VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Antanas VINDAŠIUS

ANALYSIS OF QUALITY OF SERVICE IN HETEROGENEOUS WIRELESS NETWORKS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
ELECTRICAL AND ELECTRONIC ENGINEERING (01T)

VGTU
LEIDYKLA
Vilnius TECHNIKA 2010

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2006–2010.

Scientific Supervisor

Prof Dr Habil Algimantas KAJACKAS (Vilnius Gediminas Technical Univesity, Technological Sciences, Electrical and Electronic Engineering – 01T).

VGTU leidyklos TECHNIKA 1809-M mokslo literatūros knyga
http://leidykla.vgtu.lt

ISBN 978-9955-28-652-9

antanas.vindasius@el.vgtu.lt

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Antanas VINDAŠIUS

PASLAUGŲ KOKYBĖS HETEROGENINIUOSE BEVIELIUOSE TINKLUOSE TYRIMAI

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
ELEKTROS IR ELEKTRONIKOS INŽINERIJA (01T)

Vilnius LEIDYKLA TECHNIKA 2010

Disertacija rengta 2006–2010 metais Vilniaus Gedimino technikos universitete.

Mokslinis vadovas

prof. habil. dr. Algimantas KAJACKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Abstract

The dissertation investigates wireless access network performance, ability to control and enforce quality of service (QoS) and track it at user device. The objective of presented research is to investigate wireless access network performance parameter impact on quality of service and develop reference model of QoS monitor, applied for end-user device. In order to reach the objective, the following tasks had to be solved:

- To analyze existing measures of QoS evaluation, actual achieved, perceived QoS monitoring and Service Level Agreement (SLA) enforcement.
- To analyze performance properties of contemporary wireless user access network technologies in order to determine their capabilities providing data services and employing soft and hard QoS mechanisms.
- To analyze network performance characteristics in order to determine factors, influencing quality of service in wireless access networks.
- To design principles of evaluation of actually achieved quality of service.
- To develop a reference model of QoS requirement composition and monitoring system.
- To develop and experimentally test QoS monitor for web browsing service.

First chapter of the dissertation reviews QoS definitions, categories, existing methods of QoS evaluation and formulates the problem of QoS and service level agreement enforcement measures.

Second chapter presents analysis of various access networks applicable to data and voice services. The analysis is focused on available bit rate, referred to capacity of the access network, capabilities to sustain real time and data services. This chapter presents modelling and experimental results for wireless user access networks.

Third chapter introduces the design of perceived QoS monitoring system, including monitoring agents residing at user device, tools for composing SLA, measures for tracking QoS impairments. The concept of QoS monitoring tool is proposed.

The last chapter is dedicated to presenting reference design of individual quality evaluation algorithm application to web browsing services backed up with experimental results.

Dissertation includes 112 pages of text, 37 equations, 8 tables, 54 figures and 106 references. 8 scientific articles have been published by the author on the topic of the dissertation.

Reziumė

Disertacijoje nagrinėjama belaidės prieigos tinklų savybės, gebėjimas kontroliuoti ir užtikrinti paslaugos kokybę (QoS) bei stebėti ją vartotojų įrenginiuose. Darbo tikslas – ištirti belaidės prieigos tinklų darbo parametrų įtaką paslaugos kokybei ir sukurti apibendrintą paslaugos kokybės stebėsenos priemonių, taikomų galutinio vartotojo įrenginyje, modelį. Šiam tikslui pasiekti, sprendžiami tokie uždaviniai:

- Išanalizuoti esamas paslaugos kokybės vertinimo, realiai patirtos ir suvokiamos kokybės stebėsenos bei paslaugų lygmens susitarimų (SLA) užtikrinimo priemones.
- Išanalizuoti dabartinių belaidžių vartotojo prieigos technologijų savybes teikiant duomenų perdavimo paslaugas ir taikant griežtus bei švelnius paslaugos kokybės užtikrinimo mechanizmus.
- Išanalizuoti belaidžių prieigos tinklų charakteristikas siekiant nustatyti faktorius, įtakojančius paslaugos kokybę belaidžiuose prieigos tinkluose.
- Suformuluoti realiai pasiektos paslaugos kokybės vertinimo principus.
- Sukurti apibendrintą paslaugos kokybės reikalavimų sudarymo ir stebėsenos sistemos vartotojo įrenginyje modelį.
- Sukurti ir eksperimentiškai įvertinti paslaugos kokybės stebėsenos sistemą interneto naršymo paslaugai.

Pirmajame disertacijos skyriuje pateikiama žinomų paslaugos kokybės vertinimo ir valdymo procesų apžvalga, koncentruojantis į vartotojo perspektyvą ir galimybes įvertinti realiai gaunamą ir vartotojo suvokiamą paslaugų kokybę.

Antrajame skyriuje pateikiama paplitusių belaidės prieigos tinklų technologijų analizė, siekiant nustatyti kokybę įtakojančius faktorius. Tiriami specialieji kokybės valdymo atvejai – balso perdavimas lokaliaisiais belaidžiais tinklais, taip pat perdavimas grandine. Šie scenarijai paryškina kiekybinių faktorių svarbą, kai turi būti užtikrinamos griežtos paketų delsos ribos.

Trečiajame skyriuje pateikiama galimybių sukurti tikralaikę patirtos paslaugos kokybės stebėsenos sistemą, įskaitant stebėsenos agentus vartotojo įrenginyje, SLA sudarymo bei kokybės pokyčių stebėjimo priemones, formuluojami reikalavimai ir pasiūlomas apibendrintas sistemos prototipas.

Paskutiniame skyriuje aprašomas pasyvios stebėsenos sistemos, skirtos interneto tinklalapių naršymo paslaugos kokybei vertinti vartotojo įrenginyje, modelis bei jo testavimas.

Disertacijos apimtis yra 112 puslapių, tekste panaudotos 37 numeruotos formulės, 54 paveikslai ir 8 lentelės. Rašant disertaciją buvo panaudoti 106 literatūros šaltiniai, disertacijos tema publikuoti 8 autoriaus straipsniai.

Notations

Symbols

B – throughput;
M – MSDU size;
n – capacity (voice channels);
O – overhead;
p – probability;
$Q(t)$ – quality time function;
Q_0 – quality threshold;
$q_i(t)$ – instantaneous quality time function;
R – physical transmission rate;
S – packet size; object size in bytes;
T – cycle duration; period;
t – continuous time; time expenditures;
t_i – discrete time;
$v(t)$ – physical signal time function;
ΔQ – quality change;
$\Delta\rho$ – available bit rate variation;
$\Delta\tau$ – packet delay variation;
κ – impairment factor;
π – packet loss rate;
ρ – available bit rate;

τ – delay; technological duration constants.

Abbreviations

3G – Third Generation;
ACK – Acknowledgment;
AP – Access Point;
APQoS – Accountable Perceived QoS;
AQoS – Achieved QoS;
BE – Best Effort;
CA – Collision Avoidance;
CAP – Controlled Access Period;
CCA – Clear Channel Assessment;
CCH – Control Channel;
CDF – Cumulative Distribution Function;
CFP – Contention-Free Period;
CP – Contention Period;
CSMA – Carrier Sense Multiple Access;
CW – Contention Window;
DCF – Distributed Coordination Function;
DL – Downlink;
DMOS – Difference MOS;
EDCA – Enhanced Distributed Channel Access;
GPQoS – General Perceived QoS;
GPRS – General Packet Radio Service;
GPS – Geo Positioning System;
HCCA – HCF Controlled Channel Access;
HCF – Hybrid Coordination Function;
HTTP – Hypertext Transfer Protocol;
IP, IPv4 – Internet Protocol, Internet Protocol version 4;
ITU – International Telecommunications Union;
MAC – Medium Access Control;
MCS – Modulation-Coding Scheme;
MPDU – MAC Protocol Data Unit;
MS – Mobile Station;
MSDU – MAC Service Data Unit;
NGN – Next Generation Network;
OQoS – QoS Offered by provider;
PCF – Point Coordination Function;
PESQ – Perceptual Evaluation of Speech Quality;
PF – Persistent Factor;
PHY – Physical Layer;
PIFS – PCF Inter Frame Space;
PPDU – Physical Protocol Data Unit;
PQoS – Perceived Quality of Service;

QoS – Quality of Service;
RTP – Real Time Protocol;
SCH – Service Channel;
SI – Service Interval;
SIFS – Short Inter Frame Space;
SLA – Service Level Agreement;
SPQoS – Special Perceived QoS;
STA – Wireless Station;
STDMA – Space-Time Division Multiple Access;
TBTT – Target Beacon Transmission Time;
TCP – Transmission Control Protocol;
TDM – Time Division Multiplexing;
TSF – Timing Synchronization Function;
TSPEC – Traffic Specification;
TTL – Time To Live;
TXOP – Transmission Opportunity;
UDP – User Datagram Protocol;
UL – Uplink;
UTC – Coordinated Universal Time;
VANET – Vehicular Ad-Hoc Network;
VBR – Variable Bit Rate;
WAVE – Wireless Access in Vehicular Environments;
WBSS – WAVE Basic Service Set;
WiMAX – Worldwide Interoperability of Microwave Access;
WLAN – Wireless Local Area Network;
WSM – WAVE Short Message;
WSMP – WAVE Short Message Protocol.

Contents

Introduction

The Investigated Problem

Mobile wireless network technologies migrate to multidimensional heterogeneity as they are widely used for multimedia applications. Different access technologies, such as GSM/GPRS or WCDMA, HSPA, WiFi and WiMAX may coexist and offer high efficiency for both high and low data rates, and for high and low traffic density configurations. These network technologies require to be inter-connected in an optimal manner with the ultimate objective to provide the end-user with the requested services and corresponding QoS requirements.

QoS in telecommunications networks is analyzed by many authors in various aspects (Cavender *et al.* 2008; Joskowicz *et al.* 2009; Kajackas *et al.* 2009b; Guršnys 2008; Anskaitis 2009). In general analysis, the quality of service is described as conceptual indicator Q. The indicators may be specified more accurately having particular services in mind (voice, video or data). The problems of QoS observed by end-user, have been analyzed as well, and are commonly known as Perceived QoS or QoS of Experience (ITU-T Rec. E.800 2008).

The concept of perceived QoS rises directly from used service, its quality perception and quality evaluation. Perceived quality as function $Q_i(t)$ is user's response to received signal, represented by delivered service. The QoS concept, defined by ITU-T Rec. E.800 as "satisfaction of the end user" is commonly accepted. User satisfaction is seen as one of the key components of usability and is expressed by parameters which focus on user-perceivable effects.

In realistic mobile heterogeneous environment, the quality perceived by user is a time function $Q(t)$, which is dependent on many factors: coding techniques, signal to noise (interference) ratio, availability, etc. Even though QoS management is enabled in most modern telecommunication systems, that is not enough to ensure perceived quality of service for every user. Service provider may not always be in a position to offer customers the level of QoS they have committed or may require, due to technical properties or loose service planning.

Service Level Agreement (SLA) concept was crated to define and manage relations between end-users and operators under variable achievable quality conditions. SLA is a formal negotiated agreement between a service provider and a customer, which defines allowed limits of main quality related service parameters. In context of SLA, the understanding of QoS is more formal and recognized as „degree of conformance of the service delivered to a user by a provider in accordance with an agreement between them" (ITU-T Rec. E.860 2002; ITU-T Rec. P.800 1996).

To make sure that operator follows SLA provisions, QoS level has to be tracked continuously. This may be done by software QoS agents M_{QoS}, residing in user devices and monitoring real-time quality of online services. However, such real-time user experience tracking tools at end-user devices do not exist. Some specific tools exist for monitoring physical link characteristics and link quality in terms of radio parameters, but these measures have been created purely for operator use and neither are accessible for end users, nor fully reflects the perceived QoS. The problem is that end-users have no means of establishing a quantitative quality value of the consumed telecommunications product and no quality accounting is made.

Importance of the Dissertation

QoS management in packet delivery networks have been a challenge in terms of rethinking resource management policies. Packet delivery in mobile wireless communications adds even more complexity to this task, because available channel resources are variable both in time and space.

The migration from hard QoS in circuit switching to soft QoS techniques in connectionless IP networks adds flexibility, scalability and saves network resources for operators, resulting in low cost and diversity of services for users. Real-time multimedia and interactive services have migrated to packet delivery networks and high capacity wireless technologies have allowed them to be served in mobile environment. In this context of heterogeneous networks delivering diversity of services, quality issue is becoming more and more relevant. Operators seek for more competitiveness and plan their networks to be as cost effective as possible, avoiding any commitment to user and basically offering best-effort service. On other hand, emerging next generation wireless networks are service-driven in nature and require dynamic, robust and flexible quality management, including service perception awareness.

Therefore, service-oriented access network researches play important role in understanding relations between QoS categories: QoS required by user, offered and achieved by operator and finally – QoS perceived by user.

The lack of actual perceived QoS tracking and accounting tools leads to poor SLA enforcement, causing difficulties in clarifying the demands of the user to operator and operator's accountability to the user.

Development of real-time user experienced QoS tracking tools for end-user devices is essential to modern QoS control mechanisms, accounting actual quality of service. Considering next generation networks as heterogeneous, interconnected wireless technologies, both transport and provider independent and service-driven, monitoring and accounting of actual received QoS will be critical.

The Object of Research

The object of presented research is wireless access network performance, ability to control and enforce quality of service and track it at user device.

Objective of the Work

The objective of presented research is to investigate wireless access network performance parameter impact on quality of service and develop reference model of QoS monitor, applied for end-user device.

Tasks

In order to reach the objective, the following tasks have to be solved:

1. To analyze existing measures of QoS evaluation, actual achieved, perceived QoS monitoring and SLA enforcement.
2. To analyze performance properties of contemporary wireless user access network technologies in order to determine their capabilities providing data services and employing soft and hard QoS mechanisms.

3. To analyze network performance characteristics in order to determine factors, influencing quality of service in wireless access networks.
4. To design principles of evaluation of actually achieved quality of service.
5. To develop a reference model of QoS requirement composition and monitoring system.
6. To develop and experimentally test QoS monitor for web browsing service.

Methodology of Research

To investigate the object, following methodologies are applied:

- Statistical analysis and analytical approach.
- Modelling and simulations (Opnet, NCTUns, linux tc, Matlab).
- Active and passive experimental measurements.

Scientific Novelty

Theoretical and experimental investigation resulted in following new achievements:

1. Modelling results of IEEE 802.11 based Wireless Local Area Network (WLAN) capacity evaluation including QoS enforcement overhead analysis; voice channel capacity analysis in half-duplex systems.
2. Modelling results of emergency message delay distributions in multihop topology.
3. Wireless channel models, based on experimental data from wireless network performance analysis.
4. Improved SLA definition and actually achieved QoS tracking model.
5. Reference model for QoS monitor for web browsing service.

Practical Significance of the Results

The achieved results can be used for development of achieved and perceived quality monitoring in next generation wireless access networks. Presented reference model of new SLA enforcement system can be applied for individual QoS management in service-oriented NGNs.

Presented experimental and simulation results may be used for specific service performance evaluation and adaptation, network capacity planning. Presented model of synchronous WLAN allows improving qualitative parameters of voice transmission over wireless local access networks. Simulation results of multihop networking provides base for developing delay-critical automotive applications.

Defended Propositions

1. Quality of service enforcement in wireless access networks requires additional technological resources; even though resource expenditures can be reduced by optimizing access schedulers, additional resources may not be always designated in mobile applications.
2. Control and accounting of quality of service in heterogeneous wireless networks must include passive quality monitoring at user device, creating the feedback between user and operator.
3. Real-time quality monitoring system must rely on difference of actually achieved quality and settled individual quality requirement, but not on absolute quality of service evaluation.

Approbation of the Results

8 scientific articles have been published in reviewed scientific publications by the author within the topic of research:

- 5 in reviewed *ISI Web of Science* journals;
- 3 in other reviewed scientific editions.

The author has made or contributed to 13 presentations at scientific conferences:

- International conference „ELECTRONICS“ in 2006, 2007, 2008, 2009 and 2010, Vilnius.
- Conference “Science – Future of Lithuania” in 2006, 2007, 2009, 2010, Vilnius.
- The Third International Conference on Advances in Mesh Networks MESH’2010, July 18–25, Venice, Italy.
- 14th International Telecommunications Network Strategy and Planning Symposium NETWORKS’2010, September 27–30, Warsaw, Poland.

Structure of the Dissertation

First chapter of the dissertation reviews QoS definitions, categories, existing methods of QoS evaluation and formulates the problem of QoS and service level agreement enforcement measures. In this chapter the guidelines of achieved and perceived QoS monitoring are formulated. The development of QoS monitoring concept requires deep analysis of wireless access network performance factors, service behaviour and their impact to user's perception.

Second chapter presents analysis of various access networks applicable to data and voice services. The analysis is focused on available bit rate, referred to capacity of the access network, capabilities to sustain real time and data services. This chapter presents modelling and experimental results for wireless user access networks.

Third chapter introduces the design of perceived QoS monitoring system, including monitoring agents residing at user device, tools for composing SLA, measures for tracking QoS impairments. The concept of QoS monitoring tool is proposed. The concept relies on defining the link between primary network performance factors and perceived quality for individual services.

The last chapter is dedicated to presenting reference design of individual quality evaluation algorithm and application to web browsing services backed up with experimental results.

Dissertation includes 112 pages of text, 37 equations, 8 tables, 54 figures and 106 references.

1

Methods of Quality of Service Evaluation and Control

Development of multiservice networking promotes high level of integrity and will eventually lead to ubiquitous communication, allowing users to communicate any time, anywhere, in any form, using the best connectivity available. Underlying heterogeneous wireless technologies make this communication possible and are evolving to provide more capacity, robustness and ubiquitous communication service capabilities. The problem of handling quality of services in heterogeneous networks for a long time has been understood as end-to-end control of performance parameters. However, together with migration to packet switched all-IP networking, high mobility, spectral efficiency and soft-QoS techniques, the understanding of quality of service has become more complex and now involves new concepts of service perception.

This chapter presents a review of known quality evaluation and management processes from perspective of actual user experience and analyzes possibilities to evaluate end-user perceived quality of services. The task of this research is to determine the possible ways of linking subjective and objective quality measurements and provide theoretical background for quality related network performance factor analysis.

Part of the analysis presented in this chapter have been published in (Vindašius 2006; Kajackas *et al.* 2010b).

1.1. Quality of Service Problems, Definitions and Categories

Perceived quality of service as a time function $Q_i(t)$ depends on instantaneous quality $q_i(t)$, caused by physical signal $v(t)$, which is dependent on physical communication conditions, as well as available bit rate ρ, packet loss π and latency τ – $q(t; \rho, \pi, \tau)$. $Q_i(t)$ also depends on psychological properties of the user. Let's assume that the signal $v(t)$, instantaneous quality $q(t)$ and perceived quality $Q(t)$ are linked:

$$Q(t) = \Phi[v(t); q(t); \xi], \tag{1.1}$$

where $\Phi[.]$ is the model of QoS perception and evaluation by end–user. The model has undefined factor ξ, which reflects user's psychological side of QoS perception. Factor ξ adds complexity due to hard-to-measure nature of subjective user opinions. $\Phi[.]$ stands for integral operator dependant on many variable factors and generally reflects QoS perceived by user. This operator is not clearly defined.

The initial experience of evaluating perceived quality was obtained from Mean Opinion Score (MOS) expertise. MOS methodology is defined for voice (ITU-T Rec. P.800 1996), video (ITU-T Rec. P.910 1999), multimedia (ITU-T Rec. P.911 1998) services.

The first MOS methodology was created in 1996 and dedicated for evaluation of voice codecs. Following this methodology, a group of volunteer experts are trained to evaluate the overall quality, which usually is being impaired by small defects. The experts form their opinion based on long duration speech segments. This way the influence of multiple small defects is integrated. These MOS based methods can be used for several different purposes including selection of algorithms, ranking of audiovisual system performance and evaluation of the quality level during and audiovisual connection (ANSI S3.2-1989 (R1999) 1999).

Different approach is applied for analysis of intelligibility of transmitted information or instructions (ANSI S3.2-1989 (R1999) 1999). When some segments are lost or packet loss rate exceeds a given threshold, received audio or video become unintelligible. The example of significance of intelligibility can be illustrated by sign language video transmission over cellular networks (Cavender *et al.* 2008).

Previous quality evaluation experiences show, that QoS perceived by end user $Q_i(t)$ is ambiguously linked to instantaneous quality $q_i(t)$ and may differ in wide range.

The answer to question what kind of methodology to choose and whether to focus on overall quality degradation, or emphasize the aspect of intelligibility, depends on the goals of the task. It is offered to differentiate three PQoS concepts herein:

- General Perceived QoS – GPQoS, which is applied evaluating QoS in context of communication, when person receives information of general context (voice chat, video playback, web browsing);
- Special perceived QoS – SPQoS, which is applied if specific instructions are transmitted and decisions are made based on those instructions. It may be applied to machine-to-machine communication, also various systems, dedicated to transmission of instructions;
- Accountable Perceived QoS – APQoS, which should be used to evaluate and account perceived QoS in the period of service delivery.

When analyzing any abovementioned perceived QoS concepts, it is important to set the thresholds for parameter degradation and duration of the negative effects, which are assumed to be noticeable for the user. These effects are characterized as perceptual threshold, detection threshold, noticeable audio, visual and audiovisual errors. The problems of overall acceptance of quality and general research methodology, including research on user tolerable defects is described in (Jumisko-Pyykkö *et al.* 2008).

It was discovered, that user can notice short-term quality degradation, such as individual packet loss during intelligibility tests. Our experiments show, that sometimes user can notice voice defects of extremely short duration, for example, fitting into 20 ms voice segment. Authors in (Pastrana 2004) come to similar conclusion, stating that objective error of 30 ms was audible in all contents. In video service study (Pastrana *et al.* 2004) the minimum length of the error detection threshold was 80 ms, and 200 ms long errors were visible in all contents.

However, when it comes to analysis of general quality variation in longer term, the conclusions may be different. If instantaneous quality changes from "good" to "bad" at some moment in time, the listener may not immediately notice the change. The perceived quality changes more slowly than instantaneous quality, with an approximately exponential curve with a time constant of 5 seconds for the good-to-bad transition and 15 seconds for the bad-to-good transition (ITU Study Group 12 2000). Human ability to integrate overall influence of several defects is not evaluated in most methodologies. For example voice quality MOS evaluation uses 2–3 s duration simple sentences and ITU (ITU-T Rec. P.910 1999) recommends evaluating video quality in segments of 10 s.

ITU defines four QoS related categories (ITU-T Rec. G.1000 2001): Customer's requirements of QoS, QoS Offered by Provider, QoS Achieved by Provider, QoS Perceived by Customer (Fig. 1.1).

Category of perceived QoS is subjective, depending on many subjective components, including user expectations, the importance of the task at the given moment, even the short-term or long-term experience.

Different services have different requirements for operators to achieve and have different behaviour; therefore, the objective network performance factors have complex impact on perception. Some applications may not show any better performance when network performance is improving, e. g. available bit rate, higher than required to stream live video, will not add any value from user perspective and will not increase PQoS.

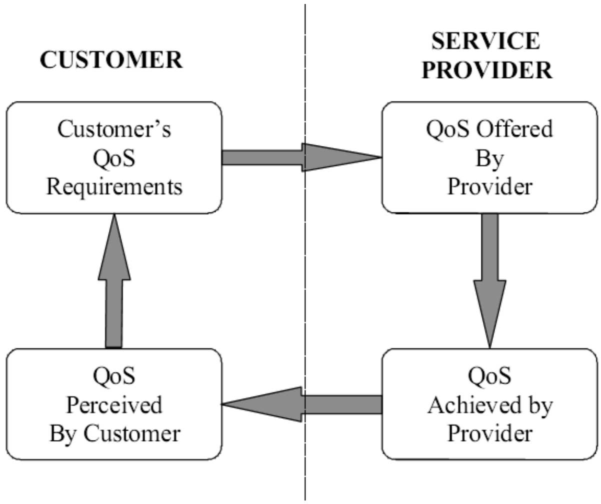

Fig. 1.1. Four QoS categories by ITU

Remaining QoS categories can be objectively measured; however their connection in today's practice of user-operator relationship is not clearly defined. Formulation of Customer's QoS requirements is usually based on network capability limitations, because only these parameters are considered by operator in terms of defining *Offered QoS*. However, from the perspective of user, technical requirements for network performance are less conceivable, since average users service quality understands as in fact *perceived quality*. Therefore requirements are formulated form service perspective, not looking into technical background of how the service was delivered.

ITU defines Perceived QoS or QoS of Experience as *"A statement expressing the level of quality that customers/users believe they have experienced"*

(ITU-T Rec. E.800 2008). This is the basic evaluation of quality from user perspective, and only if user believes, that quality experienced is lower than committed by operator, he seeks for active measurements to objectively prove the technical parameter mismatch.

Much like service–oriented thinking initiated a transformation to service-oriented architecture (SOA) in computing, same trends are seen in NGN concepts. Concept of architectural layering is native to NGN. NGNs clearly separate service and session control from the underlying transport elements. This allows carriers to choose transport elements independently from control software. The separation between access, service, and communications session control within the service layer allows each type of session to be treated independently. This allows the network to be fully service-driven. In this context, the challenges of managing QoS are becoming even stronger.

Most straight-forward examples of offered by operator QoS parameters – maximum, minimum or average bit-rate, tolerated packet loss and availability. In today's competitive telecoms market non-business users usually are offered services with virtually no commitment from operator. In order to offer low prices, attract customers and offer higher peak access network capacities, the operators are using big overbooking factors, service or throughput limitations (e.g. peer-to-peer traffic, Skype, limiting throughput when specific data volume is exceeded, etc.). This way the Offered QoS may differ from actually achieved.

Under these conditions user is unable to relate QoS Achieved by operator and perceived QoS. Furthermore, the formulation of requirements is also not easy task for user, because required QoS and offered QoS are measured and understood differently: user operates with concepts of service or application performance, when operator formulates technical network parameter sets (Table 1.1).

Table 1.1. Differences in QoS parameter understanding

User: Required and Received QoS	Operator: Offered and Achieved QoS
WWW page opening time, s	Max data rate, Mb/s
Voice quality (MOS)	Min sustained data rate, Mb/s
Video quality (MOS)	Max delay, ms
Download time of x byte file, s	Max packet loss, %
	Availability, %

QoS achieved by operator in wireless access networks are influenced by many factors, including network planning, capacity planning, user satisfaction assumptions and finally technological capabilities. Wireless access network technical capabilities are always changing due to physical parameter variations, which can not be mitigated completely in mobile environment. Usage distribution both in space and time has to be considered while planning coverage and capacities. Since network planning usually is efficiency oriented, in peak utilization times (or places) operators allow higher blocking probability or over-utilization. These factors influence the mismatch between Offered and Achieved QoS. The problem of measuring this difference in practice lies in definition of offered QoS.

In practice, offered QoS is strictly defined only in business customer SLAs, applying service guarantees, which result in high price. For the rest of the users, offered QoS parameters are unclear, just giving blurred guidance in terms of average throughput. Wireless access operators usually declare the only performance parameter – maximal achievable throughput (sometimes rather theoretical, than practically achievable).

Table 1.2. Examples of wireless operator declared maximal and minimal bit rates in user contracts

Operator (service plan)	Min. bitrate	Max. bitrate
Omnitel (Omni connect/ prepaid)	16 kbps (*„minimal average ensured service speed is about ...”*)	7.2 Mbps
Bitė Lietuva (Vodafone mobile connect/prepaid)	16 kbps (*„minimal average ensured service speed is about ...”*)	7.2 Mbps
AB LRTC (Mezon/ pre-paid)	16 kbps (*”minimal ensured service speed is …”*)	10 Mbps

By initiative of Lithuanian Communications Regulatory Authority, the internet service providers are committed to provide minimal provided bit rate in service agreements (Communications Regulatory Authority 2005) in order to give objective ground for user-operator arguments. This obligation was accepted quite formally among wireless mobile operators, making huge range between minimal and maximal declared bit rates (Table 1.2) thus solving nothing. Oddly, minimum bit rate is stated by operators as “minimal average”, which in fact is not an obligation even for these stated low bit rates, as long as averaging is not defined.

Even though offered and Achieved QoS (AQoS) are measured in same technical terms, the conclusions for particular service cannot be drawn easily, because user services or applications may not be influenced by some of the factors and harshly influenced by others.

Many network performance testing tools (web-based active measurements, specialized active measurement tools) intend to show differences between parameters, declared by operators and instantaneous performance at particular time and place of interest. However, no associations to services or PQoS are made, thus does not help to form non-technical, but service driven user requirements. This problem has to be solved integrally:

- Required QoS has to be defined in service-driven manner, better yet reflect individual needs of users;
- Required QoS interface to Offered QoS has to be defined by mapping network performance factors;
- AQoS monitoring has to be implemented in user device to monitor achieved quality;
- Only then AQoS interface to PQoS can be implemented, using knowledge about user's expectations and requirements. Making interface between AQoS and PQoS is even more challenging, because it includes mapping between user (perception) and operator (technical) domains.

Ambiguous AQoS definition proves useful neither for users nor operators. Users, guided by such agreements, cannot foresee the actual QoS and do not have grounds to terminate the agreement or complain about low quality network connection. Meanwhile operators do not have means to prove their advantages, leaving no easy objective methods for choosing operator or tracking actual received quality. For this reason many operators offer various mechanisms of "trying" services either for some time without commitment or at sales point.

Analyzing user domain and subjective evaluation of quality, concepts of individual quality (iQoS) emerge (Batkauskas 2006). Measurement and modelling results showed that a GPRS/EDGE access channel of an individual user in mobile data network should be considered as a channel with stochastically varying bandwidth.

The analysis of QoS provisioning is traditionally understood as end-to-end QoS management by technical measures in IP network segments by integrating (IntServ, originally proposed by (Braden *et al.* 1994)) or differentiating (DiffServ, originally introduced by (Blake *et al.* 1998)) the network resources. These methodologies have been analyzed and developed in numerous publications, including Lithuanian authors, focusing on scheduling and queue management problems in IP networks (Budnikas *et al.* 2005; Dekeris *et al.* 2006a; Dekeris *et al.* 2006b; Dekeris *et al.* 2007; Narbutaitė *et al.* 2008) and real-time service applications (Jankūnienė *et al.* 2005). Many objective parameter (jitter or packet loss) researches present traffic analysis without linking it to actual quality (Remeika *et al.* 2007; Ricciato *et al.* 2007; Fiedler *et al.* 2003).

Monitoring “weak” points at heterogeneous network may be a tool for operator to plan and scale network resources using traditional management protocols like SNMP (Jankūnienė *et al.* 2007).

These works cover important part of quality management and assurance in packet delivery networks from perspective of operator and performance factors, but the link to user plane is still poorly defined.

Field of subjective QoS researches are mainly focused to real-time multimedia services (Scaefer *et al.* 2002). Fully opinion based empirical researches were carried out in (Sutinen *et al.* 2005); even though the methodology cannot be applied to measurements in real time, it still defines the trends of basic service perception in various heterogeneous network contexts. Similar works analyze the influence of contextual factors and network parameters on perceived quality (Bouch *et al.* 2000; Ahmed *et al.* 2007; Boucadair *et al.* 2007), however the contributions are either operator or user based. Some integration attempts have been presented in (Galetzka *et al.* 2004; Liberal *et al.* 2005) and finally in ITU framework (ITU-T Rec. E.802 2007), which provides methodologies for QoS criteria identification and conversion to QoS parameters. The recommendation specifies procedures for defining quality objectives and presents analysis of the quality performance measurements. Following this recommendation, (Ibarrola *et al.* 2010) proposes implementation strategy for a practical service provider operation scenario, based on four QoS viewpoints of ITU (ITU-T Rec. G.1000 2001).

Customer satisfaction analysis model purely from marketing perspective was proposed in (Xiao *et al.* 2007) taking the service utility as input computed through a utility model that operates on network performance and other service criteria (service quality, service availability, and customer care). The expectation model updates the user’s future service expectation through a recurrent process, based on past expectations and current perceived utility. This concept also defines customer satisfaction as a product of perceived utility and disconformation. The principle of *service utility* has been adapted from (Anderson *et al.* 1993) and defined as antecedent of customer satisfaction. In this work used expectation and perception relationship was reversed by (Ibarrola *et al.* 2010) considering that user’s perceived QoS may affect the expectation and change the requirements. Therefore, an iterative process is required when deploying the QoS model, as suggested in ITU-T E.802 (ITU-T Rec. E.802 2007) framework.

1.2. Quality of Service Evaluation and Measurements

Perceived quality of service evaluation methods generally can be divided to subjective and objective (Fig. 1.2). Subjective methods are considered to be accurate metric, however very inconvenient, costly and not applicable for dynamic evaluation. Also, averaging opinions may introduce additional uncertainty, rising from subjective understanding and different experience of subjects. Subjective methods are standardized for audio (ITU-T Rec. P.800 1996), video (ITU-T Rec. P.910 1999) and multimedia (ITU-T Rec. P.911 1998) services.

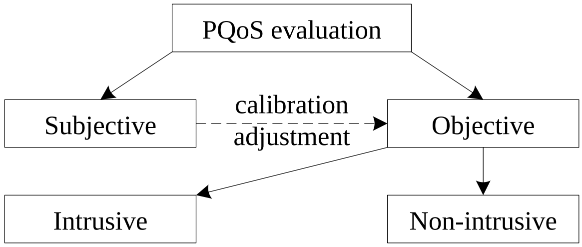

Fig. 1.2. Methods of PQoS evaluation

To make evaluation usable in networks and automate evaluation process, objective methods are attractive as they do not depend on people.

Objective QoS measuring methods have been created using the experience from studies of subjective evaluation of QoS. These algorithms analyze input signal (voice or video) and, like experts in MOS case, evaluate achieved level of QoS. Most of the common specialized QoS measurement tools today are PESQ for voice (ITU-T Rec. P.862 2001) and PEVQ for video (ITU-T Rec. J.247 2008). PESQ algorithm is designed to follow human voice perception mechanism. Algorithm takes into account human abilities to notice short distortions and ability to integrate them and forms an overall score. Input signal is divided to 32 ms chunks – phonemes. Algorithm calculates spectral characteristics and deflections from reference signal (perceptual differences) for every phoneme. The overall evaluation is based on long (8–30 s) segment of distorted and reference signal. PEVQ works in similar way, calculating the perceptual difference comparing aligned signals: all the appropriate indicators are aggregated, forming the final result – the mean opinion score. It is worth mentioning, that attempts to use PESQ for shorter voice segments, creates uncertainty which can be eliminated by special design of input signals (Kajackas *et al.* 2009a; Anskaitis 2009; Guršnys 2008).

Basically PESQ and PEVQ algorithms estimate both short-term noticeable distortions and overall general quality of received signal. The trust level of those estimations is obtained from result comparison with subjective MOS. Both calculated and MOS evaluations strongly correlate. Both PESQ and PEVQ are intrusive methods, relying on sent and received signal comparison.

In context of networking and dynamic evaluation, intrusive methods are not suitable, since reference signal is not present at receiving end (end-user device). It can be seen, that quality evaluation results can vary depending on service, measurement method and measurement tools. Furthermore, special conditions are required for abovementioned measurements. PESQ and PEVQ algorithms are invasive and may give evaluation only when original (reference) signal is available. Therefore, current QoS evaluation methods cannot be applied for service quality monitoring in real-time.

Non-intrusive or passive methods have a great advantage, but pose a challenge. Revising previous attempts to propose non-intrusive algorithms, the methodology can be defined either as parameter based or signal based. The assessment is performed without any reference signal, but applying algorithms to received (distorted) signal. Parameter based methods use network performance and service parameters – like jitter, loss, loss length, coding – as input and try to map them to PQoS. This category is represented by E-model (ITU-T Rec. G.107 2000) and pseudo subjective quality assessments – PSQA (Mohamed *et al.* 2002; Mohamed *et al.* 2004; Rubino *et al.* 2006). The E-Model is an empirical set of formulas originally designed for telephony networks, and even though it is actually used in IP networks, it has been showed (Hall 2001) that it is not good enough for user perceived quality assessment. The PSQA approach uses a statistical learning algorithm (Rubino 2005) to learn the mapping between parameters and user perceived quality. The mechanism shows strong dependence on subjective tests' results for calibration/training.

1.3. Quality of Service Evaluation Guidelines

ITU formulates (ITU-T Rec. G.1000 2001) the challenge of IP-related QoS as lack of proven, robust and scalable standard mechanisms for:

- dynamic allocation of resources;
- assuring that required end-to-end network performance objectives are in fact achieved;
- seamless signalling of desired end-to-end QoS across both network and peer interfaces;
- performance monitoring of IP-based networks and services that are consistent with methods used for network and service planning, and also meaningful to the user experience;
- rapid and complete restoration of IP layer connectivity following severe outages (or attacks) of heavily loaded networks.

Therefore, the important part of QoS enforcement concept is monitoring of QoS-related factors, which are determined by network load and technological properties as well as availability and user perception.

Emerging service-driven heterogeneous wireless access networks with layered architecture of service and transport elements raise a challenge in management and evaluation of quality of service.

To development of QoS management model should start by identifying those QoS criteria that are most relevant to end users. The guidelines of the QoS criteria are provided in ITU-T Recommendation E.802 (ITU-T Rec. E.802 2007):

- QoS criteria and parameters should be considered on a service-by-service basis.
- QoS criteria should be specified on an end-to-end basis (i.e., user terminal to user terminal).
- QoS criteria should be specified in terms understandable to customers.
- Different customer population segments may result in a various range of priorities for the performance parameters.
- The QoS profile of a customer segment may be time-dependent, and the service provider must ascertain the customer's changing requirements.

An iterative process is required when deploying the QoS model, as suggested in ITU-T E.802.

1.4. Trust and Reliance upon Network

Common understanding of QoS concept relies on overall user satisfaction of the service received. Operator providing certain grade of service eventually earns user's trust or distrust. Depending on service type and application, reliance upon service may be considered as important subjective factor. While using security critical services like online banking or shopping, the trust in network security capabilities may be equally or even more important than bit-rate, latency and packet loss rate achieved by access network. Even though objective factors may indicate sufficient QoS level, certain service may be considered unusable.

Many issues of wireless access network security have been reported and corrected over the years (Chandra *et al.* 2009). Users often refer to wireless medium as insecure. This opinion is implied by nature of free-to-access wireless transmission. In recent years, when extended cryptographic algorithms have been applied to wireless transmission, their security level often exceeds the security of wired user access lines. Even more, migration to all-IP wireless network architectures allows using all time-proven IP security measures overlaid on link-layer security protocols.

Separate class of user-managed wireless access still poses a threat (Vindašius 2006). Even though it is originated from misuse or lack of expertise instead of technological problems, it still may have influence on user satisfaction through degraded availability or other network performance factors. User might be unaware of impairments appearing due to insecure WLAN access point.

Wide area wireless networks operate in licensed spectrum bands and employ advanced authentication or encryption algorithms.

Combining some of the general aspects of the user satisfaction models (Xiao *et al.* 2007; ITU-T Rec. E.802 2007; Ibarrola *et al.* 2010), service availability is one of four general QoS criteria: network QoS, service availability, customer care and content.

1.5. Conclusions of Chapter 1

- Even though quality of service management measures are integrated and used in different levels of modern communication systems, the actual perceived quality of service is variable and may dissatisfy the requirements.
- Main causes of perceived QoS variations in mobile wireless networks are user mobility and variable network cell load. Areas with impaired quality usually form due to radio propagation conditions, influenced by buildings, fading, interference, or due to network planning properties and overutilization of network resources.
- In today’s mobile wireless networks, end-user has no measures to determine perceived QoS.
- In reality, the real perceived quality of service is close to impossible to measure applying common methodologies in real-time – during service delivery.
- Quality monitor requires researching access network performance and defining performance factor relation to quality of service.

2

Quality of Service Related Wireless Access Network Performance Analysis

The task of this chapter is to analyze properties of current wireless access networks in terms of QoS management and QoS-related factors. The research covers major types of wireless access technologies – local area access networks (LAN) based on IEEE 802.11 and metropolitan area access networks (MAN) based on IEEE 802.16.

Analysis of QoS related parameters is extended to special case applications – voice transmission over wireless access also wireless multihop transmission scenarios. These applications highlight qualitative factor importance, when strict parameter thresholds have to be met. Together with primary performance factors, access channel capacity for services is evaluated.

Results presented in this chapter have been published by author in (Kajackas *et al.* 2007; Kajackas *et al.* 2009a; Kajackas *et al.* 2009b; Vindašius *et al.* 2010).

2.1. Applying WLAN to Voice Communication

Initial versions of WLAN standard (ANSI/IEEE Std 802.11 1999) were implemented for typical infrastructure applications with light traffic and mobile users. WLAN equipment implements packet transmission and packet switching function. The common method of customer's access to the WLAN is contention-based random multiple-access.

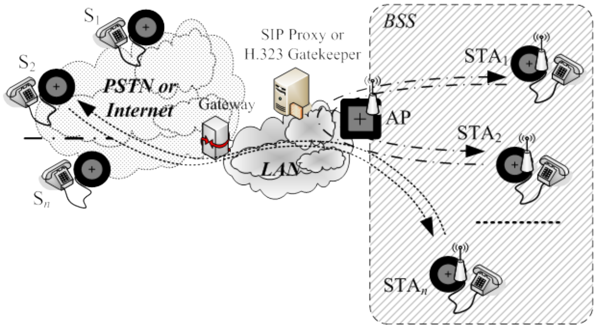

Fig. 2.1. Network with customer access based on IEEE 802.11

To support wireless architectures, the IEEE 802.11 MAC offers two operating modes. Contention mode employs distributed coordination function (DCF) and typically is used for distributed ad-hoc network. The second operation mode is contention free and employs point coordination function (PCF). PCF defines a centralized access for an infrastructure network. PCF support time-bounded services as well as transmission of asynchronous data.

The common scenario of this task is the connection of the wireless customers to external network (Fig. 2.1). The aim of research is to analyze the capabilities to provide voice services in customer access network based on IEEE 802.11. The support of telephony services allows acceptable packet loss which can emerge because of the time limitations between CFP and CP. Analysis is made considering that the PCF is implemented and the beacon rate is synchronized with voice coding rate.

The infrastructure access network consists of AP and nodes (STA's). AP acts as a bridge between the wireless terminals or STA and the wired part of the network. AP is considered to be connected to the wired network via a link that has a higher capacity than the system capacity of the WLAN, all users are fixed (to support high quality physical channel, the antennas may be used).

The main consideration in this research is focused on features of signal transmission of few user conversations at the same time through IEEE 802.11 based customers access WLAN. The offered method is applicable to analyze technological expenditures for RT traffic in IEEE 802.11 infrastructure networks by applying ITU recommendation P.59 (ITU Rec. P.59 1993). The modelling results are based on Opnet Modeller simulations.

There are many research works where wireless or radio link access from different aspects are analyzed – WLAN protocols and its features, voice and data transmissions, features of usage in the specific conditions. For WLAN protocol modernization or perfection many task solutions are proposed. A number of previous studies have evaluated the capacity in IEEE 802.11 networks for voice traffic and real time traffic in general. Many works (Bianchi 2000; Medepalli *et al.* 2004; Banchs *et al.* 2001) studied the use of DCF to support VoIP, to determine delays and jitter margins. It is shown, that the capacity to accommodate voice traffic in DCF is very limited. DCF is not efficient in supporting the delay-sensitive voice traffic. The contention-based nature and exponential backoff mechanism can not guarantee that voice packet is successfully delivered within the delay bound.

Controlled access is more suitable for voice traffic delivery, because of its smaller overhead and guaranteed performance. The capacity of a system that uses the PCF for CBR and VBR voice traffic was analyzed in (Chen *et al.* 2002) and (Veeraraghavan *et al.* 2001). The VBR voice traffic was simulated using ON-OFF voice source model.

Specific 11e related papers (Ferre *et al.* 2004) focuses on a throughput analysis and delay of the DCF/PCF of the IEEE 802.11 and 802.11e MAC. The results are data transmission oriented and presented for different packet lengths and different number of users. Throughput performances are also detailed for the EDCF.

2.1.1. Voice Source Models

The simplest voice model considers independent voice sources, which can be constant bit rate (CBR) and variable bit rate (VBR). Other way of designing voice models is composition of conversational speech by imitating dialogs.

The CBR voice source transmits packets at the constant rate κ. The size of the packets is determined by the bit rate of the codec R_{cod} and the packetization period τ_{pac}.

The voice source alternates between talk spurts and silent periods. Function of Voice Activity Detection (VAD) is implemented in modern voice codecs, generating VBR traffic. VBR voice traffic can be represented by an ON-OFF model. At ON state, whereas voice packets are generated periodically at the constant rate κ, while no voice packets are generated at OFF state. The durations of the ON and OFF states typically are modelled as independently and exponentially distributed random variables, with duration parameters τ_{ON} and τ_{OFF}, respectively. Each voice source is modelled by a two-state discrete-time Markov chain. At time instant t, a voice source has voice packet to send (source is at ON state) with probability: $p_{ON}= \tau_{ON}/(\tau_{ON}+ \tau_{OFF})$, and has no voice packets (source is at OFF state) with probability $p_{OFF}=1- p_{ON}$. VBR voice source model is suitable for analysis of one direction transmissions.

To imitate the conversation of two speakers applying this ON-OFF model, the assumption has to be made that the users speak independently. Realistically, the assumption would be incorrect since speakers talk alternately during the conversation. And only on rare particular occasions they either both talk or both listen.

For WLAN applications conversational model when two users (A and B) are talking dependently is more applicable.

Summarized model of the conversation between two users A and B like superposition between state of the conversation and the state of speech activity is composing and may be defined by four state Markov chain. In the conversation model possible states are: A talking B silent, A silent B talking, both talking, both silent (ITU Rec. P.59 1993). The probabilities of these states are indexed by p_{A0}, p_{0B}, p_{AB}, p_{00} respectively.

According to ITU recommendation P.59, the durations of states are identically distributed exponential random variables with means 854 ms, 854 ms, 226 ms and 456 ms respectively. According to these durations calculated probabilities are: $p_{A0} = 0.357$, $p_{0B} = 0.357$, $p_{AB} = 0.095$, $p_{00}=0.191$.

The channel occupation times are evaluated incorrectly, if two-state ON-OFF speech model is used while analyzing IEEE 802.11 networks, where transmission in both directions take place in same physical channel. Two-state model fails to evaluate precisely time periods of mutual talk or mutual silence. In this study conversational model is used while analyzing WLAN characteristics.

Voice coding and transmission processing time diagram is shown in Fig. 2.2. At the sender, the analogue voice signal is encoded by a codec which determine voice frame duration T_F of coding. RTP/UDP/IP headers are appended to coded voice packet, and then voice packets are transmitted over the network. Packetization procedure ends at the moment t_{pac}.

Voice packets with period T_{pac} are transmitted periodically during each CFP period. T_{pac} may be equal to T_F or multiple T_F, in case WLAN data packet carries several voice frames.

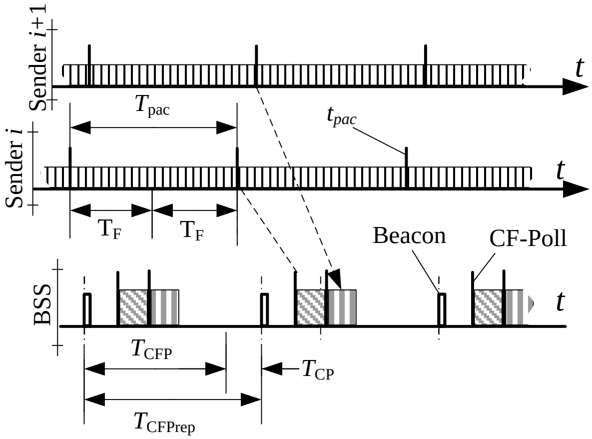

Fig. 2.2. Diagram of voice coding and transmission processing

Voice frame duration T_{PAC} and T_{CFPrep} are independent. T_{pac} is conditioned by voice codec (for GSM and AMR codecs T_{pac} = 20 ms), while T_{CFPrep} is adjustable WLAN parameter. To overcome voice packet loss, we suggest performing transmission synchronically, when

$$T_{pac} = T_{CFPrep} \tag{2.1}$$

and in case of carrying multiple voice frames:

$$a \cdot T_F = T_{CFPrep}\,. \tag{2.2}$$

In analysis and simulations presented herein, the condition (2.1) is assumed to be met.

The T_{CFPrep} is controlled by AP, which defines the timing for the entire BSS by transmitting beacons (Fig. 2.2) according to the T_{CFPrep} attribute within the timing synchronization function (TSF) (ANSI/IEEE Std 802.11 1999) of AP. TSF is a local AP timer which synchronizes the TSF of every other station in the BSS. TSF defines a series of *TBTT*s exactly $T_{CFPRate}$ time units apart. At each *TBTT*, the AP shall schedule a beacon as the next frame for transmission.

While transmitting voice packets in PCF, the queuing delay is inevitable, because the transmission to particular destined STA takes place strictly according to the polling list.

2.1.2. Modelling Voice over WLAN

Customer access modelling was based on voice source synchronization with beacon interval in PCF mode, according to (Kajackas *et al.* 2006).

The modelling of voice applied customer access was implemented in two stages: the modelling of conversational speech and modelling of IEEE 802.11 hotspot and wireless PCF clients.

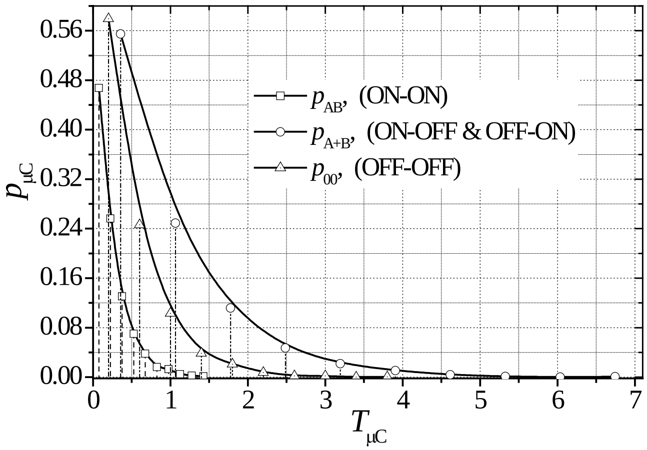

Fig. 2.3. The distribution of voice model state durations $T_{\mu C}$

The first stage includes the modelling of conversational speech, which follows four state ON-OFF model as described in (ITU Rec. P.59 1993), having transitions from double talk to mutual silence trough single talk states of one or another speaker. Model was implemented in *Matlab*. The outcome of first stage simulation expresses the timeline of speech activity in forward and backward directions. Unlike two-state ON-OFF model, the two speech directions are dependent and connected in a way of conversational dialogue. Thus the algorithms implemented as *Matlab* scripts produce voice source in pairs – one for STA and other for corresponding user on the wired part of the network. The distributions of model state durations are shown in Fig. 2.3.

The generated conversational speech flows are stored as external data in form of talk-silence cycle. Although all speech pairs were created by using the same parameters, they are generated separately and are statistically independent.

The second modelling stage includes models and simulations in Opnet Modeler (Opnet Technologies 2007). The network setup (Fig. 2.1) was implemented using standard Ethernet station, AP and wireless IEEE 802.11 station models. User profile configuration includes voice over IP calls, which are initiated at the beginning of simulation and continues to the end.

Additionally few editions were made to the original Opnet STA IEEE 802.11 MAC layer:

- In CFP operation mode the STA could be polled more than once in the same CFP period. After AP have finished to poll all STAs listed in its polling list, but still some CFP time remains and there is some data to send, the AP starts the polling over again resulting some STAs to be polled more than once. This kind of operation had to be eliminated in order to maintain the precision of voice packet delays and channel capacity. Modified model ends contention free period prematurely rather than starts polling once more.
- All STAs could send data in CP even if it is PCF enabled STA. Original model was modified to prevent DCF transmissions in PCF enabled stations in order no to lose any voice packets, which would not be included in statistics while calculating CFP durations. Despite of modifications, STA registration related traffic is allowed during DCF to maintain regular registration routine.
- All upper layer traffic in MAC layer were handled equally – stored in a single buffer. Separation was done only in terms of coordination function type – separate buffers for PCF and DCF traffic, but no differentiation within PCF buffer. This makes the voice packets to be collected and stored in buffer until sent together with poll (in case of AP) or until poll is received and data can be sent with an acknowledgement to the poll (if this is a STA). In case of overload in the PHY, when more STAs are registered than can be served by AP, the packets accumulate in the buffer, thus effecting delays or even buffer overflows. Such principle is not suitable for delay sensitive applications. It is acceptable rather to drop packet than to store it for unlimited amount of time (usually it is limited only by size of buffer). Packet drop would affect voice quality, but only for the particular flow, while storing and delaying packet may negatively affect other voice flows introducing unacceptable delay and jitter. Furthermore, if the delayed packet is not usable, but still transmitted, the bandwidth used for transmission is wasted. The model was modified in such way, that it would drop the data packet if it failed to be transmitted due to overload, i.e. it was not transmitted because previous STAs used up all available CFP time. In other words, all data frames older than CFP repetition time are dropped.

Opnet simulations were executed using prepared conversational speech models to imitate talking users. G.711 was used as voice codec with voice activity detector (VAD) activated in order to produce VBR traffic: voice flow bandwidth – 64 kbps; packet rate – 50 pps. Voice payload frame size is forced to be equal to CFP repetition time: $T_{pac} = T_{CFPrep} = 20$ ms. The voice payload of 20 ms produces 160 byte voice frames. This may be considered as multiple voice frame packing in WLAN packet with multiplier $a = 2$, since in this case two codec sample intervals of 10 ms form the 160 byte payload. The IP/UDP/RTP header of 40 bytes size is added to voice frame before passing to WLAN layer.

Since *OPNET* is discrete time simulator, it does not require any additional effort to keep the voice sample length and CFP repetition time exactly the same. Thus, no additional packetization time fluctuations over CFP repetition time can possibly occur. This means that we always precisely keep the equality (2.1) during the simulations.

The first set of simulations was intended to show the wireless channel occupation time while different number of STAs operate in the PCF controlled network. The simulations were performed for IEEE 802.11b physical data rates 5.5 and 11 Mbps, whereas all other parameters including CFP repetition and voice packetization times are constant.

In order to simplify the modelling, the simulations were combined with some calculations. The simulations were performed for STA numbers up to 10, next the distributions of channel occupation time were obtained from simulation results. The distributions for STA counts over 10 were calculated using convolution of already obtained distributions.

It was done for the following reasons. Firstly the simulation of great number of nodes is a time and processing power consuming task, therefore it is more efficient to perform longer simulations to obtain higher precision, than to run short simulations of high quantity of nodes. Secondly, it is not possible to show visually the distributions which extend over CFP occupation time only by simulation. When actual required CFP occupation time exceeds maximum allowed CFP duration, this indicates the network overload, therefore probability distributions are distorted due to packet loss. This kind of solution does not bring any bias to the results whatsoever.

Let the $p_l(t_{oc})$ and $p_j(t_{oc})$ be the cannel occupation time t_{oc} distribution functions, when STA count is l and j respectively.

While applying linear system model, the channel occupation time is not limited. The distributions of channel occupation time are calculated using rules of adding casual variables for the STA count $n=l+j$. Then the convolution can be written as:

$$p_n(t_{oc\Sigma}) = \sum_{t_{oc}} p_l(t_{oc})\, p_j(t_{oc\Sigma} - t_{oc}). \tag{2.3}$$

To keep the maximum possible accuracy, the distributions were produced using precise t_{oc} values, obtained from simulations. Since those values differ in constant steps, the discrete function convolution has been applied directly, without re-sampling frequency counts consequently avoiding any bias. It is easy to notice that the step size is equal to difference between data and control packet (empty packet) sizes.

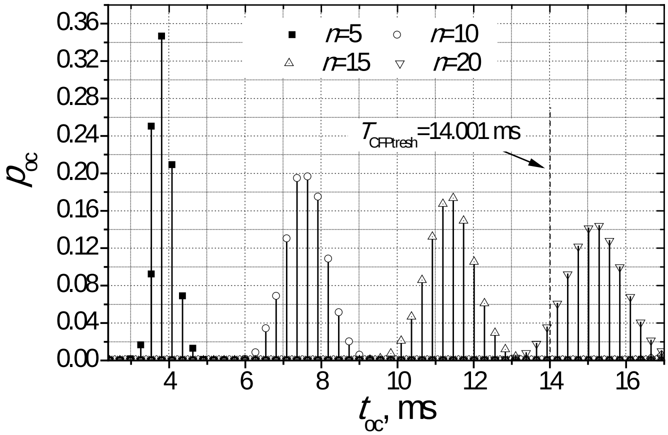

Fig. 2.4. Distributions of CFP occupation time t_{oc} rate for 5.5 Mbps PHY rate of IEEE 802.11b

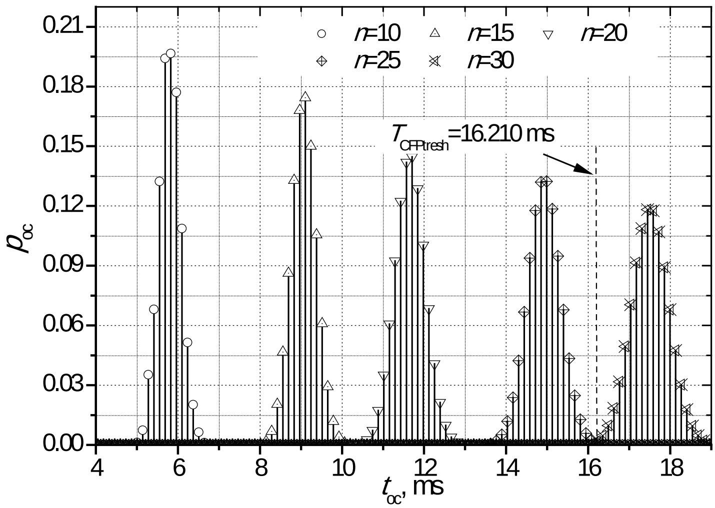

Fig. 2.5. Distributions of CFP occupation time t_{oc} rate for 11 Mbps PHY rate of IEEE 802.11b

Distributions of CFP occupation time rate are shown in Fig. 2.4 and Fig. 2.5. The graphs show distributions of different STA number under different physical data rate.

Standard (ANSI/IEEE Std 802.11 1999) defines minimal allowed CP duration T_{CP}:

$$T_{\mathrm{CPmin}} = T_{\mathrm{PPDUmax}} + 2\tau_{\mathrm{SIFS}} + 2\tau_{\mathrm{TimeSlot}} + 8\tau_{\mathrm{ACK}}. \tag{2.4}$$

At least that much transmission time has to be dedicated for CP in T_{CFPrep} cycle. When the CFP repetition time duration T_{CFPrep} is chosen, the allowed CFP duration has to consider the deduction of minimal CP time and can be marked as threshold T_{CFPtresh}:

$$T_{\mathrm{CFPtresh}} = T_{\mathrm{CFPrep}} - T_{\mathrm{CPmin}}. \tag{2.5}$$

According to (ANSI/IEEE Std 802.11 1999) and (Kajackas *et al.* 2007) when frame of CFP micro-cycle equal to User Data payload (all in ON-ON state) is transmitted, the maximum duration of CFP may be expressed:

$$T_{\mathrm{CFPmax}} = nT_{\mathrm{AB}} + \tau_{\mathrm{Beacon}} + \tau_{\mathrm{CF\text{-}End}}, \tag{2.6}$$

$$T_{\mathrm{AB}} = 2(\tau_{\mathrm{PPDU0X}} + \tau_{\mathrm{SIFS}}), \tag{2.7}$$

where n – count of micro-cycles in the CFP (the n is equal to count of pollable STAs in WLAN); τ_{Beacon} – beacon frame duration; $\tau_{\mathrm{CF\text{-}End}}$ – CF-End subtype of Control type frame duration (Kajackas *et al.* 2007a).

The voice transmissions in analyzed network model are lossless until the following condition is met:

$$T_{\mathrm{CFPmax}} \le T_{\mathrm{CFPtresh}}. \tag{2.8}$$

From (2.8), (2.5) and (2.7) follows the expression of capacity for transmission system without losses:

$$n_0 \le \frac{T_{\mathrm{CFPrep}} - T_{\mathrm{CPmin}} - \tau_{\mathrm{Beacon}} - \tau_{\mathrm{CF\text{-}End}}}{T_{\mathrm{AB}}}. \tag{2.9}$$

Consequently, if the T_{CFPrep} = 20 ms, the maximum allowed CFP duration or CFP threshold T_{CFPtresh} for IEEE 802.11b 5.5 Mbps, and 11 Mbps is 14.401 ms and 16.210 ms respectively.

Using these parameters we get n_0 12 and 21 channels for 5.5 Mbps and 11 Mbps PHY throughput respectively.

The system capacity n_{c} can be revised and increased by declaring the allowed packet loss probability. This task is solved as follows. The simulation results are processed to obtain channel occupation duration distributions, which are complemented by calculated (2.3) distributions, when $n>n_0$. The results are presented in Fig. 2.6 and Fig. 2.7.

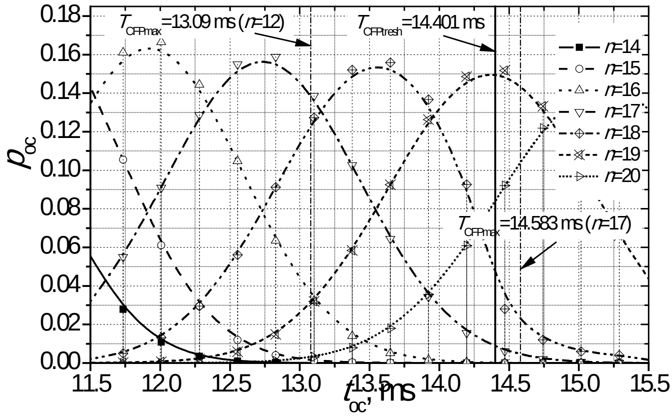

Fig. 2.6. Distributions of CFP occupation time t_{oc} rate in range of maximal CFP duration for 5.5 Mbps PHY rate

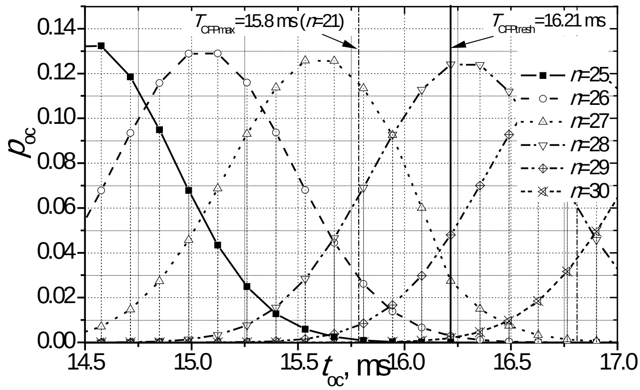

Fig. 2.7. Distributions of CFP occupation time t_{oc} rates in range of maximal CFP duration for 11 Mbps PHY rate

The capacity evaluation from the provided graphs may be observed flexibly in the means of allowed packet loss probability. If distribution of channel occupation duration does not fit into timeframe up to $T_{CFPtresh}$, it leaves some distribution bars behind. This means that under certain load (number of STAs) the last voice streams will experience packet loss, which probability is equal to sum of unfitted values.

From Fig. 2.6 it can be seen, that up to 16 STAs distributions fit perfectly within T_{CFPtresh} limits, thus no packet loss is ever introduced in the particular scenario. It can also be seen that 17 STA distribution leaves some values over T_{CFPtresh}, but the sum of it is very small – the calculations show, that the loss probability for 17 STA stays below 0.8 %, which is certainly acceptable as will not have noticeable effect on voice quality. In case of 18 STAs, the T_{CFPtresh} cuts the bigger part of distribution, thus the packet loss is expected to be greater. The calculation shows almost 4.5 % packet loss probability. So channel number is extended from 12 to 17 or 18 depending what packet loss probability we are willing to accept.

As for 11 Mbps PHY rate (Fig. 2.7), the distributions up to 25 STA show no packet loss, distribution of 26 STA shows 0.4 % packet loss probability, and 5.4 % in case of 27 STA. Thus in case of 11 Mbps PHY, we gain 5 channels extending from 21 to 26.

Thus on the particular situation in both PHY rate cases, the channel capacity is increased by 5 channels keeping reasonable packet loss probability (less than 1 %). The greater capacity allows serving more users or using greater oversubscription. The acceptable packet loss ratio can be chosen to allow more simultaneous voice streams instead of allowing less voice channels with no packet loss, which is not efficient way of predicting the capacity, since the probabilities of full (or near to it) channel usage are small.

Polling list is not reallocated during communication process (during voice call in our case), therefore, it is reasonable, that only the latest registered STAs will suffer from packet discards when CFP time is used up.

The analysis accomplished herein shows that it’s possible to create customer access with quality, comparable to DSL, when the IEEE 802.11 protocol equipment with PCF is used. This can be achieved by synchronization of beacon interval and voice coding cycles. Even though modelling is performed with IEEE 802.11b standard wireless network elements, it can be applicable to 11a/g/n standard extensions capable of supporting PCF.

The offered method will allow increasing capacity of the IEEE 802.11 PHY channel while keeping reasonable packet loss probability. The results of IEEE 802.11b customer access simulations for 5.5 and 11 Mbps PHY rates show that implementation of offered method improve channel capacity, thus allowing to serve more users or apply oversubscription. Note that defined capacity, reflecting the marginal abilities of wireless customer access network, is greater than capacity evaluations in DCF mode presented in other publications.

The synchronization of CFP repetition interval and period of voice codec in real IEEE 802.11 equipment has been and still remains problematic topic. When these periods are not synchronized, the voice coding window is sliding time-wise during transmission. The sliding effect can create many noteworthy scenarios; this problem is worth to be analyzed in the future works.

2.2. Quality of Service Management in User Access WLANs

Growing demand for multimedia services intend to capture major part of traffic in consumer premise networks. Wireless local area networks (WLANs) are no exception. The control over service parameters in wireless consumer access is critical, yet challenging. Transmitting miscellaneous traffic consisting of bandwidth consuming non-real-time data and time sensitive multimedia services such as voice and video requires traffic differentiation and QoS handling. However this task requires a lot of effort to gain control over shared radio channel, efficiently schedule and manage limited channel resources.

In this chapter we analyze system capacity of WLAN employing QoS enhancements. Results are compared to legacy PCF analysis.

2.2.1. QoS in Legacy IEEE 802.11 and IEEE 802.11e

Initially the wireless customer access technology, based on IEEE 802.11, was designed for bursty best-effort traffic. The basic access function of the IEEE 802.11 standard, called Distributed Coordination Function (DCF), handles all customer wireless stations (STAs) independently, organizing radio channel access in a manner of contention. The principle itself suits well for low-bandwidth bursty traffic, but still brings service degradation issues, related to fairness, unpredictable delay and jitter, high collision rate, altogether leading to inefficient channel utilization.

The first attempt to adapt IEEE 802.11 based wireless access networks to delay sensitive service flows, was Point Coordination Function (PCF), which is part of the legacy IEEE 802.11 standard (ANSI/IEEE Std 802.11 1999). Basic idea of the PCF is to schedule STA transmissions in round-robin manner, issuing polls to PCF enabled STAs and following the polling list, which is made upon STA registration to access point (AP). This scheme allows introducing some level of QoS; nevertheless it lacks flexibility and versatility.

Due to low concern on QoS management at the initial stage of IEEE 802.11 technology commercialization PCF availability in commercial products of IEEE 802.11a/b/g equipment is very limited. Many vendors have chosen not to implement PCF also due to possible compatibility issues. To date, IEEE 802.11a/b/g based WLANs are considered to be only best effort access technology for basic data transmissions, unable to differentiate and provide quality enabled services.

An effort to make the difference in understanding unlicensed wireless access quality management was made by IEEE 802.11 Task Group E (TGe), which resulted in standard amendment IEEE 802.11e (IEEE Std 802.11e-2005 2005).

CSMA in legacy IEEE 802.11 DCF appeared to be efficient for bursty traffic. Whereas PCF suits well for predictable, but not bursty transmissions. Hybrid coordination function (HCF) in IEEE 802.11e combines both techniques making transmissions of bursty and constant bit rate traffic efficient.

Amendment features include enhanced DCF and PCF MAC mechanisms namely Enhanced Distributed Channel Access (EDCA) and HCF Controlled Channel Access (HCCA). Both of them are controlled by Hybrid Coordination Function (HCF) which resides in QoS enabled AP.

EDCA should be understood rather as relative than guaranteed QoS management, because it cannot provide strictly defined QoS parameters. The priorities in EDCA are handled in stochastic manner, providing different channel access probabilities for different packet queues. In legacy DCF, the probability to access channel is the same for all STAs, because the channel sensing and capturing mechanism provides equal rights for all contending stations. All STAs would have to sense the channel as idle for DIFS period to start transmission and perform backoff otherwise to avoid collision. The backoff time is picked randomly from interval known as Contention Window (CW) and is equal for all network nodes as well (Fig. 2.8).

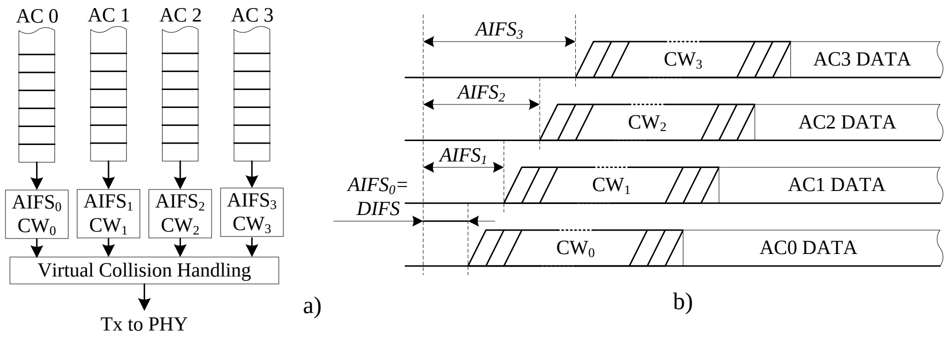

Fig. 2.8. EDCA AC handling (a) and prioritizing with AIFS (b)

In EDCA different Traffic Stream (TS) priorities are represented by different channel access timings called Arbitration Interframe Spacing (AIFS). AIFS, CW minimum and maximum values are different for each of four access categories

(AC) or packet queues (Table 2.1). In order to achieve higher channel utilization efficiency, collision handling is slightly different from DCF: if virtual collisions between different queues appear, only packet with higher priority is sent out to PHY. In case of PHY collisions, CW is not doubled as it would be in DCF, but rather incremented by fixed number called Persistent Factor (PF). Although IEEE 802.11e defines 8 separate priorities for service flows, only 4 ACs are used – the priority mapping to AC is also presented in Table 2.1.

Other important improvement in both contention and contention-free based transmission is mechanism allowing transmitting frame bursts. As frames in a burst are separated only by SIFS and do not require to contend for each packet in the burst, significant amount of channel bandwidth is saved thus improving channel utilization efficiency. For this matter, IEEE 802.11e amendment defines new type of timer – Transmission Opportunity (TXOP).

Table 2.1. IEEE 802.11e EDCA parameters (IEEE Std 802.11e-2005 2005)

Priority	AC	AIFS	CW_{min}	CW_{max}	Service
0,1,2	0	2	31	1023	BE
3	1	1	31	1023	Video
4, 5	2	1	31	63	Video
6, 7	3	1	7	15	Voice

TXOP value defines the time limit, which can be used to transmit the burst containing any number of packets as long as they fit into TXOP (including SIFSs and ACKs) and belong to the same AC (Fig. 2.9). In EDCA TXOP values are reported to STA trough beacon frames. In HCCA every poll packet from AP contains individual TXOP values, which are calculated by HC considering required QoS parameters and making sure that certain TXOP allocation will not unacceptably increase delay for other flows. The calculation algorithm is scheduler dependant.

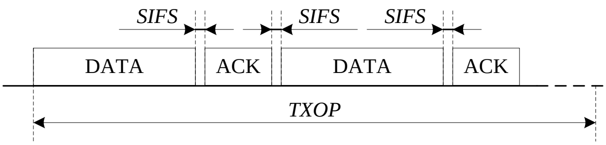

Fig. 2.9. Frame burst using TXOP

IEEE 802.11e EDCF researches (Choi *et al.* 2003) show that contention free burst usage in TXOP improves the global system performance at the cost of delay increase for certain traffic types. Detailed simulations (Dajiang *et al.* 2003) show quite efficient prioritizing in EDCA, however under high network load conditions low priority flows may suffer bandwidth starvation or complete blocking. In (Jeonggyun *et al.* 2006) EDCA is considered as not efficient for real time applications as it cannot guarantee QoS even for high priority flows – under high network load collision probability is high for all service flows not looking at their priority, which result high random delays.

Voice capacity analysis of WLANs with channel access prioritizing mechanisms is presented in (Ling *et al.* 2008). The analysis is based on EDCA and focuses on effect of CW. It can be seen, that in contention mode, the delays cannot be fully controlled.

2.2.2. Parameterized QoS with HCCA

As specified in IEEE 802.11e, contention based access EDCA is combined with contention free access, just like in legacy IEEE 802.11 where alternating contention and contention-free periods were controlled by PCF and DCF. The control of this process is dedicated to HCF, which combines EDCA and HCCA modes.

In HCF mode each superframe consists of alternating contention-free and contention periods and starts with a beacon frame, just like PCF superframe in legacy IEEE 802.11. In contention-free period AP polls STAs for data similarly as in PCF. QoS enabled STAs receive CF-poll frames containing TXOP allocation, which indicates burst length for the STA data to be sent to AP.

Contention period follows right after CFP and allows STAs to contend for the channel access. This period employs EDCA mechanism, handling virtual collisions and providing priorities for 8 different packet queues. During CP the AP may capture the channel if the need to transmit CF data arises. Waiting only for PIFS it gains priority over all contending flows and grabs the channel to send CF-poll to scheduled STA. This routine when AP polls the STA during CP is called Controlled Access Period (CAP). There can be as many CAPs as AP decides to be required to meet the QoS parameters of registered TSs. Thus the main difference form legacy IEEE 802.11 is superfame structure: CP and CFP are not strictly separated, but may alternate several times in one superframe (Fig. 2.10).

To meet required QoS parameters, AP requires to employ scheduler, which could suit for different traffic, bursty and constant flows, ensure fairness, collision free transmissions and also efficiently use spectrum. Standard proposes scheduler which is designated as reference, meaning that scheduler scheme is implementation dependent and equipment vendors may develop their own schedulers using reference as an example.

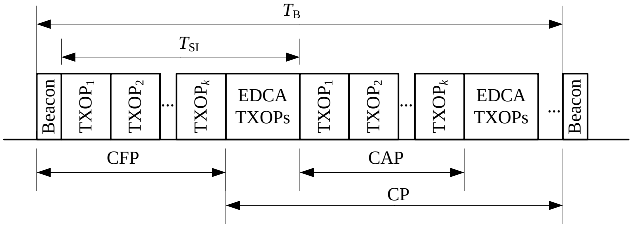

Fig. 2.10. CAP allocations in IEEE 802.11e superframe

There has been some research works on scheduler performance and scheduler proposals. Improvement proposals aim both for EDCA and HCCA, considering adoption of various traffic streams. Since EDCA seems to be easy to implement and appears to be more common access function like DCF, significant part of performance evaluations focus on EDCA operation mode only.

Contention based manner of accessing the channel in EDCA mode still results very limited QoS implementation. This drawback is also widely seen and analyzed while proposing scheduling schemes for HCF. Very few schedulers can be found for handling delay sensitive multimedia traffic. One of the attempts was J. Roy's *et al.* proposal on uplink scheduler for multimedia applications, capable of attaining the QoS requirement (Roy *et al.* 2007). The performance shows improvement in delay, throughput and channel utilization.

The main functions of reference scheduler include TS registration mechanism, resource admission and maintaining contention-free transmit operation in uplink and downlink.

Every STA registration procedure means registering every TS of which QoS parameters has to be considered. While registering, HC has to evaluate if the new stream may be admitted. It is required to check if it is possible to fulfil all QoS requirements the TS has demanded and if the admission of new TS will not interfere with QoS requirements of already registered TSs. Every TS is defined with the set of QoS parameters, namely Traffic Specification (TSPEC). Reference scheduler uses only mandatory TSPEC components: Mean Data Rate, Nominal MSDU Size, and Maximum Service Interval or Delay Bound.

The Service Interval (SI) has to be calculated at first. SI represents the maximum time interval between polling of specific TS. Reference scheduler uses the same SI for all admitted TSs, which is equal to minimal T_{SImax} value of all TSs (2.10). The SI is recalculated every time new TS joins the network – in case T_{SImax} of new stream is lower than current T_{SImax}, HC updates SI value to lower, otherwise SI remains unchanged. Also SI is adjusted to first lower submultiple of TBTT.

$$T_{\mathrm{SImax}} \le \min(T_{\mathrm{SImax}\,i}), \tag{2.10}$$

where i is the i-th QoS enabled STA.

Next, the TXOP value is calculated for the new stream. Unlike SI, the TXOP is unique for every STA. TXOP is calculated only from TSPEC parameters of respective TS:

$$T_{\mathrm{TXOP}\,i} = \max\left(\frac{N_i \cdot L_i}{R_i} + O, \frac{M}{R_i} + O\right), \tag{2.11}$$

where N_i is the count of MSDUs of i-th STA which fit into SI duration if transmitted at mean data rate (ρ); L_i – nominal MSDU size of i-th STA; R_i – physical transmission rate of the i-th STA; M – maximum allowed MSDU size, equal to 2324 bytes (IEEE Std 802.11e-2005 2005); O – overhead expressed in time units.

Overhead includes all control packets and interframe spacings required to deliver the frame. In reference scheduler the overhead includes polling frames, PLCP and MAC overhead for data frame, ACK frames and interframe spacings:

$$O = t_{\mathrm{PIFS}} + t_{\mathrm{poll}} + t_{\mathrm{SIFS}} + t_{\mathrm{DATA\,MAC}} + t_{\mathrm{SIFS}} + t_{\mathrm{ACK}}. \tag{2.12}$$

N_{i} can be calculated as ceiling of the number of MSDUs that arrived at ρ during the SI:

$$N_i = \left\lceil \frac{T_{\mathrm{SImax}} \cdot \rho_i}{L_i} \right\rceil. \tag{2.13}$$

The rounding up to the first integer is necessary to make sure that whole packet with overhead will fit into TXOP. As mentioned, the frame will not be allowed to access the medium if the time required sending it out and receiving ACK exceeds TXOP limits.

When SI and TXOP are calculated, HC checks whether TS should be allowed to register. This is done by evaluating if the new TS together with already admitted TSs will not extend beyond TBTT. The TS is admitted if the following inequality proves to be correct:

$$\frac{T_{\mathrm{TXOP}\,k+1}}{T_{\mathrm{SI}}} + \sum_{i=1}^{k} \frac{T_{\mathrm{TXOP}\,i}}{T_{\mathrm{SI}}} \le \frac{T_{\mathrm{B}} - T_{CP}}{T_{\mathrm{B}}}, \tag{2.14}$$

where $T_{\mathrm{TXOP}\,k+1}$ is the TXOP value of new TS; $T_{\mathrm{TXOP}i}$ – TXOP value, of i-th admitted TS; T_{B} – beacon interval; T_{CP} – time used for EDCA traffic; k is the number of already admitted TSs.

2.2.3. HCCA Capacity Estimation

Estimating channel utilization effectiveness for various traffic, usually breaks down to following types: real-time and non real-time, constant bit rate (CBR) and variable bit rate (VBR). Since this analysis is focused on handling real-time traffic applications, VBR voice conversation have been selected. Number of simultaneous VoIP calls, supported by the single wireless channel, can be a good measure of channel utilization effectiveness when using VBR delay sensitive service.

TDM-like polling schemes usually show great performance on CBR traffic, as cyclic polling ensures low latency and do not introduce significant overhead. However, VBR is more challenging, since handling bursty traffic requires adaptive scheduling in order to minimize the overhead, introduced by polling-acknowledging cycles without carrying any data.

The expected VoIP channel capacity is tightly bonded with the overhead introduced by the scheduler. The main differences in overhead comparing PCF and HCCA are the following:

The frame structure, allowing CAPs, is different from legacy PCF, thus latency requirements can be fulfilled even if inter-packet spacing needs to be smaller than beacon period.

TXOP bursting allows transmitting several packets separated by SIFS (Fig. 2.9), thus increasing channel utilization and reducing overhead.

Block acknowledgement scheme, introduced in IEEE 802.11e allows transmitting several packets and rather acknowledging them in blocks than each one separately.

TXOP bursting and block-ACK obviously are not helping to increase channel utilization efficiency in our case, since it is required to avoid packet grouping in real-time traffic. Thus the overhead would basically depend on frame sequences, employed by scheduler.

IEEE 802.11e amendment adds many more frame sequences to ones defined in legacy PCF and DCF routines. Those sequences include various mechanisms for polling, acknowledging, bursting, piggybacking ACKs and data to poll frames and so on. Building frame sequence in legacy PCF is straight-forward – basically only very few possible sequences are available: CF-poll, CF-poll+Data or CF-poll+CF-ACK+Data frame from AP side and CF-ACK or CF-ACK+Data frame from STA side in response to the poll. IEEE 802.11e amendment introduces much more complexity to frame sequences to achieve flexibility and higher efficiency.

General HCCA reference scheduler routine starts with HC issued CF-poll. CF-Poll contains a TXOP limit for polled STA in its QoS Control field. The CF-poll containing frame is not allowed to carry data unless aggregation subfield in the associated TSPEC is set to 1, meaning that aggregation of separate TSs is allowed.

This is reasonable, since HC polls are issued to STAs, not TSs. When separate TSs requires to be scheduled separately (no aggregation), the data from STA is carried on separate CF-ACK+QoS Data frames in TXOP frame sequence and acknowledged by AP with CF-ACK frame. Furthermore, TSs usually are defined separately for uplink (UL) and downlink (DL). Thus for single connection two TSs are required which are specified in TSPEC and scheduled independently.

Different TSs scheduling and TXOP handling creates different frame sequence scenarios; however the scope of this paper is only one connection per STA for real-time traffic. All background best-effort traffic is assumed to use EDCA. Example of considered frame sequences is shown in Fig. 2.11. It shows common polling routine for UL (a) and DL (b), also packet transmission in both directions with enabled aggregation and piggybacking (c).

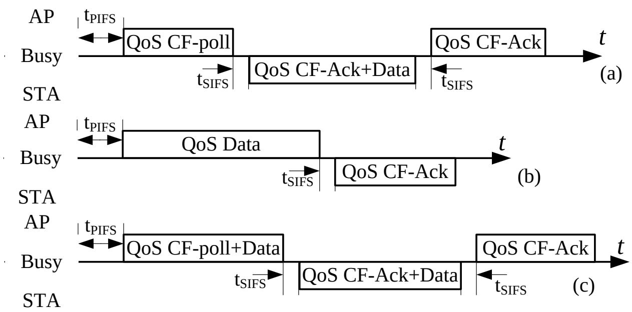

Fig. 2.11. Example frame sequences in HCCA for mutual talk state

Modelling voice transmission scenario, sending one packet per poll is most likely to happen, when SI does not exceed packetization time T_{pac}. In this case, TXOP will be granted for one packet only considering required overhead.

2.2.4. Voice Channels in HCCA

Studied network system is based on methodology presented in (Kajackas *et al.* 2007b). AP has a wired connection to the PSTN or IP network through a SIP proxy or H.323 gatekeeper. All calls are made from wireless nodes to outside network. The speech model is based on VBR four-state implementation according to (ITU Rec. P.59 1993). The wireless system, based on IEEE 802.11b (PHY 11 Mbps), is also IEEE 802.11e enabled running in HCCA mode for all wireless

STAs. Only reference scheduler will be evaluated in HCCA simulations. Also note that voice source and SI cycles are in perfect synchronization.

Wireless channel capacity was evaluated implementing voice source model and basic HCCA reference scheduler routines in *Matlab* environment.

Most common G.711 codec with voice activity detection is used for the analysis. During talk bursts codec generates 64 kbps data stream packing 20 ms voice samples into 120 B packets. After adding RTP/UDP/IP headers, we have 200 B packets sent at 80 kbps rate. In our case we sent homogenous voice traffic, so maximum expected MSDU can be set equal to nominal MSDU. However, setting M to higher values will not have effect on capacity, despite of longer allocated TXOPs as showed in (2.11).

Assuming alternating silence and talk periods mean data rate in long time period would be much lower than 80 kbps. However, to ensure predictable transmission on talk bursts, mean data rate equal to maximum data rate has to be assumed. Setting lower mean data rate will not gain any voice channel capacity whatsoever; because time required sending one maximum MSDU will be used in TXOP calculations (2.11). Mandatory TSPEC parameters for all modelled TSs are presented in Table 2.2.

Table 2.2. TSPEC parameters for simulated TSs

Parameter	Value
Nominal MSDU size (L)	200 B
Maximum MSDU size (M)	200 B
Mean Data Rate (ρ)	80 kbps
Maximum Service Interval (T_{SImax})	20 ms

It is easy to notice, that setting T_{SImax} equal to 20 ms will force the scheduler act as TDM-like algorithm, allocating TXOPs equal to time, required to send one packet including scheduling overhead and interframe spacings. Simulations showed that using VBR, TXOPs were often unused and transmission time passed to EDCA. However, the spare channel capacity cannot be used for additional voice channels, because of TSPEC restrictions.

The TS allocation algorithm controls STA registration by checking if TS with particular TSPEC can be allowed by maintaining (2.14) inequality.

TS allocation algorithm with and without TS aggregation registered 16 and 11 STAs (i.e. voice channels) respectively during simulation. These numbers show the capacity limited by TS admission mechanism rather than by wireless channel. Contention free transmission time (T_{CFP}), used for voice channels is presented in Fig. 2.12. Remaining transmission time up to T_{SImax} value is left for EDCA traffic.

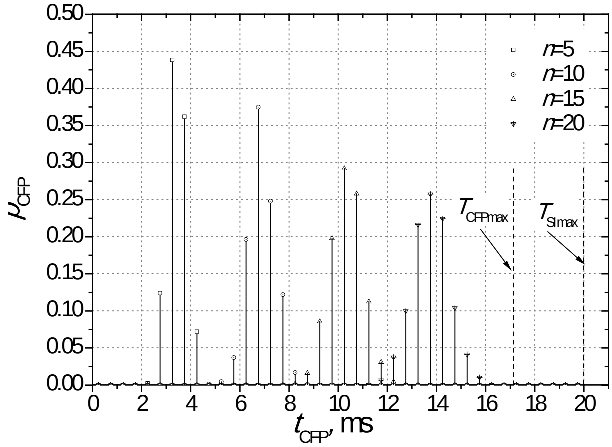

Fig. 2.12. Distributions of captured CFP durations t_{CFP}, with TS aggregation enabled

Contention period should be long enough for one maximum PPDU to transmit:

$$T_{\mathrm{CPmin}} = t_{\mathrm{PPDUmax}} + 2t_{\mathrm{SIFS}} + 2t_{\mathrm{TimeSlot}} + 8t_{\mathrm{ACK}} \,, \tag{2.15}$$

$$T_{\mathrm{CFPmax}} = T_{\mathrm{SI}} - T_{\mathrm{CP\,min}} \,. \tag{2.16}$$

T_{CPmin} value is slightly different form one calculated for original IEEE 802.11b, due to larger maximum frame body (12 bytes have been added by 11e amendment making totally 2324 bytes) and larger MAC header due to QoS related information (2 bytes have been added by 11e amendment making MAC header totally 36 Bytes length).

Additional voice channels can be gained by allowing reasonable packet loss to voice streams, i.e. granting access to more STAs than is allowed by default allocation algorithm. This hardly can be done by tuning the mandatory TSPEC parameters. Nominal or maximum MSDU size will not give any positive effect as those parameters are used for TXOP calculation and setting lower values may block the TS traffic. Lowering mean data rate does not make any sense either – when the system is tuned for one packet delivery per SI, TXOP value will not change even if TSPEC shows less than one packet per SI. Tuning T_{SImax} value may have harsh effect on quality and performance, because setting it lower than

T_{pac} will increase the overhead and decrease channel capacity; meanwhile setting T_{SImax} higher than T_{pac} will cause packet grouping thus may increase jitter.

Capacity increase may be achieved by modifying the scheduler itself and making it suitable for VBR traffic. This was analyzed in numerous publications.

However, capacity increase can be also achieved by modifying admission control method. This way can be advantageous, because much less complexity is introduced.

The VoIP channel capacity increase, related to this kind of modification, can be seen from Fig. 2.13 and Fig. 2.14 for transmissions without and with TS aggregation respectively. The occupied transmission time distributions were obtained from simulations with modified admission control mechanism, registering as many TSs, as could be fitted in T_{CFPmax}. In case TS aggregation is not used, the simulations show that no packet loss is introduced up to 15 STA, 16-th STA experience 0.45 % packet loss, 17-th STA VoIP session shows 4.15 % packet loss.

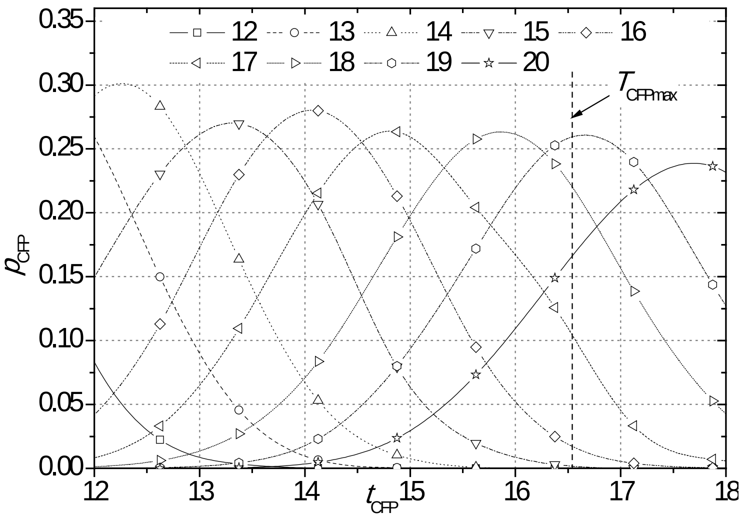

Fig. 2.13. Distributions of captured CFP durations t_{CFP} for 12–20 STA

In case of TS aggregation, no packet loss is introduced up to 20 STA, and 21-st STA experience 0.87 % packet loss probability.

Even though aggregation helps to suppress overhead, technical expenditures of reference scheduler are higher comparing to legacy PCF when one packet per poll scenario is used for real-time services.

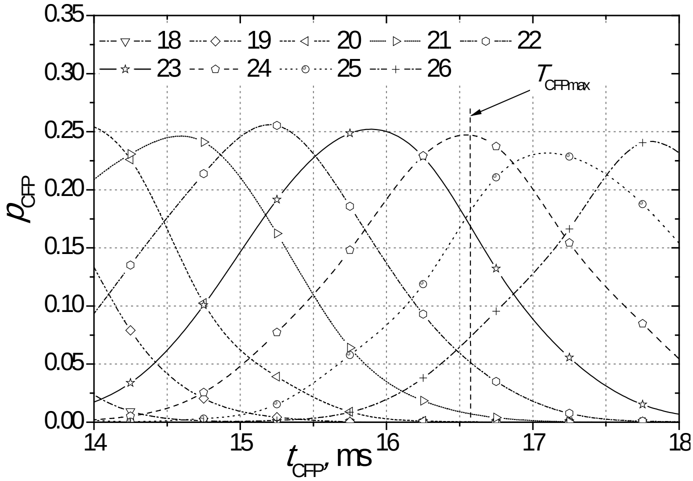

Fig. 2.14. Distributions of captured CFP durations t_{CFP} for 18–26 STA, with TS aggregation enabled

The simulation results show, that VoIP channel capacity can be extended using TS aggregation, ACK piggybacking and modifying TS admission algorithm. Due to higher overhead of reference scheduler, the capacity is lower than legacy PCF. Yet using HCCA allows much more flexibility, since TSs can be registered with certain delay bound requirements without modifying beacon period, which is not configurable in operation.

Table 2.3. VoIP capacity of simulated IEEE 802.11b/e channel

Access method (IEEE 802.11b PHY 11 Mbps)	Number of VoIP channels	Packet loss tolerance
IEEE 802.11e HCCA	11	0 %
802.11e HCCA, TS aggregation, piggybacking	16	0 %
HCCA, modified admission control algorithm	16	0.45 % for 16th STA 0 % for other STAs
HCCA, TS aggregation, piggybacking, modified admission control	21	0.87 % for 21st STA 0 % for other STAs
Legacy IEEE 802.11 PCF	26	0.4 % for 26th STA 0 % for other STAs

Allocation mechanism of TGe reference scheduler is not suitable for VBR traffic, not only because polling is arranged in TDM manner, but also because of TS allocation restrictions. Simulations show, that only by modifying TS allocation algorithm, system capacity can be improved by additional 5 voice channels. Capacity evaluation results are summarized in Table 2.3.

DL and UL scheduling efficiency is another way to increase the system capacity. Most common way of implementing TS scheduling is separate for UL and DL. Therefore technological expenditures to deliver single packet increase and system capacity decreases significantly. Aggregation of TSs may be used to solve this problem.

Aggregation bit in MAC header means that TSs within STA can be aggregated and scheduled together. Using aggregation and piggybacking, additional 5 voice channels may be gained comparing to separate UL/DL scheduling. The aggregation makes sense only when one UL and one DL TSs are used, this can be achieved if EDCA is used for any additional (non real-time data) services.

2.3. Managing Delays in Multi-hop Configuration

Road accidents and traffic jams are two most important problems on the roads. Most road accidents happen because of human error and could be avoided if drivers would be informed about the accident ahead at least several seconds before. Traffic jams could be decreased if traffic management organizations could receive detailed information about vehicles and their destinations and advise the driver to take alternative routes.

The answer for the mentioned problems above is inter-vehicle communication – wireless access in vehicular environments (WAVE). Recently WAVE is attracting much attention from industry and academia. The base for WAVE is IEEE 802.11p standard draft, which together with IEEE 1609.1/2/3/4 describes inter-vehicle communication. IEEE 802.11p amendment is intended for highly mobile vehicular environments with fast moving nodes. Communication mode is also different from usual Wi-Fi. In 802.11p not just different radio channels are defined, but also there is time division into two time channel slots: control channel (CCH) and service channel (SCH). Synchronization of CCH and SCH is done using Global Positioning Systems (GPS) receiver’s universal time clock (UTC) signals.

Using inter-vehicle communication the car suffering from accident or the car passing the accident is sending warning messages. There are several communication scenarios and one of them is multi-hop communication, where information travels from accident place to the cars which will cross it. This information routing is called geounicast, because information is sent to the relevant cars using travel path and current coordinates from GPS.

The most important task for the emergency warning system is to deliver warning messages on time. There can be several warning message types, but sudden brake or crash in front warning messages have to be delivered soonest. To calculate the permissible delivery time we refer to the recommendation to drivers to keep the distance from the front car same as half of the cars speed, which brings time between vehicles positions equal to 1.8 s. That means that after the crash in 1.8 s the following car should stop. The average reaction time of the drivers to accidents on the road is 1.8 s. Warning messages should arrive to the destinations faster than 1.8 s (how much faster should be answered by doing investigation on driver reaction to emergency warnings in the car). This principle is used by analyzing simulation results.

In this chapter the delay values of multi-hop link, based on legacy IEEE 802.11 and emerging IEEE 802.11p standard are analyzed. The results, obtained from simulations in NCTUns 5.0 environment, show delay distributions of emergency messages, broadcasted in multi-hop manner.

Multi-hop chain research is presented in (Jerbi *et al.* 2008) and is based on experiments with real cars using IEEE 802.11b technology. Different scenarios have been tested and results analyzed. Using 3 and 6 cars in the multi-hop chain is shown influence of hop count. Authors concludes, that multi-hop chain suites the needs of Vehicular Ad-Hoc Network (VANET). Though optimistic results, there are no hints to IEEE 802.11p, which differs from IEEE 802.11b. There were no background traffic generated, which influence the performance of network.

Packet delay in legacy IEEE 802.11 is analyzed in (Khalaf *et al.* 2006). Two transmission scenarios are presented: single-hop and multi-hop. Theoretical curves are compared with simulated. Therefore, there are some differences from inter-vehicle communication. The received packets are acknowledged, which is not the case in WAVE, where information is broadcasted.

Information dissemination in the network should be considered by building up the WAVE communication scenarios. A unified approach for disseminating data about different types of events in a vehicle network is presented in (Cenerario *et al.* 2008). This approach is not concentrating to a specific type of information, but it is unified approach based on encounter probability calculation, which gives a reason for simulated network described in this paper.

Two MAC methods have been evaluated according to their ability to meet real-time deadlines in (Bilstrup *et al.* 2009). IEEE 802.11p carrier sense multiple access (CSMA) was examined through simulation and conclusion was made, that CSMA is unsuitable for real-time data traffic. The second evaluated algorithm self-organizing time division multiple access (STDMA) will always grant channel access regardless of the number of competing nodes. Regardless the results of (Bilstrup *et al.* 2009), we show that standard CSMA suits the needs of WAVE (real-time deadlines is important, but we show, that the time limits are quite high for emergency messages to be transferred).

GeoMAC protocol, presented in (Kaul *et al.* 2008), exploits spatial diversity, inherent in a vehicular channel. Forwarder selection for transmission over the next hop is enabled in a distributed manner via geobackoff, which selects forwarders in decreasing order of spatial progress. Simulated network consists just of one hop chain, which does not answer to real life situation, but gives a clear overview of the possibilities of GeoMAC.

2.3.1. Vehicular Ad-Hoc Network Based on IEEE 802.11p

The upcoming IEEE 802.11p standard PHY has some differences of other IEEE 802.11a/b/g standards. As stated in (Bilstrup, et al., 2009), IEEE 802.11p will make use of the PHY supplement IEEE 802.11a and the MAC layer QoS amendment from IEEE 802.11e. WAVE PHY uses Orthogonal Frequency Division Multiplexing (OFDM). Radio frequency is similar to IEEE 802.11a and is allocated from 5.85 to 5.925 GHz into several 10 and 5 MHz channels. For USA communication channels are already defined and can be found in IEEE 802.11p standard and for Europe channel allocation is still in progress.

WAVE MAC is also specific and is described in IEEE 1609.4 standard. There is timing allocation of channels. Control Channel (CCH) is defined for emergency message transmission and for service advertisement and Service Channel (SCH) is responsible for all other information transmission. In the CCH time frame all stations should stop transmission and listen to this channel and receive/transmit emergency messages. During SCH channel time frame stations can use all other radio channels to transmit all types of information. Channels are divided into 50 ms frames. Time synchronization of channels is done using GPS universal time clock (UTC) signal. Emergency messages are sent by using WAVE Short Message Protocol (WSMP) described in IEEE 1609.3 standard.

The communicating nodes in VANET are moving fast and they should be ready for transmission as soon as possible. The WAVE Basic Service Set (WBSS) provider first transmits WAVE Announcement action frames, for which the WBSS users listen. That frame contains all information necessary to join a WBSS. Unlike infrastructure and ad-hoc 802.11 BSS types, the WAVE users do not perform authentication and association procedures before participating in the WBSS. To join the WBSS, only configuring according to the WAVE Announcement action frame is required. In addition, a node in WAVE mode shall generate a Clear Channel Assessment (CCA) report in response to a CCA request to know the time-varying channel state precisely.

Scenario of 5 lanes highway (Fig. 2.15) is used in this research. Following the idea of (Bilstrup, et al., 2009), that vehicle velocity is different in different lanes, following velocities are used: 19.4 m/s (70 km/h), 25 m/s (90 km/h), 30.5 m/s (110 km/h), 36.1 m/s (130 km/h) and 41.6 m/s (150 km/h).

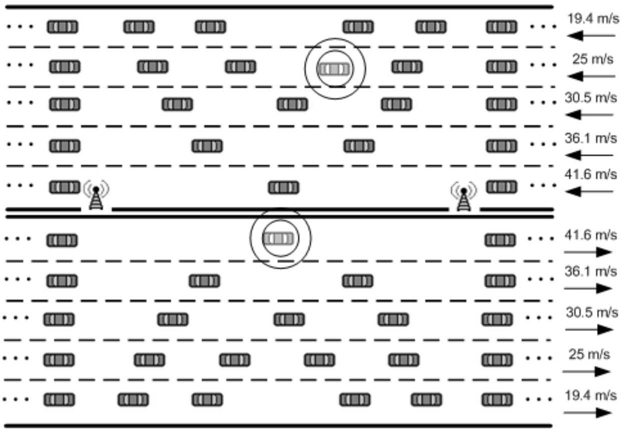

Fig. 2.15. Highway scenario (5 lanes)

According to described conditions there are ~100 vehicles in one communication range and this number is reflected in the simulation.

Simulations were performed in NCTUns 5.0 network simulation tool (Wang *et al.* 2008) under Linux Fedora Core 9 OS. NCTUns was chosen for its advanced IEEE 802.11 model library and ability to integrate with any Linux networking tools.

With the simulations we intend to investigate the delays experienced by the multi-hop link in vehicle ad-hoc scenarios. All simulations are based on IEEE 802.11a PHY and MAC, however the inferences about 11p performance can be drawn as well, since the contention mechanism is the same. In our scenarios only one type (priority) emergency message transmission is simulated, no other non-critical data transmissions are used; therefore the behaviour of IEEE 802.11a and IEEE 802.11p/IEEE 1609 is very similar. WSM transmission method is broadcasting, which does not require acknowledgements. WSMs in IEEE 802.11p/IEEE 1609 case may be transmitted in both CCH and SCH using legacy CSMA/CA. Thus, considering contention only between emergency messages, the results are valid both for legacy IEEE 802.11a and IEEE 802.11p/IEEE 1609.

The delays, introduced by CSMA/CA, theoretically can be evaluated by time expenditures calculation (Pavilanskas 2007). EDCA access mechanism is used for uncoordinated transmission (Kajackas *et al.* 2009c). In this case, time required to send the packet consists of actual packet transmission duration, inter-frame times and medium access delay:

$$t_{exp} = t_{AIFS} + rand(CW) \cdot t_{slot} + t_{packet}, \tag{2.17}$$

where t_{exp} represents total time expenditures for one packet transmission, t_{AIFS} – time required for Distributed Inter Frame Space ($t_{\mathrm{AIFS}} = 9 \cdot t_{\mathrm{slot}}$ for IEEE 802.11e AC0), CW – Contention Window, t_{slot} – slot time ($t_{\mathrm{slot}} = 9$ µs for OFDM, IEEE 802.11a), t_{packet} – time required for data and overhead transmission consisting of preamble, 30-byte MAC header transmission time – t_{MAC} and 4-byte Frame Check Sequence – t_{FCS}:

$$t_{packet} = t_{PLCP} + t_{MAC} + t_{MSDU} + t_{FCS}. \tag{2.18}$$

Since no acknowledgement is required for broadcasting, no other expenditures take place.

Contention window defines the set of possible delays for back-off algorithm. Every collision in wireless channel results congestion window to double, shifting from minimum value of $CW_{\mathrm{min}} = 15$ to maximum of $CW_{\mathrm{max}} = 1023$ slots for AC0 access category.

IEEE 802.11a PHY was modified to support IEEE 802.11p PHY rates. In simulations we use the lowest possible 3 Mbps PHY rate. Lowest modulation gives the best reliability and transmission range. Considering always changing radio environment on the roads due to unexpected obstacles (large vehicles, blocking the signal, rapid fading due to movement, etc.), the ability to use higher modulations is unpredictable and may lead to failure of transmission, thus the simulations are designed for worst-case radio transmission scenario. However, the presented results can be theoretically recalculated for any other PHY rate.

Emergency messages are simulated as 500 byte UDP packets. Following the idea of (Bilstrup *et al.* 2009), packet length of 100 bytes is just long enough to distribute the position, direction and speed, but due to security overhead, the packets are likely longer. According to that, packet length of 500 bytes is chosen. Messages are routed through the network using IPv4. Since no movement is simulated whatsoever, we use static routes to make controllable transmission through hops. Because all simulations are generally done on IP network, the initial TTL value is modified to make hopping through large number (greater than 64) of hops possible.

All the transmissions in the simulated network use layer 2 broadcasting.

Theoretically, using PHY rate of 3 Mbps and 500 byte payload (plus 8 byte UDP header, 20 byte IPv4 header and 8 byte LLC to form single MSDU), according to (2.17) and (2.18), time expenditures for single emergency message delivery can vary from 1.621 ms to 1.756 ms if no collisions effect contention window and wireless medium is always free to access. With more hops, the variation is higher.

2.3.2. Single Emergency Message Transmission Simulation

First scenario (Fig. 2.16) simulates single emergency message transmission through multi-hop chain. All nodes are located within radio transmission range and operating in the same radio channel, therefore they share the channel with equal rights. Since all the packets are being transmitted as broadcasts, they are received by all stations and not acknowledged. To control the "hopping" to one direction and to avoid broadcast storms, we filter packet forwarding and route them hop-by-hop.

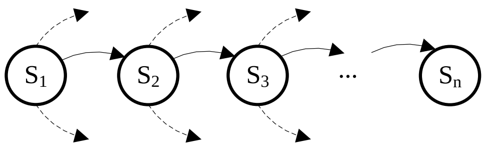

Fig. 2.16. Single emergency message transmission through multi-hop chain

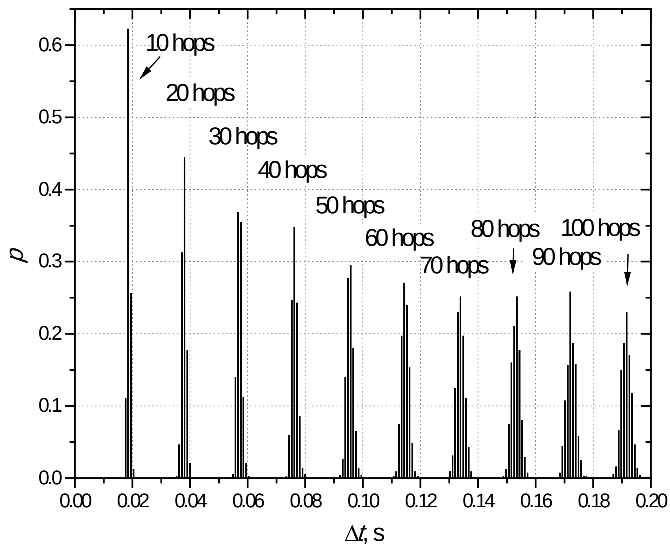

Fig. 2.17. Delay Δt rate distributions for different hop number

The delay was measured at every node and delay distributions are presented in Fig. 2.17. The mean delay for 100 hops reaches 189.3 ms, minimum and maximum values respectively 184.4 ms and 194.0 ms. The delay and delay fluctuations are relatively small due to low channel utilization. There is only one packet in the system at any given moment, therefore no contention takes place.

However, this scenario is not realistic in VANETs and is presented to give understanding of transmission delays in perfectly controlled environment and to evaluate the minimal influence of MAC layer and physical transmission of signals. This scenario can be considered as a worst-case for reliability and a best-case for traffic load.

Another set of simulations demonstrates how channel utilization influences the delay spread.

There are many investigations on efficient message broadcasting, and for the simulations we take into account, that data dissemination with broadcasts can be controlled in the network.

Our presented simulations are broadcasting solution independent and may be used to evaluate solution influence on transmission delay over different number of hops. The concept of “background traffic” has to be understood as an overhead, created by broadcasting method. Network topology remains the same, but more traffic is introduced into network as background traffic along with emergency message stream. Background traffic is generated by neighbouring nodes on the same radio channel and has the same characteristics as measured (emergency message) traffic.

One of the problems in emergency message transmission in VANETs is reliable and at the same time efficient and robust broadcasting. Inevitably it has to have significant overhead to ensure guaranteed reception. On the other hand, the overhead has to be reduced in order not to over utilize the radio channel, which will eventually lead to reception failures or extreme reception delays. Guaranteed reception can be achieved by acknowledging, however the messages have to spread fast, therefore there is no time for seeking best route in node mesh or confirming the reception. Broadcast messages cannot be acknowledged, thus the reliability has to be ensured by repeated broadcasts and neighbour retransmissions. This way the channel can be easily flooded with broadcasts degrading network performance with excessive delays.

Fig. 2.18 shows delay distributions for different number of hops when light background traffic of 100 kbps has been applied. The delays are more spread and shifted, however the influence is relatively small due to low channel utilization: for 100 hops the mean delay increases by 12 ms and maximum delay – by nearly 30 ms. By increasing the background traffic further, delay distributions shift and spread more.

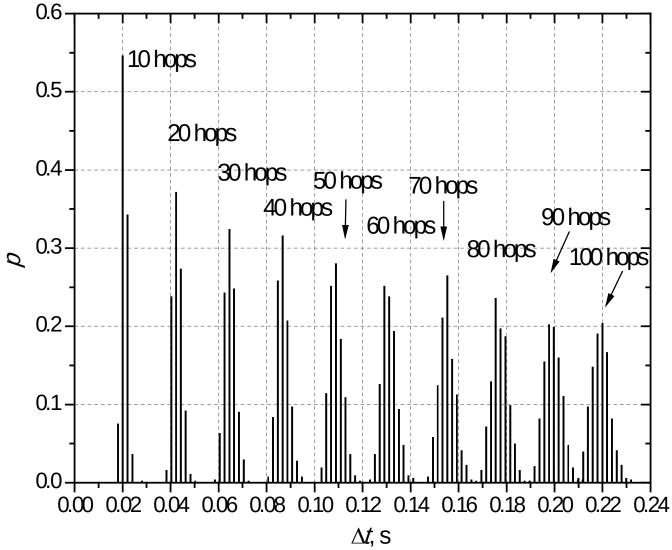

Fig. 2.18. Delay Δ*t* rate distributions for different hop number with 100 kbps background traffic

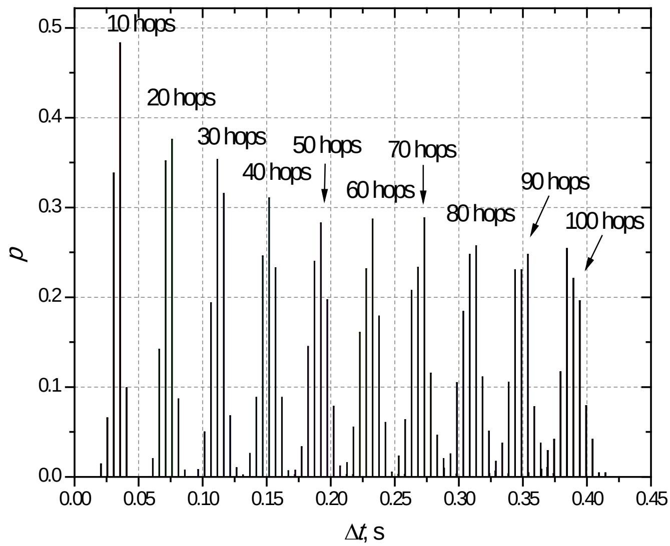

Fig. 2.19. Delay Δ*t* rate distributions with 1 Mbps background traffic

Fig. 2.19 shows delay distributions with 1 Mbps background traffic and Fig. 2.20 – with 2 Mbps background traffic. Those graphs do not include lost packets. With significant background traffic, the contention for transmission becomes harsh and collision probability increases causing packet loss. Since broadcast packets are never acknowledged, lost packets are not resent and hopping through node chain brakes. Fig. 2.21 shows the probability for packet to survive different number of hops.

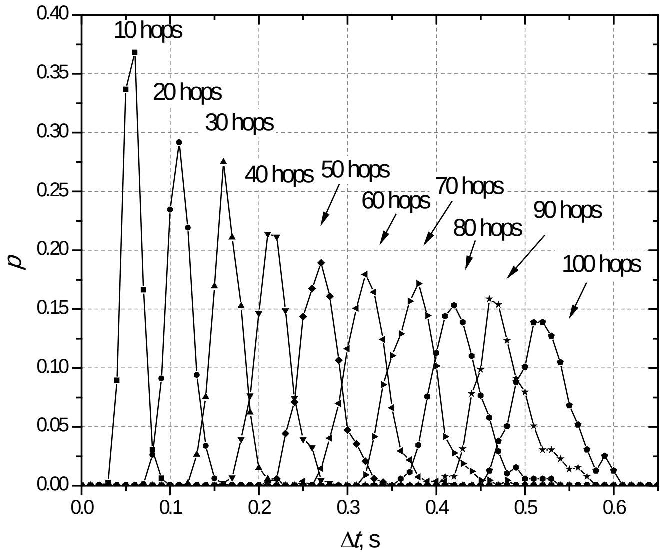

Fig. 2.20. Delay Δt rate distributions with 2 Mbps background traffic

The summary of results for 100 hops is presented in Table 2.4. It is shown, that by increasing background traffic the mean delay is growing proportionally, but standard deviation is increasing. This means, that with growing background traffic the delay can vary in wider time range.

Table 2.4. Results summary for 100 hops

Background traffic, kbps	Mean delay,s	Minimum delay, s	Maximum delay, s	Standard deviation
0	0.189	0.184	0.194	0.00173
100	0.218	0.209	0.230	0.00383
1000	0.386	0.355	0.411	0.00839
2000	0.523	0.458	0.603	0.03083

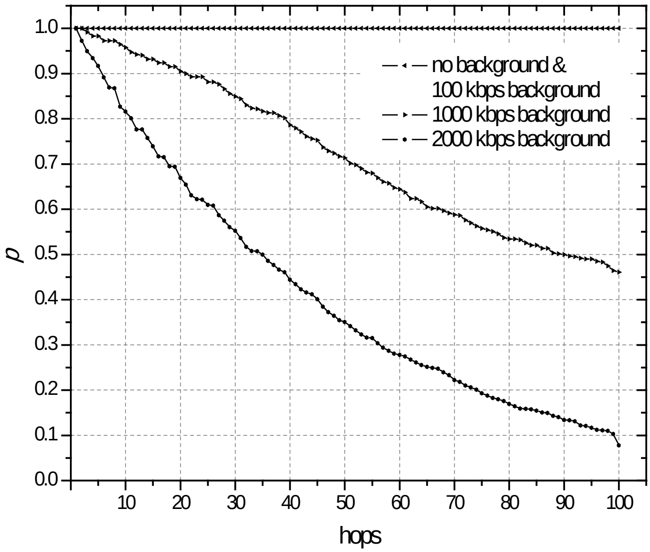

Fig. 2.21. Packet survive probability *p* for different hop number

2.3.3. Controlled Flood Scenario

One of the ways to improve reliability of multi-hop links is to make redundant paths to every node of the network. Flooding the network with broadcasts may seem the reliable way to ensure message reception for every network node. Since the transmit range is not always known due to ever-changing environment, every node in the network has to retransmit (rebroadcast) emergency message assuming that it may be at the transmission range edge of the message initiator. For this scenario an algorithm, controlling the floods must be employed, otherwise packet loops will cause broadcast storms (similarly as in looped Ethernet) which eventually will lead to channel congestion. One of the ways to avoid loops could be GPS coordinate tracking and making sure, that broadcasts are being forwarded only in one direction (similar as in (Kaul *et al.* 2008)). This can be tricky considering vehicle movement. Another simple way – logging retransmitted node IDs: all nodes, retransmitting broadcast packets, put their ID into the frame body; before resending received packet, node always searches this ID list; if own ID is found, the packet is dropped assuming it is in the transmission loop.

We implement this controlled flood scenario in NCTUns 5.0 using same nodes and traffic characteristics as defined in previous chapter. The network topology is depicted in Fig. 2.22.

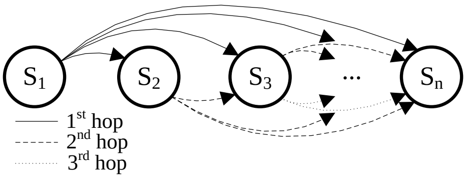

Fig. 2.22. Controlled flood scenario

S_1 is the originator of emergency message, which is broadcasted through the network. Every other node broadcasts the same message again following basic rule: if source ID is lower than own ID, then message should be broadcasted. Otherwise – received packets have to be dropped.

This way the network is flooded with the message copies, but no broadcast loops appear. This scenario can be considered as a worst-case for traffic load and a best-case for reliability.

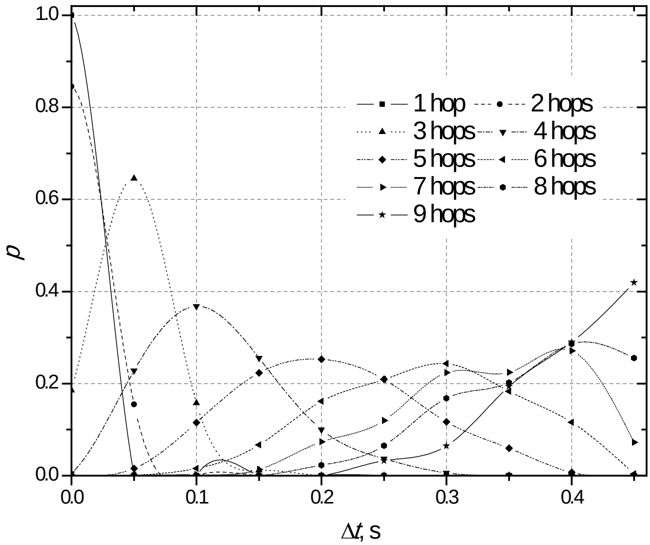

Fig. 2.23. Delay Δt rate distributions in 10-node controlled flooded scenario

Delay distributions for 10 and 20 hops scenario are presented in Fig. 2.23 and Fig. 2.24. The delays were measured at every node.

Since the broadcasts from any node are received by all other nodes and retransmitted by all with the ID higher than source ID, increasing node (hop) number, the packet copies in the system grows exponentially. It can be seen, that 10 node scenario shows quite reasonable delays, reaching 500 ms for all 9 hops, however doubling node number in the scenario results in excessive delay increase, mean value reaching almost 4 seconds for 9 hops and 7 seconds for 19 hops.

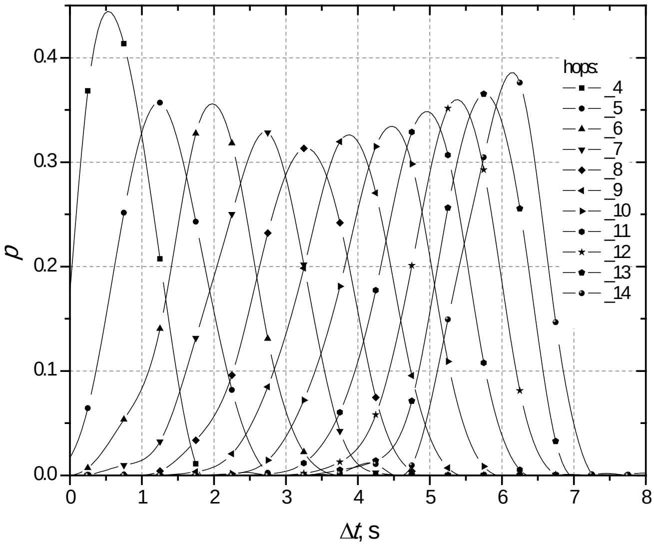

Fig. 2.24. Delay Δ*t* rate distributions in 20-node controlled flooded scenario

Permissible delay for the first car line (closest to accident place), is less than 1.8 s. This time is the reference for result analysis.

There is just one packet in the multi-hop chain in single message transmission simulation scenario. In this scenario, even with a big background traffic (2 Mbps) the maximal delay is 0.6 s, which is in permissible range – less than 1.8 s.

Analyzing controlled flood scenario simulation results, delay up to 7 s is found. The results for 10 nodes chain (Fig. 2.23) can come up to 0.5 s and are satisfying the permissible delay. But for the 20 node chain (Fig. 2.24), the delay can come up to 7 s. This scenario shows, that just communication path of 4 nodes satisfies the permissible delay. This means, that for the 20 nodes scenario the first car, following crashed car, should get the emergency warning maximum after 4 nodes in multi-hop chain. If the car after accident is in the second row, the permissible time grows up to 3.6 s. This means, that second car can get the emergency message from the chain of maximum 7 nodes.

The delay in IEEE 802.11 multi-hop transmission depends on following major components: physical signal transmission, which depends on PHY rate and distance; and contention, which depends on channel utilization. The problematic of emergency message transmission is two-fold: transmission has to be reliable and transmission delays have to stay in strict limits.

Presented three sets of simulations show the emergency message delay dependency on hop count in channel utilization best-case, delay dependencies on different loads and reliability best-case.

Simulations show, that single message propagation is in permissible range even for 100 nodes. For controlled flood scenario node number increases delay exponentially. Therefore growing node number influences the delay time and the chain for emergency message transmission is getting smaller to satisfy the permissible delay results. Communication chain length is also dependent on the car position from the accident place.

Message broadcasting methodology has to be chosen carefully, taking into account the traffic overhead. This problem is illustrated with controlled flooding scenario.

2.4. Performance of MAN Wireless Access Networks

High spectral efficiency and flexible QoS support are usually referred as the key factors for success of emerging next generation wireless networks. Limited spectrum licenses and continuously growing demand for high network capacities are forcing operators to seek for spectrum-efficient solutions. Today's telecommunication market shows great interest in personalized multimedia services, interactive real-time audio and video applications, therefore the differentiation and quality guarantee are a must. Together with the need to serve different applications and maintain certain grade of QoS comes the complexity of QoS management. Migrating to all-IP technologies, such as WiMAX or LTE helps operators to gain flexibility and scalability; however IP-related quality challenges remain.

Most of the WiMAX roll-outs today use mobile version of IEEE 802.16 standard, namely IEEE 802.16e-2005. Based on OFDM TDD and OFDMA PHY, WiMAX provides high spectral efficiency, scalability and may provide high data throughputs.

There are four QoS classes defined in IEEE 802.16 standard: unsolicited grant service (UGS), Real-time polling service (rtPS), non real-time polling service (nrtPS) and best effort (BE). In addition to four classes specified by IEEE 802.16-2004, later amendment IEEE 802.16e-2005 introduced extended real time polling service (ertPS) class. As the different QoS classes are implemented using different scheduling and bandwidth allocation methods, the resources required are also different. Standard does not define the implementation of schedulers.

Radio resource management in mobile environment always poses a challenge. Enforcing advanced schedulers and admission control often is a trade-off between optimum resource usage and actual achieved QoS for particular applications. For sake of flexibility and cost, soft-QoS mechanisms are employed, which allow preserving spectrum, but leaves QoS management a challenging task. Available bit rate of radio channel may vary in wide range due to varying nature of RF conditions and user movement, therefore the scheduling of transmissions have to adapt fast and network planning has to be done carefully.

To analyze the scatter of network performance parameters, the trace of bandwidth and delay values were collected experimentally in live and loaded WiMAX network in Vilnius, operated by operator LRTC. Main network parameters are presented in Table 2.5.

Table 2.5. Tested WiMAX network configuration

Parameter	Value
Standard	IEEE 802.16e-2005
Frequency band, GHz	3,5
Channel Width, MHz	10
Subcarriers	1024
DL:UL ratio, symbols	29:18
MIMO	2x2
Scheduler	BE
QoS settings	None

Available throughput was measured with continuous UDP stream. TCP test reflects the variations of available bit rate, however fails to provide accurate instantaneous values if used in movement. TCP congestion control mechanism adapts too slow if moving at vehicular speeds and handing over from BS to BS.

Fig. 2.25 shows experimental data of available bit rate measurement in mobile WiMAX network. The results were collected measuring UDP traffic throughput while travelling at vehicular speed (total distance – 11 km, total time – 22 min average speed 30 km/h).

The data presented here illustrates the problematic of varying network conditions. In mobile applications time and place are linked, therefore variations in time and space create unpredictable fluctuations, caused by changing radio environment and handovers to base stations with unknown resource utilization.

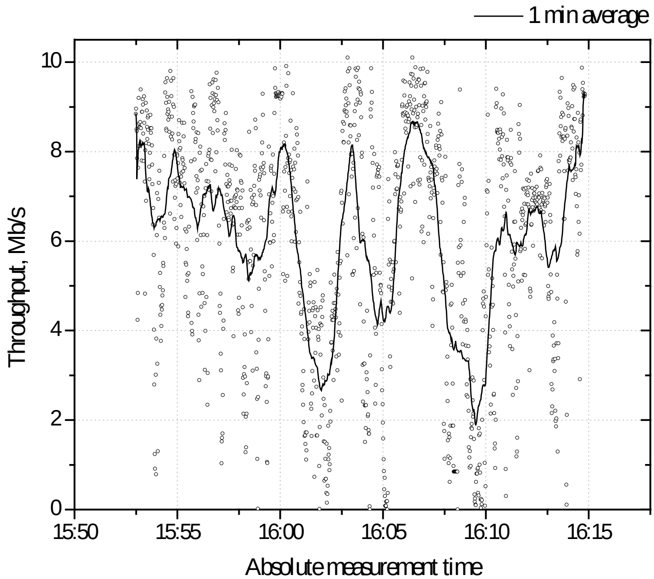

Fig. 2.25. Achieved throughput measured in movement

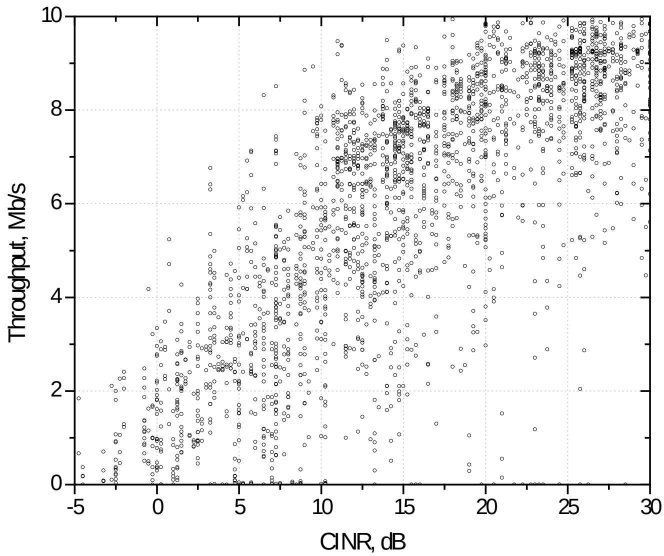

Fig. 2.26. Throughput relation to CINR values

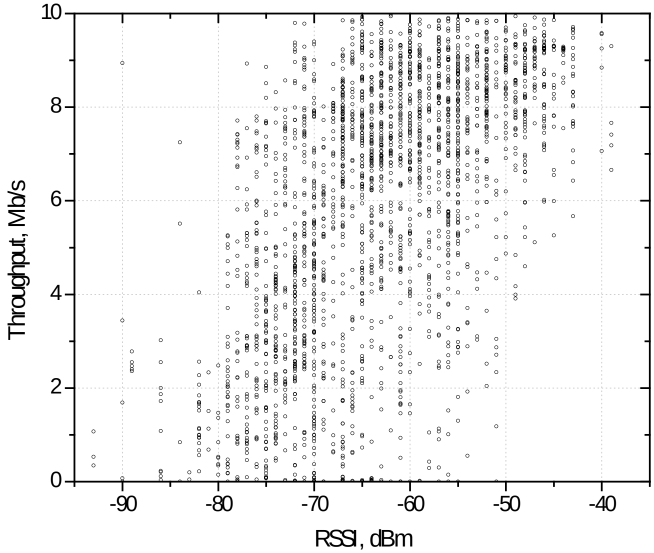

Fig. 2.27. Throughput relation to RSSI values

Throughput values have been read every second. It can be seen (Fig. 2.25) that average in relatively long time (1 minute average, continuous line) shows satisfactory service, while short time values (1 second average) may drop drastically during handovers or at high interference areas.

Since mobile network performance depends on many variable factors, quality of service relation to physical network parameters can not be followed.

Fig. 2.26 and Fig. 2.27 show measured throughput dependency on physical access network parameters – RSSI and CINR. It can be seen, that deviations are high in higher CINR and RSSI values, therefore can not be unambiguously mapped.

2.5. Conclusions of Chapter 2

- QoS provisioning in WLANs require additional technological expenditures. The overhead can be managed by improving scheduling techniques and adapting them to service specific behaviour.
- Modelling voice channels over half-duplex transmissions gains more accuracy if conversational four-state speech model is applied.
- Synchronization of beacon interval in IEEE 802.11 PCF with voice coding interval can increase voice channel capacity in WLANs while main-

taining strict quality requirements. For example, in 11 Mbps PHY case, capacity is improved by 4 channels with no packet loss, or 5 channels with 0.004 packet loss probability.

- Applying traffic stream aggregation, piggybacking and modifying admission control algorithm in IEEE 802.11e HCCA (11 Mbps PHY), system capacity for VBR voice channels can be improved nearly by factor of 2.
- Multihop configuration for such delay-sensitive applications as emergency message broadcasting in VANETs has to employ efficient broadcasting protocols, optimizing network load and reliability. Simulation results show, that marginal underutilized IEEE 802.11 based multihop network can deliver messages in less than 0.2 s over 100 hops, whereas higher utilization may lead to 0.35 s and 0.6 s with packet survival possibility of 0.4 and 0.1 for 1/3 and 2/3 channel utilization factor respectively.
- Experimental analysis of wireless wide area networks, like WiMAX, show unpredictable fluctuations of throughput. These hardly manageable fluctuations are caused by changing radio environment, handovers and unknown resource utilization.
- Experimental measurements in WiMAX network show that actual achieved quality of service cannot be directly judged by physical parameters like RSSI and CINR, measured at user device
- Specialized monitoring systems have to be created in order to evaluate actual quality of service, achieved by operator and perceived by end-user.

3

Reference Design of Quality Monitor

This chapter presents the analysis of opportunities to implement real-time monitoring of QoS achieved by operator and perceived by user and proposes a reference design of QoS monitoring system.

In order to implement full operator-user SLA enforcement based on actually achieved and user perceived QoS, monitored quality level has to be accounted and compared to requirements. Although accounting mainly has to be used to track fulfilment of SLA commitments, it may be employed for numerous secondary reasons, including data collection for operator’s network optimization, statistical data for telecommunications regulator or as a personal record for user.

Designed test-bed of actual achieved QoS analysis and tracking offers advanced capabilities both for SLA composition, including service independent network performance emulation for live service trial, and SLA enforcement in terms of non-invasive real-time monitoring of specific factors influencing QoS.

Analysis and part of the results presented in this chapter have been published by author in (Kajackas *et al.* 2010b) and (Vindašius 2010).

3.1. Quality Monitor Requirements and Opportunities

Wireless access network influence on quality of service in general is two-fold – influenced by two main factors:

- the radio environment and technological properties of the transmission system;
- the short-term and average long-term utilization of resources in access network.

Both of these factors vary in time and are influenced by network planning efficiency. For sake of economical advantage, some low-populated areas are intentionally provided with lower capacity and poor radio performance because of high distances to base stations in large cells. Whereas dense and demanding areas are provided with dense small cell placements, which provide more overall throughput and ensures robust radio performance by employing high efficiency modulation-coding schemes (MCS) enabled by high achievable CINR.

Even though service degradation cause is different, both factors influence the service quality and cannot be distinguished from perspective of end user experience. Tracking physical parameters of radio link tells only part of the story, because utilization of BS is neglected. Furthermore, user applications have different requirements, which usually cannot be directly converted to low-level physical parameters. Inevitably the analysis of QoS impairments has to rely on user-perspective, therefore the need of monitoring and indication of user perceived QoS is required.

The measures capable of evaluating perceived quality during service delivery do not yet exist. Analysis of this problem has to start with forming primary requirements. Before forming fundamental model for perceived QoS monitoring system, we have to consider the following:

1. PQoS monitoring system has to be universal and not very complex, to fit into various user devices, which have limited processing power and memory capabilities.
2. Currently accepted QoS evaluation and measurement methods as PESQ, PEVQ for voice and other cannot be applied directly during service delivery and cannot be applied for real-life services quality monitoring.
3. The system has to evaluate quality impairments or quality degradation ΔQ, but not the absolute quality; the solution has to be focused to variations from defined SLA.
4. The data about primary quality impairments during actual service delivery have to be obtained indirectly – related to parameters, used to describe network performance. Available bit rate ρ, packet loss π and la-

tency τ can be used initially for this purpose. These are well known and most influential parameters on voice and video quality.

5. The same measurable variations of parameters ρ, π and τ may have different effect on perceived QoS depending on applications, information being transmitted, signal performance, coding, etc. In general, the value of short term quality variations depends on informational value of lost or corrupted packets (Kajackas, *et al.*, 2009a).
6. The monitor has to be defined according to the task – which of the categories (PQoS, SPQoS or APQoS) is intended to evaluate. Relations between parameter variations and perceived quality have to be defined accordingly.

Following the requirements, we state following basic rules for creating algorithm of QoS monitoring system.

Initially, time scale is segmented:

$$t_{i+1} = t_i + T^x \quad , \quad i = 0,1,2..., \tag{3.1}$$

where T^x is segmentation period, depending on monitor target and requirements. For GPQoS evaluation, T^x is chosen according to MOS requirements (ITU-T Rec. P.800 1996; ITU-T Rec. P.910 1999). For other cases, evaluating SPQoS, T^x may be equal to service quality noticeable impairment periods (Pastrana *et al.*, 2004; Pastrana 2004). Impairment periods for voice and video are different.

Primary quality impairments and corresponding network performance parameters ρ, π and τ should be defined for every measurement period (t_i, t_i+T^x) as well as degradation of quality:

$$\Delta Q(t_i) = Q_0 - Q(t_i) \,. \tag{3.2}$$

Q_0 in (3.2) shows the quality level offered by service provider or defined in SLA.

Relations between quality degradation $\Delta Q(t_i)$ and network performance parameters ρ, π and τ are complex and usually cannot be clearly expressed. This problem can be solved in several ways depending on task of the monitor.

In order to find noticeable impairment periods, the quality degradation $\Delta Q(t_i)$ can be expressed as following polynomial:

$$\Delta Q(t_i) \approx \frac{\delta}{\delta\rho}\Phi(\rho,\pi,\tau)\cdot\Delta\rho + \frac{\delta}{\delta\pi}\Phi(\rho,\pi,\tau)\cdot\Delta\pi + \frac{\delta}{\delta\tau}\Phi(\rho,\pi,\tau)\cdot\Delta\tau, \tag{3.3}$$

where $\delta/\delta x\ \Phi[.]$ is functional derivative of x; $\Delta\rho$, $\Delta\pi$, and $\Delta\tau$ – deviations of parameters ρ, π, and τ respectively.

Equation (3.3) expresses the degradation of quality as a sum of factors, influencing QoS. Such model, when overall QoS degradation is expressed as linear sum of separate impairments, correspond to ITU-T E-model (ITU-T Rec. G.107 2000).

Another way of implementing QoS degradation monitor is based on fact that quality degrading factors ρ, π, and τ vary randomly and the accumulated effect on quality change is stochastic as well. Under these conditions, one of the possible solutions is proposed in paper (Kajackas *et al.* 2009b) which allows calculating distributions of the de facto perceived quality, using traces of lost frames and using conditional rates of quality categories. Some examples of such experimentally estimated conditional rates of quality classes are presented in (Kajackas *et al.* 2009b).

An overall variation of quality is evaluated by aggregating short-time noticeable quality impairments. Various aggregation mechanisms can be applied; however, simple averaging is not suitable, because average value fails to reflect actual perceived quality. Classification algorithms are more suitable for aggregation.

Two-class classification model can be successfully employed, keeping in mind, that quality and its metrics are not strict. Quality of service can be classified to two levels by setting a threshold of allowed quality degradation ΔQ_0. First level represents satisfactory quality of service, where quality degradation is smaller than allowed ($\Delta Q < Q_0$). Second level represents degraded (unsatisfactory) quality ($\Delta Q \geq Q_0$). Service delivery time is then divided to set of intervals $T^{\Delta Q<Q_0}$ and $T^{\Delta Q \geq Q_0}$ which aggregates to total durations of satisfactory (T_h) and degraded (T_m) quality:

$$\begin{cases} T_h = \sum_{i=1}^{n} T_i^{\Delta Q < \Delta Q_0} \\ T_m = \sum_{i=1}^{n} T_i^{\Delta Q \geq \Delta Q_0} \end{cases} \qquad (3.4)$$

To provide evaluation for end-user as single value (score), it may be presented as perceived quality coefficient, derived from ratio of normal and degraded quality durations T_h and T_m:

$$\kappa = \frac{T_h}{T_h + T_m}. \qquad (3.5)$$

Applying this perceived quality evaluation method, users may be offered different tolerated ΔQ_0.

Functional structure of proposed PQoS monitor is shown in Fig. 3.1. User service data flow from communications channel arrives to testing tool set, integrated into user device, where three main factor groups are measured: ρ, π, and τ. PQoS analysis tool has the task of classification and evaluation how these factors influence PQoS of particular service. PQoS monitor processes the PQoS evaluations, stores, performs accounting functions, and also indicates PQoS level to user.

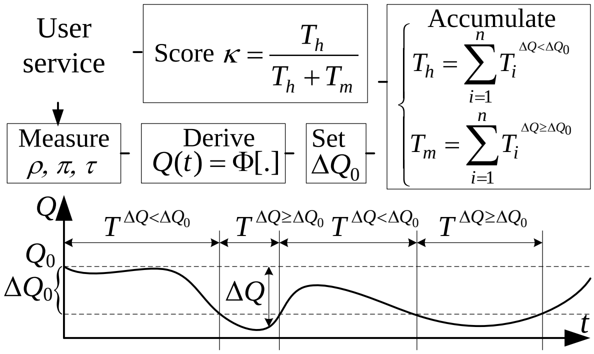

Fig. 3.1. Functional structure of PQoS monitor

It is important not only indicate the QoS impairments, but also collect and store the measurement information for later analysis to discover the cause of impairment. For this reason system architecture should include low-level physical radio parameter and possibly GPS logging capability.

The form of algorithm (3.3) as well as the structure of PQoS monitor is universal and applicable to different services. It is obvious that the derivatives $\delta/\delta x\ \Phi[.]$ are different for different service types and have to be defined separately: each factor may have a particular effect on voice transmission, but different effect on data services, e.g. web browsing. The structure of PQoS monitor may also be different for particular user devices – video, voice terminals. For example, voice quality on mobile circuit switched networks depends on bandwidth and packet loss; meanwhile latency variations are not common.

There are two main problems applying (3.3) algorithm. Firstly, it has to be defined for what period factor ρ, π, and τ characteristics have to be measured. Secondly, it has to be clear what influence particular factors have on quality.

Solving first problem, it is worth noticing that bit rate, packet loss and delay variations are easily obtainable by measuring these parameters at user device. This information usually is accessible to developers and vendors of mobile user devices. For example, in cellular networks packet loss is tracked with so called Bad Frame Indication (BFI), this information is used in voice decoding process, enabling lost frame substitution functions (3GPP TS 46.001 V8.0.0 2008). Analysis shows that BFI meets the requirements for packet loss measurement, therefore can be used to determine voice quality degradation in real time.

The available bit rate per user in mobile wireless network depends on coding scheme, which is adapted by network, therefore is always known. This algorithm can be implemented in end user devices using BFI as lost frame indication.

Voice or video transmission may not care about available bit rate as long it is higher than required by codec, however packet loss or delay may be critical impairment. Well known AMR codecs can adapt codec rate to available bit rate at the expense of quality, therefore the influence of this factor would be direct. On other hand, quality of web browsing is influenced both by available bit rate and packet delay, but degradation of one factor may be compensated with another.

To create descriptions of $\delta/\delta x\ \Phi[.]$, service models are needed, therefore services of interest have to be analyzed deeply. The goal of service research and simulation is to translate network impairments into human perceived QoS.

Individual monitors of perceived QoS will create the possibility for the user to obtain objective information and indication whether operator delivers sufficient level of QoS under SLA commitment. These monitoring systems may also help users to choose service provider.

3.2. Reference Design of Quality Monitoring System

Traditional interaction between end-user and network operator or service provider is formalized in terms of SLA provisions. However, common non-business customer contracts include barely formalized quality requirements or none at all. Upon subscription, user presents or picks from presets of SLA definitions, composed of network performance parameters, which are usually limited by max bandwidth. In special cases other parameters such as packet-loss rate, guaranteed and max bandwidth and network latency boundaries can be defined. Usually those metrics are accompanied by network availability parameters, maximum network outage times and troubleshooting durations.

This traditional way of composing SLA is not convenient for average user, because it does not indicate actual service performance and requires user to have high level of technical expertise. On other hand, if user demands any non-standard quality requirements, SLA customization results in high expenses. And most importantly, the link to actually achieved or perceived QoS is missing. The interaction between user and QoS requirements (defining SLA, service subscription) also interaction between user and achieved or perceived QoS (enforcing SLA, using service) can be understood as ambiguous subjective relations (Fig. 3.2).

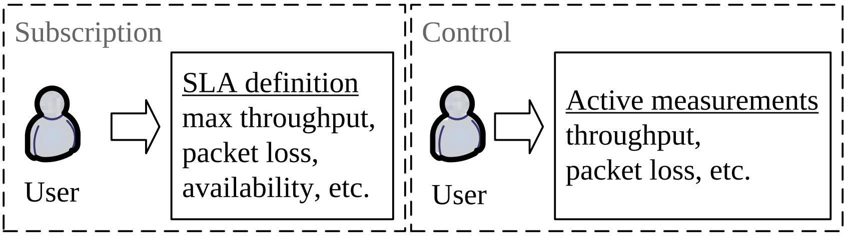

Fig. 3.2. Traditional user and SLA interaction upon subscription and control

The control of how SLA provisions are enforced is not user-friendly as well. There are no objective means of tracking how the requirements are met. User may initiate active measurements, obtain required values and complain to operator for particular inconsistency with SLA. User can use variety of on-line bandwidth or packet loss measuring tools; however, the results poorly reflect the performance of particular service. The measurement result varies in time and place; also, is logged by user, who does not have precise methodology and measures. Therefore, the argument with operator may become difficult and time consuming procedure. SLA uncertainties and lack of enforcement measures turn out only during arguments between unsatisfied user and operator.

To solve this problem, the concept of forming requirements and tracking the fulfilment has to be redesigned for sake of flexibility, complying individual needs and accountability.

Following the requirements defined in 3.1, the reference design of advanced service-aware QoS monitoring and accounting system is proposed (Fig. 3.3). The subscription to service or, in other words, SLA composition is defined not by user directly, but rather by “QoS training” module, which uses user input through designed interface, service descriptors and network capabilities information. SLA is presented or composed by user in more understandable way of trying service and providing evaluation.

This approach rises from need to provide customizable user interface to meet the challenge of linking *QoS required by user* in user (subjective) domain and *QoS offered by operator* in operator (objective) domain.

Proposed SLA enforcement system is depicted in Fig. 3.4. The purpose of this system is to track, indicate and account actually received QoS by passively collecting network performance parameters. User interaction is not needed at this stage, the monitoring is done in real-time and provides indication to user, also stores, accounts and provides this information to operator or telecommunications regulator if needed or requested. PQoS module, having SLA provisions and passive measurement results makes a decision on PQoS level and informs user if he/she wishes to. This evaluation is also stored for statistical and accounting needs of operator and/or regulator. Raw passive measurement results of network performance parameters may also be stored and provided to operator for network optimization and troubleshooting purposes.

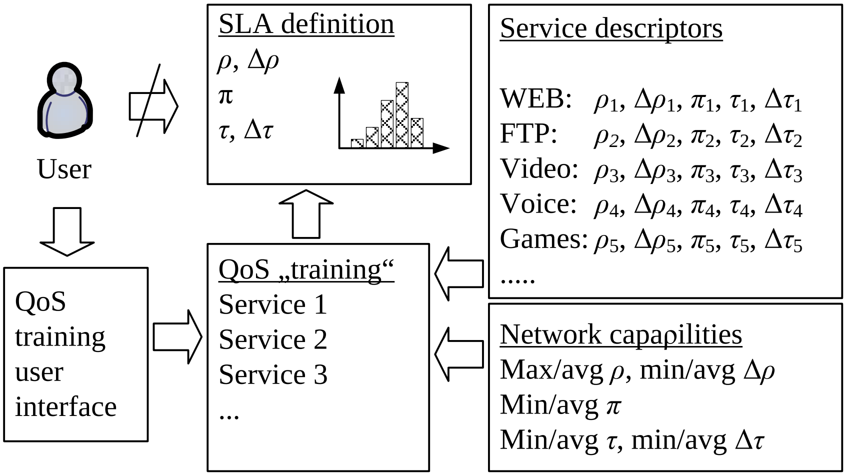

Fig. 3.3. Proposed subscription to service scheme

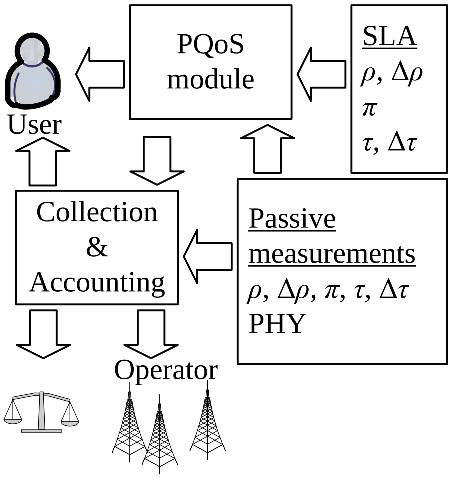

Fig. 3.4. Proposed SLA enforcement scheme

Since perceived QoS is ambiguously related to network performance factors, one of the key functions of QoS evaluation module is to define that relation. Three primary network performance factors can be defined:

- available bit rate ρ and its fluctuations $\Delta\rho$;
- packet loss rate π;
- packet delays τ and delay variations $\Delta\tau$.

The definition of PQoS relation to these parameters depends not only on service types, but also on specific user requirements.

The concept of QoS evaluation traditionally refers to assigning particular score, much like mean opinion score in MOS concept. However, historically idea of MOS came to life for voice codec evaluation, thus having a reference in this scheme is a must. When introducing individual QoS ratings, the defined SLA may be considered as a reference. For this purpose composing SLA must include some QoS module training algorithm, which adds complexity, but introduces easily understandable interface for user.

3.2.1. Composing and Managing Service Level Agreement

There are two ways of defining SLA. We can use high-level parameters, which relatively defines required QoS grade in form of ρ, $\Delta\rho$, π, τ, $\Delta\tau$ parameters, or simplified score, similar to MOS.

This way the conversion to low-level network performance factors will be done at user device by PQoS tracking agent. This conversion has to use service descriptions and SLA requirements as well as measured low-level parameters. If we consider that services are known prior to subscription (in practice that is usually the case) and service behaviour does not change, it is more efficient to do the conversion at the SLA definition stage. The advantage of this method is that user devices usually have limited processing power and complexity of the algorithms should be reduced as much as possible.

If we make SLA to include only low-level network factor definitions for every service, the PQoS module will have the simple task of comparing those factors to actually measured at the user device and providing information about compliancy to user and network operator agreement. This way the complexity of composing SLA increases; however it may be required to do only once or every time SLA change is initiated.

Therefore, composing SLA must include several stages:

- defining low-level network performance factor impact on different services – creating service descriptors;
- defining network capabilities and efficient use of resources;
- providing intelligent interface for user upon subscription;

- performing “training” of PQoS module – converting user experience needs to low-level network performance factors.

Naturally, considering unpredictable nature of wireless propagation, operator may want to define distribution of parameter values rather than committing to keep the performance indicators under certain threshold.

3.2.2. Service Descriptors

The input of different network performance factors to quality of different services is different. Service descriptors should be understood as customizable service parameter set, used for linking perceived quality of the particular service to network performance parameters, measurable at user device. Such parameters as link latency may be directly unavailable to user device, because of lack of timing reference; however indirect indication, like round-trip time can be used.

In order to define the precise PQoS dependency on network performance parameters, service requirements and sensitivity on those parameters has to be evaluated.

There has been many analysis on real-time service performance, impacted by wireless link impairments – packet loss (Batkauskas 2006) and delay, also QoS evaluation on specific real-time applications, such as online games (Scaefer *et al.* 2002).

As for non-real-time data services, the influence can be simple for one type of services and complex for another.

Bulk data transfer can be considered as most obvious and clearly understandable data service type. The quality of these services has straight forward connection to network factors. Reachable average goodput being the only criterion for user satisfaction, it is directly influenced by available bit rate at the access network. However, knowing that most common implementation for these services is TCP, transport protocol related behaviour has to be considered.

TCP performance on wireless links has been analyzed in (Pavilanskas 2005). In practice, available bit rate cannot be directly mapped to achieved goodput of the application for the following reasons:

- TCP performance may be degraded due to rapid changes in available bit rate of the link
- TCP settings of user device may not be optimal for the link
- TCP receive, send and congestion window settings.

Web browsing is another common non-real-time service. It relies on HTTP which is transported by TCP.

The quality of this service may be expressed in variety of ways. However, the most objective criteria is web page download time. Guidelines for web browsing perceived quality have been introduced in ITU Rec. G.1030 Annex A (ITU-T Rec. G.1030 2005), presenting opinion models and MOS mapping to site opening times. As perceived quality depends on expectations and user experience, later replications of ITU's experiments gave different results (Ibarrola *et al.* 2009). Earlier experiments (Dellaert *et al.* 1999) showed even bigger differences in session times and user satisfaction due to lack of experience and much lower expectations of the users. Relation between expectation and perception can be understood as two-way connection: (Ibarrola *et al.* 2010) presents reversed view of (Anderson *et al.* 1993) expectation and perception relationship and considers that PQoS may affect the expectation and will eventually change the requirements.

Considering HTTP nature of requesting every object from server, not only available bit rate, but also link delay would have an influence on web page download time. Even though today's web pages usually contain contents form other servers, making parallel downloads of the objects, the principle still relies on request-response routine.

The presented example of web browsing service descriptor intends to show the link between measurable parameters to perceived parameters.

In order to specify the input significance from these two parameters to perceived quality of web browsing, testing of different delay and bit-rate links was performed. Basic test setup is presented in Fig. 3.5.

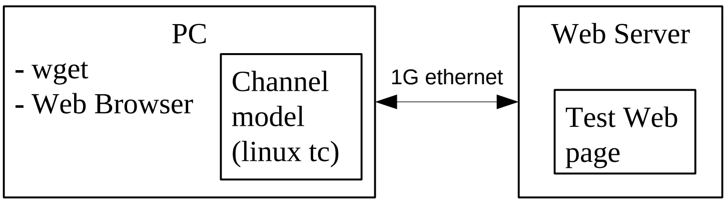

Fig. 3.5. Web service test system setup

The test was performed using Traffic Control (TC) (Kuznetsov 1999) capabilities under Linux OS. Range of presented link speeds and delays does not intend to represent different access technologies, but to show the significance of two factors to web browsing service performance. The title page of popular news portal (330 KB total size, 32 objects) was used as a test page. To eliminate any network or server load related impairments, local web server with copies of this page was used. Links simulated are symmetrical – same bit rate for uplink and downlink, the delay introduced is also symmetrical – RTT/2 for uplink and RTT/2 for downlink.

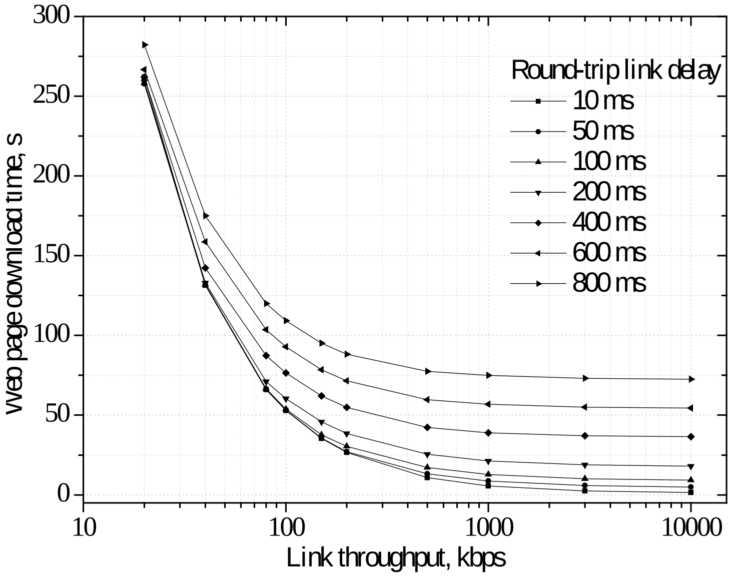

Fig. 3.6. Web page download time dependency on link throughput and delay

In Fig. 3.6 we can see that link throughput has significant influence only to some threshold. Beyond 200–500 kbps it does not make much improvement to page download time. Meanwhile, the influence of link delay slightly increases along with increasing link throughput. However, the influence of delay factor is far greater for link speed above 100 kbps, e. g. reducing delay by half from 200 ms to 100 ms RTT for 500 kbps link decreases the web page loading time by more than 8 s (or more than 30 %), and same effect would be reached by increasing link throughput 10 times.

Web browsing speed in low latency lines, such as most wired, cable or optical, also benefit from link throughput increase only to some point. Page download time drops by more than half (from 5.5 s to 2.5 s) going from 1 Mbps to 3 Mbps on 10 ms RTT link, however further improvement is less significant – 1.5, 1.2, 1.1 s for 10, 30, 100 Mbps links respectively.

Commonly used wireless access technologies introduce high delays, especially in uplink direction. More bandwidth for wireless means more radio recourses which are limited. However, wisely employing low-latency schedulers for such services as web browsing, it is possible to save bandwidth and improve service quality at the same time.

3.2.3. Modelling Wireless User Access Channel

As shown in Chapter 2, wireless user access link characteristics are influenced by position in the network and time. Both radio conditions and resource utilization ratio creates non-stationary qualitative changes in access link.

Implementing proposed SLA definition scheme requires subscriber to actually “try” services. This way the link between achieved and perceived QoS is kept. Since PQoS has many subjective components, immediate interface between subscriber and his services is a must in such system. To present real-time experience of different services we develop wireless channel model based on technological properties and actual network performance.

Traditional technological modelling of wireless access in isolated modelling environments, such as Opnet, Ns, NCTUns or other, may be suitable, but have some major disadvantages:

- Missing link to actual services and live users. Those systems can be used to evaluate performance factors and create metrics; however, the link to actual PQoS is missing.
- Complex and often inaccurate modelling of actual network implementations. Network simulation environments offer precise technological simulation and are useful for service behaviour testing. However, tuning actual network performance are implemented by modelling additional sources, users or node utilization simulation. This introduces complexity and is less flexible.
- Analysis has to include actual achieved QoS by operator, which is usually impaired by network planning, backhaul, access radio solutions and many other factors, which cannot be considered in detail while modelling.

Even though NCTUns has a different concept of using host PC TCP/IP stack (Wang *et al.* 2007) and allows interconnecting actual Linux based applications, it still introduces limitations of modelled network complexity and high hardware requirements for realistic emulation. However this tool is a good alternative for technology-specific and more accurate channel emulations, and may serve well to develop service adjustments to network technology or vice versa.

Much simpler solution can be applied to mimic actual network performance by „recording“ and later „playing back“ performance parameters. This way channel model is composed from actual experimental data, and can reflect different sorts of correlations: specific time, specific place or area, specific route through coverage zone.

The process of crating channel model is depicted in Fig. 3.7. The model, however, has simplifications, which may lead to behaviour inaccuracies for some applications and services. Therefore, experimental collection of network performance parameters has to be done in universal manner. For instance, actively measuring available bit rate ρ, measuring method has to be carefully selected – TCP can poorly reflect instantaneous channel bandwidth variations due to slow adaption, especially in scenarios with motion and handovers. Packet delay τ in some access technologies, such as GPRS, EDGE, 3G depends on packet size due to scheduling principles employed. All this service-specific behaviour has to be considered while obtaining measurement results.

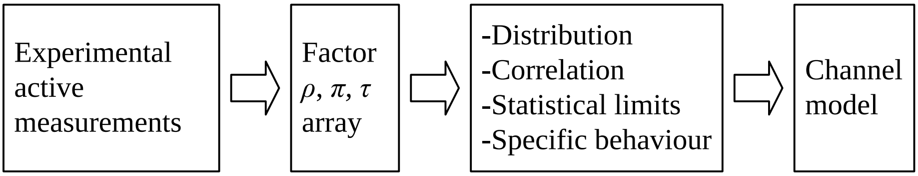

Fig. 3.7. Procedure of forming channel model

Test prototype of such system was implemented using available tools of Linux TC. Active network performance parameters were obtained from active measurements of WiMAX and 3G user access channels, using common user premise equipment.

3.3. Performance Evaluation by Active Experimental Measurements

The objective of first procedure on the way to creating relevant channel model is to collect network performance factor samples – instantaneous values of available bit rate *B*, packet loss *R* and delay τ.

Bandwidth measurements should rely on UDP traffic, to avoid transport protocol related inaccuracies. Bandwidth dependency on packet size is well known issue in WLANs (Šaltis 2004), however herein analyzed WiMAX/2G/3G networks do not have such excessive correlation. Higher layer overhead related dependency always exists, however it is not noticeable when comparing to scheduling and layer 2 acknowledging.

Experimental results of stationary continuous 24-hour run (working day) of available bit rate measurement in operational (commercial) WiMAX network is presented in Fig. 3.8. The long term fluctuations (white line represents throughput smoothed per minute basis) reflect base station utilization changes during the day.

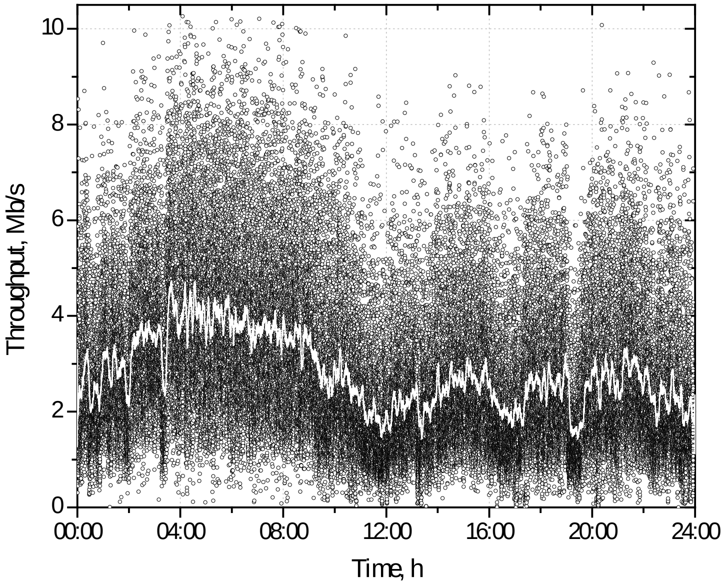

Fig. 3.8. Experimental results of available bit-rate measurement in commercial WiMAX radio access network

Real-time services, such as voice and video, require low jitter and latency. Most of the VoIP software bases their quality evaluation purely on jitter, calculated from timestamps in RTP headers. Additional buffers at receiving end are used to minimize jitter at cost of absolute latency, however absolute latency cannot be increased too much in interactive transmissions such as regular voice call or teleconference. ITU specifies (ITU-T Rec. G.114 2003) 150 ms as a threshold for mouth-to-ear delay in voice stream for complete user satisfaction. Commonly used VoIP client software can adjust buffers dynamically to optimize latency/jitter ratio.

Voice packets in IP networks have highly variable packet-interarrival intervals. Recommended practice is to count the number of packets that arrive late and create a ratio of these packets to the number of packets that are successfully processed. You can then use this ratio to adjust the jitter buffer to target a predetermined, allowable late-packet ratio. This adaptation of jitter buffer sizing is effective in compensating for delays (Davidson *et al.* 2006).

To model latency characteristics, round-trip evaluation is not informative enough for most real-time applications, especially when it comes to asymmetric (both bandwidth and delay) links.

One-way latency measurement is challenging, since time synchronization between source and sink is required. As mentioned, in most VoIP applications only jitter (based on RTP timestamps) and packet loss (based on RTP sequence number) are tracked, meanwhile end-to-end delay is technically never known, but observed subjectively by user in interactive conversation. In practice, jitter calculation techniques described in (Schulzrinne *et al* 2003) are widely used.

End-to-end latency includes not only transmission delay, but also packetization, encoding-decoding and other forwarding delays (processing at end devices).

End-to-end one way delay may be evaluated from RTP time stamps as long as source and sink clocks are in sync. There are several methods for syncing PCs: using NTP servers (The Network Time Protocol 2009) – public or dedicated, using common reference clocks. GPS clocking or pulse-per-second signal can be used for synchronizing as well, especially useful in distributed systems.

For stationary testing purposes, sending and receiving on interfaces at same PC seems as attractive option. This functionality can be easily implemented by patching Linux kernel. Patch (Anastasov 2003) implements routing of traffic between local IP addresses externally via multiple interfaces. This way traffic source and sink resides in same PC with perfect sync, since same system clock is used for packet depart and arrival time acquisition. The testing setup is presented in Fig. 3.9.

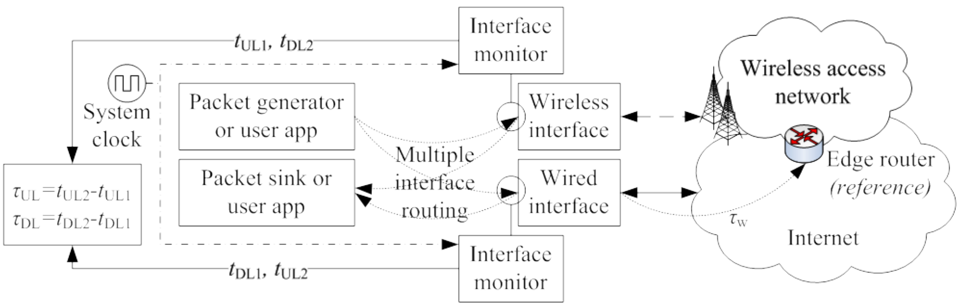

Fig. 3.9. One-way delay testing setup

System uses interface monitors, implemented with tcpdump (Tcpdump/libcap, 2010), to acquire packet departure time from wired interface (t_{DL1}) and arrival to wireless interface (t_{DL2}). Calculated difference τ_{DL} represents packet delay on downlink (from wireless interface perspective) direction; uplink direction end-to-end delay is calculated accordingly, using time values t_{UL1} and t_{UL2}.

Total delay of the packet, travelling from wired to wireless interface or vice versa, includes cumulative delay of network path to wireless link. Considering most likely service scenarios, the traffic from wireless operator will flow to service provider networks through operator peering points. Thus, eliminating all other network segments, except wireless link itself is not reasonable – additional delay, caused by BS backhauls, aggregation routers or congestions in wired part of wireless operator has an impact to service quality, therefore has to be taken into account. Also, precise topology of network is not known to network users.

For the abovementioned reasons, first known router near wireless link was considered reference point for delay measurements, i.e. delay, introduced between this router and wired interface in testing system (τ_w), was deducted from end-to-end delay measurement results. First known router can be easily discovered by tracing IP hops with tools like traceroute(Traceroute, 2010). Next, considering symmetrical delay in wired networks, τ_w is evaluated as half ping to router round-trip time. Naturally, setup has to be optimized to reach as small and stable τ_w value as possible.

System for testing routines was deployed in academic network, operated by LITNET, having stable delay and throughput to all wireless networks of interest. Minimum bandwidth provided to wired interface was 94 Mbps downlink and 45 Mbps uplink, and delay to reference router 1.9 ms mean with less than 0.1 ms standard deviation.

Delay dependency on packet size is observable in 2G/3G networks (Fabini *et al.* 2008) and must be considered for particular service testing.

Packet loss measurement can be implemented by means of tracking either identification field on IP header, by tracking RTP sequence number or by placing identification into payload. IP header tracking is most universal method due to independence from using services and overlaid protocols. Packet loss observation is done independently during latency testing. It is worth noticing that in wireless network access links with packet loss control, packet loss is usually seen as latency increase as lost packets or fragments are resent. However, delay intolerant applications like VoIP, consider delayed packet lost after given threshold. Following this rule, wireless schedulers can be adjusted accordingly. For instance, WiMAX service flow carrying voice packets can be applied with different HARQ (hybrid automatic resend request) settings, forcing to give up packet retransmission if predefined delivery threshold is not met.

3.4. Channel Model Implementation

Linux TC is used to configure Traffic Control in the Linux kernel, including shaping, scheduling, policing and dropping techniques.

To implement this particular channel modelling prototype, several disciplines were used:

- Queuing discipline qdisc (qdisc 2006) with classless token bucket filter for available bit rate modeling;
- Network emulator netem (Hemminger 2009), for introducing delay, delay variations and loss.

Netem provides capability to emulate delays based on predefined distributions, also to fix correlation value. Since measurement results do not fall into standard predefined distributions (available predefined distributions are: normal, pareto, pareto-normal), custom distributions have to be created. For this purpose *maketable* utility was used. *maketable* is a part of *iproute2* (The Linux foundation 2010) compilation.

End-to end delay can be approximated and modelled by following steps:

1. Calculating variable distribution from experimental data
2. Providing Distribution, Mean, Standard deviation and optionally Correlation coefficient to Netem.

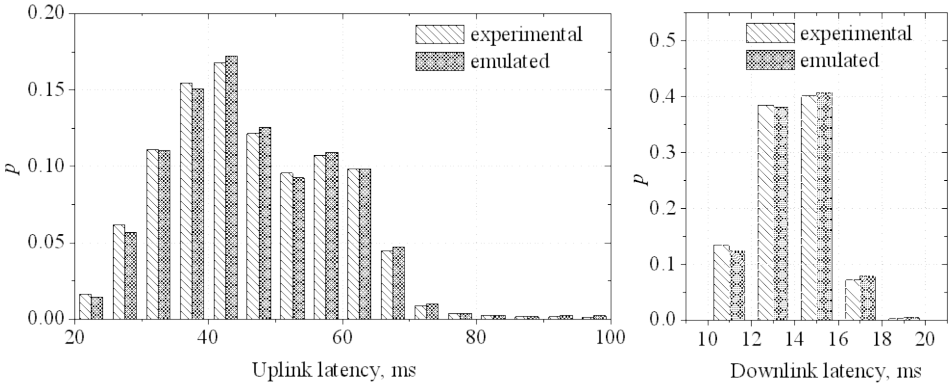

Fig. 3.10. Delay distribution of underutilized WiMAX BE link

Fig. 3.10 shows delay distribution of tested WiMAX single user link with best-effort scheduler. Showed experimental data were collected in user-wise and interference-wise isolated environment. While showed emulated results were obtained testing channel model, composed using *tc*. End-to-end delay samples were collected applying Fig. 3.9 **s**etup for WiMAX access network segment and later applying same scheme for emulated channel model. Artificial voice stream was used as packet source, simulating G.711 voice over IP packets, carrying 160 byte payloads (equal to 20 ms voice samples), followed by 12 byte RTP header and encapsulated with 8 byte UDP and 20 byte IP headers.

It can be seen (Table 2.1) that emulated channel reflects closely statistical properties of delay samples for WiMAX link.

Table 3.1. Statistical properties of collected delay distributions in WiMAX BE link

	Mean	Standard deviation	Minimum	Median	Maxi-mum
Experimental (uplink)	46.35	12.35	21.5	44.6	122
Emulated (uplink)	46.59	12.28	22.1	44.8	96
Experimental (downlink)	13.88	1.53	10.6	13.9	19.9
Emulated (downlink)	14.06	1.53	10.8	14.1	20.2

Packet loss rate emulation for this particular case is not needed as no packet loss occurred during experimental monitoring. This is natural considering isolated nature of the performed tests. In commercially deployed WiMAX networks packet drop rate depends on radio environment, handovers and retransmit policy. Current version of WiMAX employs HARQ (hybrid automatic resend request) algorithm, which can control lost packet rate individually for each service flow. Since we used best-effort service flows for all experiments with no strict packet arrival deadline, all packet losses caused delay increase on link layer and could not be seen for higher layers.

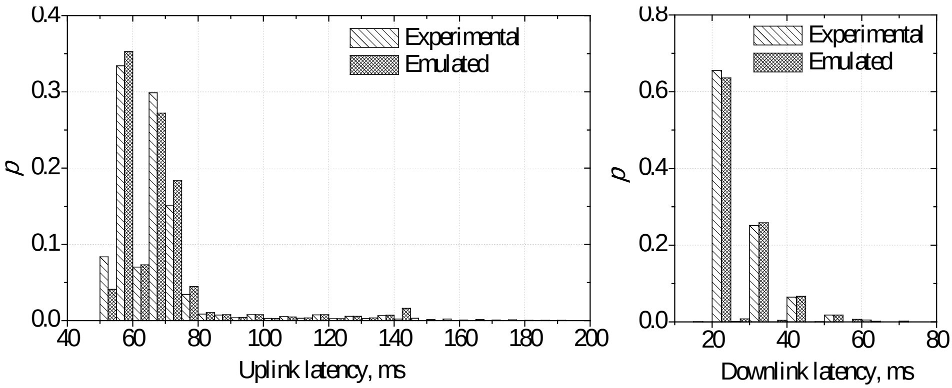

Fig. 3.11. Delay distribution of underutilized HSDPA link

However, if radio environment experience longer degradation, packet loss is inevitable, therefore usually appear in bursts. Experimental results of commercial WiMAX network in Vilnius showed packet loss in strong interference areas.

Same experiments were performed in 3G/HSDPA network. There was no possibility to run experimental measurements in isolated environment, therefore commercial network was used. The distribution of uplink delays is a little more complex, therefore emulated channel shows bigger error than in WiMAX case (Fig. 3.11), however, it sufficiently reflects the distribution and statistical properties.

3.5. User Interface and Requirement Collection

User input to SLA definition still remains difficult task. The link between user perception and network/service performance parameters is ambiguous due to different subjective understanding and different experience of the user. For this reason, it may not always be the best solution to rely on mean opinion as it hardly reflects the individual needs.

One of the ways to map user expectations and requirements to SLA is through presenting the service samples for user and obtaining his/her experience.

To simplify this task, operator may use predefined sets of network performance parameters and present them for service evaluation. Eventually this capability of dynamically defining SLA may become part of users self-provisioning of the services. User may select their required quality level for particular services at self-care portals and apply different requirements for short term or permanently.

The purpose of user requirement collection system may be considered as broader operator and technology independent comparison solution. It might be applied not only selecting certain level of service for single operator, but also evaluating how different access technologies fit user's needs. In emerging next generation access networks, so-called vertical handovers (handovers between different access technologies) play important role in always-best-connected concept. Therefore it is important to take a complex look at technological limits of various wireless access networks. For example, some user needs or services may require more advanced access technology than operator may offer.

The proposed architecture of PQoS module training through user interface is presented in Fig. 3.12.

User forms requirements by forming feedback to presented services using different network performance parameter presets. Presets are designed according to technological capabilities and minimum service requirements, also can be based on statistical (average user) data.

Network performance presets are used to configure user access channel model, which is technology dependent and reflects realistic network performance. The model must be service and therefore application independent. User through different services and applications reports his experience which then is mapped to network performance parameter sets.

User access channel model implementation may be done through specific modelling packages, such as NCTUns, or simplified to qdisc and netem routines using tc in linux PC. Channel model may be implemented on the same PC where user applications and services are run using virtual interfaces, or channel model may be implemented separately as communication link bridge.

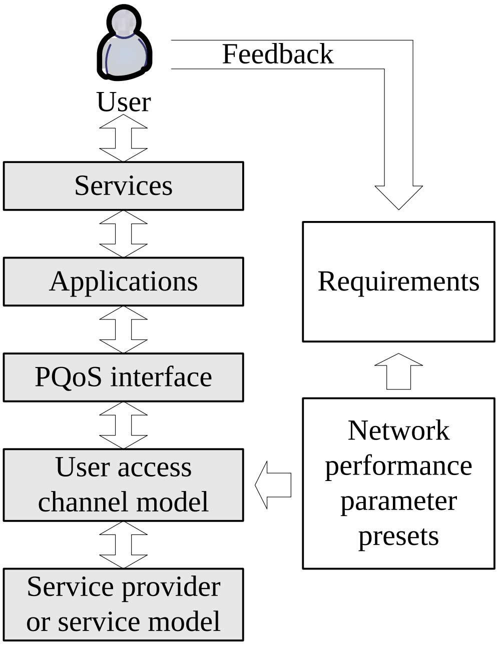

Fig. 3.12. Proposed PQoS training scheme using user access channel model

Known PQoS evaluation models require either averaging opinions, involving subjective evaluation by experts, which is expensive and inaccurate, or reference is required in case software is applied for evaluation. None of these methods can be applied for real-time PQoS evaluation at receiving end.

The idea of representing perceived quality with measurable network performance factors can be used to map personal evaluation or software based evaluation (such as PESQ).

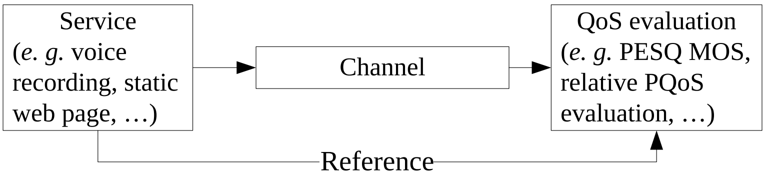

Fig. 3.13. Evaluating QoS for mapping to network performance factors

The mapping is based on comparison results of reference and distorted in channel service samples (Fig. 3.13). The advantage of this method is that mapping only has to be done once – prior to setting up QoS evaluation module in user device. This procedure can be called QoS module "training", hence the module is given individual ratings for particular objective parameter settings.

3.6. Conclusions of Chapter 3

- The traditional way of composing SLA is not convenient for average user, because it does not indicate actual service performance and requires user to have high level of expertise.
- Proposed SLA composition method is defined by user indirectly by QoS training module, which uses user input through designed interface, service descriptors and network capabilities information.
- Proposed SLA enforcement system to track, indicate and account actually received QoS is implemented by passively collecting network performance parameters.
- Proposed structure of perceived quality monitor implementation may consist of three parts: test tools for network impairments, quality analysis tool and quality monitor.
- PQoS evaluation can rely on three primary network performance factors: available bit rate B and its fluctuations ΔB; packet loss rate R; packet delays τ and delay variations $\Delta\tau$.
- PQoS evaluation has to be service specific, therefore must include unique service descriptors. Analysis of web browsing service quality dependency on available throughput and link delay shows that page opening time is exponentially dependant on both throughput and delay.
- Individual monitors of perceived QoS will create the possibility for the user to obtain objective information and indication whether operator delivers sufficient level of QoS under SLA commitment.

4

Application of Perceived Quality Tracking to Web Browsing

To make sure that operator follows SLA provisions, QoS level has to be tracked continuously. This may be done by software QoS agents, residing in user devices and monitoring real-time quality of online services.

This chapter presents development and testing of passive non-intrusive agent, residing in user device for obtaining channel parameters, which can be used for evaluation of perceived quality of web browsing service. Experimental data shows that quality perceived by user depends on available bit rate and delay of the physical access link. The mechanism for defining QoS requirements and rel-time tracking is based on perceived QoS monitoring concepts, developed in chapter 3. Passive method for user perceived quality estimation allows evaluating if actual quality of service achieved by network meets the quality declared by operator, and allows discovering if it is degraded.

With passive monitoring tool the measurement data can be processed in real time and presented to perceived quality analysis module, which, using information about access network capabilities and user's service level agreement with operator, can monitor, indicate quality impairments and account quality grade.

Results presented in this chapter have been published in (Kajackas *et al.* 2010a).

4.1. Evaluating Quality of Web Browsing Service

Many telecommunication services, including Web browsing, currently are provided to users over cellular networks. It is natural to mobile networks that throughput and quality varies depending on user location and time. Therefore, the time to locate and download a Web page with a browser application is also variable together with quality of service perceived by user. Users seek for trusted connections to be sure they can rely upon connectivity service for high importance Web operations such as online banking and shopping.

However, customer premise equipment cannot provide capability to keep track of objective data for evaluating perceived quality of service (Kajackas *et al.* 2005), (Guršnys 2008). It is justifiable, since the precise estimation of online service quality is a problem difficult to solve – a service has many interdependent quality attributes influenced by several contextual factors.

Web browsing traffic makes 50–70 % of all HTTP traffic in today's internet networks (Li *et al.* 2008). Web browsing quality is related to user access technology characteristics.

General issues on Web browsing quality have been discussed in many works (Li *et al.* 2008; Gunduz *et al.* 2003; ITU-T Rec. G.1030 2005). The Web browsing model is bidirectional and differentiates between user requests and responses. Single Web session can be divided into several phases. Browsing process starts when user requests a Web page, whereas server responds with a data packet burst. User needs reading time between Web page loadings; therefore in single browsing session there can be many alternating download and idle periods. Research and simple Web browsing experience indicates that content value and page opening time are two things of major importance to the user.

Web page download delay depends on the user's estimate of the size and page processing time. Graphical elements make pages to open slowly, which might be acceptable on an index page, viewed for the first time, but not on navigational pages. If a user is interested in the content of a page, he will likely spend more time there compared to other pages in his session. The time spent on a page is a good measure of the user's interest in that page, providing an implicit rating for it (Gunduz *et al.* 2003). An important observation in perceived Web browsing quality modelling is the fact, that the *expected* maximal session time will dominate the perceived quality (ITU-T Rec. G.1030 2005).

Perceived Web browsing quality evaluation often is based on Opinion Model for Web-browsing Application, proposed by ITU (ITU-T Rec. G.1030 2005). This model suggests relating user perceived quality under MOS (*Mean Opinion Score*) scale with service processing time – weighted session time T_{sw}, by applying regression line (4.1). Rules of T_{sw} calculation are described in (ITU-T Rec. G.1030 2005).

$$MOS = A - H \cdot \ln T_{sw}. \tag{4.1}$$

This ITU methodology of Web browsing quality evaluation is applicable directly to Web page and specialized quality evaluation tool development, when special dedicated Web pages are used for browsing quality evaluation. In reality, every Webpage is represented by different data volume and thus page download times are different. Therefore, measured session time is not a suitable characteristic for user perceived quality evaluation, because session time depends not only on link conditions, but also on particular Web page characteristics. This way, methodology described in (ITU-T Rec. G.1030 2005) cannot be applied directly to Web browsing quality evaluation, because *expected session time* is unknown. However, this methodology is a clear guideline for developing other methods.

Most important factors, influencing session time, is size of HTML object, server processing and transfer delays. Session time noticeably depends on transfer link throughput. This time is also stretched by bottlenecks – low throughput links in the transmission path.

3G and 3.5G technologies have become very popular recently. Real throughput of such access technologies reaches several Mb/s. High throughput determines good Web browsing quality, however in some places 3G network is not available and user is switched to GPRS/EDGE network. This way throughput offered by cellular network degrades and becomes bottleneck of the transmission link.

This paper proposes a method to evaluate if user perceived Web browsing quality fits the network characteristics, declared by operator. If perceived quality is impaired, the degree of impairment is determined.

The throughput of access link can be expressed as:

$$B = B_{dec} - \Delta B\,, \tag{4.2}$$

where B_{dec} is nominal declared throughput of access link, ΔB – throughput degradation. Thus the Web session time can be expressed as:

$$T_{sw} = \left(T_0 + \frac{S_s}{B_{dec} - \Delta B} \right), \tag{4.3}$$

where T_0 is the sum of all waiting (processing) times, S_s – data volume transferred in given session. By joining (4.1) and (4.3) we get Web browsing process MOS evaluation:

$$MOS = A - H \cdot \ln\left(T_0 + \frac{S_s}{B_{dec} - \Delta B}\right), \tag{4.4}$$

which may be approximated by:

$$MOS = MOS_0 - \delta \cdot \frac{\Delta B}{B_{dec}}. \tag{4.5}$$

In formula (4.10) MOS_0 represents MOS score, when observed access link throughput is equal to declared one:

$$MOS_0 = H \cdot \ln\left(T_0 + \frac{S_s}{B_{dec}}\right). \tag{4.6}$$

δ is proportional coefficient, estimating throughput degradation influence on quality of service, and κ (4.7) refers to impairment factor, caused by throughput degradation, κ is inversely proportional to ΔB:

$$\kappa = \delta \cdot \frac{\Delta B}{B_{dec}}. \tag{4.7}$$

It is noticeable, that MOS expression (4.5) is similar to E-model (ITU-T Rec. G.107 2000) expression for voice quality estimation.

Several research works quality of service ties to throughput. Paper (Fiedler *et al.* 2003) proposes the use of throughput as QoS indicator perceived by a video conferencing application in the presence of a bottleneck. Another publication (Joskowicz *et al.* 2009) presents expression for video clip, coded in MPEG-2, evaluation by heuristic equation $DMOS = m/(a\ bitrate)^n$, where DMOS (Difference MOS) is the quality metric, with values between 0 and 1, bitrate is expressed in Mb/s , for Low Movement $n = 1.125$ and $m = 0.21$, $a = 3.57$.

Most common link throughput methods intend to measure average throughput and works actively – by presenting special server, where user sends and receives known data blocks. Throughput then is calculated dividing transferred data volume by transmission time. Unfortunately, active measurement is not suitable for user perceived quality evaluation.

While searching for passive access link throughput measurement methodology, it is worth noticing that this task is similar to bottleneck bandwidth measuring problem (Benko *et al.* 2004; Ricciato *et al.* 2007). The bottleneck bandwidth sets the upper limit on how quickly the network can deliver the sent data to the receiver. Thus, available bandwidth never exceeds bottleneck bandwidth.

4.2. Passive Web Browsing Monitor

Web page retrieval, simply called Web browsing, is based on mutual user-server communication. User generates requests for particular pages by entering domain names or clicking on links. Server responds with data, containing complete Web page including all prerequisites it may have. What is seen as a single Web page in fact is complex set of objects – script files, images, inline text, style sheets, etc.

Today's internet browsing is based on HTTP protocol version 1.1, which uses single TCP session to communicate between user and Web server. Therefore, the effectiveness of HTTP is fully dependant on TCP/IP stack performance. Unlike HTTP version 1.0, version 1.1 can reuse TCP connection to transfer Web objects, therefore TCP handshake and slow start has to be experienced only once, no matter how many objects are held in particular Web page. This routine is presented in Fig. 4.1.

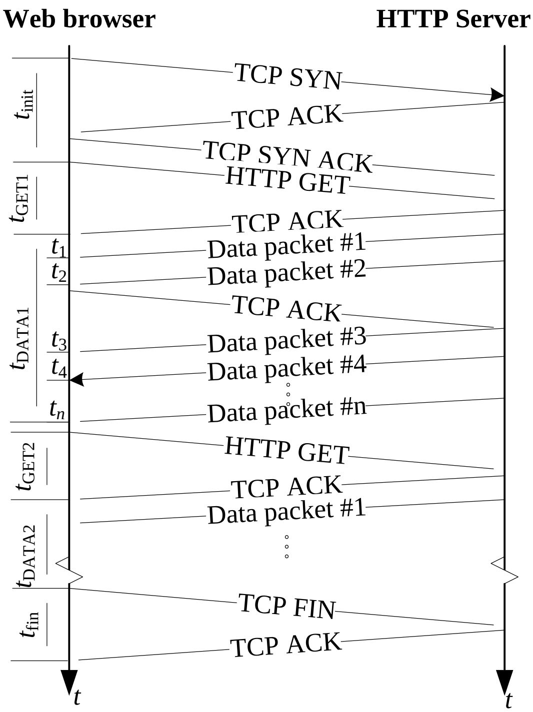

Fig. 4.1. Packet-level view of Web page retrieval

Additional time is taken by TCP connection establishment (t_{init}) and release (t_{fin}), however this is done only once at the page opening beginning and lasts as long as you browse on same server.

HTTP connection setup starts with TCP SYN packet, initiated by user. It is the beginning of three-way TCP hand-shake (TCP SYN, TCP ACK, TCP SYN ACK). After successful connection, user sends page request, which is called HTTP GET and usually is small enough to fit in one packet. Server acknowledges the request with TCP ACK, this way concluding page request procedure, which is measured as t_{GET}. Next, server starts sending Web page with required objects to show on user's Web browser. Object downloads follow standard TCP acknowledgement routine. Duration of the phase, where data of single object is being downloaded, is measured as $t_{\text{DATA}n}$. In HTTP 1.1 next object download can be carried out on the same TCP session, therefore HTTP GET message can be sent right away for next Web page object. TCP connection is released with two-way user initiated handshake: TCP FIN, TCP ACK messages.

TCP data delivery mechanism is important to consider, because it allows adapting and dynamically utilizing all available channel bit rate. However the adaptation requires time, therefore bit rate fluctuations can be observed and available channel resources underutilized at some periods of slow start and congestion avoidance.

We have collected a large array of object sizes by grabbing random Web-page objects from random internet Web pages (Fig. 4.2).

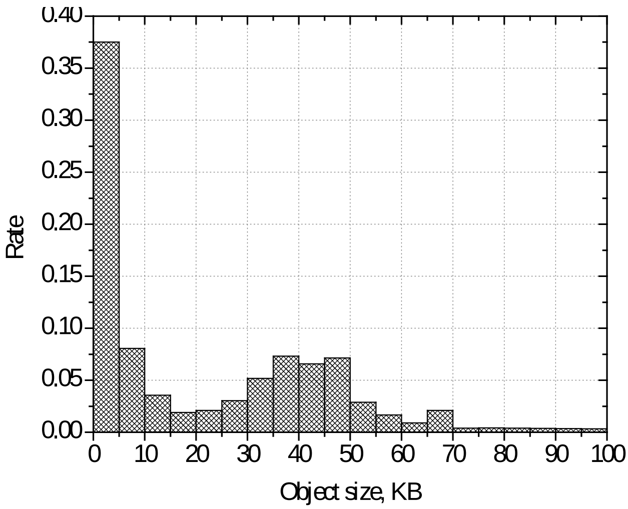

Fig. 4.2 Distribution of object sizes in random internet Web pages

It can be seen, that over one third of Web page objects have size under 5 KB, and second most commonly distributed range is from 30 to 50 KB. We can see, that most common object sizes and relatively small.

The size of t_{GET} will be influenced by physical channel bandwidth and channel delay. The higher packet delay is introduced in transmission channel, the higher t_{GET} value will be observed, therefore meaning that user will have to wait more time for object to be downloaded. This time will directly influence satisfaction, thus the perceived quality as well. Meanwhile, t_{DATA} represents the time, which is used to actually download the object. This value is also dependent on available channel bit rate, but again, the delay introduced in channel will have high impact if object size is as small as several TCP packets.

4.3. Experimental Test-bed and Results

We developed a Web browsing tracker tool to evaluate how radio access network performance influences Web browsing quality. It acts as passive agent on user device (Fig. 4.3), which logs HTTP GET response time t_{GET} and Web page object transfer times t_{DATA}. To do so, agent tracks every packet leaving and entering network interface and looks for HTTP headers (Fig. 4.4). Depending on type of transaction (request or response), separate timers are used to measure t_{GET} and t_{DATA}. An effective bandwidth, experienced at that particular time, is calculated afterwards. Object retrieval speed can be calculated easily, since Web page object sizes are known. However, few objects can be downloaded at same time (if content is distributed among different hosts), for this reason agent additionally keeps track of total interface bandwidth, allowing to measure total access link bandwidth.

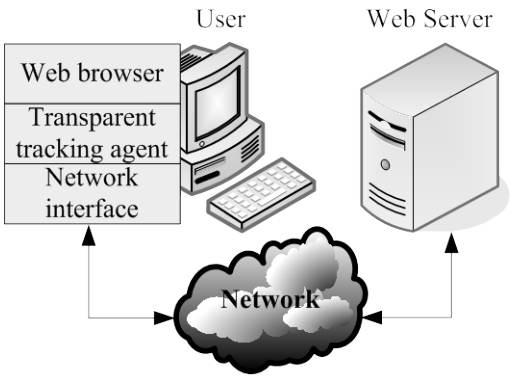

Fig. 4.3. Web page browsing quality estimation experimental setup

The tool is transparent to any user application software; therefore user may use any Web browser or Web page retrieval method. Operation of the tool is based on jnetpcap (Jnetpcap 2010) libpcap (Tcpdump/libcap 2010) libraries, which allow reading packet headers and retrieving the contents. The tool has been written on java programming language and tested under Linux OS.

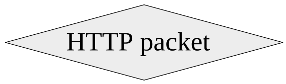

Fig. 4.4. Simplified algorithm of perceived quality tracking agent

Experimental measurements with the prototype of tool were performed by downloading popular news portal *lrytas.lt* index page with inline objects – images, scripts, etc. To eliminate server load influence, make testing environment fully controllable and to focus on access link evaluation, the dedicated server was used with precise original Web page copies.

To automate Web page browsing process and make long testing sessions possible (24-hour, week or longer runs), additional software, based on *wget* (wget 2009) was developed. This software simulates Web browsing by requesting to download predefined Web pages as a regular Web browser would do. The tool additionally continuously logs time, received signal strength indicator (RSSI) and coordinates for later data analysis. This allows data to be analyzed based on different correlations: throughput over time, place or RSSI.

Several sets of experimental results were collected. All measurements were made on real GPRS access link using usb-dongle modem in stationary environment with satisfactory network conditions: RSSI between -75 and -85 dBm. The modem was forced to work in GPRS-only mode. GPRS channel have been chosen as high latency low bandwidth channel example, to show the impact of both bandwidth and delay. From logged t_{GET} and t_{DATA} values we can see the instantaneous object retrieval time values over period of time. However, those values are not informative, since object sizes are different. To make the measurements object size independent, we calculate achieved throughput:

$$B_n = \frac{S_n}{t_{GET_n} + t_{DATA_n}}, \tag{4.8}$$

where B_n is the achieved throughput of n-th object, S_n – n-th object size, t_{GETn} + t_{DATAn} – n-th object retrieval time, measuring from the moment of request departure to moment of last object data packet arrival.

Instantaneous object download throughput values show great fluctuations over time. In Fig. 4.5 we can observe great range of achieved object throughputs, scattered from less than 10 kbps to almost 70 kbps. Even average of single Web page objects (average is shown in Fig. 4.6 as continuous line) show short-term and long-term fluctuations. Long-term fluctuations represent alternating network load: throughput decreases due to higher access network utilization at peak times – roughly from 10 a.m. to 22 p.m.; at this time maximum achievable throughput also decreases. Short-term fluctuations usually are caused by changing radio environment. Even though the presented results have been taken in stationary conditions, quality of the radio link varies due to fading effects, interference or weather conditions, as well as non-stationary nature of channel resource provision.

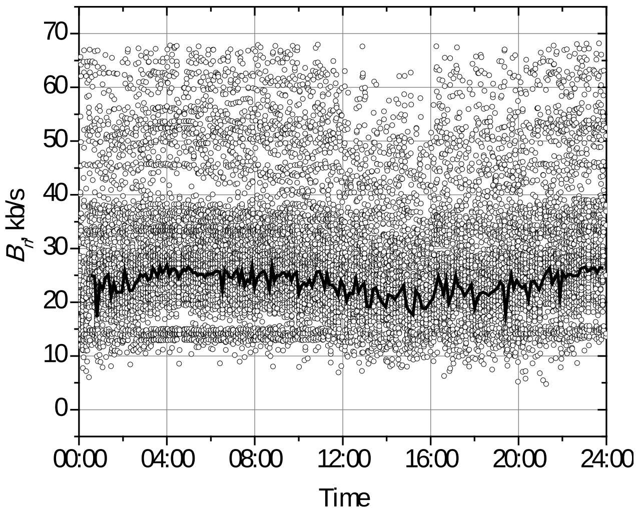

Fig. 4.5. Instantaneous object download throughput B_n during 24-hour measurement period

The wide range of achievable object download throughputs is also related to http transmission specifics. Analysis of separate object achieved throughputs shows that intensity of fluctuations is more evident on small objects.

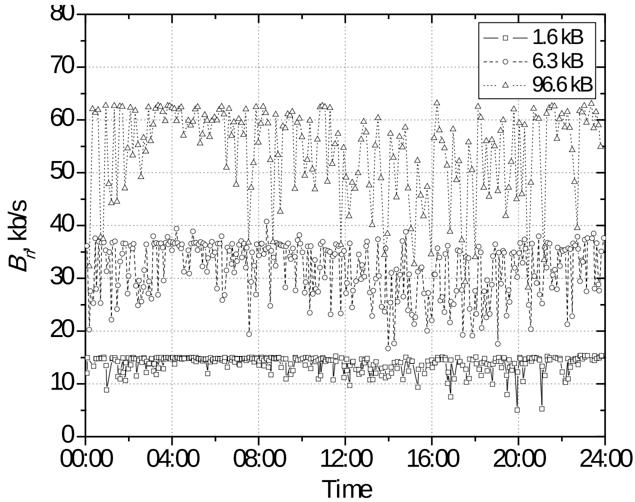

Fig. 4.6. Instantaneous different size object download throughput B_n

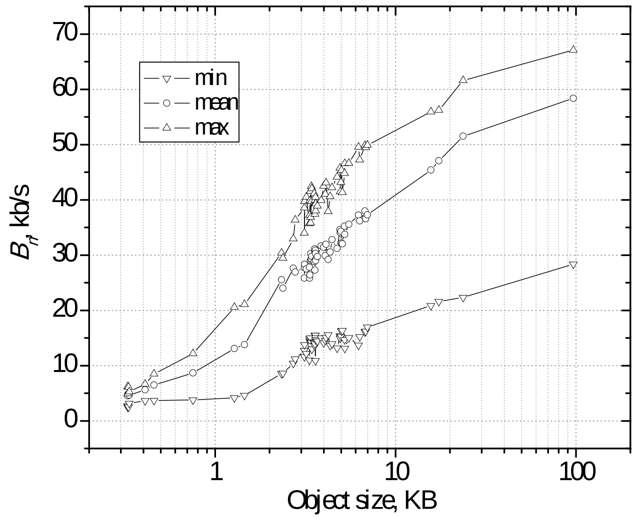

Fig. 4.7. Download throughput B_n dependency on object size

Fig. 4.6 presents instantaneous download speeds of 1.6 kB, 6.3 kB and 96.6 kB objects. It can be seen that smaller objects achieve significantly lower throughput, however they are much more stable even at period of higher link utilization rate.

If we look at minimum, average and maximum achievable throughputs for objects of different sizes in Fig. 4.7, we can see clear tendency of higher throughputs for bigger objects. The statistics has been collected from measured Website *lrytas.lt* title page, which contained over 60 different size objects.

The reason again lies in HTTP transfer principle. Every object is requested by GET command. Even though GET carries relatively small amount of data, it has to reach the server before object data download starts. Considering GET as fixed time overhead to any object download time period, the short objects suffer from relatively smaller achieved throughput.

4.4. Analysis of Available Bit Rate

Maximum achievable throughput or average throughput of the object cannot be used as browsing quality evaluation, because different size objects behave differently even in channel with the same available bandwidth. Therefore we propose to use lower-level statistics for available bit rate evaluation to mitigate result dependency on Web page design characteristics.

Our developed monitoring agent is capable of logging every packet size S_n (to exclude overhead, we take TCP payload into account only), packet arrival times t_1, t_2, … t_n and inter-packet time $\Delta t_n = t_{n+1}$-t_n (Fig. 4.1). Available bit rate at given moment of n-th packet transmission will be calculated as follows:

$$B_n = \frac{S_n}{\Delta t_n}. \tag{4.9}$$

Since transport of the HTTP traffic is implemented on TCP, Δt_n and therefore B_n values will vary a lot when packet pair includes waiting for ACK as congestion control related delay.

Max and average B_n values can be calculated per page object or whole Web page. The major disadvantage to perform calculations on whole Web page basis is that short-time network impairments may be integrated over long page opening time. On other hand, more precise evaluations than per-page score will not be needed for user, however it may integrate short term impairments and eventually will lead to inaccuracy of quality evaluation of analysis software. Calculations, based on object basis is not reasonable due to risk unequally handle different size objects – again, smaller objects will have lower probability of reaching available channel bit rate.

Fig. 4.8 presents the cumulative distribution function of all B_n measurements of complete *index.html* Web page including all Web page objects. Comparing values, obtained at peak times (higher access network utilization), and off-peak times (underutilized access network), we can see different distribution of available bit rates. Peak time measurements have approx. 45 kbps mean and 95 kbps max value, while off-peak measurement show over 57 kbps mean and over 102 kbps max value.

Note, that maximum observed bit rate values are higher than TCP goodput, reachable in GPRS link. Inter-packet time observation based bit rate measurement represents throughput per-packet basis, which naturally reaches higher values. Measured average packet-pair bit rate values are comparable with those observed with net TCP throughput.

Therefore used methodology allows evaluating both short-term and long-term available bit rate in wireless user access link. Short-term evaluation allows ignoring TCP rate control and focusing on access network related impairments. This way it will be possible to avoid unwanted correlation between Web page structure and perceived quality of service evaluation.

Binding perceived quality to object achieved bit rate has a major drawback not only in terms of sensitivity to web page design, but also in designing individual ratings.

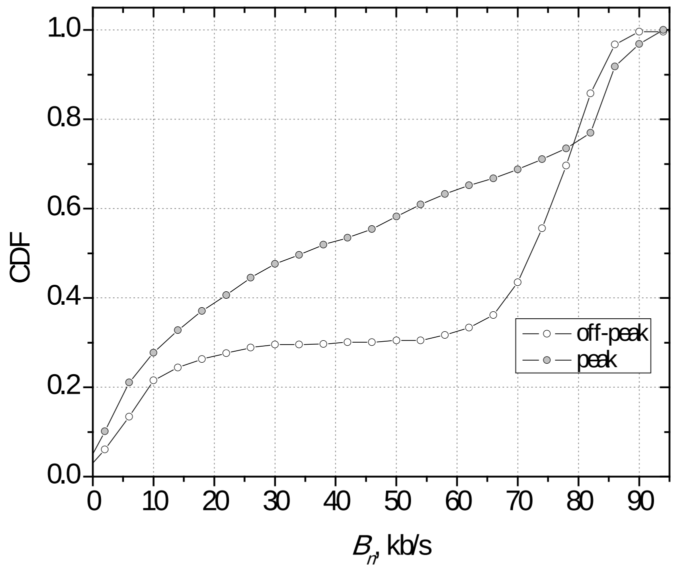

Fig. 4.8. Cumulative distribution function of measured available bit rate

Following the guidelines for designing quality monitor, defined in Chapter 1 and Chapter 3, measured factors have to be easily obtainable at user device and quality degradation easily detected without extensive calculations. Furthermore, it has to be individually designed in terms of requirements for network or service provider.

Naturally, page opening time is the parameter perceived by user in web browsing. ITU recommends mapping MOS to session time. However, it is not wanted to consider full web page download as session time, because users usually do not wait for full web page to download, but rather start reading the loaded and displayed top of the page, clicking heading links, or navigating to other pages if the content does not fit their current needs.

Previous user web browsing behaviour studies (Andrews *et al.* 2006) show that user perceived quality may correlate poorly with website opening time, because once information is displayed user is interested in, the rest of the loading process is either unnoticed or not cared about. This is considered whereas designing the large news portal pages – the full page may take a while to load, but user gets instantaneous experience if initial viewing area is displayed rapidly even if it composes 10% of a whole page.

For this reason, perceived quality evaluation has to be done in shorter intervals and then aggregated. Page consists of independent web objects, which can be evaluated independently and then aggregated. Simple averaging can be applied for aggregating individual object evaluation, since most of the web pages retrieve objects in serial manner. This way several delayed objects are allowed to be integrated over total averaging period, which is equal to threshold of noticeability. Clearly, the threshold has to be higher than one object retrieval time and less than whole page download time.

Considering that tolerable delay (thus perceived quality as well) highly depends on task and interest of the user (Dellaert *et al.* 1999; DiClemente *et al.* 2003; Otto *et al.* 2000), the threshold of noticeability can be related to uninterrupted experience threshold, which is equal to 1 s (ITU-T Rec. G.1030 2005). Further experiments with actual users are required to predict the threshold, however empirically derived value will have the same problem of poorly reflecting individual QoS needs. For example, a key finding of the research (Ibarrola *et al.* 2010) in web browsing case was also that even though the relations between objective QoS (delay) and perceived quality were strong, the overall satisfaction of a user was closely linked to other contextual parameters, like the user's previous experiences or his/her expectations.

Considering web page opening time as sum of object download times allows focusing on shorter time intervals:

$$T_p = \sum_{i=1}^{n} t_i \,, \tag{4.10}$$

where T_p is web page opening time, t_i is object download time of i-th object and n – number of objects in web page. In turn, object download time consists of object request and object download delays. First time component is dependant only on network performance factor τ, whereas latter one is dependant on object size and available bit rate ρ:

$$t_i(\rho_i, \tau_i, S_i) = t_{GET}(\tau_i) + \frac{S_i}{\rho_i} \,. \tag{4.11}$$

Using (4.11) for setting acceptable quality threshold and quality monitoring, allows equalizing ρ and τ input to perceived quality of web browsing. Handling ρ and τ independently would be incorrect, because one parameter can compensate for another. The relation is dependant on object size S.

In scope of single technology, ρ and τ are linked via π, which reflects packet loss in wireless link. Reliable transport protocols such as TCP will lower transmission bit rate due to loss or random delay, which in turn is caused by loss in PHY and L2 retransmits.

Finding relation between ρ and τ would allow binding them and using as impairment factor pair. However, considering heterogeneous network environment, ρ and τ may change independently in case of handover from one wireless access technology to another.

Uplink bandwidth is not considered, since bottleneck in uplink direction is fully reflected by increased t_{GET} time. Note, that t_{GET} is influenced by web server utilization as well.

The algorithm was implemented in user device for web browsing tracking and as in previous test, same static webpage replica was used in controlled (server utilization wise) environment.

Logged t_{GET}, t_{DATA} and S_i values are used to evaluate object retrieval threshold using (4.16) and are aggregated using different time periods.

Measurements have been taken in stationary position with satisfactory radio conditions, under influence of changing network utilization. The tool has been validated for two marginal access technologies – low throughput, high latency GPRS and high throughput, moderate latency WiMAX.

For the first test let's assume the user have tested and agreed to minimal service level defined as:

$$T_{\max}(S) = 0.8 + \frac{S}{40 \cdot 10^3} \,. \tag{4.12}$$

Next, the tool is logging and monitoring service quality in terms of meeting defines threshold. The tool have been monitoring web browsing service for 24 hours, calculating T_i and aggregating values. The CDF of object download time difference from T_{max} (4.12) aggregated for different time intervals is shown in Fig. 4.9. It can be seen, that longer aggregation durations have the similar effect to averaging – the impairment may be hidden by adjacent values.

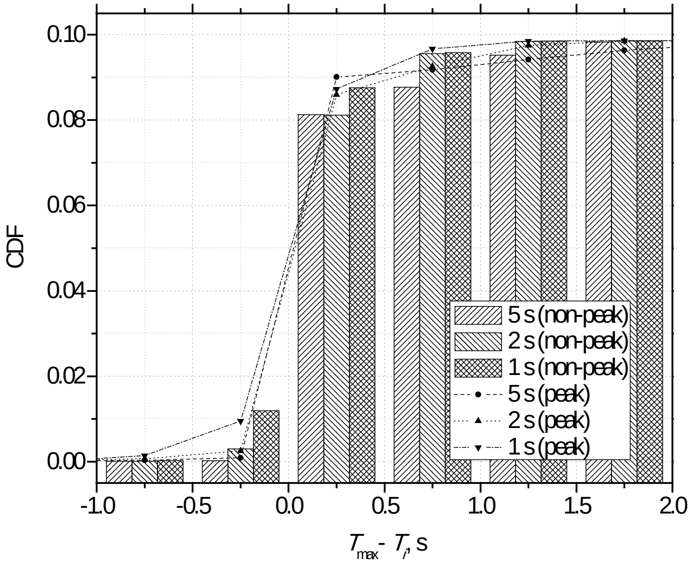

Fig. 4.9. CDF of object download time difference from required, aggregated for different time intervals

It can be seen, that perceived quality degradation can vary in wide range depending on noticeability period. Therefore, this period has to be defined together with objective parameters in “training” phase when user is presented service samples.

Analysis of peak and non-peak periods of network utilization is expected to show differences in T_{max} compliancy. This can be seen observing percentage of T_i values, which failed to meet T_{max} (Fig. 4.10).

Fig. 4.12 shows measured object download duration (T) values at peak and non-peak network utilization periods and regression lines (R) – polynomials, fitted to measurement results. Again, averaging over all measurement period shows little difference between R_{peak} and $R_{non\text{-}peak}$ (note, that difference for smaller packets is more noticeable), both being smaller than T_{max}, thus meeting the requirements. However, individual monitoring of T_i compliancy detects periods of degraded quality, when required threshold is not met.

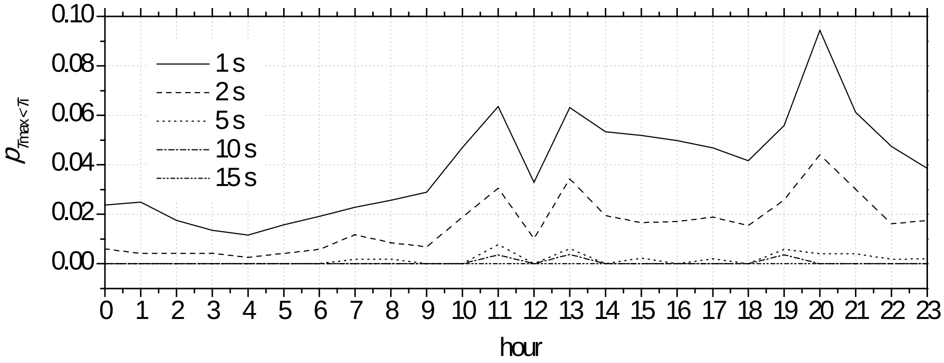

Fig. 4.10. Rate of degraded quality ($T_{\max}$<T_i) values during test in low speed link

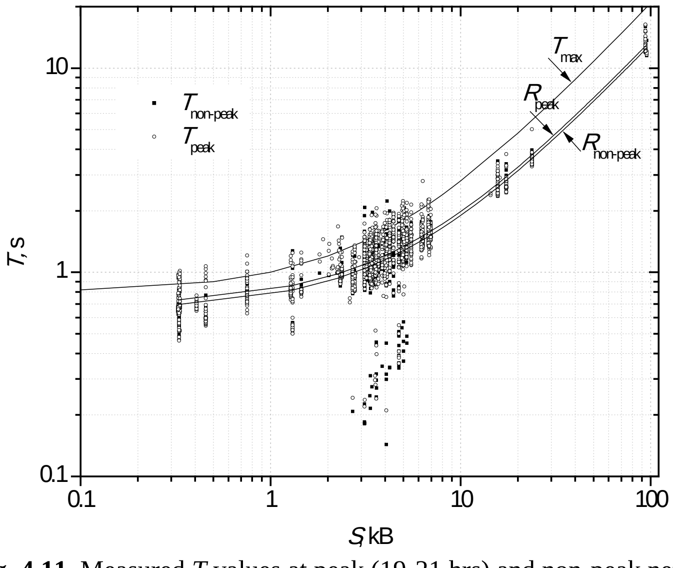

Fig. 4.11. Measured T values at peak (19-21 hrs) and non-peak network utilization periods and regression lines in low speed link

Second test was performed in high-speed moderate latency network, defining a threshold for $T_{\max}$ as:

$$T_{\max}(S) = 0.2 + \frac{S}{4 \cdot 10^6}. \quad (4.13)$$

The CDF of observed $T_{\max}$-T_i differences (Fig.4.13) are less dynamic, however, the observer impairment dependency on noticeability period (1, 2 or 5 second) can be seen in ratio of T_i values which failed to meet the threshold ($T_{\max}$-T_i <0).

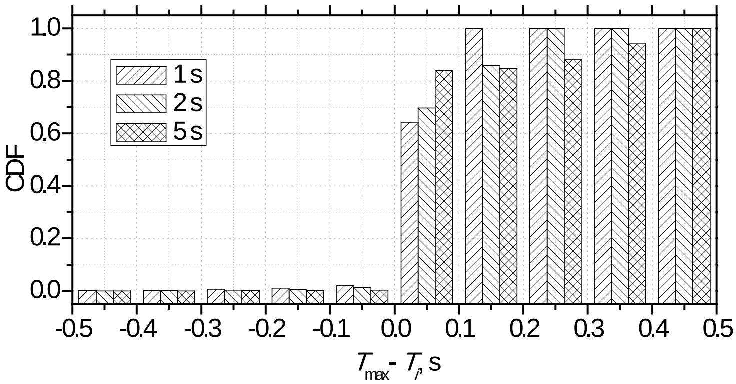

Fig. 4.12. CDF of object download time difference from required, aggregated for different time intervals

To implement web browsing PQoS monitoring mechanism to full QoS enforcement system, individual quality level in terms of web browsing quality has to be defined following proposed SLA definition (Fig. 3.12) scheme. User is presented with web service trial using channel model using network performance parameter presets. Acceptable parameter settings are considered as quality threshold. Parameters ρ and τ are recorded and in user device are used not directly, but rather via (4.11) relation, which allows one parameter compensate for another and maintain the level of perceived quality, but not strict level of objective parameters. Furthermore, using web page object size as variable allows dissociate measurements from web page properties and focus on achieved performance of network.

With passive monitoring tool the measurement data can be processed in real time and presented to perceived quality analysis module, which, using information about access network capabilities and user's service level agreement with operator, can monitor, indicate quality impairments and account quality grade.

Today's Internet Web pages are complex object structures, composed of different objects and their behaviour may differ from site to site even with same access network conditions. Therefore it is proposed to evaluate available bit-rate of the radio access channel by tracking packet sizes and calculating time period between received packet pairs. This way available channel bit rate is evaluated objectively and can be used for quality of service in access network indication for user. Even though perceived quality of Web browsing is proportional to average bit rate of the channel, additional research is needed to define the exact input of channel bandwidth to perceived Web browsing quality. This research shall include analysis of bit rate evaluation dynamics and perceived quality classification by assigning quality scores.

Future work will include perceived quality indication algorithms, which base their decisions on objective passive real-time measurements at user device.

4.5. Conclusions of Chapter 4

- Efficient and user-friendly implementation of passive perceived quality monitor can rely on passive non-intrusive software agent, residing in user device.
- Developed software agent, obtaining channel available bit-rate and GET-ACK delay can be used for evaluation of perceived quality of Web browsing service.
- The experiments with web pages show that aggregated web page object retrieval time can be used as perceived quality indicator of web browsing. Web page download time is not suitable for there is no normalization on expected opening time and is page structure dependant.
- Using web page object size as variable allows dissociate measurements from web page properties (structure, object size distribution) and focus on achieved performance of network.
- Passive PQoS monitoring tool provides indication, relative to requested quality of service, defined prior to service monitoring, therefore is individual.

General Conclusions

1. Heterogeneous wireless networks show highly variable unpredictable fluctuations of wireless channel resources. Fluctuations are caused by changing radio environment and base station utilization. Experiments show, that physical parameters measured at user device do not directly reflect actual received quality of service.
2. Quality of service enforcement in wireless access networks requires additional technological resources, which can be managed by selecting efficient access scheduling. Simulations of voice over wireless local access networks show that number of serviced voice channels can be increased by 90 % by applying additional technological measures – aggregation, acknowledge piggybacking and modified admission control. However, the number of serviced voice channels is still 24 % smaller than in legacy WLAN case.
3. Using multihop transmission in ah-hoc wireless networks, packet delay characteristics are controlled by optimizing wireless channel resource assignment to reliability and throughput of the transmission. Simulation results show, that keeping channel utilization under 2/3, packet transmission delay over 100 IEEE 802.11 nodes is as low as 0.6 s.
4. Quantitative network performance characteristics have complex impact on perception of quality of service. Currently missing link between user

and operator has to be defined by designing individual quality of service evaluation system, capable of detecting impairments during service delivery.

5. Experimental research of developed passive monitoring system application to quality evaluation of web browsing service shows, that software agents in user device can be used for evaluation of perceived quality of web browsing service by obtaining channel available bit-rate and delay between HTTP GET and ACK messages.

References

3GPP TS 46.001 V8.0.0. 2008. Technical Specification Group Services and System Aspects, Full rate speech, Processing functions.

Ahmed, T. *et al.* 2007. End-to-end quality of service provisioning through an integrated management system for multimedia content delivery, *Computer Communications* 30(3): 638–651.

Anastasov, J. 2003. Send-To-Self interface flag. [Online] 2003. [Cited: 02 15, 2010.] http://www.ssi.bg/~ja/send-to-self.txt.

Anderson, E. W.; Sullivan, M. W. 1993. The Antecedents and Consequences of Customer Satisfaction for Firms, *Marketing Science* 12: 125–43.

Andrews, M.; Cao, J.; McGowan, J. 2006. Measuring Human Satisfaction in Data Networks, in *Proceedings on INFOCOM 2006 25th IEEE International Conference on Computer Communications*. Barcelona, 1–12.

ANSI S3.2-1989 (R1999). 1999. Method for Measuring the Intelligibility of Speech over Communications System.

ANSI/IEEE Std 802.11. 1999. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.

Anskaitis, A. 2009. *Koduoto balso kokybės tyrimas* [Analysis of quality of coded voice signals]: Doctoral Dissertation. Vilnius Gediminas Technical University. Vilnius: Technika.

Banchs, A. *et al.* 2001. Service differentiation extensions for elastic and real-time traffic in 802.11 wireless LAN, in *Proc. IEEE Workshop High Performance Switching and Routing*. Dallas, 245-249.

Batkauskas, V. 2006. *Investigation of evaluation and improvement methods of service quality in mobile networks*: Doctoral Dissertation. Vilnius Gediminas Technical University. Vilnius: Technika.

Benko, P.; Malicsko, G.; Veres, A. 2004. A Large-scale, Passive Analysis of End-to-End TCP Per-formance over GPRS, in *Proceedings of IEEE Infocom'2004*, vol. 3. Hong Kong, 1883–1892.

Bianchi, G. 2000. Performance Analysis of the IEEE 802.11 Distributed Coordination Function, *IEEE Selected Areas in Communications* 18(3): 535–547.

Bilstrup, K. *et al.* 2009. On the Ability of the 802.11p MAC Method and STDMA to Support Real-Time Vehicle-to-Vehicle Communication, *EURASIP Journal on Wireless Communications and Networking* 2009: 1–13.

Blake, S. *et al.* 1998. An Architecture for Differentiated Services. RFC 2475.

Boucadair, M. *et al.* 2007. A Framework for End-to-End Service Differentiation: Network Planes and Parallel Internets, *IEEE Communication Magazine* 45(9): 134–143.

Bouch, A. *et al.* 2000. Of Packets and People: A User-Centered Approach to Quality of Service, in *Proceedings of 8th IEEE IWQoS*. Pittsburgh, PA, USA, 189–197.

Braden, R,; Clark, D.; Shenker, S. 1994. Integrated Services in the Internet Architecture: an Overview. RFC 1633.

Budnikas, A; Jankūnienė, R. 2005. The analysis of QoS assurance in the inter-domain services, in *Proceedings of the 27th international conference on Information Technology Interfaces*. Cavtat, Croatia, 533–536.

Cavender, A. *et al.* 2008. MobileASL: Intelligibility of sign language video over mobile phones, *Disability and Rehabilitation: Assistive Technology* 3(1): 93–105.

Cenerario, N.; Delot, T.; Ilarri, S. 2008. Dissemination of information in intervehicle ad hoc networks, in *IEEE Intelligent Vehicles Symposium*. Eindhoven, 736—768.

Chandra, P. *et al.* 2009. *Wireless security*. Oxford: Elsevier.

Chen, D. *et al.* 2002. Supporting VBR VoIP traffic in IEEE 802.11 WLAN in PCF mode, in *Proceedings of OPNETWork*. Washington, DC, USA, 1–6.

Choi, S. *et al.* 2003. IEEE 802.11e contention-based channel access (EDCF) performance evaluation, in *IEEE International Conference on Communications*. Anchorage, Alaska, 1151–1156.

Communications Regulatory Authority. 2005. *Elektroninių ryšių paslaugų teikimo taisyklės*. 12 23, 2005.

Dajiang, H.; Shen, C. Q. 2003. Simulation study of IEEE 802.11e EDCF, in *57th IEEE Semiannual Vehicular Technology Conference*. Orlando, 685–689.

Davidson, J. *et al.* 2006. *Voice over IP Fundamentals*. 2nd Edition. Indianapolis: Cisco Press.

Dellaert, B.; Kahn, B. E. 1999. How Tolerable is Delay? Consumers' Evaluations of Internet Web Sites after Waiting, *Journal of interactive marketing* 13(1): 41–54.

Dekeris, B.; Adomkus, T.; Budnikas, A. 2006a. Analysis of QoS Assurance Using Weighted Fair Queueing (WFQ) Scheduling Discipline with Low Latency Queue (LLQ), in *Proceedings of the 28th International Conference on Information Technology Interfaces*. Zagreb, 507–512.

Dekeris, B.; Adomkus, T.; Budnikas, A. 2006b. Assurance of video conference services with combination of weighted fair queuing scheduling discipline and low latency queue, in *Proceedings of the 5th international symposium Communication Systems, Networks and Digital Signal Processing (CSNDSP)*. Patras, Greece, 884–887.

Dekeris, B.; Narbutaitė, L.; Adomkus, T. 2007. A New Adaptive Fair Queueing (AFQ) Scheduler for Support SLA, in *29th International Conference on Information Technology Interfaces*. Cavtat, 597–602.

DiClemente, D.; Hantula, D. 2003. Optimal foraging online: Increasing sensitivity to delay, *Psychology & Marketing* 20(9): 785–809.

Fabini, J.; Reichl, P.; Poropatich, A. 2008. Measurement-Based Modeling of NGN Access Networks from an Application Perspective, in *Proceedings of the 14th GI/ITG Conference on Measurement, Modeling, and Evaluation of Communication Systems*. Dortmund, Germany, 45–59.

Ferre, P. *et al.* 2004. Throughput Analysis of IEEE 802.11 and IEEE 802.11e MAC, *Wireless Communications and Networking Conference WCNC2004*, vol. 2. Atlanta, 783–788.

Fiedler, M. *et al.* 2003. Identification of performance degradation in IP networks using throughput statistics, in *Proceedings of the 18th International Teletraffic Congress.* Berlin, 399–407.

Galetzka, M. *et al.* 2004. User-Perceived Quality of Service in Hybrid Broadcast and Telecommunication Networks, in *Proc. 5th Workshop Digital Broadcasting*. Erlangen, 39–44.

Gunduz, S.; Tamer, O. M. 2003. A Poisson model for user accesses to Web pages, *Lecture Notes in Computer Science* 2869: 332–339.

Guršnys, D. 2008. *Balso kokybės vertinimo metodai ir priemonės mobiliojo ryšio sistemoms* [Voice quality evaluation methods and means for mobile communications]: Doctoral dissertation. Vilnius Gediminas Technical University. Vilnius: Technika.

Hall, T.A. 2001. Objective Speech Quality Measures for Internet Telephony, in *Proc. of SPIE Voice over IP (VoIP) Technology*. Denver: 128–136.

Hemminger, S. 2009. Netem. The Linux Foundation. [Online] 2009. [Cited: 01 04 2010.] http://www.linuxfoundation.org/collaborate/workgroups/networking/netem.

Ibarrola, E. *et al.* 2009. Web QoE Evaluation in Multi-Agent Networks: Validation of ITU-T G.1030, in *Fifth International Conference on Autonomic and Autonomous Systems*. Valencia: 2009.

Ibarrola, E. *et al.* 2010. Quality of Service Management for ISPs: A Model and Implementation Methodology Based on the ITU-T Recommendation E.802 Framework. *IEEE Communications Magazine* 48(2): 146–153.

IEEE Std 802.11e-2005. 2005. Medium Access Control (MAC) Quality of Service Enhancements.

ITU Rec. P.59. 1993. Artificial conversational speech.

ITU Study Group 12. 2000. France Telecom Study of the relationship between instantaneous and overall subjective speech quality for time-varying quality speech sequences: influence of a recency effect. Contribution D.139.

ITU-T Rec. E.800. 2008. Definitions of terms related to quality of service.

ITU-T Rec. E.802. 2007. Framework and Methodologies for the Determination and Application of QoS Parameters.

ITU-T Rec. E.860. 2002. Framework of a service level agreement.

ITU-T Rec. G.1000. 2001. Communications quality of service: A framework and definitions.

ITU-T Rec. G.1030. 2005. Estimating end-to-end performance in IP networks for data applications. Annex A. Opinion model for Web-browsing applications.

ITU-T Rec. G.107. 2000. The E-model, a computational model for use in transmission planning.

ITU-T Rec. G.114. 2003. International telephone connections and circuits – General Recommendations on the transmission quality for an entire international telephone connection. One-way transmission time.

ITU-T Rec. J.247. 2008. Objective perceptual multimedia video quality measurement in the presence of a full reference.

ITU-T Rec. P.800. 1996. Methods for subjective determination of transmission quality.

ITU-T Rec. P.862. 2001. Perceptual evaluation of speech quality (PESQ).

ITU-T Rec. P.910. 1999. Subjective video quality assessment methods for multimedia applications.

ITU-T Rec. P.911. 1998. Subjective audiovisual quality assessment methods for multimedia applications.

Jankūnienė, R.; Gvergždys, J.; Budnikas, A. 2007. Monitoring the Quality of Heterogeneous MAN Performance, *Electronics and Electrical Engineering* 2(74): 69–74.

Jankūnienė, R.; Slanys, R.; Budnikas, A.; Gudonavičius, L. 2005. Analysis of Voice over IP Networks Service Quality Parameters Information, *Electronics and Electrical Engineering* 7(63): 16–21.

Jeonggyun, Y.; Sunghyun, C. 2006. Comparison of modified dual queue and EDCA for VoIP over IEEE 802.11 WLAN, *European Transactions on* Telecommunications 17(8): 371–382.

Jerbi, M.; Senouci, S.M. 2008. Characterizing Multi-Hop Communication in Vehicular Networks, in *IEEE Wireless Communications and Net-working Conference*. Las Vegas, 3309-3313.

Jnetpcap. 2010. jNetPcap OpenSource A Libpcap/WinPcap Wrapper. [Online] 2010. [Cited: 15 02 2010.] http://www.jnetpcap.org.

Joskowicz, J. *et al.* 2009. A Mathematical Model for Evaluating the Perceptual Quality of Video, *Lecture Notes In Computer Science* 5630: 1883–1893.

Jumisko-Pyykkö, S.; Kumar, V.; Korhonen, J. 2006. Unacceptability of Instantaneous Errors in Mobile Television: From Annoying Audio to Video, in *ACM International Conference Proceeding Series*.Vol 159. Helsinki, 1–8.

Jumisko-Pyykkö, S.; Vadakital, V. K.; Hannuksela, M. M. 2008. Acceptance Threshold: A Bidimensional Research Method for User-Oriented Quality Evaluation Studies, *International Journal of Digital Multimedia Broadcasting* 2008: 1–20.

Kajackas, A.; Anskaitis, A. 2009a. An Investigation of the Perceptual Value of Voice Frames, *Informatica* 20(4): 487–498.

Kajackas, A.; Anskaitis, A.; Guršnys, D. 2005. Individual Quality of Service concept in Next Generations Telecommunications Networks, *Electronics and Electrical Engineering* 4(60): 11–16.

Kajackas, A.; Anskaitis, A.; Guršnys, D. 2009b. Estimating Individual QoS, *Traffic and QoS Management in Wireless Multimedia Networks: COST290 Final Report: Lecture Notes in Electrical Engineering*. 31: 180–183.

Kajackas, A.; Pavilanskas, L. 2006. Analysis of the Technological Expenditures of Common WLAN Models, *Electronics and Electrical Engineering* 8(72): 19–24.

Kajackas, A.; Pavilanskas, L. 2007. Analysis of the Connection Level Technological Expenditures of Common WLAN Models, *Electronics and Electrical Engineering* 2(74): 63–68.

Karlsson, A. *et al.* 1999. Radio link parameter based speech quality index-SQI in *Proc. IEEE Workshop on Speech Coding*. Porvoo, 147–149.

Kaul, S. *et al.* 2008. GeoMAC: Geo-backoff based co-operative MAC for V2V networks, in *IEEE International Conference – Vehicular Electronics and Safety*. Columbus, USA, 334–339.

Khalaf, R.; Rubin, I. 2006. Throughput and Delay Analysis in Single Hop and Multihop IEEE 802.11 Networks. in *3rd International Conference on Broadband Communications, Networks and Systems*. San Jose, CA, USA, 1–9.

Kuznetsov, A. N. 1999. Linux Traffix Control. Linux Traffix Control. [Online] 1999. [Cited: 21 01 2010.] http://linux.die.net/man/8/tc.

Li, W.; Moore, A. W. 2008. Classifying HTTP Traffic in the New Age. in *SIGCOMM'08* poster. Seattle.

Liberal, F. *et al.* 2005. Application of a PQoS Based Quality Management Model to Identify Relative Importance of the Agents, in *Proc. IEEE ICICS*. Bangkok, 239–243.

Ling, X. *et al.* 2008. Voice Capacity Analysis of WLANs with Channel Access Prioritizing Mechanisms, *IEEE Communications Magazine* 46(1): 82–89.

Medepalli, K. *et al.* 2004. Voice Capacity of IEEE 802.11b, 802.11a and 802.11g Wireless LANs, in *Proceedings of IEEE Glob2004*, vol. 3. Dallas, 1549–1553.

Mohamed, S.; Rubino, G. 2002. A Study of Real-Time Packet Video Quality Using Random Neuronal Networks, *IEEE Trans. on Circuits and Systems for Video Technology* 12(12): 1071–1083.

Mohamed, S.; Rubino G.; Varela, M. 2004. Performance evaluation of real-time speech through a packet network: a Random Neural Networks-based approach, *Performance Evaluation* 57: 141–162.

Narbutaitė, L.; Dekeris, B. 2008. Triple Play services packet scheduling performance evaluation, *Electronics and Electrical Engineering* 6(86): 85–88.

Opnet Technologies. 2007. Opnet Modeler. [Online] 2007. [Cited: 20 04 2007.] http://www.opnet.com.

Otto, J.; Najdawi, M.; Caron, K. 2000. Web-user satisfaction: An exploratory study, *Journal of Organizational and End User Computing* 12(4): 3–11.

Pastrana, R. *et al.* 2004. Sporadic Frame Dropping Impact on Quality Perception, *Human Vision and Electronic Imaging* 5292: 182–193.

Pastrana, R.; Gicquel, J.; Colomes, C.; Hocine. C. 2004. Sporadic Signal Loss Impact on Auditory Quality Perception. Measurement of Speech and Audio Quality in Networks. On-line Workshop. [Online] 2004. [Cited: 10 24, 2009.] http://wireless.feld.cvut.cz/mesaqin2004/contributions.html.

Pavilanskas, L. 2007. *Adaptation of Wireless Access MAC Protocol for Real Time Packet Flows*: Doctoral dissertation. Vilnius Gediminas Technical University. Vilnius: Technika.

Pavilanskas, L. 2005. TCP modeling in wireless LAN, *Electronics and Electrical Engineering* 5(61): 78–83.

qdisc. 2006. Classless Queuing Disciplines. Traffic Control HOWTO. [Online] 2006. [Cited: 04 01, 2010.] http://linux-ip.net/articles/Traffic-Control-HOWTO/classless-qdiscs.html.

Ricciato, F.; Vacirca, F.; Svoboda, P. 2007. Diagnosis of Capacity Bottlenecks via Passive Monitoring in 3G Networks: an Empirical Analysis, *Computer Networks: The International Journal of Computer and Telecommunications Networking* 51(4): 1205–1231.

Remeika, I.; Činčikas, G. 2007. The Analysis of Network Resources for Real–time Services. *Electronics and Electrical Engineering* 1(73): 55–58.

Roy, J.; Vaidehi, V.; Srikanth, S. 2007. A QoS Weight Based Multimedia Uplink Scheduler for IEEE 802.11e WLAN, in *ICSCN '07 International Conference on In Signal Processing, Communications and Networking*. Chennai, 446–451.

Rubino, G. 2005. Quantifying the quality of audio and video transmissions over the Internet: the PSQA approach, *Communication Networks and Computer systems* 2005: 235–250.

Rubino, G.; Varela, M.; Bonnin, J. 2006. Controlling Multimedia QoS in the Future Home Network Using the PSQA Metric, *The Computer Journal* 49(2): 137–155.

Scaefer, C. *et al.* 2002. Subjective Quality for Multiplayer Real-Time Games, in *Proc. First Workshop on Network and System Support for Games*. Braunschweig, 74–78.

Schulzrinne, H. *et al.* 2003. RTP: A Transport Protocol for Real-Time Applications. RFC3550.

Sutinen, T.; Ojala, T. 2005. Case Study in Assessing Subjective QoS of a Mobile Multimedia Web Service in a Real Multi-access Network, *Lecture Notes in Computer Science* 3552: 298–312.

Šaltis, A. 2004. *Radijo technologijos vartotojų prieigų tinkluose* [Radio Technologies in Customer's Access Networks]: Doctoral dissertation. Vilnius Gediminas Technical University. Vilnius: Technika.

Tcpdump/libcap. 2010. Tcpdump/libcap public repository. [Online] 2010. [Cited: 01 04 2010.] http://www.tcpdump.org/.

The Linux foundation. 2010. iproute2. [Online] 2010. [Cited: 04 01, 2010.] http://www.linuxfoundation.org/collaborate/workgroups/networking/iproute2.

The Network Time Protocol. NTP: The Network Time Protocol. The Network Time Protocol. [Online] [Cited: 10 09 2009.] http://www.ntp.org/.

Traceroute. 2010. Traceroute public repository. [Online] 2010. [Cited: 03 10, 2010.] http://sourceforge.net/projects/traceroute/.

Veeraraghavan, M.; Cocker, N.; Moors, T. 2001. Support of voice services in IEEE 802.11 wireless LANs, in *Proceedings of IEEE INFOCOM'01*. Anchorage, AK , USA, 488–497.

Wang, S. Y.; Chou, C. L.; Lin, C. C. 2007. The Design and Implementation of the NCTUns Network Simulation Engine, *Simulation Modelling Practice and Theory* 15: 57–81.

Wang, S.Y.; Lin, C.C. 2008. NCTUns 5.0: A Network Simulator for IEEE 802.11(p) and 1609 Wireless Vehicular Network Researches, in *Vehicular Technology Conference*. Calgary, BC, 1–2.

Wget. 2009. GNU WGET. [Online] 2009. [Cited: 28 09 2009.] http://www.gnu.org/software/wget.

WiMAX Forum. [Online] [Cited: January 10, 2010.] http://www.wimaxforum.org.

Xiao, J.; Boutaba, R. 2007. Assessing Network Service Profitability: Modeling From Market Science Perspective, *IEEE/ACM Trans.* 2007 15: 1307–1320.

List of Publications by the Author on the Topic of the Dissertation

Papers in the Reviewed Scientific Journals

Vindašius, A. 2006. Security State of Wireless Networks, *Electronics and Electrical Engineering* 7(71): 19–22. ISSN 1392-1215 (Thomson ISI Web of Science).

Kajackas, A.; Pavilanskas, L.; Vindašius, A. 2007. Synchronous Voice Applied Customer Access Based on IEEE 802.11, *Electronics and Electrical Engineering* 8(80): 23–28. ISSN 1392-1215 (Thomson ISI Web of Science).

Kajackas, A.; Vindašius, A. 2009a. Applying IEEE 802.11e for Real-Time Services, *Electronics and Electrical Engineering* 1(89): 73–78. ISSN 1392-1215 (Thomson ISI Web of Science).

Kajackas, A.; Vindašius, A.; Stanaitis, Š. 2009b. Inter-Vehicle Communication: Emergency Message Delay Distributions, *Electronics and Electrical Engineering* 8(96): 33–38. ISSN 1392-1215 (Thomson ISI Web of Science).

Kajackas, A.; Šaltis, A.; Vindašius., A. 2010a. User Access Impact on Web Browsing Quality, *Electronics and Electrical Engineering* 4(100): 59–64. ISSN 1392-1215 (Thomson ISI Web of Science).

Vindašius, A. 2010. Tinklų modeliavimas ir emuliavimas NCTUns aplinkoje, *Mokslas – Lietuvos ateitis* 2(1): 73–76. ISSN 2029-2341 (Index Copernicus).

Other Papers

Kajackas, A.; Vindašius, A. 2010b. Analysis and Monitoring of End-user Perceived QoS in Mobile Networks, in *14th International Telecommunications Network Strategy and Planning Symposium*. Warsaw, 25–28. ISBN 978-1-4244-6703-7.

Vindašius, A.; Stanaitis, Š. 2010. Analysis of Emergency Message Transmission Delays in Vehicular Wireless Mesh Network, in *Third International Conference on Advances in Mesh Networks*. Venice, 35–40. ISBN 978-0-7695-4092-4.

Antanas VINDAŠIUS

ANALYSIS OF QUALITY OF SERVICE IN HETEROGENEOUS WIRELESS NETWORKS

Doctoral Dissertation

Technological Sciences,
Electrical and Electronic Engineering (01T)

Antanas VINDAŠIUS

PASLAUGŲ KOKYBĖS HETEROGENINIUOSE BEVIELIUOSE TINKLUOSE TYRIMAI

Daktaro disertacija

Technologijos mokslai,
elektros ir elektronikos inžinerija (01T)

2010 10 12. 10,5 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“,
Saulėtekio al. 11, 10223 Vilnius,
http://leidykla.vgtu.lt
Spausdino UAB Ciklonas,
J. Jasinskio g. 15, Vilnius