(IP/UDP/RTP): $L_\mathrm{h} = 3\,\mathrm{byte}$

For the AMR speech codec, it has been assumed that the frames are packed according to the RTP payload format defined in [Sjoberg et al. 2007], with octet-aligned mode, no frame CRCs, and no interleaving, i.e., an RTP payload header of 1 byte header + 1 byte ToC entry for each frame.

Codec	R_c [kbit/s]	T_f [ms]	L_c [bit]	$L_{plh,f}$ [bit]	L_f [bit]	$L_{plh,p}$ [bit]	N_f	r_c	IPv4/UDP/RTP L_p [byte]	IPv4/UDP/RTP R_p [kbit/s]	IPv6/UDP/RTP L_p [byte]	IPv6/UDP/RTP R_p [kbit/s]	ROHC L_p [byte]	ROHC R_p [kbit/s]
PCM	64.00	20	1280	0	1280	0	1	1	200	80.0	220	88.0	163	65.2
PCM	64.00	10	640	0	640	0	1	1	120	96.0	140	112.0	83	66.4
PCM	64.00	5	320	0	320	0	1	1	80	128.0	100	160.0	43	68.8
PCM	64.00	1	64	0	64	0	1	1	48	384.0	68	544.0	11	88.0
G.729	8.00	30	240	0	240	0	1	1	70	18.7	90	24.0	33	8.8
G.729	8.00	20	160	0	160	0	1	1	60	24.0	80	32.0	23	9.2
G.729	8.00	10	80	0	80	0	1	1	50	40.0	70	56.0	13	10.4
iLBC	13.33	30	400	0	400	0	1	1	90	24.0	110	29.3	53	14.1
iLBC	15.20	20	304	0	304	0	1	1	78	31.2	98	39.2	41	16.4
AMR	4.75	20	95	8	103	8	1	1	54	21.6	74	29.6	17	6.8
AMR	5.15	20	103	8	111	8	1	1	55	22.0	75	30.0	18	7.2
AMR	5.90	20	118	8	126	8	1	1	57	22.8	77	30.8	20	8.0
AMR	6.70	20	134	8	142	8	1	1	59	23.6	79	31.6	22	8.8
AMR	7.40	20	148	8	156	8	1	1	61	24.4	81	32.4	24	9.6
AMR	7.95	20	159	8	167	8	1	1	62	24.8	82	32.8	25	10.0
AMR	10.20	20	204	8	212	8	1	1	68	27.2	88	35.2	31	12.4
AMR	12.20	20	244	8	252	8	1	1	73	29.2	93	37.2	36	14.4
AMR-WB	6.60	20	132	8	140	8	1	1	59	23.6	79	31.6	22	8.8
AMR-WB	8.85	20	177	8	185	8	1	1	65	26.0	85	34.0	28	11.2
AMR-WB	12.65	20	253	8	261	8	1	1	74	29.6	94	37.6	37	14.8
AMR-WB	14.25	20	285	8	293	8	1	1	78	31.2	98	39.2	41	16.4
AMR-WB	15.85	20	317	8	325	8	1	1	82	32.8	102	40.8	45	18.0
AMR-WB	18.25	20	365	8	373	8	1	1	88	35.2	108	43.2	51	20.4
AMR-WB	19.85	20	397	8	405	8	1	1	92	36.8	112	44.8	55	22.0
AMR-WB	23.05	20	461	8	469	8	1	1	100	40.0	120	48.0	63	25.2
AMR-WB	23.85	20	477	8	485	8	1	1	102	40.8	122	48.8	65	26.0

Table C.1: Packet sizes for different media codecs

D

Wireless Packet Transmission Standards: UMTS and WLAN

In this appendix, a short overview of the data link and physical layers of the wireless transmission standards UMTS and WLAN are given.

D.1 UMTS Packet-Switched (PS) Channels

The complete settings of a UMTS transport channel are specified by a *Radio Access Bearer* (RAB) definition. Typical examples for RABs are given in [3GPP TR 25.993 2008], e.g., specific packet channels for transmitting certain AMR modes using header compression. The relevant sub-layers of the UMTS link and physical layers are shortly described in the following.

D.1.1 Packet Data Convergence Protocol

The *Packet Data Convergence Protocol* (PDCP) [3GPP TS 25.323] receives the IP packets from the network layer and is responsible for optional header compression, e.g., according to the ROHC standard as described in Section 2.4.2. A one byte PDCP header is added including a header compression identifier denoting the applied compression algorithm. Different packet flows, e.g., the RTP and RTCP streams of a VoIP call, are distinguished by a context identifier added by the header compression algorithms. If configured by upper layers, the PDCP may omit the PDCP header if no header compression is applied. The packets are finally handed to the transmission buffer of the Radio Link Control (RLC) layer.

D.1.2 Radio Link Control Protocol

Every transmission time interval (TTI), the *Radio Link Control* (RLC) layer [3GPP TS 25.322] is forming an RLC PDU (protocol data unit) of a fixed size by inserting packets from its transmission buffer. The packets may be segmented or concatenated to fit into the PDU, and information of the packet boundaries within the PDU are placed in the RLC header.

The RLC protocol specifies different possible modes of operation. The *Unacknowledged Mode* is used for conversational packet-switched speech transmission because of the strong delay demands of this scenario. In contrast to the *Acknowledged Mode*, there are no retransmissions of PDUs, and residual bit errors in a PDU after channel decoding will lead to the loss of the contained packets. Because of the possible segmentation of the IP packets, the loss of a single RLC PDU can result in the loss of several packets. Streaming scenarios with less strict delay demands might use the *Acknowledged Mode* with retransmissions of PDUs.

D.1.3 Medium Access Control (MAC) Protocol

The *Medium Access Control* (MAC) protocol specification is defined in [3GPP TS 25.322]. The MAC serves as an interface between the RLC layer and the physical layer, mapping the logical channels to physical layer transport channels, e.g., broadcast channels, shared channels, dedicated transport channels, and high speed downlink shared channels.

D.1.4 Physical Layer and Channel Coding

The definition of the physical layer (Layer 1) transmission includes channel coding techniques for error protection and a CRC for error detection. For UMTS, two different channel coding schemes are specified in [3GPP TS 25.212]: a convolutional coder (rate 1/2 or 1/3) and a Turbo coder (rate 1/3). The rate-1/3 *Turbo coder* consists of two parallel concatenated convolutional codes (PCCC), coupled by the Turbo code internal interleaver. The two constituent codes are recursive and have a constraint length of 4. The iterative decoding scheme typically uses the Log-MAP algorithm with 4 iterations. The maximum usable block length for the Turbo coder is 5114 bits. The *convolutional coder* of rate 1/3 has a constraint length of 9, is non-systematic and non-recursive, and encodes blocks of maximal 504 bits, which are terminated by 8 tail bits.

D.2 Wireless LAN (IEEE 802.11)

The WLAN (Wireless Local Area Network) transmission standard, which is standardized by the ANSI and IEEE as Std 802.11 [IEEE Std 802.11 2007], defines medium access control (MAC) and physical layer functions for the wireless transmission of data in the 2.4 and 5 GHz ISM band. WLAN appears to higher layers,

i.e. the Logical Link Control (LLC), as standard IEEE 802 LAN (Ethernet). Therefore, the WLAN standard handles station mobility and provides other untraditional functionality within the MAC layer in order to meet the reliability assumptions of the LLC about lower layers.

The WLAN protocol defines different variants, e.g., 802.11a, 802.11b, and 802.11g, which differ in the frequency band and the utilized modulation schemes, thereby providing a range of transmission data rates from 1, 2, 5.5 and 11 Mbit/s up to 54 Mbit/s. The higher the data rate, the more prone the system gets to transmission errors when the channel quality decreases. In standard systems, the choice of the modulation scheme is therefore adapted to the current channel quality, leading to a possible variation of the data rate within the same connection. The standard 802.11e [IEEE Std 802.11e 2005] specifies several improvements in QoS, e.g., faster retransmission for delay sensitive applications.

D.2.1 Medium Access Control (MAC) Protocol

The WLAN transmission protocol defines the MAC (medium access control) and physical layer for the transmission of IP packets. The MAC layer receives the IP packets from the upper layers and forms MAC protocol data units (MPDU), which consist of the MAC header (24 byte), the IP packet, and a frame check sequence (4 byte).

The WLAN MAC initiates retransmission of unacknowledged frames in unicast scenarios, but not in case of multicast as the number of necessary transmission attempts would increase considerably with an increasing number of receivers.

D.2.2 Physical Layer Convergence Protocol (PLCP)

The Physical Layer Convergence Protocol (PLCP) defines the mapping of MPDUs into a framing format suitable for transmission on the physical medium and allows the MAC to operate with minimum dependence on the PMD (physical medium dependent) sublayer.

The PLCP frame format consists of the PLCP preamble, the PLCP header, and the PLCP payload. The PLCP preamble contains a synchronization bit pattern and a start frame delimiter, together requiring 144 bits. The following PLCP header indicates the length of the MPDU, the modulation type that will be used too transmit the MPDU, and a 16 bit CRC, in total requiring 48 bit. PLCP preamble and header are always transmitted using the 1 Mbit/s DBPSK modulation scheme, resulting in a transmission time of $\tau_{\mathrm{PLCP}} = 192\,\mu$s. The following PLCP payload consists of the MPDUs and is transmitted at the chosen WLAN transmission rate. The PLCP payload transmission time depends on the used modulation scheme, i.e. the WLAN transmission rate R_{ch}:

$$\tau_{\mathrm{MPDU}} = \frac{L_{\mathrm{MPDU}}}{R_{\mathrm{ch}}}. \tag{D.1}$$

This results in a total packet transmission time of

$$\tau_{\mathrm{p}} = \tau_{\mathrm{PLCP}} + \tau_{\mathrm{MPDU}}. \tag{D.2}$$

D.2.3 Physical Layer

Depending on the WLAN variant and data rate, different modulation schemes are used on the physical channel. While 802.11b uses *Direct-Sequence Spread Spectrum* (DSSS) for the 1 and 2 Mbit/s and *Complementary Code Keying* (CCK) for 5.5 and 11 Mbit/s, 802.11a and 802.11g, an enhancement of 802.11b, utilize *Orthogonal Frequency-Devision Multiplexing* (OFDM) as multi-carrier modulation scheme which is less susceptible to multipath interference and achieves a higher spectral efficiency. On the different sub-channels, standard single-carrier schemes are applied, i.e., BPSK, QPSK, 16-QAM or 64-QAM.

E

Deriving Channel Models for UMTS & WLAN

E.1 Model Training and Assessment

The determination of an accurate channel model requires a sufficiently large data set of observations from simulations or real-life measurements. The parameters of the channel model are derived from or trained with this set of data. The complexity of the training process depends on the considered channel model. In the following, suitable training procedures for the relevant channel models will be described. Finally, different measures will be discussed which allow to assess whether the experimental data and the theoretical models agree. With these measures, the different models will be compared regarding their *goodness of fit*.

E.1.1 Deriving Model Parameters from Channel Measurements and Simulations

The parameters of the simplified Gilbert model can be derived straight forward from a given measurement sequence, because the state sequence is directly observable from the loss trace. For the Gilbert and the Generalized Gilbert-Elliott models, this state sequence is not observable from the loss trace. Because of the given error probabilities in states G and B it is not clear in which state a packet has been lost or received. These models therefore form *Hidden Markov Models* (HMM).

For the estimation of the parameters of the latter models, the *Baum-Welch* algorithm [Baum et al. 1970; Welch 2003] has been used. It performs a maximum likelihood estimation of the model parameters for a hidden Markov model, i.e., the transition and emission (here loss) probabilities.

E.1.2 Channel Model Adaptation Based on Feedback Reports

For application and transmission scenarios where the channel behavior is expected to be fairly constant, the channel model may be determined in advance and used for the design and parameterization of the system. Other scenarios may experience unpredictable or changing channel characteristics. Here, the system either needs to be designed for the worst expected channel behavior, or the system parameterization needs to update itself in the course of transmission. The latter requires an update of the channel model based on the current loss statistics of the transmission.

In general, the channel model can be implemented on either side of the transmission chain, at the sender or at the receiver. An implementation at the *sender* would require detailed feedback reports from the receiver about the experienced loss statistics, e.g., realized with feedback reports according to the Real-Time Transport Control Protocol (RTCP) [Schulzrinne et al. 2003]. The exact details required for a model update depend on the implemented model, whereas the necessary frequency of the reports depends on the change rate of the channel characteristics. Based on the updated model, the sender then adapts the parameterization of the transmission parameters, e.g., codec rate, redundancy scheme, amount of redundancy, etc., in order to achieve the best possible system performance under the given data rate and delay constraints. If the channel model is implemented at the *receiver*, the measured loss statistics can be directly used for the channel model update and do not need to be transmitted. Instead, the receiver either transmits the resulting channel model parameters to the parameter optimization routine at the sender, or the receiver performs this optimization itself and transmits the chosen parameter set back to the sender.

The adaptation of the system parameterization to the changing channel characteristics needs a frequent update of the underlying channel model. The update of the channel model itself requires a feedback channel from the receiver to the sender for exchanging the loss statistics or the derived parameters. The updated channel model is based on the current settings of the transmission time interval and packet size as it models the according transmission statistics. An adaptation of the model to different transmission time intervals and packet sizes according to Section 3.2 is therefore limited to lower resolutions than the current setting.

E.1.3 Testing the Goodness of Fit

The *goodness of fit* of a model can be assessed by graphical and numerical measures. The advantage of graphical measures is that they show the complete data set at once and therefore allow to see how good the model fits the data, in which ranges it might deviate, or if it is in general not suitable to model the form of the distribution. Numerical measures, on the other hand, compress the whole information into a single numerical value, which can then be objectively compared for different models.

In the following, both graphical and numerical measures are introduced which will then be used to compare the different channel models for the considered network

scenarios in Appendix E.2. The goodness of the channel models shall be assessed with respect to the following properties of the loss process: a) the loss rate, b) the average burst length, and c) the distribution of loss and receive lengths, i.e., the number of consecutively lost or received packets which will be referred to as burst and gap lengths, respectively.

E.1.3.1 Graphical Evaluation of the Goodness of Fit

The distributions of burst and gap lengths have been calculated from the measured loss traces and determined theoretically for the different models from the models' parameters. For each model, the calculated and predicted distributions will be plotted for comparison in the same diagram and another plot will depict the difference between the two.

E.1.3.2 Chi-Square Goodness of Fit Test

In addition to the graphical plots, the goodness of fit of the models shall be objectively assessed by the chi-square test [Press et al. 1992; Papoulis and Pillai 2002]. This test can be used to determine how likely a given data set is drawn from a postulated distribution function. The data set needs to be discrete or previously partitioned into a discrete number of bins.

Suppose that M_i is the number of events observed in the ith bin, with the total number of observations $N = \sum_{i=1}^{m} M_i$ over all considered m bins. The observations are compared to the expected number of events in the ith bin, Np_i, with the probability of occurrence p_i according to the distribution function under test. The statistic used in this test is known as Pearson's statistic. It is calculated as

$$q = \sum_{i=1}^{m} \frac{(M_i - Np_i)^2}{Np_i}. \tag{E.1}$$

A large value of q indicates that it is rather unlikely that the M_i are drawn from the postulated distribution defined by the p_i. In the calculation of the sum, all terms where $M_i = Np_i = 0$ are omitted. Under the hypothesis H_0 that the data set is drawn from the postulated distribution function, the random variable q has a $\chi^2(m-1)$ distribution. Thus, the hypothesis H_0 is accepted if

$$q < \chi^2_{1-\alpha}(m-1), \tag{E.2}$$

with the significance level α, usually chosen as $\alpha = 0.05$.

The considered data sets of the channel characteristics are discrete run lengths of successively lost and received packets, i.e., a binning of the data is not necessary. The bins i then refer to the length of a loss or a gap, i.e., the number of successively lost or received packets. The M_i are derived by analyzing the loss traces of a particular channel for the run lengths of bursts and gaps and by subsequently measuring their histograms. For the channel model under test, the probability p_i for the burst length i is calculated according to the respective probability $P_{\mathrm{b}}(i)$, specified in the model definitions in Section 3.1. For the gap lengths the calculation is done accordingly.

E.2 Simulation and Modeling of UMTS and WLAN Channels

Different realistic channels have been considered for applying the methodologies proposed in this work. In particular, two different network technologies have been considered: the cellular UMTS network and the Wireless LAN transmission technology (cf. Appendix D). Bit level simulations with dedicated simulation software have been carried out to determine realistic loss patterns for different channels, transmission parameters (e.g. packet size and transmission time interval), as well as for different channel qualities. Based on these patterns, respective channel models have been determined as described in Appendix E.1. The following sections detail the specific settings of the transmission channels, measurements and simulations, as well as an evaluation of the resulting channel models.

E.2.1 UMTS Channel Model

Several packet-switched (PS) channels have been standardized by 3GPP for use in Release 5 or later in UMTS networks. In this work we assume, e.g., a packet transmission of a voice call arriving in the UMTS network and being routed to an end user over the UMTS air interface. For this scenario, a dedicated UMTS downlink channel (DTCH) as defined in [3GPP TR 25.993 2008, Sec. 7.1.123], has been simulated on bit level by using a UMTS reference implementation in the Synopsys System Studio Software [Synopsys 2007]. As transmission scenario, the "outdoor to indoor and pedestrian test environment" defined in [ETSI TR 101 112 1998] and its according propagation model was selected. The radio access bearer (RAB) has 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 is intended for transmitting IP/UDP/RTP packets containing AMR encoded speech frames and using 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 rate 1/3 Turbo code. If a 16 bit CRC detects residual bit errors after channel decoding, the transport block and the contained IP packet is discarded.

The performance of Turbo codes increases with the length of the applied interleaver. Turbo codes are therefore mainly used for high-data-rate services. Low rate circuit-switched voice services usually use convolutional codes. Nevertheless, studies show that turbo codes still offer some modest gains with respect to convolutional codes with a frame size as low as 100 bits (see, e.g., discussion in [Lee et al. 2000]). Third-generation systems therefore allow Turbo codes to be used for almost all data rates. The dedicated UMTS channel from [3GPP TR 25.993 2008, Sec. 7.1.123] is explicitly specified for packet-switched voice services and utilizes Turbo coding. The interleaver of the Turbo code in the UMTS standard scrambles the bits of a transmission block (40-5114 bit) and does not involve preceding or following blocks. Hence, if a single packet is transmitted per transmission block, no additional delay is introduced by the interleaver itself.

E_c/I_{or} [dB]	Loss Statistics		Channel Model Parameters			
	P_{pl}	$\overline{b}$	$P_{t,GB}$	$P_{t,BG}$	$P_{e,G}$	$P_{e,B}$
-10.0	0.001030	1.072288	0.000930	0.922163	0.000083	0.940025
-13.0	0.007190	1.145118	0.003304	0.770122	0.003156	0.947434
-15.5	0.025390	1.258391	0.011330	0.691746	0.009515	0.994640
-17.0	0.048660	1.381944	0.023910	0.645534	0.013423	0.999996
-20.0	0.129770	1.831813	0.079906	0.542665	0.001631	1.000000

Table E.1: Packet loss rates P_{pl}, mean burst lengths $\overline{b}$, and corresponding channel model parameters determined from simulated 17.6 kbit/s packet-switched UMTS channel for different channel qualities E_c/I_{or}.

Model	Burst lengths			Gap lengths		
	N	q	$\chi^2_{0.95}(N-1)$	N	q	$\chi^2_{0.95}(N-1)$
Simplified Gilbert	9	129.53	15.507	209	324.29	242.65
Gilbert-Elliott	9	36.287	15.507	209	324.42	242.65
4-state, $G_{min} = 3$	9	668.34	15.507	209	398.23	242.65
4-state, $G_{min} = 5$	9	295.07	15.507	209	500.93	242.65
4-state, $G_{min} = 10$	9	57.878	15.507	209	715.62	242.65
4-state, $G_{min} = 15$	9	38.942	15.507	209	826.15	242.65

Table E.2: χ^2 Goodness of fit test: UMTS, $E_c/I_{or} = -17$ dB

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 as explained in Appendix E.1. The determined parameters, as well as the resulting packet loss rates P_{pl} and mean burst lengths $\overline{b}$, are given in Table E.1.

The goodness of fit of different channel models for the case of $E_c/I_{or} = -17$ dB is graphically analyzed in Figure E.1 for the simplified Gilbert model, in Figure E.2 for the generalized Gilbert-Elliott model, and in Figure E.3 for the 4-state model with $G_{min} = 15$. Considering the distributions of both burst and gap lengths, the Gilbert-Elliott model provides the closest fit for the measured data. This is also confirmed by the lowest values of the Pearson's statistic q from the χ^2-test shown in Table E.2. Here, N shows the largest non-zero burst or gap length in the simulation.

E.2.2 WLAN Channel Model

For the simulation of a Wireless LAN (WLAN) channel, the bit level IEEE 802.11a simulation model from The MathWorks' Simulink Communications Blockset [The MathWorks] has been used. The simulation employs a dispersive multipath fading channel with a maximum Doppler shift of 200 MHz. The transmission data rate of the channel is set to a fixed rate of 6 Mbit/s. Different channel qualities (SNRs) have been simulated and the corresponding packet loss traces have been recorded. Based on the simulated error patterns we determined a channel model with a high

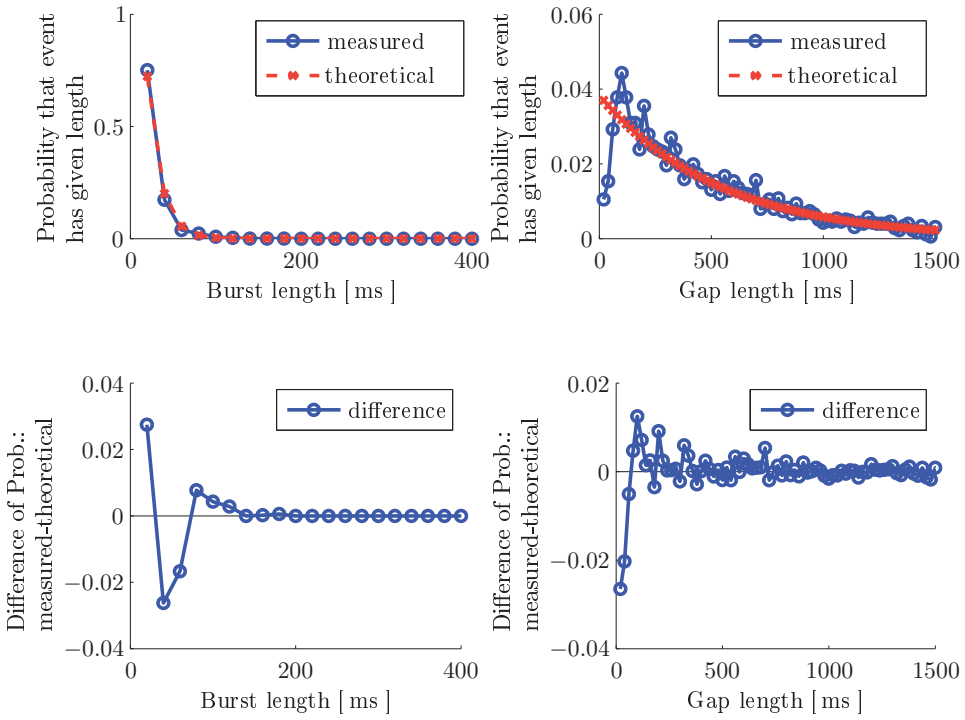


Figure E.1: Simplified Gilbert model; UMTS, $E_c/I_{or} = -17\,\mathrm{dB}$

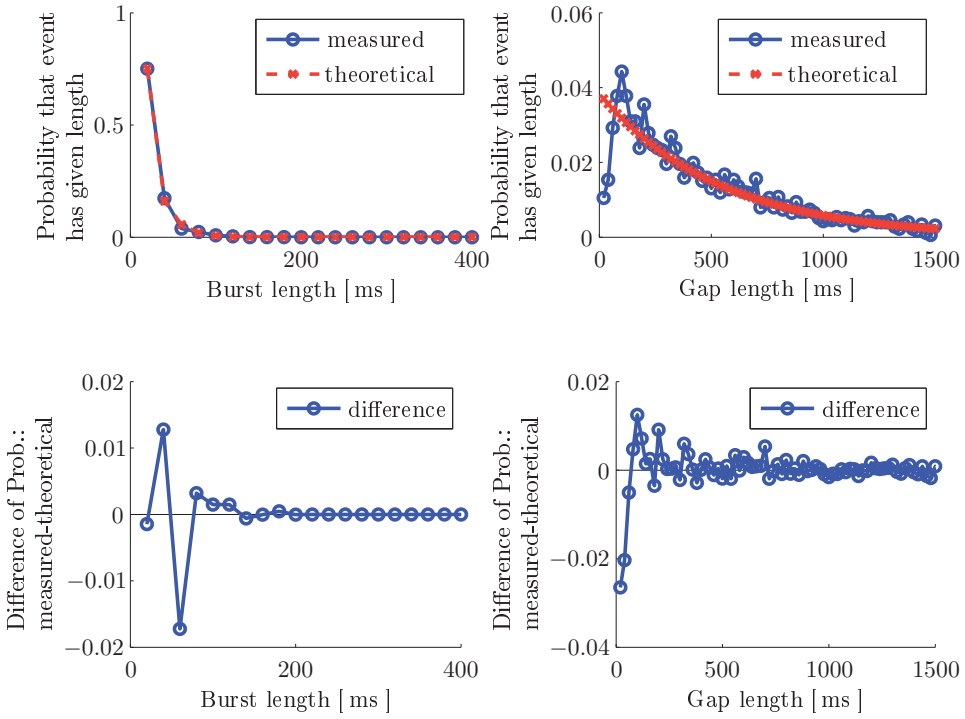


Figure E.2: Generalized Gilbert-Elliott model; UMTS, $E_c/I_{or} = -17\,\mathrm{dB}$

SNR [dB]	Loss Statistics		Channel Model Parameters			
	P_{pl}	$\overline{b}$	$P_{\mathrm{t,GB}}$	$P_{\mathrm{t,BG}}$	$P_{\mathrm{e,G}}$	$P_{\mathrm{e,B}}$
15.0	0.102043	2.651747	0.036868	0.272487	0.000002	0.856205
20.0	0.041908	1.657073	0.023012	0.448553	0.003924	0.782311
25.0	0.019143	1.254851	0.008282	0.513822	0.008245	0.695282
30.0	0.012407	1.141754	0.003369	0.522375	0.008065	0.685686

Table E.3: Packet loss rates P_{pl}, mean burst lengths $\overline{b}$, and corresponding channel model parameters determined from simulated WLAN channel for different channel qualities (SNR). Base channel model with $T'_{\mathrm{TTI}} = \tau'_{\mathrm{p}} = 0.08\,\mathrm{ms}$.

resolution ($T'_{\mathrm{TTI}} = \tau'_{\mathrm{p}} = 0.08\,\mathrm{ms}$) which can be adapted to different packet lengths and TTIs according to Sec. 3.2.

Figure E.4 shows the loss rates P_{fl} and mean burst lengths $\overline{b}$ of the transmitted data units for different SNR values on the channel. The measured data points are marked with DA, followed by the transmission time interval T_{TTI} (0.08, 1, 5, 10, or 20 ms) and the packet transmission time τ_{p} which reflects the packet size (0.08, 0.16, 0.24, or 0.32 ms). The data points 'DA 0.08, 0.08' therefore reflect the measurements at the maximum resolution. The other data points 'DA T_{TTI}, τ_{p}' have been derived by downsampling this original measurement. The curves marked with GE show the predicted loss rates and mean burst lengths of the respective Gilbert-Elliott channel models. The base channel model has been trained on the simulated loss pattern with $T'_{\mathrm{TTI}} = \tau'_{\mathrm{p}} = 0.08\,\mathrm{ms}$ as explained in Section E.1.1. The other models have been obtained by applying the adaptations to different transmission time intervals and packet sizes from Section 3.2, i.e., the state transition probabilities have been calculated according to (3.25) and the transition dependent error probabilities according to (3.30). A close fit of model and data points can be observed.

For an SNR of 20 dB, Figures E.5–E.7 show the goodness of fit of the simplified Gilbert model, the Gilbert-Elliott model, and the 4-state model, respectively, for the base channel of highest resolution. The generalized Gilbert-Elliott model shows the closest fit between data and model. After an exemplary adaptation of the model to a new transmission time interval of $T_{\mathrm{TTI}} = 10\,\mathrm{ms}$ and a packet transmission time of $\tau_{\mathrm{p}} = 0.2\,\mathrm{ms}$, the goodness of fit of model and data is visualized in Figures E.8 and E.9 for the simplified Gilbert model and the Gilbert-Elliott model, respectively. It turns out that especially the distribution of the gap lengths cannot be modeled sufficiently by the simplified Gilbert model. The Gilbert-Elliott model still shows a good fit of model and data after adaptation, and this also applies to the other combinations of T_{TTI} and τ_{p}, as shown clearly by the results of the χ^2-test presented in Table E.4.

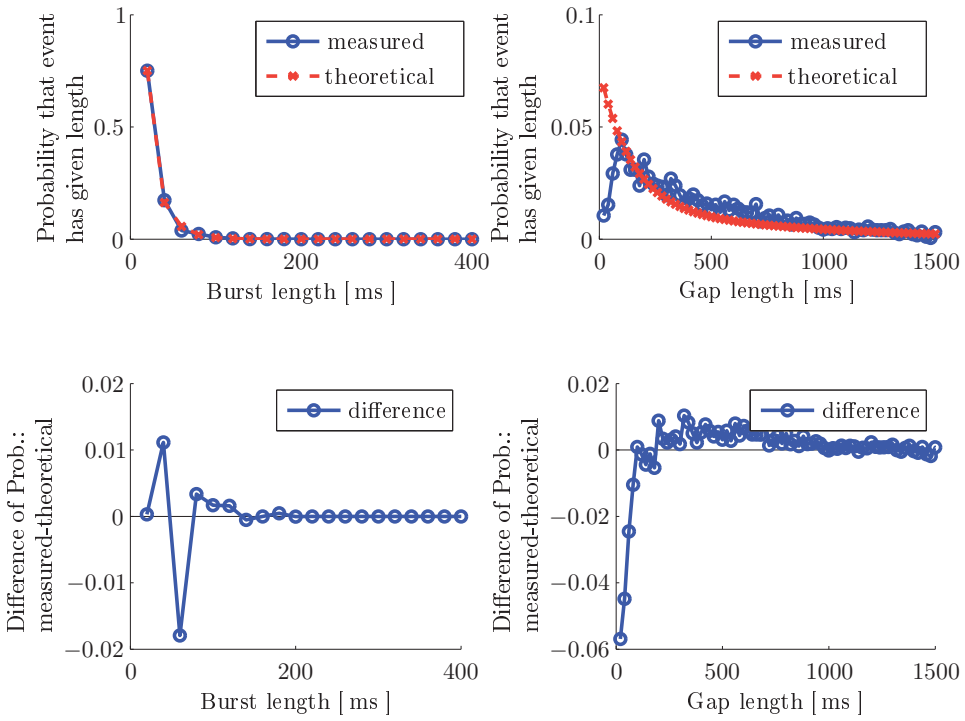


Figure E.3: 4-states model; $G_{\min} = 15$; UMTS, $E_c/I_{or} = -17\,\mathrm{dB}$

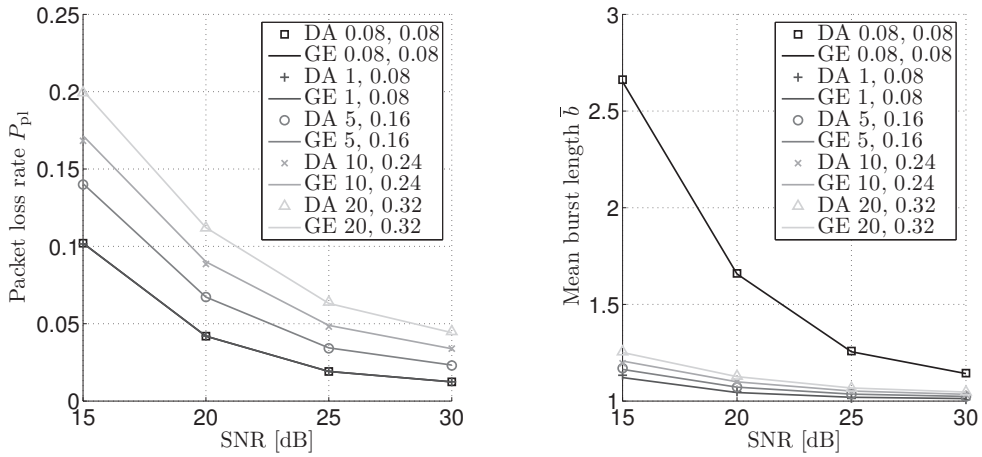


Figure E.4: Packet loss rates P_{pl} and mean burst lengths $\overline{b}$ on simulated WLAN channel for different channel qualities (SNR). DA – based on measured data points; GE – predicted by the respective Gilbert-Elliott model

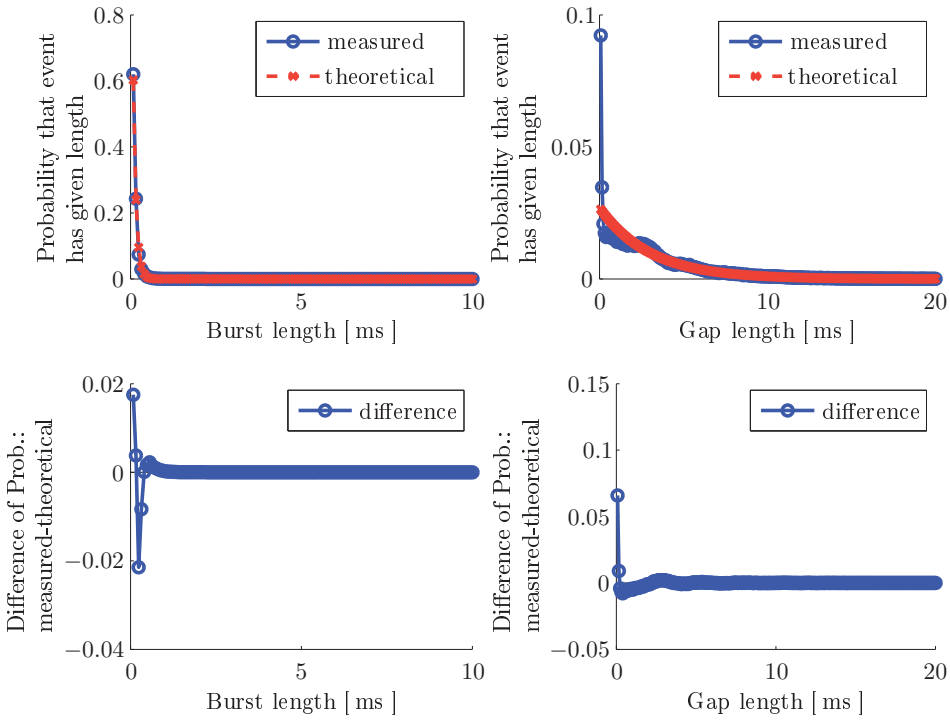


Figure E.5: Simplified Gilbert model; WLAN, $SNR = 20\,\mathrm{dB}$

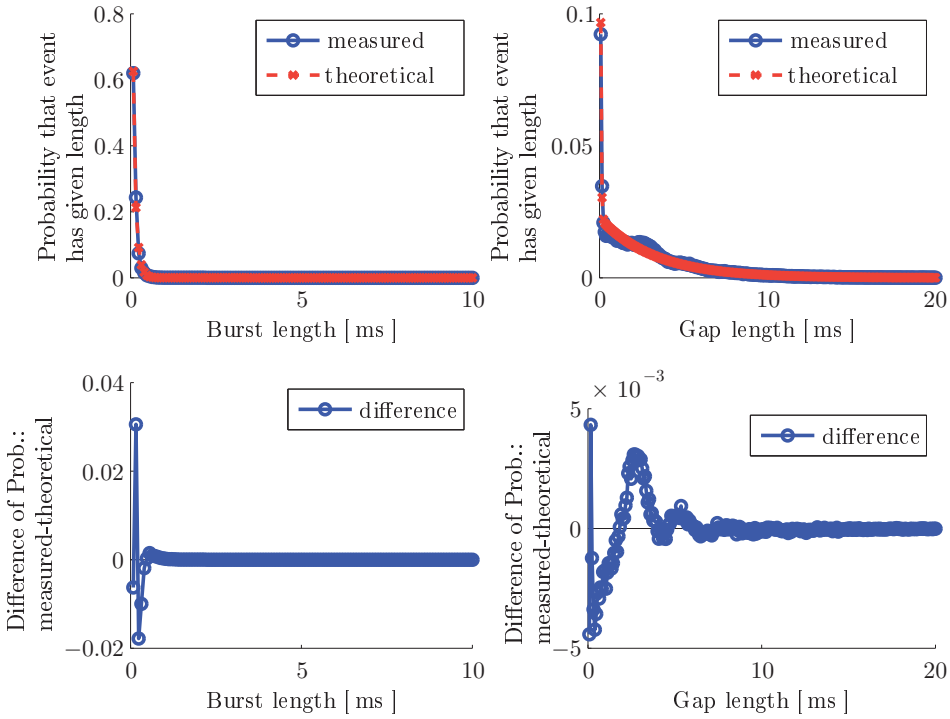


Figure E.6: Generalized Gilbert-Elliott model; WLAN, $SNR = 20\,\mathrm{dB}$

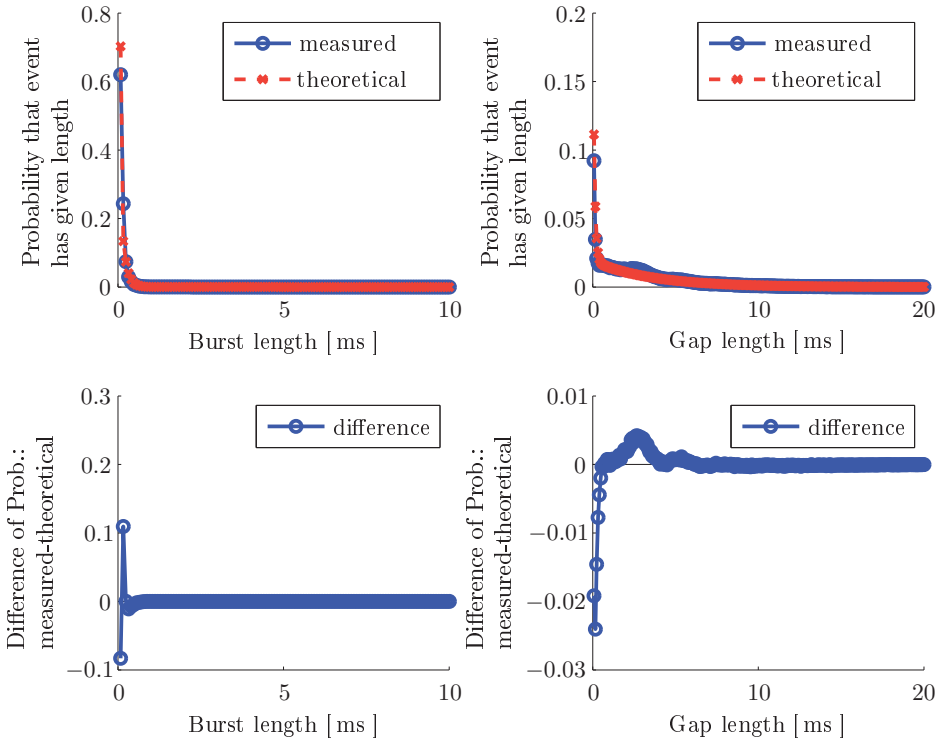


Figure E.7: 4-states model; $G_{\min} = 5$; WLAN, $SNR = 20\,\mathrm{dB}$

Model	Burst lengths			Gap lengths		
	N	q	$\chi^2_{0.95}(N-1)$	N	q	$\chi^2_{0.95}(N-1)$
$T_{\mathrm{TTI}} = 0.08\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.08\,\mathrm{ms}$						
Simplified Gilbert	27	416214.55	38.89	434	37575.62	482.51
Gilbert-Elliott	27	70709.24	38.89	434	3900.55	482.51
$T_{\mathrm{TTI}} = 5\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.15\,\mathrm{ms}$						
Simplified Gilbert	5	176.48	9.49	145	1361.50	173.00
Gilbert-Elliott	5	7.52	9.49	145	179.01	173.00
$T_{\mathrm{TTI}} = 5\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.2\,\mathrm{ms}$						
Simplified Gilbert	5	893.45	9.49	129	4174.20	155.40
Gilbert-Elliott	5	7.01	9.49	129	334.06	155.40
$T_{\mathrm{TTI}} = 10\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.2\,\mathrm{ms}$						
Simplified Gilbert	4	333.30	7.81	95	1942.98	117.63
Gilbert-Elliott	4	2.59	7.81	95	139.57	117.63
$T_{\mathrm{TTI}} = 10\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.31\,\mathrm{ms}$						
Simplified Gilbert	4	926.46	7.81	84	3842.29	105.27
Gilbert-Elliott	4	0.98	7.81	84	111.20	105.27
$T_{\mathrm{TTI}} = 20\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.31\,\mathrm{ms}$						
Simplified Gilbert	4	564.37	7.81	64	1891.05	82.53
Gilbert-Elliott	4	0.62	7.81	64	67.54	82.53
$T_{\mathrm{TTI}} = 20\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.52\,\mathrm{ms}$						
Simplified Gilbert	6	6926.44	11.07	44	5819.50	59.30
Gilbert-Elliott	6	1.57	11.07	44	52.35	59.30

Table E.4: χ^2 goodness of fit test: WLAN, $SNR = 20\,\mathrm{dB}$

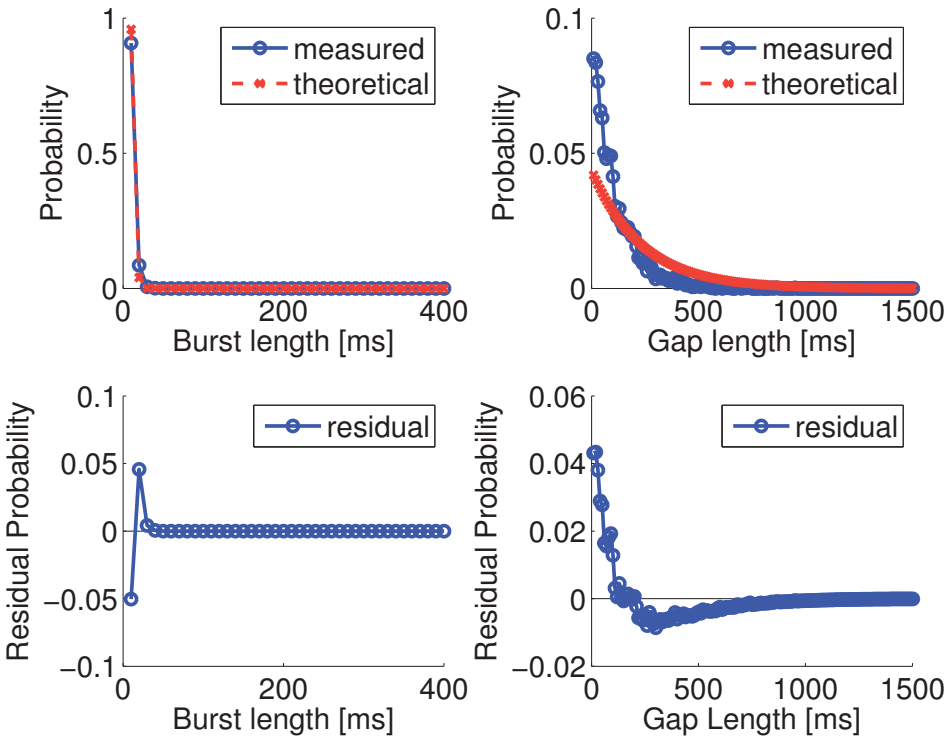


Figure E.8: Simplified Gilbert model; WLAN, $SNR = 20\,\mathrm{dB}$; $T_{\mathrm{TTI}} = 10\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.2\,\mathrm{ms}$

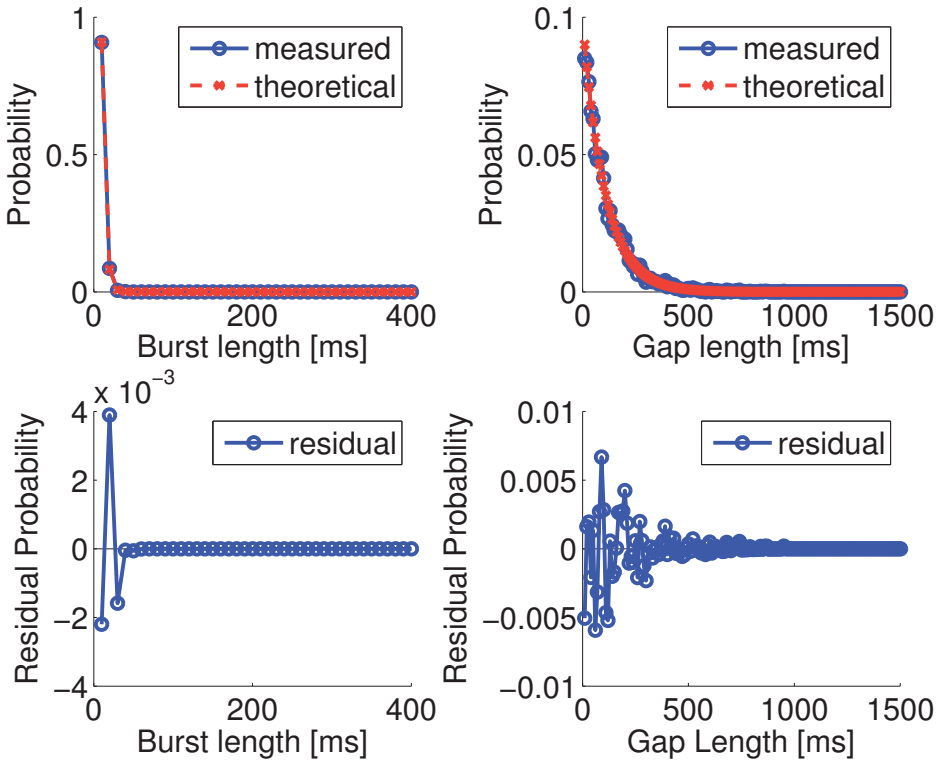


Figure E.9: Generalized Gilbert-Elliott model; WLAN, $SNR = 20\,\mathrm{dB}$; $T_{\mathrm{TTI}} = 10\,\mathrm{ms}$; $\tau_{\mathrm{p}} = 0.2\,\mathrm{ms}$

F

Probabilities of Specific Loss Patterns in the Extended Gilbert-Elliott Model

In this appendix, several assisting formulas are introduced which allow a more compact notation in the theoretical determination of residual loss distributions when applying forward error correction (FEC) schemes, as discussed in Chapter 4. All calculations assume an extended Gilbert-Elliott model as introduced in Chapter 3.

The determination of the error correction capabilities of different FEC schemes requires the probabilities of occurrence of specific loss events, i.e., loss patterns of successive packets with groups of successively lost packets, groups of successively received packets, and groups of packets that are arbitrarily lost or received. For example, consider the following pattern: $\rho = \{1\ x^2\ 0\}^3$. In this notation, x stands for a packet which is either received or lost, 0 stands for a received packet, and 1 for a lost packet. An exponent denotes that a specific part of a pattern or the whole pattern itself occurs a certain number of times in direct sequence. Hence, pattern $\{1\ x^2\ 0\}^3$ denotes a 3-fold occurrence of the pattern $\{1\ x^2\ 0\}$, which itself consists of a lost packet, followed by 2 packets that are each either lost or received, followed by a received packet. The probability of occurrence of this pattern in dependence on the channel states is denoted as $P_{XY}^{\mathrm{pat}}(\{1\ x^2\ 0\}^3)$, with $X, Y \in \{\mathrm{G}, \mathrm{B}\}$. X denotes the channel state for the first packet in the pattern, Y the channel state for the packet directly following the pattern. The states G and B are the two states of the Gilbert-Elliott model as introduced in Section 3.1.3, which has possibly been adapted from a base model to a specific packet transmission time interval and packet size according to Section 3.2.

In Appendix F.1, the probability of occurrence of a compound loss pattern like $\{1\ x^2\ 0\}$ is derived. The probability of a repeated occurrence of such a pattern, as in $\{1\ x^2\ 0\}^3$, is calculated in Appendix F.2. The derivations assume already calculated

probabilities of occurrence of m losses in n consecutive packets, $P_{XY}(m,n)$, as derived for the extended Gilbert-Elliott model in Section 3.4.

F.1 Probability of Compound Loss Patterns

In the following, the probability of occurrence of a compound loss pattern, consisting of lost, received, and arbitrarily lost or received packets, will be explained and illustrated with examples.

The probability of occurrence of a group of p successively lost packets is given as the probability of losing all p packets in a group of p successive packets as derived in Section 3.4:

$$P_{XY}^{\text{pat}}(\{1^p\}) = P_{XY}(p,p), \tag{F.1}$$

for a given state X at the first packet and a given state Y at the packet directly following the group, i.e., the $(p+1)$-th packet, with $X, Y \in \{\text{G}, \text{B}\}$.

In the same way, the probability of occurrence of a group of p successively received packets is given as the probability of losing no packets in a group of p successive packets:

$$P_{XY}^{\text{pat}}(\{0^p\}) = P_{XY}(0,p). \tag{F.2}$$

Finally, the probability of occurrence of an arbitrary pattern of p lost and/or received packets is calculated by the following sum of probabilities:

$$P_{XY}^{\text{pat}}(\{x^p\}) = \sum_{i=0}^{p} P_{XY}(i,p). \tag{F.3}$$

If the states X and Y shall be arbitrary, this probability results to one, i.e., $P^{\text{pat}}(\{x^p\}) = 1$.

Hence, the probabilities $P_{XY}^{\text{pat}}(\{1^p\})$, $P_{XY}^{\text{pat}}(\{0^p\})$, and $P_{XY}^{\text{pat}}(\{x^p\})$ describe probabilities of certain run lengths of lost, received, or arbitrary packets, respectively. The probability of occurrence of the event comprising two of these run lengths, i.e., the loss pattern consisting of the p-fold occurrence of s_1 directly followed by the q-fold occurrence of s_2, with $s_1, s_2 \in \{0, 1, x\}$ is calculated as

$$P_{XY}^{\text{pat}}(\{s_1^p\ s_2^q\}) = P_{X\text{G}}^{\text{pat}}(\{s_1^p\})\, P_{\text{G}Y}^{\text{pat}}(\{s_2^q\}) + P_{X\text{B}}^{\text{pat}}(\{s_1^p\})\, P_{\text{B}Y}^{\text{pat}}(\{s_2^q\}). \tag{F.4}$$

with $X, Y \in \{\text{G}, \text{B}\}$ and $s_1, s_2 \in \{0, 1, x\}$. Here, the state the channel is in at the $(p+1)$-th packet needs to be considered in the calculation, i.e., both possible states, G and B, need to be taken into account. Analogously, the probability of occurrence of any compound loss pattern consisting of several different run lengths of lost, received and arbitrary packets can be calculated by iteratively combining two patterns ρ_1 and ρ_2 according to the following general rule:

$$P_{XY}^{\text{pat}}(\{\rho_1\ \rho_2\}) = P_{X\text{G}}^{\text{pat}}(\rho_1)\, P_{\text{G}Y}^{\text{pat}}(\rho_2) + P_{X\text{B}}^{\text{pat}}(\rho_1)\, P_{\text{B}Y}^{\text{pat}}(\rho_2). \tag{F.5}$$

If the channel states X and Y are considered arbitrary, the following probability results:

$$P^{\mathrm{pat}}(\{\rho_1\ \rho_2\}) = \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P^{\mathrm{pat}}_{XY}(\rho_1)\, P^{\mathrm{pat}}_{YZ}(\rho_2). \tag{F.6}$$

The following examples show the calculation of probabilities of different loss patterns using the previously derived equations. The first example calculates the probability of receiving p successive packets and losing the following q packets. The channel is in an arbitrary state at the first packet and also at the packet directly following the pattern:

$$\begin{aligned} P^{\mathrm{pat}}(\{0^p\ 1^q\}) &= \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P^{\mathrm{pat}}_{XY}(\{0^p\})\, P^{\mathrm{pat}}_{YZ}(\{1^q\}) \\ &= \sum_{\substack{X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},X}\, P_{XY}(0,p)\, P_{YZ}(q,q). \end{aligned} \tag{F.7a}$$

The second example calculates the probability of losing q successive packets after p arbitrarily lost or received packets under the condition of being in state X at the first packet of the pattern. The channel state for the packet directly following the pattern shall be Y. The probability is calculated with (F.1) and (F.3) as

$$\begin{aligned} P^{\mathrm{pat}}_{XY}(\{x^p\ 1^q\}) &= P^{\mathrm{pat}}_{X\mathrm{G}}(\{x^p\})\, P^{\mathrm{pat}}_{\mathrm{G}Y}(\{1^q\}) + P^{\mathrm{pat}}_{X\mathrm{B}}(\{x^p\})\, P^{\mathrm{pat}}_{\mathrm{B}Y}(\{1^q\}) \\ &= \sum_{i=0}^{p} \Big(P_{X\mathrm{G}}(i,p)\, P_{\mathrm{G}Y}(q,q) + P_{X\mathrm{B}}(i,p)\, P_{\mathrm{B}Y}(q,q) \Big). \end{aligned} \tag{F.7b}$$

The third and last example calculates the probability of losing p packets, followed by q packets which are each either lost or received, and finally a group of r consecutive packets which are all lost. The state of the channel at the first packet of the pattern and also at the packet directly following the pattern is arbitrary:

$$\begin{aligned} P^{\mathrm{pat}}(\{1^p\ x^q\ 1^r\}) &= \sum_{\substack{W,X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},W}\, P^{\mathrm{pat}}_{WX}(\{1^p\})\, P^{\mathrm{pat}}_{XY}(\{x^q\})\, P^{\mathrm{pat}}_{YZ}(\{1^r\}) \\ &= \sum_{\substack{W,X,Y,Z \\ \in\{\mathrm{G,B}\}}} P_{\mathrm{s},W}\, P_{WX}(p,p) \Big(\sum_{i=0}^{q} P_{XY}(i,q) \Big) P_{YZ}(r,r). \end{aligned} \tag{F.7c}$$

F.2 Probability of a Repeated Occurrence of a Specific Pattern

Assume an arbitrary pattern ρ of lost, received, and arbitrary packets for which the probabilities of occurrence $P^{\mathrm{pat}}_{XY}(\rho)$, $X, Y \in \{\mathrm{G}, \mathrm{B}\}$, have been calculated as derived

in the previous section. For such an arbitrary pattern of lost and received packets, the probability of an n-fold consecutive occurrence of this pattern, denoted as ρ^n, can be calculated recursively taking the channel states at the first packets of each occurrence of ρ into account:

$$P_{XY}^{\mathrm{pat}}(\rho^n) = P_{X\mathrm{G}}^{\mathrm{pat}}(\rho)\, P_{\mathrm{G}Y}^{\mathrm{pat}}(\rho^{n-1}) + P_{X\mathrm{B}}^{\mathrm{pat}}(\rho)\, P_{\mathrm{B}Y}^{\mathrm{pat}}(\rho^{n-1}), \tag{F.8}$$

for $n > 1$ and with $X, Y \in \{\mathrm{G}, \mathrm{B}\}$. For $n = 1$, the equation therefore reverts to the probability of the pattern itself, i.e., $P_{XY}^{\mathrm{pat}}(\rho^1) = P_{XY}^{\mathrm{pat}}(\rho)$.

G

Concatenation of Channel Models

In a transmission scenario of a heterogeneous network, the packets may pass several transmission links of different characteristics. In a majority of cases, the transmission can be separated into three network parts. First, there is the uplink transmission from the user device into the network of the service provider. This can be a LAN connection, a WLAN transmission to an access point, or a wireless transmission in a mobile communication network like UMTS. Then there is the transmission within the core network of the service provider, possibly transferred via gateways into the core network of the communication partner's service provider. Finally, the downlink transmission to the user device of the communication partner may again be a wireless link of a mobile network, a WLAN access, or a LAN connection.

Such an end-to-end transmission scenario in a heterogeneous network can be modeled by a single Gilbert-Elliott model, trained from measurements or simulations of the considered end-to-end transmission. Alternatively, an end-to-end model can be derived from models of some network parts which may already be available, e.g., a model of the wireless access channel and another model for the core network. The procedure of combining several models into a realistic model of the end-to-end transmission requires the same time base for every model part, i.e., the same transmission time interval and the same packet size. If necessary, the models can be adapted as explained in Section 3.2 to meet this precondition. The different model parts are then concatenated. Depending on the type of involved models, this may be achieved by a straightforward combination of the models into a new model of the same type, or it may involve constructing a new and more complex model. The following sections will consider the concatenation of different model types, which reflect the most relevant network scenarios:

- Wireless access (Gilbert-Elliott model) and core network (Bernoulli model)

- Wireless uplink (Gilbert-Elliott model), core network (Bernoulli model), wireless downlink (Gilbert-Elliott model)

The described scenarios consider always two channel models which are concatenated. The concatenation of three channel parts can be calculated stepwise by first concatenating two models and then concatenating the resulting model with the third.

G.1 Concatenation of Bernoulli Models

The concatenation of two channels models with independent packet losses of probabilities $P_{\mathrm{e}}^{(1)}$ and $P_{\mathrm{e}}^{(2)}$ results again in a Bernoulli model with the following error probability:

$$P'_{\mathrm{e}} = P_{\mathrm{e}}^{(1)} + P_{\mathrm{e}}^{(2)} - P_{\mathrm{e}}^{(1)} P_{\mathrm{e}}^{(2)}. \tag{G.1}$$

G.2 Concatenation of Simplified Gilbert Model and Bernoulli Model

The concatenation of a simplified Gilbert channel and a Bernoulli channel, i.e., a channel with independent packet losses of probability P_{e}, is achieved by adapting the state transition probabilities of the simplified Gilbert model:

$$P'_{\mathrm{t,GB}} = P_{\mathrm{t,GB}} + P_{\mathrm{e}} - P_{\mathrm{t,GB}} \cdot P_{\mathrm{e}}, \tag{G.2a}$$
$$P'_{\mathrm{t,GG}} = P_{\mathrm{t,GG}} \left(1 - P_{\mathrm{e}}\right) = 1 - P'_{\mathrm{t,GB}} \tag{G.2b}$$
$$P'_{\mathrm{t,BG}} = P_{\mathrm{t,BG}} \left(1 - P_{\mathrm{e}}\right) \tag{G.2c}$$
$$P'_{\mathrm{t,BB}} = P_{\mathrm{t,BB}} + P_{\mathrm{e}} - P_{\mathrm{t,BB}} \cdot P_{\mathrm{e}} = 1 - P'_{\mathrm{t,BG}}. \tag{G.2d}$$

G.3 Concatenation of Gilbert(-Elliott) Model and Bernoulli Model

In the scenario of a wireless access channel, modeled by a Gilbert model or a Gilbert-Elliott model (with loss probabilities $P_{\mathrm{e,G}}$ and $P_{\mathrm{e,B}}$), and a core network, modeled by a Bernoulli model with loss probability P_{e}, the concatenation of the models results in a Gilbert-Elliott model with adjusted loss probabilities in the two states:

$$P'_{\mathrm{e,G}} = P_{\mathrm{e,G}} + P_{\mathrm{e}} - P_{\mathrm{e,G}} \cdot P_{\mathrm{e}}, \tag{G.3a}$$
$$P'_{\mathrm{e,B}} = P_{\mathrm{e,B}} + P_{\mathrm{e}} - P_{\mathrm{e,B}} \cdot P_{\mathrm{e}}. \tag{G.3b}$$

The transition probabilities of the Gilbert-Elliott model remain unaffected.

G.4 Concatenation of Simplified Gilbert Models

Assume a transmission over two channel parts which are both modeled by a simplified Gilbert model. Let $P^{(1)}_{\mathrm{t},ij}$ be the state transition probabilities of the first model, and $P^{(2)}_{\mathrm{t},ij}$ those of the second model, with the states $(i,j) \in \{\mathrm{G},\mathrm{B}\}$. These two models can be concatenated and described with a single simplified Gilbert model with new state transition probabilities. In this new model, a packet is received if both contributing models are in state G, and a packet is lost if either one or both of the models are in state B. Hence, the state probabilities of the new model result to

$$P'_{\mathrm{s,G}} = P^{(1)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{s,G}}, \tag{G.4a}$$

$$P'_{\mathrm{s,B}} = P^{(1)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{s,B}} + P^{(1)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{s,G}} + P^{(1)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{s,B}}. \tag{G.4b}$$

The new model is in state G if both contributing models are in state G. Consequently, a state transition from G to B in the new model results if either one or both contributing models change to state B. The new model is in state B if one or both contributing models are in state B. The computation of the state transition probability from B to G in the new model therefore needs to take these three cases into account. The state transition probabilities of the new model then result to

$$P'_{\mathrm{t,GB}} = P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{t,GB}} + P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{t,GG}} + P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{t,GB}}, \tag{G.5a}$$

$$P'_{\mathrm{t,GG}} = 1 - P'_{\mathrm{t,GB}} = P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{t,GG}}, \tag{G.5b}$$

$$\begin{aligned} P'_{\mathrm{t,BG}} = \Big(& P^{(1)}_{\mathrm{s,G}} \cdot P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{t,BG}} + P^{(1)}_{\mathrm{s,B}} \cdot P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{t,GG}} \\ & + P^{(1)}_{\mathrm{s,B}} \cdot P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{t,BG}} \Big) \Big/ P'_{\mathrm{s,B}}, \end{aligned} \tag{G.5c}$$

$$P'_{\mathrm{t,BB}} = 1 - P'_{\mathrm{t,BG}}. \tag{G.5d}$$

G.5 Concatenation of Gilbert-Elliott Models

Consider a scenario where the end-to-end transmission involves two wireless access networks which are both modeled by independent Gilbert-Elliott models. The states of both models will most likely have different loss probabilities. An accurate description of the concatenated channel therefore requires a new model with four different states reflecting all possible state combinations of the contributing models. The increase of the number of states, however, leads to a significant increase of the computational complexity for the calculation of loss distributions and error correction capabilities. In the following, the concatenated model will therefore be approximated by a standard Gilbert-Elliott model with two states.

For this purpose it is assumed that the new model is in state G if both contributing models are in state G, and in state B when at least one of the contributing models is in state B. This assumption leads to the following state probabilities:

$$P'_{\mathrm{s,G}} = P^{(1)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{s,G}}, \tag{G.6a}$$

$$P'_{\mathrm{s,B}} = P^{(1)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{s,B}} + P^{(1)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{s,G}} + P^{(1)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{s,B}}. \tag{G.6b}$$

Consequently, the state transition probabilities compute as

$$P'_{\mathrm{t,GB}} = P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{t,GB}} + P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{t,GG}} + P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{t,GB}}, \tag{G.7a}$$

$$P'_{\mathrm{t,GG}} = 1 - P'_{\mathrm{t,GB}} = P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{t,GG}}, \tag{G.7b}$$

$$\begin{aligned} P'_{\mathrm{t,BG}} = \Big(& P^{(1)}_{\mathrm{s,G}} \cdot P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{t,BG}} + P^{(1)}_{\mathrm{s,B}} \cdot P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{t,GG}} \\ & + P^{(1)}_{\mathrm{s,B}} \cdot P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{t,BG}} \Big) \Big/ P'_{\mathrm{s,B}}, \end{aligned} \tag{G.7c}$$

$$P'_{\mathrm{t,BB}} = 1 - P'_{\mathrm{t,BG}}. \tag{G.7d}$$

The state and state transition probabilities are therefore calculated as for the concatenation of simplified Gilbert models (cf. Appendix G.4). When considering Gilbert-Elliott models, however, the loss probabilities for the states are not necessarily 0 and 1. If both contributing models are in state G, i.e., the new model is also in state G, the probability of a packet loss is defined by the respective probabilities of the two models. For the calculation of the loss probability in state B, the different possible state combinations of the contributing models have to be taken into account. Hence, the state dependent loss probabilities for the new model result in:

$$P'_{\mathrm{e,G}} = P^{(1)}_{\mathrm{e,G}} + P^{(2)}_{\mathrm{e,G}} - P^{(1)}_{\mathrm{e,G}} \cdot P^{(2)}_{\mathrm{e,G}} \tag{G.8a}$$

$$\begin{aligned} P'_{\mathrm{e,B}} = \Big(& P^{(1)}_{\mathrm{s,G}} \cdot P^{(2)}_{\mathrm{s,B}} \cdot \left(P^{(1)}_{\mathrm{e,G}} + P^{(2)}_{\mathrm{e,B}} - P^{(1)}_{\mathrm{e,G}} \cdot P^{(2)}_{\mathrm{e,B}} \right) \\ & + P^{(1)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{s,G}} \cdot \left(P^{(1)}_{\mathrm{e,B}} + P^{(2)}_{\mathrm{e,G}} - P^{(1)}_{\mathrm{e,B}} \cdot P^{(2)}_{\mathrm{e,G}} \right) \\ & + P^{(1)}_{\mathrm{s,B}} \cdot P^{(2)}_{\mathrm{s,B}} \cdot \left(P^{(1)}_{\mathrm{e,B}} + P^{(2)}_{\mathrm{e,B}} - P^{(1)}_{\mathrm{e,B}} \cdot P^{(2)}_{\mathrm{e,B}} \right) \Big) \Big/ P'_{\mathrm{s,B}}. \end{aligned} \tag{G.8b}$$

In case the models have been previously adapted for different packet sizes according to Section 3.2, the loss probabilities are given in the transition dependent form of (3.30). The loss probabilities of the concatenated model then also have to

be calculated in transition dependent form as

$$P'_{\mathrm{e,GG}} = P^{(1)}_{\mathrm{e,GG}} + P^{(2)}_{\mathrm{e,GG}} - P^{(1)}_{\mathrm{e,GG}} P^{(2)}_{\mathrm{e,GG}} \tag{G.9a}$$

$$\begin{aligned} P'_{\mathrm{e,GB}} = &\Big(P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{t,GB}} \cdot \left(P^{(1)}_{\mathrm{e,GG}} + P^{(2)}_{\mathrm{e,GB}} - P^{(1)}_{\mathrm{e,GG}} P^{(2)}_{\mathrm{e,GB}} \right) \\ &+ P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{t,GG}} \cdot \left(P^{(1)}_{\mathrm{e,GB}} + P^{(2)}_{\mathrm{e,GG}} - P^{(1)}_{\mathrm{e,GB}} P^{(2)}_{\mathrm{e,GG}} \right) \\ &+ P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{t,GB}} \cdot \left(P^{(1)}_{\mathrm{e,GB}} + P^{(2)}_{\mathrm{e,GB}} - P^{(1)}_{\mathrm{e,GB}} P^{(2)}_{\mathrm{e,GB}} \right) \Big) \Big/ P'_{\mathrm{t,GB}} \end{aligned} \tag{G.9b}$$

$$\begin{aligned} P'_{\mathrm{e,BG}} = &\Big(P^{(1)}_{\mathrm{s,G}} P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BG}} \cdot \left(P^{(1)}_{\mathrm{e,GG}} + P^{(2)}_{\mathrm{e,BG}} - P^{(1)}_{\mathrm{e,GG}} P^{(2)}_{\mathrm{e,BG}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,G}} P^{(2)}_{\mathrm{t,GG}} \cdot \left(P^{(1)}_{\mathrm{e,BG}} + P^{(2)}_{\mathrm{e,GG}} - P^{(1)}_{\mathrm{e,BG}} P^{(2)}_{\mathrm{e,GG}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BG}} \cdot \left(P^{(1)}_{\mathrm{e,BG}} + P^{(2)}_{\mathrm{e,BG}} - P^{(1)}_{\mathrm{e,BG}} P^{(2)}_{\mathrm{e,BG}} \right) \Big) \Big/ \left(P'_{\mathrm{s,B}} P'_{\mathrm{t,BG}} \right) \end{aligned} \tag{G.9c}$$

$$\begin{aligned} P'_{\mathrm{e,BB}} = &\Big(P^{(1)}_{\mathrm{s,G}} P^{(1)}_{\mathrm{t,GG}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BB}} \cdot \left(P^{(1)}_{\mathrm{e,GG}} + P^{(2)}_{\mathrm{e,BB}} - P^{(1)}_{\mathrm{e,GG}} P^{(2)}_{\mathrm{e,BB}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BB}} \cdot P^{(2)}_{\mathrm{s,G}} P^{(2)}_{\mathrm{t,GG}} \cdot \left(P^{(1)}_{\mathrm{e,BB}} + P^{(2)}_{\mathrm{e,GG}} - P^{(1)}_{\mathrm{e,BB}} P^{(2)}_{\mathrm{e,GG}} \right) \\ &+ P^{(1)}_{\mathrm{s,G}} P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BG}} \cdot \left(P^{(1)}_{\mathrm{e,GB}} + P^{(2)}_{\mathrm{e,BG}} - P^{(1)}_{\mathrm{e,GB}} P^{(2)}_{\mathrm{e,BG}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,G}} P^{(2)}_{\mathrm{t,GB}} \cdot \left(P^{(1)}_{\mathrm{e,BG}} + P^{(2)}_{\mathrm{e,GB}} - P^{(1)}_{\mathrm{e,BG}} P^{(2)}_{\mathrm{e,GB}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BB}} \cdot P^{(2)}_{\mathrm{s,G}} P^{(2)}_{\mathrm{t,GB}} \cdot \left(P^{(1)}_{\mathrm{e,BB}} + P^{(2)}_{\mathrm{e,GB}} - P^{(1)}_{\mathrm{e,BB}} P^{(2)}_{\mathrm{e,GB}} \right) \\ &+ P^{(1)}_{\mathrm{s,G}} P^{(1)}_{\mathrm{t,GB}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BB}} \cdot \left(P^{(1)}_{\mathrm{e,GB}} + P^{(2)}_{\mathrm{e,BB}} - P^{(1)}_{\mathrm{e,GB}} P^{(2)}_{\mathrm{e,BB}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BB}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BG}} \cdot \left(P^{(1)}_{\mathrm{e,BB}} + P^{(2)}_{\mathrm{e,BG}} - P^{(1)}_{\mathrm{e,BB}} P^{(2)}_{\mathrm{e,BG}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BG}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BB}} \cdot \left(P^{(1)}_{\mathrm{e,BG}} + P^{(2)}_{\mathrm{e,BB}} - P^{(1)}_{\mathrm{e,BG}} P^{(2)}_{\mathrm{e,BB}} \right) \\ &+ P^{(1)}_{\mathrm{s,B}} P^{(1)}_{\mathrm{t,BB}} \cdot P^{(2)}_{\mathrm{s,B}} P^{(2)}_{\mathrm{t,BB}} \cdot \left(P^{(1)}_{\mathrm{e,BB}} + P^{(2)}_{\mathrm{e,BB}} - P^{(1)}_{\mathrm{e,BB}} P^{(2)}_{\mathrm{e,BB}} \right) \Big) \Big/ \left(P'_{\mathrm{s,B}} P'_{\mathrm{t,BB}} \right) \end{aligned} \tag{G.9d}$$

If the loss probabilities of one of the contributing models are not given in transition dependent form (i.e., it has not been adapted to a new packet size), consider the following analogy between the transition and state dependent loss probabilities for this case prior to applying (G.9):

$$P_{\mathrm{e,GG}} = P_{\mathrm{e,BG}} = P_{\mathrm{e,G}} \tag{G.10a}$$

$$P_{\mathrm{e,GB}} = P_{\mathrm{e,BB}} = P_{\mathrm{e,B}} \tag{G.10b}$$

Note that the resulting state of the transition determines the loss probability, because the loss process in this work is defined as 'first state transition, then receive/loss determination'. The goodness of fit of this approximation compared to the accurate model can be assessed as explained in Appendix E.1.3.

If there are some further independent losses expected to occur in the core network, this Bernoulli model can be combined with the resulting Gilbert-Elliott model as explained in Appendix G.3.

H

Perceived Quality Assessment and Prediction

The overall quality as perceived by a user of a speech conversation or multimedia streaming application is determined by several factors which are specific to the application and depend on transmission parameters and network behavior.

For a speech conversation, the overall perceived quality involves not only the audible quality of the received signal, but also comprises conversation related factors such as intelligibility and interactivity. The latter is particularly sensitive to the experienced end-to-end transmission delay. The measure of interest is therefore the perceived quality of the conversation. For streaming services, e.g., of music signals, on the other hand, the perceived quality is mainly determined by the signal quality at the receiver as the effect of delay is not significant unless it exceeds 1-2 seconds.

The perceived signal quality depends on the utilized codec, which determines the base quality, and is further influenced by signal distortions, e.g., echo and background noise, as well as transmission errors and the utilized algorithms for error protection and recovery. In packet transmission, packet losses may lead to the loss of complete signal segments (frames), for which an estimation is regenerated by the packet/frame loss concealment algorithm at the receiver. The frequency and length of such segment losses and the capability of the concealment algorithm determine to a large extent the perceived quality in a packet transmission scenario.

Depending on the considered scenario, the determination of quality should be done in different ways, ranging from expensive subjective tests to standardized objective measurement algorithms. A general overview and classification of available standards for the assessment of the quality of speech and music signals as well as voice conversation is given in Appendix H.1.

The two central topics of this work, the optimal parameterization of transmission schemes and the development of PLC algorithms, both require an assessment of the resulting quality, although with different requirements on the measurement method. The developed PLC algorithms for speech signals can be evaluated and compared

using the speech quality measure PESQ (*P*erceptual *E*valuation of *S*peech *Q*uality), standardized by the ITU in [ITU-T Rec. P.862 2001]. PESQ is an intrusive quality measure since it requires a reference signal, which is available here in form of the error free speech signal. The optimization of the system parameterization, on the other hand, requires a quality measure which can predict the effect of different loss rates and distributions, and for conversational applications also the effect of the end-to-end delay, on the resulting quality. For such a prediction of the quality from system parameters, the ITU has developed the E-model [ITU-T Rec. G.107 2005].

The following sections give a short overview of the two quality measurement methods applied in this work, the PESQ algorithm and the ITU-T E-model. Since the ITU-T E-model standard [ITU-T Rec. G.107 2005] does not yet provide impairment factors describing the coding distortions of all modes of the Adaptive Multi-Rate (AMR) speech codec, these are derived in Appendix H.5 from PESQ measurements according to the methodology defined in [ITU-T Rec. P.834 2002].

H.1 Means of Assessing and Predicting the Perceived Quality

The available and standardized means for assessing and predicting the perceived quality of a speech or music signal can be classified into subjective listening tests, intrusive objective measurements, and non-intrusive objective measurements.

Subjective listening tests for evaluating speech quality are generally performed according to the guidelines given by the ITU in [P.800 1996]. The specific type of test depends on the effects under study. Listening only tests will be carried out to evaluate, e.g., the quality of a new speech codec or the performance of different packet loss concealment algorithms. If the objective of the test, however, is to evaluate the quality of a conversation also including delay or echo effects, for example, a conversational test needs to be conducted. In order to get statistically meaningful results, the tests need to be carried out with a sufficiently large group of subjects as well as presented stimuli, which makes these tests time consuming and expensive. Another disadvantage is that the results of such tests cannot be exactly reproduced.

In order to overcome the disadvantages of subjective tests, *objective measurement techniques* have been developed which avoid the need for extensive and expensive tests and provide results which are reproducible by others. Objective measurement techniques can be divided into two different classes, intrusive and non-intrusive techniques.

Methods for *intrusive quality measurement* are in general more accurate than non-intrusive methods. They require a reference signal to compare the signal under consideration with and utilize models of the human auditory system in order to evaluate the auditory difference between the two signals. For speech signals, the PESQ algorithm (*P*erceptual *E*valuation of *S*peech *Q*uality) [ITU-T Rec. P.862 2001] has become the most commonly used intrusive measurement method and a de-facto

standard. It is computationally complex and considers only distortions by coding or transmission errors. It does not consider quality degradation by delay effects. What PESQ is for speech signals, the PEAQ algorithm (*P*erceptual *E*valuation of *A*udio *Q*uality) tries to become for audio signals in general and especially music signals.

Although they are less accurate than intrusive methods, there is a demand for *non-intrusive quality measurement* techniques. These techniques do not require a reference signal and are therefore applicable in running systems, e.g., in order to predict the currently achieved signal quality. There are methods which monitor or predict the quality from available network or other system parameters (e.g. loss rate, etc.). A widely used representative of this class is the so-called E-model [ITU-T Rec. G.107 2005], which is described in more detail in Appendix H.3. Other methods try to predict the quality from the degraded signal itself, e.g. [ITU-T Rec. P.563 2004].

Main applications of such non-intrusive measurement methods are to monitor the voice quality in life calls by network operators in order to check whether the network performs as desired, or even to adaptively control and optimize the resulting quality of service (QoS) based on the predicted perceived quality.

H.2 Objective Speech Quality Evaluation with PESQ

For the objective measurement of the quality of a speech signal, the ITU has standardized an algorithm called PESQ[1] in [ITU-T Rec. P.862 2001]. The PESQ algorithm operates in the frequency domain and uses a model of the human auditory system to evaluate the quality of a speech signal in comparison to a reference signal. The PESQ software expects two input files, the undistorted reference speech file and the speech file that has to be rated. The resulting output value lies within a range of 1 (worst) to 4.5 (best) and reflects an estimation of the *Mean Opinion Score* (MOS) that would result from a listening test. These values are termed MOS-LQO and MOS-LQS, LQ for *listening quality*, and O and S indicating an *objective* and *subjective* value, respectively. For a detailed definition of the MOS terminology see [ITU-T Rec. P.800.1 2006].

The PESQ algorithm uses a psycho-acoustical model to estimate the rating of an average human listener. However, it can only evaluate those signal distortions which it has been designed for, i.e., especially coding distortions and transmission impairments like bit errors, as well as packet loss and packet loss concealment (in particular with CELP codecs). For a list of further factors for which PESQ has demonstrated acceptable accuracy, see [ITU-T Rec. P.862 2001]. It does not consider other effects of quality degradation such as delay effects.

[1] Perceptual Evaluation of Speech Quality

H.3 Prediction of Quality with the ITU-T E-Model

For assessing different types of quality impairments, including frame losses and delay, the ITU has standardized the E-model [ITU-T Rec. G.107 2005], a non-intrusive computational model for speech quality prediction. The model is based on different system parameters and network characteristics and does not require an actually transmitted signal and according reference signal. The E-model serves as a computational model for transmission planning of modern telecommunication networks and considers combined effects of different types of impairment occurring simultaneously in a connection. It is based on the principle assumption that individual transmission impairments can be transformed into "psychological factors", and that these factors are additive on a "psychological scale".

The E-model derives a unidimensional quality index from a matrix of network parameters. The model output is the 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 [ITU-T Rec. G.107 2005], the calculation of the rating factor becomes

$$R = 93.2 - I_d - I_{e,eff}\,, \tag{H.1}$$

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. A possible further factor, the advancement factor, which reflects any possible user toleration in the given circumstances and is added as positive number, has been neglected here.

H.3.1 Delay impairment factor I_d

The delay impairment factor is calculated as

$$I_d = I_{dte} + I_{dle} + I_{dd} \tag{H.2}$$

and consists of three terms: the talker echo delay impairment I_{dte}, depending on echo path loss and delay; the listener echo delay impairment I_{dle}, depending on round-trip delay on 4-wire loop; and the loss of interactivity due to an overall delay I_{dd}. I_{dte} and I_{dle} are not considered in this work and therefore set to 0. Of particular interest in the consideration of packet-based speech transmission is the last term, I_{dd}, which describes the loss of interactivity in a conversation due to an overall delay d. It is calculated as

$$I_{dd} = \begin{cases} 0 & ;d \leq 100 \text{ ms} \\ 25\left((1+X^6)^{\frac{1}{6}} - 3(1+\frac{X}{3}^6)^{\frac{1}{6}} + 2\right) & ;d > 100 \text{ ms} \end{cases} \tag{H.3}$$

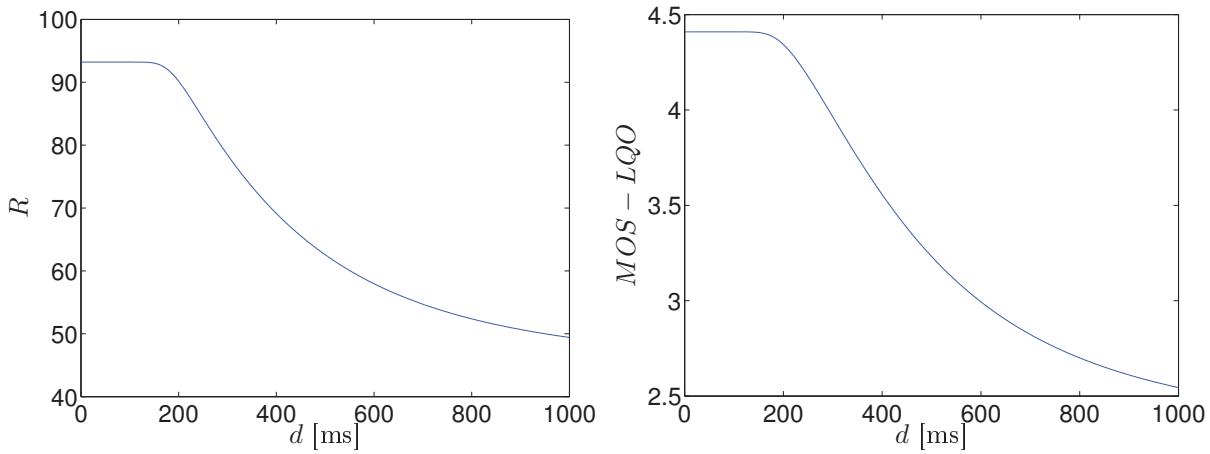


Figure H.1: Effect of the delay impairment factor I_d

with $X = \log_2\left(\frac{d}{100}\right)$. To save computational complexity, this relation can and should be stored as a lookup table with a resolution of a few milliseconds in a device's memory.

For the overall delay d, consider the total one-way mouth-to-ear delay, including algorithmic encoding/decoding delay, packetization delay, network delay (transmission, propagation, queuing), and de-jitter delay (receiver playout buffer).

The effect of the overall delay d on the resulting quality of the conversation, expressed in the rating factor R, is shown in Figure H.1. A clear threshold effect can be observed. Below a total delay of about 200 ms there is no quality degradation, above 200 ms the quality of a speech conversation starts to decrease considerably because the interactivity of the conversation gets affected.

H.3.2 Effective Equipment Impairment Factor $I_{e,eff}$

The effective equipment impairment factor $I_{e,eff}$, as defined in [ITU-T Rec. G.107 2005], extends the original equipment impairment factor, I_e, to reflect the impairment due to packet losses and concealment algorithms in addition to codec distortions.

The equipment impairment factor I_e [ITU-T Rec. G.113 2007] describes effects of digital processes other than pure PCM, i.e., low bit rate codecs. So far, only narrowband codecs are considered. It gives a relative degradation in comparison to other impairments occurring in a connection. This particular impairment factor is the framework for most common non-waveform codecs based on subjective listening-only tests. There is one drawback, tandems of multiple codecs of the same or different types have shown not to be simply additive. In some cases, order effects seem to apply. This is currently examined at ITU in more detail. The equipment impairment factors I_e for some speech codecs are given in Table H.1.

The frame loss dependent equipment impairment factor $I_{e,eff}$ includes the equipment impairment factor for codec distortions I_e and is defined in [ITU-T Rec. G.107

Codec	kbit/s	I_e	Codec	kbit/s	I_e
PCM (G.711)	64	0	G.729	8	10
ADPCM (G.726)	40	2	GSM FR	13	20
	32	7	AMR / EFR	12.2	5
	24	25	G.723.1	5.3	19
	16	50		6.3	15

Table H.1: Equipment impairment factors I_e for some speech codecs.

2005] as

$$I_{e,eff} = I_e + (95 - I_e) \cdot \frac{100 \cdot P_{\mathrm{fl}}}{\frac{100 \cdot P_{\mathrm{fl}}}{BurstR} + B_{\mathrm{pl}}}, \tag{H.4}$$

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.

H.3.3 Categories of Speech Quality and According Rating Factors

The resulting rating factor R of the E-model can be converted into an estimated MOS_{CQE} value as defined in [ITU-T Rec. G.107 2005].

The conversion from MOS_{CQE} to R is defined as:

$$\begin{array}{ll} R = 0 & \text{for } MOS_{\mathrm{CQE}} = 1 \\ R = \frac{20}{3}\left(8 - \sqrt{226}\cos\left(h + \frac{\pi}{3}\right)\right) & \text{for } 1 < MOS_{\mathrm{CQE}} < 4.5 \\ R = 100 & \text{for } MOS_{\mathrm{CQE}} \geq 4.5 \end{array} \tag{H.5}$$

with the helping terms

$$h = \frac{1}{3}\,\mathrm{arctan2}(y, x) \tag{H.6}$$

$$x = 18566 - 6750\, MOS_{\mathrm{CQE}} \tag{H.7}$$

$$y = 15\sqrt{-903522 + 1113960\, MOS_{\mathrm{CQE}} - 202500\, MOS_{\mathrm{CQE}}^2} \tag{H.8}$$

Accordingly, the conversion of R to MOS_{CQE} is achieved with

$$MOS_{\mathrm{CQE}} = 1 + 0.035\, R + R\,(R - 60)\,(100 - R) \cdot 7 \cdot 10^{-6} \tag{H.9}$$

R and MOS_{CQE} value can be categorized according to [ITU-T Rec. G.109 1999] into several categories of speech transmission quality as shown in Table H.2. [ITU-T Rec. G.109 1999] also gives some examples of typical scenarios: An ISDN connection would result in $R = 94$ (Best), while a PSTN connection would achieve $R = 82$ (High). A connection between a mobile and a PSTN subscriber would result in $R = 72$ at the mobile side (Medium) and $R = 64$ at the PSTN side (Low). Finally, a VoIP connection with the G.729A speech codec, 2% packet loss rate, and an end-to-end delay of 300 ms would result in $R = 55$ (Poor), i.e., unsatisfactory behavior.

R value range	MOS_{CQE} range	Quality category
$90 < R < 100$	$4.34 < MOS_{\mathrm{CQE}} < 4.5$	Best
$80 < R < 90$	$4.03 < MOS_{\mathrm{CQE}} < 4.34$	High
$70 < R < 80$	$3.60 < MOS_{\mathrm{CQE}} < 4.0$	Medium
$60 < R < 70$	$3.10 < MOS_{\mathrm{CQE}} < 3.60$	Low
$50 < R < 60$	$2.58 < MOS_{\mathrm{CQE}} < 3.10$	Poor

Table H.2: Categories of speech transmission quality: Range of R and MOS_{CQE}.

H.4 Deriving Equipment Impairment Factors from Instrumental Models

In general, there are two approaches specified by the ITU for the derivation of equipment impairment factors. The first methodology, described in [ITU-T Rec. P.833 2001], is based on the results of auditory listening-only tests. It is therefore able to describe the degradation as experienced by a human listener. However, it comes with the general disadvantages of subjective tests which are expensive and not completely reproducible. An alternative methodology which is based on instrumental models ("objective methods"), e.g. PESQ (cf. Appendix H.2), is given in [ITU-T Rec. P.834 2002]. This approach requires that the auditory models used in the instrumental tests provide valid estimations of auditory judgments for the considered codecs. The advantage of this method is that it is 100% reproducible if speech data and processing algorithms are precisely defined.

First, the set of speech files delivered with [ITU-T Rec. P.834 2002] is processed with the codec or algorithm for which the equipment impairment factor shall be determined. The quality of the processed files is then assessed by an objective quality measure and the resulting MOS-LQO values are transformed to rating factors R, e.g., as detailed in Appendix H.3.3 for PESQ. For each processed file, a preliminary equipment impairment factor K is then calculated as

$$K = R(\mathrm{G.711}) - R(\mathrm{processed\ file}), \tag{H.10}$$

with R(G.711) determined in the same way from the according PCM encoded file. The equipment impairment factor can then be calculated according to

$$I_e = \frac{K - b}{a}. \tag{H.11}$$

The coefficients a and b depend exclusively on the chosen instrumental model and can therefore be seen as a kind of "fingerprint" of this model, providing a reliable derivation of equipment impairment factors from its MOS-LQO values. The derivation of a and b are explained in [ITU-T Rec. P.834 2002]. For the PESQ quality measure, the following values have been standardized in [ITU-T Rec. P.834 2002]: $a = 0.5226$ and $b = 7.8502$.

AMR mode	MOS-LQO	I_e
12.2 kbit/s	4.031	5
10.2 kbit/s	3.921	9
7.95 kbit/s	3.754	15
7.4 kbit/s	3.724	16
6.7 kbit/s	3.616	20
5.9 kbit/s	3.503	23
5.15 kbit/s	3.360	27
4.75 kbit/s	3.270	29

Table H.3: Equipment impairment factors I_e (rounded) for the AMR modes obtained with PESQ according to the methodology in [ITU-T Rec. P.834 2002].

H.5 Deriving Equipment Impairment Factors for the AMR Speech Codec from PESQ Measurements

Provisional planning values for equipment impairment and packet loss robustness factors of different speech codecs are given in [ITU-T Rec. G.113 2007]. The studies in Chapter 5 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. Therefore, the proposed methodology from [ITU-T Rec. P.834 2002] has been utilized in this work to determine the equipment impairment factors for the other modes with the objective speech quality measure PESQ [ITU-T Rec. P.862 2001]. A more detailed overview of this methodology is given in Appendix H.4.

The equipment impairment factors for the 8 AMR modes have been determined according to the standardized procedure defined in [ITU-T Rec. P.834 2002] using the objective quality measurement tool PESQ [ITU-T Rec. P.862 2001]. The results are listed in Tab. H.3. The equipment impairment factor for the 12.2 kbit/s AMR mode exactly resulted in the value already defined in [ITU-T Rec. G.113 2007], i.e., $I_e = 5$. The values for the other modes could not be validated – this would require extensive listening tests – and have to be considered preliminary until official values are standardized by the ITU.

Also for the burst sensitivity factor B_{pl}, only the value for the 12.2 kbit/s AMR mode is given in [ITU-T Rec. G.113 2007]. This factor is codec dependent, i.e., it depends on interframe dependencies and the implemented PLC scheme. It is assumed that the standard concealment of the AMR codec is used. In lack of standardized values, it is further assumed that the same factor $B_{\mathrm{pl}} = 10$ applies for all modes of the AMR codec until more precise values are standardized. This is justifiable because the AMR codec modes have a similar general structure, apply the same frame loss concealment, and therefore also show similar effects of error propagation.

I

Deutschsprachige Kurzfassung

Kommunikationsnetze entwickeln sich mit hoher Geschwindigkeit zu sogenannten "all-IP" Netzwerken, die aus flexiblen Kernnetzen mit hoher Übertragungskapazität sowie verschiedenen Zugangsnetzen mit drahtgebundenen und drahtlosen Zugangstechnologien bestehen. Die Datenübertragung in diesen Kern- und Zugangsnetzen erfolgt durchgehend paketvermittelt auf der Basis von gemeinsamen, standardisierten Übertragungs- und Signalisierungsprotokollen, die auf der IP Protokollfamilie basieren. Die Paketvermittlung ermöglicht Netzbetreibern und unabhängigen Dienstanbietern die flexible Realisierung diverser Dienste und Anwendungen, wie z.B. Sprach-, Musik- und Videoübertragung, sowie weitere mobile Internetdienste wie z.B. Email, Web-Browsing, Instant Messaging, etc. Mit der Entwicklung neuer DSL- und Mobilfunk-Technologien steigt zudem die zur Verfügung stehende Übertragungsrate in den Zugangsnetzen deutlich an. In mobilen Zugangsnetzen wird die Entwicklung hin zu UMTS-HSDPA und UMTS-LTE einen nahezu allgegenwärtigen mobilen Zugang bieten, mit Datenraten von einigen hundert kbit/s bis hin zu mehreren Mbit/s in zukünftigen Systemen. Als Konsequenz aus dieser Entwicklung wird eine zunehmende Konvergenz von Fest- und Mobilfunknetzen entstehen, die dem Nutzer von unterschiedlichen Endgeräten und an nahezu jedem Ort Zugang zu allen seinen Diensten bieten wird.

Eine solche Konvergenz von Netzwerken und Diensten wird erst durch die paketvermittelte Übertragungstechnologie ermöglicht, die jedoch auch neue technische Herausforderungen an die Realisierung von Echtzeitdiensten wie Telephonie (Voice over IP), sowie Musik- und Video-Streaming stellt. Hauptprobleme sind eine variable Übertragungsverzögerung und Paketverluste, die z.B. durch Übertragungsfehler auf Mobilfunkkanälen oder Überlastung von Netzwerkknoten entstehen können. Bestehende Systeme setzen verschiedene Techniken ein, um diesen Übertragungsstörungen entgegen zu wirken:

- Die Varianz in der Übertragungsverzögerung (Jitter) wird in der Regel durch einen Empfangspuffer kompensiert, allerdings auf Kosten einer erhöhten Ende-zu-Ende Verzögerung.

- Verfahren zur Vorwärtsfehlerkorrektur (engl. *forward error correction*, FEC) werden zur Rekonstruktion von Signalrahmen bei Paketverlusten eingesetzt. Signalrahmen sind Segmente des ggf. codierten Signals mit einer normalerweise konstanten Länge.

- Durch FEC nicht rekonstruierbare Signalrahmen werden schließlich mit Hilfe von modellgestützten Methoden der Signalverarbeitung am Empfänger geschätzt (sogenanntes *Packet Loss Concealment*, PLC).

Ein noch nicht zufriedenstellend gelöstes Problem besteht in der analytischen Bestimmung der am besten geeigneten FEC Methode und ihrer Parametrierung (z.B. Verhältnis zwischen Nutz- und Paritätsinformation), sowie weiterer Systemparameter (z.B. Codec, Länge des Signalsegments per Paket, Länge des Empfangspuffers) für ein gegebenes Übertragungsszenario. Die optimale Wahl der Systemparameter hängt zum einen von den Anforderungen der Anwendung ab, z.B. hinsichtlich Signalqualität und tolerierbarer Verzögerung. Zum anderen wird die optimale Wahl der Parameter von der Charakteristik der Übertragungsfehler und der Übertragungsverzögerung auf dem Übertragungskanal bestimmt, der zudem gewissen Beschränkungen hinsichtlich der zur Verfügung stehenden Übertragungsrate unterliegt.

Hauptgegenstand der vorgelegten Dissertation ist die Untersuchung und Entwicklung von Ansätzen zur Optimierung der Sprach- und Audioübertragung in paketvermittelten Netzen mit drahtlosen Zugangstechnologien.

Im ersten Teil der Arbeit wird hierzu eine Methodik zur Auswahl der optimalen Übertragungsparameter und FEC Methoden entwickelt, die als Grundlage ein flexibles, erweitertes Gilbert-Elliott Kanalmodell für Paketverluste benutzt. Basierend auf diesem Modell werden analytische Berechnungen der verbleibenden Rahmenverlustrate für verschiedene FEC Techniken angestellt. Als Optimierungskriterium wird schließlich die erwartete Qualität der Anwendung für den Nutzer herangezogen, d.h. neben der Signalqualität wird für Kommunikationsdienste insbesondere auch die resultierende Signalverzögerung berücksichtigt. Die Anwendung der entwickelten Methodik wird anschließend am Beispiel verschiedener realistischer Szenarien demonstriert.

Bei drahtlosen Übertragungskanälen mit begrenzter Übertragungsrate kann das einsetzbare FEC Verfahren meist nicht alle verlorenen Signalrahmen rekonstruieren. Für solche Verluste wird im zweiten Teil der Arbeit ein neues verbessertes *Packet Loss Concealment* Verfahren für Sprachcodecs entwickelt, die auf dem CELP (*Code Excited Linear Prediction*) Codierverfahren basieren. Der vorgeschlagene Algorithmus bestimmt im Sender geeignete Nebeninformationen für jeden Rahmen, die im Payload des nachfolgenden Paketes zusammen mit dem folgenden Signalrahmen

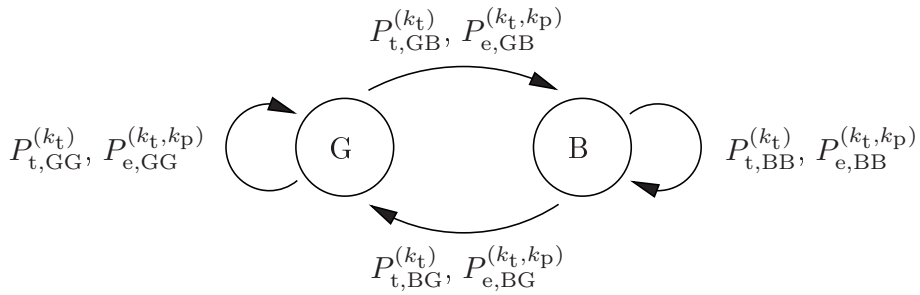


Abbildung I.1: Erweitertes Gilbert-Elliot Modell: Zustandsmodell mit zwei Zuständen: G (gerige Verlustrate) und B (höhere Verlustrate); Übergangswahrscheinlichkeiten $P_{\mathrm{t},XY}^{(k_\mathrm{t})}$ und übergangsabhängige Fehlerwahrscheinlichkeiten $P_{\mathrm{e},XY}^{(k_\mathrm{t},k_\mathrm{p})}$ $(X, Y \in \{\mathrm{G}, \mathrm{B}\})$ in Abhängigkeit von den Anpassungsfaktoren zum Basismodell für Übertragungsinterval und Paketgröße, k_t und k_p.

übertragen werden. Im Falle eines Paketverlustes hilft diese Information dem Empfänger dabei, die Codec-Parameter des verlorenen Rahmens zu schätzen und ein Ersatzsignal zu generieren. Für die effiziente Übertragung der Nebeninformation wird ein neues Konzept vorgestellt, das darauf basiert, die Nebeninformation mittels steganographischer Methoden in den codierten Bits des nachfolgenden Rahmens zu verstecken, und somit keine zusätzliche Datenrate erfordert.

Die wesentlichen Ergebnisse dieser Arbeit werden in den nachfolgenden Abschnitten zusammenfassend erklärt.

Erweitertes Gilbert-Elliott Kanalmodell für heterogene Paketnetzwerke

Die optimale Wahl der Übertragungsparameter für eine paketvermittelte Übertragung von Sprach- und Musiksignalen erfordert ein zuverlässiges Modell der Fehlercharakteristik des betrachteten Übertragungskanals. Zu diesem Zweck werden verschiedene Modelle für drahtlose Paketübertragungskanäle im Bezug auf ihre Eignung untersucht. Verschiedene Kanalsimulationen und -messungen zeigen, dass für drahtlose Übertragungskanäle die resultierende Paketverlustrate für eine bestimmte Anwendung von der Größe und dem Übertragungsintervall der übertragenen Pakete abhängt. Für größere Pakete ist die Wahrscheinlichkeit höher, Restbitfehler zu enthalten und somit verworfen zu werden. Die Wahl einer kürzeren Länge des übertragenen Signalrahmens je Paket führt zwar zu kleineren Paketen und somit weniger Verlusten, allerdings werden dann mehr Pakete pro Zeiteinheit übertragen, was aufgrund der benötigten Paket-Header in einer höheren Datenrate resultiert. Diese Abhängigkeit zwischen der Verlustwahrscheinlichkeit und der Größe und Frequenz der übertragenden Pakete muss daher beim Vergleich verschiedener FEC Methoden im Rahmen der Optimierung berücksichtigt werden. Dies erfordert ein geeignetes Kanalmodell, das sich an die jeweiligen Paketgrößen und Übertragungsintervalle

anpassen lässt.

Die Untersuchungen in diesem Kapitel zeigen, dass sich das verallgemeinerte Gilbert-Elliott Modell zwar grundsätzlich gut als Basismodell für Paketverluste auf drahtlosen Übertragungskanälen eignet, sich jedoch in seiner Standardform nicht an verschiedene Paketgrößen anpassen lässt und daher eine Modifikation erfordert. Zur Anpassung dieses Modells an verschiedene Paketgrößen werden neue Berechnungen angestellt, die schließlich zu einem erweiterten Modell führen, bei dem die Verlustwahrscheinlichkeit vom jeweiligen Zustandsübergang abhängt (siehe Abbildung I.1). Das resultierende, erweiterte Gilbert-Elliott Modell stellt im nachfolgenden Abschnitt der Arbeit die Basis für die analytischen Berechnungen der Fehlerkorrektureigenschaften von FEC Verfahren dar und sorgt für die erforderliche Vergleichbarkeit der verschiedenen Techniken. Als Vorbedingung für die Adaptierbarkeit des Modells ist lediglich erforderlich, dass das Basismodell eine ausreichend hohe Auflösung besitzt, d.h., für kleine Paketgrößen und kurze Übertragungsintervalle trainiert sein muss. Dies muss bei der Messung oder Simulation des Übertragungskanals berücksichtigt werden.

Analyse von Verfahren zur Vorwärtskorrektur auf Anwendungsebene

In heterogenen Netzwerken haben Anwendungen normalerweise keinen direkten Einfluss auf die in der physikalischen Übertragungsschicht verwendeten Kanalcodierungsalgorithmen. Dies gilt insbesondere für heterogene Netzwerke mit verschiedenen Übertragungskanälen entlang des Übertragungspfades, einschließlich drahtloser und mobiler Zugangstechnologien. In solchen Netzwerken spielt die Nutzung von zusätzlichen Fehlerschutzmechanismen auf Anwendungsebene eine entscheidende Rolle für die Kontrolle der Qualität von Diensten der Sprach- und Musikübertragung.

In der vorliegenden Dissertation werden verschiedene FEC Techniken untersucht, die in realen Systemen zu diesem Zweck zum Einsatz kommen: Reed-Solomon (RS) Block-Codes, Exklusiv-Oder (XOR) Verknüpfung von Signalrahmen sowie Rahmenwiederholung. Für jedes Verfahren werden neue analytische Berechnungen zur Rate und Verteilung der verbleibenden Verluste an Signalrahmen nach Auslöschungskorrektur durch das FEC Verfahren angestellt. Die resultierenden Wahrscheinlichkeitsfunktionen hängen von den Parametern des eingeführten Kanalmodells ab (einschließlich der Adaption an die entsprechende Paketgröße und das resultierende Übertragungsintervall), sowie von der Rahmenlänge des Codecs, den Parametern des FEC Verfahrens selber und der betrachteten Paketierungsstrategie. Zwei unterschiedliche Paketierungsstrategien werden für die Übertragung der FEC Rahmen betrachtet: Übertragung in separaten Paketen, also als unabhängiger

Paketstrom, oder gemeinsam mit einem folgenden Signalrahmen in einem nachfolgenden Paket (sogenanntes *piggybacking*). Die theoretischen Betrachtungen der Fehlerkorrektureigenschaften ermöglichen einen fairen Vergleich verschiedener Fehlerschutzverfahren und bilden somit die Basis für eine optimale Systemparametrierung.

Systemoptimierung für Sprach- und Musikübertragung in Paketnetzwerken

In diesem Abschnitt der Arbeit werden das erweiterte Kanalmodell und die analytischen Berechnungen der Restfehlerwahrscheinlichkeiten für verschiedene FEC Techniken auf praktische Anwendungs- und Übertragungsszenarien angewandt: Musik-Streaming über Wireless LAN und Sprachkommunikation über WLAN, UMTS und heterogene Paketnetzwerke.

Multicast Musik-Streaming über Wireless LAN

Für das Multicast-Streaming von MP3 codierten Musiksignalen zeigt sich, dass die Verwendung von FEC im Bezug auf die benötigte Übertragungsrate als deutlich effizienter anzusehen ist als ein erneutes Übertragen verlorener Pakete (außer wenn nur eine kleine Zahl an Empfängern betrachtet wird). Die optimale Wahl und Parametrierung des FEC Verfahrens selber hängt von der beobachteten Kanalqualität ab. Da die Signalverzögerung für eine Streaming Anwendung nicht von entscheidender Bedeutung ist (etwa 2 Sekunden können toleriert werden), orientiert sich die Optimierung alleine an der empfangsseitigen Signalqualität, die von der resultierenden Verlustrate bestimmt wird. Ein systematischer Reed-Solomon (RS) Blockcode mit geeigneter Coderate und Blocklänge zeigt sich unter den untersuchten Verfahren als am besten geeignet, eine geringe Rahmenverlustrate bei gleichzeitig guter Datenrateneffizienz zu erzielen. Für den betrachteten WLAN Kanal ist z.B. ein (8,4) RS-Code in der Lage, bei einem SNR von 20-25 dB eine Restfehlerrate von nahezu 0% zu erreichen. Bei einem geringeren SNR von nur 15-20 dB auf dem Kanal ist hierfür ein (8,2) RS-Code erforderlich. Die schwache Anforderung an die Verzögerung kann durch die Wahl des RS-Codes mit relativ großer Blocklänge und somit auch Verzögerung zur Erzielung einer geringen Verlustrate und daraus resultierend hohen Signalqualität genutzt werden.

Voice over IP über Wireless LAN

Sprachkommunikationsanwendungen wie Voice over IP (VoIP) verlangen eine deutlich geringere Ende-zu-Ende Verzögerung als Musik-Streaming, die 300 ms nicht überschreiten sollte. Eine zu hohe Verzögerung würde die Interaktivität zwischen den Kommunikationspartnern behindern und hätte dadurch einen direkten negativen Einfluss auf die empfundene Konversationsqualität. Für eine VoIP Verbindung

über WLAN erweist sich die im WLAN Standard spezifizierte automatische erneute Übertragung von verlorenen Datenblöcken in Übertragungsschicht 2 als sehr effizient für die Rückgewinnung verlorener Pakete. Da eine solche Neuübertragung die Signalverzögerung erhöht, sollten jedoch die Mechanismen zur Qualitätssteigerung gemäßIEEE 802.11e Standard angewandt werden, die eine höhere Priorität der zu wiederholenden Pakete sicherstellen. Die optimale Anzahl der Übertragungsversuche für ein einzelnes Paket zeigt sich als abhängig vom Kanal-SNR und der übertragenen Rahmenlänge je Paket bei Verwendung des PCM Sprachcodecs. Bei einem SNR von 20 dB auf dem betrachteten WLAN Kanal und einer Rahmenlänge von 5 ms je Paket ergibt sich eine Restfehlerrate von etwa 6%, wenn keine wiederholte Übertragung erlaubt ist. Für größere Rahmenlängen von 20-30 ms steigt die Verlustrate schnell auf 11-14%. Ein zweifacher Versuch einer Neuübertragung nach Paketverlust reduziert die Rahmenverlustrate zu nahezu 0% für alle betrachteten Rahmenlängen von 5-30 ms. Die zusätzliche Anwendung von FEC Techniken zum Schutz gegen Paketverluste von Ende zu Ende ist nur dann erforderlich, wenn die Pakete noch zusätzlich über ein Anschlussnetz mit einer beträchtlichen Paketverlustrate übertragen werden. In der Arbeit wird gezeigt, dass für ein solches Szenario der optimale Ansatz darin besteht, das FEC Verfahren auf die erwarteten Paketverluste allein im Anschlussnetz zu entwerfen und gleichzeitig die schnellen Neuübertragungen im drahtlosen Zugangsnetz zu verwenden.

Voice over IP über UMTS Paketkanäle

Die Übertragung von VoIP Diensten über UMTS Paketkanäle, die eine geringere Übertragungsrate als WLAN Kanäle besitzen, erfordert die Nutzung eines Sprachcodecs zur Datenkompression, z.B. den *Adaptive Multi Rate* (AMR) Codec. Eine Komprimierung der Paket-Header mit geeigneten Standardverfahren (z.B. ROHC nach IETF RFC 5795) ist häufig zusätzlich erforderlich. Die Beschränkung der Übertragungskapazität lässt nicht viel Raum für eine zusätzliche Anwendung von FEC Verfahren. Die Strategie, die daher für dieses Szenario als sinnvoll eingeschätzt wird, besteht darin, die Rate des Sprachcodecs zu verringern und somit Platz zu schaffen für einen zusätzlichen Fehlerschutz. Zu diesem Zweck wird der AMR Codec verwendet, ein Multiratencodec mit 8 verschiedenen Codierraten von 4.75 kbit/s bis 12.2 kbit/s. Durch Wahl eines Modus mit geringerer Codierrate ist es möglich, zusätzliche FEC Rahmen im selben Paket zu übertragen und dabei die Paketgröße und die erforderliche Übertragungsrate konstant zu halten. In Abhängigkeit von der Fehlercharakteristik auf dem Kanal kann somit ein optimaler Kompromiss zwischen der Grundqualität des Sprachsignals (bestimmt durch den verwendeten Codiermodus) und der Fehlerrobustheit erreicht werden, die durch das verwendete FEC Verfahren bestimmt wird. Es wird gezeigt, dass für eine bestmögliche Konversationsqualität die folgenden Kriterien gemeinsam betrachtet werden müssen, um die optimale Parametrierung für den betrachteten Kanal zu bestimmen:

- die Signalgrundqualität,

- die resultierende Verteilung an Rahmenverlusten und
- die Gesamtsignalverzögerung, die entscheidend durch das FEC Verfahren mitbestimmt wird.

Der Einfluss dieser verschiedenen Kriterien auf die resultierende Qualität wird mit Hilfe des von der ITU standardisierten *E-Model* [ITU-T Rec. G.107 2005] bestimmt. Das *E-Model* ist ein Berechnungsmodell zur Vorhersage der empfunden Qualität, welches auf der prinzipiellen Annahme basiert, dass verschiedene Störfaktoren in psychologische Faktoren transformiert und nachfolgend auf einer psychologischen Skala addiert werden können.

Für den betrachteten UMTS Kanal ergibt die analytische Betrachtung, dass bei einer Paketverlustrate von mehr als 2% die optimale Gesamtqualität erzielt wird, wenn der AMR Codec statt im 12.2 kbit/s im 6.7 kbit/s Modus betrieben wird und in jedem Paket eine zusätzliche Kopie des codierten Rahmens übertragen wird, der zeitlich 3 Rahmenlängen zurück liegt. Mit diesem Verfahren können einzelne Paketverlustereignisse von bis zu 3 aufeinanderfolgenden Paketen ausgeglichen werden. Bei einer geringeren Paketverlustrate als 2% wird hingegen kein FEC benötigt. Die komplette Übertragungsrate sollte stattdessen für den Sprachcodec aufgewendet werden, der folglich im 12.2 kbit/s Modus betrieben werden sollte. Die wenigen Paketverluste können durch das entsprechende Standard *Packet Loss Concealment* Verfahren für den AMR Codec am Empfänger genügend gut verdeckt werden.

Der betrachtete UMTS Paketkanal ist von 3GPP explizit für paketbasierte Sprachdienste spezifiziert worden und nutzt die Turbo Kanalcodierungstechnik. Turbo Codes werden im Allgemeinen hauptsächlich für Dienste mit hohen Datenraten eingesetzt, da ihre Leistungsfähigkeit von der Länge des verwendeten Interleavers abhängt. Nichtsdestotrotz haben Studien gezeigt, dass Turbo Codes auch bei Rahmenlängen von nicht weniger als 100 Bits noch immer einen moderaten Gewinn gegenüber Faltungscodierung erzielen können (siehe z.B. [Lee et al. 2000]). Der Interleaver für den Turbo Code im UMTS Standard verwürfelt die Bits eines einzelnen Übertragungsblocks, ohne vorangegangene und nachfolgende Blöcke einzubeziehen. Der Interleaver verursacht daher keine zusätzliche Verzögerung, was die Nutzung von Turbo Codes auch für die Voice over IP Übertragung in UMTS ermöglicht.

Voice over IP über heterogene Netzwerke mit variablen Paketlaufzeiten

Eine weitere Einsatzmöglichkeit für FEC Verfahren zeigt sich bei der Betrachtung des Szenarios einer Voice over IP Übertragung über ein Paketnetzwerk mit beträchtlicher Varianz in der Paketübertragungslaufzeit. Vorwärtsfehlerkorrekturverfahren sind in der Lage, nicht nur gestörte und verlorene Rahmen wiederzugewinnen, sondern zudem auch einzelne Signalrahmen, die eine hohe Übertragungsverzögerung

erfahren haben und auf die der Empfänger nicht mehr warten kann. Diese Eigenschaft kann somit zur Reduzierung der Länge des Empfangspuffers genutzt werden, wodurch die Ende-zu-Ende Verzögerung reduziert und in Folge die Konversationsqualität erhöht wird. Die optimale Parametrierung von FEC Verfahren und Länge des Empfangspuffers muss gemeinsam bestimmt werden und wird in der Dissertation beispielhaft an verschiedenen Kanalqualitäten gezeigt.

Packet Loss Concealment mit Nebeninformation

Der Einsatz von FEC Techniken unterliegt in realen Systemen gewissen Beschränkungen, die z.B. durch die tolerierbare Verzögerung einer Anwendung oder die verfügbare Übertragungsrate auf dem Kanal bestimmt werden. Das einsetzbare FEC Verfahren ist daher zumeist nicht in der Lage, Rahmenverluste komplett ausschließen zu können. Im Empfänger wird daher ein leistungsfähiges Packet Loss Concealment (PLC) Verfahren benötigt, welches verlorene Signalrahmen schätzen und somit den Verlust verdecken kann. Im zweiten Teil der vorgelegten Dissertation werden zu diesem Zweck neue PLC Verfahren für CELP basierte Sprachcodecs entwickelt, die sich insbesondere für Mobilfunkkanäle eignen.

In einem ersten neuen Ansatz (1) wird die Schätzmethode für die Codec Parameter eines verlorenen Rahmens an die Stimmhaftigkeit des betrachteten Sprachsegmentes angepasst. Hierzu wird zunächst die Stimmhaftigkeit des vorangegangenen und des nachfolgenden Rahmens am Empfänger anhand des Pitchverlaufs innerhalb des Rahmens geschätzt. Nachfolgend wird für jeden einzelnen Parameter eine für den entsprechenden Übergang von stimmhaft-stimmhaft, stimmhaft-stimmlos, stimmlos-stimmhaft oder stimmlos-stimmlos geeignete Schätzmethode verwendet, z.B. Extrapolation, Interpolation oder Wiederholung des Parameters. Es wird gezeigt, dass dieses adaptive Verfahren eine deutliche Verbesserung zu Standardverfahren bietet, welche zumeist das Signal extrapolieren und gleichzeitig die Amplitude absenken, um mögliche Signalstörungen zu verschleiern.

Aufbauend auf diesem Ansatz wird ein weiteres Verfahren (2) entwickelt, das die Auswahl eines geeigneten Schätzverfahrens für die einzelnen Parameter jedes Rahmens bereits im Sender bestimmt und als Nebeninformation zusammen mit dem nachfolgenden Rahmen zum Empfänger überträgt. Im Vergleich zum vorherigen Verfahren wird hier eine nochmals deutlich verbesserte Schätzung erzielt, da aus einer Reihe von Verfahren das tatsächlich genaueste für den entsprechenden Rahmen am Empfänger verwendet wird. Die Leistungsfähigkeit der entwickelten Verfahren wird im Vergleich zum Standard Verfahren des AMR Codecs in Tabelle I.1 anhand der mit dem standardisierten PESQ Algorithmus [ITU-T Rec. P.862 2001] berechneten objektiven Signalqualität verdeutlicht. Verglichen werden verschiedene Szenarien mit unterschiedlichen Verlustraten und zufälliger Verteilung der Verluste (Szenario A) bzw. künstlicher Einschränkung der Verlustposition auf bestimmte

Packet Loss Concealment Verfahren	PESQ MOS-LQO für verschiedene Szenarien A-E				
	A: $P_{\text{fl}} = 9.8\%$ zufälliger Übergang	**B:** $P_{\text{fl}} = 10\%$ nur stimmlos-stimmlos	**C:** $P_{\text{fl}} = 3.2\%$ nur stimmlos-stimmhaft	**D:** $P_{\text{fl}} = 3.0\%$ nur stimmhaft-stimmlos	**E:** $P_{\text{fl}} = 8.6\%$ nur stimmhaft-stimmhaft
Standard Verfahren	2.52	3.04	2.96	3.44	2.61
Ansatz (1) – Stimmhaftigkeit	2.74	3.39	3.00	3.47	2.84
Ansatz (2) – Nebeninformation	2.95	3.37	3.28	3.65	3.00

Tabelle I.1: Vergleich der Packet Loss Concealment Verfahren für den AMR Codec, 12.2 kbit/s Modus: Mittleres MOS-LQO für verschiedene Simulationsszenarien (Spalten), d.h., verschiedene Verlustraten P_{fl} und künstliche Beschränkung auf bestimmte Übergänge der Stimmhaftigkeit. Der mittlere PESQ MOS-LQO Wert für fehlerfreie Übertragung (d.h. lediglich die Codecverzerrungen werden bewertet) beträgt für die verwendeten Sprachdateien 4.03.

Übergänge der Stimmhaftigkeit im Signal (Szenarien B bis E). In Szenario B sind z.B. ausschließlich Rahmen an stimmlos-stimmlos Übergängen im Signal verloren. Es zeigt sich, dass Ansatz (1) insbesondere bei stimmlos-stimmlos sowie stimmhaft-stimmhaft Übergängen die Qualität verbessert (höherer PESQ MOS-LQO Wert). Ansatz (2) steigert die Qualität bei stimmhaft-stimmhaft Übergängen noch weiter und sorgt zudem auch für eine deutliche Verbesserung bei stimmhaft-stimmlos und stimmlos-stimmhaft Übergängen.

Zur Übertragung der Nebeninformation in Ansatz (2) ist lediglich eine geringe zusätzliche Datenrate von 400-1300 bit/s erforderlich, je nachdem in welchem Umfang Nebeninformationen zur Korrektur des Schätzfehlers übertragen werden sollen. Dieser sendergestützte Ansatz kann daher als eine Zwischenlösung klassifiziert werden zwischen datenratenintensiven senderbasierten Verfahren, zu denen die im ersten Teil der Dissertation analysierten FEC Verfahren gehören, und rein empfangsbasierten Packet Loss Concealment Verfahren, die ohne zusätzliche Datenrate auskommen. Die entwickelte Methode ist daher insbesondere für drahtlose Übertragungskanäle mit beschränkter Datenrate geeignet.

Abschließend wird in der Arbeit ein spezieller Algorithmus zur Übertragung der Nebeninformation vorgestellt, der dafür keine zusätzliche Datenrate erfordert. Der Ansatz benutzt hierzu die steganographischen Methoden aus [Geiser and Vary 2008] für eine versteckte Übertragung der Nebeninformation in den codierten Bits des AMR Sprachcodecs. Es wird gezeigt, dass mit diesem Verfahren die Nebeninformation für das empfangsseitige Packet Loss Concealment auch über herkömmliche leitungsvermittelte Teile der Übertragungskette, wie z.B. einen entsprechenden

GSM oder UMTS Kanal, übertragen werden kann. Diese kann dann vom Empfänger extrahiert und genutzt werden, sofern er dies unterstützt. Unterstützt der Empfänger dies nicht, bleibt die Nebeninformation verborgen und wird vom Empfänger ignoriert. Zur Erzielung einer gewissen Robustheit gegenüber möglichen Bitfehlern in der Übertragung wird eine geeignete Kanalcodierung auf die Bits der Nebeninformation angewandt, z.B. ein BCH-Code zum Fehlerschutz und ein zusätzlicher CRC (*cyclic redundancy check*) zur Fehlererkennung. Die Parametrierung dieses Fehlerschutzes hängt dabei von der verfügbaren Datenrate innerhalb des versteckten Bitstroms von 2 kbit/s ab. Bei einer Datenrate für die Nebeninformation von 400-1300 bit/s stehen für den Fehlerschutz 700-1600 bit/s zur Verfügung.

Die entwickelten Verfahren für ein verbessertes *Packet Loss Concealment* stellen somit eine geeignete Ergänzung zu der im ersten Teil der Arbeit entwickelten Methodik zur Bestimmung der am besten geeigneten FEC Verfahren und Systemparameter dar.

BIBLIOGRAPHY

3GPP TR 25.993 (2008). Typical examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) supported by Universal Terrestrial Radio Access (UTRA).

3GPP TS 23.228 . IP Multimedia Subsystem (IMS); Stage 2.

3GPP TS 25.212 . Multiplexing and channel coding (FDD).

3GPP TS 25.322 . Radio Link Control (RLC) protocol specification.

3GPP TS 25.323 . Packet Data Convergence Protocol (PDCP) Specification.

3GPP TS 26.090 . Adaptive Multi-Rate (AMR) Speech Transcoding.

3GPP TS 26.091 . Substitution and muting of lost frames for Adaptive Multi Rate (AMR) speech traffic channels.

3GPP TS 26.190 . AMR Wideband speech codec; Transcoding Functions.

3GPP TS 26.191 . Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Error concealment of erroneous or lost frames.

Adrat, M. (2003). *Iterative Source-Channel Decoding for Digital Mobile Communications, PhD Thesis, RWTH Aachen University*, number 16 in *Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN) (Vary, P., ed.)*, Verlag Mainz in Aachen.

Agiomyrgiannakis, Y. and Stylianou, Y. (2005). Coding with Side Information Techniques for LSF Reconstruction in Voice Over IP, *Proc. of the Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, PA, USA.

Andersen, S., Duric, A., Astrom, H., Hagen, R., Kleijn, W. and Linden, J. (2004). Internet Low Bit Rate Codec (iLBC), RFC 3951.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statist.*, vol. 41, no. 1, pp. 164–171.

Berlekamp, E. (1968). *Algebraic Coding Theory*, New York, NY: McGraw-Hill Book.

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and Weiss, W. (1998). An Architecture for Differentiated Service, RFC 2475 (Informational). Updated by RFC 3260.

Bolot, J.-C. (1993). End-to-end packet delay and loss behavior in the Internet, *SIGCOMM '93: Conference Proceedings on Communications Architectures, Protocols and Applications*, ACM Press, New York, NY, USA, pp. 289–298.

Bolot, J.-C. (1995). Characterizing the End-To-End Behavior of the Internet: Measurements, Analysis, and Applications.

Bolot, J.-C., Crepin, H. and Garcia, A. V. (1995). Analysis of audio packet loss in the internet, *Network and Operating System Support for Digital Audio and Video*, pp. 154–165.

Bolot, J.-C., Fosse-Parisis, S. and Towsley, D. (1999). Adaptive FEC-based error control for Internet telephony, *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, New York, NY, vol. 3, pp. 1453–1460.

Bormann, C., Burmeister, C., Degermark, M., Fukushima, H., Hannu, H., Jonsson, L.-E., Hakenberg, R., Koren, T., Le, K., Liu, Z., Martensson, A., Miyazaki, A., Svanbro, K., Wiebke, T., Yoshimura, T. and Zheng, H. (2001). RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed, RFC 3095 (Proposed Standard). Updated by RFCs 3759, 4815.

Braden, R. (1989). Requirements for Internet Hosts - Communication Layers, RFC 1122 (Standard). Updated by RFCs 1349, 4379.

Braden, R., Clark, D. and Shenker, S. (1994). Integrated Services in the Internet Architecture: an Overview, RFC 1633 (Informational).

Braden, R., Zhang, L., Berson, S., Herzog, S. and Jamin, S. (1997). Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification, RFC 2205 (Proposed Standard). Updated by RFCs 2750, 3936, 4495.

Bruhn, S., Grancharov, V., Kleijn, W. B., Klejsa, J., Li, M., Plasberg, J., Pobloth, H., Ragot, S. and Vasilache, A. (2008). The FlexCode Speech and Audio Coding Approach, *ITG Fachtagung Sprachkommunikation*, Aachen.

Chen, Y.-L. and Chen, B.-S. (1997). Model-based multirate representation of speech signals and its application to recovery of missing speech packets, *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, pp. 220–231.

Chibani, M., Gournay, P. and Lefebvre, R. (2005). Increasing the Robustness of CELP-Based Coders By Constrained Optimization, *Proc. of the Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, PA, USA, vol. 1, pp. 785–788.

Chibani, M., Lefebvre, R. and Gournay, P. (2006). Resynchronization of the Adaptive Codebook in a Constrained CELP Codec After a Frame Erasure, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006*, vol. 1, pp. I–I.

Clark, A. (2001). Modeling the effects of burst packet loss and recency on subjective voice quality, *IPtel 2001 Workshop*.

Clevorn, T. (2006). *Turbo DeCodulation: Iterative Joint Source-Channel Decoding and Demodulation, PhD Thesis, RWTH Aachen University*, number 24 in *Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN) (Vary, P., ed.)*, Verlag Mainz in Aachen.

Clüver, K. and Noll, P. (1996). Reconstruction of Missing Speech Frames Using Subband Excitation, *Proc. International Symposium on Time-Frequency and Time-Scale Analysis*, Paris, pp. 277–280.

Comroe, R. and Costello, D., J. (1984). ARQ Schemes for Data Transmission in Mobile Radio Systems, *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 4, pp. 472–481.

Cottrell, L. (2009). ICFA SCIC Network Monitoring Report, *Technical report*, ICFA SCIC Monitoring Working Group.

Cox, R. V., de Campos Neto, S. F., Lamblin, C. and (Editors), M. H. S. (2009). ITU-T Coders for Wideband, Superwideband, and Fullband Speech Communication, *IEEE Communications Magazine*, vol. 47, no. 10, pp. 106–137.

D. De Vleeschauwer, J. Janssen, E. D. and Petit, G. (2000). Tolerable delay bounds for low bit rate voice transport, *Proceedings of the XVIIth World Telecommunications Congress, incorporating the International Switching Symposium 2000 (WTC/ISS2000)*, Birmingham, UK.

Deering, S. and Hinden, R. (1998). Internet Protocol, Version 6 (IPv6) Specification, RFC 2460 (Draft Standard).

Elliott, E. (1963). Estimates of error rates for codes on burst-noise channels, *The Bell*

System Technical Journal, vol. 42, pp. 1977–1997.

ETSI. Spec. GSM 06.60: Enhanced Full Rate (EFR) speech transcoding.

ETSI TIPHON TS 101 329-5 Annex E. QoS Measurements for Voice over IP.

ETSI TR 101 112 (1998). *Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03)*.

Fairhurst, G. and Wood, L. (2002). Advice to link designers on link Automatic Repeat reQuest (ARQ), RFC 3366 (Best Current Practice).

Fingscheidt, T. (1998). *Softbit-Sprachdecodierung in digitalen Mobilfunksystemen, PhD thesis, RWTH Aachen University*, number 9 in *Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN) (Vary, P., ed.)*, Verlag Mainz in Aachen. (in German).

Fingscheidt, T. and Perez, J. G. (2002). An Interpolative Decoding Approach for Speech Streaming Services and Voice Over IP, *Proceedings of 4th International ITG Conference on Source and Channel Coding*, Berlin.

Fingscheidt, T., Vary, P. and Andonegui, J. (1998). Robust speech decoding: can error concealment be better than error correction?, *Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of the 1998 IEEE International Conference on*, Seattle, WA, vol. 1, pp. 373–376.

Finlayson, R. (2008). A More Loss-Tolerant RTP Payload Format for MP3 Audio, RFC 5219 (Proposed Standard).

FlexCode (2009). Project within the European Commission's Sixth Framework Programme "Information Society Technologies", http://www.flexcode.eu.

Frossard, P. (2001). FEC performance in multimedia streaming, *IEEE Communications Letters*, vol. 5, no. 3, pp. 122–124.

Geiser, B. and Vary, P. (2008). High rate data hiding in ACELP speech codecs, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2008*, Las Vegas, Nevada, USA, pp. 4005–4008.

Geiser, B., Mertz, F. and Vary, P. (2008). Steganographic Packet Loss Concealment for Wireless VoIP, *ITG-Fachtagung Sprachkommunikation*, Aachen, Germany.

Gilbert, E. N. (1960). Capacity of a burst-noise channel, *The Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265.

Gournay, P., Rousseau, F. and Lefebvre, R. (2003). Improved packet loss recovery using late frames for prediction-based speech coders, *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, vol. 1, pp. 3498–3503.

Goyal, V. K. and Kovacevic, J. (1998). Optimal multiple description transform coding of gaussian vectors, *Proc. Data Compression Conference DCC '98*, pp. 388–397.

Grossman, D. (2002). New Terminology and Clarifications for Diffserv, RFC 3260 (Informational).

Gunduzhan, E. and Momtahan, K. (2001). Linear prediction based packet loss concealment algorithm for PCM coded speech, *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 778–785.

Hagen, R. (1994). Spectral quantization of cepstral coefficients, *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vol. I, pp. I/509–I/512.

Handley, M., Jacobson, V. and Perkins, C. (2006). SDP: Session Description Protocol, RFC 4566 (Proposed Standard).

Henkel, W. (1989). Another description of the Berlekamp-Massey algorithm, *IEE Proceedings I Communications, Speech and Vision*, vol. 136, no. 3, pp. 197–200.

Horn, U., Herzogenrath, E. E., Stuhlmuller, K., Link, M. and Girod, B. (1999). Robust Internet video transmission based on scalable coding and unequal error protection, *Signal Processing: Image Communication*, vol. 15, pp. 77–94.

IEEE Std 802.11 (2007). IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.

IEEE Std 802.11e (2005). Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications; Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, IEEE Std 802.11e-2005.

IEEE Std 802.16 (2004). IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems.

ISO/IEC 11172-3:1993 (1993). Information Technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio, ISO/IEC 11172-3:1993.

ISO/IEC 13818-7:2006 (2006). Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC), ISO/IEC 13818-7:2006.

ISO/IEC 7498-1:1994 (1994). Information Technology–Open Systems Interconnection–Basic reference model: The Basic Model, ISO/IEC 7498-1:1994.

ITU-T Rec. G.1020 (2006). Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks.

ITU-T Rec. G.107 (2005). The E-model, a computational model for use in transmission planning.

ITU-T Rec. G.109 (1999). Definition of categories of speech transmission quality.

ITU-T Rec. G.113 (2007). Transmission impairments due to speech processing.

ITU-T Rec. G.114 (2003). One-way transmission time.

ITU-T Rec. G.711 (1988). Pulse code modulation (PCM) of voice frequencies.

ITU-T Rec. G.711 Appendix I (1999). A high quality low-complexity algorithm for packet loss concealment with G.711.

ITU-T Rec. G.718 (2008). Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s, pre-published.

ITU-T Rec. G.722 (1988). 7 kHz audio-coding within 64 kbit/s.

ITU-T Rec. G.722 Appendix IV (2008). A low-complexity algorithm for packet loss concealment with G.722.

ITU-T Rec. G.726 (1990). 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).

ITU-T Rec. G.729 (1996a). Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP).

ITU-T Rec. G.729 (1996b). Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP).

ITU-T Rec. G.729.1 (2007). G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729.

ITU-T Rec. H.323 (2006). Packet-based multimedia communications systems.

ITU-T Rec. P.563 (2004). Single-ended method for objective speech quality assessment in narrow-band telephony applications.

ITU-T Rec. P.800.1 (2006). Mean Opinion Score (MOS) terminology.

ITU-T Rec. P.833 (2001). Methodology for derivation of equipment impairment factors

from subjective listening-only tests.

ITU-T Rec. P.834 (2002). Methodology for the derivation of equipment impairment factors from instrumental models.

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

ITU-T Rec. P.862.2 (2005). Wideband extension to Rec. P.862 for the assessment of wideband telephone networks and speech codecs (prepublished).

ITU-T Rec. X.200 (1994). Information technology - Open Systems Interconnection - Basic Reference Model: The basic model.

Jayant, N. and Christensen, S. (1981). Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure, *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 29, no. 2, pp. 101–109.

Jelinek, M. and Salami, R. (2007). Wideband Speech Coding Advances in VMR-WB Standard, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1167–1179.

Jiang, W. and Ortega, A. (2000). Multiple description speech coding for robust communication over lossy packet networks, *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, New York, NY, vol. 1, pp. 444–447.

Jiang, W. and Schulzrinne, H. (2002). Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss, *NOSSDAV '02: Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video*, ACM Press, New York, NY, USA, pp. 73–81.

Johansson, I., Frankkila, T. and Synnergren, P. (2002). Bandwidth efficient AMR operation for VoIP, *IEEE Workshop on Speech Coding*, Tsukuba, Ibaraki, Japan.

John S. Garofolo, e. a. (1993). TIMIT Acoustic-Phonetic Continuous Speech Corpus, Linguistic Data Consortium, Philadelphia.

Jonsson, L.-E. and Pelletier, G. (2004). RObust Header Compression (ROHC): A Compression Profile for IP, RFC 3843 (Proposed Standard). Updated by RFC 4815.

Jonsson, L.-E., Sandlund, K., Pelletier, G. and Kremer, P. (2007). RObust Header Compression (ROHC): Corrections and Clarifications to RFC 3095, RFC 4815 (Proposed Standard).

Karam, M. and Tobagi, F. (2001). Analysis of the delay and jitter of voice traffic over the Internet, *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, Anchorage, AK, vol. 2, pp. 824–833.

Kövesi, B. and Ragot, S. (2008). A low complexity packet loss concealment algorithm for ITU-T G.722, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2008*, pp. 4769–4772.

Kubin, G. and Kleijn, W. B. (1999). Multiple-description coding (MDC) of speech with an invertible auditory model, *Proc. IEEE Workshop on Speech Coding*, pp. 81–83.

Lahouti, F., Lahouti, F. and Khandani, A. K. (2007). Soft reconstruction of speech in the presence of noise and packet loss, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 44–56.

Larzon, L.-A., Degermark, M., Pink, S., Jonsson, L.-E. and Fairhurst, G. (2004). The Lightweight User Datagram Protocol (UDP-Lite), RFC 3828 (Proposed Standard).

Lee, L.-N., Hammons, A. R., J., Sun, F.-W. and Eroz, M. (2000). Application and standardization of turbo codes in third-generation high-speed wireless data services, *IEEE Transactions on Vehicular Technology*, vol. 49, no. 6, pp. 2198–2207.

Lefebvre, R., Philippe, G. and Salami, R. (2004). A study of design compromises for speech coders in packet networks, *Proc. of the Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, pp. 265–8.

Li, A. (2007). RTP Payload Format for Generic Forward Error Correction, RFC 5109 (Proposed Standard).

Liang, Y., Farber, N. and Girod, B. (2001). Adaptive playout scheduling using time-scale modification in packet voice communications, *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, Salt Lake City, UT, vol. 3, pp. 1445–1448.

Lin, S. and Costello, D. (2004). *Error Control Coding: Fundamentals and Applications*, Prentice Hall.

Linde, Y., Buzo, A. and Gray, R. (1980). An algorithm for vector quantizer design, *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95.

Martin, R., Hoelper, C. and Wittke, I. (2001). Estimation of missing LSF parameters using Gaussian mixture models, *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*, vol. 2, pp. 729–732.

Massey, J. (1969). Shift-register synthesis and BCH decoding, *IEEE Transactions on Information Theory*, vol. 15, no. 1, pp. 122–127.

Mertz, F. and Vary, P. (2006). Packet Loss Concealment with Side Information for Voice over IP in Cellular Networks, *ITG-Fachtagung Sprachkommunikation*, Kiel, Germany.

Mertz, F. and Vary, P. (2008). Efficient Wireless VoIP in Heterogeneous Packet Networks, *ITG-Fachtagung Sprachkommunikation*, Aachen, Germany.

Mertz, F., Engelke, U., Vary, P., Taddei, H. and Varga, I. (2005). Applicability of UDP-Lite for Voice over IP in UMTS Networks, *16th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2005)*, Berlin, Germany.

Mertz, F., Taddei, H., Varga, I. and Vary, P. (2003). Voicing Controlled Frame Loss Concealment for Adaptive Multi-Rate (AMR) Speech Frames in Voice-over-IP, *Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH)*, Geneva, Switzerland, pp. 1077–1080.

Moon, S. B., Kurose, J. and Towsley, D. (1998). Packet audio playout delay adjustment: performance bounds and algorithms, *Multimedia Systems*, vol. 6, pp. 17âĂŞ28.

Murthi, M. N., Rodbro, C. A., Andersen, S. V. and Jensen, S. H. (2006). Packet Loss Concealment with Natural Variations using HMM, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006*, vol. 1, pp. I–I.

OECD. OECD Broadband Portal, http://www.oecd.org/sti/ict/broadband.

P.800, I.-T. R. (1996). Methods for subjective determination of transmission quality.

Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables and Stochastic Processes*, 4 edn, McGraw Hill Higher Education.

Pelletier, G. (2005). RObust Header Compression (ROHC): Profiles for User Datagram Protocol (UDP) Lite, RFC 4019 (Proposed Standard). Updated by RFC 4815.

Pelletier, G. and Sandlund, K. (2008). RObust Header Compression Version 2 (ROHCv2): Profiles for RTP, UDP, IP, ESP and UDP-Lite, RFC 5225 (Proposed Standard).

Perkins, C., Hodson, O. and Hardman, V. (1998). A Survey of Packet Loss Recovery Techniques for Streaming Audio, *IEEE Network*, vol. 12, no. 5, pp. 40–48.

Perkins, C., Kouvelas, I., Hodson, O., Hardman, V., Handley, M., Bolot, J., Vega-Garcia, A. and Fosse-Parisis, S. (1997). RTP Payload for Redundant Audio Data, RFC 2198 (Proposed Standard).

Postel, J. (1980). User Datagram Protocol, RFC 768 (Standard).

Postel, J. (1981a). Internet Protocol, RFC 791 (Standard). Updated by RFC 1349.

Postel, J. (1981b). Transmission Control Protocol, RFC 793 (Standard). Updated by RFC 3168.

Pradhan, S. S., Puri, R. and Ramchandran, K. (2004). n-channel symmetric multiple descriptions - part i: (n, k) source-channel erasure codes, *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 47–61.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn, Cambridge University Press.

Puri, R., Pradhan, S. S. and Ramchandran, K. (2005). n-channel symmetric multiple descriptions-part ii:an achievable rate-distortion region, *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1377–1392.

Ramjee, R., Kurose, J., Towsley, D. and Schulzrinne, H. (1994). Adaptive playout mechanisms for packetized audio applications in wide-area networks, *Proc. IEEE INFOCOM '94. Networking for Global Communications. th*, pp. 680–688 vol.2.

Rey, J., Leon, D., Miyazaki, A., Varsa, V. and Hakenberg, R. (2006). RTP Retransmission Payload Format, RFC 4588 (Proposed Standard).

Rodbro, C. A., Murthi, M. N., Andersen, S. V. and Jensen, S. H. (2006). Hidden Markov model-based packet loss concealment for voice over IP, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1609–1623.

Rosen, E., Viswanathan, A. and Callon, R. (2001). Multiprotocol Label Switching Architecture, RFC 3031 (Proposed Standard).

Rosenberg, J., Qiu, L. and Schulzrinne, H. (2000). Integrating packet FEC into adaptive voice playout buffer algorithms on the Internet, *Proc. IEEE Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM 2000*, vol. 3, pp. 1705–1714 vol.3.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and Schooler, E. (2002). SIP: Session Initiation Protocol, RFC 3261 (Proposed Standard). Updated by RFCs 3265, 3853, 4320.

Sanneck, H., Stenger, A., Ben Younes, K. and Girod, B. (1996). A new technique for audio packet loss concealment, *Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity*, London, pp. 48–52.

Schmalen, L., Schlien, T. and Vary, P. (2010). The FlexCode Source-Channel Coding Approach for Audio and Speech Transmission, *Proceedings of International ITG Conference on Source and Channel Coding*, Informationstechnische Gesellschaft (ITG), VDE Verlag GmbH, Siegen, Germany.

Schroeder, M. and Atal, B. (1985). Code-excited linear prediction (CELP): High-quality speech at very low bit rates, *Proc. IEEE International Conference on ICASSP '85. Acoustics, Speech, and Signal Processing*, vol. 10, pp. 937–940.

Schulzrinne, H. and Casner, S. (2003). RTP Profile for Audio and Video Conferences with Minimal Control, RFC 3551 (Standard).

Schulzrinne, H., Casner, S., Frederick, R. and Jacobson, V. (2003). RTP: A Transport Protocol for Real-Time Applications, RFC 3550 (Standard).

Shetty, N. and Gibson, J. D. (2006). Improving the Robustness of the G.722 Wideband Speech Codec to Packet Losses for Voice Over WLANs, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006*, vol. 5, pp. V–V.

Sjoberg, J., Westerlund, M., Lakaniemi, A. and Xie, Q. (2007). RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs, RFC 4867 (Proposed Standard).

Sun, L. and Ifeachor, E. (2004). New models for perceived voice quality prediction and their applications in playout buffer optimization for VoIP networks, *Communications, 2004 IEEE International Conference on*, vol. 3, pp. 1478–1483.

Synopsys (2007). *Synopsys System Studio, User Guide*, Syopsys.

The MathWorks . MATLAB – The Language of Technical Computing, The MathWorks, Inc., Natick, MA, USA.

Tobagi, F. (2004). Voice over IP: the challenges behind the vision, *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 410–414 Vol.1.

Tosun, L. and Kabal, P. (2005). Dynamically Adding Redundancy for Improved Error Concealment in Packet Voice Coding, *13th European Signal processing Conference. EUSIPCO 2005*, Antalya, Turkey.

Vaillancourt, T., Jelinek, M., Salami, R. and Lefebvre, R. (2007). Efficient frame erasure concealment in predictive speech codecs using glottal pulse resynchronisation, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2007*, vol. 4, pp. IV–1113–IV–1116.

Vary, P. and Geiser, B. (2007). Steganographic wideband telephony using narrowband speech codecs, *Conference Record of Asilomar Conference on Signals, Systems, and Computers (ACSSC)*, Pacific Grove, CA, USA.

Vos, K., Jensen, S. and Soerensen, K. (2010). SILK Speech Codec, Internet-Draft, IETF.

Wah, B., Su, X. and Lin, D. (2000). A survey of error-concealment schemes for real-time audio and video transmissions over the Internet, *IEEE International Symposium on Multimedia Software Engineering*, pp. 17–24.

Wang, J. and Gibson, J. (2001). Parameter interpolation to enhance the frame erasure robustness of CELP coders in packet networks, *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*, Salt Lake City, Utah, USA, vol. 2, pp. 745–748.

Welch, L. (2003). Hidden Markov Models and the Baum-Welch Algorithm, *IEEE Information Theory Society Newsletter*.

Yajnik, M., Moon, S., Kurose, J. and Towsley, D. (1999). Measurement and modelling of the temporal dependence in packet loss, *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, New York, NY, vol. 1, pp. 345–352.