TAMPERE UNIVERSITY OF TECHNOLOGY
Degree Programme in Communication Engineering

MOZHDEH GHOLIBEIGI

CROSS-LAYER PER-FLOW QoE EVALUATION FOR VoIP IN WIRELESS SYSTEMS

MASTER OF SCIENCE THESIS

Examiners:
Prof. Yevgeni Koucheryavy
Dr. Dmitri Moltchanov
Examiners and topic approved in the Computing and Electrical Engineering Faculty Council meeting on 6th April, 2011

Abstract

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology, Department of Communication Engineering
Gholibeigi, Mozhdeh: Cross-layer per-flow QoE evaluation for VoIP in wireless systems
Master of Science Thesis, 52 pages, One Appendix pages
April 2011
Major: Communication Engineering
Examiners: Professor Yevgeni Koucheryavy and Dr. Dmitri Moltchanov
Keywords: Cross-layering, quality of user experience (QoE), voice over IP (VoIP), wireless networks, performance optimization

The Internet as the biggest worldwide network provides a huge range of facilities and conveniences. Different types of communication applications are one of the most manifest conveniences provided up to date. Among others wireless VoIP as one of the most adhered applications which compliances anywhere/anytime communication capability is of special interest and attention. Performance optimization in the context of real-time applications such as VoIP is one of the most principal and yet challenging issues. Wireless environments aggravate conditions and tensity of issues due to inherent uncertainties, vulnerabilities and time-varying characteristics. The layered structure of the communication protocol stack is not well-suited for wireless environments by setting isolated layers and encumbering limitations. Whereas, if different layers of the protocol stack not only neighbouring ones can communicate and exchange information and make appropriate functionality decisions based on obtained information may be it becomes more straightforward to achieve optimized performance at any given instant of time. This is the basis for cross-layer design. In this thesis work we proposed a cross-layer performance evaluation frame work for a wireless VoIP flow of interest from the end-user perspective. As, quality perception is the most momentous aspect of it. In this frame work we used the E-model for formulating and measurement of perceived speech quality. In our work we considered the effect of underlying layers parameters and processes contributing in performance evaluation on the performance provided to the IP layer. As, performance evaluation is carried out at the IP layer. IP packet loss probability and transmission delay as the effects of the wireless channel, FEC and ARQ error concealment mechanisms at the data-link layer, queuing process at the IP layer, losses due to buffer overflow and at the end perceived quality evaluation through simple packet loss rate model and the integrated loss metric model (Clark's model) were considered through extensive simulations. We realized that the Clark's model which takes into account the effect of loss correlation provides more accurate performance estimates. As a general and important result we concluded that by designing and developing dynamic performance control systems such as rate control or resource allocation which dynamically adapt to time-varying real-time traffic and wireless channel conditions we can achieve better performance at any given instant of time. This can be considered as further studies and as an extension of this thesis work.

Preface

This thesis is based upon studies conducted during November 2010 to April 2011 at the Department of Communication Engineering, Faculty of Information Technology, Tampere University of Technology (TUT), Finland.

In the first place I would like to record my gratitude to Dmitri Moltchanov for his supervision, advice, and guidance from the very early stage of this thesis work as well as giving me invaluable experiences throughout the work. Further, I would like to gratefully thank my co-supervisor Yevgeni Koucheryavi for his great co-operation and help. Pete, I am grateful in every possible way and hope to keep up our collaboration in the future.

Words fail me to express my appreciation to my husband Morteza whose love, dedication and valuable advice in science discussion has taken the load off my shoulder.

Finally, my special thanks go to my darling family, my father Parviz, my mother Jamileh, my sister Mozhgan and my beloved niece Anis who have inseparably supported me abroad. My parents deserve special mention for their inseparable support and prayers.

Mozhdeh Gholibeigi
April 2011

Table of Contents

List of Acronyms

AMC	Adaptive Modulation and Coding
ARQ	Automatic Repeat Request
BER	Bit Error Rate
CBR	Constant Bit Rate
CELP	Code Excited Linear Prediction
CRTP	Compressed Real Time Protocol
DSP	Digital Signal Processing
FEC	Forward Error Correction
HARQ	Hybrid ARQ
ICMP	Internet Control Message Protocol
ILBC	Internet Low Bit rate Codec
IR-HARQ	Incremental Redundancy-HARQ
MIMO	Multiple Input Multiple Output
MOS	Mean Opinion Score
NACF	Normalized Auto-Correlation Function
OSI	Open Systems Interconnection
PAMS	Perceptual Analysis Measurements System
PCM	Pulse Code Modulation
PDU	Protocol Data Unit
PESQ	Perceptual Evaluation of Speech Quality
PLR	Packet Loss Rate
PSQM	Perceptual Speech Quality Measure
PSTN	Public Switched Telephone Network
QoE	Quality of user Experience
QoS	Quality of Service
RLP	Radio Link Protocol
RS	Reed Solomon
RTCP	Real-Time Control Protocol
RTP	Real-Time Protocol
SAP	Service Access Point
SBP	Switched Bernoulli Process
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VoIP	Voice over Internet Protocol
WCDMA	Wideband Code Division Multiple Access
WEH	Wireless Extension Header

Chapter 1

Introduction

The Internet is a compulsion providing lots of facilities in various aspects of today's life. Different types of communication applications are one of the most manifest conveniences provided up to date. The growing level of expectance of people and competition between various providers of Internet communications result in lots of advanced communication technologies and this spontaneously opens the doors toward newer and state of the art achievements daily. One of the most popular and prevalent Internet communication applications is Voice over Internet Protocol (VoIP). As communication between people is indispensable nowadays, VoIP has attracted the interest of research communities so much, especially over the last decades. The structure of Voice over IP networks differs from conventional telephone networks and the voice quality is affected by a wider variety of network impairments and can significantly vary even during a call session. Therefore, monitoring the voice quality is essential in order to have a proper view of the network conditions to be able to make correct decisions such as resource allocation and rate adaptation with the final aim of perceived quality optimization. In this context cross-layer approaches are highly used as effective means breaking the traditional rule of the OSI layered reference model which restricts the relations between protocols and processes into the adjacent layers only and as a result smoothing the way for further optimizations.

Recalling the difference between quality of service (QoS) and perceived quality known as quality of user experience (QoE), in this thesis work we propose a cross-layer QOE evaluation frame work for a VoIP flow of interest in a wireless system. It evaluates the effect of lower layer protocols and processes on the performance provided for VoIP flows at the IP layer.

1.1. Wireless VoIP architecture and its applications

VoIP is one of the widely used applications of packet-switched networks. As a general definition VoIP is the real-time delivery of small voice packets over packet-switched networks like Internet. Compared to traditional circuit-switched telephone networks, it is noticeably more economic and most often can offer a satisfactory level of speech quality. However, it takes time for VoIP technology to have a perceived speech quality being comparable to quality provided by PSTN telephony.

Today's life strictly necessitates anywhere any time accessibility to communication services. As a result, wireless technologies and VoIP over wireless systems as one of the most populars of them have noticeably advanced recently. Although there are many aspects of VoIP technology to be studied, the most outstanding one is the final adjudication of the perceived speech quality by the end-user. Hence, most researches are naturally concentrated on the perceived voice quality, known as quality of user experience (QoE).

Figure 1.1 shows the protocol stack of the VoIP traffic in wire line and wireless networks.

Application Layer	RTP	RTCP
Transport Layer	UDP	
Network Layer	IP	
Data Link Layer	IEEE 802.3	IEEE 802.11 x

Figure 1.1: *VoIP protocol stack.*

RTP (Real-time Transport Protocol) [31] supports VoIP in the Application layer. It is the protocol used to deliver delay-sensitive real-time data. RTP provides payload type identification; sequence numbering; time stamping and delivery monitoring services. In the Transport layer VoIP is supported by UDP (User Datagram Protocol) that does not guarantee quality of service (QoS).

RTCP (Real-time Control Protocol) [31] is the control protocol responsible to monitor quality of service by transmitting information about the contributing users in a session.

The compressed voice sample is packed as an IP packet along with a header in the IP layer to be routable over IP networks. Encapsulation of IP packets in frames is the responsibility of IEEE 802.3 [33] or 802.11 for wire line and wireless networks respectively in the data-link layer. Framing, error control and flow control are the services supported by these two data-link layer protocols.

Figure 1.2 shows a wireless VoIP system architecture. We survey different components of it, introducing also, impairments caused by each one on voice communications.

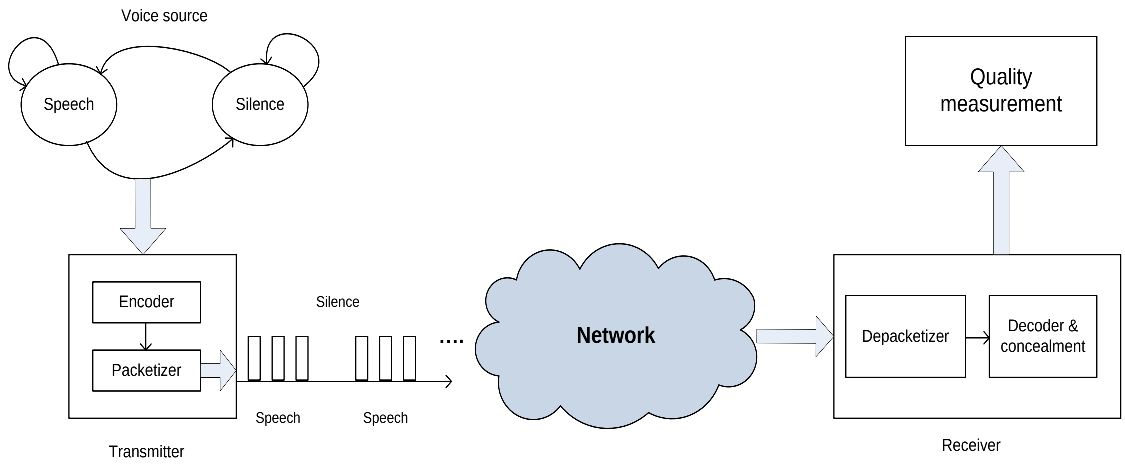


Figure 1.2: *VoIP system architecture.*

VoIP architecture components

Speech is a slow-varying analogue signal swapping between speech and silence states periodically. Exponential distribution has been proved as an appropriate distribution to model and formulate these states in simulations [32]. As a result of being an analogue signal, speech signal must be digitized and packetized at the transmitter to be ready to transmit over IP networks. At the other end, the receiver is responsible to carry out the converse transformation.

Voice codecs

IP telephony utilizes a lot of voice codecs with different bit rates and complexities. Some standard ITU-T voice codecs are G.711, G.723, G.726, G.728, and G.729. ILBC (Internet Low Bit rate Codec) is one of the recent ‘‘IETF’’ voice codecs which is license free, but it is not yet widely used compared to ITU-T codecs. Codecs affect the required bandwidth as they determine the payload size of the packets. Increasing the payload size reduces the number of sent packets. Therefore, reducing the number of required headers, results in reduced bandwidth consumption but increased latency on the other hand.

The most common sampling-based mechanism is G.711 utilizing pulse code modulation (PCM). Some codecs such as G.723 and G.729 are based on code excited linear prediction (CELP). Table 1 summarizes common codec specifications, we used in our simulations. These are codecs with silence suppression capabilities, implemented using voice activity detection (VAD) system. In VoIP systems, both talk and silence periods are packetized and RTP packets are sent by VAD only upon voice detection. Generally, VAD is able to reduce the bandwidth consumption by approximately *40%* [30].

VAD is used to decrease the data rate. The mean of speech and silence periods, modelled by exponential distribution, is affected by VAD system characteristics. For instance, most of VAD systems introduce some additional time around the length of a speech period to avoid speech truncation which is known as hangover time. The longer the hangover time, the longer the speech and silence states as well.

VoIP packets are indicated by payload and IP/UDP/RTP headers. The payload is compressed using digital signal processing (DSP) and has variable length based on the codec. The headers have a constant length of *40* bytes which is rather high for example compared to the *20* bytes of payload for G.729 codec. It is possible to compress these headers to two or four bytes using RTP header compression (cRTP). As an example, bandwidth consumption without cRTP is around 24 kbps for a default G.729 VoIP call, while it is about *12* kbps when cRTP is enabled [20]. In general, the bandwidth consumption by a VoIP call is affected by codec type, samples per packet, VAD, and cRTP.

Table 1.1*: Some common codecs specifications.*

Codec & Bitrate (Kbps)	Codec sample size (Bytes)	Codec sample Interval (ms)	Mean Opinion Score (MOS)	Voice payload size (Bytes)	Voice payload size (ms)	Packets per second (PPS)	BW MP or FRF.12 (Kbps)	BW w/cRTP Mp or FRF.12 (Kbps)	BW Ethernet (Kbps)
G.711 (64 kb/s)	80	10	4.1	160	20	50	82.8	67.6	87.2
G.728 (16 kb/s)	10	5	3.61	60	30	34	28.5	18.4	31.5

As it is obvious from Figure 1.2, packetizing is the next operation carried out on coded speech signal. As it is known, any packet includes headers and payload. The voice packet contains headers from different layers of the VoIP protocol stack including *12* Byte from RTP in the Application layer, *8* Byte from UDP in the Transport layer, *20* Byte from IP in the IP layer and also headers from the data-link layer protocols and the payload as a piece of coded speech signal. All VoIP packets have the same size.

After signal conditioning, it is ready to be sent over the channel. The VoIP packets pass through the wireline channel and after reaching the access point, they must continue the way on the wireless channel. They are subject to delay and maybe loss, as results of being sent over packet switched networks and wireless channels. Delays (fixed and variable) are mostly results of codec processing, serialization, propagation and the components along the path of the voice packets. Variable delays are mainly due to intermediate nodes (queuing and processing delays). Therefore, a buffer is required to be used in order to buffer enough packets to have a ceaseless playout at the receiver, in the presence of time-varying delays. So, packets exceeding the predicted playout delay must be dropped according to two types of mechanisms used for this purpose: fixed and adaptive.

In the fixed mechanism the overall delay due to network transmission and buffering is the same for all packets and is predicted beforehand. This delay must be calculated in such a way, resulting in the best possible speech quality at any instant of time.

Considering large buffering delay results in loss reduction (as a consequence of exceeding the arrival delay limit), but tardy communication as well. On the other hand, small buffering delay improves interplay but increased loss results in perceived quality depression. The trade-off between the overall delay and the buffer loss must be considered to choose the fixed playout delay. For this to be achieved, the network delay must be reasonably predictable. Often, it is not possible, due to time-varying delays in the network. Therefore, this mechanism is not appropriate for VoIP systems, as it cannot respond to time-varying delay timely. The second method is more appropriate for real-time applications such as VoIP, due to dynamic adaptation of the playout delay for each speech period or even for a single packet, according to time-varying characteristics of packet-switched networks.

As it is seen in Figure 1.2, the buffer delivers the voice packets to the depacketizer to be stretched and delivered to the decoder. The responsibility of the decoder is converting digital signal to analogue speech signal.

1.2. Motivation

To enhance the performance of VoIP and extremely putting upon its significant benefits, there is always space for further research and investigation. Research communities try to surmount the gap between the performance provided by the traditional circuit-switched telephone network and packet-switched voice over IP. In this way various weak points and impairments must be precisely analyzed and solutions must be proposed to alleviate their effects.

1.2.1. Wireless VoIP quality impairments

Perceived speech quality is defined as the quality perceived by the end user, known as quality of user experience (QoE). There are two ways to evaluate the perceived speech quality referred to as subjective and objective evaluations. In subjective evaluation real end users are requested to evaluate the perceived voice quality based on the mean opinion score (MOS) quality metric in a range from *1* to *5* for worth and best perceived quality respectively. Subjective quality evaluation is not practical, as lots of end users in similar conditions must be requested to do evaluation. Also, the evaluation of end users is not accurate enough as for instance switch from good to bad periods is perceived instantly by end users while, switch from bad to good period is perceived in a longer time period than it is. For this reasons objective quality tests are being used. These tests provide quality measurement mechanisms in which the quality metric can be mapped to MOS.

Although the cost-effectiveness of VoIP is the most outstanding advantage of it, but its quality is not yet comparable to the traditional PSTN telephony. VoIP packets traversing through their path are subject to a lot of impairments. They are even more severe in the case of wireless systems due to inherent uncertainties and vulnerabilities of wireless mediums. Generally, packet loss and end-to-end delay are considered as the most devastating impairments.

- **Packet Loss**

Packet loss is the most severe and sensible impairment. Overflow at intermediate nodes of the network, overflow at the playout buffer and congestion of the network links are the main reasons for losses to be occurred. Additionally, sending voice packets on IP networks may result in disordered voice packets that would be dropped by the receiver. Therefore, packet loss is not avoidable in best-effort IP networks. The type of the coding algorithm used by codecs (for instance, FEC) significantly affects the voice quality in the presence of packet loss.

- **End-to-end delay**

The overall end-to-end delay imposed to voice packets results in aggrieved and inconvenient interaction between two participating end users. It includes the delay imposed by coding and decoding processes, the delay imposed by packetization process, the delay imposed by the network (transmission time, propagation and buffering delay at intermediate network nodes) and the playout buffering delay. The human ear is not sensitive to delays less than *100ms*. Delays longer than *300ms* are obviously sensible and annoy the end users interactivity. Therefore, the maximum end-to-end delay must be kept under a certain level, typically *150ms* [27]. The delay imposed by the network is the longest one and in this work we consider that.

1.2.2. Methods of packet error recovery

Loss as a result of packet loss or bit errors has the most severe effect on the perceived quality of voice. There are several mechanisms introduced and developed for error recovery. Here, we consider two main of them.

- **Forward Error Correction (FEC)**

FEC is an error recovery scheme employed at the data-link layer. As the name implies, FEC [26] is an error recovery mechanism that does not rely on the transmitter for error correction and loss recovery. The redundant data required for loss recovery is transmitted along with data packets. There are two types of redundant data as media-independent and media-dependent. In the first type there is no need to know the type of the original data and the original data is sent along with the redundant data to the receiver. In the latter case media-dependent redundant packets are used to recover the lost original data packet.

The coding mechanisms used for transmitting the redundant data packets usually use less bandwidth than mechanisms used for transmitting data packets. Waste of bandwidth for transmission of redundant data in the case that no errors have been occurred, is the main weak point of FEC. Hence, it is not bandwidth-efficient and also causes to increased end-to-end delay.

– **Automatic retransmission request (ARQ)**

As the name implies, ARQ [22] is based on retransmission of lost or erroneous packets by the transmitter. ARQ error recovery process could be divided into three steps: at the first step, the transmitter or the receiver detects the lost data. The second step is acknowledgement step. Acknowledgements regarding received or lost data are transmitted by the receiver to the transmitter. The last one is retransmission step, indicating data retransmitted by the transmitter. Despite its significant advantages, such as efficiency and robustness, ARQ leads to some problems in delay-sensitive real-time applications such as VoIP.

In our work we utilized Hybrid ARQ mechanisms known as Type I and Type II. These error concealments mechanisms integrate FEC and ARQ schemes together. In the first case both FEC error correction and ARQ error detection bits are transmitted along with the original data packets. At the receiver side FEC bits are decoded at first. If the channel is in good condition and all errors are correctable, the receiver accepts the data block, otherwise, if the channel is not in a good condition and all errors are not correctable, the receiver realizes that via ARQ redundant bits, rejects the data block and requests the sender to retransmit the data block. In the Type II HARQ only ARQ bits are transmitted at first transmission and if the data block is error-free, there is no need to send FEC bits, while if there is any error FEC bits are also transmitted in further retransmissions. Therefore, Type II HARQ does not suffer from capacity loss like Type I HARQ. Since, FEC bits which are more bandwidth-consuming compared to ARQ redundant bits are sent only upon request.

1.2.3. Cross-layering

Cross-layer design breaks the rules of the traditional layered structure of the OSI (open systems interconnection) reference model in which functions of layers are isolated and independent from each other and each layer communicates only with its neighbour layers. The outstanding advantage of the layered paradigm is the possibility of developing standard modular components for each layer separately resulting in simplified integration. However, this traditional layered structure is not well-suited for wireless communication systems. As, no adaptation is purposed to the time-varying characteristics of wireless systems and this deficiency may lead to significant performance degradation, especially in the case of real-time applications such as VoIP.

The aim of cross-layer design is performance improvement by joint-layer optimization mechanisms such as information exchange between non-adjacent layers. Cross-layer design utilizes the inter-relation between the knowledge and processes of different layers. In other words, a layer takes into account the information provided by other layers to make dynamic decisions regarding its operation with the aim of better adaptation of the system to the time-varying characteristics of wireless communication mediums.

Cross-layer designing has attracted the interest of research communities over the recent years and a lot of cross-layer frameworks have been developed for a wide range of applications with the main aim of performance optimization. For example, in [14] a cross-layer framework for performance optimization of single-cell voice over WiFi communications has been proposed; in [13] a cross-layer performance optimization framework for real-time video streaming in ad-hoc networks has been developed and in [25] cross-layer performance enhancement of multimedia applications over satellite has been purposed.

1.2.4. Problem statement

As mentioned earlier mechanisms such as FEC and ARQ are used to cope with impairment factors. Nonetheless, they are not sufficient to individually cope with time-varying traffic and channel conditions and improve the performance of wireless VoIP in terms of perceived quality significantly. Therefore, it is required to somehow optimize these error concealment mechanisms such that reflecting dynamic wireless system conditions and performance metrics. The interaction between these mechanisms and various components of wireless VoIP system must be considered and analyzed too. The integration of these error concealment mechanisms could be considered as a way for performance improvement.

We utilized a cross-layering approach to carry out a joint optimization between different layers of wireless VoIP protocol stack through extensive simulations. Further, we tried to analyze and discover the effect of various performance metrics and inter-relation between them.

1.3. Objectives

The aim of this thesis work is to simulate and evaluate a cross-layer approach between the data-link layer and the IP layer of the VoIP protocol stack to improve QOE for wireless VoIP flows and analyze the effect of interactions between various setting and conditions on the perceived quality of wireless VoIP.

1.4. Contributions

In this thesis work we propose a cross-layer performance evaluation frame work. As a result of the cross-layer basis, it is possible to investigate and analyze the effect of various contributing parameters and processes belonging to different layers of the protocol stack on performance evaluation of a VoIP flow of interest in wireless environments.

To parameterize the quality assessment procedure the E-model is chosen as the most appropriate method introduced up to date for packet-switched networks. We estimate the R-factor as the output of the E-model based on the simple packet loss rate model and the more advanced model known as Clark's model (integrated loss metric). Then, it can be mapped to the well-known perceived speech quality metric known as mean opinion score (MOS).

In addition to considering IP packet loss and transmission delay as the effects of the wireless transmission medium, the contribution of FEC and two types of hybrid ARQ as error concealment mechanisms at the data-link layer on performance evaluation at the IP layer are studied to achieve more precise estimates.

1.5. Thesis structure

The rest of the thesis is organized as follows. In chapter 2 the background of VoIP applications and perceived quality evaluation mechanisms for VoIP are introduced. In chapter 3 we consider to the context of cross-layer design approaches for wireless systems and various aspects of it. In chapter 4 our cross-layered-based performance evaluation model and its different parts are introduced. In chapter 5 we discuss the outcome of our project work and simulations. Finally in chapter 6 we present a short conclusion to conclude the thesis.

Chapter 2

Background of VoIP applications

In this chapter we review the background of VoIP applications emphasizing the growing role of them in today's life and also present some perceptual quality measurement mechanisms have been used up to date.

2.1. Growing popularity of VoIP applications

As the Internet has turned into a universal network new Internet-compatible communication applications dominate many traditional applications such as public switched telephone network (PSTN). VoIP is one of great importance due to its significant revenue and ease of use. These factors lead to growing popularity of VoIP and consequently open the doors to dependent research areas such as quality improvement and evaluation.

2.2. Quality prospect of VoIP

For VoIP to be a tolerable alternative to the traditional PSTN, it is required to provide an aceptive level of perceived speech quality. VoIP packets traversing through their path are subject to impairments such as delay and loss. In the case of wireless systems the effect of these impairments are even more severe. Therefore, the quality of voice needs to be evaluated somehow. Quality assessment can be carried out from the network perspective, known as quality of service (QoS) [12] or user perspective, known as quality of user experience (QoE) [18]. In the context of VoIP, the final adjudication of perceived speech quality by end users is the most important. Hence, it is rational to mainly consider quality assessment from the user perspective. In order to assess the quality of voice communication in the presence of impairments, it is essential to consider the individual factors and overall effects of the impairments to provide quantitative measurements reflecting the subjective rating known as mean opinion score (MOS).

MOS is a subjective quality evaluation metric defined in ITU-T P800 standard [24] which is introduced to provide a numerical measure of the perceived quality of human speech at the receiving end ranges from *1* (bad) to *5* (excellent) as demonstrated in Table 2. It is estimated by averaging the results of a set of subjective tests where a number of humans grade the heard audio quality of test sentences. A listener is required to give each sentence a rating using grades from *1* to *5*. The MOS is the mean of all scores set by individuals.

***Table 2.1**: MOS quality scales.*

Quality scale	Score	User satisfaction & listening effort
Excellent	5	Very satisfied & no effort
Good	4	Satisfied & no significant effort
Fair	3	Some dissatisfied users & moderate effort
Poor	2	Dominant dissatisfaction & significant effort
Bad	1	No meaning

Although there are some methods to measure the perceived speech quality for a VoIP system, they can be categorized into two main mechanisms as subjective quality assessment and objective quality assessment.

2.2.1. Subjective quality measurement

Subjective quality measurement needs to provide a large group of people in similar conditions and request them to grade the perceived voice quality from *1* to *5*. It requires much time and it is difficult to provide the same conditions for all users. Additionally, it is not very accurate as the quality perception by different people may differ noticeably. Thus, it is expensive and unrepeatable.

2.2.2. Objective quality measurement

Objective tests are automatic and do not require real end users to evaluate quality in real environments. As a result they are repeatable and do not depend on environmental conditions. There are several objective methods to measure the voice quality. Some of them are based on long term averages of statistics such as packet loss, delay and jitter. These mechanisms do not reflect the quality perceived by users. Perceptual analysis measurements system (PAMS), perceptual evaluation of speech quality (PESQ) [28] and perceptual speech quality measure (PSQM (+)) are examples of these methods. They require both the original (reference) speech signal and the degraded version to carry out quality assessment by means of digital signal processing (DSP) algorithms. These mechanisms are not suitable for voice quality evaluation on a data network, as they are not developed considering data networking and they cannot reflect time-varying networking impairments such as loss, delay and jitter. The E-model [35,3] which is an ITU-T standard is developed as a transmission planning tool not only a measurement tool. It does not require the original speech signal to perform quality measurement. In this thesis we consider the E-model for perceived quality assessment.

2.2.2.1. E-model

The E-model was originally developed for PSTN planning. It is an objective performance model which covers the effect of random (independent) packet loss, after revision. The introduced closed-form model makes it more applicable to (narrowband) VoIP network planning. As presented in Figure 2.1 the E-model combines the perceived effect of all impairments based on the fact that they are additive. A single scalar known as transmission rating factor (R-factor) is the output of the E-model computed based on channel and equipment impairments.

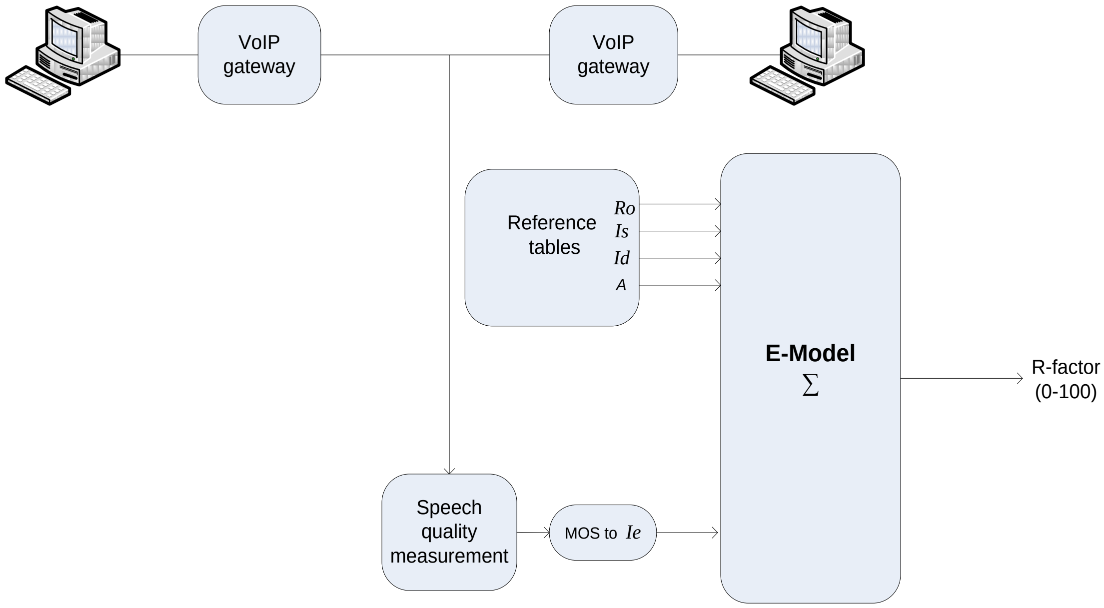


***Figure 2.1**: Using the E-Model for VoIP quality assessment.*

R-factor can be mapped to MOS quality values according to the following equations as shown in Figure 2.2.

$$
\begin{aligned}
&MOS = 1 && \text{if } R <= 0 \\
&MOS = 1 + 0.035\,R + R\,(R-60)\,(100-R) * 7 * 10^{-6} && \text{if } 0 < R < 100 \\
&MOS = 4.5 && \text{if } R >= 100
\end{aligned}
\qquad (2.1)
$$

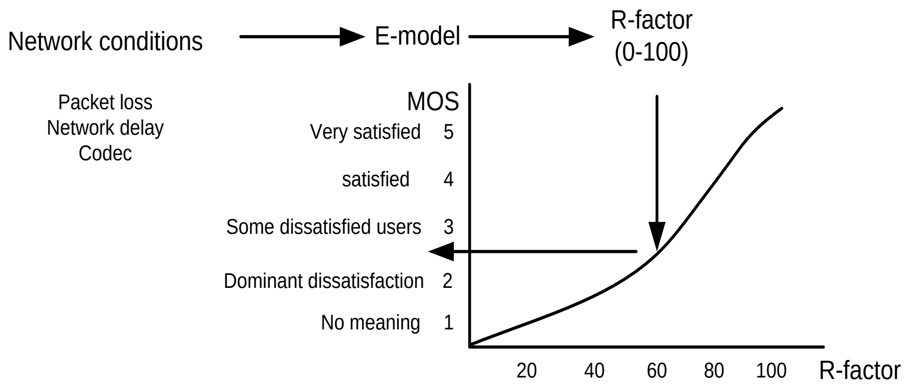


Figure 2.2: *Mapping MOS to R-factor.*

The basic equation for the E-model and the meaning of each parameter are as follows.

$$R = R_0 - I_s - I_d - I_e + A \tag{2.2}$$

R is the overall network quality rating. R_0 represents noise and loudness in terms of the signal to noise (*S/N*) ratio at *0 dBr* point. I_s indicates the sum of all impairments which are more or less simultaneous with voice signal transmission (for example, sidetone, coding and compression are included in I_s). R_0 and I_s are inherent to the transmitted voice signal and are not affected by transmission over the network. I_d is the sum of all impairments delayed after voice signal transmission such as loss of interactivity and echo. I_e stands for impairment of equipment (e.g. low bit-rate codecs). A is the advantage factor indicating sacrificed users who accept the voice quality considering the easy access to the service. R-value ranges from *0* to *100* for poor to excellent quality respectively.

– **Delay impairment I d**

The quality degradation due to one-way delay (mouth-to-ear) is formulated by I_d as: [23]

$$I_d = 0.024\, T_a + 0.11\, (T_a - 177.3)\, H\, (T_a - 177.3) \tag{2.3}$$

Where T_a is the one-way delay in milliseconds and the function of *H(x)* is as follows.

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{2.4}$$

The delay impairment factor experienced in VoIP routes is usually less than the maximum tolerable delay (*150ms*, [27]). In this case I_e is the dominant impairment factor.

– **Equipment impairment I_e**

Impairments due to low rate codecs and packet losses are captured by equipment impairment factor I_e. The effect of I_e has been found using subjective experiments [1]. However, as demonstrated in [4] relying just on the first-order statistics of the packet loss process may result in different perceived speech qualities. The correlation between packet losses is the reason behind that. By using codecs with packet loss concealment capability (which is an optional feature) it is easy to cope with single packet losses and reduce their effect using extrapolation of the reconstructed signal. However, in the case of lost bursts extrapolation does not help and leads to undesired results. Therefore, it is required to somehow take into account the effect of packet loss correlation.

2.2.2.1.1. Packet loss rate (PLR)

Packet loss rate is rather a simple way to predict the perceived speech quality as the percentage of lost packets vs. the total number of transmitted ones. In this thesis work we use this model and more advanced Clark's model to evaluate the perceived speech quality of VoIP and compare achieved results.

2.2.2.1.2. Clark's model

Clark [1] defined two loss and loss-free states in the packet loss statistics, to take in to account the effect of loss correlation. According to this model, the system remains in the loss state as long as there are no more than m successfully received packets between two loss events. If more than m packets are successfully received the system switches to the loss-free state. The threshold *m* is affected by the type of the codec and extrapolation capabilities of it. Loss-related impairment I_e is measured in the case of state transition. Then, the average of loss and loss-free states impairments is considered as the overall loss-related impairment.

This model is extended in [4] and [34] and it includes the effect of delayed perception. Changes in quality levels and state transitions are not instantly perceived by humans. For instance, transition from loss-free to loss state is usually sensed faster compared to the inverse case. According to [4] the effect of this delayed perception can be well modelled based on exponentially decaying functions with suitable time constants as demonstrated in Figure 2.3. The time-averaged loss-related impairments can be obtained by integrating over all probable durations of loss-free and loss states [1]. Then, the integral speech quality can be computed by substituting it to (2.2) [4]. [8]

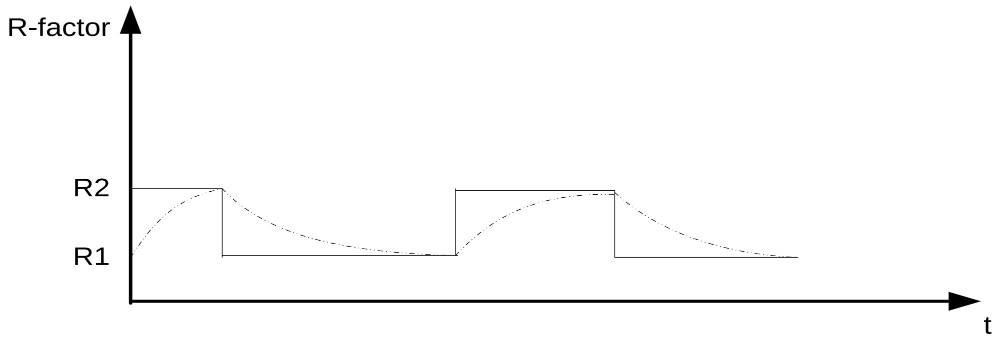


Figure 2.3*: R-factor computation based on* [4].

Chapter 3

Cross-layer design of wireless systems

In this chapter we briefly review the general principles behind cross-layering design and discuss the reasons for exploiting.

3.1. Why cross-layering

Optimizing the performance of applications in wired networks is often achievable by controlling performance degradation as a result of packet forwarding procedures. In the case of wireless networks the performance degradation caused by incorrect symbol reception at the air interface must be taken into account. These errors propagating to higher layers often contribute to significant end-to-end performance degradation. Therefore, the air interface is considered as a ’weak point’ in any end-to-end performance assurance model proposed for IP-based wireless networks up to date.

Some advanced mechanisms such as adaptive modulation and coding (AMC) scheme, multiple-in multiple-out (MIMO) antenna design, different forward error correction (FEC) and automatic repeat request (ARQ) procedures, Transport layer error concealment functionality and etc are employed by wireless access technologies to improve the performance of information transmission over wireless channels. To make decisions regarding the protocol parameters offering the best possible performance for a given channel and traffic conditions, wireless access mechanisms demand for new approaches for designation of the protocol stack including cross-layer performance optimization capabilities.

Different layers of the protocol stack must be able to communicate and exchange control information to optimize the performance of applications running over wireless channels. In both ITU-T OSI reference model and TCP/IP model the functionalities of each layer of the protocol stack are isolated. Each layer in these models is responsible for distinct functions and communicates only with its neighbouring layers through request-response primitives defined for service access points (SAP) and the same layer of a peer communication entity. The layers do not know about the specific functions of other layers.

Despite its efficiency in wired networks, the layered structure of the protocol stack is not very well-suited for wireless networks. Therefore, new organization of the protocol stack at the air interface is required to optimize the performance of applications running over wireless channels. It does not mean to totally redesign the protocol stack and direct interfaces between adjacent layers are always preferable, along with direct communications between non-adjacent layers. Indeed, the network layer and layers above it often need direct interfaces to the data-link layer for handover. Specially, data-link layer protocols must be informed about higher layers including network and transport layers' parameters and vice versa in order to decide on traffic management issues.

In this context we can define the cross-layer design of the protocol stack as a design breaking the traditional rules of the layered structure of communication protocols. A number of cross-layers proposals have been introduced up to date. Some types of cross-layer interactions are as follows. Merging of adjacent layers, vertical calibration of parameters across layers, design coupling and creating new interfaces [16]. The task of definition and implementation of two or more adjacent protocols in the protocol stack is referred to as merging of adjacent layers. As a result of implementing these schemes there is no need for new interfaces. However, they are characterized by more complicated implementations. Vertical calibration of parameters across layers means that the parameters of protocols belonging to different layers are adjacent during the runtime and some performance metrics can be optimized. Hence, new interfaces between non-adjacent layers are required to be introduced. In the case of design coupling as the name implies some protocols are made aware of the operational parameters of each other at the design stage while there is no information exchange between non-adjacent layers during the runtime. According to another approach, new interfaces can be established for upward and downward exchange of information between non-adjacent layers at the runtime.

As it can be realized, the exchange of information between different layers of the protocol stack during the runtime or at the design step is the common ultimate goal of mentioned approaches.

In [39] a comprehensive review of cross-layer design approaches can be found. In [38] some cross-layer design examples are presented.

3.2. Cross-layer signalling

Realizing communication between non-adjacent layers requires an appropriate signalling scheme. Using the signalling scheme and appropriate interfaces exchange of control information between different layers is feasible. Although various cross-layer signalling mechanisms are introduced up to date, but they can be categorized as in-band and out-of-band signalling in general [7]. For instance, in [15] the use of wireless extension header (WEH) of IPv6 protocol is proposed to communicate between TCP and radio link protocol (RLP) in wireless IP-enabled networks. The cross-layer signalling mechanism of this approach is in-band as it utilizes IP data packets for information exchange between the transport and the data-link layer. Authors in [29] proposed the use of ICMP messages for communication between different layers of the protocol stack. A new ICMP message is generated in the case of change of a certain parameter. In these two cross-layer signalling schemes only few layers can participate in information exchange processes. According to the approach proposed in [5] a network service gathering, managing and distributing information about current parameters is used at mobile hosts. This central service is accessible by all protocols interested in certain parameters. Therefore, this scheme introduces a new service separated from the layers of the protocol stack. Authors in [17] proposed the use of local profiles to achieve information. The concepts of last two approaches are more or less similar, but in the latter case the information is stored locally. Less overhead and delay are positive results of that approach. This scheme is further extended in [10] as active local profiles. Implementing control procedures optimizing the performance wireless applications is the extended responsibility of these active profiles in addition to storage only. To make possible direct communication between non-adjacent layers of the protocol stack without annoying processing at intermediate layers a dedicated cross-layer signalling scheme is proposed in [40]. However, it imposes more complexity on the protocol stack.

Out-of-band signalling is claimed to be more efficient and reasonable by authors in [40] as it does not suffer from unnecessary processing of signalling messages at intermediate layers of the protocol stack. The structure of in-band signalling messages makes them often non-appropriate for providing upward and downward signalling together. It is necessary to mention that the common final aim of all these cross-layer signalling approaches is optimization of protocol parameters at different layers with the aim of performance improvement at any instant of time based on time-varying channel and traffic conditions.

The concept of distributed or centralized performance control must be taken into account to jointly design cross-layer signalling and performance optimization schemes. Distributed performance control is applied by in-band cross-layer signalling schemes proposed in [15,29,40]. Based on these approaches exchanged information between layers is used by performance control units of those layers to control appropriate parameters of them dynamically. Therefore, significant modifications are required to be carried out at layers of the protocol stack. It means that separate performance optimization subsystems are required to be implemented at participating layers and also some modifications must be implemented at other layers of the protocol stack which are passage ways for information exchange. This may lead to some problems [37]. The reason is that independent decisions for changing associated parameters by each layer may result in undesired and not expected consequences. Additionally, the delay imposed by information exchange between non-adjacent layers may be quite significant.

Out-of-band signalling schemes proposed in [5,17] implement a central external performance optimization unit which different layers of the protocol stack send their appropriate parameters to that entity through specified interfaces. This external centre performs an overall performance optimization considering all parameters sent by different layers of the protocol stack. These optimized parameters are then sent to associated layers. Hence, these approaches are based on an external cross-layer performance optimization system utilizing out-of-band signalling scheme.

3.3. Cross-layer design

As is in any approach cross-layer design of the protocol stack may also lead to problems and challenges [7]. The handling and management of the modular layered structure of the communication protocol stack is quite straight forward in wired networks [37]. The number of layers, functionalities of each layer and interfaces to adjacent layers must be specified at the design stage. Additionally, the layered structure of the protocol stack results in simplified implementation and manufacturing. In the layered structure of the protocol stack each layer communicates only with its neighbouring lower and higher layers meaning that each layer receives a certain set of services from its adjacent lower layer and provides it to its adjacent higher layer. Therefore, the responsibilities of protocols at each layer are predefined and isolated development of them is possible.

Design and implementation complexity of a system may significantly increase as a result of cross-layer design of the protocol stack and non-predicted multi-layer interactions. These consequences may result in non-clear overall functionality of the system and high manufacturing costs. Therefore, to guarantee the stability of the system additional efforts are required to be taken.

Based on all these discussions it seems that considering a reasonable trade-off between the layered structure and the cross-layer performance optimization of wireless channel is a must. While cross-layer performance optimization may satisfies the short-term goals in terms of better performance [37], clear layered structure eventuates long-term benefits. As an example we can mention low per-unit performance optimization cost [37]. Hence, in the context of cross-layering we must try to make less cross-layer interactions and also isolate the performance control system from the protocol stack as much as it is feasible.

Chapter 4

VoIP System model

In this chapter we introduce our cross-layer performance evaluation model, its sections, performance evaluation process and utilized mechanisms.

4.1.Inter-layer QOE optimization

In this project work we used a cross layer approach to evaluate the performance of wireless VoIP flows in terms of R-factor quality metric. As we discussed earlier, in cross-layering design parameters and knowledge are exchanged between different layers of protocol stack. It does not mean that all layers must be involved. Often it is sufficient to consider only some particular layers to achieve a significant performance improvement. According to our thesis work we model wireless channel characteristics at the PHY layer using bit error process and transmission delay and then extend their effect to the IP layer. This approach takes into account FEC and ARQ as error correction mechanisms implemented at the data-link layer. As a punch line, we consider the effect of wireless channel bit error and delay propagated to the IP layer on the perceived quality of VoIP flow with FEC/ARQ implemented at the data-link layer.

4.2.Model description

In this section we introduce the model in detail describing its parts and the whole performance evaluation process.

4.2.1. Sections of the model

The system under consideration in this thesis is shown in Figure 4.1. We assume that a certain number of VoIP flows of the same priority share a wireless link. Data generated by a number of sources in wired network are packetized using RTP, UDP, and IP at the end systems and then arrive at the wireless link of interest. The size of all packets is assumed to be N bytes including all headers. The buffering which is done at the IP layer is limited to the capacity of K IP packets. When there is at least one packet in the buffer and the channel is free for transmission the head-of-line packet is scheduled to the data-link layer. Between these two layers packets are segmented into v frames. Then, FEC code of Reed-Solomon (RS) type with the symbol length of m_s bits that can correct up to l incorrectly received symbols is applied and these frames are then scheduled to the ARQ process. It is assumed that the protocol data units (PDU) of the ARQ protocol consist of exactly one codeword referred to as frames. The frame size is assumed to be equal to m_f symbols. Non-persistent implementation of ARQ protocol is considered in this work and we distinguish between two cases: (i) the number of retransmission attempts is limited for a packet (ii) the number of retransmissions is limited for a single frame. When a packet is successfully transmitted or lost as a result of insufficient number of retransmission attempts the channel is made free for another packet that is queued at the IP layer. The HARQ (hybrid ARQ) procedure allows to precisely controlling the delay introduced by imperfect wireless channel conditions which is important for real-time applications such as VoIP. It is known as Type I HARQ system. The following assumptions are taken into account about the operation of the HARQ system: (i) feedback frames (negative and positive acknowledgements) are always correctly received, (ii) feedback delay is ignorable and (iii) the probability of undetected error is negligibly small. Considering the first two assumptions functionality of stop-and-wait, go-back-n, and selective repeat ARQ implementations become identical. These assumptions which are used in many studies are suitable for high-speed wireless channels with small propagation delay (see e.g. [19]).

Actually, these assumptions are not fundamental. For instance, a certain packet size distribution can be considered instead of the fixed size of a packet. More than one HARQ system can also be assumed, e.g. one on top of another. Modifications to the HARQ model can be taken into account to capture Type II HARQ functionality. Multiple HARQ implementations running in parallel can also be analyzed. Further, the model can be extended to consider other types of FEC codes. The maximum number of retransmission attempts required to transmit a single packet (successfully or not) is affected by the type of the FEC code which may change the mean packet transmission time. This eventually affects the amount of buffer space required at the IP layer to store arriving IP packets [6]. As a result, a trade-off must be taken into account between the packet loss gain obtained using the FEC code with better error correction and the amount of the buffer space required to store packets. All these refinements can be considered as extensions to the presented model. The reason for taking all these assumptions is to concentrate more on per-source performance evaluation which is the main goal of this thesis and also getting rid of a number of unnecessary input parameters. Nevertheless, possible extensions of the model are discussed in appropriate sections. [6]

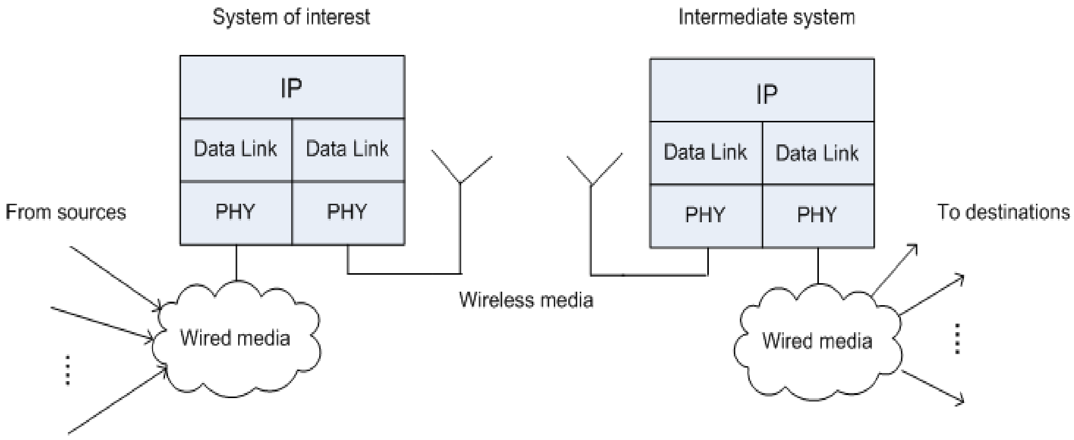


Figure 4.1: *The system model.*

We continue our discussions as follows. Two types of packet loss may occur in the system that we distinguish between them. Losses occur either as a result of excessive number of retransmission attempts performed at the data-link layer or the buffer overflow at the IP layer. Loss process caused by excessive number of retransmissions is considered at first and the service process of the buffering system is derived. A cross-layer modelling approach considering segmentation and reassembly of PDUs between neighbour layers of the protocol stack and error correction mechanisms implemented at the data-link layer is used. Further, we concentrate on the IP layer queuing system. In our simulations we apply losses caused by excessive number of retransmission attempts and those occurring as a result of buffer overflow at the IP layer.

4.2.2. Service process of the wireless channel

Performance evaluation of applications in IP networks is carried out at the IP or higher layers. Hence, for considering the effect of wireless channel such as packet delay and packet loss on the application performance they must be extended to the IP layer at which performance is evaluated and cannot be directly used. The mechanisms and processes which are used by underlying layers such as data-link error correction techniques and segmentation and reassembly between adjacent layers must be taken into account to have a precise extension.

To model the packet service process the cross-layer approach developed in [11] is used. Based on this model the wireless channel characteristics are represented using the bit error process and transmission delay and then extended probabilistically to the IP layer. Autocorrelational properties of the bit error process and error correction mechanisms of the data-link layer including both FEC and ARQ are taken into account by the model. The basic steps of the model are briefly presented here.

4.2.2.1. Bit error process

The bit error process which is denoted by *{$W_E(l)$, $l = 0, 1, . . .$}*, $W_E(l) \in \{ 0, 1 \}$ is modelled using the discrete-time Markov modulated process with irreducible Markov chain *{$S_E(l)$, $l = 0, 1, . . .$}*, $S_E(l) \in \{ 0, 1 \}$, with *1* and *0* standing for incorrect and correct bit reception, respectively [6]. Mean value and lag-1 normalized autocorrelation coefficients are used to parameterize the bit error process which is assumed as a switched Bernoulli process (SBP).

$$\alpha_E = \big(1 - K_E(1)\big)E[\,W_E]$$
$$\beta_E = \big(1 - K_E(1)\big)(1 - E[\,W_E])$$

$$\begin{cases} f_{1,E}(1) = 0 \\ f_{2,E}(1) = 1 \end{cases} \tag{4.1}$$

$f_{1,E}(1)$ and $f_{2,E}(1)$ are bit error probabilities in states *1* and *2* and α_E and β_E are transition probabilities from state *1* to state *2* and vice versa. $K_E(1)$ is the lag-1 autocorrelation of bit error process and $E[W_E]$ is the mean of bit error observations.

First and second-order statistical characteristics in terms of the bit error rate (BER) and normalized autocorrelation function (NACF) are captured by the model. If wireless channel behaves piecewise stationary as reported in a number of recent studies this model may represent statistical characteristics of covariance stationary parts with geometrically decaying autocorrelations. Under this condition, (4.1) is interpreted as a model for limited duration of time during which mean value and NACF of bit error observations remain constant. We can refer to [2,9] to get more information about non-stationary wireless channel statistics.

4.2.2.2. Symbol error process

It is quite simple as the bit error process. However, as a RS decoder assumes a symbol as lost if at least one bit of it is received incorrectly it is required that the process of correct and incorrect reception of RS symbols to be characterized at first [6].

The process *{ $W_N(n)$, $n = 0, 1, . . .$ }* , $W_N(n) \in \{ 0, 1, . . . , m_S\}$ describes the number of incorrectly received bits in consecutive bit patterns with length m_S and the index of the process denotes successive time intervals of length $m_S.\Delta$ which Δ is the transmission time of a single bit. Again, Markov chain can be used to model this doubly-stochastic process as *{$S_N(n)$, $n = 0, 1, . . .$}*, $S_N(n) = S_E(l) \in \{ 0, 1 \}$. It can be parameterized via parameters of the bit error process. m_S -step transition probabilities of the modulating Markov chain *{$S_E(l)$, $l = 0, 1, . . .$}* with exactly k, $k = 0, 1, . . . , m_S$, incorrectly received bits are required to be determined at first [6].

4.2.2.3.Frame error process

The same procedures are repeated for formulating the frame error process [6]. The length of a frame including those used for error correction is assumed to be m_F. The frame error process $\{ W_F(t), t = 0, 1, . . . \}$, $W_F(t) \in \{ 0, 1 \}$ can be obtained assuming that up to l incorrectly received symbols can be corrected by FEC code. The index here indicates the consecutive time intervals of length $m_F . m_S . \Delta$.

4.2.2.4. IP packet transmission process

Naturally, the packet service process can be given as the number of time slots of length $m_F . m_S . \Delta$ required to transmit a single IP packet [6]. Truncated Type I HARQ which limits the number of frame retransmissions for a packet to r and extensions are considered by the model. It is clear that the delay must be defined without regard to whether a packet is successfully transmitted or not. Since, the channel is occupied for a random number of time slots in both cases and this accordingly affects the waiting time of IP packets in the buffer. Therefore, it can be said that the delay of a packet is either time to be successfully transmitted or time till being lost as a result of excessive number of retransmissions.

Persistent Type I HARQ system which does not limit the number of retransmissions and non-persisted HARQ are considered by the model. The delay and loss metrics for non-persisted HARQ case can be obtained by finding the IP packet distribution for persistent Type I HARQ assuming independence between successive packet transmission times [6]. For further detailed analyze and information we refer to [6].

The memory of the bit error process in this model is limited to lag $m_F . m_S . v . \Delta$, which $m_F . m_S . v . \Delta$ indicates the length of the packet in bits. This value is sufficiently large and it is possible to accurately capture the memory of the bit error process. As an example, we assume that the NACF of the bit error process decays according to a single geometrical term. To compute the threshold m at which the model is still valid we can use $K_E(m) = \lambda^m$, in which λ is the lag-1 NACF. It is clear that even for highly correlated bit error processes with $\lambda = 0.9$ the correlation is ignorable for lags larger than *30* as $0.9^{30} \approx 0.042$. Based on this simple but important observation we conclude that the packet loss process can be accurately characterized by the single packet loss probability metric [6].

4.2.2.5. Extensions of the utilized framework

In this section we briefly discuss possible extensions of the framework we used in our thesis work [6]. As, it can be extended to capture various characteristics of modern wireless technologies. The algorithm presented to model the process of segmentation and reassembly with Type I persistent/non-persistent HARQ can be utilized to estimate the performance metrics provided to the IP layer, if more than one HARQ system are defined for a certain technology. For instance, two HARQ systems indicated by

R_1 and R_2 could be implemented at the data-link layer. Therefore, an IP packet is first segmented into a number of frames by the R_1 HARQ system and correction bits are added to those frames. Further, they are dismissed to the R_2 system one after another. The frames are also segmented by R_2 HARQ system into even smaller pieces of data known as code words and then the channel coding FEC bits are added to them. In this situation the performance provided by the R_2 HARQ system to the R_1 system can be analyzed based on the bottom-up approach and the frame error probability and the probability function of the delay experienced by frames can be obtained. Similarly, it is possible to exploit the same framework to get the IP packet loss probability and the probability function of the time required to transmit a single IP packet.

Some HARQ systems limit the number of permitted retransmissions for a single frame instead of limiting the maximum transmission time for a single packet. It means that if a frame is incorrectly received in r retransmission attempts the whole packet associated to that frame is discarded. The reason for pursuing such strategy is that most probably the physical connection is lost when r consecutive retransmission attempts fail. The utilized framework can also be used to analyze this case [6]. The extension to the case of Type II HARQ system with incremental redundancy (IR-HARQ) can be easily handled by the utilized framework [6]. Only redundancy symbols are carried by retransmission attempts in such systems. These symbols which are added to the original data symbols and sent in the first transmission attempt increase the probability of successful frame reception. Type II HARQ systems are most often non-persistent and limit the number of frame retransmissions in a packet. To formulate the functionality of such systems it is enough to re-compute the frame and packet loss probabilities [6].

This framework can also be used when a single packet is transmitted over a number of channels using multiple-in multiple-out (MIMO) transmission system. Assuming that all the sub channels are independent of each other they can be modelled using separate bit error models [6].

The extension to the case of variable packet size can also be performed. It is assumed that $a(k)$ is the packet size distribution measured in the number of frames. Obviously, the packet loss probability and the IP packet transmission time depend on the number of frames in a single IP packet. For each k delay distributions can be computed separately. Finally, weighting these distributions with corresponding packet size probabilities it is possible to obtain the averaged metric of interest. This approach works well when sizes of successive IP packets are not dependent. When the packet sizes are correlated the analysis is more complicated and requires exact knowledge of the arrival process.

To increase the throughput of HARQ systems operating in stop-and-wait regime a number of ARQ instances are sometimes run in parallel. This is beneficial for systems with non-negligible RTT, where waiting periods are filled with frame transmissions from different ARQ instances (see e.g. WCDMA air interface of UMTS system). In this case rough approximation can be obtained considering continuous transmission of frames at the wireless channel. Intuitively, this approximation becomes better when the number of ARQ instances and the number of frames in a single packet get higher [6].

4.2.3. Performance evaluation model

To evaluate the system performance it is required to model various sections and processes of the system contributing in performance qualifying.

4.2.3.1. Arrival process (VoIP traffic model)

As mentioned earlier, in this work we use two codecs with silence suppression capabilities implemented using voice activity detection (VAD) system as follows. G.711 with raw rate *64 Kbps* and G.728 with raw rate *16 Kbps*.

We are interested in voice payload size measured in bytes and milliseconds. This will give us the size of the packet and time interval between packets. For example for G.711 codec the size of the packet will be *200* bytes which consist of *160* bytes of payload and *40* bytes of RTP/UDP/IP headers. Thus, for G.711 codec, packets of *200* bytes are generated each *20ms* resulting in *50* packets per second. We note that the redundancy added by codecs should be taken into account in estimation of the minimum rate required for the channel.

The human speech is generally modelled by ON and OFF periods, belonging to the active and silence periods in human speech respectively. The VoIP source in the ON state generates packets representing human speech. Voice activity detection tries to suppress pauses in two-way voice conversation. One-way traffic after VAD application looks as shown in Figure 4.2, where SIP stands for silence indication packets. The effect of SIP packets is not taken into account in our system.

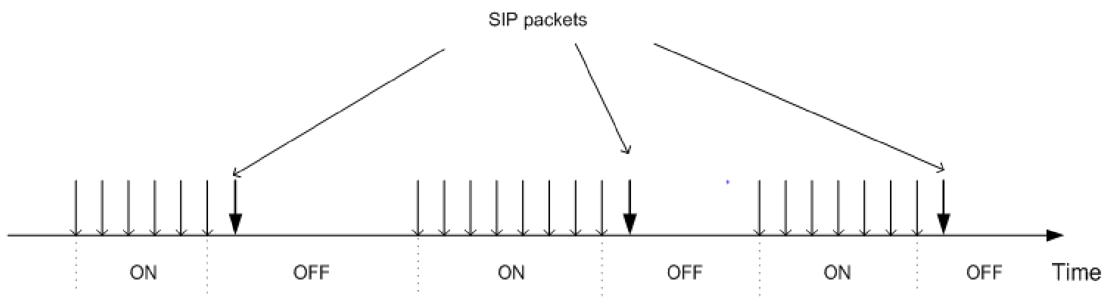


Figure 4.2*: Voice source ON/OFF pattern.*

We have the following facts about VAD:

- silence is for about *60%* of time
- talking is for approximately *40%* of time
- during talking period packet are equally separate in time (*20ms* for G.711)
- no packets are generated during silence period

- mean talking duration is *1.4s*.
- mean silence duration is *2.3s*.
- talking and silence periods are approximately geometrically distributed.

These observations (specifically, the last one) allow us to use Markov model to capture the voice source with VAD system (sojourn time in the state of discrete-time Markov chain is exponentially distributed). Naturally, the time slot duration of such model is assumed to be the time interval between two packets in the talking state (e.g. *20ms* for G.711 codec). The length of the speech and silence periods is exponentially characterized as $f(t) = \lambda\, e^{-\lambda t}$, where $1/\lambda$ is the mean of the speech or silence period. In this work we assume N VoIP sources considering one as the source of interest and others as background. Traffic arrivals at the input of the buffer are assumed as batches of several IP packets.

4.2.3.2. Queuing system

In this work a finite capacity buffering system (K) measured in IP packets is assumed as shown in Figure 4.3. Discrete event simulation is used for defining arrival and service events of IP packets. In the case of IP packet arrival happening the queue length is increased by the number of IP packets in the arriving batch, filling the queue. On the other hand, in the case of service event the queue length is decreased by one as one IP packet leaves the buffer. The loss probability in the case of buffer overflow is implemented using drop tail. As the name implies, according to this simple queue management algorithm the newly arriving packets are discarded if the queue is completely filled.

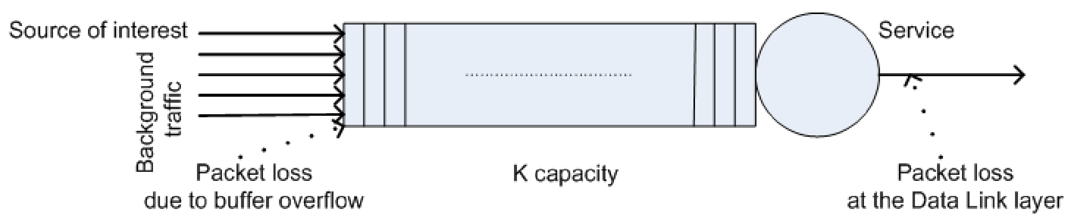


***Figure 4.3**: The queuing system.*

4.2.4. Performance evaluation

In this section the effect of IP packet loss due to buffer overflow and wireless channel transmission and the delay experienced by IP packets due to transmission over the wireless channel (probability function of IP packet service times) on the performance qualification of the system are studied.

4.2.4.1. Per-source loss performance

Since the number of IP packet arrivals from both source of interest and background sources are virtually unlimited, there can be infinitely many lost packets resulting in significant performance degradation.

In this work all possible cases for accommodation of IP packets of the source of interest by the system are considered. Consequently, the loss probability of IP packets of the source of interest in the case of inadequate or no accommodation space in the system is simulated. As discussed earlier, the loss event due to packet transmission over the wireless channel can be characterized by a single loss probability parameter. The effect of this loss is considered in our simulations. Since we carry out per-source performance evaluation in this work, only loss of IP packets from the source of interest is of our attention for simulation. In the simulations the lost IP packets from the source of interest are indicated by ones, while successfully accommodated IP packets from the source of interest are demonstrated by zeros.

4.2.4.2. Per-source delay performance

Transmission of IP packets over the wireless medium imposes some delay in IP packet submission which is important for real-time delay-sensitive applications such as VoIP and affects the overall performance negatively. We note to the fact that the delay caused by IP packet buffering is not that important. As, according to experiments this delay is most often less than *150ms* [27] and this amount of delay is not sensible by human ear. Therefore, it is ignorable and we do not consider that delay in our simulations.

The probability function of the number of time slots required to transmit IP packets (delay distribution) is obtained through simulations and after conversion to time is applied as the service time distribution for our discrete event simulator. Indeed, if the IP packet of the source of interest is accommodated by the system at the place i it must wait till all IP packets ahead of it will be served by the system. As it is shown in Figure 4.4 this is covering the residual service time of an IP packet currently being served; the sum of $(i - 2)$ service times of the IP packets ahead of the tagged one and the service time of the tagged IP packet. By knowing all these quantities the distribution of the waiting time can be obtained as the convolution of $(i - 1)$ service times and the residual service time.

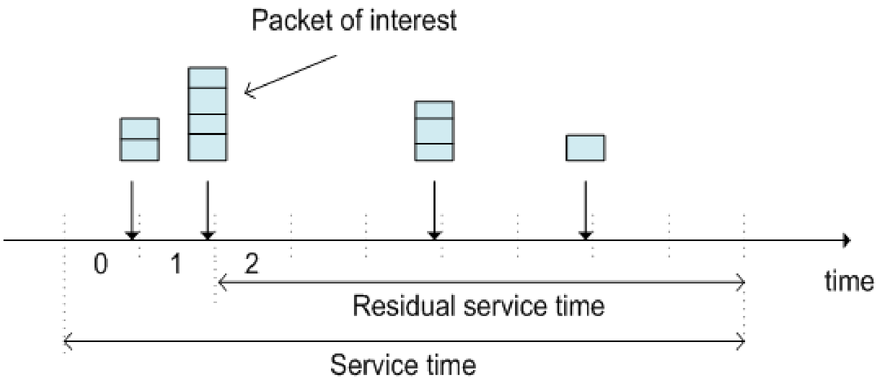


Figure 4.4*: Residual service time from the packet of interest point of view* [6].

4.3. Simulating the performance evaluation model

If we want to express the general objective of our simulations, it can be summarized as the performance evaluation of a wireless VoIP flow at IP layer of the protocol stack, considering the effect of loss and delay as the most significant impairments experienced by IP packets at buffer and the Physical Layer (wireless channel) in the presence of FEC and ARQ error recovery mechanisms implemented at the data-link layer. Indeed, we consider the effect of the lower layer protocols on the performance provided to the IP layer. Since these three layers of the protocol stack are contributing in performance qualification, our model is a cross-layer performance evaluation system. The effect of all mentioned events and mechanisms are considered through simulations as follows. At first the effect of IP packet loss at the wireless medium as a single packet loss probability value and the delay experienced by IP packets due to transmission over the wireless channel as the probability function of the number of time slots required to transmit them are obtained through simulations as the effects of the wireless medium (Physical layer). Two types of ARQ error recovery mechanisms are considered as Hybrid ARQ Type I and Type II and the loss and delay performance metrics are computed for each of them based on our choice. Additionally, at this step we set bit error rate (ranges from *0.01* to *0.1*), number of ARQ retransmission attempts (ranges from *2* to *10*), the lag-1 NACF of the bit error process and the IP packet size for our simulations.

At the second step we simulate the queuing system through a discrete-event simulator. The probability function of the number of time slots required to transmit IP packets obtained in the previous step is used as the service times for this discrete-event simulation. VoIP traffic is simulated as exponentially distributed ON-OFF sources demonstrating the speech and silence periods with means of *1.4* and *2.3* seconds respectively. The arrivals for discrete-event simulator are characterized by these exponentially distributed ON periods. In the case of buffer overflow IP packets will be dropped. Therefore, the effect of packet loss at the IP layer is also considered. Then, we get a trace of ones and zeros by saving them in an output file as the output indicating lost and successfully received IP packets respectively.

Thereafter, in the third step these traces are evaluated through another simulation to realize good/bad periods and compute some quality evaluation parameters such as the number of loss and no-loss states (dependent to used codec type), the overall loss ratio and the loss ratio for loss and no-loss states, R-factor quality metric through simple packet loss rate (PLR) model and integrated loss model (Clark) and other parameters of E-model to compare their efficiency and delicacy in performance estimation.

The logic behind realization of good and bad periods during a VoIP session is that if the number of successive successfully received IP packets between two lost ones is at least some predefined value (*16* in our simulations), then that period is realized as a good one from the perceived quality point of view. Otherwise, if the number of consecutive successfully received IP packets between two lost IP packets is less than *16*, that period is realized as a bad one.

In this thesis work we concentrate on the effect of some important determinative parameters such as the number of ARQ retransmission attempts, the type of ARQ error concealment mechanism, codec type, bit error rate, lag-1 normalized autocorrelation function of the bit error process and the error correction capability of the FEC code on the perceived VoIP quality.

We obtained a lot of diagrams as the results of simulations that help us to better understand the effect of various performance metrics on quality perception and interdependencies between them. In the next chapter we present some of these diagrams.

Chapter 5

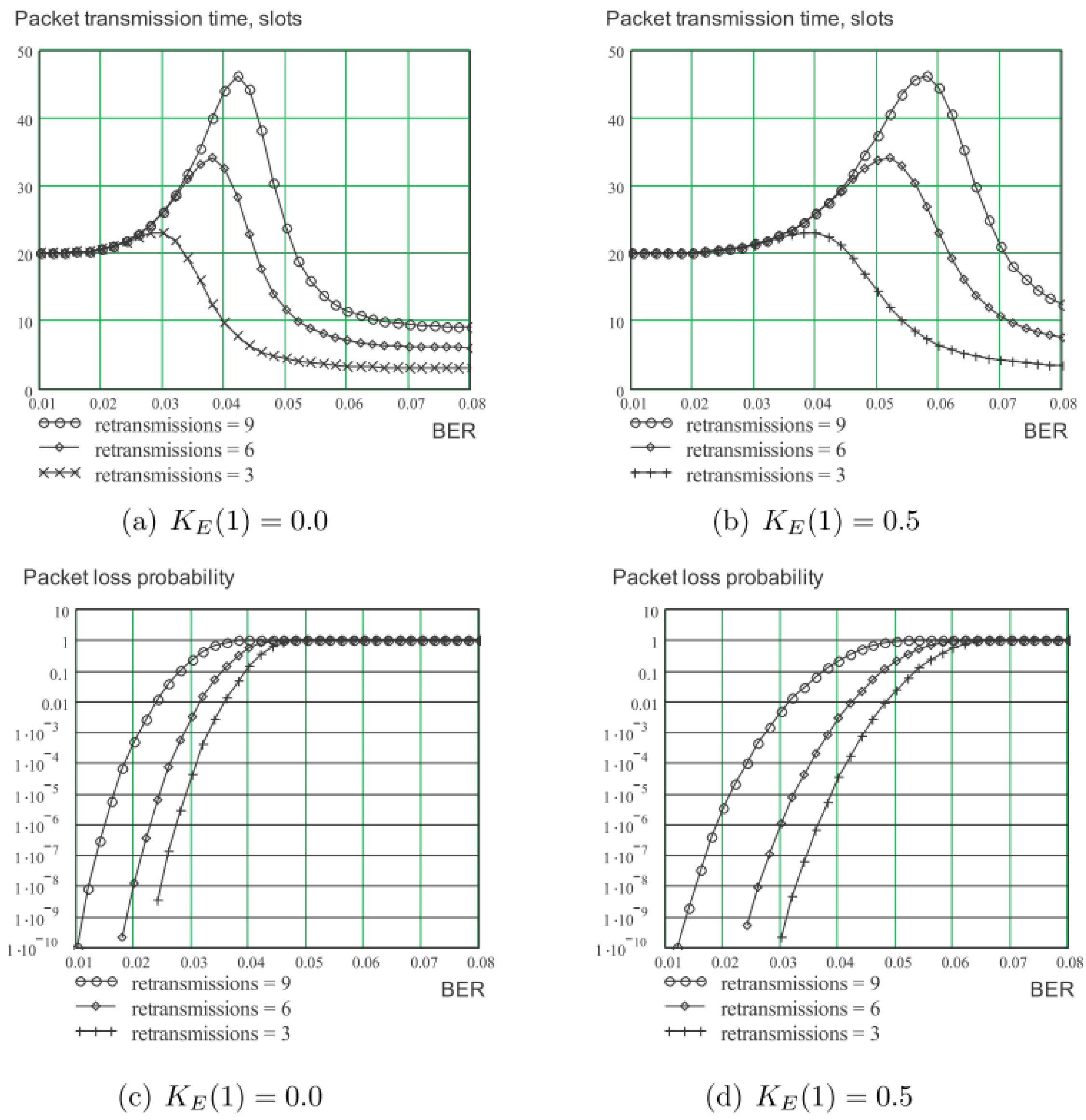
Results and discussion

In this chapter we present diagrams obtained as results of our simulations. They help us to better understand the effect of various performance metrics and processes on quality perception and interdependencies between them, not achievable through analytical analysis.

5.1. Simulation results and discussion

At first we refer to the spacious diagrams presented in [6] which are obtained as a result of simulating the wireless channel effects indicated by probability function of IP packet transmission times in terms of the number of required time slots and IP packet loss probability as a single parameter.

At first, we consider the effect of the number of ARQ retransmission attempts. In Figure 5.1 the diagrams for the mean packet transmission time and the packet loss probability due to transmission over the wireless channel are presented as a function of number of retransmissions and bit error rate for packet size set to *400* bytes, RS code *(40, 20)* and lag-1 NACF set to *0.0* and *0.5*. As it is clear from the diagrams, the mean packet transmission time (presented as the number of time slots) is more or less constant for low BERs. The reason is the sufficient correction capability of the FEC code associated with the HARQ scheme to successfully decode a frame in its first few transmission attempts. However, by increasing the bit error rate the mean packet delay starts to grow and eventually approaches its peak value which is known as the turning point. Most IP packets are successfully received before reaching this point while after this point most of them are lost as a result of excessive number of retransmission attempts made for a certain frame. The interesting point in these diagrams is that the packet loss probability grows quite rapidly before the turning point while it remains approximately constant after this point, even by increasing the bit error rate. The more interesting fact is exponentially fast degradation of the mean packet delay after this point when BER increases even further. The natural reason behind these behaviours is that all IP packets are dropped as a result of failure to transmit their first frames.



(a) $K_E(1) = 0.0$ (b) $K_E(1) = 0.5$

(c) $K_E(1) = 0.0$ (d) $K_E(1) = 0.5$

Figure 5.1: *packet delay and loss response for different number of retransmissions* [6].

As it is quite clear from the diagrams, by increasing the number of retransmission attempts the turning point occurs for higher bit error rates. Actually, we can realize that if we consider a system with unlimited number of ARQ retransmission attempts starting from the turning point the mean packet delay would still grow exponentially fast. By utilizing a system with Type I HARQ which limits the number of ARQ retransmission attempts for a packet there would be a limit for growth of the mean packet delay. As it can be realized from the diagrams the behaviour of the packet loss probability is relatively similar for all truncated Type I HARQ schemes. Although Type II HARQ system behaves approximately similar to Type I HARQ, the performance of it is better than that of Type I HARQ for higher bit error rates. The reason for such behaviour is that each retransmission carrying new information increases the probability of successful frame reception. However, for small values of bit error rate Type I HARQ system outperforms Type II HARQ in terms of both IP packet loss probability and the mean IP packet transmission time.

At the end, we can refer to Figure 9 in [6] demonstrating the effect of the IP packet size on loss and delay performance metrics. For this experiment the number of retransmission attempts was set to *9* and RS FEC code was chosen as *(40, 20)*. The diagrams are obtained for *0.0* and *0.5* values of lag-1 NACF. Actually, the effect of the IP packet size is predictable and higher IP packets lead to worse performance in terms of the IP packet loss probability and the mean packet transmission time. However, the interesting point is the different magnitude of the effect for IP packet loss and delay performance metrics. As it can be seen from the diagrams, compared to the IP packet loss probability the mean IP packet transmission time is affected quite more. In overall, changing the size of the IP packet is one of the ways to affect the performance of applications in wireless environments.

In all of the presented diagrams there is a common fact and it is that in the case that the FEC code is not well-optimized for the system, for instance in the case that bit error rate is quite high for a given IP packet size, FEC code, and the number of retransmission attempts, the mean packet delay decreases significantly to a certain value. However, this phenomenon does not lead to better performance due to approaching of the IP packet loss probability to one. It is remarkable that ARQ protocols operating over media with non-negligible bit error rate are characterized by such property and although for Type II HARQ system this effect is qualitatively similar, the magnitude of effect is noticeably smaller.

Since the distribution of the IP packet transmission delay at the wireless channel serves as the service time distribution of the discrete event simulator for queuing system at the IP layer it is beneficial to be considered in detail. The effect of changing BER for constant lag-1 NACF of the bit error process and the effect of lag-1 NACF of the bit error process for constant BER on the structure of the probability function of the IP packet transmission delay are discussed in [6]. From the diagrams demonstrated in [6] it can be realized that in addition to the mean IP packet transmission time higher moments are affected too. Generally, it can be concluded that both bit error rate and lag-1 NACF value of the bit error process may affect the structure of probability function of IP packet transmission time, while their effect differs as discussed below.

Actually, for low values of bit error rate, e.g. *0.01* the probability function of the wireless channel transmission time (delay) approaches to zero. As, in such situation the FEC code has the ability to correct most errors in few additional retransmissions. Naturally, for higher bit error rates the mean of the IP packet transmission time increases and its probability function spans more widely around the mean. The evident reason for such behaviour is the need for more retransmission attempts to transmit a packet successfully. Nonetheless, most of IP packets are still successfully received in this situation. However, by increasing the bit error rate even further the probability function of IP packet transmission time approaches to the maximum number of allowable retransmission attempts. As an example we refer to the bit error rate *0.05* in

Figure 10 (a) in [6]. In this situation most of IP packets get lost and for even higher bit error rates all of them are lost. In this case the probability function of IP packet transmission time degrades to a certain value.

To analyze the effect of autocorrelation we refer to Figure 10 (c) in [6]. The effect of small values of lag-1 NACF of the bit error process (less than *0.2*) on the probability function of IP packet transmission time is not significantly sensible. It is noticeable that by increasing the value of lag-1 NACF of the bit error process more and more IP packets are received successfully and the effect on the probability function of IP packet transmission time is more significant and distinguishable. It can be interpreted as the phenomenon of grouping of correctly and incorrectly received IP packets. Even less retransmissions attempts are required for successful reception of IP packets in the case of even higher values of lag-1 NACF. These observations are evidences for concluding that the presence of autocorrelation in the bit error process leads to better performance at the higher layers (IP layer).

After analyzing the effects of the wireless channel, at this step we present and analyze the diagrams obtained as results of simulations of IP packet buffering at the IP layer and performance evaluation based on R-factor perceived quality metric for both simple packet loss rate model and the more advanced integrated loss metric Clark's model. Firstly, it is required to mention the parameter types and values used in our simulations. As summarized earlier in Table 1, we used two types of standard ITU-T voice codecs. Two types of hybrid ARQ error recovery mechanisms were supposed. Bit error rate ranges from *0.01* to *0.1* and the maximum number of retransmissions is limited to *10*. In our simulations *0.0*, *0.5* and *1.0* lag-1 normalized autocorrelations were assumed as none, mid and severe levels for auto correlated bit error observations. We used the constant IP packet size of *200* Bytes in our simulations. However, the effect of IP packet size on performance control was discussed as a reference to [6]. The number of VoIP flows was set to *5* and *20*. As the maximum buffer capacity *5* and *20* IP packets were used. The rate of the wireless channel in our simulations was computed based on the FEC code, codec rate and the number of VoIP flows accordingly.

Figure 5.2 demonstrates the graphs for two different types of codecs obtained with Type I ARQ, capacity of *20* IP packets, *5* VoIP flows. The x and y axes represent the bit error rate and the average length of loss period respectively.

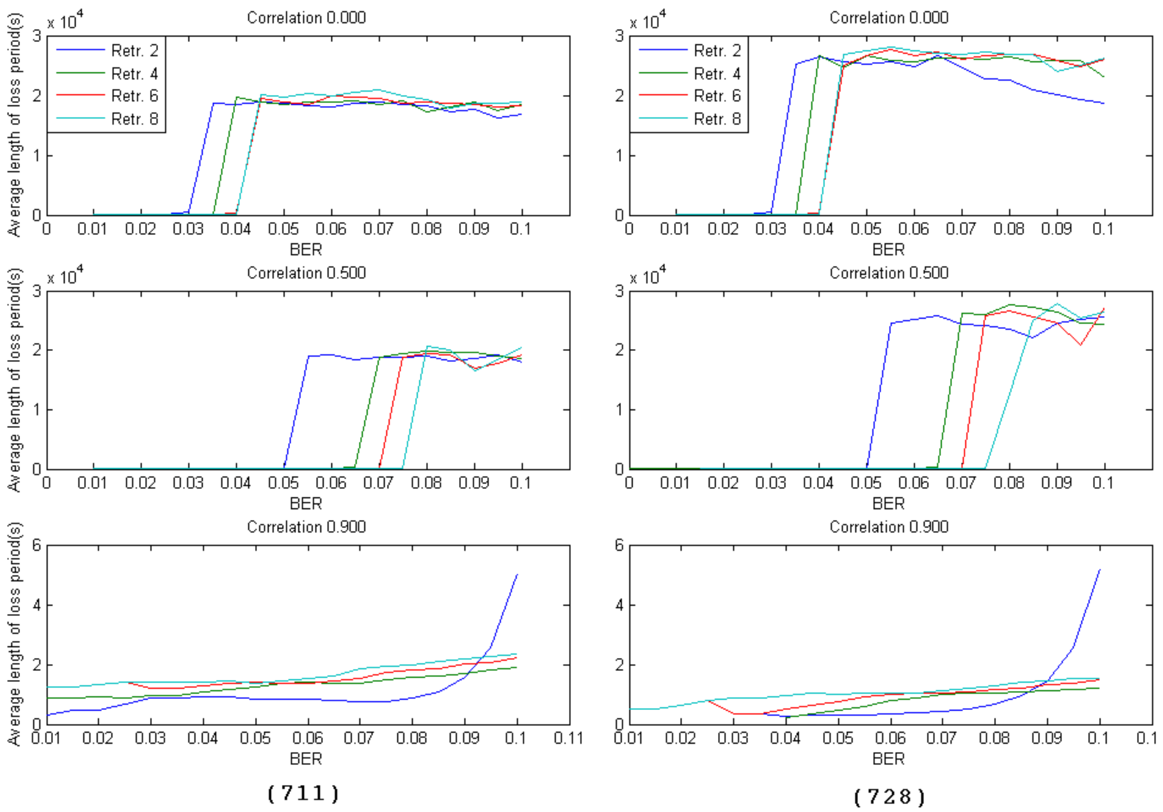


Figure 5.2: *The average length of loss period for different types of codecs with Type I HARQ, capacity of 20 and 5 flows.*

By increasing the bit error rate the average length of loss period increases accordingly, as it is expected. It is also obvious from the graphs by increasing the number of ARQ retransmission the average length of loss period decreases reasonably. As, the losses occurring at the data-link layer decrease.

As mentioned earlier the positive effect of bit error correlations is also evident in these graphs. Additionally, one can see that the codec G.711 outperforms G.728 in terms of the average length of loss period.

Figure 5.3 represents the probability of packet loss ratio for two types of codecs with settings the same as the previous case. As it is expected one can see the positive effect of highly correlated bit error observations and more retransmission attempts. It is also clearly obvious that the codec G.728 outperforms G.711 in terms of the probability of packet loss ratio.

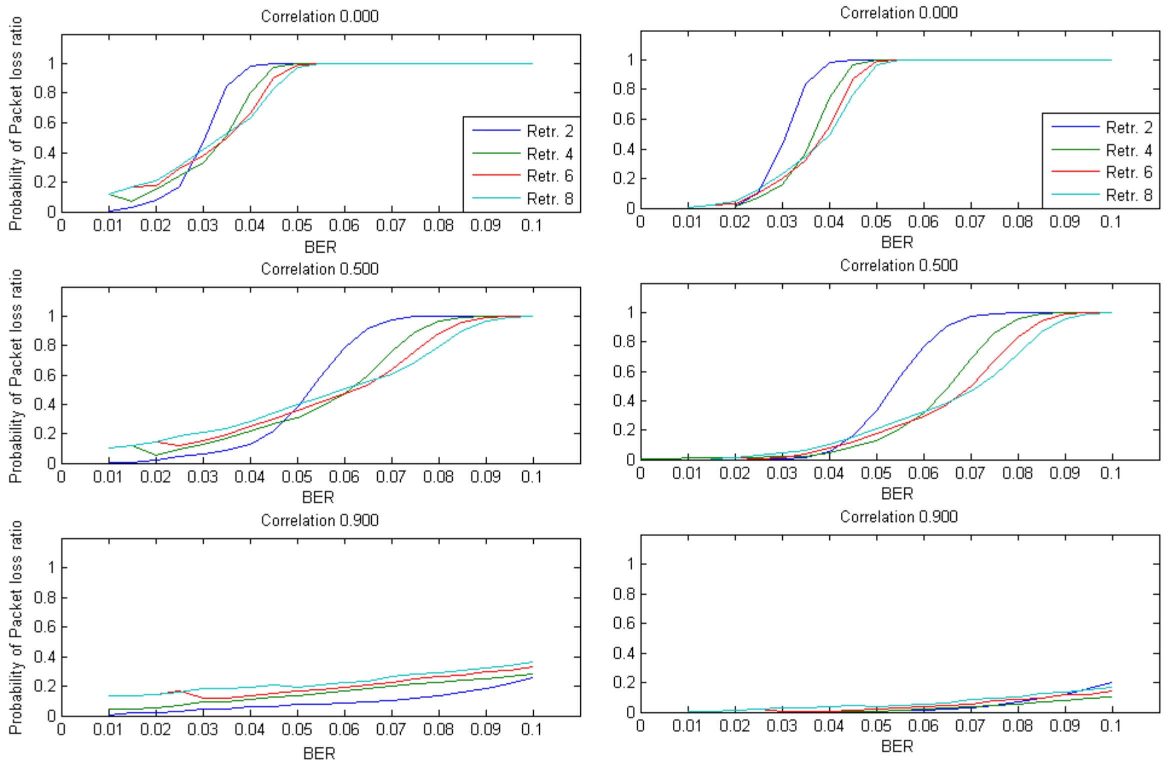
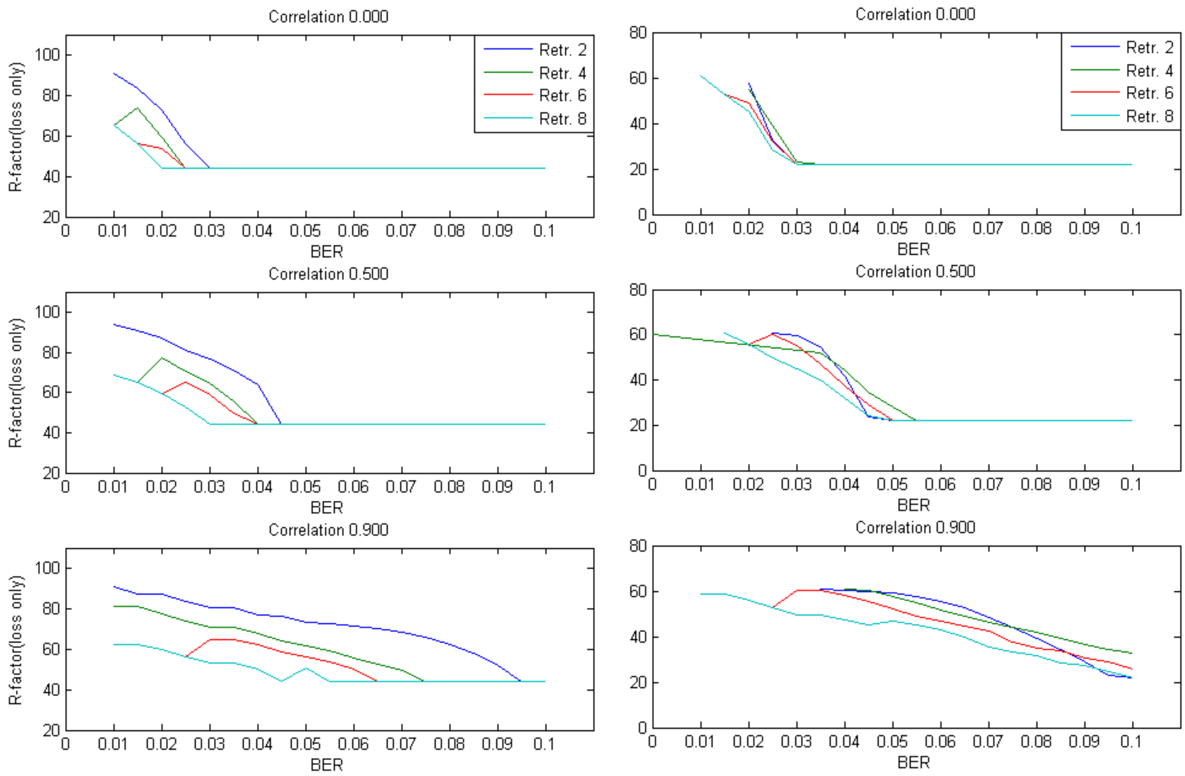


Figure 5.3: *The probability of packet loss ratio for different types of codecs with Type I HARQ, capacity of 20 and 5 flows.*

In Figure 5.4 R-factor quality metric for two types of codecs is represented obtained based on the packet loss rate approach with the same settings for the previous cases. As mentioned earlier in this simple model loss impairment factor is computed based on the ratio of the lost packets to all transmitted ones. The graphs clearly show the positive effect of higher correlations of bit error process. The interesting point in these graphs is the degradation of R-factor quality metric by increasing the number of ARQ retransmission attemps. The reason for such behaviour is that by increasing the number of retransmission attempts the number of packet losses occurring at the data-link layer decreases as well. However, this results in the increased time till successful delivery or packet drop due to the excessive number of retransmissions and this leads to the increased length of the buffering queue accordingly. Eventually, the increased probabilities of buffer overflow and packet losses eventuate to R-factor degradation. One can realize the complicated interplay between packet losses at the data-link layer and the IP layer. Due to complex relations and inter-dependencies among various processes and parameters in different layers of the protocol stack, we chose simulation approach to address them and come up with reasonable quality estimations.

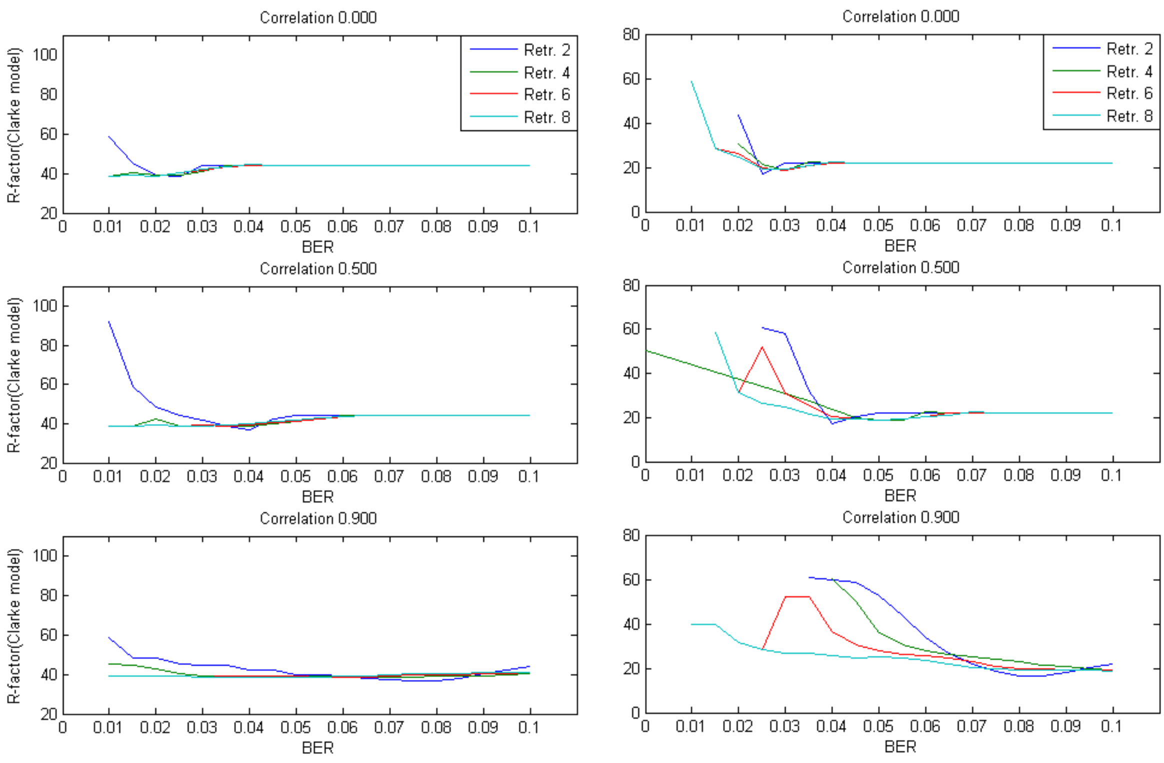


(711) (728)

Figure 5.4: *R-factor quality metric (PLR model) for different types of codecs with Type I HARQ, capacity of 20 and 5 flows.*

Figure 5.5 shows the same graphs obtained based on the Clark's model (integrated loss metric). This model is the more advanced one compared to the simple packet loss rate approach which achieve more precise performance estimation by grouping loss statistics as loss and no-loss periods and taking into account the effect of loss correlation by computing I_e loss impairment factor as the time averaged of I_e's of all periods. I_e

In addition to the positive effect of the bit error correlation, one can realize the effect of the increased number of ARQ retransmission attempts on R-factor degradation also in these graphs.



(711) (728)

***Figure 5.5**: R-factor quality metric (Clark's model) for different types of codecs with Type I HARQ, capacity of 20 and 5 flows.*

The graphs demonstrated in Figure 5.4 and Figure 5.5 show how far these two approaches are from each other. This is also obvious from the graphs in Figure 5.4 and Figure 5.5 that codec G.711 generally outperforms G.728 codec in terms of perceived quality factor. However, it does not mean that G.711 is the bet codec for use in VoIP systems. Since the system conditions and statistics are highly dynamic and this affects the system requirements dependently. For instance, high bandwidth consumption is considered as the weak point of G.711 codec.

The graphs in Figure 5.6 represent the average length of loss period for G.711 codec obtained with different buffer capacities and no. of flows. The positive effects of highly correlated bit errors and increased number of ARQ retransmissions on performance improvement in terms of the average length of loss period can be seen in these graphs. It can be realized from the graphs that the positive effect of the increased buffer size is more evident for higher values of bit error correlations.

Figure 5.7 represents the probability of packet loss ratio for G.711 codec with Type I HARQ, different buffer capacities and no. of flows. In addition to the positive effect of the highly correlated bit errors and increased number of ARQ retransmission attempts, one can realize the expected positive effect of higher buffer capacities on performance improvement in terms of the probability of packet loss ratio. The positive effect of the increased buffer size is also obvious in these graphs. Please note that the scales of axes are different from each other in some figures.

In Figure 5.8 R-factor quality metric based on the Clark's model with Type II HARQ, capacity of 20 and 5 flows for different types of codecs is demonstrated. One can realize that codec G.711 outperforms G.728 codec in terms of R-factor quality metric. It can be realized that Type II HARQ acts a bit better than Type I and results in higher values of R-factor.

In Figure 5.9 which represents the graphs for the probability of packet loss ratio with Type II HARQ, capacity of 20 and 5 flows for different types of codecs, in addition to the evident positive effects of highly correlated bit errors and more ARQ retransmission attempts, one can see that for higher numbers of ARQ retransmissions codec G.728 act better than G.711 in terms of the probability of packet loss ratio. Furthermore, by a bit attention it can be realized that for higher numbers of ARQ retransmission attempts Type II HARQ outperforms Type I HARQ in terms of the probability of packet loss ratio.

The average length of loss period with Type II HARQ, capacity of *20* and *5* flows for different types of codecs is demonstrated in Figure 5.10. By comparing these graphs with similar ones represented for Type I HARQ, one can realize that Type II acts significantly better than Type I, especially for higher ARQ retransmission attempts. There is no need to express the positive effects of highly correlated bit errors and higher numbers of ARQ retransmission attempts. As they are quite clear.

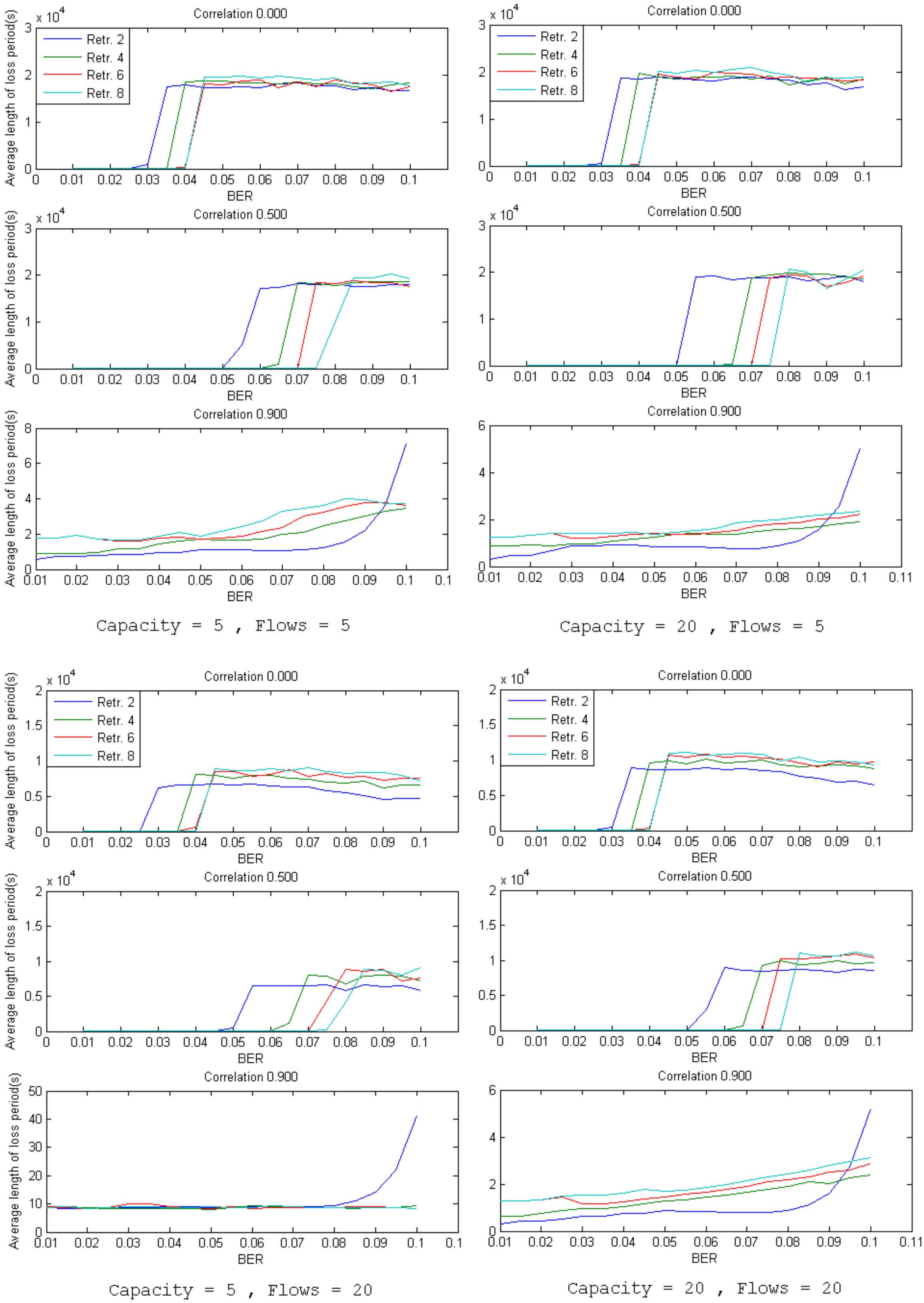


***Figure 5.6**: The average length of loss period with Type I, different buffer capacities and no. of flows for G.711 codec.*

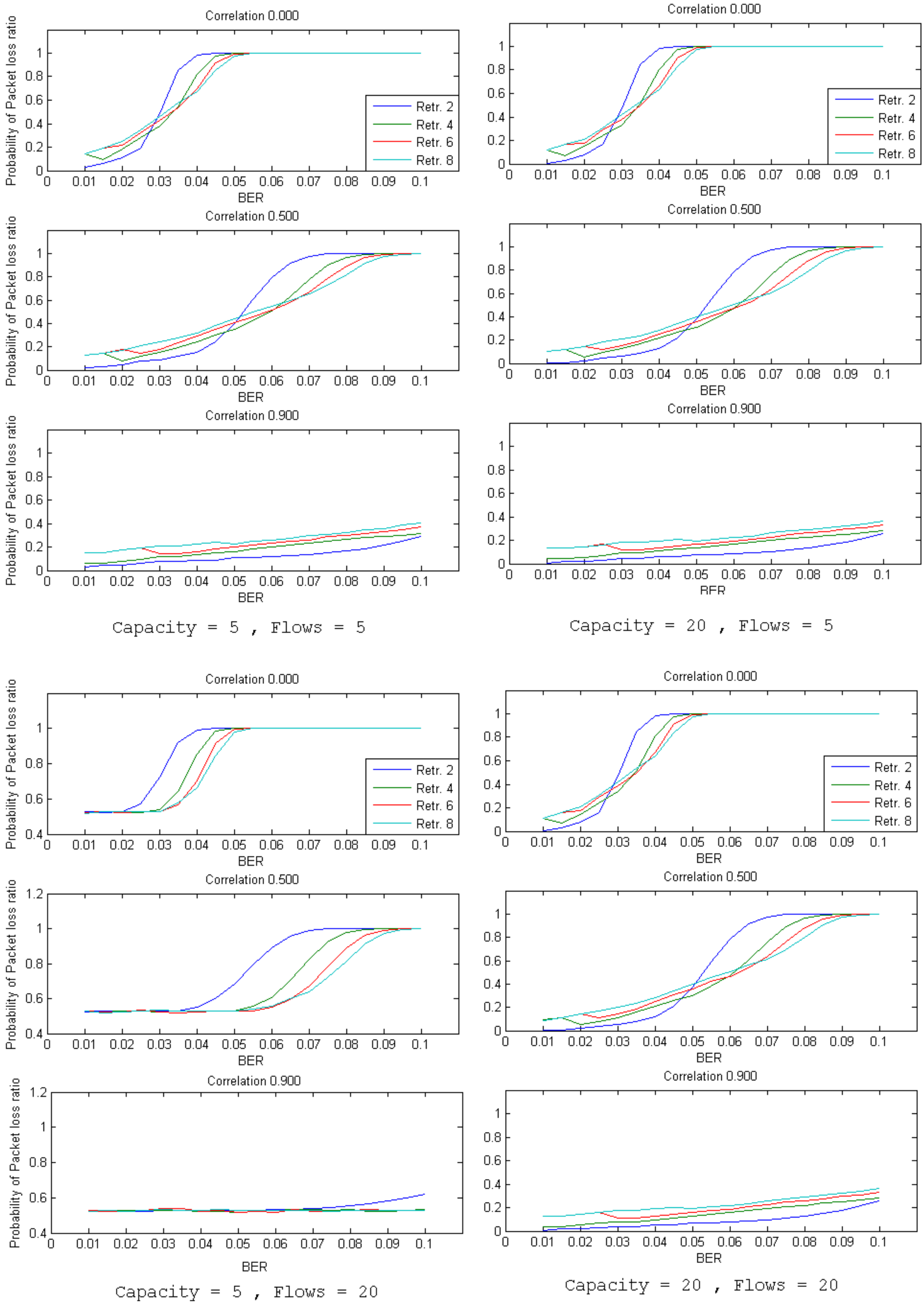
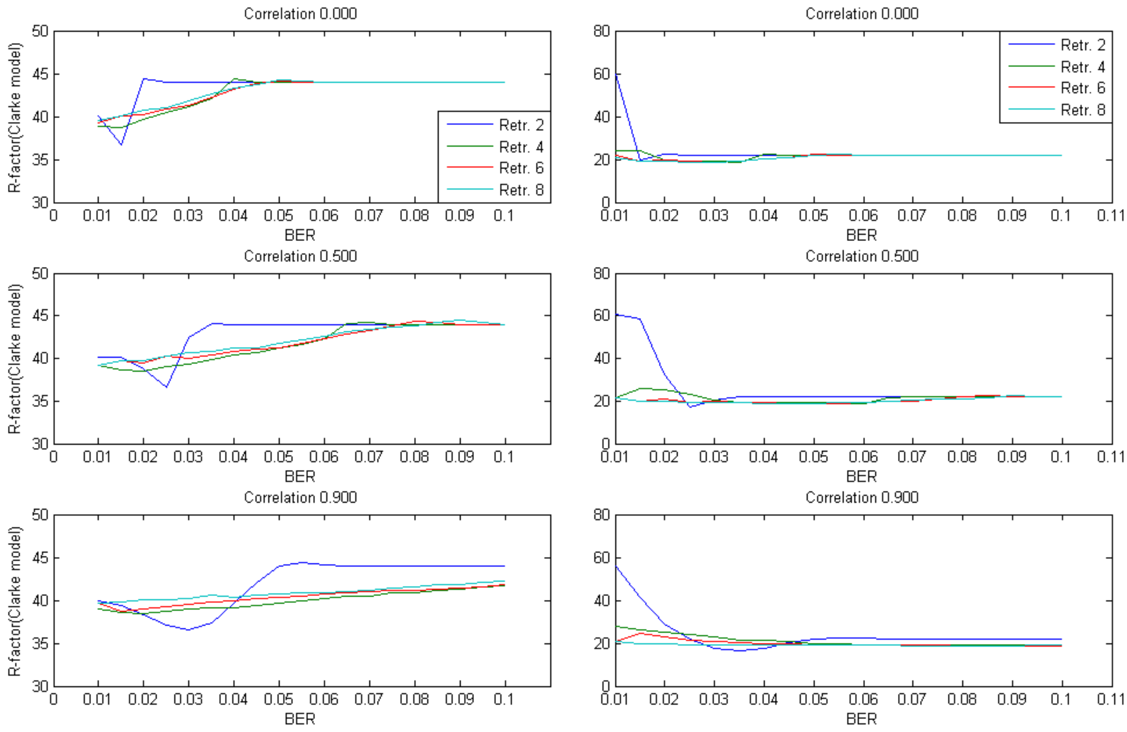
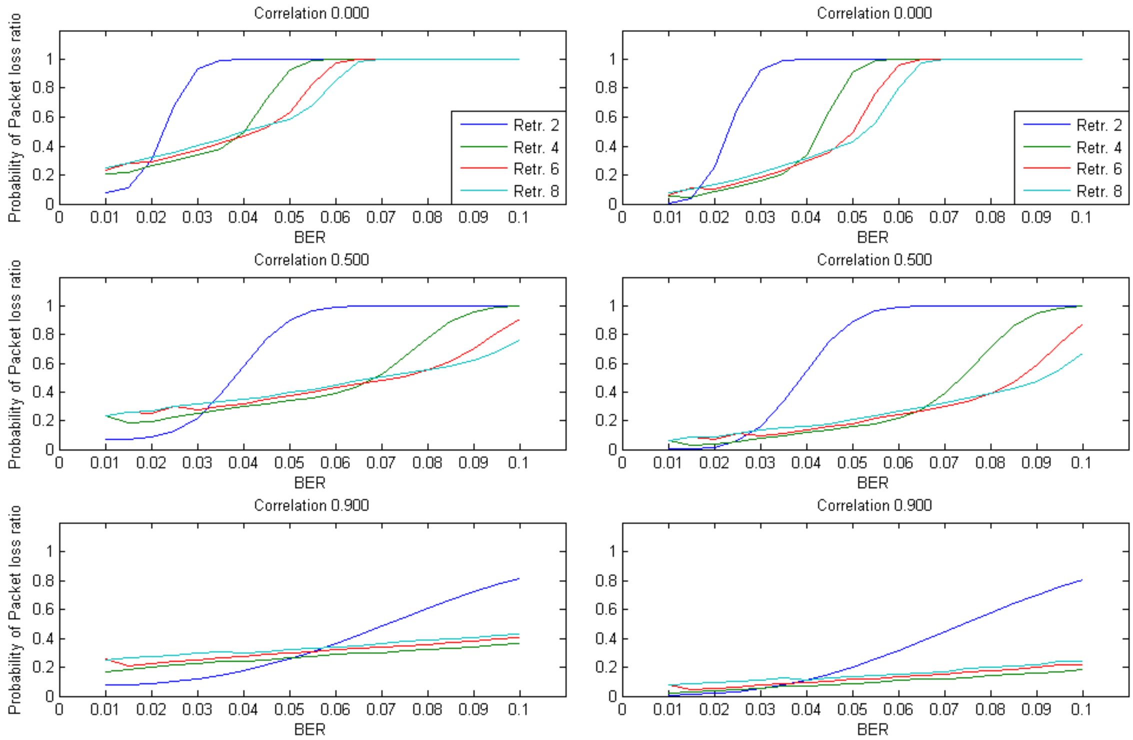


Figure 5.7: *The probability of packet loss ratio with Type I HARQ, different buffer capacities and no. of flows for G.711 codec.*



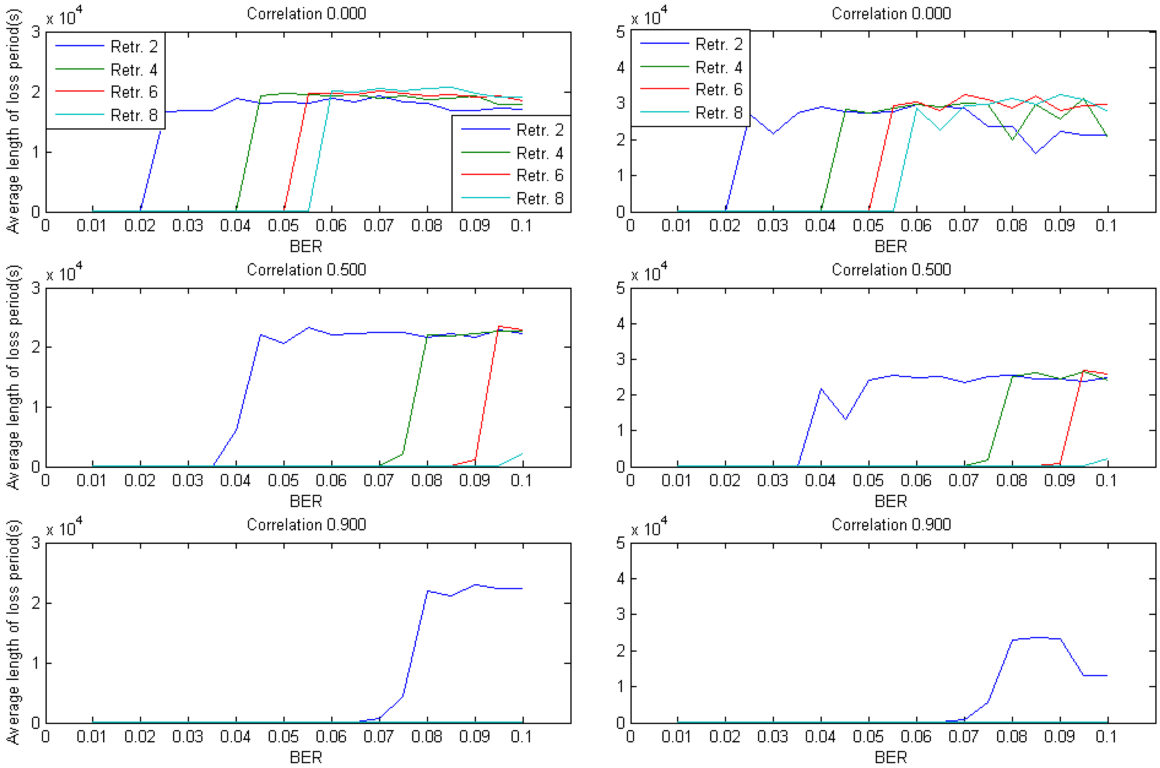
(711) (728)

***Figure 5.8**: R-factor quality metric based on the Clark's model with Type II HARQ, capacity of 20 and 5 flows for different types of codecs.*



(711) (728)

***Figure 5.9:** The probability of packet loss ratio with Type II HARQ, capacity of 20 and 5 flows for different types of codecs.*



(711) (728)

Figure 5.10*: The average length of loss period with Type II HARQ, capacity of 20 and 5 flows for different types of codecs.*

R-factor quality metric based on the PLR model for G.711 codec with Type I HARQ and different values for capacity and no. of flows is represented in Figure 5.11. The interesting point in these graphs is that for higher values of bit error correlations the positive effect of the increased buffer capacity is more evident and sensitive.

R-factor quality metric based on the Clark's model for G.711 codec with Type I HARQ and different values of capacity and no. of flows is demonstrated in Figure 5.12. In these graphs the positive and negative effects of the increased buffer space and increased number of flows are more evident for less ARQ retransmission attempts.

By comparing the graphs in Figure 5.13 representing R-factor quality metric based on the PLR model with capacity of *20*, no. of flows *5* and Type II HARQ for different types of codecs with similar ones represented in Figure 5.4 for Type I HARQ, one can realize that better R-factors are achieved with Type I HARQ, especially for G.711 codec.

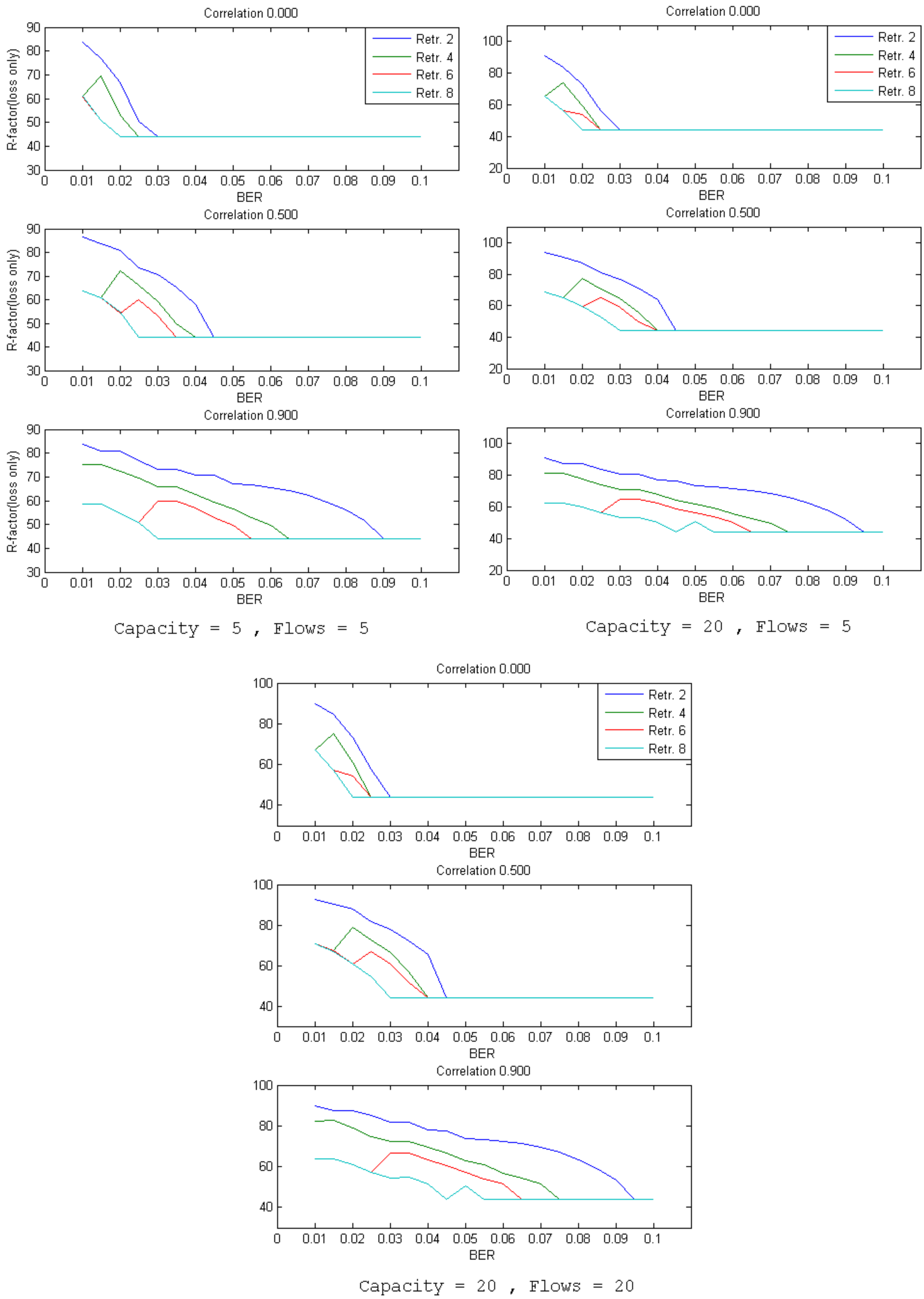


Figure 5.11: *R-factor quality metric based on the PLR model for G.711 codec with Type I HARQ and different values for capacity and no. of flows.*

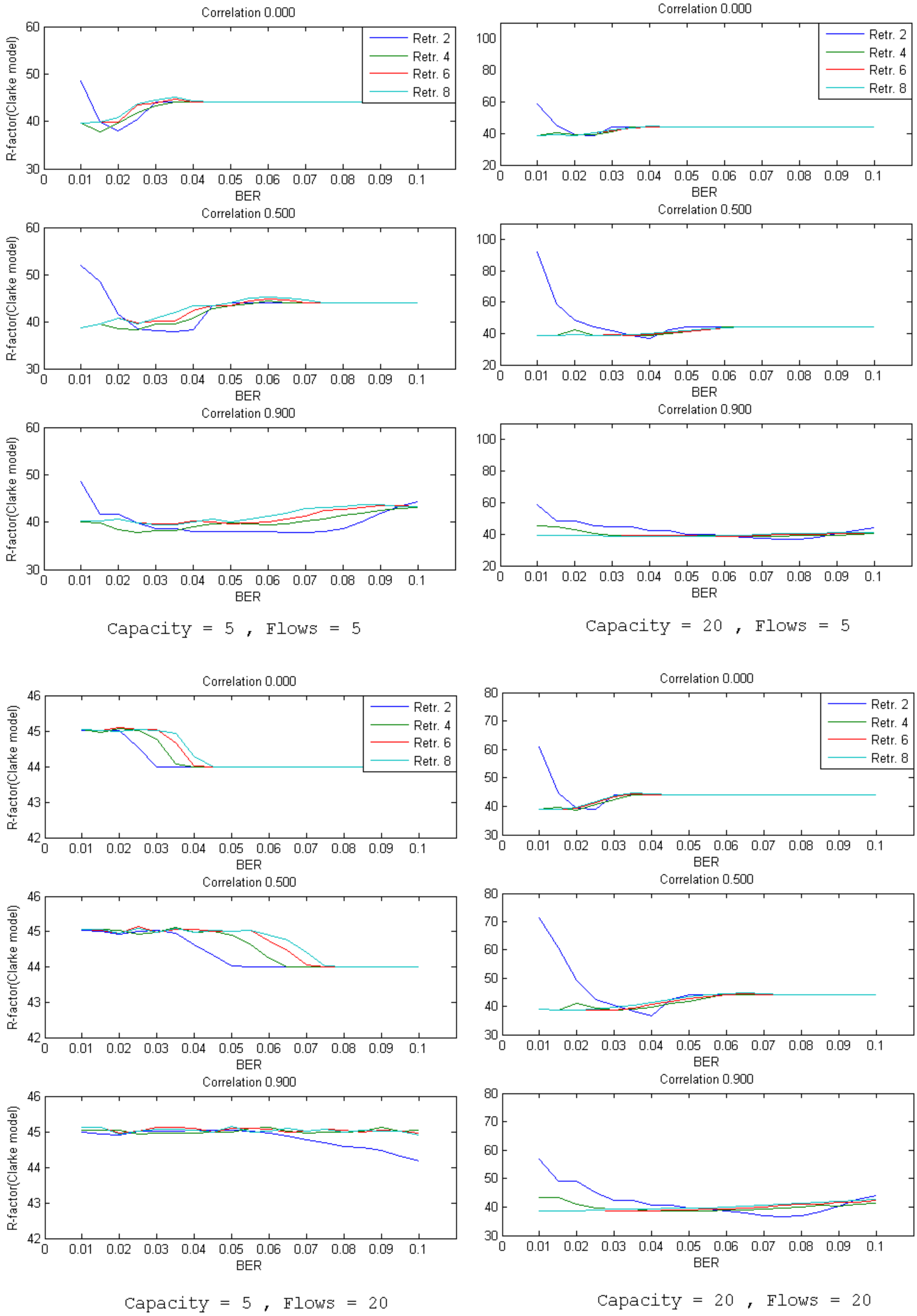
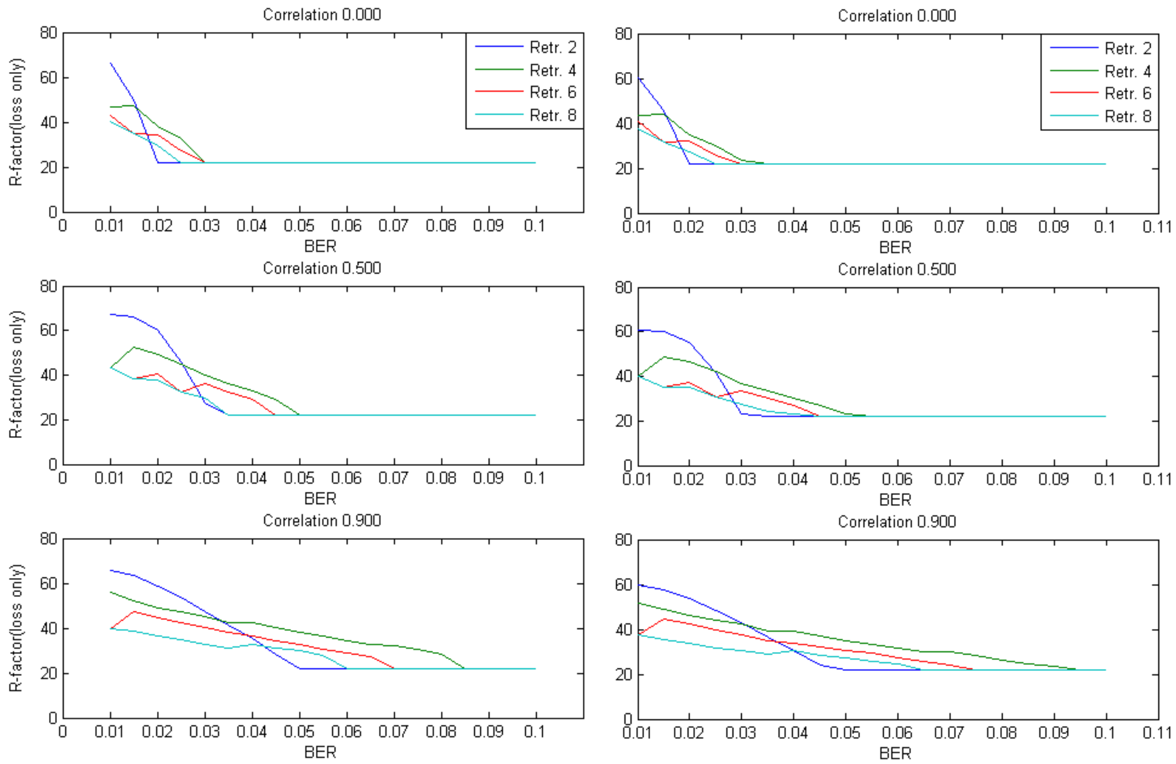


***Figure 5.12**: R-factor quality metric based on the Clark's model with Type I HARQ and different values of capacity and no. of flows for G.711 codec.*



(711) (728)

***Figure 5.13**: R-factor quality metric based on the PLR model with capacity of 20, no. of flows 5, Type II HARQ for different types of codecs.*

As a general realization from presented graphs, we realize that it is not possible to choose some parameter values and processes types and lay in that these are the best ones to use in the system. It arises from highly dynamic wireless environments and time-varying statistics and properties of real-time traffics and wireless channels. Therefore, exploiting dynamic performance control systems is a must to achieve optimized performance at any given instant of time.

Chapter 6

Conclusions

In this thesis work a methodology to evaluate per-source performance parameters of wireless VoIP was proposed. We considered a covariance stationary Markov channel (smooth channel) model. Although this property rarely holds in practice due to time-varying characteristics of the wireless medium, the statistical characteristics of wireless channels stay approximately the same in the case of assuming short travel distances or small time durations (see e.g. [36] Ch. 5 for further discussion).

In this thesis work we provided a simulation approach to evaluate per-source performance parameters of VoIP flows multiplexed over a single wireless channel. We modelled buffering process at the IP layer using an *M/G/1/K* queuing system. In this model it is possible to find out the effect of many parameters and processes including statistical properties of the wireless channel, bit error rate, correlation of bit error observations, FEC code, codec, various ARQ mechanisms, the number of ARQ retransmission attempts, buffer space and the number of VoIP flows. Full losses occurring as a result of the buffering process at the IP layer and imperfect error concealment at the data-link layer are considered in our model. As a result of considering the effects of contributing underlying layers processes and parameters including the utilized service process obtained as a cross-layer function of underlying layers, our approach is a cross-layer quality evaluation frame work.

In this thesis work we used E-model as a rather well-suited approach for perceived quality evaluation. R-factor quality metric of interest which is the output of E-model computed based on the Clark's model and simple packet loss rate (PLR) approach. Indeed, the difference between these two approaches is the calculation of I_e loss impairment factor. In PLR model it is computed simply based on the ration of lost packets to all transmitted ones, while the more advanced Clark's model divides loss statistics into loss and no-loss periods and I_e is computed by time averaging loss impairments of all these periods. Simulation results show the difference between these two mechanisms and that the Clark's model which takes into account the effect of loss correlation provides rather more accurate performance estimation. As presented in [21] it is possible to extend the E-model to cover the impairment effect of delay jitter and delayed perception.

The effects of the wireless channel including packet loss and transmission delay and local error correction mechanisms such as ARQ and FEC on the performance provided to the IP layer are rather predictable. Higher correction capabilities of FEC codes and higher numbers of ARQ retransmission attempts increase the acceptable operation region of a wireless channel. Based on observations and simulation results the operation range of the systems implementing non-persistent Type I and Type II HARQ can be separated into two sub ranges as overflow-dominated and wireless-dominated modes by the turning point. In the first mode IP packet losses are mostly results of buffer overflow, while in the latter case imperfect error correction mechanisms are the main reason for performance degradation. Due to significant performance degradation in the wireless dominated mode, we conclude that the design and development of dynamic performance control systems implementing different strategies such as dynamic source rate adaptation, resource allocation, variable IP packet size and adaptive ARQ schemes is a reasonable solution to cope with this severe depression and provide optimized quality at any given instant of time.

Proposing a dynamic performance optimization system providing performance metrics adaptive to time-varying wireless channel and real-time traffic conditions may be considered as an extension of this work for further studies. It means to provide an adaptive VoIP performance optimization structure with some controllable parameters and mechanisms such that in the case of noticeable performance decay adopting a policy to provide the best possible performance at any given instant of time in the presence of time-varying wireless channel and real-time traffic conditions.

Appendices[1]:

Symbol	Meaning
$E[W_X]$	Mean of $\{W_X(n)$, n = 0, 1, 2, ...}
$K_X(1)$	Lag-1 NACF of {W_X(n), n = 0, 1, 2, ...}
$\alpha_E\ and\ \beta_E$	Transition probabilities of the bit error model
m_S	Number of bits in a RS symbol
m_F	Number of RS symbols in a frame
v	Number of frames in a packet
N	Number of bytes in a packet
$f(1)$	Frame error probability
r	Number of retransmission attempts
l	Error correction threshold
K	Capacity of the queuing system (in packets)

[1] *Obtained from* [6]

References:

[1] A. Clark, “Modelling the effects of burst packet loss and recency on subjective voice quality”, In IP Telephony Workshop, page available at: http://www.ocecpr.org.cy/media/documents/Reports/, 2001.

[2] A. Konrad, B. Zhao, A. Joseph, R. Ludwig, “Markov-based channel model algorithm for wireless networks”, Wireless Networks, 9(3):189–199, 2003.

[3] “Application of the Emodel: a planning guide”, ITU-T Recommendation G.108, SEP 1998.

[4] A. Raake, “ Short- and long-term packet loss behavior: Towards speech quality prediction for arbitrary loss distributions”, IEEE Trans. Audio, Speech and Language Proc., 14(6):1957–1968, 2006.

[5] B.J. Kim, “A network service providing wireless channel information for adaptive mobile applications”, In Proc. IEEE ICC’01, pages 1345–1251, Helsinki, June 2001.

[6] D. Moltchanov, ”Aggregated and per-source performance modelling for broadband wireless channels”.

[7] D. Moltchanov, “Cross-layer performance control of wireless channels using active local profiles”.

[8] D. Moltchanov,” Problems arising in evaluating perceived quality of media applications in packet networks”

[9] D. Moltchanov, “State description of wireless channels using change-point statistical tests”, In Proc. WWIC’2006, pages 275–286, Bern, Switzerland, May 2006.

[10] D. Moltchanov, “The structure of the reactive performance control system for wireless channels”, In Proc. NEW2AN’06, pages 164–176, 2006.

[11] D. Moltchanov, Y. Koucheryavy, J. Harju, “Loss performance model for wireless channels with autocorrelated arrivals and losses”, Comp. Comm., 29(13–14):2646–2660, Aug. 2006.

[12] E. D. Puschita, T. P. Palade, “QoS Perspective for Wireless Scenarios”, Broadband Europe 2005, Whitepapers, 5 pages, Bordeaux, 12-14 December 2005.

[13] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, B. Girod, “CROSS-LAYER DESIGN OF AD HOC NETWORKS FOR REAL-TIME VIDEO STREAMING”, STANFORD UNIVERSITY.

[14] F. Granelli 1 , D. Kliazovich, J. Hui, and M. Devetsikiotis, “Performance Optimization of Single-Cell Voice over WiFi Communications Using Quantitative Cross-Layering Analysis”.

[15] G. Wu, Y. Bai, J. Lai, and A. Ogielski, “Interactions between TCP and RLP in wireless Internet”, In Proc. IEEE GLOBECOM’99, pages 661–666, Rio de Janeiro, Dec. 1999.

[16] Information technology – coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s – Part 1: Systems. Technical report, ISO/IEC 11172-1, 1993.
[17] K. Chen, S. Shan, K. Nahrstedt, “Cross-layer design for accessibility in mobile ad hoc networks”, Wireless Pers. Comm., 21(1):49–76, Apr.2002.
[18] K. Kilkki, “Next Generation Internet and Quality of Experience”, EuroFGI IA.7.6 Workshop on Socio-Economic Issues of NGI, Santander, Spain, June, 2007.
[19] Dmitri aggregated [10] L. Badia, M. Rossi, M. Zorzi, “SR ARQ packet delay statistics on markov channels in the presence of variable arrival rate”, IEEE Trans. Wireless Comm., 5(7):1639–1644, July 2006.
[20] L. Besacier, C. Bergamini, D. Vaufreydaz, E. Castelli, The effect of speech and audio compression on speech recognition performance, IEEE Multimedia, 2001.
[21] L. Ding, R. A. Goubran, “Speech Quality Prediction in VoIP Using the Extended E-Model”.
[22] Leon-Garcia and Widjaja, Communication Networks: Fundamental Concepts and Key Architectures, McGraw-Hill, 2000, ISBN 0070228396
[23] L. Sun, E. Ifeachor, "New Methods for Voice Quality Evaluation for IP Networks", Proc. of 18th International Teletraffic Congress (ITC18), Berlin, Germany, SEP 2003.
[24] Methods for subjective determination of transmission quality. ITU-T Rec. P.800. 1996.
[25] M. Gineste, N. Van Wambeke, E. Exposito, C. Chassot, L. Dairaine, “A Cross-Layer Approach to Enhance QoS for Multimedia Applications Over Satellite”
[26] M. Luby, L. Vicisano, J. Gemmell, L. Rizzo, M. Handley and J. Crowcroft, RFC3453: The Use of Forward Error Correction (FEC) in Reliable Multicast, DEC 2002.
[27] “One-way transmission time”, ITU-T recommendation G.114, ITU-T, 1996.
[28] “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs”, ITU-T Recommendation, P.862.
[29] P. Sudame and B. Badrinath, “On providing support for protocol adaptation in mobile wireless networks”, Mob. Netw. and Appl., 6(1):43–55, Jan./Feb. 2001.
[30] R. Reynolds, A. Rix, Quality VoIP—an engineering challenge, BT Technology Journal 19 (2) (2001).
[31] Schulzrinne H., Castner S., Frederick R and Jacobson V.,RFC 1889: RTP: a Transport Protocol for Real-time Applications, 1996
[32] Smiesko, Exponential Model of Token Bucket System, Communications 4/2003, pp. 66-70.
[33] Tanenbaum A.S. Computer networks, Prentice-Hall, 1996, ISBN 0-13-394248-1
[34] T. Bohnert, D. Staehle, G.-S. Kuo, Y. Koucheryavy, and E. Monteiro, “Speech quality aware admission control for fixed IEEE 802.16 wireless MAN”, In IEEE ICC, page to appear, 2008.

[35] “The E-model, a computational model for use in transmission planning”, ITU-T Recommendation G.107 (05/2000).
[36] T. Rappaport, “Wireless communications: principles and practice”, Prentice Hall, 2nd edition, 2002.
[37] V. Kawadia, P.R. Kumar, “A cautionary perspective on cross-layer design”, IEEE Pers. Comm., 12(1):3–11, Feb. 2005.
[38] V. Raisinghani and S. Lyer, “Cross-layer design optimizations in wireless protocol stacks”, IEEE Comp. Comm., 27:720–724, 2004.
[39] V. Srivastava and M. Motani, “Cross-layer design: a survey and the road ahead”, IEEE Comp. Comm., 43(12):112–119, Dec. 2005.
[40] Q. Wang, M. Abu-Rgheff, “Cross-layer signalling for next-generation wireless systems”, In Proc. WCNC’03, pages 1084–1089, March 2003.